Received: 16 October 2018 | Revised: 8 March 2019 | Accepted: 15 April 2019

DOI: 10.1002/mgg3.718

ORIGINAL ARTICLE

Molecular Genetics & Genomic Medicine Open Access WILEY

Effect quantification and value prediction of factors in noninvasive detection for specific fetal copy number variants by semiconductor sequencing

Chunhua Zhang[1] | **Bo Liang**[2] | **Longwei Qiao**[1] | **Liming Xuan**[3] | **Hong Li**[1] | **Quanze He**[1] | **Xiaojuan Wu**[1] | **Jiafeng Lu**[1] | **Bin Yu**[4] | **Ting Wang**[1]

[1]The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, China

[2]State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

[3]Basecare Medical Device Co., Ltd, Suzhou, China

[4]Changzhou Women and Children Health Hospital affiliated to Nanjing Medical University, Changzhou, China

Correspondence
Ting Wang, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou 215000, Jiangsu Province, China.
Email: biowt@163.com

Bin Yu, Changzhou Women and Children Health Hospital affiliated to Nanjing Medical University, Changzhou 213003, Jiangsu Province, China.
Email: ybcz0519@163.com

Funding information
Suzhou Science and Technology Support Program (SYS201649), Suzhou Key Medical Center (SZZX201505), Jiangsu Maternal and Children Health Care Research Project (F201603).

Abstract

Background: The detection limit of noninvasive prenatal testing (NIPT) by next generation sequencing for any given fetal copy number variants (CNV) can be influenced by several factors. In this study, we quantified the effects and predicted the value of parameters for CNVs detection by NIPT.

Methods: Genomic DNA from patient's leucocytes with 3.16 Mb microdeletion in 22q11.21 was mixed with DNA from aborted fetal tissues without CNV at various concentrations by an enzyme digestion method. Abnormal DNA at 0% served as negative control. Sequencing of mixture samples (at 0%, 4%, 12%, and 20%) by Ion Proton Sequencer was performed at flow 500, with WISECONDOR as the pipeline in CNV-calling and bin of 500, 750 and 1,000 kb for counting unique reads. The parameters were evaluated with Box–Behnken design. The region with Z score $\leqq -3$ was marked as a potential microdeletion.

Results: The equation of Z score depending on fetal fraction, unique read number and bin size was obtained by Box–Behnken design. The negative effect was quantified as the coefficient in the equation. The smallest values of these parameters were defined as 4 M unique read number, and 10.08% fetal DNA concentration at bin of 750 kb for detecting subchromosomal microdeletion of 3.16 Mb.

Conclusion: The quantification of effect and value of parameters as well as the method used in this study can benefit the establishment of quality standards for CNVs detection and interpretation of CNVs detection results.

KEYWORDS
fetal copy number variants, microdeletion, next generation sequencing, noninvasive prenatal testing

Zhang and Liang contributed equally to this work.

1 | INTRODUCTION

Fluorescence in situ hybridization (Buysse et al., 2009) and chromosomal microarray analysis (CMA) (Carter, 2007) are widely used classical prenatal diagnosis methods for fetal copy number variants (CNVs) detection. However, invasive procedures (through chorionic villous sampling or amniocentesis) used for these classical prenatal diagnosis methods are associated with pregnancy loss risk (Buysse et al., 2009). Noninvasive prenatal testing (NIPT) by next generation sequencing provides a viable alternative to the existing methods for CNVs detection with higher accuracy and resolution (Zhao, Wang, Wang, Jia, & Zhao, 2013). NIPT used for aneuploidy screening to detect CNV ≥5 Mb shows good performance with sensitivity of 90.0%, while sensitivity for detecting CNV <5 Mb is only 14.3% (Li et al., 2016). The detection limit of NIPT for any given CNV is influenced mostly by fetal fraction, CNV size, coverage, as well as biological and technical variability of the CNV region (Zhao et al., 2015). In this study, using simulated DNA samples, we tried to quantify the effects and predict the value of parameters when NIPT is applied for CNV detection.

What's already known about this topic?

The application of Noninvasive prenatal testing (NIPT) in detection of fetal copy number variation (CNV) has been reported. In general, NIPT in aneuploidy screening could detect the smallest microdeletion of 3 Mb in size, and has good performance for detecting CNV larger than 5 Mb. Fetal fraction, unique read number, size of fetal CNV, and the biological and technical variability of event region are reported as the factors affecting the sensitivity of CNV detection. The limit of CNV detection can be improved by increasing the fraction of abnormal DNA as well as the sequencing depth.

What does this study add?

In this study, the Box–Behnken design was utilized to obtain the equation of *Z* score depending on fetal fraction, unique read number and bin. The negative effect of these parameters was quantified as the coefficient in the equation, through which the smallest values of these parameters could be defined as 4 M unique read number, and 10.08% fetal DNA concentration at bin of 750 kb for detecting subchromosomal microdeletion of 3.16 Mb. Thus, our study quantifies the effects and predicts the value of important parameters in NIPT for fetal CNV detection, which is extremely valuable for clinical application.

2 | MATERIALS AND METHODS

2.1 | Ethical compliance

This study was approved by the Institutional Review Boards of The Affiliated Suzhou Hospital of Nanjing Medical University (No. 011) in May 2016. Written informed consent for this study was obtained from all participants.

2.2 | DNA Preparation

DNA extraction from leukocytes and aborted tissues was performed with QIAGEN DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. To simulate the DNA mixture derived from maternal peripheral blood, genomic DNA obtained from the leukocyte of patients (XY) with 3.16 Mb microdeletion in 22q11.21 [the DiGeorge syndrome chromosome region, Online Mendelian Inheritance in Man 188400] was mixed with DNA from aborted fetal tissues (XX) without CNV at concentrations of 0%, 4%, 12%, and 20% (w/w %) (Table S1). Abnormal DNA at concentration of 0% was used as negative control in each run of sequencing.

2.3 | Setting main parameters of sequencing

The artificial mixture of genomic DNA at different ratios was subjected to gDNA fragment library preparation with Ion Xpress™ Plus Fragment Library Kit (Life Technologies, Foster, CA) according to the manufacturer's protocol (Knierim, Lucke, Schwarz, Schuelke, & Seelow, 2011). The library was subjected to sequencing using an Ion Proton sequencer at 500 flows according to the manufacturer's instructions (Life Technologies, Carlsbad, CA). Raw reads obtained from the Ion Torrent Suite Software were of different lengths and were trimmed according to the sequencing quality value of >15. After raw reads with read length <50 bp was filtered out, the remaining raw reads were aligned to the human reference genome (hg19) using BWA software. Reads that were unmapped, had multiple primary alignment records, or duplicated, were removed. The number of unique read obtained was counted from each bin on each chromosome. The LOESS regression was used to compute the corrected number of unique reads in each bin, to eliminate the effect of GC bias in different simulated samples (Yin et al., 2015). The *Z* score method was used to compare the normalized read frequencies of the simulated samples to the normalized read frequencies in the same bin in a reference set of the control normal samples (Straver et al., 2014).

Sample preparation, sequencing, and data processing were performed according to previous publication (Yin et al., 2015) with a few modifications: Firstly, sequencing of four mixture samples (at 0%, 4%, 12%, 20%) by Ion Proton Sequencer (Life Technologies, Carlsbad, CA) was performed at flow 500 respectively; Secondly WISECONDOR (Straver et al., 2014) was used as the pipeline implemented in CNV-calling, and thirdly, bin of 500, 750, and 1000 kb was used for counting unique read. The region with *Z* score ≦−3 was marked as a potential microdeletion.

2.4 | Evaluation of parameters

The following parameters were evaluated with Box–Behnken design: unique read number (A, related to coverage), fetal DNA concentration (B), and window width (C, also named as bin). Design Expert Software (version 8.0.6 trial, Stat-Ease, Inc) was employed to perform the Box–Behnken design. The significance of each coefficient was statistically analyzed by $p < 0.05$.

3 | RESULTS

The read length distribution of the artificial DNA mixture is provided in Figure S1. In the artificial DNA mixture, DNA derived from leukocytes of patient with 3.16 Mb deletion in chr22q11.21 was mixed in some ratios to simulate the fetal fraction in DNA derived from maternal plasma at known concentration, resulting in the observed patterns of read length distribution of the artificial mixture. Similar to the read length distribution of DNA derived from maternal plasma for aneuploidy screening, the median of the read length distribution for the artificial mixture was also located within 150–200 bp, in addition to similar read length distribution pattern.

The Box–Behnken results are shown in Table S2. The simulated fetal fraction ranged from 4% to 20% with unique read number ranging from 4 to 10 M and bin size ranging from 500 to 1,000 kb. The mean *Z* score and cv % was obtained at different conditions. A multiple quadratic regression model was established based on the results: $R1 = -4.72 - 0.83 \times A - 0.88 \times B - 0.81 \times C - 0.59 \times A \times B - 0.80 \times A \times C + 0.84 \times A^2 + 0.42 \times C^2$ (Final Equation in Terms of Coded Factors). The variance analysis of the model is as shown in Table 1. A, B, and C had negative effects on the *Z* score. The interactions between A and B as well as A and C also presented negative effects. The existence of A^2 and C^2 had positive effects on *Z* score with bigger *p* value. The statistical results implied that this model could be used to navigate the design space. (Table 1).

The simulated models by Box–Behnken design are shown in Figure 1. When B stayed in the region ranging from 4% to 20% with A changing in the scope of 4–10 M, the response surface diagram showed the highest *Z* score (Figure 1a). There were many solutions for the target *Z* score (−3) with a desirability of 1 (Figure 1b).

Based on the response surface model, we obtained 24 optimization solutions of the *Z* score −3 for microdeletion of 3.16 Mb (Table S3). The solution of fetal DNA concentration (code −0.24, predicted value 10.08%) and unique read number (code −1, predicted value 4 M) was recommended by the software, which can be set as the lowest quality specifications that must be met, when detecting fetal CNV ≧3.16 Mb.

4 | DISCUSSION

In this study, the equation of *Z* score depending on the fetal fraction, unique read number and bin size was obtained by Box–Behnken design. The negative effects of A, B and C

TABLE 1 ANOVA for response surface reduced quadratic model analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	*df*	Mean Square	*F* Value	*p*-value Prob > *F*	Model status	R^2	Adjusted R^2	Predicted R^2
Model	24.77	7.00	3.54	37.81	<0.0001	Significant	0.97	0.94	0.76
A–A	5.49	1.00	5.49	58.71	<0.0001				
B–B	6.14	1.00	6.14	65.63	<0.0001				
C–C	5.28	1.00	5.28	56.43	<0.0001				
AB	1.42	1.00	1.42	15.13	0.0037				
AC	2.54	1.00	2.54	27.18	0.0006				
A^2	2.97	1.00	2.97	31.69	0.0003				
C^2	0.75	1.00	0.75	7.98	0.0199				
Residual	0.84	9.00	0.09						
Lack of fit	0.84	5.00	0.17						
Pure error	0.00	4.00	0.00						
Cor total	25.61	16.00							

Note: The coefficients were considered statistically significant at $p < 0.05$.

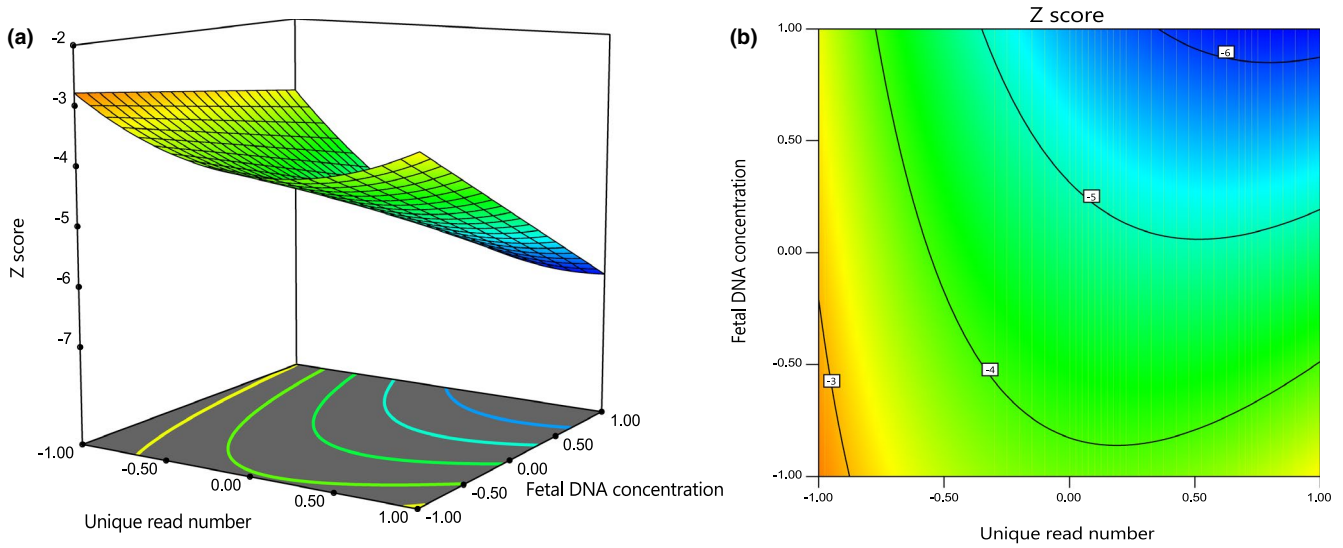

FIGURE 1 (a) Response surface diagram of *Z* score corresponding to fetal DNA concentration and unique read number for sample CMA0217-P1. (b) Desirability and 95% prediction intervals of *Z* score corresponding to fetal DNA concentration and unique read number

were quantified as coefficients in the equation, and the smallest values of these parameters were defined as unique read number of 4 Mb and fetal DNA concentration of 10.08% at fixed bin of 750 kb for detecting microdeletion of 3.16 Mb.

The smaller the *Z* score, the larger the absolute value, and the more favorable it is to report the CNV. When the *Z* score is greater than −3, the pipeline WISECONDOR reports absence of CNV. The negative effect of fetal DNA concentration and unique read number has been reported in a previous study (Zhao et al., 2015). We further found that bin size had significant negative effect on *Z* score for microdeletions, interactions of A and B as well as A and C were also significant negative factors. In contrast to other publications that characterized the effect of factors as positive or negative (Lo et al., 2016), this study quantified the effect of fetal DNA concentration, unique read number and bin as coefficients of −0.83, −0.88, and −0.81. Moreover, the value of B (fetal DNA concentration) can be as large as possible because there is no B^2 existing in the equation. Clinically, we used this conclusion to verify specious CNV. As shown in Figure S2, sample 1 and 2 showed plausible CNV; re-sequencing was performed after enrichment of fetal fraction by cfDNA size selection, *Z* scores increased with the increasing fetal fraction correspondingly. Thus, the plausible CNVs should be existing reliably and finally were confirmed by CMA. It is indicated that resequencing after DNA size selection by gel electrophoresis could be an effective validation method for CNV detection.

The simulated sample in this study was not prepared by mixing the DNA from the fetal plasma and maternal plasma due to sampling difficulty, but rather was prepared similar to that used for negative control preparation in aneuploidy screening. In addition, we only assessed a single genotype (3.16 Mb deletion at 22q11.2). Nonetheless, the quality control specification obtained from such mixtures is expected to extend the application scope to CNVs ≥3.16 Mb since CNV size has a negative effect on CNV detection (Zhao et al., 2015). Taking CNVs ≥3.16 Mb in Yin's research as examples (Yin et al., 2015), among the samples with sequencing results satisfying the quality control specification, the sensitivity for CNVs ≥3.16 Mb reaches 97.3% (36/37, with one deletion of 4.95 M in Chr21 not reported) higher than 75% (when values of parameters meeting the original quality control standard of unique read number ≥2.5 M, fetal DNA concentration ≥3.5% in aneuploidy screening). Since samples representing the biological and technical variability of the event region are difficult to mimic and quantify (Li et al., 2016), they were not involved in this study. Taken together, the predicted quality control specification for detecting CNVs ≥3.16 Mb by the simulated model needs to be verified and adjusted by more clinical samples. This method is expected to predict detection conditions for smaller size of CNVs with specific locations, sizes and genotype. Further, nested deletions, such as those in 22q11.2 present as a significant fraction of Velocardiofacial syndrome (Kong, Cheng, To, & Leung, 2019), will be further explored in a subsequent study.

Most studies retrospectively reported the clinical performance of general NIPT in CNV screening (Liu et al., 2016), while in this study, we positively designed experiments to quantify the effect of key factors involved in microdeletion detection and determine the value of parameters. Such a method can benefit the establishment of quality standards for CNV detection and help doctors and technicians in interpreting the CNV detection results.

COMPETING INTEREST

None declared.

ORCID

Chunhua Zhang https://orcid.org/0000-0003-2391-9290
Quanze He https://orcid.org/0000-0003-4642-117X
Bin Yu https://orcid.org/0000-0001-7796-7314

REFERENCES

Buysse, K., Delle Chiaie, B., Van Coster, R., Loeys, B., De Paepe, A., Mortier, G., … Menten, B. (2009). Challenges for CNV interpretation in clinical molecular karyotyping: Lessons learned from a 1001 sample experience. *European Journal of Medical Genetics*, *52*(6), 398–403. https://doi.org/10.1016/j.ejmg.2009.09.002

Carter, N. P. (2007). Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature Genetics*, *39*(7 Suppl), S16–S21. https://doi.org/10.1038/ng2028

Knierim, E., Lucke, B., Schwarz, J. M., Schuelke, M., & Seelow, D. (2011). Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS ONE*, *6*(11), e28240. https://doi.org/10.1371/journal.pone.0028240

Kong, C. W., Cheng, Y. K. Y., To, W. W. K., & Leung, T. Y. (2019). Prevalence of chromosomal abnormalities and 22q11.2 deletion in conotruncal and non-conotruncal antenatally diagnosed congenital heart diseases in a Chinese population. *Hong Kong Medical Journal*, *25*(1), 6–12. https://doi.org/10.12809/hkmj187552

Li, R., Wan, J., Zhang, Y., Fu, F., Ou, Y., Jing, X., … Liao, C. (2016). Detection of fetal copy number variants by non-invasive prenatal testing for common aneuploidies. *Ultrasound in Obstetrics and Gynecology*, *47*(1), 53–57. https://doi.org/10.1002/uog.14911

Liu, H., Gao, Y., Hu, Z., Lin, L., Yin, X., Wang, J., … Wang, W. (2016). Performance evaluation of NIPT in detection of chromosomal copy number variants using low-coverage whole-genome sequencing of plasma DNA. *PLoS ONE*, *11*(7), e0159233. https://doi.org/10.1371/journal.pone.0159233

Lo, K. K., Karampetsou, E., Boustred, C., McKay, F., Mason, S., Hill, M., … Chitty, L. S. (2016). Limited clinical utility of non-invasive prenatal testing for subchromosomal abnormalities. *American Journal of Human Genetics*, *98*(1), 34–44. https://doi.org/10.1016/j.ajhg.2015.11.016

Straver, R., Sistermans, E. A., Holstege, H., Visser, A., Oudejans, C. B., & Reinders, M. J. (2014). WISECONDOR: Detection of fetal aberrations from shallow sequencing maternal plasma based on a within-sample comparison scheme. *Nucleic Acids Research*, *42*(5), e31. https://doi.org/10.1093/nar/gkt992

Yin, A. H., Peng, C. F., Zhao, X., Caughey, B. A., Yang, J. X., Liu, J., … Zhang, K. (2015). Noninvasive detection of fetal subchromosomal abnormalities by semiconductor sequencing of maternal plasma DNA. *Proceedings of the National Academy of Sciences*, *112*(47), 14670–14675. https://doi.org/10.1073/pnas.1518151112

Zhao, C., Tynan, J., Ehrich, M., Hannum, G., McCullough, R., Saldivar, J. S., … Deciu, C. (2015). Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma. *Clinical Chemistry*, *61*(4), 608–616. https://doi.org/10.1373/clinchem.2014.233312

Zhao, M., Wang, Q., Wang, Q., Jia, P., & Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: Features and perspectives. *BMC Bioinformatics*, *14*(Suppl 11), S1. https://doi.org/10.1186/1471-2105-14-S11-S1

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Zhang C, Liang B, Qiao L, et al. Effect quantification and value prediction of factors in noninvasive detection for specific fetal copy number variants by semiconductor sequencing. *Mol Genet Genomic Med*. 2019;7:e718. https://doi.org/10.1002/mgg3.718