VNU Journal of Science: Com. Science & Com. Eng., Vol. 39, No. 2 (2023) 101-108

VNU Journal of Science:
Computer Science and Communication Engineering

Journal homepage: http://www.jcsce.vnu.edu.vn/index.php/jcsce

Original Article

A Method to Create NIPT Samples with Turner Disorder to Evaluate NIPT Algorithms

Lê Sỹ Vinh*

VNU University of Engineering and Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

Received 16 October 2023
Revised 27 November 2023; Accepted 28 November 2023

Abstract: Noninvasive prenatal test (NIPT) is a widely used screening method to detect numerical aberrations on autosomal chromosomes with high sensitivity and specificity. However, predicting abnormalities on sex chromosomes is much more challenging due to complicated genomic characteristics of sex chromosomes. Turner disorder (XO) is one of the most common disorders due to the missing of one X chromosome in females. A number of large-scale retrospective studies have showed that the positive detection rate of Turner disorder is considerably high and the false negative rate has not been well evaluated due to the lack of available positive Turner samples. To solve the problem, we present a novel method to create positive Turner samples from negative samples that can be easily obtained from NIPT testing centers. We applied the method to create 600 positive Turner samples and examined and the performance of WisecondorX, CNVkit, and VINIPT algorithms on the samples. Experiments show that the sensitivity of WisecondorX, VINIPT, and CNVkit in detecting positive Turner samples are 100%, 100%, and 99.5%, respectively. We also evaluated the performance of the algorithms on 500 negative XO samples. The VINIPT and CNVkit algorithms have very high specificity in identifying negative XO samples (i.e., 99.8% for VINIPT and 99.6% for CNVkit), while WisecondorX has a lower specificity of 96.8%. The study opens an easy way for researchers to assess the performance of NIPT algorithms on screening the Turner disorder.

Keywords: NIPT, cfDNA analysis, Turner disorder, XO aberration, WisecondorX, CNVkit.

1. Introduction

Genetic tests are becoming routine tests for investigating or screening a number of diseases. Each person has 23 pairs of chromosomes, i.e., 46 chromosomes. The numerical chromosome abnormalities (e.g., missing one chromosome or having an additional chromosome) result in different disorders. The well-known disorders related to autosomal chromosomes include

* Corresponding author.
E-mail address: vinhls@vnu.edu.vn

https://doi.org/10.25073/2588-1086/vnucsce.1611

trisomy 13 (Patau syndrome or having an additional copy of chromosome 13), trisomy 18 (Edwards syndrome or having an additional copy of chromosome 18), and trisomy 21 (Down syndrome or having an additional copy of chromosome 21). The numerical chromosome abnormalities on sex chromosomes consist of monosomy X (XO or Turner syndrome, i.e., completely missing one copy of X chromosome) and trisomy X syndrome (XXX or having an additional X chromosome) in females; Klinefelter syndrome (having an additional X chromosome) and Jacobs syndrome (having an additional Y chromosome) in males.

The NIPT has been widely used for screening numerical chromosomal abnormalities by analyzing the cell-free DNA (cfDNA) in the maternal blood. The NIPT is usually performed within the 10^{th} week and the 12^{th} week of the pregnancy. At that time, the amount of fetal cfDNA accounts for around 10% of the total cfDNA in the maternal blood. A number of computational methods have been proposed to detect numerical chromosomal aberrations from the cfDNA data such as NIFTY [1], Wisecondor [2] and its improvement WisecondorX [3], CNVkit [4], or triSure [5]. Some of them are commercial software and not publicly available for testing.

The NIPT algorithms have high sensitivity (i.e., sensitivity is the number of correctly predicted positive samples over the number of positive samples) and high specificity (i.e., specificity is the proportion of correctly predicted negative samples out of all negative samples) for detecting trisomy (i.e., having an addition chromosome) on chromosomes 13 (T13), 18 (T18), and 21 (T21) [6–8]. An intensive retro-perspective study of 146958 samples showed a sensitivity of 100% for T13, 98.24% for T18 and 99.17% for T21 [9]. The specificity of the NIPT algorithms for detecting trisomy on chromosomes 13, 18, and 21 is very high, i.e., around 99.9%. The performance of the NIPT algorithms on sex chromosomes is much worse than that on autosomal chromosomes. This might be due to sequencing bias of guanine and cytosine on the sex chromosomes, a partially loss of X chromosome in ageing women, or a large number of homologous genes between two sex chromosomes X and Y.

The accumulated frequency of numerical aberrations on sex chromosomes (SCA) is estimated to be about 1 in 500 live births. Detecting SCA by phenotypes of fetus during the pregnancy might be difficult, therefore, NIPT plays an important role in screening SCA [10-12]. The positive screening rate of SCA by NIPT ranges from 0.66% to 0.68%; and positive predictive rate is considerable low, i.e., between 36.9% and 40.5% [11, 12].

The monosomy X (called XO or Turner syndrome) indicates a condition that a female loses a chromosome X. The XO disorder affects about 1 in 2000 live-born females [13]. The general symptoms of XO disorder are small stature, amenorrhea and infertility. The current large-scale studies showed that about 0,32% of cases are predicted to be positive with the XO disorder [11, 12], but positive predictive value (PPV - the ratio of truly positive cases over the number of predicted positive cases) is very low, e.g., only 12.5% in [12] or 21% in [11].

The specificity of NIPT on detecting negative samples can be estimated by large-scale retrospective studies. However, the sensitivity (or false negative rate) of NIPT, one of the most important indicators, is not easily measured. The false negative samples are mainly detected based on customer reports. It is impractical to perform karyotypes or follow up a large number of customers to determine false negative results of SCA.

To date, the sensitivity of NIPT has been typically measured from small (and not publicly available) positive datasets. An alternative approach to assess the sensitivity of NIPT methods is to employ the simulation data. Recently, we have proposed a simple computational approach to simulate positive samples with aberrations on autosomal chromosomes to evaluate the sensitivity of NIPT algorithms [14]. In this paper, we introduce a

novel method for generating positive samples with the Turner disorder from negative male samples. The generated positive XO samples were used to evaluate the sensitivity of different NIPT algorithms.

2. Methods

2.1. Non-invasive prenatal testing pipeline

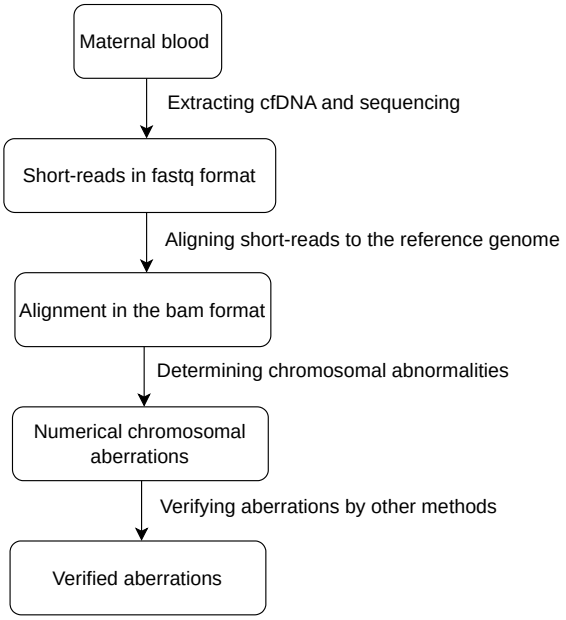

Figure 1. The flowchart of NIPT procedure.

The NIPT consists of both wet-lab to sequence genomic data and bioinformatic analyses to detect chromosomal aberrations from short-reads data. The NIPT procedure as illustrated in Figure 1 includes a number of steps:

- Blood from the mother is obtained between the 10^{th} week and the 12^{th} week. This step might be re-conducted if the amount of fetal DNA in the mother's blood is not enough to reliably detect chromosomal aberrations.

- cfDNA in the mother's blood are extracted and sequenced using high-throughput sequencing machines. In practice, the cfDNA are sequenced at a low coverage, i.e., 0.1x to 1x. Either single-end or paired-end sequencing could be applied.

- Short-reads obtained from the sequencing step are aligned with the reference genome to determine their genomic locations. The alignment file is the input for NIPT algorithms to determine numerical chromosomal aberrations. The NIPT algorithms estimate the percentage of fetal DNA (note that the main proportion of short-reads come from the DNA of mother and only a small part of the short-reads is from the placenta of fetus). If the fetal DNA fraction is not sufficient (i.e., smaller than a cut-off threshold e.g., 5%), we need to re-take the blood from the mother, re-sequence the sample and reanalyze the data. Normally, the fetal DNA increases with the time of pregnancy, i.e., the fetal DNA in the mother's blood at the 12^{th} week is normally more than that at the 10^{th} week.

- The positive predictive value of the NIPT algorithms might be low, therefore, positive results from NIPT should be verified by other methods such as ultrasound or karyotype examinations; and must be carefully explained by genetic consultants for pregnant women.

2.2. NIPT algorithms

A number of algorithms have been proposed to determine the number of copies of chromosomes by analyzing the number of short-reads on the chromosomes (called read coverage). There are a number of challenges in detecting the numerical aberrations on fetal chromosomes. First, the number of short-reads from the fetus accounts for only a small part in comparison with that from the mother. Thus, it is difficult to detect a change in the fetal DNA. The NIPT is normally conducted with a low coverage data, i.e., read coverage ranging from only $0.1x$ to $1x$. The low coverage data might lead to the inconsistency of NIPT algorithms. We also note that detecting aberrations on sex chromosomes is more difficult than that on autosomal chromosomes due to the complicated genomic characteristics of sex chromosomes.

The NIPT algorithms contain the main following steps:

- *Read alignment and quality control step*: In this first step, the short-reads obtained from next-generation sequencing (NGS) are aligned with a reference genome (e.g., the reference GRCh37 or the reference GRCh38) using genome alignment programs such as BWA [15] or Bowtie 2 [16]. The alignment process ensures that short-reads are mapped to their corresponding genomic locations. To enhance the accuracy of downstream analyses, a number of quality control might be applied. These include the removal of duplicated reads (i.e., if some reads are mapped into the same location, only one is kept), reads that are ambiguously mapped to multiple genomic locations are ignored. Eliminating such artifacts is crucial to ensure that the downstream analyses are based on reliable data.

- *Read counting per bin step*: The genome is divided into a series of equally sized bins. These bins can vary in size, but we should use large bin sizes for low coverage data (e.g., 100,000 base pairs) to ensure that each bin contains a sufficient number of reads. The number of reads mapping to each bin is counted from the alignment data. The read coverage of bins provides the basis for analyzing the aberrations on bins and chromosomes. To improve the reliability of the data, data correction methods such as GC-content correction might be applied to account for any bias in the read coverage.

- *Fetal DNA percentage estimation step:* Let's denote f as the fetal DNA fraction in the sample, representing the proportion of reads originating from the fetus. For example, $f = 10\%$ indicates that 10% of the reads originate from the fetal DNA, while the remaining 90% of reads are from the maternal DNA. Different methods have been proposed to estimate the fetal DNA [17]. The SeqFF method has high accuracy and applicable to both male and female fetuses.

- *Statistical testing for aberrations*: Statistical tests are conducted to determine whether the read coverage in bins of a test sample significantly deviates from what is expected in normal (reference) samples. To mitigate variations of read coverage among multiple samples, within-sample analysis methods can be employed [2, 3]. These methods compare a specific bin b in the test sample to a set of reference bins located on other chromosomes within the same test sample. The set of reference bins of b are selected such that they have similar behaviors or characteristics with b. The aberration scores of bins on a specific chromosome C are combined to calculate an overall aberration score. This score is often represented as a z-score, which quantifies the extent to which the read coverage in chromosome C deviates from the normal expectation. If the z-score for chromosome C surpasses a predefined threshold (either higher or lower), it can be considered as an abnormal chromosome. Otherwise, it is categorized as normal.

- *Aberration detection techniques*: The Wisecondor [2] uses Stouffer's z-score sliding window approach to segment and determine aberrations on chromosomes. It involves sliding a window along the chromosome, calculating z-scores for each window, and identifying significant deviations. The WisecondorX [3] or CNVkit [4] uses circular binary segmentation algorithm [18] instead of Stouffer's technique to detect segment aberrations that overcomes the running time burden of the Wisecondor algorithm.

The WisecondorX algorithm might produce a considerable number of false positive samples when using the overall z-score to predict chromosome aberrations. The NIPT algorithms might be sensitive with reference samples used to create their reference panels. The VINIPT algorithm has been introduced to overcome the limitations [14]. To achieve this, it identifies proper reference samples and establishes different reference panels. It also combines both Wisecondor and WisecondorX algorithms to decide if a chromosome is abnormal. These

enhancements collectively help reduce the false positive rate of the VINIPT algorithm.

2.3. Generating positive Turner samples

A normal female has two copies of chromosome X. The female with Turner disorder (XO) has only one chromosome X. Specifically, a negative XO sample is a female sample with two copies of X chromosome; while a positive sample of XO aberration is a female sample with one copy of X chromosome. The percentage of live-born females with the Turner disorder is about 0.05%, however, the positive detection rate of XO from NIPT is normally much higher, i.e., estimated around 0.32% [11, 12]. In other words, the positive predictive value of XO disorder from NIPT are very low, e.g., only 12.5% in [12] or 21% in [11].

The positive Turner samples are rare and not publicly available for evaluating NIPT algorithms. This prevents the improvement of current NIPT algorithms or the development of new NIPT algorithms for accurately analyzing XO aberrations. We can solve the obstacle by creating positive XO samples from negative male NIPT samples. We note that a negative male sample (i.e., NIPT sample with normal male fetus) has one X chromosome and one Y chromosome. If we turn the negative male sample into a female sample by removing its Y chromosome, the obtained female sample now has only one X chromosome and can play as a positive XO sample. Technically, given a negative male NIPT sample M, we remove all short-reads in Y chromosome of M (i.e., short-reads that are mapped to the Y chromosome) to create a female sample M'. The sample M' is now considered as a female sample with one X chromosome (i.e., having the Turner disorder).

We collected a dataset of 600 male samples from singleton pregnancies that have clear negative predictions from NIPT screening and no customer reports about false results as used in our current study [14]. The samples were sequences by the MGI sequencing platform with single-end reads of size 50 bps. The 600 male samples were used to create 600 positive samples with XO aberrations that were used to evaluate the sensitivity of the NIPT algorithms.

We also collected a dataset of 500 negative female samples from singleton pregnancies (i.e., clear negative predictions from NIPT screening and no customer reports about false results) as used in our current study [14] to test the specificity of the NIPT algorithms on identifying negative XO samples.

3. Results

First, we analyze the distribution of fetal DNA fractions and read coverages of both negative and positive samples (see Table 1). The fetal DNA fractions of negative samples range from 5.1% to 16.6% with a mean of 10.1%; while the fetal fractions of positive XO samples are from 4.6% to 22.2% with a mean of 9.7%. The mean fetal DNA fractions of the samples are both around 10%. The negative samples contain from 4.4 million (~0.08x) to 36.9 million (~0.61x) of short-reads. The mean read coverage is about 0.23x. The positive XO samples consist of fewer short-reads (i.e., the mean read coverage of about 0.19x).

Figure 2 illustrates the distribution of the fetal DNA fractions of both negative female samples and positive XO samples. The read coverages of the samples are displayed in Figure 3.

Table 1. The distribution of fetal DNA fractions and read coverages of negative and positive XO samples

	Fetal DNA fraction			Read coverage (#short-reads)		
	Min	Mean	Max	Min	Mean	max
Negative female samples	5.1%	10.1%	16.6%	4.4M	14.2M	36.9M
Positive XO samples	4.6%	9.7%	22.2%	4.3M	11.6M	35.9M

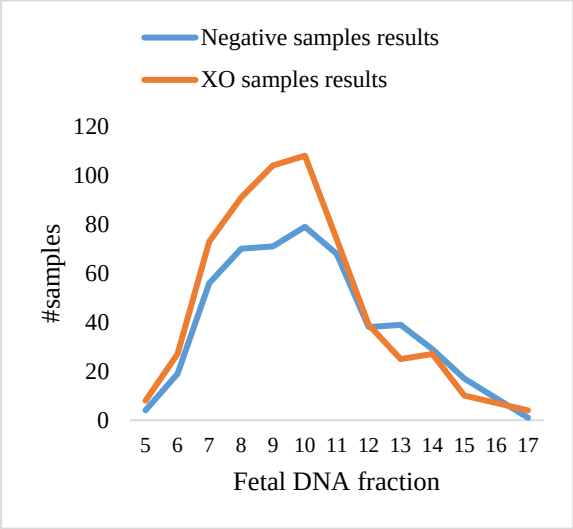

Figure 2. The fetal DNA fractions of negative and positive XO samples.

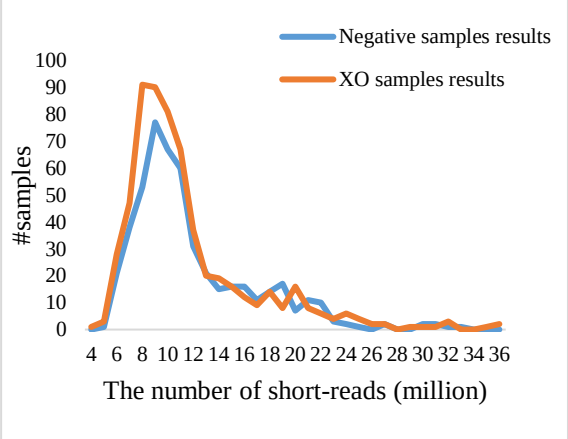

Figure 3. The distribution of read coverages of negative and positive XO samples.

We examined the ability of the NIPT algorithms on detecting negative samples. To this end, WisecondorX, CNVkit and VINIPT were tested on 500 negative female samples. Table 2 presents the number of true negative and false positive samples resulted from the three algorithms.

WisecondorX correctly identified 484 negatives; and falsely assigned 16 negative samples as positive samples. Its overall specificity of detecting negative XO samples is 96.8%. The CNVkit algorithm has a better specificity than WisecondorX. It produced two false positive samples (i.e., a specificity of 99.6%). The VINIPT algorithm could determine 499 over 500 negative samples and falsely assigned only one negative sample as a positive sample, i.e., its specificity is 99.8%.

Table 2. The performance of NIPT algorithms on negative female samples

	True negative	False positive	Specificity
WisecondorX	484	16	96.8%
CNVkit	498	2	99.6%
VINIPT	499	1	99.8%

We measured the sensitivity of the NIPT algorithms on detecting positive XO samples. The WisecondorX, CNVkit and VINIPT algorithms were evaluated on the 600 positive XO samples. Table 3 shows the number of true positive and false negative samples predicted from the algorithms. WisecondorX and VINIPT correctly determined all positive samples, thus, they have a sensitivity of 100%. The CNVkit algorithm has a lower sensitivity than WisecondorX and VINIPT. It correctly identified 597 positive samples and falsely assigned 3 positive samples as negative samples, i.e., its sensitivity is 99.5%.

Table 3. The performance of NIPT algorithms on positive XO samples

	True positive	False negative	Sensitivity
WisecondorX	600	0	100%
CNVkit	597	3	99.5%
VINIPT	600	0	100%

We illustrate z-scores from the VINIPT algorithm on both negative and positive XO samples in Figure 4. We see that the negative samples and the positive XO samples form two different groups. There is only one negative sample that belongs to the group of positive samples indicating that this negative sample is wrongly predicted as a positive sample. The distribution of z-scores from the WisecondorX is displayed in Figure 5. Basically, the group of

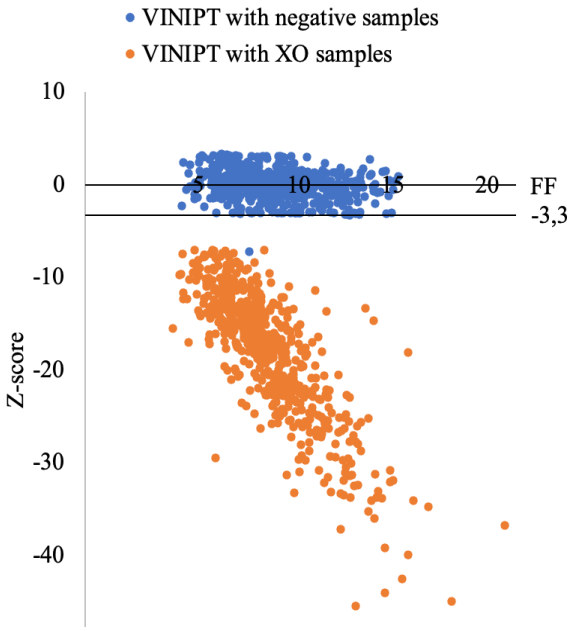

Figure 4. The distribution of z-scores from the VINIPT algorithm on both negative and positive XO samples. FF indicates the fetal DNA fraction.

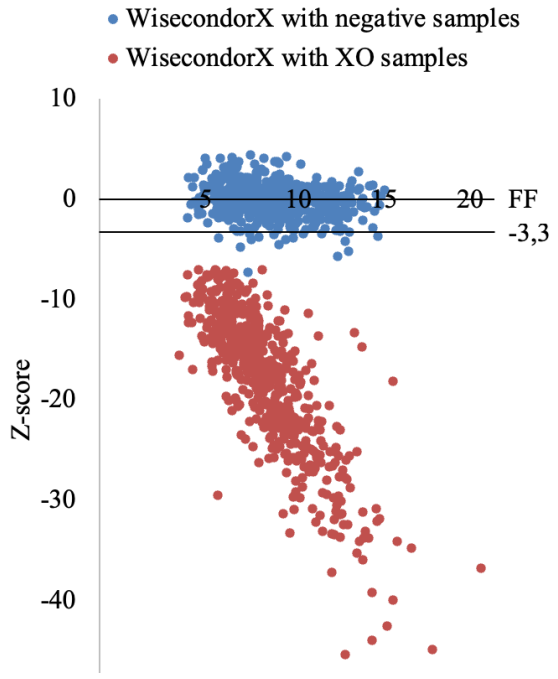

Figure 5. The z-score distribution of the WisecondorX algorithm on negative and positive XO samples. FF indicates the fetal DNA fraction.

negative samples and that of positive samples are separated. However, we observe a number of negative samples that lie below the z-score cut-off threshold of -3.3 indicating that the negative samples were falsely assigned as positive samples. The CNVkit algorithm does not output z-score information for the illustration.

4. Discussions

Screening chromosomal aberrations in fetal genomes during the first weeks of pregnancy plays an important in prenatal diagnosis. Noninvasive prenatal testing based on cfDNA in the mother's blood has been widely used due to its considerably high accuracy, cheap and easy to implement in practice.

In this paper, we presented a new way to create positive samples with the Turner disorder from negative male samples. We might create a positive female sample with XO aberration from a negative female sample by removing a number of reads on its X chromosome. To do this, we have to estimate the fetal DNA fraction by computational methods. The fetal DNA fraction might be not accurately estimated, especially for samples with low fetal DNA fractions, that makes generating positive samples not as real as clinical positive samples.

We created a dataset of 600 positive Turner samples. We used the 600 positive Turner samples to benchmark the performance of NIPT algorithms. Experiments showed that WisecondorX and VINIPT had high sensitivity of detecting XO aberrations. WisecondorX has high sensitivity, however, its specificity on determining negative XO samples needs to be improved. The CNVkit algorithm has higher specificity than WisecondorX, however, its sensitivity is not as high as that of WisecondorX and VINIPT. The VINIPT algorithm performed well on identifying negative XO samples as well as detecting positive XO samples.

The NIPT algorithms have very high sensitivity on detecting positive XO samples generated by our method. It is worth to note that the generating positive XO samples have exactly one copy of X chromosome. In practice, a positive XO sample might loss a part of X chromosome (called microdeletion) that makes

the NIPT algorithms more difficult to detect positive XO samples with microdeletions.

Acknowledgements

Some technical supports from Nguyen Huy Tinh and Nguyen Van Huynh are highly appreciated.

References

[1] F. Jiang et al., "Noninvasive Fetal Trisomy (NIFTY) test: an Advanced Noninvasive Prenatal Diagnosis Methodology for Fetal Autosomal and sex Chromosomal Aneuploidies," BMC Med Genomics, Vol. 5, no. 1, p. 57, Dec.2012, https://doi.org/10.1186/1755-8794-5-57.

[2] R. Straver, E. A. Sistermans, H. Holstege, A. Visser, C. B. M. Oudejans, and M. J. T. Reinders, "WISECONDOR: Detection of Fetal Aberrations from Shallow Sequencing Maternal Plasma Based on a Within-Sample Comparison Scheme," Nucleic Acids Res, vol. 42, No. 5, pp. e31–e31, Mar. 2014, https://doi.org/10.1093/nar/gkt992.

[3] L. Raman, A. Dheedene, M. De Smet, J. Van Dorpe, B. Menten, "WisecondorX: Improved Copy Number Detection for Routine Shallow Whole-Genome Sequencing," Nucleic Acids Res, Vol. 47, no. 4, pp. 1605–1614, Feb. 2019, https://doi.org/10.1093/nar/gky1263.

[4] E. Talevich, A. H. Shain, T. Botton, B. C. Bastian, "CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing," PLoS Comput Biol, Vol. 12, no. 4, p. e1004873, Apr. 2016, https://doi.org/10.1371/journal.pcbi.1004873.

[5] M.-D. Phan et al., "Establishing and Validating Noninvasive Prenatal Testing Procedure for Fetal Aneuploidies in Vietnam," The Journal of Maternal-Fetal & Neonatal Medicine, Vol. 32, no. 23, pp. 4009–4015, Dec. 2019, https://doi.org/10.1080/14767058.2018.1481032.

[6] T. Mokveld, Z. Al-Ars, E. A. Sistermans, and M. Reinders, "A Comprehensive Performance Analysis of Sequence-Based within-Sample Testing NIPT Methods," PLoS One, vol. 18, no. 4, p. e0284493, Apr. 2023, https://doi.org/10.1371/journal.pone.0284493.

[7] P. Paluoja et al., "Systematic Evaluation of NIPT Aneuploidy Detection Software Tools with Clinically Validated NIPT Samples," PLoS Computational Biology, Vol. 17, No. 12, 2021, https://doi.org/10.1371/journal.pcbi.1009684.

[8] H. Zhang et al., "Non-Invasive Prenatal Testing for Trisomies 21, 18 and 13: Clinical Experience from 146 958 Pregnancies," Ultrasound in Obstetrics & Gynecology, Vol. 45, No. 5, pp. 530–538, May 2015, https://doi.org/10.1002/uog.14792.

[9] H. Zhang et al., "Non-Invasive Prenatal Testing for Trisomies 21, 18 and 13: Clinical Experience from 146 958 Pregnancies," Ultrasound in Obstetrics & Gynecology, Vol. 45, No. 5, pp. 530–538, May 2015, https://doi.org/10.1002/uog.14792.

[10] Y. Zou et al., "Performance of Expanded non-Invasive Prenatal Testing for Fetal Aneuploidies and copy Number Variations: A prospective Study from a Single Center in Jiangxi province, China," Front Genet, vol. 13, Jan. 2023, https://doi.org/10.3389/fgene.2022.1073851.

[11] N. Guo, M. Cai, M. Lin, H. Xue, H. Huang, L. Xu, "Positive Predictive Value of Noninvasive Prenatal Testing for sex Chromosome Abnormalities," Mol Biol Rep, vol. 49, no. 10, pp. 9251–9256, Oct. 2022, https://doi.org/10.1007/s11033-022-07754-x.

[12] X. Lu, C. Wang, Y. Sun, J. Tang, K. Tong, and J. Zhu, "Noninvasive Prenatal Testing for Assessing Foetal sex Chromosome Aneuploidy: a Retrospective Study of 45,773 cases," Mol Cytogenet, vol. 14, no. 1, p. 1, Dec. 2021, https://doi.org/10.1186/s13039-020-00521-2.

[13] C. Patrick, "XO Syndrome," in Encyclopedia of Child Behavior and Development, S. Goldstein and J. A. Naglieri, Eds., Boston, MA: Springer US, 2011, pp. 1583–1584. https://doi.org/10.1007/978-0-387-79061-9_3118.

[14] T. Nguyen, H. Nguyen, M. Pham, and V. Le, "An Efficient Computational Method to Create Positive NIPT Samples with Autosomal Trisomy" https://www.biorxiv.org/content/10.1101/2023.11.24.568620v1.

[15] H. Li and R. Durbin, "Fast and Accurate Long-Read Alignment with Burrows–Wheeler Transform," Bioinformatics, vol. 26, no. 5, pp. 589–595, Mar. 2010, https://doi.org/10.1093/bioinformatics/btp698.

[16] Langmead, "Bowtie2,"Nat Methods, Vol. 9, No. 4, pp. 357–359, 2013, https://doi.org/10.1038/nmeth.1923.Fast.

[17] D. M. ,Beek et al., "Comparing Methods for Fetal Traction Determination and Quality Control of NIPT Samples," Prenat Diagn, vol. 37, no. 8, pp. 769–773, Aug. 2017, https://doi.org/10.1002/pd.5079.

[18] A. B. Olshen, E. S. Venkatraman, R. Lucito, M. Wigler, "Circular Binary Segmentation for the Analysis of Array-Based DNA Copy number data," Biostatistics, Vol. 5, No. 4, pp. 557–572, 2004, https://doi.org/10.1093/biostatistics/kxh008.