Application Note

Noninvasive prenatal paternity testing by means of SNP-based targeted sequencing barcoded with Unique Molecular Indices

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

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

Introduction

Paternity testing is conducted to determine the biological linkage between a child and an alleged father, and can be done either before or after the birth of the child. There are numerous reasons why prenatal testing is sometimes preferable to postnatal testing with both invasive and non-invasive sampling methods employed to recover genetic material for subsequent DNA analysis. Invasive methods such as chorionic villus sampling or amniocentesis do however have an associated risk of miscarriage and infection and are therefore not recommended unless to aid in diagnosis of severe genetic disorders.

Non-invasive methods involve sampling from maternal blood where circulating cell free DNA (cfDNA) from the fetus is present (1). Whilst short tandem repeats (STRs) are used routinely for paternity testing, their use with cfDNA is hindered by the fragmented nature of cfDNA which makes amplification of the relatively large STR amplicons problematic. Therefore, single nucleotide polymorphisms (SNPs), which require much smaller DNA amplicons, have been used increasingly for paternity testing from cfDNA.

There is a relatively low proportion of fetal DNA to maternal DNA in cfDNA samples. This, and associated challenges with analysis and interpretation of data, mean that significantly larger panels have been required to generate sufficient and reliable sequencing data. These larger panels are essential to enable the profiling of fetal genotypes for the inclusion/exclusion calculations when comparing the SNP profile from the putative father. Such large SNP arrays involve additional technical challenges and in turn increase the overall time and cost of the analysis.

Here we describe, for the first time, the use of Unique Molecular Indices (UMIs) as part of a QIAseq® targeted sequencing panel for use with prenatal paternity testing. The use of UMIs allows higher levels of PCR and sequencing error corrections and therefore assures a significantly higher confidence in the final sequencing data for accurate paternity calling and discrimination. This in turn enables use of much smaller numbers of SNPs to generate the required probability of paternity, making the sequencing workflow simpler and faster.

Methods

Collection of samples

Peripheral blood samples were obtained from 15 pregnant mothers. 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). Maternal peripheral blood samples (approximately 10 ml) were collected in cell-free DNA collection tubes (Roche). Peripheral blood samples from adult males were collected in Vacuette® blood collection tubes (Greiner Bio-One). Buccal samples were collected using flocked swabs (Copan Diagnostics).

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). Maternal cfDNA was extracted from maternal plasma using the Maxwell RSC LV ccfDNA Custom Kit (Promega). Concentrations of the extracted genomic DNA and cfDNA were measured using the NanoDrop Lite spectrometer (Thermo Fisher Scientific) 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.

Selection of SNPs

An initial panel of 356 SNPs with minor allele frequencies greater than 0.30 and covering all 22 autosomes was selected as target-SNPs for sequencing. This panel was selected based on population genetics data from the 1000 Genomes Project (**www.1000genomes.org**).

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 10–20 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 8-cycle multiplex PCR with custom-designed QIAseq Targeted DNA Panel primers (QIAGEN) using a Thermocycler C1000 system (BioRad). 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).

Sequencing data processing

The smCounter2 pipeline, as previously described (2), was used for calling of low-frequency variants (in this instance the fetal SNP genotypes) from the QIAseq-based targeted sequencing data. The UMIs incorporated as part of the QIAseq chemistry enabled error correction for most of the sequencing and PCR errors. 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 (3).

Calculation of PI and posterior probability of paternity

Using 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. (4). Only SNPs with sequencing depth >100X 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.

Results and Discussion

Paternity testing using targeted sequencing was applied to 15 alleged family cases. In each case, the full panel of target-SNPs was sequenced, the genotypes were determined, effective-SNPs (i.e., with a coverage of over 100X) 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 1), 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 (i.e., 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; see Table 1). The fetal fraction was determined to be greater than 4.5% (of the total sequenced DNA), above the threshold of 4.0% required to support the validity of non-invasive prenatal test results. Subsequently, all paternity results were 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 1. Paternity testing using targeted sequencing

				Targeted sequencing						Validating test		
Case	Fetus gender	Gestational age (week)	Fetal fraction (%)	Effective SNPs*	Depth X†	Mismatch number‡	CPI (log)	Paternity probability (%)	Decision	Test§	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.

* SNPs with sequencing depth >100X 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.

† Average sequencing depth of the effective-SNPs in maternal cfDNA.

‡ 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.

§ 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.

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 to hundreds. 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 2). 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 1). 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 accurately assess close relative-derived false paternity-inclusion cases.

The exclusion of 60 unrelated men when tested as alleged father in each of the 15 cases further verified the specificity of the method (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 1), revealing the increased power of the novel method compared with Y-STR analysis.

Table 2. Accuracy of targeted sequencing in SNP genotyping

Case	Probability range (%)	No. of SNPs	Correct genotypes*	Incorrect alleles†	Missed alleles‡	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
	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
	< 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. The case numbers are as listed in Table 1.

* Number of SNP genotypes consistent between the two sources of fetal DNA.

† Number of alleles detected in cfDNA but not in fetal genomic DNA.

‡ Number of alleles detected in fetal genomic DNA but not in cfDNA.

Table 3. Paternity tests with close male relatives

Case	No. of effective SNPs*	Sequencing depth X†	Opposing homozygosity‡	Genetic inconsistency‡	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 1.

* 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.

† Average sequencing depth of the effective-SNPs in maternal cfDNA.

‡ 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*		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	137–149	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 1. The values displayed are the average (or median) and range obtained for the set of unrelated men in each case.

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

The data presented here supports our hypothesis that relatively small numbers of SNPs (<200, versus previous panels with over 1000) can be used to generate strong paternity probabilities enabling confident inclusion or exclusion of putative fathers. In order to successfully sequence a high percentage of SNPs in a panel and avoid the need to have large redundancy in a greatly expanded SNP panel, we have described a novel approach, using Unique Molecular Indices (UMIs). This approach, when combined with the smCounter2 pipeline for analysis, ▷

enables the effective removal of sequencing and PCR errors from data analysis and absolute quantitation of SNP alleles by grouping sequencing reads with the corresponding UMI sets, This provides higher confidence in the prediction of maternal-fetal SNP genotypes and, as a result, the paternity could be determined with smaller panel of SNPs and the sequencing cost could be reduced.

By enabling the use of much smaller numbers of SNPs, this workflow allows easier and more cost-effective utilization of non-invasive prenatal paternity testing, relative to larger SNP panels. This makes cfDNA prenatal testing a viable alternative to other, higher risk, prenatal testing options such as amniocentesis, in more cases and potentially to more laboratories and, in turn, to more families.

References

1. Lo, Y.M. et al. (1997) Presence of fetal DNA in maternal plasma and serum. The Lancet **350**, 485-487.
2. Xu, C. et al. (2019) smCounter2: an accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers. Bioinformatics **35**, 1299-1309.
3. Goya, R. et al. (2010) SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics **26**, 730-736.
4. Buckleton, J., Clayton, T. and Triggs, C. (2005) Parentage Testing. In: Triggs, C., Buckleton, J. and Walsh, S., eds. *Forensic DNA Evidence Interpretation*. 1st ed. Boca Raton: CRC Press, p 534.

Ordering Information

Product	Contents	Cat. no.
QIAseq Targeted DNA Custom Panel (96)	Kit containing ALL reagents (except indexes) for targeted DNA sequencing; custom panel for 96 samples	333525
QIAseq 96-Index I Set A (384)	Box containing oligos, enough for a total of 384 samples, for indexing up to 96 samples for targeted panel sequencing on Illumina platforms; one of four sets required for multiplexing 384 samples	333727
QIAseq Library Quant Assay Kit	1 tube of Primer Mix (500 µl), 1 bottle of Dilution Buffer (30 ml), 1 tube DNA Standard (100 µl), and 5 tubes GeneRead qPCR SYBR Green (1.35 ml) for sample library quantification prior to NGS	333314

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at www.qiagen.com or can be requested from QIAGEN Technical Services or your local distributor.