

Article

Noninvasive Prenatal Paternity Testing with a Combination of Well-Established SNP and STR Markers Using Massively Parallel Sequencing

Xuefeng Shen ^{1,2}, Ran Li ^{1,2} , Haixia Li ^{1,2}, Yu Gao ³, Hui Chen ^{1,2}, Ning Qu ^{1,2}, Dan Peng ^{1,2}, Riga Wu ^{2,*}  and Hongyu Sun ^{1,*}

- ¹ Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China; shenxf3@mail2.sysu.edu.cn (X.S.); liran6@mail2.sysu.edu.cn (R.L.); lhaix@mail3.sysu.edu.cn (H.L.); chen258@mail2.sysu.edu.cn (H.C.); quning@mail2.sysu.edu.cn (N.Q.); pengdan5@mail2.sysu.edu.cn (D.P.)
- ² Guangdong Province Translational Forensic Medicine Engineering Technology Research Center, Sun Yat-Sen University, Guangzhou 510080, China
- ³ Department of Obstetrics, The Sixth Affiliated Hospital of Sun Yat-Sen University, Guangzhou 510630, China; gaoy57@mail.sysu.edu.cn
- * Correspondence: wurg3@mail.sysu.edu.cn (R.W.); sunhy@mail.sysu.edu.cn (H.S.)

Abstract: Cell-free fetal DNA (cffDNA) from maternal plasma has made it possible to develop noninvasive prenatal paternity testing (NIPPT). However, most studies have focused on customized single nucleotide polymorphism (SNP) typing systems and few have used conventional short tandem repeat (STR) markers. Based on massively parallel sequencing (MPS), this study used a widely-accepted forensic multiplex assay system to evaluate the effect of noninvasive prenatal paternity testing with a combination of well-established SNP and STR markers. Using a ForenSeq DNA Signature Prep Kit, NIPPT was performed in 17 real parentage cases with monovular unborn fetuses at 7 to 24 gestational weeks. Different analytical strategies for the identification of paternally inherited allele (PIA) were developed to deal with SNPs and STRs. Combined paternity index (CPI) for 17 real trios as well as 272 unrelated trios was calculated. With the combination of SNPs and A-STRs, 82.35% (14/17), 88.24% (15/17), 94.12% (16/17), and 94.12% (16/17) of real trios could be accurately determined when the likelihood ratio (LR) threshold for paternity inclusion was set to 10,000, 1000, 100, and 10, respectively. This reveals that simultaneous surveys of SNP and STR markers included in the ForenSeq DNA Signature Prep Kit offer a promising method for NIPPT using MPS technology.

Keywords: forensic genetics; noninvasive prenatal paternity testing; short tandem repeat; single nucleotide polymorphism



Citation: Shen, X.; Li, R.; Li, H.; Gao, Y.; Chen, H.; Qu, N.; Peng, D.; Wu, R.; Sun, H. Noninvasive Prenatal Paternity Testing with a Combination of Well-Established SNP and STR Markers Using Massively Parallel Sequencing. *Genes* **2021**, *12*, 454. <https://doi.org/10.3390/genes12030454>

Academic Editor: Emiliano Giardina

Received: 9 February 2021

Accepted: 20 March 2021

Published: 22 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Prenatal paternity testing is of great importance in certain situations, such as in the investigation of pregnancy due to rape or incest. One possible method is to perform paternity testing through amniocentesis and chorionic villus sampling at the gestational age of week ten and beyond [1–3]. However, the use of utensils through the cervix or abdomen is invasive, and such sampling can lead to a risk of complications, including but not limited to limb reduction defects [4], fetal respiratory disturbances [5], and miscarriage [6].

Fortunately, the discovery of cell-free fetal DNA (cffDNA) in maternal plasma [7] enables the possibility of noninvasive prenatal paternity testing (NIPPT) [8]. Aside from the noninvasiveness of cffDNA-based paternity testing that avoids risks associated with invasive procedures, encouraging results also showed that cffDNA can be detected at the gestational age of week four at the earliest [9,10]. Therefore, the development of highly sensitive and accurate NIPPT methods which are easy to popularize is crucial for forensic genetic researchers.

Short tandem repeat (STR) is a conventionally used genetic marker in paternity testing, and capillary electrophoresis (CE)-based STR length polymorphism analysis is the most commonly utilized typing method. Thus, early attempts in NIPPT began with STR; however, investigators were only able to detect few numbers of informative paternally derived alleles from autosomal STRs (A-STRs) [11,12]. On the one hand, stutter artifacts were produced frequently during the STR genotyping, which may influence the identification of paternally derived alleles. Since cffDNA generally comprises <20% of cell-free DNA (cfDNA) in maternal plasma [13], masking of cffDNA by dominant maternal DNA further complicated the testing. On the other hand, cffDNA is usually <160 bp in length [14,15], which is shorter than many STR amplicons detected by the CE method.

In recent years, many genetic markers, such as single nucleotide polymorphism (SNP) [16], DNA methylation [17], deletion/insertion polymorphism-short tandem repeat (DIP-STR) [18] and microhaplotype [19], were used to develop NIPPT methods. Among these, SNP typing by high-throughput sequencing or microarray technology was the most frequently used method in NIPPT due to their short amplicon sizes. For example, Chang et al. established a 5457 SNP typing system based on massively parallel sequencing (MPS) for NIPPT [20]. Qu et al. selected 1795 SNPs and performed NIPPT on the basis of the Illumina HiSeq platform [21]. However, the polymorphism of SNP markers was limited and thousands of markers are recommended to achieve a higher accuracy [16]. Moreover, whether the popular SNP markers or the new microhaplotype markers and self-designed panels are unfavorable for data communication and utilization of public reference database lacks evaluation. As a result, there is a lack of widely acknowledged panels and unified evaluation criteria, resulting in uncertainty on the evidence value of NIPPT.

STR markers are the main markers for exclusion-based relationship analysis and many STR DNA databases have been built and are still going on. If the advancement of well-established STR markers could be taken, the application of NIPPT could be broadened. To achieve this goal, both the problem of long amplicon size and high rate of stutters using the conventional CE-based STR typing method should be solved, so that paternally derived alleles could not only be detected but also be distinguished easier from stutters. Fortunately, MPS enables genotyping of combinations of multiple types (e.g., autosomal STRs, X-STRs, Y-STRs and SNPs) and large numbers of markers with shorter amplicon lengths than the CE method. Among MPS panels developed for forensic research, the ForenSeq DNA Signature Prep Kit (Verogen., San Diego, CA, USA), which is based on Illumina MiSeq FGx platform, was widely used and evaluated [22]. Also, a sequence simplification method for MPS-based STR sequence polymorphism analysis was developed, which may bring improvement for mixture deconvolution under stutter interference [23]. For unbalanced mixtures, the minor contributor's alleles could be identified even when the ratios were as low as 1:1000. Therefore, we investigated the utility of the ForenSeq DNA Signature Prep Kit in NIPPT and the usefulness of this sequence simplification strategy, to test the effect of noninvasive prenatal paternity testing with a combination of well-established SNP and STR markers.

2. Materials and Methods

2.1. Sample Collection and DNA Extraction

A total of 17 pregnant women with singleton pregnancies, including eight in their first trimester (7 to 12 gestational weeks) and nine in their second trimester (13 to 24 gestational weeks), were enlisted at the Sixth Affiliated Hospital of Sun Yat-sen University. Their husbands (alleged fathers) were also included in the study. Three milliliters of peripheral blood were collected in EDTA-containing tubes. Amniotic fluid or chorionic villi were collected for medical purpose by professional operations at the hospital. Supplementary Table S1 shows the characteristics of the donors. All samples were anonymously collected after informed consent was obtained. All the experimental processes in this study strictly followed ethical research principles, and all methods were performed in accordance with

the relevant guidelines and regulations. This study was approved by the Ethics Committee of Sun Yat-sen University (No. [2019]62).

For both the pregnant women and their husbands, blood stain papers were prepared. To obtain the DNA of the child, maternal plasma was isolated from peripheral blood by a two-step centrifugation method [24,25]. cfDNA from 1-milliliter plasma samples was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) on a QIAvac 24 Plus vacuum manifold (Qiagen, Hilden, Germany) according to the manufacturer's instructions and subsequently eluted to 20 microliters. Genomic DNA (gDNA) was extracted from the amniotic fluid using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) while that from chorionic villus was extracted via the Chelex-100 method. The extracted DNA was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, San Francisco, CA, USA) with a Qubit 3.0 fluorometer (Thermo Fisher Scientific, San Francisco, CA, USA) according to the manufacturer's instructions.

2.2. Library Preparation and Massively Parallel Sequencing

DNA libraries were prepared with 1 ng input DNA or two punched blood stained papers (1.0 mm²) using the ForenSeq DNA Signature Prep Kit (DNA Primer Mix A, DPMA; Verogen., San Diego, CA, USA) according to the manufacturer's recommendations. Sequencing was performed on a MiSeq FGx system (Verogen., San Diego, CA, USA) using the MiSeq FGx Reagent Kit (Verogen., San Diego, CA, USA) following the manufacturer's instructions.

2.3. Pedigree Confirmation by CE-Based STR Typing

Both parents and their fetus (amniotic fluid or chorionic villus) were genotyped using the Goldeneye 25A kit (Peoplespot, Beijing, China), which included 23 A-STR loci, according to the manufacturer's instructions. Then, the PCR-products were separated on an AB 3500xL Genetic Analyzer (Thermo Fisher Scientific, San Francisco, CA, USA) and analyzed using the GeneMapper ID-X Version 1.5 (Thermo Fisher Scientific, San Francisco, CA, USA). Paternity index (PI) values were calculated according to the technical specification for the paternity test (GB/T 37223-2018) implemented in China [26]. Paternity inclusion was affirmed with a combined paternity index (CPI) value of greater than 10,000, whereas paternity exclusion was affirmed with a CPI of less than 0.0001.

2.4. Genotype Calling

Genotype calling of SNP and STR markers was performed with STRait Razor 3.0 [27] and an in-house R script. The workflow for genotype calling is shown in Figure 1. For nonplasma samples, a threshold of the locus-specific depth of 15× and 10×, and a heterozygote ratio of 0.2 and 0.35 were applied for SNP and STR, respectively. Heterozygote ratio was calculated by dividing the lower coverage allele by the higher coverage allele at each locus.

For plasma samples, a minimum depth of 50× was required for SNP allele calling. We set 2% as the threshold to distinguish paternally derived alleles from background noise in cfDNA as studied before [20]. We also formulated a rule to define an expected paternally inherited allele (*e*PIA) and observed paternally inherited allele (*o*PIA). Specifically, when a fetus was identified to have an allele that the mother did not have at a locus, this allele was defined as an *e*PIA. For example, if the mother showed the homozygous genotype of AA and the fetus showed the heterozygous genotype of AB, allele B was an expected paternally derived allele, or to say *e*PIA. Then, if the plasma was also identified to have an allele that the mother did not have at that locus, this allele was defined as an *o*PIA. For the above example, if the number of reads of the minor SNP allele B in corresponding plasma exceeded the noise threshold, allele B was an observed paternally derived allele, or to say, *o*PIA, and the inferred genotype was supposed as AB.

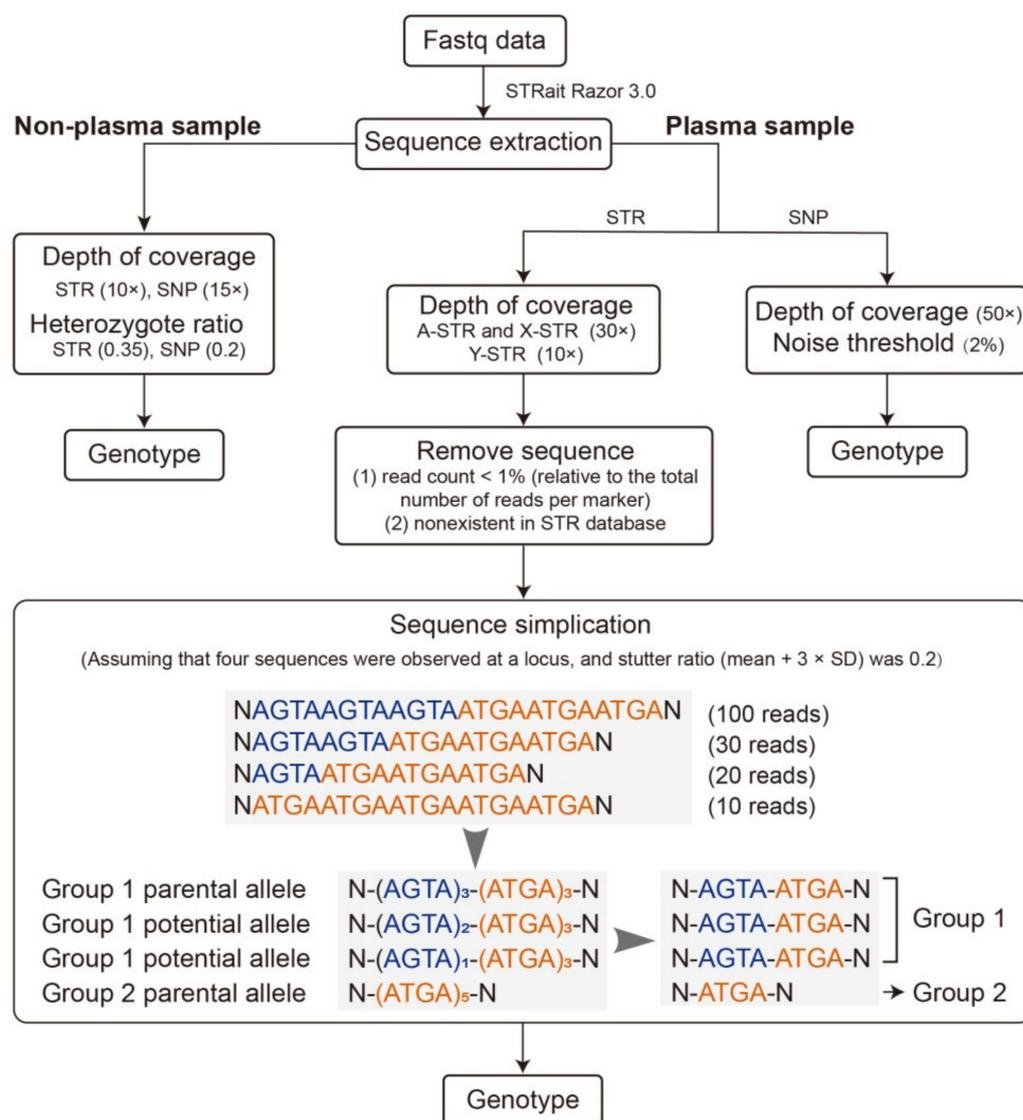


Figure 1. The workflow for genotype calling.

For A-STR and X-STR allele calling in plasma samples, a minimum depth of 30× was required, while it was 10× for Y-STR. In STR analysis, *ePIA* and *oPIA* have the same definition as that in the analysis of SNPs, but the process of identifying alleles in plasma was more complicated. A novel strategy, inspired by our previous study about the characteristics of stutter variants in STR genotyping [23], was developed for the identification of STR alleles in plasma. Firstly, sequences with read counts of <1% were excluded for each locus, assuming that they were mainly PCR-derived noises or sequencing errors. Secondly, sequences that were not observed in the self-established STR sequence database, which consisted of 1527 alleles based on four Chinese populations, one Nigerian population [28], and four major American populations [29], were eliminated. All the alleles detected in nonplasma samples of both parents were added to the STR sequence database to avoid exclusion of novel alleles. Thirdly, a sequence simplification process for the remaining sequences was performed following Li et al.’s strategy [23]. According to this method, sequences with the same simplified architecture were categorized into one group and the sequence with the highest read count was considered as the parental allele of that group. For nonparental sequences within a specific group, sequences sharing the same simplified architecture as the parental allele were considered as potential paternal alleles if their read counts—(1) exceeded the mean plus three-fold of the stutter ratio standard deviation (SD) for alleles at the $N - 1$ and $N + 1$ positions; and (2) were higher than half the read

counts of their corresponding parental alleles (e.g., for the sequence of $N - 2$ position, its corresponding parental allele is the sequence of $N - 1$ position) for alleles at other stutter positions. Finally, the genotype of the plasma was determined with a combination of the parental alleles and potential alleles at one locus for a specific sample. As a result, *o*PIAs were identified as the combination of the parental alleles and potential alleles minus maternal alleles.

2.5. Data Analysis

The dropout rate and drop-in rate for STRs and SNPs were obtained from a comparison between *o*PIAs and *e*PIAs. The fetal fraction in cfDNA was calculated using the formula:

$$\frac{2 \times n_{\text{father}}}{n_{\text{father}} + n_{\text{mother}}} \quad (1)$$

where n_{father} and n_{mother} represent the reads number of the *o*PIA and maternal allele for one SNP, respectively. The fetal fraction of a plasma was estimated by averaging the fractions of all loci. Detection rate of the *e*PIA for SNPs was computed as:

$$\frac{n_{\text{correct}}}{n_{\text{expected}}} \quad (2)$$

where n_{correct} stands for the number of *o*PIAs consistent with paternal alleles of the reference fetus while n_{expected} stands for the number of *e*PIAs.

2.6. PI and CPI Calculation for NIPPT

To verify discrimination power of the system, the paternity of 16 unrelated subjects (the alleged fathers outside the studied family) were tested for each case, resulting in a total of 272 (16 alleged fathers \times 17 families) negative tests. PI values were calculated using the R package relMix (Available online: <http://cran.r-project.org/web/packages/relMix> (accessed on 14 December 2020)), in which both mutations and the probabilities of dropout and/or drop-in were considered [30]. The CPI was calculated as the product of PIs and an LR threshold of 10,000, 1000, 100, and 10 was used as the threshold for confirmation of the parent–child relationship.

3. Results

3.1. Performance of the ForenSeq DNA Signature Prep Kit

The ForenSeq DNA Signature Prep Kit (DPMA) allows targeted genotyping of 152 forensically relevant genetic markers in a single reaction, including 94 iiSNPs, 27 A-STRs, seven X-STRs, and 24 Y-STRs. As shown in Figure 2, these markers are diffusely distributed in different chromosomes, and the amplicon length of over 90% of SNP loci and nearly 80% of A-STR loci were shorter than 160 bp. However, for X-STR loci and Y-STR loci, approximately 86% and 66% of the loci exceeded 160 bp.

Prior to library generation, parent–offspring relationships were confirmed by STR genotyping with Goldeneye 25A system for all 17 families. Then, 68 samples from 17 families were sequenced with three sequencing runs. The average cluster density was 563 k/mm² and the average percentage of the clusters passing filter, phasing, and prephasing were 97.95%, 0.17% and 0.11%, respectively. The average number of reads for each sample was 85,899.

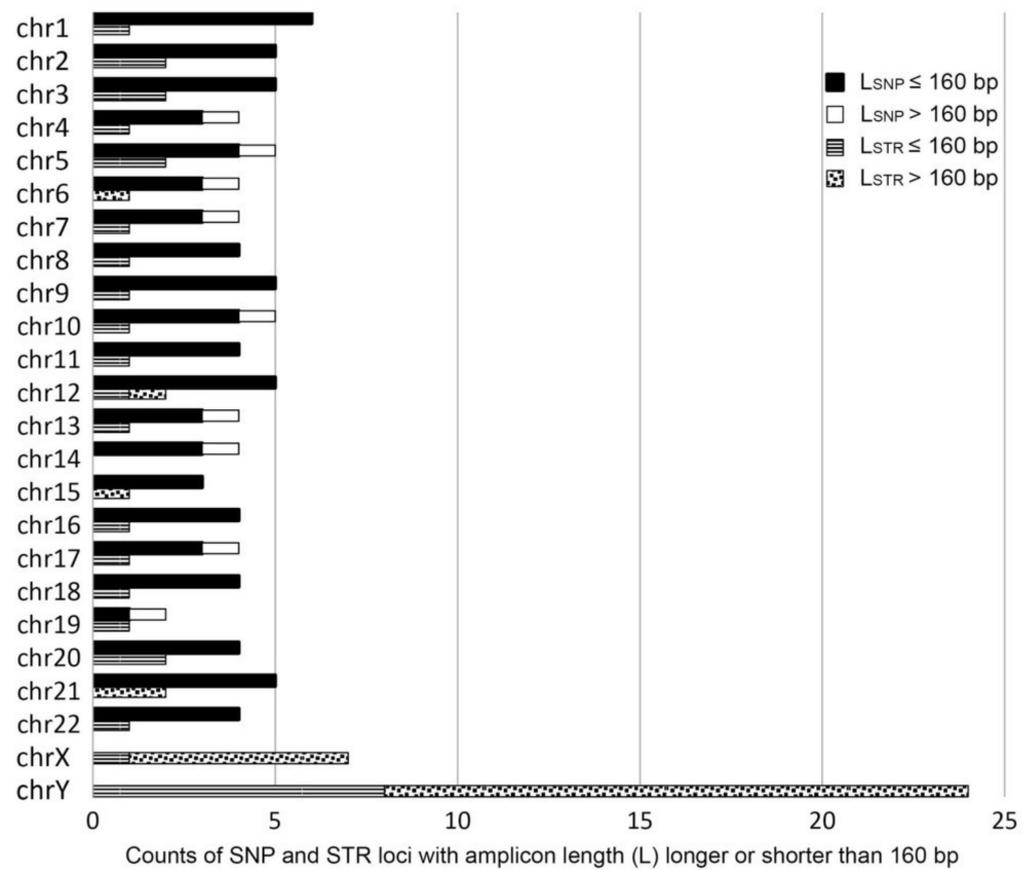


Figure 2. The density of single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) across different chromosomes and the number of SNP and STR loci with amplicon length (L) longer or shorter than 160 bp.

3.2. PIA Identification from Plasma for NIPPT

For SNPs, a total of 313 *ePIAs* were expected in plasma across 17 families and the average number of *ePIAs* was 18 (range—10 to 22). The sequencing result of plasma showed that a total of 240 *oPIAs* were identified across 17 families with and the average number of *oPIAs* was 14 (range—1 to 21) (Table 1). That is, approximately 76.36% of paternal fetal SNP alleles could be detected on average. The fetal fraction in the plasma ranged from 0.32% to 50.3% (Supplementary Table S2). Full paternal fetal SNP alleles could be identified in cases with high fetal fractions; for example, in case 17 (12.57 gestational weeks; fetal fraction = 50.3%). In contrast, only one paternally derived fetal SNP allele was attained in case 14, which was at 7 gestational weeks with a fetal fraction of 0.32%. Among 240 *oPIAs*, 239 were consistent with profiles of the reference fetuses with one remaining false positive allele (fetal fraction = 4.69%) observed at rs2076848 in case 9. The overall error rate was 0.4%, similar to the error rate estimated by Christiansen et al. [10].

With respect to STRs, the average number of *ePIAs* for A-STRs, X-STRs, and Y-STRs was 17 (range—11 to 21), two (range—0 to 4), and three (range—0 to 24), respectively. The average number of *oPIAs* of A-STRs, X-STRs, and Y-STRs was eight (range—2 to 17), one (range—0 to 2), and 13 (range—11 to 15), respectively (Table 1). Twenty-three of 144 *oPIAs* at A-STRs were found to be false positives. Thereinto, 14, four, and five alleles resulted from $N - 1$ stutter, $N + 1$ stutter, and others (including stutters at other positions except for $N - 1$ and $N + 1$, microvariant allele, etc.), respectively (Supplementary Table S3). The drop-out rate and drop-in rate for A-STRs were 58% and 16%, respectively. For X-STR, 50% of *oPIAs* were found to be false positive and 75% of them resulted from $N - 1$ stutter. Y-STR profiles observed in the male fetuses of case 3 and case 21 totally matched those of biological fathers.

Table 1. Number of expected paternally inherited allele (*ePIAs*) and observed paternally inherited allele (*oPIAs*) identified in maternal plasma samples.

| Case | SNPs | | A-STRs | | X-STRs | | Y-STRs | |
|------|--------------|---------------------|--------------|---------------------|--------------|--------------------|--------------|--------------|
| | <i>ePIAs</i> | <i>oPIAs</i> | <i>ePIAs</i> | <i>oPIAs</i> | <i>ePIAs</i> | <i>oPIAs</i> | <i>ePIAs</i> | <i>oPIAs</i> |
| 1 | 18 | 14 | 21 | 7 + 1 ¹ | 2 | 1 ¹ | 0 | 0 |
| 2 | 21 | 16 | 21 | 9 + 2 ¹ | 0 | 0 | 27 | 11 |
| 3 | 21 | 14 | 20 | 7 + 2 ¹ | 3 | 0 | 0 | 0 |
| 4 | 17 | 15 | 19 | 10 | 4 | 1 + 1 ¹ | 0 | 0 |
| 5 | 19 | 16 | 18 | 9 | 1 | 1 | 0 | 0 |
| 6 | 18 | 14 | 19 | 3 + 3 ¹ | 4 | 0 | 0 | 0 |
| 7 | 10 | 9 | 16 | 7 + 3 ¹ | 3 | 1 ¹ | 0 | 0 |
| 8 | 17 | 14 | 14 | 4 | 2 | 1 ¹ | 0 | 0 |
| 9 | 19 | 18 + 1 ¹ | 16 | 8 + 2 ¹ | 4 | 1 | 0 | 0 |
| 10 | 16 | 14 | 14 | 5 + 2 ¹ | 4 | 0 | 0 | 0 |
| 11 | 15 | 9 | 16 | 5 + 1 ¹ | 3 | 2 ¹ | 0 | 0 |
| 12 | 20 | 14 | 11 | 7 | 3 | 1 | 0 | 0 |
| 13 | 22 | 18 | 17 | 8 + 2 ¹ | 3 | 1 + 2 ¹ | 0 | 0 |
| 14 | 21 | 1 | 15 | 1 + 1 ¹ | 2 | 0 | 0 | 0 |
| 15 | 19 | 14 | 17 | 4 + 3 ¹ | 3 | 1 | 0 | 0 |
| 16 | 22 | 21 | 13 | 10 + 1 ¹ | 0 | 0 | 27 | 15 |
| 17 | 18 | 18 | 19 | 17 | 2 | 2 | 0 | 0 |

¹ False positive *oPIA*, which is inconsistent with the profile of the reference fetus.

3.3. Influencing Factors on the Detection Rate

For SNPs, the detection rate of *ePIAs* differed considerably, which ranged from 4.76% at case 14 to 100% at case 17. Also, the fetal fraction in the plasma varied greatly in a range between 0.32% and 50.3% (Supplementary Table S2). In order to explore the influencing factors of detection rate, dispersion plots were constructed using the value of the fetal fraction, length of the amplicon, and gestational week. As shown in Figure 3, the detection rate was correlated with fetal fraction and length of the amplicon, but not correlated with the gestational week.

3.4. CPIs with SNP and STR Typing for NIPPT

The average $\text{Log}_{10}\text{CPI}$ value of the 17 families was 3.10 (range—0.28 to 6.27) when only 94 iiSNPs were applied and was 4.73 (range—0.64 to 8.86) when only STRs were analyzed (Table 2). With an LR threshold of 10,000, 1000, 100, and 10 [31], a total of three, nine, 13, and 16 out of 17 family cases' paternity could be determined using SNPs, respectively, and the values were six, 11, 14, and 14, respectively, when only A-STRs were used. X-STRs and Y-STRs were not considered for the PI calculation since little effective information could be obtained for cases presented in this study. As there is no significant LD in the Chinese population for these iiSNPs and A-STRs after Bonferroni correction [22,32], the PI values of iiSNPs and A-STRs were multiplied to calculate the CPI values. When combining the SNP and A-STR marker, 82.35% (14/17), 88.24% (15/17), 94.12% (16/17) and 94.12% (16/17) of real family cases could be accurately determined with an LR threshold for inclusion set as 10,000, 1000, 100 and 10, respectively (Table 3). The three cases with $\text{Log}_{10}\text{CPI} < 4$ were case 6, case 14 and case 15, with the $\text{Log}_{10}\text{CPI}$ were 3.25, -0.36 and 2.94, respectively. As shown in Figure 4 and Table 2, the $\text{Log}_{10}\text{CPI}$ values obtained from 272 unrelated trios were all below four and separated significantly from the values of corresponding real trios except in case 14, which had the lowest fetal fraction and detection rate among 17 families.

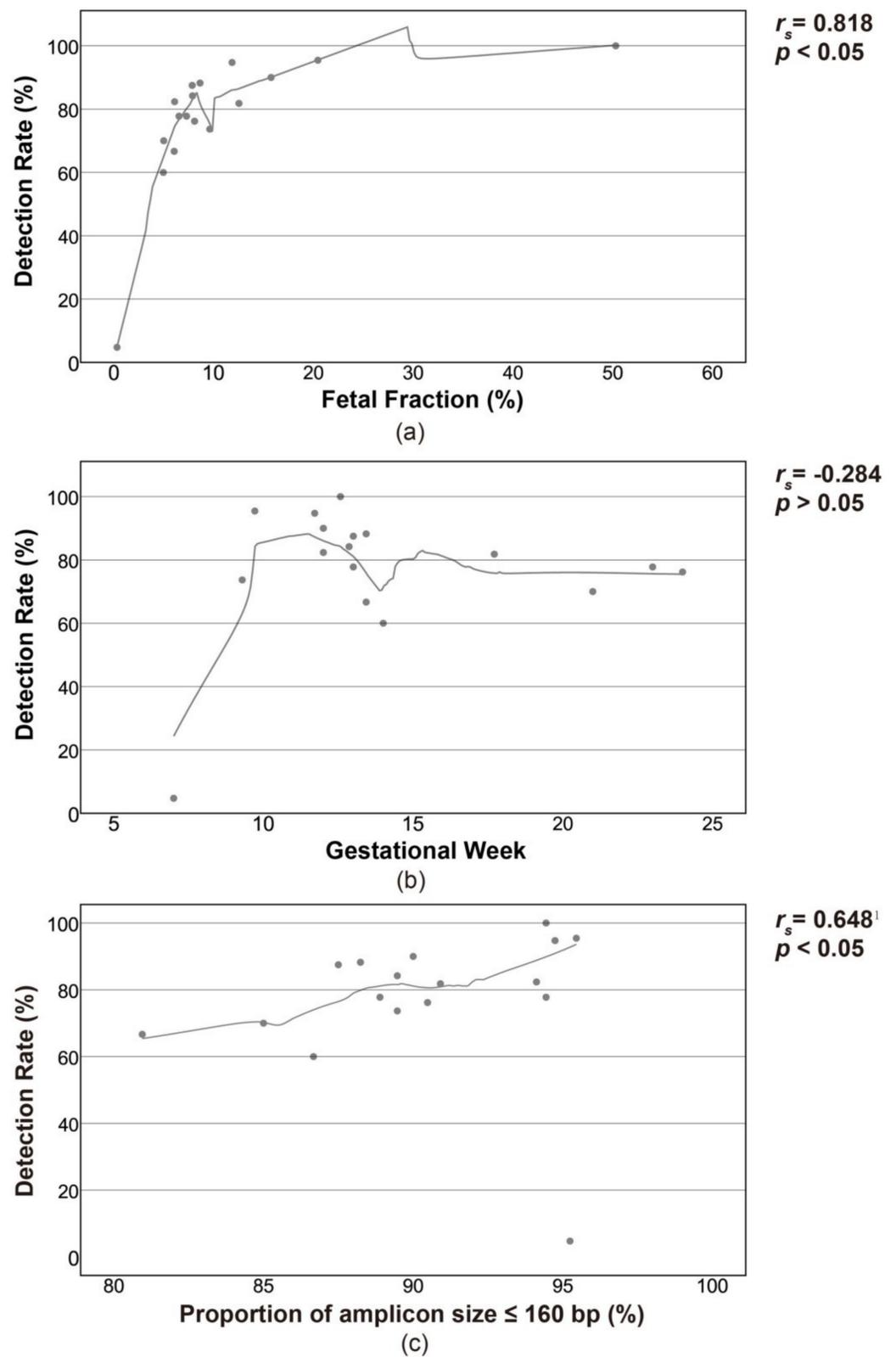


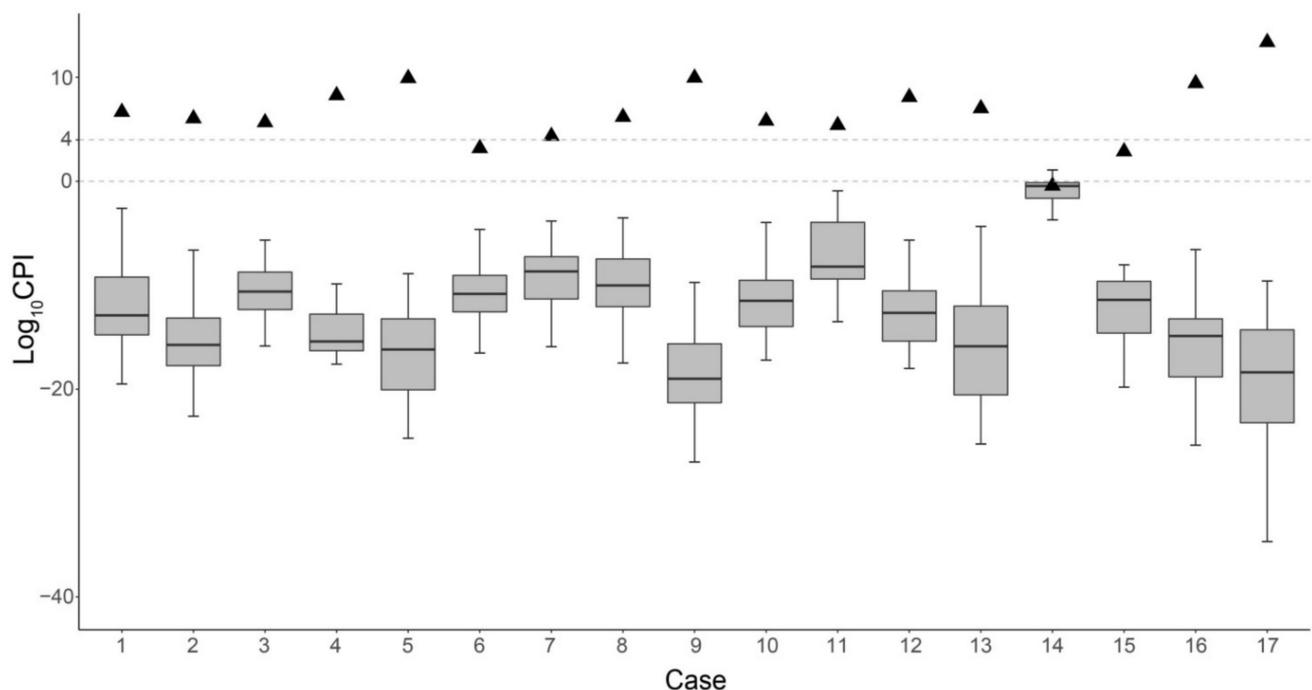
Figure 3. Correlation between detection rate for ePIAs and fetal fraction (a), gestational week (b) and amplicon size (c). ¹ An outlier was excluded from statistical analysis.

Table 2. Log₁₀CPI values for 17 families.

| Case | Sum of <i>o</i> PIAs for SNPs and A-STRs | Log ₁₀ CPI Based on SNP Typing | Log ₁₀ CPI Based on A-STR Typing | Log ₁₀ CPI Based on SNP and A-STR Typing |
|------|--|---|---|---|
| 1 | 22 | 2.91 | 3.82 | 6.73 |
| 2 | 27 | 2.65 | 3.47 | 6.12 |
| 3 | 23 | 1.78 | 3.96 | 5.74 |
| 4 | 25 | 3.86 | 4.48 | 8.34 |
| 5 | 25 | 3.74 | 6.25 | 9.99 |
| 6 | 20 | 3.08 | 0.17 | 3.25 |
| 7 | 19 | 1.47 | 2.95 | 4.42 |
| 8 | 18 | 3.92 | 2.34 | 6.26 |
| 9 | 29 | 6.27 | 3.76 | 10.03 |
| 10 | 21 | 3.50 | 2.38 | 5.88 |
| 11 | 15 | 1.79 | 3.68 | 5.47 |
| 12 | 21 | 3.39 | 4.78 | 8.17 |
| 13 | 28 | 2.63 | 4.47 | 7.10 |
| 14 | 3 | 0.28 | −0.64 | −0.36 |
| 15 | 21 | 2.24 | 0.70 | 2.94 |
| 16 | 32 | 4.42 | 5.08 | 9.50 |
| 17 | 35 | 4.60 | 8.86 | 13.46 |

Table 3. Combined paternity index (CPI) distribution for real trios and unrelated trios with the combination of SNPs and autosomal STRs (A-STRs).

| LR Threshold | Real Parentage (<i>n</i> = 17) | Unrelated (<i>n</i> = 272) |
|--------------|---------------------------------|-----------------------------|
| >10 | 94.12% | 0.36% |
| >100 | 94.12% | 0 |
| >1000 | 88.24% | 0 |
| >10,000 | 82.35% | 0 |

**Figure 4.** Log₁₀CPI distribution in noninvasive prenatal paternity testing (NIPPT). Triangles represent the Log₁₀CPI values obtained from 17 real trios. Boxplots show the distribution of Log₁₀CPI values obtained from 272 unrelated trios.

4. Discussion and Conclusions

Paternity testing is often necessary in a wide range of situations to provide DNA evidence for the forensic community. Therein, antenatal paternity testing is of special use when investigating crimes related to rape or incest. Compared with invasive prenatal paternity testing by amniocentesis and chorionic villus sampling, noninvasive prenatal paternity testing based on cfDNA is less harmful and has stronger timeliness because testing could be conducted earlier, which reduces potential risks to both mother and fetus. However, most studies have focused on self-designed SNP typing systems with loci that have not been included in reference databases or commercial kits. Data communication and standard establishment were difficult, let alone the application of these panels in forensic practice. To take advantage of well-established STR markers, we investigated the utility of a standard forensic multiplex assay system, which enables genotyping of SNPs and STRs simultaneously, in NIPPT.

SNP typing results showed that, on average, 76.36% of paternal SNP alleles could be detected, reaching up to 14 *o*PIAs per case. Detection rates of *e*PIAs for SNPs varied among different cases but tended to increase with fetal fractions and proportions of short amplicons. Since cfDNA is mostly DNA fragments less than 160 bp, panels consisting of shorter amplicons will inevitably yield a higher success rate compare to that including a higher proportion of long amplicons. The strong correlation between detection rate and gestational week, which was reported in previous studies, was not observed in this work [10,20]. The most probable reason is that the pregnant women enrolled in the study were not from various gestational ages but concentrated at 12 to 14 gestational weeks, and the differences in detection rates were individual differences.

Although the existence of cfDNA in maternal plasma allows noninvasive prenatal paternity testing, fetal DNA generally comprises <20% of cfDNA in maternal plasma [13], which demands sensitive techniques and appropriate analytical methods to be detected. In the past years, SNP markers were mostly used in NIPPT due to their short amplicon sizes and absence of stutters [33]. However, thousands of markers are recommended to achieve higher accuracy, due to the limited polymorphisms of SNP markers [16]. Moreover, whether it is the widely-explored SNP marker or the newly-recruited microhaplotype marker, self-designed panels are unfavorable for data communication and utilization of public reference database, and usually lack evaluation. STR markers, by contrast, have multiple alleles and high information content and thus, are the gold standard marker for relationship analysis [34]. With the same number of markers, STRs are superior to SNPs in conventional parentage testing. However, the length of amplicons in CE typing and the presence of stutters both limited the application of STR markers in NIPPT [11,12]. STR markers are the main markers for the forensic community, both now and in the future. If the advancement of well-established STR markers could be taken, the application of NIPPT could be broadened. In the present study, ForenSeq DNA Signature Prep Kit, which targeting well-established STRs and SNPs, and with shorter amplicons than the common CE typing method, was applied for NIPPT for the purpose of increasing the STR genotyping success rate. Besides, investigation of paternally derived fetal STR alleles was performed with a novel strategy, which may bring improvement for mixture deconvolution under stutter interference [23]. However, the result demonstrated that there still was a risk of misidentifying sequencing error or stutter artifacts as paternal alleles due to the complexity of STR variations and the shortage of PCR-based methods. As the molecular architecture of STRs contributes to the generation of stutters [35], including more A-STRs to improve discriminating power may become a solution. Furthermore, a bioinformatics pipeline based on unique molecular identifier (UMI) technology was developed recently, which allows PCR or sequencing errors to be identified efficiently and, subsequently, increase the accuracy of allele interpretation during MPS-based STR genotyping [36]. Therefore, recruitment of this method in NIPPT with well-established STR markers could be conducted in the next step.

The result also showed that a single type of marker in the ForenSeq DNA Signature Prep Kit may not be sufficient for NIPPT, and a combination of multiple types of markers (SNPs and STRs) will improve the testing efficiency. For the three cases with paternity that could not be determined with the LR threshold of 10,000, the stage of pregnancy should be the main reason. In fact, they are amongst the earliest gestational weeks within this study. The earlier the gestational age, the higher probability of low fetal fraction in the plasma, which would further determine a lower *o*PIA detection rate and, subsequently, low CPI. Therefore, when using this panel to do NIPPT for women at early gestational weeks, false negatives may occur, and it should be more cautious to exclude paternity. Even so, this study still indicated that the ForenSeq DNA Signature Prep Kit could solve a high proportion of NIPPT cases.

However, the application of NIPPT should be very cautious. From a technical aspect, either false negative or false positive PIAs might occur, which will further affect the testing result and make it inconclusive. In this condition, additional tests must be done before determining paternity inclusion or exclusion. Furthermore, ethical and social issues should always be considered [37–40]. One of the major aspects to consider is the protection of under-age children who are not able to express their intention and are vulnerable in the face of setbacks [37]. For noninvasive paternity testing which involves embryos, both rights of the embryos and the parents should be protected. Therefore, the testing should be conducted even more cautiously, especially in countries where termination of pregnancy is allowed for nonmedical reasons. The participants' ability to deal with the outcome of NIPPT should be considered, and they should always be warned of the potential implications of such testing. Overall, NIPPT should be conducted within the context that highly sensitive and accurate NIPPT methods are developed and ethical aspects are fully considered. To our knowledge, this is the first study exploring the utilization of the ForenSeq DNA Signature Prep Kit in NIPPT. It revealed that NIPPT was possible with a combination of well-established SNPs and STRs. The results of this study may facilitate the standardization of NIPPT in forensic practice. However, more efforts are needed on the selection of effective markers and development of interpretation strategies, to aid in the application of NIPPT in forensic applications.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2073-4425/12/3/454/s1>, Table S1: Characteristics of the donors, Table S2: Statistical parameters of cases in the test, Table S3: Characteristics of false positive *o*PIAs.

Author Contributions: Conceptualization, H.S.; Methodology, X.S. and R.L.; Software, R.L. and X.S.; Validation, X.S., R.L. and H.L.; Formal Analysis, X.S., R.L. and H.L.; Investigation, X.S., R.L., Y.G., H.C., N.Q. and D.P.; Resources, X.S., H.C. and D.P.; Data Curation, X.S. and R.L.; Writing—Original Draft Preparation, X.S.; Writing—Review and Editing, H.S. and R.W.; Visualization, H.S., R.W. and X.S.; Supervision, H.S.; Project Administration, H.S. and R.W.; Funding Acquisition, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the National Natural Science Foundation of China (Grants No. 81971798) and the Natural Science Foundation of Guangdong Province (Grants No. 2019A15150).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Zhongshan School of Medicine, Sun Yat-Sen University (protocol code 2019-062, date of approval 10 June 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the restriction of privacy.

Acknowledgments: We are grateful to all volunteers who provided the samples used in this study. We thanked Xueling Ou for collecting samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brambati, B.; Simoni, G.; Fabro, S. *Chorionic villus Sampling: Fetal Diagnosis of Genetic Diseases in the First Trimester*; M. Dekker: New York, NY, USA, 1986; p. 310.
2. Brambati, B.; Simoni, G.; Travi, M.; Danesino, C.; Tului, L.; Privitera, O.; Stioui, S.; Tedeschi, S.; Russo, S.; Primignani, P. Genetic diagnosis by chorionic villus sampling before 8 gestational weeks: Efficiency, reliability, and risks on 317 completed pregnancies. *Prenat. Diagn.* **1992**, *12*, 789–799. [[CrossRef](#)] [[PubMed](#)]
3. Elchalal, U.; Ben Shachar, I.; Peleg, D.; Schenker, J.G. Maternal Mortality following Diagnostic 2nd-Trimester Amniocentesis. *Fetal Diagn. Ther.* **2004**, *19*, 195–198. [[CrossRef](#)]
4. Firth, H.; Boyd, P.; Chamberlain, P.; MacKenzie, I.Z.; Morriss-Kay, G.; Huson, S. Analysis of limb reduction defects in babies exposed to chorionic villus sampling. *Lancet* **1994**, *343*, 1069–1071. [[CrossRef](#)]
5. Cederholm, M.; Haglund, B.; Axelsson, O. Infant morbidity following amniocentesis and chorionic villus sampling for prenatal karyotyping. *BJOG Int. J. Obstet. Gynaecol.* **2005**, *112*, 394–402. [[CrossRef](#)]
6. Akolekar, R.; Beta, J.; Picciarelli, G.; Ogilvie, C.; D’Antonio, F. Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: A systematic review and meta-analysis. *Ultrasound Obstet. Gynecol.* **2015**, *45*, 16–26. [[CrossRef](#)] [[PubMed](#)]
7. Lo, Y.M.D.; Corbetta, N.; Chamberlain, P.F.; Rai, V.; Sargent, I.L.; Redman, C.W.; Wainscoat, J.S. Presence of fetal DNA in maternal plasma and serum. *Lancet* **1997**, *350*, 485–487. [[CrossRef](#)]
8. Tong, Y.-K.; Lo, Y.D. Diagnostic developments involving cell-free (circulating) nucleic acids. *Clin. Chim. Acta* **2006**, *363*, 187–196. [[CrossRef](#)] [[PubMed](#)]
9. Yu, S.C.; Lee, S.W.; Jiang, P.; Leung, T.Y.; Chan, K.A.; Chiu, R.W.; Lo, Y.D. High-Resolution Profiling of Fetal DNA Clearance from Maternal Plasma by Massively Parallel Sequencing. *Clin. Chem.* **2013**, *59*, 1228–1237. [[CrossRef](#)]
10. Christiansen, S.L.; Jakobsen, B.; Børsting, C.; Udengaard, H.; Buchard, A.; Kampmann, M.-L.; Grøndahl, M.L.; Morling, N. Non-invasive prenatal paternity testing using a standard forensic genetic massively parallel sequencing assay for amplification of human identification SNPs. *Int. J. Leg. Med.* **2019**, *133*, 1361–1368. [[CrossRef](#)]
11. Tang, D.-L.; Li, Y.; Zhou, X.; Li, X.; Zheng, F. Multiplex fluorescent PCR for noninvasive prenatal detection of fetal-derived paternally inherited diseases using circulatory fetal DNA in maternal plasma. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2009**, *144*, 35–39. [[CrossRef](#)]
12. Wagner, J.; Džijan, S.; Marjanovic, D.; Lauc, G. Non-invasive prenatal paternity testing from maternal blood. *Int. J. Leg. Med.* **2008**, *123*, 75–79. [[CrossRef](#)] [[PubMed](#)]
13. Zhou, Y.; Zhu, Z.; Gao, Y.; Yuan, Y.; Guo, Y.; Zhou, L.; Liao, K.; Wang, J.; Du, B.; Hou, Y.; et al. Effects of Maternal and Fetal Characteristics on Cell-Free Fetal DNA Fraction in Maternal Plasma. *Reprod. Sci.* **2015**, *22*, 1429–1435. [[CrossRef](#)]
14. Fan, H.C.; Blumenfeld, Y.J.; Chitkara, U.; Hudgins, L.; Quake, S.R. Analysis of the Size Distributions of Fetal and Maternal Cell-Free DNA by Paired-End Sequencing. *Clin. Chem.* **2010**, *56*, 1279–1286. [[CrossRef](#)]
15. Pan, M.; Chen, P.; Lu, J.; Liu, Z.; Jia, E.; Ge, Q. The fragmentation patterns of maternal plasma cell-free DNA and its applications in non-invasive prenatal testing. *Prenat. Diagn.* **2020**, *40*, 911–917. [[CrossRef](#)]
16. Jiang, H.; Xie, Y.; Li, X.; Ge, H.; Deng, Y.; Mu, H.; Feng, X.; Yin, L.; Du, Z.; Chen, F.; et al. Noninvasive Prenatal Paternity Testing (NIPAT) through Maternal Plasma DNA Sequencing: A Pilot Study. *PLoS ONE* **2016**, *11*, e0159385. [[CrossRef](#)]
17. Ou, X.; Wang, H.; Qu, D.; Chen, Y.; Gao, J.; Sun, H. Epigenome-wide DNA methylation assay reveals placental epigenetic markers for noninvasive fetal single-nucleotide polymorphism genotyping in maternal plasma. *Transfusion* **2014**, *54*, 2523–2533. [[CrossRef](#)]
18. Moriot, A.; Hall, D. Analysis of fetal DNA in maternal plasma with markers designed for forensic DNA mixture resolution. *Genet. Med.* **2019**, *21*, 613–621. [[CrossRef](#)] [[PubMed](#)]
19. Bai, Z.; Zhao, H.; Lin, S.; Huang, L.; He, Z.; Wang, H.; Ou, X. Evaluation of a Microhaplotype-Based Noninvasive Prenatal Test in Twin Gestations: Determination of Paternity, Zygosity, and Fetal Fraction. *Genes* **2020**, *12*, 26. [[CrossRef](#)]
20. Chang, L.; Yu, H.; Miao, X.; Zhang, J.; Li, S. Development and comprehensive evaluation of a noninvasive prenatal paternity testing method through a scaled trial. *Forensic Sci. Int. Genet.* **2019**, *43*, 102158. [[CrossRef](#)] [[PubMed](#)]
21. Qu, N.; Xie, Y.; Li, H.; Liang, H.; Lin, S.; Huang, E.; Gao, J.; Chen, F.; Shi, Y.; Ou, X. Noninvasive prenatal paternity testing using targeted massively parallel sequencing. *Transfusion* **2018**, *58*, 1792–1799. [[CrossRef](#)]
22. Guo, F.; Yu, J.; Zhang, L.; Li, J. Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System. *Forensic Sci. Int. Genet.* **2017**, *31*, 135–148. [[CrossRef](#)] [[PubMed](#)]
23. Li, R.; Wu, R.; Li, H.; Zhang, Y.; Peng, D.; Wang, N.; Shen, X.; Wang, Z.; Sun, H. Characterizing stutter variants in forensic STRs with massively parallel sequencing. *Forensic Sci. Int. Genet.* **2020**, *45*, 102225. [[CrossRef](#)] [[PubMed](#)]
24. Chim, S.S.C.; Tong, Y.K.; Chiu, R.W.K.; Lau, T.K.; Leung, T.N.; Chan, L.Y.S.; Oudejans, C.B.M.; Ding, C.; Lo, Y.M.D. Detection of the placental epigenetic signature of the maspin gene in maternal plasma. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14753–14758. [[CrossRef](#)] [[PubMed](#)]
25. Tong, Y.K.; Ding, C.; Chiu, R.W.; Gerovassili, A.; Chim, S.S.; Leung, T.Y.; Leung, T.N.; Lau, T.K.; Nicolaidis, K.H.; Lo, Y.D. Noninvasive Prenatal Detection of Fetal Trisomy 18 by Epigenetic Allelic Ratio Analysis in Maternal Plasma: Theoretical and Empirical Considerations. *Clin. Chem.* **2006**, *52*, 2194–2202. [[CrossRef](#)] [[PubMed](#)]

26. State Administration for Market Regulation, China National Standardization Administration. Specification of Parentage Testing, GB/T 37223-2018, Issued on 2018-12-28. Available online: <https://www.chinesestandard.net/PDF/English.aspx/GBT37223-2018> (accessed on 28 December 2018).
27. Woerner, A.E.; King, J.L.; Budowle, B. Fast STR allele identification with STRait Razor 3.0. *Forensic Sci. Int. Genet.* **2017**, *30*, 18–23. [[CrossRef](#)]
28. Wu, R.; Peng, D.; Ren, H.; Li, R.; Li, H.; Wang, N.; Shen, X.; Huang, E.; Zhang, Y.; Sun, H. Characterization of genetic polymorphisms in Nigerians residing in Guangzhou using massively parallel sequencing. *Forensic Sci. Int. Genet.* **2020**, *48*, 102323. [[CrossRef](#)]
29. Novroski, N.M.; King, J.L.; Churchill, J.D.; Seah, L.H.; Budowle, B. Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Sci. Int. Genet.* **2016**, *25*, 214–226. [[CrossRef](#)]
30. Hernandis, E.; Dórum, G.; Egeland, T. relMix: An open source software for DNA mixtures with related contributors. *Forensic Sci. Int. Genet. Suppl. Ser.* **2019**, *7*, 221–223. [[CrossRef](#)]
31. Ge, J.; Budowle, B. How many familial relationship testing results could be wrong? *PLoS Genet.* **2020**, *16*, e1008929. [[CrossRef](#)]
32. Peng, D.; Zhang, Y.; Ren, H.; Li, H.; Li, R.; Shen, X.; Wang, N.; Huang, E.; Wu, R.; Sun, H. Identification of sequence polymorphisms at 58 STRs and 94 iSNPs in a Tibetan population using massively parallel sequencing. *Sci. Rep.* **2020**, *10*, 1–12. [[CrossRef](#)]
33. Yang, D.; Liang, H.; Gao, Y.; Lin, S.; He, Z.; Gao, J.; Sun, H.; Li, Q.; Ma, X.; Ou, X. Noninvasive fetal genotyping of paternally inherited alleles using targeted massively parallel sequencing in parentage testing cases. *Transfusion* **2017**, *57*, 1505–1514. [[CrossRef](#)] [[PubMed](#)]
34. Parson, W. Age Estimation with DNA: From Forensic DNA Fingerprinting to Forensic (Epi)Genomics: A Mini-Review. *Gerontology* **2018**, *64*, 326–332. [[CrossRef](#)]
35. Schlötterer, C.; Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **1992**, *20*, 211–215. [[CrossRef](#)]
36. Woerner, A.E.; Mandape, S.; King, J.L.; Muenzler, M.; Crysap, B.; Budowle, B. Reducing noise and stutter in short tandem repeat loci with unique molecular identifiers. *Forensic Sci. Int. Genet.* **2020**, *51*, 102459. [[CrossRef](#)] [[PubMed](#)]
37. Moray, N.; Pink, K.E.; Borry, P.; Larmuseau, M.H. Paternity testing under the cloak of recreational genetics. *Eur. J. Hum. Genet.* **2017**, *25*, 768–770. [[CrossRef](#)]
38. Caenazzo, L.; Comacchio, A.; Tozzo, P.; Rodriguez, D.; Benciolini, P. Paternity testing requested by private parties in Italy: Some ethical considerations. *J. Med. Ethics* **2008**, *34*, 735–737. [[CrossRef](#)]
39. Evans, M.I.; Andriole, S.; Pergament, E.; Speer, J.; Curtis, J.; Britt, D.W. Paternity Balancing. *Fetal Diagn. Ther.* **2013**, *34*, 135–139. [[CrossRef](#)] [[PubMed](#)]
40. Toya, W. Ethical, Legal and Social Issues in Japan on the Determination of Blood Relationship via DNA Testing. *Asian Bioeth. Rev.* **2017**, *9*, 19–32. [[CrossRef](#)]