

ARTICLE OPEN

FetalQuant^{SD}: accurate quantification of fetal DNA fraction by shallow-depth sequencing of maternal plasma DNAPeiyong Jiang^{1,2,4}, Xianlu Peng^{1,2,4}, Xiaoxi Su^{1,2}, Kun Sun^{1,2}, Stephanie CY Yu^{1,2}, Weng In Chu^{1,2}, Tak Y Leung³, Hao Sun^{1,2}, Rossa WK Chiu^{1,2}, Yuk Ming Dennis Lo^{1,2} and Kwan Chee Allen Chan^{1,2}

Noninvasive prenatal testing using massively parallel sequencing of maternal plasma DNA has been rapidly adopted in clinical use worldwide. Fetal DNA fraction in a maternal plasma sample is an important parameter for accurate interpretations of these tests. However, there is a lack of methods involving low-sequencing depth and yet would allow a robust and accurate determination of fetal DNA fraction in maternal plasma for all pregnancies. In this study, we have developed a new method to accurately quantify the fetal DNA fraction by analysing the maternal genotypes and sequencing data of maternal plasma DNA. Fetal DNA fraction was calculated based on the proportion of non-maternal alleles at single-nucleotide polymorphisms where the mother is homozygous. This new approach achieves a median deviation of 0.6% between predicted fetal DNA fraction and the actual fetal DNA fraction using as low as 0.03-fold sequencing coverage of the human genome. We believe that this method will further enhance the clinical interpretations of noninvasive prenatal testing using genome-wide random sequencing.

npj Genomic Medicine (2016) 1, 16013; doi:10.1038/npjgenmed.2016.13; published online 11 May 2016

INTRODUCTION

The discovery of circulating cell-free fetal DNA in maternal plasma¹ has catalysed a series of new avenues for noninvasive prenatal testing (NIPT), including fetal RhD blood group genotyping,^{2,3} fetal sex determination for sex-linked disorders,⁴ chromosomal aneuploidy detection^{5–10} and diagnosis of monogenic diseases.^{11–16} The accuracy of result interpretation in these tests relies on the presence of adequate amounts of fetal DNA in a maternal sample, commonly expressed as the fetal DNA fraction. The fetal DNA fraction is directly taken into consideration in the diagnostic algorithms in many clinical applications, for example, the detection of chromosomal aneuploidies¹⁷ and the determination of monogenic disease inheritance.^{11–16} In particular, the fetal DNA fraction is a key parameter for determining whether the imbalance between wildtype and mutant molecules in maternal plasma is statistically significant in the diagnosis of monogenic diseases.^{11,12,14} In such analyses, fetal DNA fraction is integrated in the mathematical models that are used in the relative mutation dosage¹⁸ and relative haplotype dosage¹² approaches to determine the theoretical thresholds for classifying the inheritance of monogenic disorders in an unborn fetus through the analysis of maternal plasma DNA.

To date, several methods have been developed for estimating the fetal DNA fraction in a maternal plasma sample. Most of these methods are based on the quantification of fetal-specific sequences that are not present in the mother's genome, for example, chromosome-Y sequences^{5,19–21} and paternally inherited single-nucleotide polymorphism (SNP) alleles. However, the detection of chromosome-Y sequences is only applicable for pregnancies with male fetuses. For the analyses using fetal-

specific alleles, informative SNP loci where the mother is homozygous (denoted as having two A alleles, i.e., AA) and the fetus is heterozygous (denoted as AB) would be identified and the fetal DNA fraction is calculated based on the ratio of the paternally inherited fetal allele (B allele) and the allele shared between the mother and the fetus (A allele). Most of these methods would require the genotypic information of both parents to identify the informative SNP loci.^{12,22} The acquisition of paternal genotype could present practical difficulties because (a) only maternal blood samples would be collected for prenatal testing in most clinical settings and (b) the paternal genotype information may not be accurate due to non-paternity.²³ In this regard, we have previously developed a method that does not require prior parental genotype information.²⁴ In that method, SNP loci showing two different alleles in plasma are first identified. Bayesian statistical analysis is applied to determine the fetal DNA fraction based on the ratios of the two alleles.²⁴ As this approach requires a sequencing depth of over 120× to ensure that the fetal-specific allele can be detected, targeted sequencing, for example, through the use of hybridization- or amplicon-based enrichment systems, would be required.^{17,22}

NIPT for chromosomal aneuploidies has been rapidly adopted for clinical service in >90 countries globally.^{25,26} One widely used approach for performing NIPT for aneuploidies is random sequencing of plasma DNA of pregnant women.⁷ In this approach, maternal plasma DNA is randomly sequenced with a relatively low-sequencing depth and mapped to a reference genome. An aneuploid chromosome would lead to an increased or reduced representation of the chromosome in the maternal plasma. For example, a trisomy 21 fetus would release an increased amount of

¹Centre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China; ²Department of Chemical Pathology, The Chinese University of Hong Kong, Hong Kong SAR, China and ³Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong SAR, China.

Correspondence: P. Jiang (jiangpeiyong@cuhk.edu.hk)

⁴These authors contributed equally to this work.

Received 1 December 2015; revised 15 March 2016; accepted 4 April 2016

chromosome 21 sequences into maternal plasma. Using this approach, the overall detection rate of trisomy 13, 18 and 21 was reported to be 98.9% with ~0.2-fold sequencing coverage of the human genome.²⁷ It would be useful if the fetal DNA fraction can be accurately and robustly determined in maternal plasma samples subjected to shallow-depth random sequencing. In this regard, size-based¹⁰ and methylation-based^{28,29} methods have been developed. However, these methods are generally less precise and accurate than methods based on fetal-specific alleles for the estimation of fetal DNA fraction. Recently, a new approach for fetal DNA fraction estimation based on random sequencing of the maternal plasma has been developed through analysing the tag densities within different windows,³⁰ however, this approach

might have reduced robustness for measuring the fetal DNA fraction < 5%.^{30,31}

In this study, we proposed a new method for determining fetal DNA fraction in maternal plasma by sequencing the maternal plasma DNA at a shallow depth, for example, 0.03-fold coverage of the human genome. This sequencing depth is readily achievable for most of the routine clinical service for NIPT of chromosomal aneuploidies. Therefore, this new method can be easily adapted to the protocols currently used by laboratories offering NIPT service. The resultant availability of accurate fetal DNA fraction information would be useful for quality control and might be incorporated into the diagnostic algorithms to improve diagnostic performance.

This new method was named as FetalQuant^{SD}. 'SD' stands for 'shallow depth' of sequencing data, which was used to highlight the difference from our previous algorithm 'FetalQuant'²⁴ that uses high-depth sequencing data (e.g., targeted sequencing).

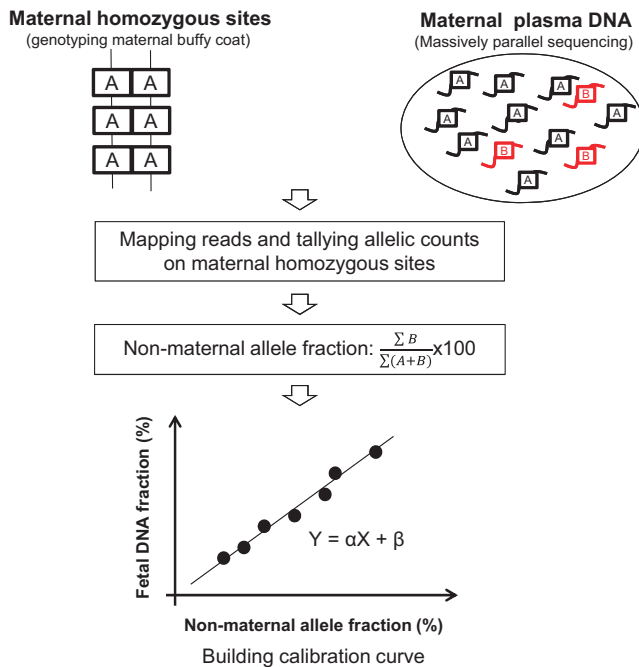


Figure 1. Schematic illustration of the principle used in FetalQuant^{SD}. Sequence reads were aligned to the human reference genome and compared with the sites where the maternal genotypes were homozygous. The non-maternal allele fraction can be inferred by aggregating all the reads carrying an allele different from the corresponding maternal allele across the maternal homozygous sites in a genome-wide manner.

RESULTS

Principle

The principle of this method is illustrated in Figure 1. Briefly, maternal blood cells were genotyped using microarray-based genotyping technologies to identify SNP loci where the pregnant woman is homozygous. Then, sequenced reads with non-maternal alleles were identified from the maternal plasma DNA-sequencing results. These non-maternal alleles would potentially represent paternally inherited fetal alleles. However, a small proportion of these non-maternal alleles could be caused by sequencing errors in maternal plasma and/or genotyping errors in maternal genomic DNA. Assuming that the error rates are relatively constant across different cases, the fetal DNA fraction would be proportional to the fraction of non-maternal alleles measured in maternal plasma.

Correlation between fetal DNA fraction and the proportion of non-maternal alleles in maternal plasma

A linear relationship was observed between the proportion of non-maternal alleles in maternal plasma and the actual fetal DNA fraction for the training data set consisting of 23 samples ($R^2 = 0.99$ and $P < 0.0001$, linear regression, Figure 2a). Thereby, we built a linear regression model to describe how the actual fetal DNA fraction is correlated with the proportion of non-maternal alleles in the maternal plasma, deriving the following equation:

$$\hat{F} = 18.9X - 6.6$$

where \hat{F} is the estimated fetal DNA fraction and X is the percentage of non-maternal alleles in the plasma sample. To

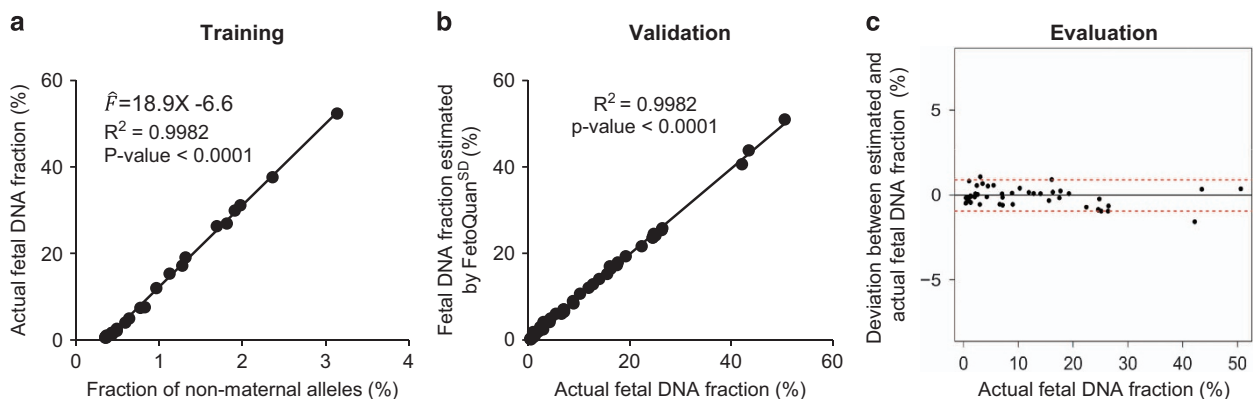


Figure 2. Linear regression model construction and validation. The regression model was constructed using the training data set (a) and validated in an independent data set (b). The deviations between the estimated and actual fetal DNA fraction were shown in c. Horizontal red lines represent the 95% confidence interval of deviations.

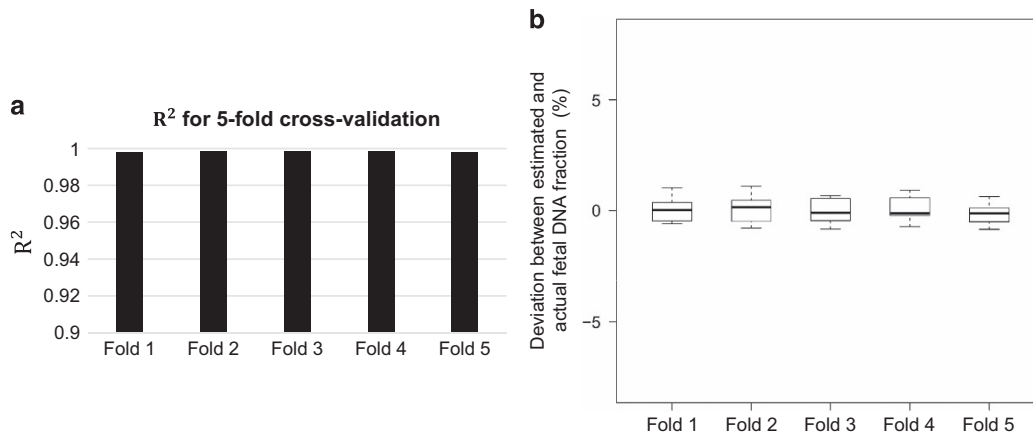


Figure 3. Fivefold cross-validation analysis. (a) R^2 of linear regression at each fold. (b) Boxplot of deviations between the estimated and actual fetal DNA fraction at each fold.

evaluate the accuracy of this regression model, we further applied it to an independent validation data set.

Accuracy of fetal DNA fraction estimation

The estimated fetal DNA fractions correlated well with the actual fetal DNA fractions ($R^2=0.99$, $P<0.0001$, linear regression, Figure 2b). The median of absolute deviation from the actual fetal DNA fraction was 0.4% (range: -1.6 to 1.1% , Figure 2c). The 95% confidence interval for the deviation was from -0.95 to 0.9% (Figure 2c).

Fivefold cross-validation analysis

Fivefold cross-validation analysis was conducted to demonstrate the robustness of the linear regression model deduced in this study. The mean values of the slopes and intercepts for these linear regression models across the fivefold cross-validation results are 19.0 (range: 18.879–19.063) and 6.62 (range: 6.53–6.68), respectively, which is close to the aforementioned linear model ($\bar{F} = 18.9X - 6.6$) deduced from 23 samples. Moreover, the mean of R^2 values across all folds was 0.998 (Figure 3a) and the mean value of the absolute deviations is 0.5% (Figure 3b), suggesting that this linear regression model could be well reproduced in the independent validation subsets.

Factors affecting the accuracy

Sequencing depth. To further demonstrate how sequencing depth would affect the measured fetal DNA fraction, we performed downsampling analysis. We randomly selected paired-end reads of 4, 3, 2, 1, 0.5 and 0.1 million per sample independently each time from the original sequencing data in the validation data set and repeated the aforementioned fetal DNA fraction prediction. As a result, for data using 1 million reads, the median of the absolute deviation reaches 0.61% (95% confidence interval: -1.93% to 1.52%), which visually provides a similar performance compared with that of 4 million reads (Figure 4).

The number of SNPs. We further explored how the number of SNPs would affect the accuracy of fetal DNA fraction estimation when 1 million reads were used. Thus, 2,000, 1,250, 1,000, 750, 500 and 250 K SNPs were randomly selected from the full data set. A total of 750 K SNPs were sufficient to give an accurate prediction showing a median deviation of 0.52% (95% confidence interval: -2.19 to 1.77% ; Figure 5).

Impact of sequencing depth and the number of SNPs upon the accuracy. Because both the number of SNPs and sequencing depth would influence the accuracy of fetal DNA fraction measurement, we repeated the aforementioned simulation analyses to investigate the accuracies corresponding to combinations of different number of SNPs and sequencing depths. Figure 6 shows the deviation at 95% confidence interval at a given number of SNPs and a particular sequencing depth. For example, 8 million reads and 300 K SNPs could give a deviation of $\pm 1.8\%$.

DISCUSSION

The accurate interpretation of the result of NIPT is affected by the amount of fetal DNA in the maternal plasma sample. Thus, the accurate measurement of the fetal DNA fraction is crucial for such testing. In this study, we have developed a new methodology to estimate the fetal DNA fraction, leveraging on the relationship between the fetal DNA fraction and the fraction of non-maternal alleles present in the plasma of a pregnant woman. Good linearity between the actual fetal DNA fraction and the fraction of non-maternal alleles was observed in the maternal plasma. These results demonstrated that the genotyping and sequencing errors were relatively constant in our hands. The predictive ability of this method has been validated in an independent data set.

Notably, the sequencing depth allowing an accurate fetal DNA fraction estimation can be as low as 0.03-fold human-genome coverage as demonstrated in the downsampling analysis. Our results suggested that this method could be robustly applied to samples undergoing NIPT for chromosomal aneuploidy detection using shallow-depth random sequencing.^{27,30} The number of SNPs required can be as little as 300,000 when 8 million reads are used (Figure 6). The small number of SNPs required can allow us to perform multiplex genotyping of maternal buffy coat samples to get maternal genotype information in a cost-effective manner. For example, the HumanCore-24 v1.0 DNA Analysis Kit (Illumina, San Diego, CA, USA) is capable of genotyping 48 samples, each for 300,000 SNPs, at a cost of USD 2,400. In other words, it would cost an extra USD 50 for each sample to obtain the maternal genotype information for fetal DNA fraction estimation. According to the simulation, 300,000 SNPs would achieve an accurate fetal DNA fraction as suggested by the deviation of $\pm 1.8\%$ (Figure 6). The sample throughput of genotyping is greater than 2,800 per week. Therefore, we believe that this method should be practicable in actual clinical use.

The accuracy of the fetal DNA prediction using this method should be higher than two previous non-polymorphism-based

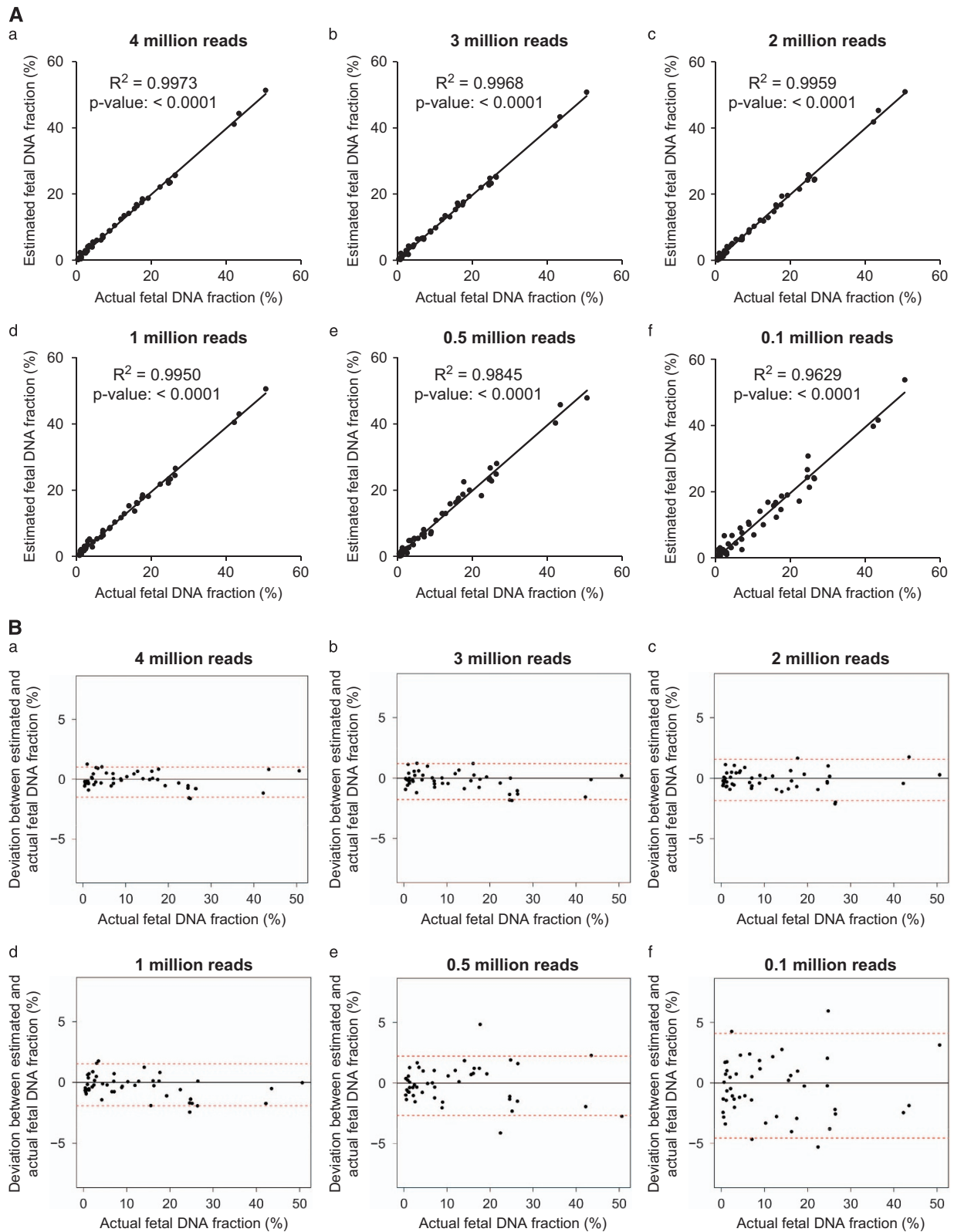


Figure 4. Evaluation on the effect of sequencing depth on estimated fetal DNA fraction. **(A)** Comparison between the estimated and actual fetal DNA fraction at different sequencing depths. **(B)** Deviations between the estimated and actual fetal DNA fraction at different sequencing depths. Horizontal red lines represent the 95% confidence interval of deviations.

approaches.^{10,30} The reported correlation coefficients between the measured and the actual fetal DNA fraction were 0.83 and 0.93 in these previous studies^{10,30}, whereas the value in our approach is 0.99. Furthermore, the new method, FetalQuant^{SD}, can accurately

measure the fetal DNA fraction of even below 5% as suggested by the median deviation of 0.6% (95% confidence interval: -1.2 to 1.7%) using 1 million reads (Figure 4B). This ability to measure low fetal DNA fractions is particularly important because the accurate

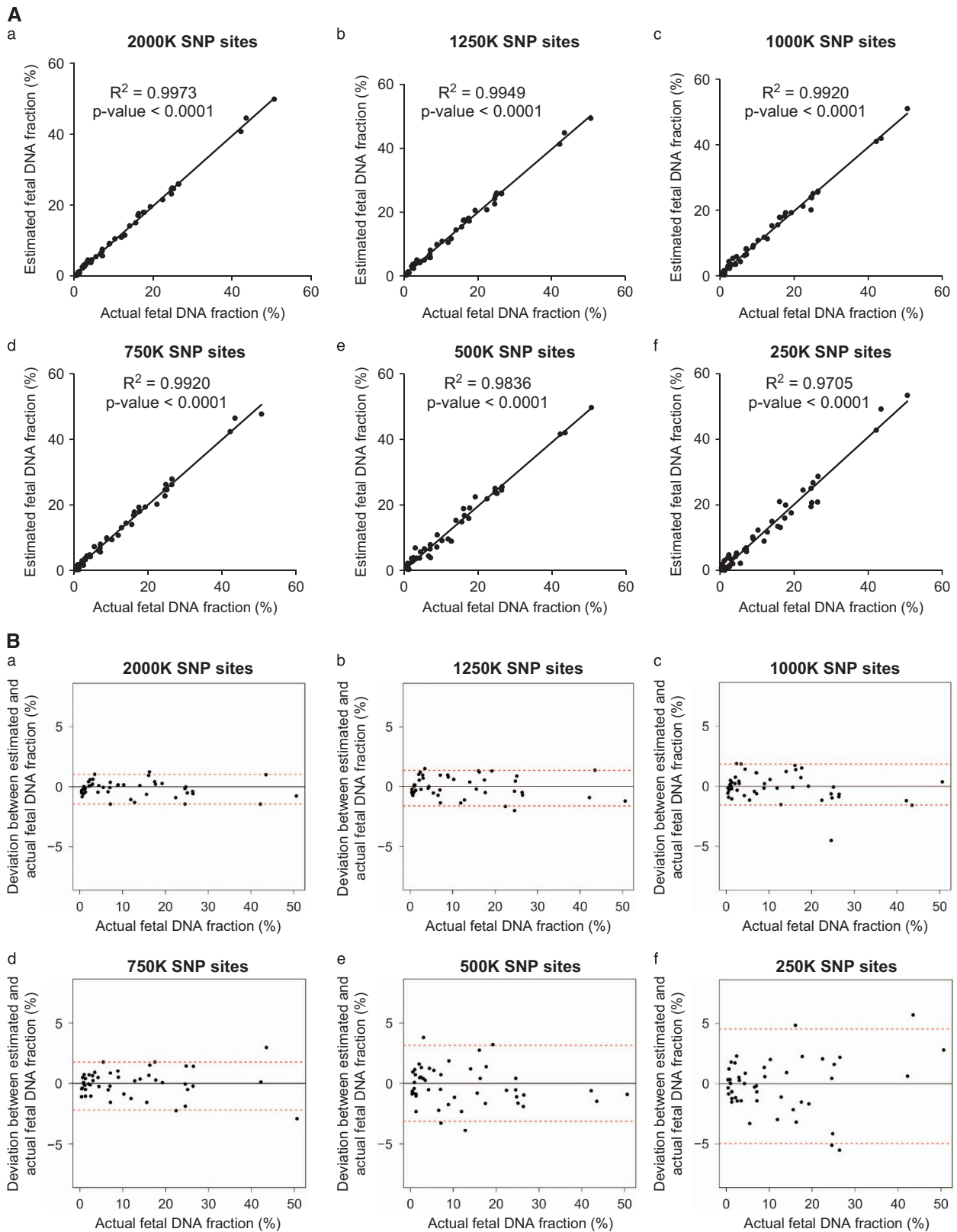


Figure 5. Evaluation on the effect of SNP count on estimated fetal DNA fraction. **(A)** Comparison between the estimated and actual fetal DNA fraction at different number of SNPs. **(B)** Deviation between the estimated and actual fetal DNA fraction at different number of SNPs. Horizontal red lines represent the 95% confidence interval of deviations.

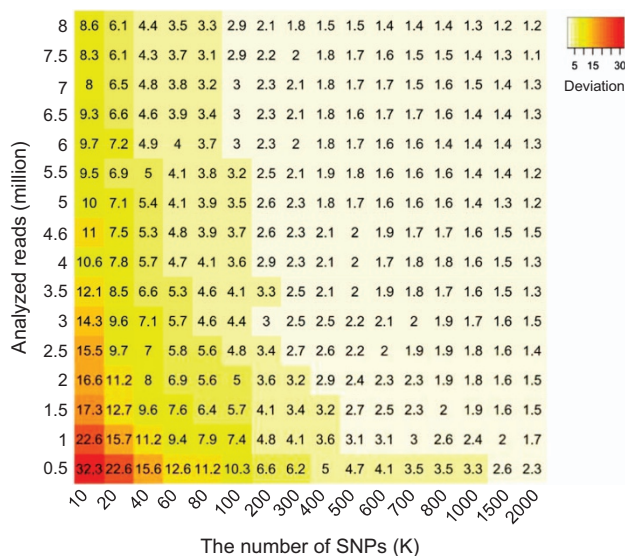


Figure 6. Impact of sequencing depth and the number of SNPs on the accuracy of fetal DNA fraction estimation. The number in each grid indicates half the width of the 95% confidence interval of deviations at a given sequencing depth and number of SNPs. The gradient colour on each grid of the heat map denotes the performance of each combination.

estimation of fetal DNA fraction will allow us to identify samples with low fetal DNA fractions so as to reduce the chance of false-negative results.^{5,27}

However, with different sequencing and genotyping platforms used, the training process would need to be repeated to deduce a new set of parameters for the linear regression model. In addition, there are two extra factors that might affect the fetal DNA fraction estimation, namely the uneven GC content and allele-specific copy-number variations. However, even though the uneven GC content across the human genome has been reported to affect the quantification of sequence reads across different genomic regions,³² the allelic ratio metric used in this study would control for the GC bias because two quantifications derived from the same region are compared. As the frequency of copy-number variations³³ and SNPs³⁴ in the human genome is relatively low, the specific bias caused by certain copy-number variations would contribute little to the current linear regression model.

With the rapidly reducing cost and increasing accessibility of personal genome-sequencing analysis,³⁵ individual genotype information will become more readily available in the near future. In the event that a pregnant subject already has genotyping information generated from a previous analysis, the maternal genotype-assisted fetal DNA fraction estimation would be readily integrated into currently existing approaches used in NIPT without any additional cost. Therefore, this method would serve as an accurate and robust method for quality control for NIPT and may lead to algorithms for improved diagnostic performance.

MATERIALS AND METHODS

Subjects

In our previous study regarding high-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing,³⁶ we collect predelivery and serial postdelivery maternal plasma. As a result, 70 plasma samples from our previous study,³⁶ which were recruited with informed consent from 12 women with uncomplicated singleton pregnancies were analyzed in this study. The blood cells were genotyped using the BeadChip array (Illumina) and the maternal plasma DNA samples were sequenced using the HiSeq 2000 platform (Illumina) with a 50-cycle

paired-end mode.³⁶ On an average, 1.94 million (range: 1.92 to 1.95 million) homozygous loci were obtained for each case for the 2.35 million SNPs on the BeadChip array. A median of 139.9 million (range: 44 to 188 million) alignable nonduplicated reads were obtained for each plasma sample.

To evaluate the performance of the fetal DNA fraction prediction, the estimated fetal fraction was compared with the fetal DNA fraction that was determined through the use of the maternal and fetal genotypes as the gold standard (deemed as the actual fetal DNA fraction) according to the study by Lo *et al.*¹²

Calculation of fetal DNA fraction

The samples were randomly divided into a training set (23 samples) and an independent validation set (47 samples). Linear regression was performed to determine the relationship between the actual fetal DNA fraction and the fraction of non-maternal alleles in maternal plasma. The actual fetal DNA fraction (F , deemed as the gold standard) was deduced by comparing the aligned sequence reads to the sites where the maternal genotypes were homozygous (AA) and the fetal genotypes were heterozygous (AB) using the following formula.¹²

$$F = \frac{2p}{p+q} \times 100$$

where p is the number of sequenced reads carrying fetal-specific alleles (i.e., allele B) and q is the number of sequenced reads carrying alleles shared by the mother and the fetus (i.e., allele A). The fraction of non-maternal alleles were calculated by comparing the aligned sequence reads to the sites where maternal genotypes were homozygous but fetal genotypes were not required to be known.

Fivefold cross-validation

Seventy samples were randomly partitioned into 5 equal-size subsets, with 14 samples in each subset. Fivefold cross-validation was performed in such a way that a single subset was retained as the validation data for testing the model, whereas the remaining subsets were used as the training data to construct the linear regression model regarding the relationship between the actual fetal DNA fraction and non-maternal allele fraction. The cross-validation process was repeated five times (also referred to as folds). Then R^2 values and deviations from the fivefold cross-validation results were used to evaluate the robustness of the deduced linear regression model.

Computational simulation for studying the impact of sequencing depth and the number of SNPs upon the accuracy

Sixty five samples with over 80 million reads each were used to study how sequencing depth and the number of SNPs would impact upon the accuracy of fetal DNA fraction estimation through the use of down-sampling analysis (Figure 6). For each combination of a given SNP count and sequencing depth, the 95% confidence interval for deviations between the estimated and actual fetal DNA fraction were calculated according to the down-sampling analysis by 20 times. Then, we used a heat map to graphically visualise half the width of the 95% confidence interval in order to demonstrate the influence of sequencing depth and the number of SNPs on the accuracy of fetal DNA fraction estimation.

Implementation of FetalQuant^{SD}

FetalQuant^{SD} was implemented using Perl (<https://www.perl.org/>) and R (<https://www.r-project.org/>) languages. It was designed to run on a x86_64 GNU/Linux platform. Perl script was used to calculate the allelic ratio and R language was used to construct the linear regression model. The source code was available at <http://www.cuhk.edu.hk/med/cpy/Research/FetalQuantSD/>.

Availability of data and materials

The sequence data for the subjects studied in this work who have consented to data archiving have been deposited in the European Genome-Phenome Archive (EGA), www.ebi.ac.uk/ega, hosted by the European Bioinformatics Institute (EBI), www.ebi.ac.uk (accession no. EGAS00001001611).

ACKNOWLEDGEMENTS

We thank Lisa Y.S. Chan and Yongjie Jin for performing the sequencing and Wing Shan Lee for performing sequencing alignment. This work was supported by the University Grants Committee of the Government of the Hong Kong Special Administrative Region, China, under the Areas of Excellence Scheme (AoE/M-04/06) and a Sponsored Research Agreement from Sequenom.

CONTRIBUTIONS

P.J., K.C.A.C., R.W.K.C., and Y.M.D.L. designed research; P.J., X.P., H.S., and K.C.A.C. performed research; P.J., X.P., X.S. K.S., S.C.Y.Y., W.I.C, T.Y.L., K.C.A.C., R.W.K.C., and Y.M.D.L. analyzed data; P.J., X.P., K.C.A.C., R.W.K.C., and Y.M.D.L. wrote the paper.

COMPETING INTERESTS

R.W.K.C. and Y.M.D.L. received research support from Sequenom. R.W.K.C. and Y.M.D.L. are consultants to Sequenom. R.W.K.C., K.C.A.C. and Y.M.D.L. hold equities in Sequenom. R.W.K.C., K.C.A.C. and Y.M.D.L. are founders of Xcelom. P.J., R.W.K.C., K.C.A.C. and Y.M.D.L. have filed patents/patent applications based on the data generated from this work. The remaining authors declare no conflict of interest.

REFERENCES

- Lo, Y. M. D. *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet* **350**, 485–487 (1997).
- Finning, K. *et al.* Effect of high throughput RHD typing of fetal DNA in maternal plasma on use of anti-RhD immunoglobulin in RhD negative pregnant women: prospective feasibility study. *BMJ* **336**, 816–818 (2008).
- Lo, Y. M. D. *et al.* Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. *N. Engl. J. Med.* **339**, 1734–1738 (1998).
- Costa, J. M., Benachi, A. & Gautier, E. New strategy for prenatal diagnosis of X-linked disorders. *N. Engl. J. Med.* **346**, 1502 (2002).
- Chiu, R. W. K. *et al.* Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ* **342**, c7401 (2011).
- Chiu, R. W. K., Cantor, C. R. & Lo, Y. M. D. Non-invasive prenatal diagnosis by single molecule counting technologies. *Trends Genet.* **25**, 324–331 (2009).
- Chiu, R. W. K. *et al.* Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc. Natl Acad. Sci. USA* **105**, 20458–20463 (2008).
- 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).
- Lo, Y. M. D. *et al.* Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc. Natl Acad. Sci. USA* **104**, 13116–13121 (2007).
- Yu, S. C. Y. *et al.* Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc. Natl Acad. Sci. USA* **111**, 8583–8588 (2014).
- Lam, K. W. *et al.* Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to beta-thalassemia. *Clin. Chem.* **58**, 1467–1475 (2012).
- Lo, Y. M. D. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).
- Ma, D. *et al.* Haplotype-based approach for noninvasive prenatal diagnosis of congenital adrenal hyperplasia by maternal plasma DNA sequencing. *Gene* **544**, 252–258 (2014).
- New, M. I. *et al.* Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. *J. Clin. Endocrinol. Metab.* **99**, E1022–E1030 (2014).
- Tsui, N. B. *et al.* Noninvasive prenatal diagnosis of hemophilia by microfluidics digital PCR analysis of maternal plasma DNA. *Blood* **117**, 3684–3691 (2011).

- Yoo, S. K. *et al.* Noninvasive prenatal diagnosis of duchenne muscular dystrophy: comprehensive genetic diagnosis in carrier, proband, and fetus. *Clin. Chem.* **61**, 829–837 (2015).
- Sparks, A. B., Struble, C. A., Wang, E. T., Song, K. & Oliphant, A. Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18. *Am. J. Obstet. Gynecol.* **206**, 319 e311–e319 (2012).
- Lun, F. M. F. *et al.* Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc. Natl Acad. Sci. USA* **105**, 19920–19925 (2008).
- Hudecova, I. *et al.* Maternal plasma fetal DNA fractions in pregnancies with low and high risks for fetal chromosomal aneuploidies. *PLoS ONE* **9**, e88484 (2014).
- Lo, Y. M. D. *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).
- Lun, F. M. F. *et al.* Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. *Clin. Chem.* **54**, 1664–1672 (2008).
- Liao, G. J. *et al.* Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. *Clin. Chem.* **57**, 92–101 (2011).
- 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).
- Jiang, P. *et al.* FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. *Bioinformatics* **28**, 2883–2890 (2012).
- Allyse, M. *et al.* Non-invasive prenatal testing: a review of international implementation and challenges. *Int. J. Womens Health* **7**, 113–126 (2015).
- Chandrasekharan, S., Minear, M. A., Hung, A. & Allyse, M. Noninvasive prenatal testing goes global. *Sci. Transl. Med.* **6**, 231fs215 (2014).
- 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).
- Chan, K. C. A. *et al.* Hypermethylated RASSF1A in maternal plasma: A universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis. *Clin. Chem.* **52**, 2211–2218 (2006).
- Chim, S. S. *et al.* Detection of the placental epigenetic signature of the maspin gene in maternal plasma. *Proc. Natl Acad. Sci. USA* **102**, 14753–14758 (2005).
- Kim, S. K. *et al.* Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenat. Diagn.* **35**, 810–815 (2015).
- Takoudes, T. & Hamar, B. Performance of non-invasive prenatal testing when fetal cell-free DNA is absent. *Ultrasound Obstet. Gynecol.* **45**, 112 (2015).
- Chen, E. Z. *et al.* Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS ONE* **6**, e21791 (2011).
- Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **16**, 172–183 (2015).
- Zhao, Z., Fu, Y. X., Hewett-Emmett, D. & Boerwinkle, E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312**, 207–213 (2003).
- Hayden, E. C. Technology: The \$1,000 genome. *Nature* **507**, 294–295 (2014).
- Yu, S. C. Y. *et al.* High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing. *Clin. Chem.* **59**, 1228–1237 (2013).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>