

## ORIGINAL ARTICLE

# A method for noninvasive detection of fetal large deletions/duplications by low coverage massively parallel sequencing

Shengpei Chen<sup>1,7†</sup>, Tze Kin Lau<sup>2†</sup>, Chunlei Zhang<sup>1†</sup>, Chenming Xu<sup>3</sup>, Zhengfeng Xu<sup>4</sup>, Ping Hu<sup>4</sup>, Jian Xu<sup>5</sup>, Hefeng Huang<sup>4</sup>, Ling Pan<sup>5</sup>, Fuman Jiang<sup>1</sup>, Fang Chen<sup>1,8</sup>, Xiaoyu Pan<sup>1,6</sup>, Weiwei Xie<sup>1</sup>, Ping Liu<sup>1</sup>, Xuchao Li<sup>1</sup>, Lei Zhang<sup>1</sup>, Songgang Li<sup>1</sup>, Yingrui Li<sup>1</sup>, Xun Xu<sup>1</sup>, Wei Wang<sup>1</sup>, Jun Wang<sup>1,8,9,10</sup>, Hui Jiang<sup>1,8\*</sup> and Xiuqing Zhang<sup>1\*</sup>

<sup>1</sup>BGI-Shenzhen, Shenzhen, China

<sup>2</sup>Fetal Medicine Centre, Paramount Clinic, Hong Kong

<sup>3</sup>Key Laboratory of Reproductive Genetics, Zhejiang University, Ministry of Education, Hangzhou, China

<sup>4</sup>State Key Laboratory of Reproductive Medicine, Center of Prenatal Diagnosis, Nanjing Maternity and Child Health Care Hospital Affiliated to Nanjing Medical University, Nanjing, China

<sup>5</sup>Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China

<sup>6</sup>School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China

<sup>7</sup>State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, China

<sup>8</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark

<sup>9</sup>King Abdulaziz University, Jeddah, Saudi Arabia

<sup>10</sup>The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark

\*Correspondence to: Hui Jiang. E-mail: jianghui@genomics.org.cn or Xiuqing Zhang. E-mail: zhangxq@genomics.org.cn

†These authors contributed equally to this work.

## ABSTRACT

**Objective** To report the feasibility of fetal chromosomal deletion/duplication detection using a novel bioinformatic method of low coverage whole genome sequencing of maternal plasma.

**Method** A practical method Fetal Copy-number Analysis through Maternal Plasma Sequencing (FCAPS), integrated with GC-bias correction, binary segmentation algorithm and dynamic threshold strategy, was developed to detect fetal chromosomal deletions/duplications of >10 Mb by low coverage whole genome sequencing (about 0.08-fold). The sensitivity/specificity of the resultant FCAPS algorithm in detecting deletions/duplications was firstly assessed *in silico* and then tested in 1311 maternal plasma samples from those with known G-banding karyotyping results of the fetus.

**Results** Deletions/duplications, ranged from 9.01 to 28.46 Mb, were suspected in four of the 1311 samples, of which three were consistent with the results of fetal karyotyping. In one case, the suspected abnormality was not confirmed by karyotyping, representing a false positive case. No false negative case was observed in the remaining 1307 low-risk samples. The sensitivity and specificity for detection of >10-Mb chromosomal deletions/duplications were 100% and 99.92%, respectively.

**Conclusion** Our study demonstrated FCAPS has the potential to detect fetal large deletions/duplications (>10 Mb) with low coverage maternal plasma DNA sequencing currently used for fetal aneuploidy detection. © 2013 John Wiley & Sons, Ltd.



Supporting information may be found in the online version of this article.

**Funding sources:** The study was funded by Shenzhen Birth Defect Screening Project Lab [JZF No. (2011) 861] approved by Shenzhen Municipal Commission for Development and Reform and Key Laboratory Project in Shenzhen [CXB200903110066A and CXB201108250096A] and Key Laboratory of Cooperation Project in Guangdong Province [2011A060906007].

**Conflicts of interest:** Shengpei Chen, Chunlei Zhang, Fuman Jiang, Fang Chen, Hui Jiang, Xiaoyu Pan, Weiwei Xie, Ping Liu, Xuchao Li, Lei Zhang, Songgang Li, Yingrui Li, Xiuqing Zhang and Wei Wang are employees of BGI-Shenzhen, and none of the other authors have any financial relationship with BGI-Shenzhen.

## INTRODUCTION

Deletion/duplication syndromes are well known to be associated with a wide range of structural and functional

abnormalities,<sup>1–3</sup> such as Cri du chat syndrome (5p deletion),<sup>4</sup> DiGeorge syndrome (22q11.2 deletion)<sup>5</sup> and Angelman syndrome (15q11–q13 deletion).<sup>6</sup> Such deletion/duplication

syndromes can be reliably detected prenatally by studying fetal DNA/cells collected by invasive procedures, using a wide range of techniques, including karyotyping, fluorescence *in situ* hybridization, comparative genomic hybridization (CGH) and array-based technologies.<sup>7</sup> However, the population-based prenatal detection of these syndromes could be difficult because there is no simple effective screening test for this group of conditions, while some patients may decline invasive tests because of the associated risk of miscarriage.<sup>8</sup> Therefore, there is a demand to develop a highly accurate noninvasive genetic test for fetal large deletion/duplication detection.

Although the presence of cell-free fetal DNA (cff-DNA) in maternal plasma was first reported until 1997,<sup>9</sup> the noninvasive detection of fetal aneuploidy has now become a reality.<sup>10-14</sup> Two proof-of-concept studies have demonstrated the possibility of fetal deletion/duplication detection from maternal plasma. Peters *et al.* reported that maternal plasma sequencing with 243M reads could identify a 4-Mb fetal deletion at 35 weeks of gestation.<sup>15</sup> Jensen *et al.* also developed a strategy for 22q11.2 syndrome detection using an average 3.83-fold of maternal plasma sequencing data.<sup>16</sup> Recently, a new study about noninvasive detection of fetal subchromosome abnormalities was reported by Anupama Srinivasan *et al.* In this study, approximately 10<sup>9</sup> tag sequencing data were obtained to identify subchromosomal duplications and deletions, translocations, mosaicism and trisomy 20 by maternal plasma sequencing in seven cases.<sup>17</sup> However, all those three studies required deep sequencing. As a comparison, the sequencing depth for noninvasive prenatal detection of fetal aneuploidy from maternal plasma is only about 0.08-fold. The cost of deep whole genome sequencing restricts their application in real clinical situation. To be clinically usable, there is a need to reduce the requirement on deep sequencing.

In noninvasive prenatal detection of fetal aneuploidies, most published studies employed basic reads counting strategy for each chromosome and simple statistics such as Z-test to

identify aneuploidies.<sup>18-20</sup> In principle, fetal large deletion/duplication can also be detected by such algorithms if the sequencing depth is increased.<sup>16</sup> However, such analytic approach does not provide enough power to detect deletion/duplication at the current sequencing depth of aneuploidy detection. The sliding window strategy has been reported to be able to accurately detect copy-number variations in human genome with relatively low coverage whole genome sequencing.<sup>21</sup> Thus, developing a sliding window-based statistic model may enable the detection of fetal large deletion/duplication detection using less sequencing data of maternal plasma.

In this study, we established a novel bioinformatics method, Fetal Copy-number Analysis through Maternal Plasma Sequencing (FCAPS), for noninvasive genome-wide detection of fetal large deletions/duplications, and the algorithm was tested in 1311 maternal plasma samples. Our study highlighted the prospect of universal and practical noninvasive screening for deletion/duplication in fetal genome, which can be incorporated into the current program of noninvasive prenatal detection of fetal aneuploidy without increasing sequencing depth.

METHODS

Overall study design

The human reference genome was firstly divided into overlapping sliding windows to ensure each window contained the same number of unique reads. The GC-bias correction was performed using sequencing data from 140 control samples. We then developed a binary segmentation algorithm for potential breakpoint localization and a dynamic threshold for signal filtering. Figure 1 summarized the flow of FCAPS analysis.

The performance of the resultant FCAPS algorithm in detecting >10 Mb deletions/duplications was first evaluated *in silico*, followed by testing 1311 maternal plasma samples with known fetal karyotype.

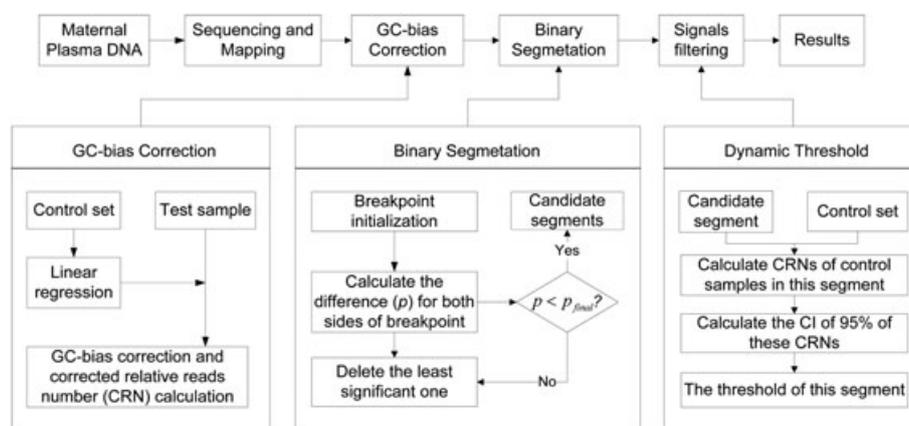


Figure 1 The pipeline for Fetal Copy-number Analysis through Maternal Plasma Sequencing (FCAPS). This figure shows the pipeline of FCAPS for fetal deletion/duplication identification using maternal plasma sequencing. GC-bias correction, binary segmentation and signals filtering using dynamic threshold form the core of FCAPS. CRN, corrected relative reads number; CI, confidence interval

### Clinical sample collection and preparation

Maternal plasma samples were collected from 1451 pregnant women with gestational age ranging from 13 to 28 weeks. For all samples, the pregnant women had received amniocentesis or chorionic villus sampling on the basis of their clinical need after peripheral blood sampling. The metaphase chromosome G-banding karyotype was performed with 350–500 bands with around 10 Mb of resolution; therefore, in this study, the G-banding karyotype was employed as gold standard to evaluate the accuracy of FCAPS test. Among those with normal karyotype, we randomly selected 140 normal samples as control set to develop the adjustment factor for GC correction, while the remaining 1311 samples as test set to assess the performance of the FCAPS pipeline. Approval was obtained from the institutional review board of BGI-Shenzhen before recruitment. Informed written consent was obtained from each participant.

Ten milliliters of peripheral venous blood was taken from each pregnant woman into tubes containing ethylenediaminetetraacetic acid, and plasma was prepared by centrifugation at 1600g for 10 min. The supernatant was transferred into sterile tubes and centrifuged for another 10 min at 14 000g. The plasma fraction was aliquoted and stored at  $-80^{\circ}\text{C}$  for future processing. DNA was isolated from 600  $\mu\text{l}$  of plasma using the QIAamp DNeasy Blood & Tissue Kit (Qiagen, Germany) according to the manufacturer's recommendations. The process of small insert size DNA library construction was in accordance with the manual of Illumina HiSeq 2000.

The single-end 50-bp (SE50) reads were mapped to the reference human genome (Hg18, Build36) using SOAP2.<sup>22</sup> After removal of the PCR duplication and non-unique mapped reads, the remaining unique reads were used for the following study.

### Window selection

The human reference genome (Hg18, Build 36) was smashed into sliding SE50 simulated reads. Simulated reads that can be uniquely mapped to the genome were reserved for following window construction. Instead of creating constant length windows, we adjusted the boundary of each window to ensure all windows shared a constant expected number of uniquely mapped simulated reads. To improve the resolution and accuracy, adjacent windows were allowed to share 99% overlap. We aimed at having an average window length of approximately 1 Mb.

### GC-bias correction

For further bioinformatics analysis, we defined the number of unique alignments within a window as  $r_{i,j}$ , where  $i \in \{1, 2, \dots, n\}$  and  $j \in \{1, 2, \dots, m\}$  represent window number and sample number, respectively. The relative reads number (RRN,  $R_{i,j}$ ) was defined as the logarithm normalized reads coverage and was expressed as  $R_{i,j} = \log_2(r_{i,j}/\bar{r}_j)$ , where  $\bar{r}_j = \frac{1}{n} \sum_{i=1}^n r_{i,j}$ . The

sequencing GC content ( $GC_{i,j}$ ) was defined as the average GC content of sequencing reads in window  $i$  of sample  $j$ .

Previous studies have shown that the sequence read coverage would be under-represented in GC-rich region and GC-poor region because of the PCR process in library preparation and cluster generation.<sup>23</sup> Here, we firstly studied the GC bias in the sliding windows we have selected in the 140 controls. Least-squares estimation was performed in the same window of different control samples for analytic expression of the bias factor between RRN and GC content. The slope and intercept of the linear regression were denoted as  $a_i$  and  $b_i$ . Therefore, the bias factor ( $\tilde{R}_{i,j}$ ) could be calculated as  $\tilde{R}_{i,j} = a_i \times GC_{i,j} + b_i$ . Consequently, the corrected reads number (CRN) in a test sample, defined as the RRN after GC-bias correction, could be calculated as  $Z_{i,j} = (R_{i,j} - \tilde{R}_{i,j} - \text{mean}_j) / \text{sd}_j$ , where  $\text{mean}_j =$

$$\frac{1}{n} \sum_{i=1}^n (R_{i,j} - \tilde{R}_{i,j}) \text{ and } \text{sd}_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R_{i,j} - \tilde{R}_{i,j} - \text{mean}_j)^2}.$$

### Segmentation algorithm for fetal large-segment deletion/duplication identification

To localize the segment breakpoints of the fetal large deletions/duplications, we merged the adjacent windows with similar CRN. A binary segmentation algorithm was used in our statistic model to achieve high sensitivity of accessing the optimized breakpoint,<sup>21</sup> in which the difference of CRN value for the left and right windows of the candidate breakpoints was calculated by an iterative algorithm. Run-test was recruited to examine CRN difference between two adjacent segments as the significance of candidate breakpoint ( $p_k$ ). Candidate with the largest  $p$ -value would be clicked off, and other  $p$ -values would be refreshed until all of the  $p$ -values were less than the genome-wide significance threshold ( $p_k < p_{\text{final}}$ ). (Full details are available in Supplementary Methods.)

### Dynamic threshold determination for final signal filtering

Chromosome ends, centromere or most of repeats regions, which are characterized by N regions, might display false positive or false negative signals. To minimize false signal of fetal deletions/duplications, we developed a dynamic threshold strategy. On the basis of the dynamic threshold, we identified the mutation type of each segment after segmentation. The dynamic threshold was calculated as 95% confidence interval (CI) of CRNs, which were located in certain segment, in control samples. (Full details are available in Supplementary Methods.)

### Sensitivity and specificity estimation of FCAPS *in silico*

Poisson-distributed random numbers were generated as sequence reads number of each window *in silico* under different conditions including cff-DNA concentration, size of the deletion/duplication and the number of reads from sequencing data. For this simulation study, the cff-DNA concentration varied from 5% to 15% with increment of 2.5%, the size of deletion/duplication varied from 1 to 15 Mb with

increment of 2 Mb and data volume was 5M, 7M and 10M, respectively. Each simulation was repeated 100 times to estimate the detection power.

Application of FCAPS in clinical samples

The developed FCAPS algorithm was tested in 1311 maternal plasma samples with known fetal karyotyping results. The karyotyping result was blinded before obtaining the results of FCAPS test. The findings from FCAPS analysis were compared against the karyotyping result, to calculate the sensitivity and specificity of FCAPS in detecting >10 Mb deletion/duplication in clinical samples.

RESULTS

Samples collection and maternal plasma DNA sequencing

Pregnant women (1451) were recruited from 15 centers around China in our multicenter study, and maternal bloods were sampled before invasive procedures. The average age of subjects was 32 years. The gestational age ranged from 10 to 28 weeks, with an average of 21 weeks. About 2~8 million SE 50-bp sequencing reads per test samples were obtained (4.42 ± 2.65 M). These reads were mapped uniquely to the

reference sequence (Hg18, Build36), covering approximately 6% of human genome (i.e. about 0.08-fold). The basic clinical information and sequencing data were shown in Table 1.

Window selection

On the basis of our methodology, the human genome was divided into a total of 308 789 sliding windows with 99% overlap as the basic observation units. There were 84 000 uniquely mapped simulated reads within each window, and the average size of the windows was 0.94 ± 0.68 Mb.

GC-bias observation and correction in FCAPS

To minimize the interference of PCR-specific bias, a regression was performed between RRN and GC content in the same window of controls. The median coefficient of determination ( $R^2$ )<sup>24</sup> was 0.776, showing a significant linear relationship between RRNs and the GC content in the same window of different control samples (Figure S1). The remarkable linear relationship enabled us to perform the GC correction and calculate the CRN. The distribution between RRN and GC content before and after normalization indicated that our GC correction strategy highly improved the data stability against GC content (Figure 2).

Table 1 Sequencing data statistics and clinical information

Samples	Control cases (n=140)	Test cases with negative FCAPS results (n=1307)	Positive Case 1	Positive Case 2	Positive Case 3	Positive Case 4
GC (%)	41.70 ± 2.23	39.82 ± 1.00	40.48	42.17	40.59	41.67
Number of reads (M)	5.00 ± 2.59	4.35 ± 2.65	6.86	6.96	7.84	2.17
Coverage (%)	5.98 ± 3.54	5.14 ± 3.38	10	8.8	8.4	2.5
cff-DNA concentration (%)	—*	—*	19.01	10.43	35.38	—*
Maternal age	37 ± 4.30	32 ± 5.33	33	29	34	38
Gestational weeks	13 ± 2.00	21 ± 2.53	28	14	20	20

cff-DNA, cell-free fetal DNA; FCAPS, Fetal Copy-number Analysis through Maternal Plasma Sequencing.

\*—, means the information was missing.

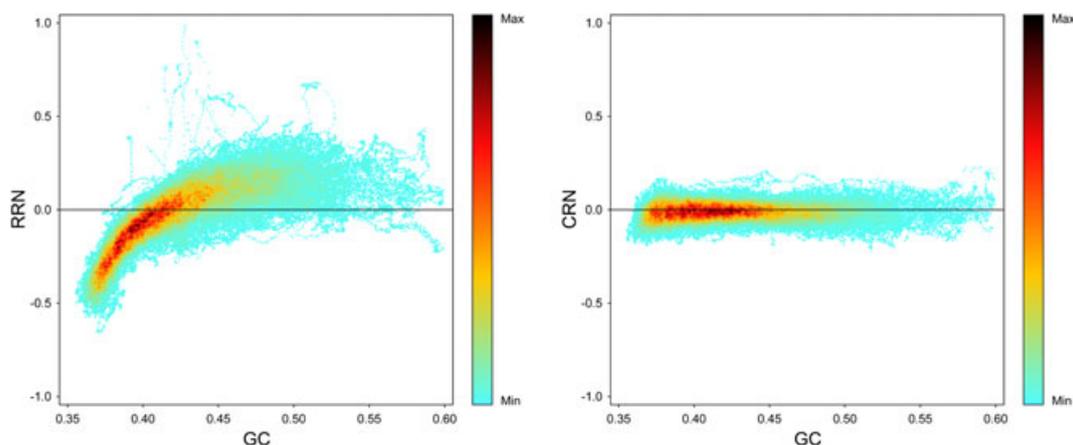


Figure 2 The distribution between relative reads number (RRN) and GC content before and after correction. The distributions of RRN (y axis) and the corrected relative reads number (CRN, y axis) were showed as heat map respectively with their sequence GC content (x axis)

### Dynamic threshold strategy for fetal deletion/duplication identification

To minimize the false signals of fetal deletions/duplications, we employed the dynamic threshold strategy. If a fixed threshold as 95% CI of CRNs in this sample ( $-1.645, 1.645$ ) was used, 77 false events of deletion/duplication ( $>10$  Mb) would have been suspected among the 1311 subjects. With dynamic threshold, only one false event was detected. For example, in Positive Case 3, a normal region on Chr4:1–10 173 290 (average CRN=1.686) would be judged as duplication by fixed threshold while identified as normal in dynamic threshold ( $-2.156, 2.139$ ). In other words, a dynamic threshold would substantially decrease the false positive rate.

### Performance of FCAPS *in silico*

We estimated the sensitivity of FCAPS *in silico* to assess the performance of this test in clinical practices (Figure 3 and Table S1). Generally, the detection power increased with higher cf-DNA concentration and more sequencing reads. On the condition of 10% cf-DNA concentration and 7M sequence reads, our simulated data showed that close to 100% of  $>10$  Mb deletion/duplication could be successfully detected (Figure 3).

### Large deletion/duplication detection by FCAPS

We tested a total of 1311 maternal plasma samples using FCAPS to identify deletion/duplication over 10 Mb. Four samples were classified as a high-risk subgroup that carried fetus with large deletion/duplication by our dynamic threshold algorithm. Totally, six events of deletion/duplication were observed in four samples, including three deletions with a mean length of 12.02 Mb, and three duplications with a mean length of 18.51 Mb (Figures 4 and S2). No fetal deletions/

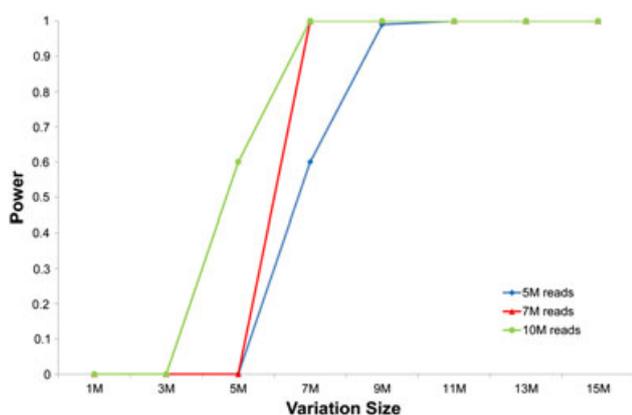


Figure 3 The power evaluation of Fetal Copy-number Analysis through Maternal Plasma Sequencing. The figure shows that the fetal deletion/duplication detection power (y axis) increases with increasing deletion/duplication size (x axis) and sequencing data size (color-code lines) when the cf-DNA concentration is constant at 10%. The lines with blue, red and green represent the change of fetal deletion/duplication detection power when the sequencing data size is 5M, 7M and 10M

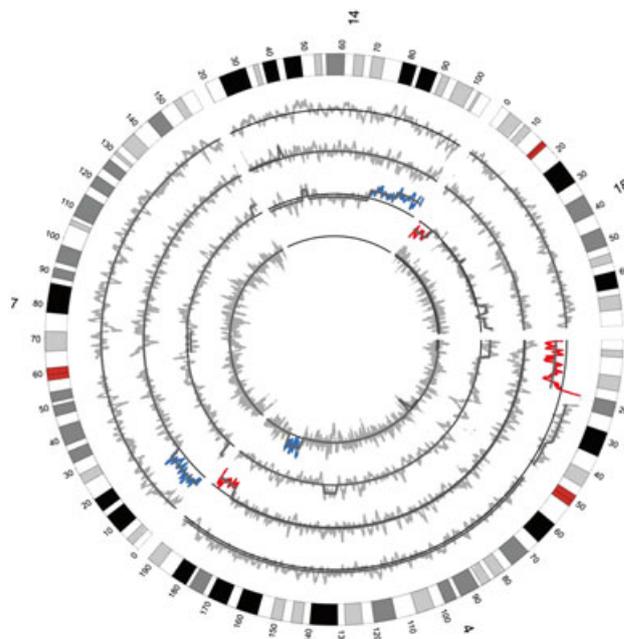


Figure 4 The performance of Fetal Copy-number Analysis through Maternal Plasma Sequencing (FCAPS) for test samples. Circular map shows the performance of FCAPS in four positive samples with deletion/duplication. The circles show chromosome no., color-code chromosome bands, Positive Case 1, Positive Case 2, Positive Case 3 and Positive Case 4 successively inwards. The color-code dots show the distribution of CRN, of which blue and red dots show duplication and deletion, respectively. The dark gray lines crossing the color-code dots show the deletions/duplications after segmentation

duplications larger than 10 Mb were detected in the remaining 1307 samples.

All samples had G-banding karyotyping, which was used as the golden standard to detect chromosomal deletions/duplications over 10 Mb. Three of four positive cases by FCAPS were confirmed by G-banding karyotyping, whereas the remaining one was a false positive (Table 2). No false negative cases were observed in the remained 1307 low-risk samples. The incidence of fetal chromosomal large deletion/duplication was 0.23%. Overall, the sensitivity and specificity for large deletion/duplication using FCAPS were 100% and 99.92%.

For the false positive case (Case 4), further examination showed that the false positive signal was nearly N regions on reference genome, which refer the regions with unknown sequence on human genome and display as 'N' in reference genome sequence. For instance, in this false positive case, the false signal on Chr4:158 281 795–169 246 069 was overlapped with the N region on reference genome: 167 795 055–167 825 054.

### Clinical outcome of three positive cases

Three positive cases were detected by our method (Figures 4 and S2, and Table 2). In the first case (Positive Case 1), we identified a 19 Mb deletion on Chr4:1–19 011 143, corresponding to cytogenetic bands of 4p16–4p15.3, which

Table 2 The results by the FCAPS analysis on the four samples found to be abnormal

Samples	FCAPS results	Karyotype results
Positive Case 1	del(4) (p16-p15.3, ~15.86 Mb)	46,XN*, del(4) (p16-p15.3)
Positive Case 2	del(4) (q34.3-q35.2, ~10.01 Mb), dup(7) (p22.3-p21.1, ~17.04 Mb)	46,XN*, del(4) (q34.3-q35.2), dup(7) (p22.3-p21.1)
Positive Case 3	dup(14) (q24.3 → qter, ~28.46 Mb), del(18) (p11.22 → pter, ~9.01 Mb)	46,XN*, der(18)
Positive Case 4	dup(4) (q32.1-q32.3) (~10.96 Mb)	46,XN*

FCAPS, Fetal Copy-number Analysis through Maternal Plasma Sequencing.

\*XN is the method of annotation of fetal sex in prenatal karyotype based on the national policy of China that prenatal determination of fetal sex is not allowed. 'XN' means there are two normal sex chromosomes, one of which is X chromosome, whereas the other is concealed so as not to reveal the fetal sex in the prenatal period.

caused Wolf-Hirschhorn syndrome.<sup>7</sup> An obvious fetus deformity was also detected by ultrasonic examination, and the decision of termination of pregnancy was made by the pregnant woman after counseling by her clinician. In the second case (Positive Case 2), we simultaneously identified a 10.01 Mb deletion on Chr4:181 243 323–191 250 465 and 17.04 Mb duplication on Chr7:1–17 074 358, both of which were confirmed by fetal array-CGH analysis. The baby was delivered with multiple anomalies and died in day 15. In the last case (Positive Case 3), we found a 28.46 Mb duplication on Chr14:77 901 695–106 360 226 and a 9.01 Mb deletion on Chr18:483 517–9 489 300. The decision of termination of pregnancy was made by the pregnant women. All these participants made the decision on the basis of the invasive test results (G-banding analysis), and the results detected by our method were consistent with that of G-banding analysis.

## DISCUSSION

In this study, we developed a practical bioinformatics method FCAPS to noninvasively detect fetal large deletion/duplication, which employed a regression-based GC correction strategy to improve the stability of diploid background, binary segmentation algorithm for breakpoint localization and dynamic threshold for signal filtering. We tested this algorithm in 1311 pregnant women to detect large deletion/duplication in fetus noninvasively. Using only 2–8 million sequencing reads, we correctly identified three pregnant women carried fetuses with deletion/duplication over 10 Mb with 99.92% of specificity and 100% of sensitivity. The positive predictive value was 75%, which is substantially higher than that of <5%<sup>25,26</sup> among conventional Down syndrome screening tests based on ultrasound or maternal serum biochemistry.

The major advantage of our approach is the substantial reduction in the required sequencing reads to less than 10 millions, compared with hundreds of millions in previous studies, making the noninvasive detection of fetal large deletion/duplication closer to reality in clinical practices. Unlike previous studies, our study firstly developed a novel bioinformatics approach to identify deletion/duplication in fetus and then tested the efficiency of this approach by both computer simulations and real clinical data.

In many previous studies, the reference genome was divided into windows with constant length in window selection, which would lead to further RRN statistics bias. For example, in the case of 1-Mb overlapping window with constant length, in Chr1: 1–1 000 000, only 27.2% of SE50 simulated reads could be mapped uniquely into this region. As a comparison, the average level of unique mapping simulated reads in all windows was as high as ~90%. In this study, we divided the reference genome into observation windows with constant expected unique reads numbers instead of constant length. Thus, uncertainties caused by sequence and mapping strategy, such as reads length or repeat sequence, could be well considered to minimize the potential bias in different regions. Moreover, by adjusting the boundary, the expected unique reads number of each observation window would be equal, leading to a more centralized RRN in further statistics.

There also are limitations in this study, especially about the relative low resolution of our approach so that only large deletions/duplications could be detected. Decipher is one of the most well-known database of chromosomal imbalance and phenotype in humans, including 64 syndromes with various deletion/duplication in genome ranging from 0.02 to 16.97 Mb.<sup>27</sup> However, only three out of the 64 syndromes in the Decipher were associated with deletion/duplication larger than 10 Mb. According to our *in silico* data, at 10% cff-DNA concentration and 7M sequence reads, our simulated data showed that close to 100% of >10-Mb deletion/duplication could be successfully detected, and the detection power reduces rapidly with the decreasing size of deletion/duplication. Fortunately, the detection power of smaller mutations could be improved by increasing sequence reads as expected (Figure 3 and Table S1). At this moment, the simplest approach to increase the resolution of FCAPS to cover fetal micro deletion/duplication is to increase the sequencing depth. Further study in bioinformatics might enable the detection of smaller deletion/duplication without the need for high sequencing depth.

With the rapid development of noninvasive detection of fetal trisomies 21, 18 and 13, this MPS-based method is widely used in both the USA and China as a second tier screening test offered to medium-risk or high-risk women on the basis of conventional prenatal screening. Because our approach could detect large deletion/duplication of fetus with the same or a slightly higher sequencing reads, it could be easily integrated into the existing fetal aneuploidy detection test. However, considering its

significant higher false positive rate, a careful post-test genetic counseling is required for the pregnant women to make a decision of invasive validation by G-banding karyotyping or aCGH.

## CONCLUSION

In summary, our study showed the great potential of noninvasive detection of fetal large deletion/duplication with ultra-low coverage of whole genome sequencing of maternal plasma. Our method, with a high sensitivity and acceptable specificity, can broaden the application for lower coverage sequencing of maternal plasma in noninvasive prenatal testing.

## REFERENCES

- Freeman JL, Perry GH, Feuk L, *et al.* Copy number variation: new insights in genome diversity. *Genome Res* 2006;16:949–61.
- McCarroll SA, Altschuler DM. Copy-number variation and association studies of human disease. *Nat Genet* 2007;39:S37–42.
- Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 2007;8:639–46.
- Lejeune J, Lafourcade J, Berger R, *et al.* 3 cases of partial deletion of the short arm of a 5 chromosome. *C R Hebd Seances Acad Sci* 1963;257:3098–102.
- Lammer EJ, Opitz JM. The DiGeorge anomaly as a developmental field defect. *Am J Med Genet* 2005;25:113–27.
- Hart H. 'Puppet' children. A report on three cases (1965). *Dev Med Child Neurol* 2008;50:564.
- Bianchi DW. From prenatal genomic diagnosis to fetal personalized medicine: progress and challenges. *Nat Med* 2012;18:1041–51.
- Simpson JL. Invasive procedures for prenatal diagnosis: any future left? *Best Pract Res Clin Obstet Gynaecol* 2012;26(5):625–38.
- Lo YM, Corbetta N, Chamberlain PF, *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet* 1997;350:485–7.
- Dan S, Wang W, Ren J, *et al.* Clinical application of massively parallel sequencing-based prenatal noninvasive fetal trisomy test for trisomies 21 and 18 in 11 105 pregnancies with mixed risk factors. *Prenat Diagn* 2012:1–8.
- Lau TK, Chen F, Pan X, *et al.* Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing. *J Matern Fetal Neonatal Med* 2012;25:1370–4.
- Palomaki GE, Deciu C, Kloza EM, *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* 2012;14:296–305.
- Norton ME, Brar H, Weiss J, *et al.* Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. *Am J Obstet Gynecol* 2012;207:137 e1–8.
- Bianchi DW, Platt LD, Goldberg JD, *et al.* Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstet Gynecol* 2012;119:890–901.
- Peters D, Chu T, Yatsenko SA, *et al.* Noninvasive prenatal diagnosis of a fetal microdeletion syndrome. *N Engl J Med* 2011;365:1847–8.
- Jensen TJ, Dzakula Z, Deciu C, *et al.* Detection of microdeletion 22q11.2 in a fetus by next-generation sequencing of maternal plasma. *Clin Chem* 2012;58(7):1148–51.
- Srinivasan A, Bianchi DW, Huang H, *et al.* Noninvasive detection of fetal subchromosome abnormalities via deep sequencing of maternal plasma. *Am J Hum Genet* 2013;92(2):167–76.
- Fan HC, Quake SR. Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. *PLoS One* 2010;5:e10439.
- Chiu RW, Chan KC, Gao Y, *et al.* Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2008;105:20458–63.
- Chiu RW, Sun H, Akolekar R, *et al.* Maternal plasma DNA analysis with massively parallel sequencing by ligation for noninvasive prenatal diagnosis of trisomy 21. *Clin Chem* 2010;56:459–63.
- Chiang DY, Getz G, Jaffe DB, *et al.* High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods* 2009;6:99–103.
- Li R, Yu C, Li Y, *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25:1966–7.
- Aird D, Ross MG, Chen WS, *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 2011;12:R18.
- Windmeijer ACCFAG. An *R*-squared measure of goodness of fit for some common nonlinear regression models. *J Econometrics* 1998;77(2):329–42.
- Greene N. First-trimester sonographic screening for Down syndrome. *Obstet Gynecol* 2004;103:1000; author reply 1.
- Audibert F, Dommergues M, Benattar C, *et al.* Screening for Down syndrome using first-trimester ultrasound and second-trimester maternal serum markers in a low-risk population: a prospective longitudinal study. *Ultrasound Obstet Gynecol* 2001;18:26–31.
- Vissers LE, Stankiewicz P. Microdeletion and microduplication syndromes. *Methods Mol Biol* 2012;838:29–75.

## WHAT'S ALREADY KNOWN ABOUT THIS TOPIC?

- Sequencing-based noninvasive prenatal detection of fetal aneuploidy has been proven to be highly accurate. However, it is still a challenge to detect fetal deletion/duplication syndrome because the interference from maternal DNA in maternal plasma.

## WHAT DOES THIS STUDY ADD?

- Here, we developed a practical bioinformatic methodology to detect fetal chromosomal deletions/duplications of >10Mb using low coverage whole genome sequencing of maternal plasma.