

TECHNICAL ADVANCE

Open Access

Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies

Fuman Jiang^{1†}, Jinghui Ren^{2†}, Fang Chen^{1†}, Yuqiu Zhou³, Jiansheng Xie⁴, Shan Dan⁵, Yue Su⁵, Jianhong Xie³, Baomin Yin³, Wen Su³, Huakun Zhang⁴, Wei Wang¹, Xianghua Chai¹, Linhua Lin², Hui Guo², Qiyun Li², Peipei Li¹, Yuying Yuan¹, Xiaoyu Pan¹, Yihan Li¹, Lifu Liu¹, Huifei Chen¹, Zhaoling Xuan¹, Shengpei Chen¹, Chunlei Zhang¹, Hongyun Zhang¹, Zhongming Tian¹, Zhengyu Zhang¹, Hui Jiang¹, Lijian Zhao¹, Weimou Zheng¹, Songgang Li¹, Yingrui Li¹, Jun Wang¹, Jian Wang¹ and Xiuqing Zhang^{1*}

Abstract

Background: Conventional prenatal screening tests, such as maternal serum tests and ultrasound scan, have limited resolution and accuracy.

Methods: We developed an advanced noninvasive prenatal diagnosis method based on massively parallel sequencing. The Noninvasive Fetal Trisomy (NIFTY) test, combines an optimized Student's t-test with a locally weighted polynomial regression and binary hypotheses. We applied the NIFTY test to 903 pregnancies and compared the diagnostic results with those of full karyotyping.

Results: 16 of 16 trisomy 21, 12 of 12 trisomy 18, two of two trisomy 13, three of four 45, X, one of one XYY and two of two XXY abnormalities were correctly identified. But one false positive case of trisomy 18 and one false negative case of 45, X were observed. The test performed with 100% sensitivity and 99.9% specificity for autosomal aneuploidies and 85.7% sensitivity and 99.9% specificity for sex chromosomal aneuploidies. Compared with three previously reported z-score approaches with/without GC-bias removal and with internal control, the NIFTY test was more accurate and robust for the detection of both autosomal and sex chromosomal aneuploidies in fetuses.

Conclusion: Our study demonstrates a powerful and reliable methodology for noninvasive prenatal diagnosis.

Keywords: Noninvasive Fetal Trisomy (NIFTY) test, Massively parallel sequencing, Autosomal aneuploidies, Sex chromosomal aneuploidies

Background

Down syndrome (Trisomy 21), Edward syndrome (Trisomy 18) and Patau syndrome (Trisomy 13) are the most clinically significant autosomal aneuploidies, and the incidence of autosomal abnormalities can be as high as one in 160 live births [1]. Turner's syndrome (45, X), Klinefelter's syndrome (47, XXY) and XYY syndrome are common sex chromosomal aneuploidies that are

associated with reproductive loss, infertility and language development delays, among others [2-4]. Sex chromosomal aneuploidies occur in one out of 500 male births and one out of 850 female births [5-8].

Conventional prenatal diagnostic methods for detecting aneuploidies, such as karyotyping, FISH and QF-PCR, which rely on invasive procedures, bear potential risks for miscarriage [9,10]. Noninvasive screening for fetal aneuploidies using maternal serum markers and ultrasound scans entails less risk, but offers limited sensitivity and specificity [11,12].

* Correspondence: zhangxq@genomics.org.cn

†Equal contributors

¹BGI- Shenzhen, Shenzhen, China

Full list of author information is available at the end of the article

When Lo et al. first reported cell-free fetal DNA (cff-DNA) in 1997 [13], and they highlighted its potential clinical utility as a biomarker because it can be detected from as early as four gestational weeks. Cell-free fetal DNA clears rapidly from the maternal circulation after delivery [14-16]. However, the fraction of fetal DNA in the maternal plasma varies from 5% to 10%, which makes it difficult to detect genetic variation in the fetus [17,18]. Conventional molecular techniques, such as allele-specific polymerase chain reaction (PCR) or quantitative real-time PCR, which aim to detect fetal chromosomal disorders, focus only on specific populations [19-21], such as fetuses with heterozygous alleles. The recent rapid development of massively parallel sequencing (MPS) technology now makes it possible to noninvasively detect fetal aneuploidies in a clinical setting [22-24]. Several recent studies demonstrated that fetal aneuploidies could be detected and quantified via high-throughput whole-genome sequencing of maternal plasma cell-free DNA combined with a standard z-score test. Prior studies by Chiu et al. and Ehrich et al. suggest that an MPS-based approach is reliable at detecting trisomy 21 [25,26].

In principle, an MPS-based approach that resolves whole genome information should be applicable for detecting aneuploidies in all of the chromosomes. Chen et al. showed, however, that such test was less successful for detecting trisomy 18 and trisomy 13 compared with trisomy 21. The mixed results may be related to the GC-bias caused by the sample preparation or sequencing procedures [27]. Quake et al. developed a method to remove the effect of GC-bias, and thus significantly improve the sensitivity of the MPS-based approach for detecting trisomy 18 and trisomy 13 [28]. Another recent study reported the possibility of detecting sex chromosomal aneuploidies using an internal chromosome control approach [29].

In this study, we developed an advanced GC-correlation methodology for an MPS-based, noninvasive fetal trisomy (NIFTY) test. Our technique has higher sensitivity and specificity than all previously reported z-score approaches for the detection of autosomal and sex chromosomal aneuploidy.

Results

Study participants and data production

We enrolled 903 pregnant women with ages ranging from 20 to 45 years. The gestational ages varied from 10 to 34 weeks, covering the first to the third trimesters. Based on the results of full karyotyping using amniotic fluid, 866 of the fetuses were euploid and 37 were aneuploid. The cases of aneuploidy included two cases of trisomy 13, 12 cases of trisomy 18, 16 cases of trisomy 21, four cases of 45,X (three typical cases of 45,X and one mosaic case of 45, X ([27]/46, XX [23])), two cases of

XXY and one case of XYY. We obtained 2-4 million reads for each sample. After alignment and filtering, the average data volume for aneuploidy detection was 1.7 million uniquely aligned reads. We constructed a comprehensive bioinformatics pipeline to scan for fetal chromosomal aneuploidies. The pipeline comprised short reads alignment, GC content correction, fetal DNA concentration estimation, t-test of a binary hypothesis, and fetal gender classification (Figure 1).

The relationship between GC content and sequencing bias

To investigate the relationship between GC content and sequence bias, we selected 300 control pregnancies with normal karyotypes and plotted the relative reads coverage for each chromosome against the corresponding GC content (Additional file 1: Figure S1). The relative reads coverage of the different chromosomes was strongly related to the inherent chromosomal GC content, and the correlation between the two factors varied among the chromosomes. We observed a significant positive correlation between the reads coverage and GC content for chromosomes with an average GC content greater than 41%, whereas we observed, a significant negative correlation between the reads coverage and GC content for chromosomes with an average GC content less than 41%. Among the chromosomes with average GC content close to 41% the reads coverage was not correlated with GC content (Additional file 2: Figure S2). To further investigate the effect of GC content on reads coverage, we examined the hidden relationship between chromosome structure and inherent GC content. We classified all of the unique 35-mers in the genome into 36 levels based on the numbers of guanine (G) and cytosine (C) bases, ranging from 0 to 35. We used the 35-mer counts to cluster the chromosomes according to their GC levels within a matrix (Additional file 3: Figure S3). Chromosomes 19 and 22 clustered together because of their higher inherent GC contents, while chromosome 4 and 13 clustered together for their lower inherent GC contents. The differences in the inherent GC content of the chromosomes combined with the sequencer-related GC-bias explained the significant correlation between reads coverage and corresponding GC content. For example, chromosome 13 has a relatively low GC content, the PCR and sequencing process enriched chromosomes with higher GC content, leading to relatively low reads coverage for chromosome 13 and thus a negative correlation between the reads coverage and GC content among the chromosomes (Additional file 1: Figure S1).

The relationship between cell-free fetal DNA concentration and gestational week

Previous work demonstrated that the cff-DNA concentration was the lynchpin in fetus aneuploidy detection

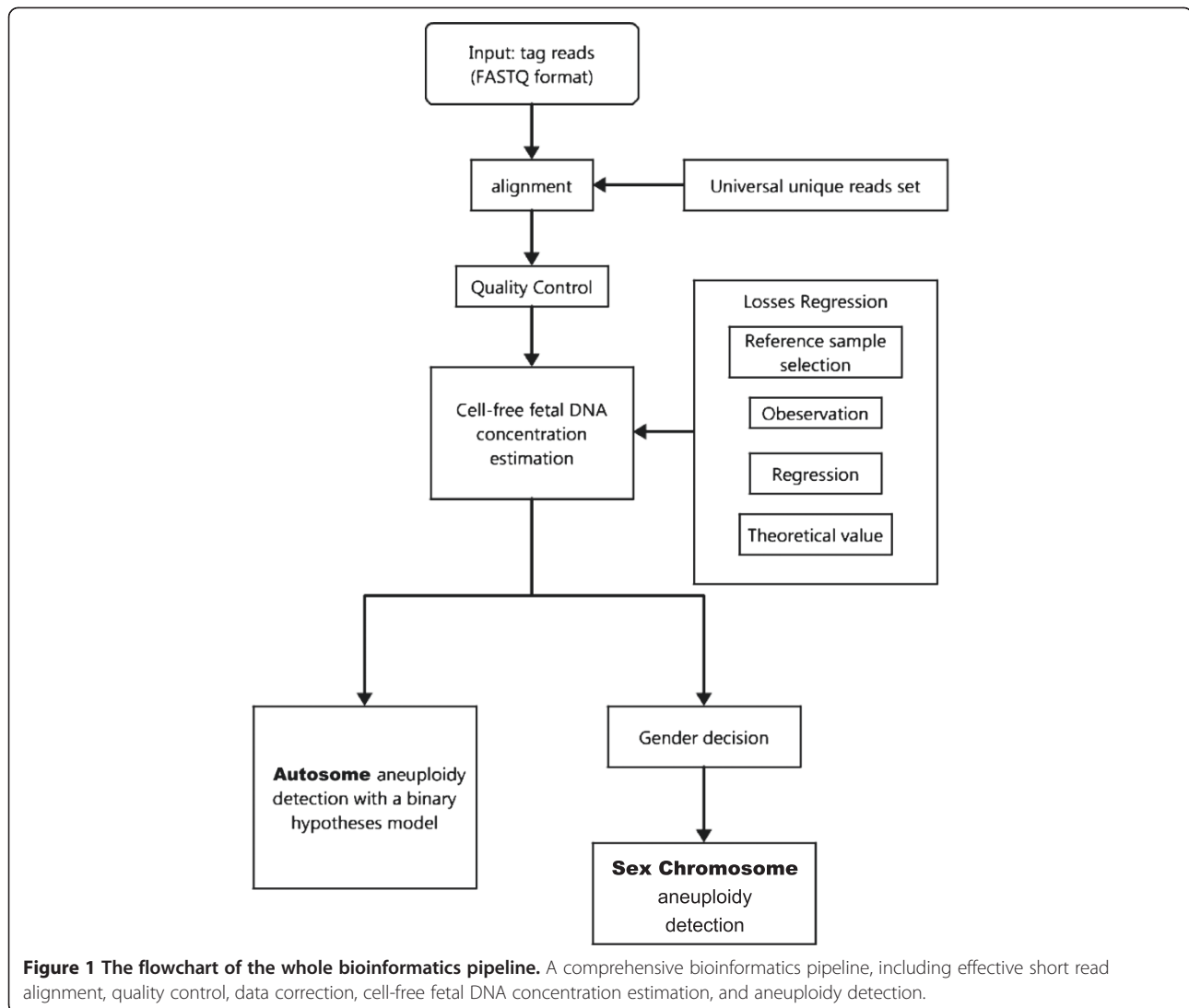


Figure 1 The flowchart of the whole bioinformatics pipeline. A comprehensive bioinformatics pipeline, including effective short read alignment, quality control, data correction, cell-free fetal DNA concentration estimation, and aneuploidy detection.

[16]. To assess the best strategy for clinical applications, we examined the relationship between cell-free fetal DNA concentration and the gestational week. We examined the 443 plasma samples with male fetal to appraise the approximate probability relationship. Using Losses regression, we found that the amount of cff-DNA increased significantly with the gestational week (Additional file 4: Figure S4). The correlation coefficient of the linear regression was only 0.1246, however, indicating that there might be a more complex mechanism driving the cff-DNA concentrations.

Quantitative description of data volatility and tags number

The volatility of the relative reads coverage was one of the major factors affecting the sensitivity and specificity of aneuploidy detection. To quantify the volatility of the relative reads coverage, we used the standard deviation of the difference between the observed and fitted relative

reads coverage. We found that for each chromosome the standard deviation was stable when the numbers of samples was larger than 100 (Additional file 5: Figure S5). We also found that the depth of sequencing strongly influenced the accuracy of aneuploidy detection. We isolated 150 plasma samples with euploid fetuses to inspect the relationship between the tags number (unique reads) and the standard deviation of relative reads coverage. On each chromosome, the standard deviation of relative reads coverage among the 150 samples was significantly correlated with the numbers of tags (Additional file 6: Figure S6). We further estimated the effects of the gestational week and the number of tags on the power of our statistical method to detect fetal aneuploidy in chromosomes 13, 18, 21 and X (Additional file 7: Figure S7). In most cases, the detection power increased with both gestational week and the number of sequencing reads. We also found that the detection power was higher when

the fetus was male; it may be due to the more accurate estimation of cff-DNA concentration.

Robust data quality control of the GC-correlation t-test

Several indicators were used to judge the quality of the sequence data. We classified these indicators into two categories: direct and indirect. The indirect indicators of the accuracy of NIFTY test came from the sequencing procedure: Q20% refers to the fraction of bases within the sequenced reads with an Illumina quality score greater than 20, and the PCR duplication rate, refers to the fraction of the reads sharing the same start position and end positions on the reference genome. The direct indicators came from the data analysis procedure and included the number of unique reads (Figure 2), the genome-wide average GC content, and the consistency between the test samples and the reference controls.

Aneuploidy detection with NIFTY test

The NIFTY test performed with 100% sensitivity and specificity for the detection of trisomy 13 (two out of two) (Figure 3a) and trisomy 21 (16 out of 16) (Figure 3c). For trisomy 18, the NIFTY test detected 12 of 12 cases and identified 890 of 891 healthy controls (Figure 3b), indicating 100% sensitivity and 99.7% specificity, corresponding to zero false negative results and a false positive rate of 0.3%. The false positive occurred in a sample from gestational week 21. Our GC-correlation t-test approach correctly detected sex chromosomal abnormalities. For Turner's syndrome, the NIFTY test identified three out of

four XO cases but failed to detect the mosaic 45, X case which was in gestational week of 25 and had a normal karyotype in 46% of the cells sampled. Thus, the sensitivity and specificity of our approach for the detection of Turner's syndrome were 75% and 99.9%, respectively; in other words, the false negative rate was 25% and false positive rate was 0.1% for 45, X detection using the NIFTY test. The test performed with 100% sensitivity and specificity for the detection of XXY (two out of two) or XYY (one out of one) (Figure 3e and f).

The NIFTY test correctly identified the sex of approximately 99.9% of the 896 fetuses, 443 male and 452 female, which did not have sex chromosomal aneuploidies. The NIFTY test was inconclusive for one fetus that was determined to be 46, XX by karyotyping.

Comparison between different aneuploidy detection approaches

To evaluate the performance of the NIFTY test in the detection of fetal aneuploidy, we compared it with the performance of three other previously reported approaches to analyse our 903 cases, with full karyotyping of the same 300 euploid cases [23,27,29]. Chiu et al. used the standard z-score approach without any GC-bias removal to detect Down syndrome [23]. Chen et al. developed a z-score approach with a different GC-bias removal strategy [27], which we named the "GC-correct z-score approach." Lau et al. previously demonstrated a internal chromosome control based z-score approach [29].

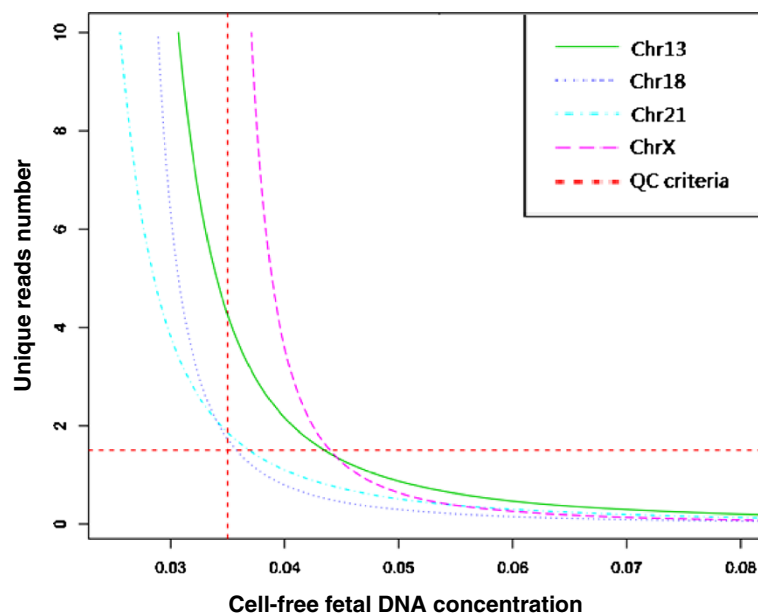


Figure 2 The required number of of unique reads for high sensitivity across different cff-DNA concentrations. For aneuploidy detection, the required number of unique reads (y-axis) increased with decreasing cff-DNA concentration (x-axis). For a 3.5% cff-DNA concentration, 1.7 million unique reads are needed to obtain high sensitivity.

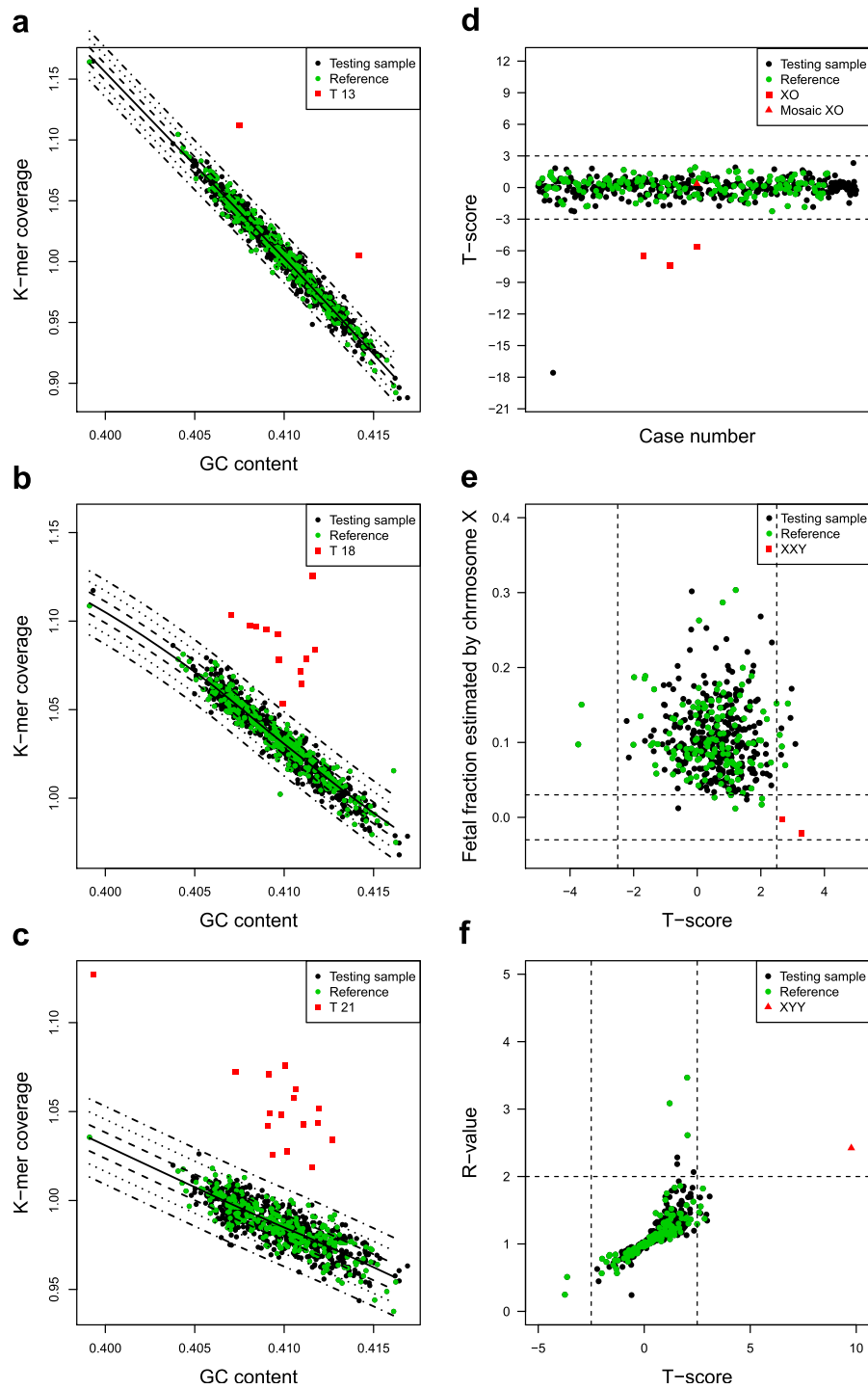


Figure 3 Fetal aneuploidy detection using NIFTY test. a-c. The k-mer coverage (y-axis) of 903 samples was plotted with corresponding GC content (x-axis) for chromosome 13,18 and 21. The solid black line is the fit between the k-mer coverage and GC content among the 300 controls. The dot-dash lines from inside to outside are the contour lines of $t=1$, $t=2$ and $t=3$ respectively. **d.** XO detection. The t-score of chromosome X for 452 cases with female fetuses and 4 XO cases is dotted. The t-score less than -2.5 indicates XO aneuploidy. **e.** XXY detection. The x-axis is the t-score of chromosome X for samples carrying male fetuses. The y-axis is the fetal fraction estimated by chromosome X. Red square points indicated XXY cases that have a t-score larger than 2.5 and the cff-DNA concentration estimated by chromosome X nearly equal to zero. **f.** XYY detection. The x-axis is t-score for chromosome X among samples carrying male fetuses. The y-axis is the R-value, i.e. the ratio of the fetal DNA fraction estimated by chromosome Y to that estimated by chromosome X. Red triangle points indicate XYY cases with t-score greater than 2.5 and R-value greater than 2. The case types are color coded (black: testing samples; green: reference samples; red: aneuploidy samples).

We used the coefficient of variation (CV) to evaluate the performances of these four approaches (Figure 4). Additionally, we found that the CV for the standard z-score approach was larger than that for other approaches among clinically relevant chromosomes (13, 18 and 21). Thus, the standard z-score approach has a low sensitivity for the detection of trisomies 13 and 18 (Table 1). The performance of the GC-correct z-score approaches and our NIFTY test were close, both demonstrated over 99% sensitivity and specificity for the detection of trisomy 13, 18 and 21 (Table 1). It was difficult to precisely detect sex chromosomal aneuploidy using the GC-correct z-score approach due to fetal gender confusion. The internal chromosome control approach displayed larger CV values for chromosomes 13, 18, and 21 and had a higher risk of false negatives related to XXY and XYY detection. In contrast, the NIFTY test had increased accuracy in the detection of sex chromosomal aneuploidies, such as XO, XXY and XYY (Table 1).

Discussion

The cost of high throughput sequencing decreased dramatically over the past few years, thus increasing its utility for clinical practices [30,31]. In this study, we demonstrated the NIFTY test is a novel MPS-based method for

noninvasively detecting fetal aneuploidies. The NIFTY test proved to be a reliable and timely method for detecting both autosomal and sex chromosomal aneuploidies, especially trisomy 21, trisomy 18, trisomy 13, XO, XXY and XYY.

GC-bias is a common issue for applications using the current massively parallel sequencing platforms, and it can be introduced either by sample preparation or the sequencing procedure [32,33]. In this study, we observed that sequencers appear to have a preferred GC ranges, and the differences in GC composition among different chromosomes can act as an intrinsic factor influencing the degree of data fluctuation. We employed GC-correlation to significantly reduce the effect of GC-bias. Using the same reagents for library construction and sequencing, and improving the image analysis software may be useful in further minimizing the GC-bias.

The small fetal DNA fraction in the maternal blood was the main limitation on sequencing dependent non-invasive prenatal diagnosis. The required number of unique reads increases exponentially when the concentration of cff-DNA falls to less than 3.5%, which is consistent with the results of a previous study [27]. An advantage of our approach is that the quality-control procedure uses the estimation of cff-DNA concentration as a key index. Thus, the quality-control procedure, improves the accuracy of

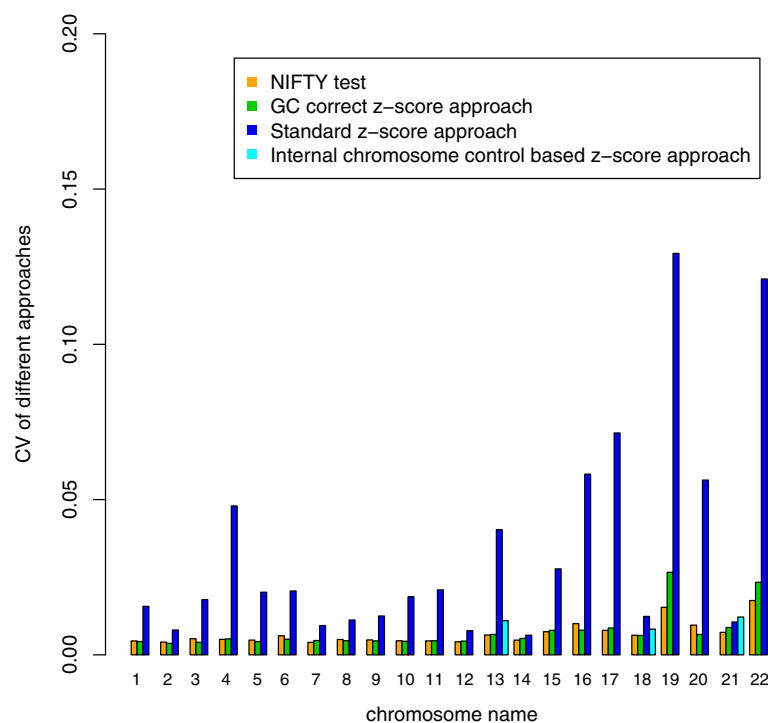


Figure 4 The Performance of three methods: coefficient of variation (CV). We calculated the CVs (y-axis) among the different chromosomes (x-axis) for the 903 samples with karyotyping. The different methods are the color-coded (Orange: NIFTY test; Green: GC correct z-score approach; Dark Blue: Standard z-score approach; Light Blue: Internal chromosome control based z-score approach). In the clinically interesting chromosomes (13, 18, 21), our approach obtained the lowest CVs, indicating a higher sensitivity.

Table 1 The performance of four approaches for detection of fetal aneuploidy

Test		Standard z-score approach		GC correct z-score approach		Internal chromosome control based z-score approach		NIFTY test	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Autosome	T13 (2)	50%(1/2)	99.9%(900/901)	100%(2/2)	100%(901/901)	100%(2/2)	99.7%(898/901)	100%(2/2)	100%(901/901)
	T18 (12)	91.7%(11/12)	100%(891/891)	100%(12/12)	99.9%(890/891)	100%(12/12)	100%(891/891)	100%(12/12)	99.9%(890/891)
	T21 (16)	93.7%(15/16)	100%(887/887)	100%(16/16)	100%(887/887)	100%(16/16)	100%(887/887)	100%(16/16)	100%(887/887)
Sex chromosome*	45, X (3 45, X, 1 45,X/46,XX)	Not available	Not available	Not available	Not available	75%(3/4)	99.8%(897/899)	75%(3/4)	99.9%(898/899)
	XXX (1)	Not available	Not available	Not available	Not available	0%(0/1)	100%(902/902)	100%(1/1)	100%(902/902)
	XXY (2)	Not available	Not available	Not available	Not available	0% (0/2)	100%(901/901)	100%(2/2)	100%(901/901)
Binary hypothesis		Not available		Not available		Available		Available	

* We could not perform sex chromosome aneuploidy detection for one case because of the failure in gender classification.

fetal aneuploidy detection. Another improvement of our methodology over previous methods is the change from sole reliance on a binary hypothesis to a more comprehensive statistical model. Our statistical model strengthens the theoretical sensitivity of the test; the NIFTY test performed with 100% sensitivity and 99.9% specificity for detecting autosomal aneuploidies and 85.7% sensitivity and 99.9% specificity for detecting sex chromosomal aneuploidies.

Using the NIFTY test we treated each chromosome as a whole, which allowed us to focus on detecting the aneuploidies. In principle, we could detect other chromosomal abnormalities resulting in serious clinical consequences, such as microdeletions and microduplications, by slicing the chromosomes into smaller fragments and increasing the number of sequencing tags. It is also possible to detect mutations, such as those underlying Mendelian diseases, in genes or regions of interest through target region capture and high depth sequencing [34].

Although we achieved high detection accuracy in a cohort of 903, the sample size in this study was a limiting factor because the incidence of aneuploidies in the general population is low. To precisely estimate sensitivity and specificity of our procedure, large-scale, multi-center clinical trials will be required in the future. Additionally, the conventional approaches for cff-DNA concentration estimation are mostly locus-specific and only applicable to limited population [20,35]. Our approach was also less accurate in assessing the cff-DNA concentration for female fetuses. Further studies should focus on developing an unbiased method to precisely estimate the fraction of cff-DNA in the maternal plasma.

Conclusions

In this study we demonstrate a robust and accurate methodology to detect fetal aneuploidies using MPS. This is the first study to systematically identify sex chromosomal aneuploidies with maternal plasma DNA sequencing. We hope the use of this method in clinical practices will contribute to reducing the number of birth defects.

Methods

Sample recruitment

From June 2009 to August 2010, we recruited a total of 903 participants prospectively from the Shenzhen People's Hospital, the Zhuhai Municipal Maternal and Child Healthcare Hospital and the Shenzhen Maternal and Child Care Center. We recruited another 19 euploid adult males for the estimation of fetal DNA fraction. Institutional Review Board approval was obtained at each site, and all participants gave informed written consent. We obtained the full karyotyping results for all samples

from regular clinical tests. We randomly selected 300 euploid samples among the karyotyping results to use as the reference controls.

Maternal plasma DNA sequencing

We collected five ml peripheral venous blood from 903 pregnant women in EDTA tubes. The tubes were centrifuged at $1,600 \times g$ for 10 min within four hours of collection. Plasma was transferred to microcentrifuge tubes and centrifuged at $16,000 \times g$ for 10 min to remove residual cells. Cell-free plasma was stored at -80°C until DNA extraction. Each plasma sample was frozen and thawed only once.

For massively parallel genomic sequencing, DNA fragments from 600 ul of maternal plasma were used for library construction according to a modified protocol from Illumina. End-repairing of maternal plasma DNA fragments was performed using T4 DNA polymerase, Klenow polymerase, and T4 polynucleotide kinase. Afterwards, A-base tailing adapters were ligated to the DNA fragments. Standard multiplex primers were introduced by 17-cycle PCR. The libraries were analysed for size distribution by Agilent Bioanalyzer and quantified using real-time PCR. Thirty-six-cycle single-end multiplex sequencing and 50-cycle single-end multiplex sequencing were used for the Illumina GAIIx and Illumina HiSeq 2000 platform, respectively.

High effective alignment with universal unique reads set

Computationally, we incised the human reference genome (HG 18, NCBI build 36) into k -mers (k refers to the length of the sequencing reads) and then aligned the k -mers back to the reference genome. All of the k -mers that could be uniquely mapped to a single position on the reference genome, the unique mapping reads, were named as the universal unique reads set. We selected the sequencing reads that could be mapped with 0-mismatch to the universal unique reads set (i.e. the tag) for our analysis.

K-mer coverage and GC-correlation

We computed the k -mer coverage for each chromosome and every sample, as $C_{i,j} = \frac{n_{i,j}}{N_i}$ where i is the ID of control samples; j is the chromosome ID; $n_{i,j}$ is the number of unique reads mapped onto chromosome j from sample i and $N_{i,j}$ was the total number of unique reads for chromosome j . Because of the differences among the samples, we normalized the data and computed the relative k -mer coverage for each sample as $r_{i,j} = \frac{C_{i,j}}{\bar{C}_i}$, where $\bar{C}_i = \frac{1}{22} \sum_{j=1}^{22} C_{i,j}$ was the average k -mer coverage of the 22 autosomes in the i -th sample.

Given the unclear mechanism of GC-bias, we performed a Losses regression to fit the relative k -mer coverage to the corresponding GC content. We denoted the fitted relative k -mer coverage as $cr'_{i,j} = f_j(GC_{i,j})$. The fitted value, which we used as the theoretical value, was vital to our statistical model for cff-DNA concentration estimation and aneuploidy detection.

Because we using a male/female data set, we had different fitted values for the analysis of sex chromosomes. We calculated the fitted relative k -mer values for the sex chromosome analysis as follows:

$$cr'_{i,j,m} = f_{j,m}(GC_{i,j}) \cdot (j = X, Y), \text{ for the fitted relative } k\text{-mer coverage from a regression of an adult male data set; and}$$

$$cr'_{i,j,f} = f_{j,f}(GC_{i,j}) \cdot (j = X, Y), \text{ for the fitted relative } k\text{-mer coverage from a regression of a fetal-female data set.}$$

Cff-DNA concentration estimation

Using the gender difference to compute the relative k -mer coverage of the sex chromosome, we estimated the cff-DNA concentrations, which denote as ϵ . Subscripts corresponding to chromosome IDs indicate concentrations estimated from different chromosomes:

$$\epsilon_{i,Y} = \frac{cr_{i,Y} - cr'_{i,Y,f}}{cr'_{i,Y,m} - cr'_{i,Y,f}}, \text{ is the estimation using the data for chromosome Y; and}$$

$$\epsilon_{i,X} = \frac{cr_{i,X} - cr'_{i,X,f}}{cr'_{i,Y,m} - cr'_{i,Y,f}}, \text{ is the estimation using data for chromosome X.}$$

Autosomal aneuploidy detection with binary hypothesis

We developed a binary hypothesis strategy to achieve a higher sensitivity and specificity. We performed two Student's t -test based on null/alternative hypotheses, and we subsequently calculated the relative logarithmic likelihood odds ratio. The null and alternative hypotheses are shown below.

For the first test:

H0 (null hypothesis): the fetal chromosome was euploid.
 H1 (alternative hypothesis): the fetal chromosome was trisomic.

$$\text{The first } t\text{-value, } t_{i,j,first} = \frac{cr_{i,j} - cr'_{i,j}}{sd_j}.$$

For the second test:

H0 (null hypothesis): the test fetal chromosome was trisomic.
 H1 (alternative hypothesis): the test fetal chromosome was euploid.

$$\text{The second } t\text{-value, } t_{i,j,second} = \frac{cr_{i,j} - cr'_{i,j}(1 + \epsilon_i/2)}{sd_j}.$$

The logarithmic likelihood odds ratio between our binary hypotheses was defined as

$$L_{i,j} = \log \left(\frac{p(t_{i,j,first}, DOF|H_0)}{p(t_{i,j,second}, DOF|H_1)} \right)$$

where DOF = the degree of freedom., We used, $|t_{i,j,first}| > 3$ and $|t_{i,j,second}| < 3$ as warning criteria. From the logarithmic likelihood odds ratio, we could make a confident judgment of autosomal aneuploidy if $L_{i,j} > 1$.

Fetal gender classification and sex chromosomal aneuploidy detection

We developed a double standard strategy with an experimental threshold and logistic regression to detect the fetal gender. The k -mer coverage on chromosome Y was an ideal choice for distinguishing genders. Based on the 300 reference controls, we considered $cr_{i,Y} < 0.04$ the threshold for identifying a female fetus, while we regarded samples with $cr_{i,Y} > 0.051$ as having a male fetus. We considered samples with $0.04 < cr_{i,Y} < 0.051$ to be gender-uncertain.

Additionally, we developed a logistic regression strategy to improve the specificity of the gender determination. We computed the probability (P_i) of that a fetus was male by the following formula:

$$\text{logit}(p_i) = \ln \left(\frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 cr_{i,X} + \beta_2 cr_{i,Y}, \text{ where the parameters } (\beta_0, \beta_1, \beta_2) \text{ were determined by regression using the 300 reference controls mentioned above.}$$

We regarded samples with $p_i > 0.8$ as having male fetuses, samples with $p_i < 0.3$ as having female fetuses, and the remaining samples as being gender-uncertain.

After gender classification, we performed XXX and XO detection on samples with a female fetus and XXY and XYY detection on samples with a male fetus.

For samples with a female fetus, we performed a t -test for chromosome abnormality detection.

$t_{i,X} = \frac{cr_{i,X} - cr'_{i,X,f}}{sd_{X,f}}$ where $sd_{X,f}$ is the standard deviation of $cr_{i,X,f} - cr'_{i,X,f}$ calculated from the reference controls with female fetuses; we expected $sd_{X,f}$ to equal zero. We considered samples with $t_{i,X}$ or $t_{i,X} < -2.5$ to be XXX or XO.

For a male fetus, we first supposed that chromosome Y is monosomic and extrapolated the fitted k -mer coverage for chromosome X, with the fetal DNA fraction estimated only by the k -mer coverage of chromosome Y. We calculated the t -score by the following formula:

$t_i = \frac{cr_{i,X} - cr'_{i,X,f} \cdot (1 - \frac{\epsilon_{i,Y}}{2})}{sd_{X,f}}$, where $\epsilon_{i,Y}$ is the estimated cff-DNA concentration using chromosome Y data, and is the standard deviation of $cr_{i,X,f} - cr'_{i,X,f}$ calculated from the reference controls carrying female fetuses with an expectation of zero. Both of these quantities are defined above.

We regarded samples with $t_i > 2.5$ as being XXY or XYY. Additionally, the cff-DNA concentration estimated

by chromosome X and Y independently is a combined marker for sex chromosomal aneuploidy detection especially XXY and XYY. For an XXY sample, not only was the $t_i > 2.5$ but also the cff-DNA concentration estimated by chromosome X was nearly zero, with a confidence interval from -0.03 to 0.03 ; For an XYY samples, not only the $t_i > 2.5$, but the R-value (Ratio of the cff-DNA concentration estimated by chromosome Y to that estimated by chromosome X) was nearly two, reflecting the fact that there were two copies of chromosome Y and only a single copy of chromosome X.

Additional files

Additional file 1: Figure S1. The correlation between sequence GC content and relative k-mer coverage. We plotted the relative k-mer coverage of each chromosome (y-axis) among our 300 controls against the corresponding sequence GC content (x-axis). Red plot are for female fetuses and black plot are for male fetuses.

Additional file 2: Figure S2. The tendency between normalized k-mer coverage and corresponding GC content. The GC content of each chromosome is listed in orders. The blue bars refer to the chromosomes that have a negative correlation between the k-mer coverage and the GC content, green bar refer chromosomes that have a positive correlation between the k-mer coverage and the GC content, and yellow bar refer to the chromosomes with no correlation.

Additional file 3: Figure S3. The reconstructed relationship between chromosomes by GC content. We reconstructed the GC-content relationship between the different chromosomes by clustering the 35-mer counts for 36 GC levels (y-axis) on the different chromosomes (x-axis). The normalized 35-mer counts, as a percentage of each chromosome, are color-coded in the heat map.

Additional file 4: Figure S4. The relationship between estimated cff-DNA concentration and gestational week. The black dots represent the cff-DNA concentrations (y-axis) plotted against the corresponding gestational week (x-axis) for the 443 samples with male fetuses.

Additional file 5: Figure S5. The standard deviation and sample numbers. The standard deviations of the difference between the observed and fitted k-mer coverage (y-axis) for different numbers of samples (x-axis). Different chromosomes are colour-coded. The standard deviation becomes stable when the number of samples is larger than 100.

Additional file 6: Figure S6. The relationship between tags number and the standard deviation of relative k-mer coverage among 150 samples. The standard deviations of the relative k-mer coverage (y-axis) declines with the increasing number of tags (x-axis) from 0.5 to 3.5 million for each chromosome.

Additional file 7: Figure S7. The aneuploidy detection power estimation. The colored contour lines show the aneuploidy detection power at different gestational weeks (x-axis) and with different numbers of unique reads (y-axis). Fetal genders are shown separately. The power is much higher when the fetus is male.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

ZXQ, RJH, WJI and WJU conceived and designed the study; WW, ZYQ, SD, SY, ZHK, XJS, SW, YBM, ZLJ, XJH, ZZY, and LLH recruited participants and separated the plasma samples; CF, LYH, PXY, XZL, ZHY, TZM, LPP, LQY, GH, and JH performed library construction and sequencing; JFM, ZWM, CXH, LYR and LSG developed the methodology and LLF, CHF, ZCL and YYY participated in the alignment of the sequencing data and data analysis; JFM and CSP wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This project is supported by Key Laboratory Project in Shenzhen, CXB200903110066A and CXB201108250096A. We sincerely thank our colleagues at the BGI-Shenzhen for sequencing. We thank Dr. Lars Bolund and Dr. TK Lau for excellent advice on the project.

Author details

¹BGI- Shenzhen, Shenzhen, China. ²The Center of Prenatal Diagnosis, Shenzhen People's Hospital, 2nd Clinical Medical College of Jinan University, Shenzhen, Guangdong, China. ³Zhuhai Institute of Medical Genetics, Zhuhai Municipal Maternal and Child Healthcare Hospital, Zhuhai, China. ⁴Central for Prenatal Diagnosis, Shenzhen Maternity and Child Healthcare Hospital, Affiliated Southern Medical University, Shenzhen, China. ⁵Department of Perinatology, Beijing Obstetrics and Gynecology Hospital-Capital University of Medical Sciences, Beijing, China.

Received: 23 December 2011 Accepted: 24 November 2012

Published: 1 December 2012

References

1. Driscoll DA, Gross S: **Clinical practice. Prenatal screening for aneuploidy.** *N Engl J Med* 2009, **360**(24):2556–2562.
2. Bouchlariotou S, Tsikouras P, Dimitraki M, Athanasiadis A, Papoulidis I, Maroulis G, Liberis A, Liberis V: **Turner's syndrome and pregnancy: has the 45, X/47, XXX mosaicism a different prognosis? Own clinical experience and literature review.** *J Matern Fetal Neonatal Med* 2011, **24**(5):668–672.
3. Leggett V, Jacobs P, Nation K, Scerif G, Bishop DV: **Neurocognitive outcomes of individuals with a sex chromosome trisomy: XXX, XYY, or XXY: a systematic review.** *Dev Med Child Neurol* 2010, **52**(2):119–129.
4. Hofherr SE, Wiktor AE, Kipp BR, Dawson DB, Van Dyke DL: **Clinical diagnostic testing for the cytogenetic and molecular causes of male infertility: the Mayo Clinic experience.** *J Assist Reprod Genet* 2011, **28**(11):1091–1098.
5. Rives N, Simeon N, Milazzo JP, Barthelemy C, Mace B: **Meiotic segregation of sex chromosomes in mosaic and non-mosaic XYY males: case reports and review of the literature.** *Int J Androl* 2003, **26**(4):242–249.
6. Park JH, Burns-Cusato M, Dominguez-Salazar E, Riggan A, Shetty S, Arnold AP, Rissman EF: **Effects of sex chromosome aneuploidy on male sexual behavior.** *Genes Brain Behav* 2008, **7**(6):609–617.
7. Tartaglia NR, Howell S, Sutherland A, Wilson R, Wilson L: **A review of trisomy X (47, XXX).** *Orphanet J Rare Dis* 2010, **5**:8.
8. Sybert VP, McCauley E: **Turner's syndrome.** *N Engl J Med* 2004, **351**(12):1227–1238.
9. Mujezinovic F, Alfirevic Z: **Procedure-related complications of amniocentesis and chorionic villous sampling: a systematic review.** *Obstet Gynecol* 2007, **110**(3):687–694.
10. Pitukijronnakorn S, Promsonthi P, Panburana P, Udomsuppayakul U, Chittacharoen A: **Fetal loss associated with second trimester amniocentesis.** *Arch Gynecol Obstet* 2011, **284**(4):793–797.
11. Kagan KO, Wright D, Valencia C, Maiz N, Nicolaides KH: **Screening for trisomies 21, 18 and 13 by maternal age, fetal nuchal translucency, fetal heart rate, free beta-hCG and pregnancy-associated plasma protein-A.** *Hum Reprod* 2008, **23**(9):1968–1975.
12. Canick J: **Prenatal screening for trisomy 21: recent advances and guidelines.** *Clin Chem Lab Med* 2011, **50**(6):1003–1008.
13. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS: **Presence of fetal DNA in maternal plasma and serum.** *Lancet* 1997, **350**(9076):485–487.
14. Simpson JL: **Choosing the best prenatal screening protocol.** *N Engl J Med* 2005, **353**(19):2068–2070.
15. Bischoff FZ, Sinacori MK, Dang DD, Marquez-Do D, Horne C, Lewis DE, Simpson JL: **Cell-free fetal DNA and intact fetal cells in maternal blood circulation: implications for first and second trimester non-invasive prenatal diagnosis.** *Hum Reprod Update* 2002, **8**(6):493–500.
16. Illanes S, Denbow M, Kailasam C, Finning K, Soothill PW: **Early detection of cell-free fetal DNA in maternal plasma.** *Early Hum Dev* 2007, **83**(9):563–566.
17. Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM: **Rapid clearance of fetal DNA from maternal plasma.** *Am J Hum Genet* 1999, **64**(1):218–224.
18. Birch L, English CA, O'Donoghue K, Barigye O, Fisk NM, Keer JT: **Accurate and robust quantification of circulating fetal and total DNA in maternal plasma from 5 to 41 weeks of gestation.** *Clin Chem* 2005, **51**(2):312–320.

19. Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, Wainscoat JS, Johnson PJ, Chang AM, Hjelm NM: **Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis.** *Am J Hum Genet* 1998, **62**(4):768–775.
20. Lo YM, Tsui NB, Chiu RW, Lau TK, Leung TN, Heung MM, Gerovassili A, Jin Y, Nicolaides KH, Cantor CR, et al: **Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection.** *Nat Med* 2007, **13**(2):218–223.
21. Tsui NB, Chiu RW, Ding C, El-Sheikhah A, Leung TN, Lau TK, Nicolaides KH, Lo YM: **Detection of trisomy 21 by quantitative mass spectrometric analysis of single-nucleotide polymorphisms.** *Clin Chem* 2005, **51**(12):2358–2362.
22. Tong YK, Ding C, Chiu RW, Gerovassili A, Chim SS, Leung TY, Leung TN, Lau TK, Nicolaides KH, Lo YM: **Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: theoretical and empirical considerations.** *Clin Chem* 2006, **52**(12):2194–2202.
23. Chiu RW, Chan KC, Gao Y, Lau VY, Zheng W, Leung TY, Foo CH, Xie B, Tsui NB, Lun FM, et al: **Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma.** *Proc Natl Acad Sci USA* 2008, **105**(51):20458–20463.
24. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR: **Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood.** *Proc Natl Acad Sci USA* 2008, **105**(42):16266–16271.
25. Ehrlich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasian L, Tim R, Lu V, McCullough R, McCarthy E, Nygren AO, et al: **Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting.** *Am J Obstet Gynecol* 2011, **204**(3):e201–e211. 205.
26. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, Chan KC, Lun FM, Go AT, Lau ET, To WW, et al: **Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study.** *BMJ* 2011, **342**:c7401.
27. Chen EZ, Chiu RW, Sun H, Akolekar R, Chan KC, Leung TY, Jiang P, Zheng YW, Lun FM, Chan LY, et al: **Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing.** *PLoS One* 2011, **6**(7):e21791.
28. 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**(5):e10439.
29. Lau TK, Chen F, Pan X, Pooh RK, Jiang F, Li Y, Jiang H, Li X, Chen S, Zhang X: **Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing.** *J Matern Fetal Neonatal Med* 2011, **25**(8):1370–1374.
30. Brownstein Z, Friedman LM, Shahin H, Oron-Karni V, Kol N, Rayyan AA, Parzefall T, Lev D, Shalev S, Frydman M, et al: **Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in middle eastern families.** *Genome Biol* 2011, **12**(9):R89.
31. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, et al: **Carrier testing for severe childhood recessive diseases by next-generation sequencing.** *Sci Transl Med* 2011, **3**(65):65ra64.
32. Cheung MS, Down TA, Latorre I, Ahringer J: **Systematic bias in high-throughput sequencing data and its correction by BEADS.** *Nucleic Acids Res* 2011, **39**(15):e103.
33. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A: **Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries.** *Genome Biol* 2011, **12**(2):R18.
34. Liao GJ, Lun FM, Zheng YW, Chan KC, Leung TY, Lau TK, Chiu RW, Lo YM: **Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles.** *Clin Chem* 2011, **57**(1):92–101.
35. Purwosunu Y, Sekizawa A, Okai T: **Detection and quantification of fetal DNA in maternal plasma by using LightCycler technology.** *Methods Mol Biol* 2008, **444**:231–238.

doi:10.1186/1755-8794-5-57

Cite this article as: Jiang et al.: Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies. *BMC Medical Genomics* 2012 5:57.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

