Received: 22 July 2019 | Revised: 3 October 2019 | Accepted: 25 October 2019

DOI: 10.1002/pd.5595

ORIGINAL ARTICLE

PRENATAL DIAGNOSIS WILEY

Noninvasive prenatal paternity testing by means of SNP-based targeted sequencing

Jacqueline Chor Wing Tam | Yee Man Chan | Shui Ying Tsang | Chung In Yau | Shuk Ying Yeung | Ka Ki Au | Chun Kin Chow

Department of R&D, Medtimes Medical Group Limited, Kwai Chung, Hong Kong

Correspondence
Jacqueline Chor Wing Tam, Medtimes Medical Group Limited, Unit B, 7/F Roxy Industrial Centre, Kwai Chung, Hong Kong.
Email: jacqueline@medtimes.com.hk

Abstract

Objective: To develop a method for noninvasive prenatal paternity testing based on targeted sequencing of single nucleotide polymorphisms (SNPs).

Method: SNPs were selected based on population genetics data. Target-SNPs in cell-free DNA extracted from maternal blood (maternal cfDNA) were analyzed by targeted sequencing wherein target enrichment was based on multiplex amplification using QIAseq Targeted DNA Panels with Unique Molecular Identifiers. Fetal SNP genotypes were called using a novel bioinformatics algorithm, and the combined paternity indices (CPIs) and resultant paternity probabilities were calculated.

Results: Fetal SNP genotypes obtained from targeted sequencing of maternal cfDNA were 100% concordant with those from amniotic fluid-derived fetal genomic DNA. From an initial panel of 356 target-SNPs, an average of 148 were included in paternity calculations in 15 family trio cases, generating paternity probabilities of greater than 99.9999%. All paternity results were confirmed by short-tandem-repeat analysis. The high specificity of the methodology was validated by successful paternity discrimination between biological fathers and their siblings and by large separations between the CPIs calculated for the biological fathers and those for 60 unrelated men.

Conclusion: The novel method is highly effective, with substantial improvements over similar approaches in terms of reduced number of target-SNPs, increased accuracy, and reduced costs.

1 | INTRODUCTION

Paternity testing is conducted to determine the biological linkage between a child and an alleged father, and it can be done either before or after the birth of the child. A common method for postnatal paternity testing is the analysis of genetic information obtained from buccal swabs or other biological samples of the child and the alleged father to generate a probability of paternity. The main difficulty in implementing this approach in prenatal paternity testing lies in the procurement of fetal DNA. Currently, fetal DNA sampling methods can be divided into invasive and noninvasive sampling. Invasive sampling includes chorionic villus sampling or amniocentesis whereby amniotic fluid is obtained. Because invasive sampling induces a risk of miscarriage and infection,[1-3] these procedures are not recommended unless to aid in diagnosis of severe genetic disorders such as those related to fetal aneuploidy.[4]

Noninvasive sampling refers to maternal peripheral blood sampling wherein fetal DNA is present as cell-free DNA (cfDNA). Since the discovery of cell-free fetal DNA in maternal bloodstream circulation,[5] a variety of cfDNA-based methods have been developed for numerous clinical applications.[6-12] In terms of prenatal paternity

testing, early attempts used short tandem repeats (STRs) as genetic markers,[11,13] but because overwhelming maternal signals effectively concealed the fetal signals of autosomal STRs, only Y-chromosome STRs (Y-STRs) could be utilized, and this restricted application to only male fetuses.[10,14] Moreover, Y-STR analysis could not exclude relationships from the same male lineage, and the high mutation rate of Y-STRs (10^{-3} to 10^{-2} per locus per generation) increased the probability of false paternity exclusions.[15]

Use of single nucleotide polymorphisms (SNPs) as genetic markers can avoid STR-associated drawbacks, and consequently, SNP-based prenatal paternity tests have recently emerged as alternatives to STR-based methods. A major challenge of SNP-based tests is the accurate genotyping of fetal SNPs in the low fetal fraction (FF; average approximately 10% at 10-13 gestation weeks[16]) in cfDNA extracted from maternal blood (maternal cfDNA). High-density array chips[17-19] and high-throughput next-generation sequencing[20-22] are efficient SNP genotyping platforms, and both have shown success in this application. The use of targeted sequencing (hybridization-based target enrichment of 5000-8000 SNPs) and a Bayesian analysis approach successfully determined paternity in 17 clinical cases.[20] Likewise, Qu et al sequenced 1795 SNPs for successful paternity determination in 34 parentage test cases.[21] Target enrichment can also be amplicon-based, which allows the important implementation of molecular barcoding through the incorporation of Unique Molecular Identifiers (UMIs). Molecular barcoding combined with deep sequencing has demonstrated reliable detection of low frequency variants as UMI-based manipulations allow efficient correction of PCR or sequencing errors.[23]

In the present study, we demonstrate that with systematic SNP selection and UMI-based error correction, the number of target-SNPs can be significantly reduced, and targeted sequencing wherein target enrichment is by multiplex amplification can be effectively employed for prenatal paternity testing. A total of 15 parentage test cases as well as 903 negative tests with close male relatives (three tests) and unrelated individuals (15×60 tests) are reported to demonstrate its validity and potential utility in forensic and clinical settings.

What's already known about this topic?

- Cell-free fetal DNA in maternal blood circulation can be used in various prenatal applications including paternity testing.
- Fetal short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) can be used as genetic markers in prenatal paternity tests.

What does this study add?

- Targeted sequencing of maternal plasma-derived cell-free DNA wherein target-SNPs enrichment was amplicon-based as a method for noninvasive prenatal paternity testing.
- A systematic SNPs selection procedure that can significantly reduce the number of target-SNPs for sequencing analysis yet retain comparable discriminating power in paternity testing.
- A novel bioinformatics algorithm to allow accurate fetal SNP genotyping from targeted sequencing data of maternal cell-free DNA.

2 | MATERIALS AND METHODS

2.1 | Collection of samples

Peripheral blood samples were obtained from 15 pregnant mothers, and peripheral blood or buccal samples were obtained from the alleged fathers, close male relatives of the alleged fathers, and 60 unrelated men. Paired amniotic fluid samples collected at 16 to 19 weeks of gestation from two of the pregnant mothers were provided by the Prenatal Diagnostic Laboratory at Tsan Yuk Hospital (Hong Kong, China), and buccal samples were collected from the newborn in three other cases. All participants were of *Han* Chinese origin. Maternal peripheral blood samples (approximately 10 mL) were collected in cell-free DNA collection tubes (Roche, Basel, Switzerland), and peripheral blood samples from adult males (approximately 5 mL) were collected in Vacuette blood collection tubes (Greiner Bio-One, Kremsmünster, Austria). Buccal samples were collected using flocked swabs (Copan Diagnostics, Murrieta, CA, USA). Only singleton pregnancies were included in the study, and gestational ages at blood sampling were 7 to 20 weeks. Written informed consent was obtained from all participants, and the study was approved by the Medtimes Medical Group Ethics Review Board.

2.2 | Extraction of DNA

Genomic DNA was extracted from peripheral blood of male adults and from buccal swab and amniotic fluid samples using the QIAamp DNA Blood Mini kit (QIAGEN, Hilden, Germany). Maternal cfDNA was extracted from maternal plasma using the Maxwell RSC LV ccfDNA Custom Kit (Promega, Fitchburg, WI, USA). Concentrations of the extracted genomic DNA and cfDNA were measured using the NanoDrop Lite spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) and the Qubit dsDNA HS Assay Kit with the Qubit fluorometer (Thermo Fisher Scientific), respectively. All procedures were performed following the respective manufacturer's protocols for the respective sample types.

2.3 | Selection of SNPs

An initial panel of SNPs with minor allele frequencies greater than 0.30 and covering all 22 autosomes was selected as target-SNPs for sequencing (Table S1). This panel was selected based on population genetics data from the 1000 Genomes Project (www.1000genomes.

org) according to a list of defining criteria (Data S1) and for practical purposes, subjectively stipulated to include only 356 SNPs.

2.4 | Library preparation and sequencing

Sequencing libraries were prepared from the extracted genomic DNA and cfDNA using the QIAseq Targeted DNA Panels Kit (QIAGEN), following the manufacturer's respective protocols for the two DNA types. Briefly, 40-ng genomic DNA or 10 to 20-ng cfDNA per sample was used for library construction. The initial steps of fragmentation, end repair, and A-tailing were followed by adapter ligation, ligation of UMIs, and sample indexing. Ligated DNA was then subjected to target enrichment by performing an eight-cycle multiplex PCR with custom-designed QIAseq Targeted DNA Panel primers (QIAGEN) using a Thermocycler C1000 system (BioRad, Irvine, CA, USA). After enrichment, the DNA fragments were further amplified using universal primers by means of a 21-cycle PCR for genomic DNA or 23-cycle PCR for cfDNA. The enriched libraries were quantified using the QIAseq Library Quant Assay Kit (QIAGEN) and multiplex, paired-end sequenced using the MiniSeq Mid Output Kit on the Illumina MiniSeq sequencer (Illumina, San Diego, CA, USA).

2.5 | Sequencing data processing

The smCounter2 pipeline, specially designed for the accurate calling of low-frequency variants from QIAseq-based targeted sequencing data,[24] was employed for data processing and variant calling. UMI tags enabled error correction for most of the sequencing and PCR errors, and a refining algorithm was used to further amend those errors that were not correctable with UMI. Briefly, sequencing reads were trimmed, and the UMI sequences identified before alignment to the reference genome with BWA-MEM,[25] followed by filtering of poorly mapped reads and UMI clustering. Duplicated reads were filtered whereby reads sharing the same UMI and aligned to the same position were represented by the consensus read. The aligned reads were then employed in variant calling in which the data were processed to generate the number of reads corresponding to the reference and alternate alleles at target-SNPs, followed by their annotation. Target-SNPs with sequencing depths less than 100× were excluded from further analysis.

The smCounter2-called target-SNP genotypes derived from genomic DNA were directly employed in downstream analysis. The allele counts generated from cfDNA were used as input for a novel Bayesian-based algorithm to predict the combined maternal and fetal genotypes (maternal-fetal genotypes) at individual SNP loci (Figure 1; Data S1). The algorithm was essentially an extension of that described in Goya et al[33] and was implemented using an in-house R script. The SNP parameters were fitted to a set of hypothetical models that varied in FF by undergoing iterative Expectation-Maximization cycles. The model with the highest likelihood determined the set of maternal-fetal genotypes for each sample as well as its estimated FF.[26] Each maternal-fetal genotype was given a posterior probability that reflected the confidence of the call. Maternal-fetal genotypes with a probability less than 99.0% were excluded from downstream paternity calculations. Samples with FF less than or equal to 2.0% were given "Inconclusive" calls.

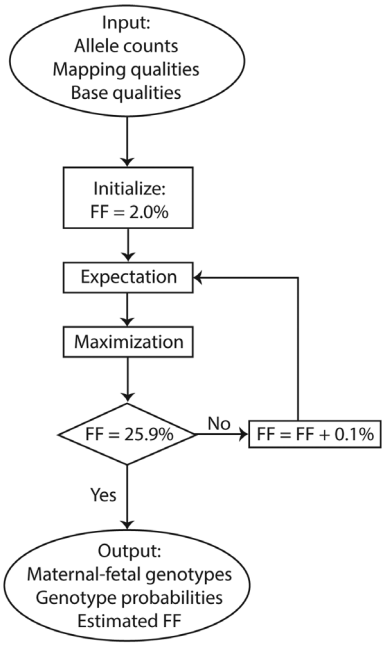

FIGURE 1 Bayesian-based algorithm in maternal-fetal genotype prediction. The workflow starts with the input of SNP parameters derived from the sequencing data into the algorithm wherein iterative Expectation-Maximization cycles for different fetal fractions (FF), increasing in increments of 0.1% per cycle for the range 2.0% to 25.9%, are performed to yield the maternal-fetal genotypes and their posterior probabilities as well as the estimated FF for the sample

2.6 | Calculation of PI and posterior probability of paternity

Given the genotypes of mother, alleged father, and fetus as well as the allele frequencies from the alleged father's population, a value for the paternity index (PI) at a particular SNP was calculated based on the method described in Buckleton et al[34] (Data S1) using the equations designed and formulated for postnatal testing (Table S2). Only SNPs with sequencing depth > 100× in both analyses of maternal cfDNA and alleged paternal genomic DNA and with maternal-fetal genotype probability ≥ 99.0% were classified as effective-SNPs and used in paternity calculations. In cases of mismatch between detected genotype and expected genotype (ie, any nonmutated genotype projected from the parental genotypes), whether in the form of genetic inconsistency where both mother and alleged father were homozygous with the same allele but the fetus was heterozygote or as opposing homozygosity where both fetus and alleged father were homozygous but of different alleles, the parameters of mutation rate and silent allele probability[27] were included in the calculations (Table S2).

As the SNPs were considered independent of each other, the combined paternity index (CPI) was expressed as the product of PIs

for all effective-SNPs. The posterior probability of paternity was subsequently given by CPI/(CPI + 1), and a posterior probability of paternity > 99.99% was taken to indicate the alleged father to be the biological father.

2.7 | STR-based paternity testing

To confirm paternity results in alleged family cases, conventional STR-based paternity testing using fetal genomic DNA extracted from amniotic fluid or buccal swabs was performed. The AmpFISTR Identifiler PCR amplification kit (Applied Biosystems) was employed, following the manufacturer's protocol and according to strict AABB standards. For alleged family cases with male fetus where amniotic fluid was not available, cross-validation of the paternity results was performed through additional testing with maternal cfDNA using the AmpFISTR Yfiler PCR amplification kit (Applied Biosystems), following the manufacturer's protocol modified for use with cfDNA. Parental genomic DNA was analyzed in parallel in each case. Capillary electrophoresis (50-cm capillary array, POP-7) of PCR amplicons from both kits were conducted in an ABI 3500 Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions and strict AABB standards, and data were analyzed using the GeneMapper ID software v.5 (Applied Biosystems). The associated paternity probabilities were calculated according to equations similar to those described above but based on frequencies of the matched STRs in the population of the alleged father.

2.8 | Specificity studies

The specificity of the paternity test (ie, its ability to identify non-paternity) was examined by testing one close male relative of the alleged father in place of the alleged father in three family cases. In addition, 60 unrelated men were tested in place of the alleged father in each of the 15 family cases.

3 | RESULTS

3.1 | Sequencing data

In the sequencing data for the 15 alleged family cases, the average number of read pairs acquired for maternal cfDNA samples was over four-fold that for paternal genomic DNA samples (1.1×10^6 vs 2.5×10^5, respectively; Table S3). The duplicated reads in cfDNA samples accounted for a much higher percentage of mapped reads (average 82.0% and 25.4% in cfDNA and genomic DNA, respectively) such that after filtration of duplicated reads, average sequencing depths at target-SNPs were 243× in cfDNA and 407× in genomic DNA. This resulted in at least 261 target-SNPs (>73.0%) in each sample covered at a minimum of 100× and qualifying for downstream analysis.

3.2 | Accuracy of SNP genotyping by targeted sequencing

The fetal SNP genotypes obtained from targeted sequencing of two maternal cfDNA samples were compared with those obtained via analysis of the paired fetal genomic DNA extracted from amniotic fluid (Table 1). There was full concordance between maternal cfDNA-derived and fetal genomic DNA-derived genotypes when only considering the 257 SNP loci with genotype probabilities > 99.99%. Full concordance was also observed in the 99.0% to 99.99% range, with no incorrectly detected alleles or missing true alleles. However,

TABLE 1 Accuracy of targeted sequencing in SNP genotyping

Case	Probability Range, %	No. of SNPs	Correct Genotypes[a]	Incorrect Alleles[b]	Missed Alleles[c]	Concordance, %
1	>99.99	135	135	0	0	100
	99.0-99.99	19	19	0	0	100
	90.0-98.99	20	19	1	0	95.0
	80.0-89.99	26	16	6	4	61.5
	<80.0	107	65	19	23	60.7
2	>99.99	122	122	0	0	100
	99.0-99.99	47	47	0	0	100
	90.0-98.99	11	8	0	3	72.7
	80.0-89.99	20	13	1	6	65.0
	<80.0	123	60	34	29	48.8

Note: Fetal genotypes determined by targeted sequencing of cfDNA extracted from maternal plasma were verified using those obtained via targeted sequencing of fetal genomic DNA extracted from amniotic fluid.

[a]Number of SNP genotypes consistent between the two sources of fetal DNA.

[b]Number of alleles detected in cfDNA but not in fetal genomic DNA.

[c]Number of alleles detected in fetal genomic DNA but not in cfDNA.

increased instances of discordance appeared in the lower probability ranges, suggesting that genotype calls with probability < 99.0% had ambiguous accuracy (Table 1). The results strongly supported that after processing with the present analysis pipeline, the fetal genotypes with high confidence calls (probability ≥ 99.0%) essentially reflected true genotypes (323/323 correct calls; 100% accuracy), demonstrating the utility of this approach to accurately determine fetal genotypes in maternal cfDNA.

3.3 | Targeted sequencing in paternity testing

Paternity testing using the targeted sequencing method (Figure 2) was applied to 15 alleged family cases. In each case, the full panel of target-SNPs was sequenced, the genotypes were determined, effective-SNPs were identified, and the paternity probability was calculated. The numbers of target-SNPs classified as effective-SNPs ranged from 108 to 174 (average 148; Table 2), corresponding to an effective-SNPs percentage ranging from 30.3% to 48.9% (average 41.6%). All test cases yielded paternity probabilities > 99.9999%, and "Inclusion" results were called (ie, the alleged father in each case was determined to be the biological father). In each case, mismatches between detected and expected genotypes were extremely low (≤2 loci; Table 2), and the fetal fraction was determined to be greater than 4.5%, above the threshold of 4.0% required to support the validity of noninvasive prenatal test results.[28] Details of Case 1 are given in Table S4 to illustrate the approach. All paternity results were subsequently either confirmed using STR-based conventional paternity tests on fetal/child genomic DNA for cases with paired amniotic fluid/buccal samples or cross-validated using Y-STR-based tests on maternal cfDNA (Table 2). All 15 STR loci (Figure S1) or at least 10 Y-STR loci were detected in each test, sufficient to provide validating results (Table S5).

3.4 | Paternity test specificity

The ability to differentiate between the biological father and closely related males of the same paternal lineage was demonstrated when a brother of the biological father was tested as the alleged father in three of the paternity-confirmed cases, and "Exclusion" calls were given for all three siblings (Table 3). Further validation of the present approach was performed by testing each of 60 unrelated men as the alleged father in place of the biological father in the 15 family cases. All tests gave calls of "Exclusion," and large numbers of mismatches were found in each test (Table 4). In addition, significant separation in CPI between that for the biological father and those for unrelated men were observed (Figure 3), indicating the high specificity of this approach in paternity testing.

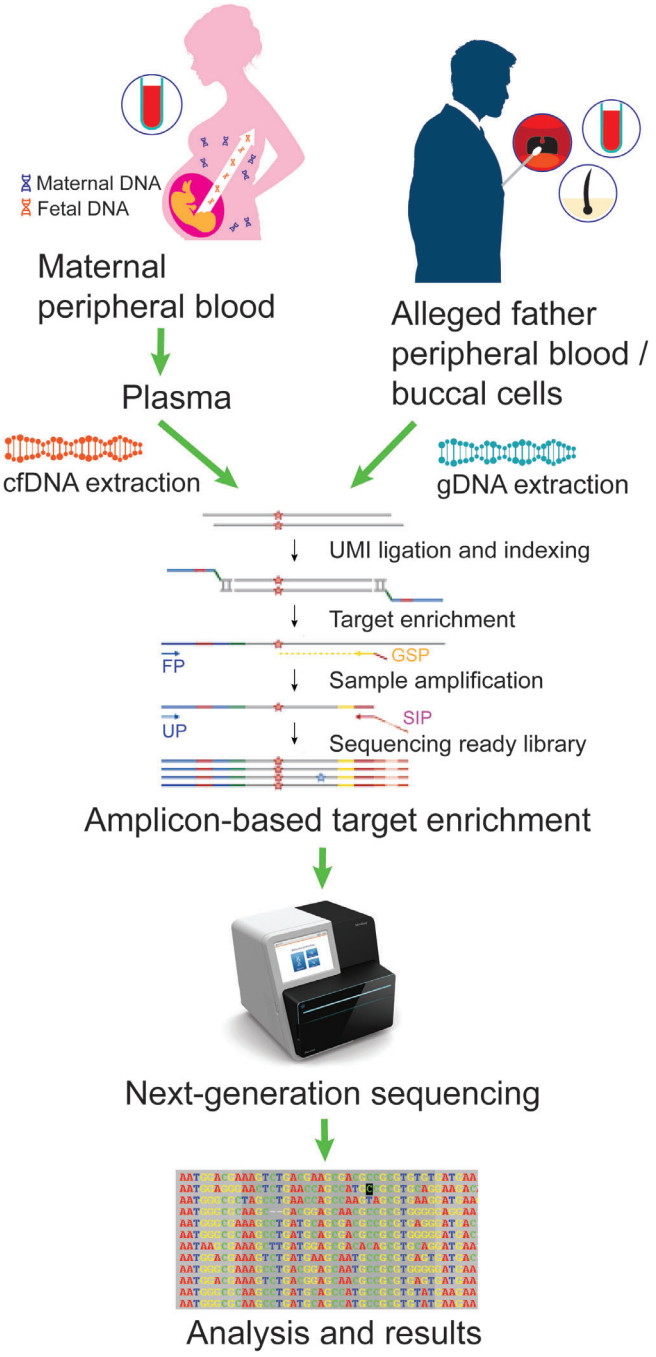

FIGURE 2 Schematic representation of the noninvasive prenatal paternity test. Both genomic DNA (gDNA) extracted from tissue samples of the alleged father and cell-free DNA (cfDNA) extracted from maternal plasma were subjected to target enrichment based on QIAseq Targeted DNA Panels with incorporation of Unique Molecular Identifiers (UMIs). The target-enriched libraries were sequenced, and target-SNPs were filtered for high confidence maternal-fetal genotype calls. The mother, alleged father, and fetus genotypes of these target-SNPs were then analyzed to generate paternity probabilities

4 | DISCUSSION

In conventional postnatal paternity tests, child DNA is sampled, and autosomal STRs are used as genetic markers. Such tests require the genotyping of only 15 to 20 STR markers to generate highly accurate results,[29] with the industry standard for paternity probability set at 99.99% to establish unambiguous paternity. Current noninvasive

TABLE 2 Paternity testing using targeted sequencing

				Targeted sequencing						Validating test		
Case	Fetus Gender	Gestational Age, wk	Fetal Fraction, %	Effective SNPs[a]	Depth, ×[b]	Mismatch Number[c]	CPI, log	Paternity Probability, %	Decision	Test[d]	Paternity Probability, %	Decision
1	M	13	10.7	139	257.2	0	12.2	>99.9999999	Inclusion	STR	99.9999977	Inclusion
2	F	16	7.1	160	293.4	1	12.8	>99.9999999	Inclusion	STR	99.9999998	Inclusion
3	M	17	7.9	170	280.6	1	18.9	>99.9999999	Inclusion	STR	99.9999998	Inclusion
4	F	17	7.0	174	418.1	1	8.6	99.9999997	Inclusion	STR	99.9999988	Inclusion
5	F	18	15.5	159	257.8	2	10.7	>99.9999999	Inclusion	STR	99.9999979	Inclusion
6	M	20	9.6	169	323.5	0	14.6	>99.9999999	Inclusion	Y-STR	99.8834	Inclusion
7	M	11	16.5	158	279.4	0	16.6	>99.9999999	Inclusion	Y-STR	99.8825	Inclusion
8	M	8	18.6	131	214.4	0	15.2	>99.9999999	Inclusion	Y-STR	99.8825	Inclusion
9	M	9	10.1	118	194.1	0	9.5	>99.9999999	Inclusion	Y-STR	99.8837	Inclusion
10	M	8	7.2	139	247.1	0	15.3	>99.9999999	Inclusion	Y-STR	99.8832	Inclusion
11	M	8	10.6	150	257.8	0	14.1	>99.9999999	Inclusion	Y-STR	99.8835	Inclusion
12	M	7	10.1	142	246.2	0	18.0	>99.9999999	Inclusion	Y-STR	99.8829	Inclusion
13	M	8	5.6	160	366.2	0	15.4	>99.9999999	Inclusion	Y-STR	99.8830	Inclusion
14	M	8	4.6	136	275.4	2	8.6	99.9999998	Inclusion	Y-STR	99.8839	Inclusion
15	M	13	5.7	108	214.5	0	11.4	>99.9999999	Inclusion	Y-STR	99.8834	Inclusion

Note: Each case included the alleged father, mother, and fetus trio, and the "Inclusion" test result determined the alleged father to be the biological father.

[a]SNPs with sequencing depth > 100× in both analyses of maternal cfDNA and alleged paternal genomic DNA, and with maternal-fetal genotype probabilities > 99.0% were classified as effective-SNPs and included in paternity calculations.

[b]Average sequencing depth of the effective-SNPs in maternal cfDNA.

[c]Number of detected fetal SNP genotypes not matching the expected genotypes derived from the genotypes of the mother and alleged father, with either opposing homozygosity or genetic inconsistency.

[d]Validating tests were STR-based (STR) if amniotic fluid or buccal cells were sampled whereupon fetal genomic DNA was used; otherwise, the tests were Y-chromosome STR-based (Y-STR), and maternal cfDNA was used. Details of validating test results are given in Table S5.

TABLE 3 Paternity tests with close male relatives

Case	Number of Effective-SNPs[a]	Sequencing Depth, ×[b]	Opposing Homozygosity[c]	Genetic Inconsistency[c]	CPI, log	Decision
3	169	280.6	8	11	−63.7	Exclusion
4	175	418.0	4	11	−57.2	Exclusion
5	132	261.0	5	11	−61.3	Exclusion

Note: One close male relative (brother) of the biological father was tested as alleged father in each of three paternity-confirmed cases. The "Exclusion" test result determined the alleged father to be excluded as the biological father. The case numbers are as listed in Table 2.

[a]SNPs with sequencing depth > 100× in both analyses of maternal cfDNA and alleged paternal genomic DNA, and with maternal-fetal genotype probabilities > 99.0% were classified as effective-SNPs and included in paternity calculations.

[b]Average sequencing depth of the effective-SNPs in maternal cfDNA.

[c]Number of detected fetal SNP genotypes not matching the expected genotypes derived from the genotypes of the mother and alleged father, with either opposing homozygosity or genetic inconsistency.

TABLE 4 Negative paternity tests with unrelated men

	Effective-SNPs		Number of Mismatches[a]		CPI, log	
Case	Median	Range	Average	Range	Average	Range
1	138	136-139	31.6	21-42	−121.0	−73.8 to −172.0
2	161	157-161	36.6	27-51	−137.0	−88.1 to −191.3
3	169	166-170	40.5	28-57	−165.8	−112.8 to −240.2
4	175	171-176	41.4	27-56	−171.1	−116.2 to −240.9
5	160	156-160	38.2	25-49	−138.7	−78.0 to −188.9
6	168	166-169	38.8	29-51	−137.6	−97.8 to −186.9
7	159	155-159	36.8	26-50	−126.6	−81.2 to −168.9
8	130	127-131	32.0	20-42	−120.1	−67.9 to −156.5
9	117	114-118	28.0	18-42	−107.0	−62.9 to −181.2
10	138	137-139	33.7	23-46	−130.4	−96.6 to −177.8
11	149	144-150	35.1	23-47	−129.8	−82.2 to −178.4
12	142	138-143	34.4	24-44	−126.8	−86.6 to −168.3
13	161	158-162	38.2	26-51	−149.2	−103.0 to −204.0
14	140	137-140	32.7	21-44	−116.9	−73.5 to −162.2
15	108	105-108	23.3	13-31	−82.7	−41.4 to −121.7

Note: Sixty unrelated men were tested as alleged father in each of the 15 paternity-confirmed cases listed in Table 2. The values displayed are the average (or median) and range obtained for the set of unrelated men in each case.

[a]Number of detected fetal genotypes not matching the expected genotypes derived from the genotypes of the mother and unrelated men tested as alleged fathers.

prenatal paternity methods based on maternal cfDNA use only Y-STRs[14] or alternatively, a large number, typically thousands, of SNPs as genetic markers.[20,21] SNP-based approaches are superior because of higher compatibility with the fragmented nature of cfDNA via the use of shorter amplicon lengths, reduced false positives and false negatives, and no fetus gender limitations. The large number of SNPs reflects both the lower paternity-differentiating power of SNPs compared with STRs, with the estimated power of 50 SNPs, having minor allele frequencies between 0.2 and 0.8, being similar to 12 STRs in postnatal tests[30] as well as the noisy data, resulting from low target allele concentrations, against which analysis of more SNPs are required to allow for filtering of low-quality data.

In the present study, we hypothesized that with systematic selection of SNP loci and accurate genotyping achieved through UMI-based targeted sequencing, the number of tested SNPs could be reduced from thousands[20,21] to hundreds. To this end, SNP selection was performed based on population genetics data and comprised selection criteria that would simplify calculations, reduce SNP redundancy, and increase discriminative power (Data S1). The criteria to increase discriminative power by selecting SNPs with high heterozygosity were consistent with previous studies,[20] and the selection process subjectively resulted in a panel of 356 target-SNPs (Table S1). Following sequencing data analysis, 108 to 174 target-SNPs were classified as effective-SNPs and included in paternity calculations (Table 2). These numbers were comparable with that reported in a previous study where an initial panel of over 1400 SNPs had been sequenced, but only 130 to 162 SNPs were used in paternity calculations,[22] suggesting that the present selection process

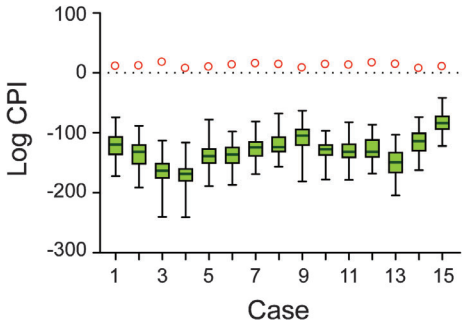

FIGURE 3 Combined paternity indices for biological father and unrelated men. The logarithm of CPIs (log CPIs) calculated for the 15 family cases were all greater than 8.0 when the biological father was tested (red circles), and each was a distinct outlier compared with the respective set of log CPIs obtained for 60 unrelated men each tested as alleged father (box-and-whisker plots). The dotted line marks where log CPI = 0, and the case numbers correspond to those given in Table 2

effectively reduced the number of redundant SNPs and increased the percentage of effective-SNPs. Although, in general, the more SNPs sequenced, the more the discriminating power of the test, in practice, the actual number of SNPs sequenced would be limited by costs and high cfDNA input, and the discriminating power would be dependent on the number of SNPs that are eventually included in paternity calculations.

Single nucleotide polymorphism-based noninvasive prenatal paternity testing is highly dependent on the accurate genotyping of fetal SNPs in maternal cfDNA. Moreover, since the higher the FF in maternal cfDNA, the more accurate the genotyping result, with a threshold set at 2% to 4%,[28,31] below which the validity of noninvasive prenatal tests would not be supported, it is important that FFs are also accurately estimated in order to support the test results for samples with high FF and to identify and exclude samples with low FF. The low fetal allele counts in the presence of high maternal allele counts translate to difficulty in reliable detection of fetal alleles and constitute a major challenge in maternal cfDNA analysis. In sequencing approaches, this issue was previously addressed by employing targeted deep sequencing wherein target enrichment was hybridization based.[20,21] To our knowledge, this is the first report where target enrichment was performed by multiplex amplification with the incorporation of UMIs. The use of UMI barcoding and the UMI-based smCounter2 algorithm for sequencing data processing enabled correction of sequencing and PCR errors that would otherwise have affected allele count determination. Moreover, a novel Bayesian-based algorithm developed in-house was used to generate the final maternal-fetal genotype calls and their associated posterior probabilities, and this allowed the removal of ambiguous calls with probabilities < 99.0% to minimize genotyping errors. This filtering process was supported by the correctness of all fetal genotype calls from maternal cfDNA above the threshold probability as verified through comparison with those derived from fetal genomic DNA (Table 1). Therefore, the analysis pipeline as a whole enabled accurate SNP genotyping and ensured that essentially only true genotypes were used in subsequent paternity calculations.

The utility of the method in noninvasive prenatal paternity testing was ultimately demonstrated through the successful determination of paternity in 15 family cases, results of which were all subsequently validated (Table 2; Table S5). The minimum logarithm of CPI and paternity probability for these cases were 8.6% and 99.9999997%, respectively, well above the lower limits for paternity inclusion and attesting to the strength of the method. Moreover, close male relatives were readily excluded as the biological father in three cases (Table 3), validating the potential to eliminate close relative-derived false paternity-inclusion cases.[27] The exclusion of 60 unrelated men when tested as alleged father in each of the 15 cases further verified the specificity of the method (Figure 3; Table 4). Notably, the paternity probabilities generated by the method were comparable with those obtained by STR analysis but much higher than those from Y-STR analysis (Table 2), revealing the increased power of the novel method compared with Y-STR analysis. Besides, although Y-STR analysis is extremely useful for paternal lineage identification,[32] it is less effective for paternity testing because of difficulties in differentiating between male relatives. The Y-STR analysis-generated probabilities are therefore more correctly patrilineality or male line probabilities, and the Y-STR paternity-inclusion results confirmed only paternal lineage relationship.

Conventional paternity tests analyze autosomal STRs in genomic DNA by performing PCR amplification and capillary electrophoresis.[29] Compared with the present approach, STR analysis does not involve complex bioinformatics, but since genomic DNA is analyzed, conventional methods cannot be applied to noninvasive prenatal testing. Moreover, although close male relative-derived false paternity-inclusion results are less problematic for autosomal STR-based methods compared with Y-STR analysis, elimination of such false positives in the present approach is a significant advantage over conventional tests.[27]

Compared with previous SNP-based targeted sequencing methods for noninvasive prenatal paternity testing, the present approach differed in terms of the analyzed SNPs, the amplicon-based nature of target enrichment, and the data analysis algorithm. Importantly, the SNPs selection process significantly reduced the number of sequenced target-SNPs yet supported comparable discriminating power, and the amplicon-based approach allowed the use of UMI barcoding to facilitate absolute quantification of SNP alleles and more reliable genotyping, both features representing significant improvements to previous methods. Moreover, where only maternally homozygous SNPs were included in paternity calculations in previous studies,[20-22] the present analysis algorithm included all SNPs with high-confidence fetal genotype calls, disregarding the maternal genotype, thus providing an added means to increase the effective-SNPs percentage. Furthermore, the sequencing of significantly reduced numbers of target-SNPs associates with reduced costs and increased cost-effectiveness.

In conclusion, despite limitations that include the testing of only a few family cases and only the *Han* Chinese population, the current

pilot study has demonstrated the utility of the approach in noninvasive prenatal paternity testing. The method has further potential for application to a range of related clinical functions such as relationship testing or screening for single gene disorders, and such feasibility could be evaluated in future studies.

ACKNOWLEDGEMENTS

The authors thank Tsan Yuk Hospital for providing the amniotic fluid samples and the Department of Molecular Laboratory at Medtimes Medical Group Limited (AABB accredited relationship testing laboratory) for performing the STR-based paternity validation assays.

CONFLICT OF INTEREST

Authors are employees of Medtimes Medical Group Limited.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Jacqueline Chor Wing Tam https://orcid.org/0000-0003-3004-9043

REFERENCES

1. Brambati B, Simoni G, Travi M, et al. Genetic diagnosis by chorionic villus sampling before 8 gestational weeks: efficiency, reliability, and risks on 317 completed pregnancies. *Prenat Diagn*. 1992;12(10):789-799.
2. Elchalal U, Shachar IB, Peleg D, Schenker JG. Maternal mortality following diagnostic 2nd-trimester amniocentesis. *Fetal Diagn Ther*. 2004;19(2):195-198.
3. Seeds JW. Diagnostic mid trimester amniocentesis: how safe? *Am J Obstet Gynecol*. 2004;191(2):607-615.
4. American College of Obstetricians and Gynecologists. ACOG practice Bulletion no. 88: invasive prenatal testing for aneuploidy. *Obstet Gynecol*. 2007;110:1459-1467.
5. Lo YM, Corbetta N, Chamberlain PF, et al. Presence of fetal DNA in maternal plasma and serum. *Lancet*. 1997;350(9076):485-487.
6. Brojer E, Zupanska B, Guz K, et al. Noninvasive determination of fetal RHD status by examination of cell-free DNA in maternal plasma. *Transfusion*. 2005;45(9):1473-1480.
7. Chiu EKL, Hui WWI, Chiu RWK. cfDNA screening and diagnosis of monogenic disorders—where are we heading? *Prenat Diagn*. 2018;38(1):52-58.
8. Chiu RW, Akolekar R, Zheng YW, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ*. 2011;342:c7401.
9. Hill M, Finning K, Martin P, et al. Non-invasive prenatal determination of fetal sex: translating research into clinical practice. *Clin Genet*. 2011;80(1):68-75.
10. Rong Y, Gao J, Jiang X, Zheng F. Multiplex PCR for 17 Y-chromosome specific short tandem repeats (STR) to enhance the reliability of fetal sex determination in maternal plasma. *Int J Mol Sci*. 2012;13(5):5972-5981.
11. Wagner J, Dzijan S, Marjanovic D, Lauc G. Non-invasive prenatal paternity testing from maternal blood. *Int J Leg Med*. 2009;123(1):75-79.
12. Zimmermann B, Hill M, Gemelos G, et al. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. *Prenat Diagn*. 2012;32(13):1233-1241.
13. Deng Z, Wu G, Li Q, et al. Noninvasive genotyping of 9 Y-chromosome specific STR loci using circulatory fetal DNA in maternal plasma by multiplex PCR. *Prenat Diagn*. 2006;26(4):362-368.
14. Barra GB, Santa Rita TH, Chianca CF, et al. Fetal male lineage determination by analysis of Y-chromosome STR haplotype in maternal plasma. *Forensic Sci Int Genet*. 2015;15:105-110.
15. Zhang S, Han S, Zhang M, Wang Y. Non-invasive prenatal paternity testing using cell-free fetal DNA from maternal plasma: DNA isolation and genetic marker studies. *Leg Med (Tokyo)*. 2018;32:98-103.
16. Ashoor G, Syngelaki A, Poon LC, et al. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics. *Ultrasound Obstet Gynecol*. 2013;41(1):26-32.
17. Li S, Liu H, Jia Y, et al. A novel SNPs detection method based on gold magnetic nanoparticles array and single base extension. *Theranostics*. 2012;2(10):967-975.
18. Li S, Liu H, Jia Y, et al. An automatic high-throughput single nucleotide polymorphism genotyping approach based on universal tagged arrays and magnetic nanoparticles. *J Biomed Nanotechnol*. 2013;9(4):689-698.
19. Ryan A, Baner J, Demko Z, et al. Informatics-based, highly accurate, noninvasive prenatal paternity testing. *Genet Med*. 2013;15(6):473-477.
20. Jiang H, Xie Y, Li X, et al. Noninvasive prenatal paternity Testing (NIPAT) through maternal plasma DNA sequencing: a pilot study. *PLoS One*. 2016;11(9):e0159385.
21. Qu N, Xie Y, Li H, et al. Noninvasive prenatal paternity testing using targeted massively parallel sequencing. *Transfusion*. 2018;58(7):1792-1799.
22. Yang D, Liang H, Gao Y, et al. Noninvasive fetal genotyping of paternally inherited alleles using targeted massively parallel sequencing in parentage testing cases. *Transfusion*. 2017;57(6):1505-1514.
23. Kou R, Lam H, Duan H, et al. Benefits and challenges with applying unique molecular identifiers in next generation sequencing to detect low frequency mutations. *PLoS One*. 2016;11(1):e0146638.
24. Xu C, Gu X, Padmanabhan R, et al. smCounter2: an accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers. *Bioinformatics*. 2019;35(8):1299-1309. https://doi.org/10.1093/bioinformatics/bty790
25. Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM arXiv*. 2013;1303:3997.
26. Jiang P, Chan KC, Liao GJ, et al. FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. *Bioinformatics*. 2012;28(22):2883-2890.
27. Borsting C, Morling N. Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Sci Int Genet*. 2011;5(3):236-241.
28. Xu H, Wang S, Ma LL, et al. Informative priors on fetal fraction increase power of the noninvasive prenatal screen. *Genet Med*. 2018;20(8):817-824.
29. Pena SD, Chakraborty R. Paternity testing in the DNA era. *Trends Genet*. 1994;10(6):204-209.
30. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Leg Med*. 2001;114(4–5):204-210.
31. Artieri CG, Haverty C, Evans EA, et al. Noninvasive prenatal screening at low fetal fraction: comparing whole-genome sequencing and single-nucleotide polymorphism methods. *Prenat Diagn*. 2017;37(5):482-490.
32. Kayser M. Forensic use of Y-chromosome DNA: a general overview. *Hum Genet*. 2017;136(5):621-635.
33. Goya R, Sun MG, Morin RD, et al. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics*. 2010;26(6):730-736.

34. Buckleton J, Clayton T, Triggs C, Testing P. In: Triggs C, Buckleton J, Walsh S, eds. *Forensic DNA Evidence Interpretation*. 1st ed. Boca Raton: CRC Press; 2005.

SUPPORTING INFORMATION

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

How to cite this article: Tam JCW, Chan YM, Tsang SY, et al. Noninvasive prenatal paternity testing by means of SNP-based targeted sequencing. *Prenatal Diagnosis*. 2020;40:497–506. https://doi.org/10.1002/pd.5595