ARTICLE

doi:10.1038/nature11251

Non-invasive prenatal measurement of the fetal genome

H. Christina Fan[1]†*, Wei Gu[1]*, Jianbin Wang[1], Yair J. Blumenfeld[2], Yasser Y. El-Sayed[2] & Stephen R. Quake[1,3,4]

The vast majority of prenatal genetic testing requires invasive sampling. However, this poses a risk to the fetus, so one must make a decision that weighs the desire for genetic information against the risk of an adverse outcome due to hazards of the testing process. These issues are not required to be coupled, and it would be desirable to discover genetic information about the fetus without incurring a health risk. Here we demonstrate that it is possible to non-invasively sequence the entire prenatal genome. Our results show that molecular counting of parental haplotypes in maternal plasma by shotgun sequencing of maternal plasma DNA allows the inherited fetal genome to be deciphered non-invasively. We also applied the counting principle directly to each allele in the fetal exome by performing exome capture on maternal plasma DNA before shotgun sequencing. This approach enables non-invasive exome screening of clinically relevant and deleterious alleles that were paternally inherited or had arisen as *de novo* germline mutations, and complements the haplotype counting approach to provide a comprehensive view of the fetal genome. Non-invasive determination of the fetal genome may ultimately facilitate the diagnosis of all inherited and *de novo* genetic disease.

Our work is based on the phenomenon of circulating cell-free DNA, whose existence and role in pregnancy was first investigated in 1948[1]. A portion of the cell-free DNA in a pregnant woman's blood is derived from the fetus[2], and this fact has enabled the development of a number of non-invasive prenatal diagnostic techniques[3]. A prominent example is the non-invasive detection of Down syndrome and other aneuploidies, which was first demonstrated by our group[4], validated by clinical trials[5–10], and is now available in the clinic. We describe here how the chromosome counting principle we invented for aneuploidy detection can be applied to non-invasive fetal genome analysis by directly counting haplotypes and even individual alleles. Others have studied the relationship between maternal and fetal cell-free DNA[11], but their approach required invasively sampled fetal material, did not determine the fetal genome, and also needed knowledge of paternal genetic data.

Measuring the fetal genome by counting parental haplotypes

Maternal plasma DNA is a mixture of maternal and fetal DNA; the fraction of fetal DNA ranges from a few percent or lower early in pregnancy to as high as ~50%[2,7], and generally increases with gestational age. Because the fetal genome is a combination of the four parental chromosomes, or haplotypes, as a result of random assortment and recombination during meiosis, three haplotypes exist in maternal plasma per genomic region: the maternal haplotype that is transmitted to the fetus, the maternal haplotype that is not transmitted, and the paternal haplotype that is transmitted. If the relative copy number of the untransmitted maternal haplotype is $1-\varepsilon$, where ε is the fetal DNA fraction, then the relative copy number of the transmitted maternal haplotype is 1, and the relative copy numbers of the transmitted and untransmitted paternal haplotypes are ε and 0, respectively (Fig. 1). Therefore, within each pair of parental haplotypes, the transmitted haplotype is over-represented relative to the untransmitted one. By measuring the relative amount of parental haplotypes through counting the number of alleles specific to each parental haplotype (referred to as 'markers'), one can deduce the inheritance of each parental haplotype and hence build the full inherited fetal genome.

Strictly speaking, the markers that define each maternal haplotype are the alleles that are present in one maternal haplotype but not in the other maternal haplotype and the two paternal haplotypes. However, because it is rare that two unrelated persons share the same long-range haplotype, that is, a haplotype much longer than the usual length of haplotype blocks observed in the population (~100 kilobases (kb)), the presence of alleles contributed by the transmitted paternal haplotype at these loci would not interfere with the measurement of representation of maternal haplotypes as long as the haplotype being considered is sufficiently long (>1 megabase (Mb)). Thus all the maternal heterozygous loci can be used to define the two maternal haplotypes (Fig. 1). This enables the measurement of relative representation of the two maternal haplotypes without the knowledge of paternal haplotypes. The relative representation of the two maternal haplotypes is the difference in the counts of markers specific to each haplotype. Even if the over-representation of the transmitted maternal haplotype is small, the over-represented haplotype can be identified provided that the counting depth exceeds the counting noise, which is governed by Poisson statistics. Supplementary Table 1 and Supplementary Fig. 1 provide estimations of counting requirement as a function of confidence of measurement and fetal DNA percentage in the clinically observed range. Because the number of markers that define each parental haplotype increases with haplotype length, the longer the phased haplotypes, the lower the average number of sampling per individual marker is required for confident determination of the over-represented parental haplotypes.

If paternal haplotypes are known, it is straightforward to determine the inherited paternal haplotypes by comparing the sum of count of

[1]Department of Bioengineering, Stanford University, Clark Center Rm E300, 318 Campus Drive, Stanford, California 94305, USA. [2]Division of Maternal-Fetal Medicine, Department of Obstetrics & Gynecology, Stanford University School of Medicine, 300 Pasteur Drive, Room HH333, Stanford, California 94305, USA. [3]Department of Applied Physics, Stanford University, Clark Center Room E300, 318 Campus Drive, Stanford, California 94305, USA. [4]Howard Hughes Medical Institute, Stanford University, Clark Center Room E300, 318 Campus Drive, Stanford, California 94305, USA.
†Current address: ImmuMetrix LLC, 552 Del Rey Avenue, Sunnyvale, California 94085, USA.
*These authors contributed equally to this work.

©2012 Macmillan Publishers Limited. All rights reserved

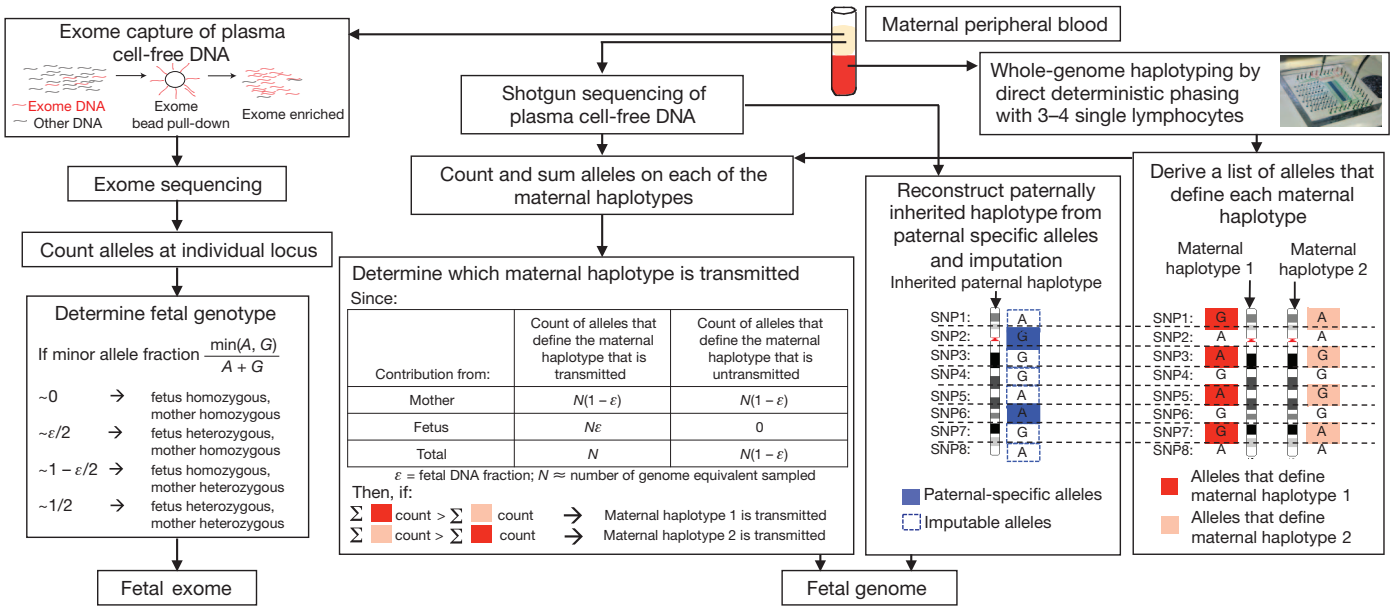

Figure 1 | Molecular counting strategies for measuring the fetal genome non-invasively from maternal blood only. Genome-wide, chromosome length haplotypes of the mother are obtained using direct deterministic phasing. The inheritance of maternal haplotypes is revealed by sequencing maternal plasma DNA and summing the count of the alleles specific to each haplotype at heterozygous loci and determining the relative representation of the two alleles. The inherited paternal haplotypes are defined by the paternal-specific alleles (that is, those that are different from the maternal ones at positions where the mother is homozygous). The allelic identity at loci linked to the paternal-specific alleles on the paternal haplotype can be imputed. Alternatively, molecular counting can be applied directly to count alleles at individual loci to determine fetal genotypes via targeted deep sequencing, such as exome-enriched sequencing of maternal plasma DNA. For illustrative purpose, each locus is biallelic and carries the 'A' or 'G' alleles.

alleles specific to each paternal haplotype (Supplementary Fig. 2), thereby revealing the entire inherited fetal genome. Supplementary Fig. 3 and the accompanying Supplementary Information show how this could be achieved using sequencing data of a synthetic mixture of DNA from a mother and daughter within a fully phased family trio[12]. However, it is not always possible to obtain paternal information; the incidence of non-paternity is estimated to be between 3% and 10%[13,14], making this a particularly delicate issue. In the absence of paternal information, the paternally inherited haplotypes can be reconstructed via linkage to observed non-maternal (that is, paternal-specific) alleles (Fig. 1).

We verified this approach on samples collected from two pregnancies. Pregnant woman P1 carried a female fetus with normal karyotype, whereas pregnant woman P2 is an individual with a ~2.85 Mb heterozygous deletion on chromosome 22 that is associated with DiGeorge syndrome. To obtain phased maternal chromosomes, we performed direct deterministic phasing (DDP)[15] on three or four maternal metaphase cells obtained by culturing maternal whole blood (Supplementary Table 2 and Supplementary Fig. 4). DDP involves microfluidic separation and amplification of individual metaphase chromosomes from single cells followed by genome-wide genotyping analysis of amplified materials, and enables each chromosome in the genome to be phased along its full length. Genomic DNA of cord blood collected at delivery was also genotyped to serve as the true reference for fetal genotypes. The true inheritance of maternal haplotypes was determined by aligning the homozygous SNPs of the fetus by cord blood genotyping against the two maternal haplotypes defined by the phased maternal heterozygous SNPs (Fig. 2). The analysis here concerns the approximately 1 million positions across the genome present on Omni1-Quad genotyping array. Phase information of the remaining genomic positions, particularly those that carry rare variants of clinical importance, can be obtained by broader array coverage or direct sequencing of amplified chromosome materials, as demonstrated previously[15].

Maternal cell-free DNA samples were shotgun-sequenced on the Illumina platform to a depth of ~52.7× (151 gigabases (Gb)), ~20.8× (59.7 Gb) and ~1.3× (3.7 Gb) haploid genome coverage for P1T1 (P1, 1st trimester), P1T2 (P2, 2nd trimester) and P2T3 (P3, 3rd trimester), respectively (Supplementary Table 2). To determine fetal inheritance of maternal haplotypes, we divided each chromosome into bins of 2.5–3.5 Mb for autosomal chromosomes and 5–7.5 Mb for chromosome X (Supplementary Table 2), with sliding steps of 100 kb, and compared the counts of alleles specific to each of the two haplotypes. Bin sizes were chosen according to the estimated sampling requirement (Supplementary Table 1) based on the sequencing depth, density of markers and fetal DNA fraction, which was estimated to be ~6%, ~16% and ~30% for P1T1, P1T2 and P2T3 by comparing relative representation of maternal haplotypes, respectively. The lower SNP array density on chromosome X required larger bin sizes for that chromosome. The over-represented maternal haplotype over the entire genome was apparent and corresponded to the maternal haplotype transmitted to the fetus (Fig. 2). Taking into account the uncertainty surrounding regions of crossovers (median ~350–450 kb per crossover, Supplementary Fig. 5), maternal inheritance of at least 99.2% of the SNPs could be deduced with at least 99.8% accuracy for all samples. Less sequencing depth also allowed the inherited maternal haplotypes to be deduced (Supplementary Fig. 6) with lower resolution of crossovers due to larger bin sizes (Supplementary Fig. 5).

The paternally inherited haplotypes were reconstructed by detection of paternal-specific alleles, followed by imputation at linked positions. We used the haplotypes of normal population documented by the 1000 Genome Project[16] as reference haplotypes for imputation. Imputation accuracy is dependent on the density of markers, and the number of identified non-maternal alleles is dependent on sequencing depth and fetal DNA fraction. At the final sequencing depth (~52.7×, ~20.8× and ~10.7× haploid genome coverage for P1T1, P1T2 and P2T3, respectively), we detected ~66–70% of the paternal-specific alleles at least once (Supplementary Table 2 and Supplementary Fig. 7). Approximately 3.4–5.6% of the non-maternal alleles were sequencing noise. Using the non-maternal markers, we deduced ~70% of the paternally inherited haplotypes with ~94–97% accuracy via imputation (Fig. 3). The loci that could not be confidently imputed reside in regions where paternal-specific alleles were not detected, in regions

©2012 Macmillan Publishers Limited. All rights reserved

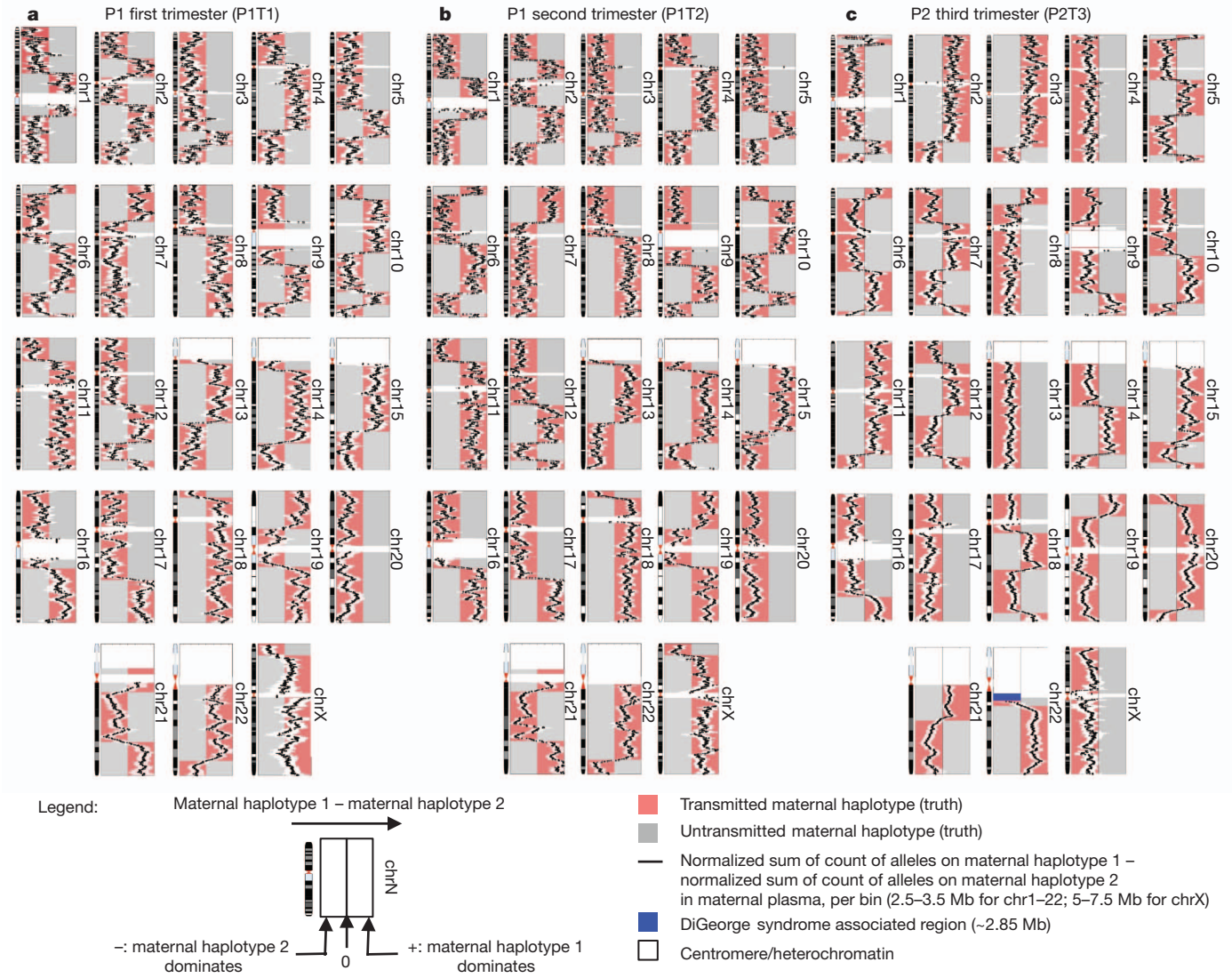

Figure 2 | Non-invasively determining genome-wide fetal inheritance of maternal haplotypes via haplotype counting of maternal plasma DNA with at least 99.8% accuracy over 99.2% of the genome in three maternal plasma samples. a–c, Each point on a black line represents the relative amount of the two maternal haplotypes evaluated using the markers lying within a bin centred at the point, and is accompanied by a white bar that corresponds to the 95% confidence interval for each measurement in P1 first trimester (**a**), P1 second trimester (**b**) and P2 third trimester (**c**). chr, chromosome. The maternal haplotypes are coloured pink or grey according to the true transmission states, as determined by fetal cord blood genotypes. Over-representation of 'maternal haplotype 2' in P2T3 maternal plasma immediately adjacent to the DiGeorge syndrome associated deletion (blue) indicates fetal inheritance of the deletion, which agrees with fetal cord blood genotype.

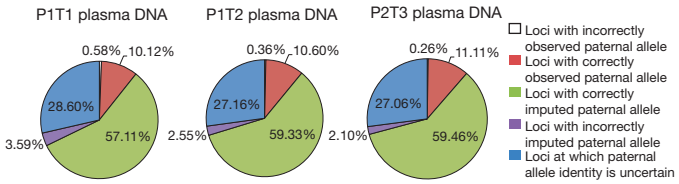

Figure 3 | Reconstruction of paternally inherited chromosomes non-invasively based on imputation using observed non-maternal alleles. The paternally inherited haplotypes were reconstructed by detection of paternal-specific alleles, followed by imputation at linked positions. At the final sequencing depth, ~66–70% of all the paternal-specific alleles were detected at least once. Using those markers, ~70% of the paternally inherited haplotypes were imputed with ~94–97% accuracy. The loci that could not be confidently imputed could in principle be completely determined by deeper sequencing and application of the counting principle directly to the individual alleles at every genomic position.

that lack paternal-specific alleles, or where the paternal alleles are associated with more than one haplotype observed in the population. In principle these regions could be completely determined by deeper sequencing and application of the counting principle directly to the local regions or the individual alleles at every genomic position, as shown below.

Counting alleles at individual loci measures fetal exome

We sought to determine clinically relevant portions of the fetal genome in maternal plasma DNA by applying the counting principle to each allele at all positions in the exome. Because the exome is two orders of magnitude smaller than the genome, less sequencing throughput is required to provide deep sequencing at individual loci and thus allows sensitive and specific detection of clinically relevant and deleterious polymorphisms that were either paternally inherited alleles or *de novo* mutations. We performed exome capture and sequencing on maternal plasma DNA samples of P1 in all three trimesters (Fig. 1 and Supplementary Fig. 9). We obtained a median coverage of 194×, 221× and 631× per position in the exome for the

©2012 Macmillan Publishers Limited. All rights reserved

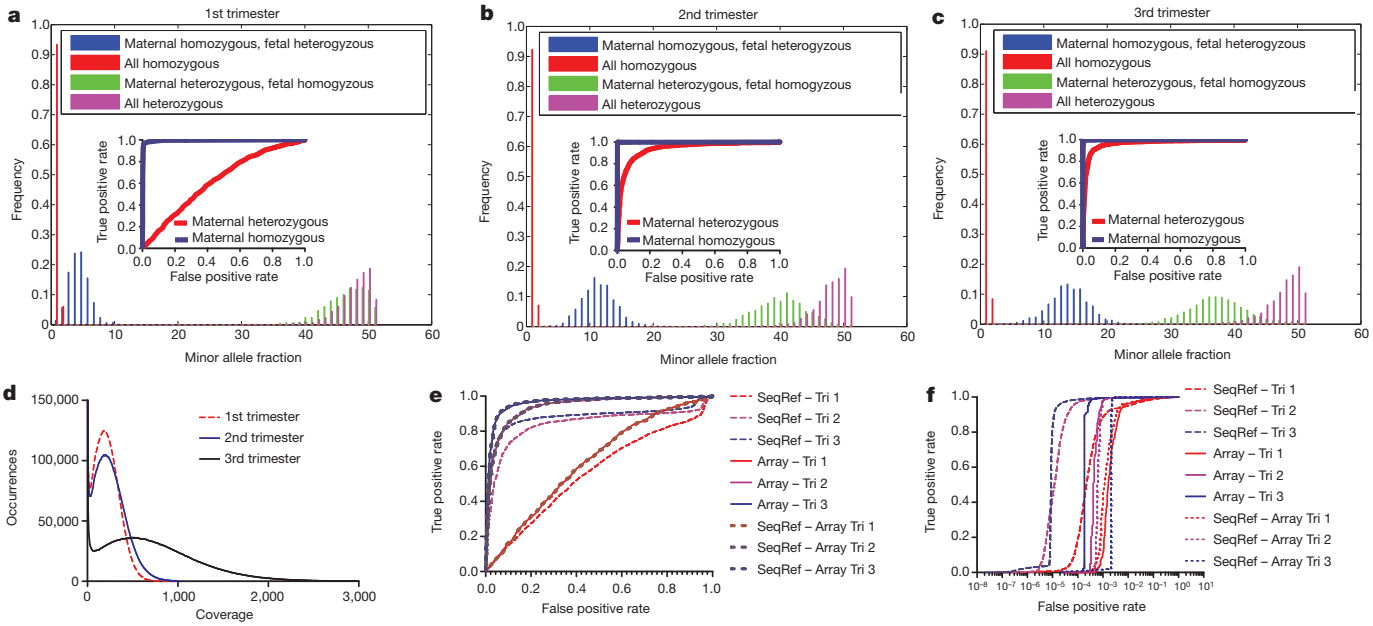

Figure 4 | Exome sequencing of P1 maternal plasma DNA in all three trimesters to determine maternal and fetal genotypes. **a**–**c**, Histograms of minor allele fraction in maternal plasma from first (**a**), second (**b**) and third (**c**) trimesters of P1 at positions that are confidently called in both plasma sequencing data and pure fetal/maternal DNA genotyping data. Insets: Receiver operating characteristic (ROC) curves of positions detecting fetal genotypes differing from maternal genotype when the maternal position is either homozygous or heterozygous. The higher the fetal fraction (~6, 20, 26% for trimester 1, 2, 3, respectively), the more the distributions are separated, and the easier it is to distinguish between the two distributions of fetal genotype. **d**, Histogram of per-position coverage, with bin size of 5. Exome positions >100× are 75%, 78% and 90% respectively for trimester 1, 2, and 3, respectively, and >200× are 48%, 56% and 84%. **e**, **f**, ROCs curves at genomic positions where mother is heterozygous (**e**) or homozygous (**f**), using either sequencing or SNP array of pure DNA as references for maternal and fetal genotypes. 'SeqRef' uses a sequenced reference, 'Array' uses a SNP array, and 'SeqRef-Array' uses a sequenced reference only at positions on a SNP array.

first, second and third trimester, respectively (Fig. 4d). After stringent data filtering to eliminate miscalled paternal-specific alleles due to limited sampling and mis-mapping to the reference genome (Supplementary Fig. 10), 75%, 78% and 90% of all exomic positions in the first, second and third trimester samples, respectively, had >100× coverage and were retained for analysis (Supplementary Table 2).

We calculated minor allele fraction, defined as the second largest nucleotide fraction divided by the sum of the two largest nucleotide fractions, at positions that are confidently called in genotyping data within the exome (Fig. 4a–c) or exome sequencing data (Supplementary Fig. 11–13) of fetal cord blood DNA and pure maternal DNA. In all three trimesters, fetal genotypes could be assigned robustly at loci where the mother is homozygous based on the separation in minor allele fraction at a depth of 200×. Paternal-specific alleles were detected with sensitivity of 96–99.8% at the specificity threshold of 99% (Fig. 4e–f and Table 1). Because the minor allele fraction at loci with paternal-specific alleles is theoretically half of the fetal DNA fraction, we estimated fetal DNA percentage to be 6.6%, 20.1%, 26.3% for the three trimesters, respectively (Supplementary Table 2). For the second and third trimester samples with higher fetal DNA fraction, fetal genotypes could be extracted for most loci at which the mother is heterozygous, as the separation in minor allele fraction for fetal homozygous and fetal heterozygous SNPs was apparent (Fig. 4a–c, e–f). For these loci, the ability to differentiate fetal heterozygosity from homozygosity depended on sequencing depth and fetal DNA fraction (Supplementary Fig. 1).

Table 1 | Exome diagnostic cut-offs and the resulting sensitivity and specificity

			Specificity cut-offs			
			Maternal homozygous		Maternal heterozygous	
	Trimester	Fetal fraction	95%	99%	85%	90%
Sensitivity	1	6%	98%	96%	25%	16%
	2	20%	99.8%	99.8%	89%	85%
	3	26%	99.7%	99.6%	96%	93%

Discussion

The molecular counting methods described here offer a gateway to comprehensive non-invasive prenatal diagnosis of genetic disease. There are substantial ethical issues associated with non-invasive prenatal genome determination, which we have not attempted to address. We will note however that there are numerous clinical scenarios where this approach would be useful. In the first or second trimester, it is possible to test for conditions that are not survivable or lead to medical complications. As technologies for pharmaceutical and surgical intervention improve, it may be possible to develop prenatal treatment or even cures for these congenital conditions.

This is illustrated by our data on P2, who is an individual with DiGeorge syndrome. Haplotyping of the maternal genome identified a ~2.85 Mb deletion on 22q11.1 that is associated with the syndrome on one copy of the maternal chromosome 22 (denoted as 'maternal haplotype 2' in Fig. 2c). Haplotype counting in maternal plasma indicated an over-representation of 'maternal haplotype 2' of the region immediately adjacent to that deletion, indicating fetal inheritance of the DiGeorge syndrome associated deletion (Fig. 2c, deletion indicated in blue). This result was confirmed by quantitative PCR of cord blood DNA (Supplementary Fig. 8). In this clinical scenario, confirmation of the deletion would argue for a fetal echocardiogram and neonatal assessment of calcium levels.

Knowledge of the fetal genotypes obtained in the third trimester enables diagnosis of conditions that would benefit from treatment immediately after delivery; these include metabolic and immunological disorders such as phenylketonuria, galactosaemia, maple syrup urine disease, and severe combined immunodeficiency. Currently, newborns with these conditions suffer as symptoms manifest themselves in the time it takes to determine the proper diagnosis and treatment,

©2012 Macmillan Publishers Limited. All rights reserved

which is often as simple as diet change. In summary, we anticipate that there is no technical barrier and many practical applications to having the entire fetal genome determined non-invasively in clinical settings.

METHODS SUMMARY

Two pregnant subjects (P1 and P2) were recruited with informed consent and approval of the Internal Review Board of Stanford University. Peripheral blood was prospectively obtained at each trimester during the course of pregnancy and post delivery. Direct deterministic phasing (DDP) was performed on three to four single cells obtained from cultures of maternal blood lymphocytes[15]. Cell-free DNA was extracted from maternal plasma during pregnancy and converted into Illumina sequencing libraries using previously established methods[4]. Exome capture was performed on cell-free DNA using SeqCap EZ v2.0 Kit (Roche NimbleGen). Genomic DNA from postpartum maternal blood cells and cord blood cells were assessed by genotyping array (Illumina's HumanOmni1-Quad) and exome sequencing to provide the reference genotypes of the mother and the fetus.

To detect the over-represented parental haplotypes, each chromosome was divided into equally sized bins with sliding window of 100 kb. The bin size was chosen such that the average count was at least that required to overcome counting noise when determining relative representation of the two maternal haplotypes. The relative representation of the maternal haplotypes was calculated using the expression $(N_{p1}/n_{p1} - N_{p2}/n_{p2})$, where N_{pi} is the number of occurrences of markers defining 'maternal haplotype *i*' within the bin counted by sequencing, n_{pi} is the total number of usable markers that define 'maternal haplotype *i*' within the bin. If the expression was positive, maternal haplotype 1 was considered inherited. If the expression was negative, maternal haplotype 2 was considered inherited. Imputation of the allelic identity on unobserved loci was calculated with Impute v1 (ref. 17) using the –haploid option.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 March; accepted 23 May 2012.
Published online 4 July 2012.

1. Mandel, P. & Metais, P. Les acides nucléiques du plasma sanguin chez l'homme. *C. R. Acad. Sci. Paris* **142,** 241–243 (1948).
2. Lo, Y. M. *et al.* Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am. J. Hum. Genet.* **62,** 768–775 (1998).
3. Bodurtha, J. & Strauss, J. F. III. Genomics and perinatal care. *N. Engl. J. Med.* **366,** 64–73 (2012).
4. Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L. & Quake, S. R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl Acad. Sci. USA* **105,** 16266–16271 (2008).
5. Sehnert, A. J. *et al.* Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. *Clin. Chem.* **57,** 1042–1049 (2011).
6. Bianchi, D. W. *et al.* Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obst. Gynecol.* **119,** 890–901 (2012).
7. Palomaki, G. E. *et al.* DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. *Genet. Med.* **14,** 296–305 (2012).
8. Palomaki, G. E. *et al.* DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. *Genet. Med.* **13,** 913–920 (2011).
9. Ehrich, M. *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.* **204,** 205.e1–211.e11 (2011).
10. Chiu, R. W. *et al.* Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *Br. Med. J.* **342,** c7401 (2011).
11. Lo, Y. M. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2,** 61ra91 (2010).
12. Fan, H. C. & Quake, S. R. In principle method for noninvasive determination of the fetal genome. Preprint at http://precedings.nature.com/documents/5373/version/1 (2010).
13. Macintyre, S. & Sooman, A. Non-paternity and prenatal genetic screening. *Lancet* **338,** 869–871 (1991).
14. Bellis, M. A., Hughes, K., Hughes, S. & Ashton, J. R. Measuring paternal discrepancy and its public health consequences. *J. Epidemiol. Community Health* **59,** 749–754 (2005).
15. Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. Whole-genome molecular haplotyping of single cells. *Nature Biotechnol.* **29,** 51–57 (2011).
16. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
17. Marchini, J. *et al.* A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.* **78,** 437–450 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors would like to thank E. Kogut and staff of the Division of Perinatal Genetics and the General Clinical Research Center of Stanford University for coordination of patient recruitment; R. Wong for initial sample processing of clinical samples; N. Neff, G. Mantalas, B. Passarelli and W. Koh for their help in sequencing library preparation and data analysis.

Author Contributions H.C.F., W.G. and S.R.Q. conceived the study. H.C.F., W.G. and J.W. performed experiments. H.C.F., W.G. and J.W. analysed the data. Y.J.B. and Y.Y.E.-S. coordinated patient recruitment. H.C.F., W.G., J.W. and S.R.Q. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.R.Q. (quake@stanford.edu).

©2012 Macmillan Publishers Limited. All rights reserved

METHODS

Prediction of counting depth requirement for determination of over-representation of transmitted maternal haplotypes. Given two distributions of Poisson random variables, one with mean of N, and the other with mean of $N(1-\varepsilon)$, where N is the cumulative sum of the count of all usable markers on the transmitted maternal haplotype, the sampling requirement of N to differentiate the two distributions can be estimated from the following expression, using the normal approximation of the Poisson distribution for large values of N:

$$\frac{N-N(1-\varepsilon)}{\sqrt{N(1-\varepsilon)+N}}=\frac{N\varepsilon}{\sqrt{N(1-\varepsilon)+N}}\geq z_\alpha$$

where z_α is the z-score associated with the confidence level of α. Thus,

$$N\geq\frac{z_\alpha^2(2-\varepsilon)}{\varepsilon^2}$$

Supplementary Table 1 present the estimated requirement of N for different values of fetal DNA fraction (ε) and level of confidence (α).

Patient samples. Two subjects, referred to as P1 and P2, were recruited to the study under approval of the Internal Review Board of Stanford University. For P1, peripheral blood was obtained during the first, second and third trimesters, and postpartum (Supplementary Table 2). For P2, peripheral blood was obtained during the third trimester and postpartum (Supplementary Table 2). Cord blood was obtained at delivery for both patients.

Whole-genome haplotyping of patient subjects. Postpartum maternal whole blood was collected into sodium heparin coated Vacutainer. Postpartum blood was used in this study because blood samples collected during pregnancy were not cryopreserved on the basis of blood culture requirement. One millilitre of whole blood was cultured with PB Max Karyotyping medium for 4 days. Direct deterministic phasing (DDP)[15] was performed on three to four single cells. Each haplotype was genotyped with Illumina's Omni1-Quad genotyping array. About 92% to 96% of the ~1 million SNPs present on the Omni1Quad BeadChip array (Illumina) (~25% are heterozygous within each individual) were phased (Supplementary Fig. 4), yielding 250–350 heterozygous markers per 3.5-Mb window. In addition to genotyping array analysis, PCR was performed on amplified materials from separated chromosome 22 of P2 to determine which maternal haplotype carried the DiGeorge syndrome associated deletion. Two regions within the deletion were tested, dgs37 and dgs40. Primer sequences are listed in Supplementary Table 3. Other rare SNPs not present on the genotyping array or not linked to loci on the array could also be phased by PCR or sequencing.

Whole-genome genotyping of the study subjects and their infants. Genomic DNA was extracted from 200 µl of postpartum maternal blood and 200 µl cord blood using QIAamp Blood Mini Kit (Qiagen), and subjected to genome-wide genotyping on Illumina's Omni1-Quad genotyping array.

Quantitative PCR confirmation of fetal inheritance of DiGeorge-associated deletion. The inheritance of the maternal haplotype carrying the deletion on chromosome 22q11.1 by the fetus of P2 was independently confirmed by quantitative real-time PCR performed on cord blood genomic DNA. The quantity of an amplicon within the deletion region (Supplementary Table 3) was compared to that of an amplicon on chromosome 1 (EIF2C1). A ratio of ~0.5 indicated that the maternal deletion was inherited.

Extraction of cell-free DNA from maternal plasma. Maternal blood were collected into EDTA coated Vacutainers. Blood was centrifuged at 1,600g for 10 min at 4 °C, and the plasma was centrifuged again at 16,000g for 10 min at 4 °C to remove residual cells. Cell-free DNA was extracted from plasma using QIAamp Blood Mini Kit (Qiagen) or QIAamp Circulating Nucleic Acid Kit (Qiagen).

Whole-genome shotgun sequencing of cell-free DNA extracted from maternal plasma. DNA was extracted from 1 to 2 ml of plasma, and subsequently converted into Illumina sequencing libraries[4] and quantified by digital PCR[18]. Sequencing was performed on the GAII and the HiSeq instruments (Supplementary Table 2). Sequences were aligned to the human genome (hg19) using CASAVA version 1.7.0. Only alleles called with quality scores >30 were used. In addition, only alleles that match previously reported variants in dbSNP were used for analyses.

Identifying the inherited parental haplotypes. Each chromosome was divided into equally sized bins with sliding window of 100 kb. The bin size was chosen such that the total number of count of markers within the bin was at least that required to overcome counting noise specifically when determining relative representation of the two maternal haplotypes.

The relative representation of the haplotype pairs of each parent was calculated using the expression $(N_{p1}/n_{p1}-N_{p2}/n_{p2})$, where N_{p1} is the number of occurrences of markers defining 'maternal or paternal haplotype 1' within the bin counted by sequencing, n_{p1} is the total number of usable markers that define 'maternal or paternal haplotype 1' within the bin, N_{p2} is the number of occurrences of markers defining 'maternal or paternal haplotype 2' within the bin counted by sequencing, n_{p2} is the total number of usable markers that define 'maternal or paternal haplotype 2' within the bin. If the expression was positive over a continuous region of 5 Mb, parental haplotype 1 was considered as inherited. If the expression was negative over a continuous region of 5 Mb, parental haplotype 2 was considered as inherited. The 95% confidence interval of relative maternal haplotype representation calculated within each bin was estimated by simulating the distribution of reads assuming the count of each maternal haplotype was the mean of a Poisson random variable.

Determining locations of recombination. The true recombination events on the maternally inherited sets of chromosomes were determined by comparing the genotype of the fetus and to the allele on each of the two maternal haplotypes at locations where the fetus is homozygous and the mother is heterozygous. In maternal plasma, a crossover event between the two maternal haplotypes giving rise to the maternally inherited chromosome in the fetus was called if in plasma DNA if two criteria were met: (1) a continuous increase or decrease in the relative representation of haplotype 1 over haplotype 2 (that is, the expression $N_{p1}/n_{p1}-N_{p2}/n_{p2}$), accompanied by a sign change, as one scanned in the direction from the p arm to the q arm of a chromosome; (2) the sign of the expression remained the same for the sliding bins 5 Mb downstream, on the basis of the fact that crossovers are rarely close to each other (positive interference).

Imputation of untyped loci on experimentally measured haplotypes. Imputation was performed using Impute v1 (ref. 17), using the –haploid option. Imputation was performed using August 2010 data from the 1000 Genome Project of the CEU population. For maternal genomes, imputation was based on the ~1 million markers phased by DDP. For paternal haplotypes, imputation was based on non-maternal alleles observed in shotgun sequencing data at locations where mother is observed and predicted based on imputation (>99% confidence) to be homozygous. Only loci with confidence of imputation >99% were considered; the allele identity for the rest were deemed uncertain. The results were compared to the true paternal haplotypes derived on the basis of the comparison of the phased maternal genome and the cord blood genotyping array data. Imputation was performed in 5-Mb segments along each chromosome.

Estimating fetal DNA fraction from maternal plasma sequencing by comparing maternal haplotype representation. Fetal DNA fraction was estimated from the over-representation of one of the maternal haplotypes. Precisely, fetal DNA fraction (ε) was estimated as $2x/(2+x)$, where x is the median absolute value of the expression $(N_{p1}/n_{p1}-N_{p2}/n_{p2})$ for all bins evaluated on either the maternal haplotypes, divided by the average marker density of the two maternal haplotypes.

Exome enrichment from maternal genomic DNA, fetal genomic DNA and cell-free DNA extracted from maternal plasma. Exome capture was performed with the SeqCap EZ v2.0 Kit (Roche NimbleGen) according to manufacturer's protocol with modifications. There are several commercially available exome kits available with varying degrees of coverage for exons, untranslated region, and microRNA regions[19]. We chose the NimbleGen platform due to its ability to capture efficiently on targeted regions and our desire for cost-efficient deep sequencing, but other platforms may perform similarly when sequenced at enough depth.

For exome enriched directly from genomic DNA extracted from maternal blood cells and cord blood, DNA was first sheared using Covaris S220 using the recommended settings for 200-base pair fragments. End repair and dA tailing reactions were cleaned up by QIAquick PCR Purification Kit (Qiagen) whereas ligation and PCR were cleaned by Agencourt Ampure XP beads (Beckman Coulter) at a 1.8× ratio of bead reagent to input volume to discard shorter adaptors, primers and ligation/PCR by-products.

Cell-free DNA extracted from approximately 3, 4, 4 and 2.5 ml of plasma were extracted from P1T1, P1T2, P1T3 and P2, respectively, was used for exome capture. For exome capture from cell-free DNA, sequencing libraries were first prepared following the NEBNext Master Mix 1 Kit (NEB). Extracted DNA was end-repaired and dA-tailed using the NEBNext kit and subsequently cleaned up with QIAquick Nucleotide Removal Kit (Qiagen) in both steps. Ligation to typical Illumina paired-end adaptors was performed at a 1:10 concentration ratio of the initial sample DNA to the adaptors. The first PCR before hybridization was carried for 18 cycles as detailed in the SeqCap protocol. Both ligation and PCR were cleaned up with Agencourt Ampure XP beads as described in the NimbleGen protocol. Prepared non-exome sequencing libraries were incubated with SeqCap kit reagents and the exome-rich sequencing library was amplified for 18 cycles in the second PCR. Libraries were quantified with digital PCR[18].

Analysis of exome sequencing data. Supplementary Fig. 10 outlined the informatics pipeline for analysing exome data. Paired-end sequencing for 100 bases on each end was performed on the HiSeq 2000 (Illumina) using v3

©2012 Macmillan Publishers Limited. All rights reserved

chemistry. Illumina's native software provided image analysis and base calling to provide FASTQ files. Those files were aligned via BWA's 'sampe' function.

Exome sequencing yielded 332, 344 and 930 million aligned reads for first, second and third trimesters, respectively (Supplementary Table 2). Because exome preparation involved more procedural steps and cycles of PCR than whole-genome shotgun sequencing preparation, we imposed a set of filters on the exome data. To remove or at least minimize bias, we opted to remove PCR duplicates on the basis of aligned location with the Picard MarkDuplicates program (the Broad Institute)[20]. In this deduplication procedure, reads with ends aligned to the exact same locations are considered PCR duplicates and amplified from same original single molecule. Deduplication helps substantially reduce bias when using paired-end and sequencing depths exceeding the sample library size. For single end-reads 100 bases long, there is only a maximum unique identification of 200 (for both directions). However, for paired end reads both ends of a DNA fragment are aligned and if fragments lengths are varied equally by 50 bases then the maximum identification library size can be 10,000, which is at least an order of magnitude above the highest coverage seen in this study. In theory it is possible to remove nearly all PCR bias if sequencing is deep enough to discover under-amplified DNA and if the theoretical identification library size is well above the actual molecular library size.

After deduplication, reads were piped through GATK (the Broad Institute) local realigner. Samtools mpileup was used to stack per position counts of different nucleotides within the exome tiles provided by the manufacturer of the SeqCap exome kit. The nucleotide count of each position was analysed against pure fetal and maternal DNA genotyping and sequencing data using custom python and MATLAB code. The minor allele fraction at each position was calculated to be the second largest nucleotide fraction divided by the sum of the two largest nucleotide fractions.

Given that fetal heterozygous genotypes at positions where maternal is homozygous can have a minor allele fraction as low as 1% on the lower end of the distribution, it is important to have more than 100× coverage to avoid classification errors occurring by chance. Beyond 100× coverage, there are also marginal improvements in sensitivity and specificity (Supplementary Fig. 14a). In addition, we filtered out misaligned regions by detecting regions with several excessively high minor allele fractions in close proximity. We filtered out 3–4 positions 40 bases apart with minor allele fractions greater than 1–5% and were able to achieve marked improvement in specificity (Supplementary Fig. 14b). Whereas filtering removes up to 4% of all positions (Supplementary Fig. 14c), it can reduce false positives by an order of magnitude at approximately the same level of sensitivity (Supplementary Fig. 14b).

Fetal DNA fraction was estimated from exome data based on minor allele fraction. The theoretical minor allele fractions are 0 for group 1 SNPs at which both mother and fetus are homozygous, $\varepsilon/2$ for group 2 SNPs of which fetus is heterozygous and mother is homozygous, $1 - \varepsilon/2$ for group 3 SNPs at which fetus is homozygous and mother is heterozygous, and 1/2 for group 4 SNPs at which both mother and fetus are heterozygous, where ε is the fetal DNA fraction. We used the median of the distribution of minor allele fraction for group 2 SNPs to provide an estimate of fetal DNA fraction.

18. White, R. A. III, Blainey, P. C., Fan, H. C. & Quake, S. R. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* **10,** 116 (2009).
19. Clark, M. J. *et al.* Performance comparison of exome DNA sequencing technologies. *Nature Biotechnol.* **29,** 908–914 (2011).
20. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **108,** 9530–9535 (2011).

©2012 **Macmillan Publishers Limited. All rights reserved**