Clinical Chemistry 57:1
92–101 (2011)

Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles

Gary J.W. Liao,[1,2] Fiona M.F. Lun,[1,2] Yama W.L. Zheng,[1,2] K.C. Allen Chan,[1,2] Tak Y. Leung,[3] Tze K. Lau,[3] Rossa W.K. Chiu,[1,2] and Y.M. Dennis Lo[1,2*]

BACKGROUND: Massively parallel sequencing has recently been used in noninvasive prenatal diagnosis. The current costs of this technology are still relatively expensive, however, and sample throughput is still relatively low when it is used as a molecular diagnostic tool. Rather than nonselectively sequencing the genome, target enrichment provides a logical approach for more efficient and cost-effective massively parallel sequencing because it increases the proportion of informative data from the targeted region(s). Existing applications of targeted sequencing have mainly been qualitative analyses of genomic DNA. In this study, we investigated its applicability in enriching selected genomic regions from plasma DNA and the quantitative performance of this approach.

METHODS: DNA was extracted from plasma samples collected from 12 pregnant women carrying female fetuses. The SureSelect Target Enrichment System (Agilent Technologies) was used to enrich for exons on chromosome X. Plasma DNA libraries with and without target enrichment were analyzed by massively parallel sequencing. Genomic DNA samples of the mother and fetus for each case were genotyped by microarray.

RESULTS: For the regions targeted by the enrichment kit, the mean sequence coverage of the enriched samples was 213-fold higher than that of the nonenriched samples. Maternal and fetal DNA molecules were enriched evenly. After target enrichment, the coverage of fetus-specific alleles within the targeted region increased from 3.5% to 95.9%.

CONCLUSIONS: Targeted sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles at genomic regions of interest and is a powerful method for measuring the proportion of fetal DNA in a maternal plasma sample.

© 2010 American Association for Clinical Chemistry

Currently, fetal genetic materials for definitive prenatal diagnoses of fetal genetic and chromosomal diseases are typically collected through invasive procedures. Unfortunately, the procedures are associated with a risk of fetal loss. The identification in 1997 of cell-free fetal DNA in maternal plasma *(1)* has facilitated the development of noninvasive prenatal diagnosis (NIPD)[4] *(2)*, which is free of procedure-related risks; however, the coexistence in maternal plasma of a minor population of fetal DNA within a major background of maternal DNA *(3)* has posed challenges for extending NIPD applications beyond those focusing on the detection of paternally inherited fetal alleles *(1, 4)*. One such challenging area is the use of NIPD for fetal chromosomal aneuploidies, which entails the detection of aberrant quantities of fetal DNA derived from an aneuploid chromosome.

The recent availability of massively parallel sequencing enhances the precision of DNA quantification to an unprecedented level *(5, 6)*. Our group *(7, 8)* and others *(9)* have demonstrated its feasibility in measuring small proportional increments of chromosome 21 (chr21) DNA in maternal plasma of pregnancies carrying trisomy 21 (T21) fetuses. One of the major limitations of the use of massively parallel sequencing for routine prenatal diagnosis is the relatively high cost. Besides waiting for a decline in sequencing costs, one might take another logical approach, which would be to increase the proportion of informative data in the output. In T21 detection, for example, we are particularly interested in DNA fragments from chr21, which

[1] Centre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, [2] Department of Chemical Pathology, and [3] Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong.
* Address correspondence to this author at: Department of Chemical Pathology, Prince of Wales Hospital, 30–32 Ngan Shing St., Shatin, New Territories, Hong Kong SAR, China. Fax +852-2636-5090; e-mail loym@cuhk.edu.hk.
Received July 27, 2010; accepted November 2, 2010.
Previously published online at DOI: 10.1373/clinchem.2010.154336
[4] Nonstandard abbreviations: NIPD, noninvasive prenatal diagnosis; chr21, chromosome 21; T21, trisomy 21; chrX, chromosome X; SNP, single-nucleotide polymorphism; GR, genomic representation.

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

Foresight EX1009
Foresight v Personalis

represents only about 1.5% of the entire genome. If we could enrich these target DNA fragments to a higher proportion, we would require much less sequencing effort to generate an equivalent amount of data from chr21.

The principle of targeted sequencing is to selectively amplify or capture targeted regions from a DNA sample before sequencing. Targeted sequencing is initially applied to detect population genetic variation, e.g., for genetic-association studies. Therefore, its current application in genomics research is frequently aimed at solving qualitative problems (e.g., genotyping or mutation detection) *(10, 11)*. The application of targeted sequencing in maternal plasma DNA for NIPD, however, involves a number of previously unaddressed quantitative considerations *(5, 7)*. Another issue of concern is that the current targeted-sequencing protocols are designed primarily for genomic DNA (e.g., derived from tissues or cells). Thus, the application of targeted sequencing to plasma DNA still requires investigation, especially when one considers the substantial differences between genomic DNA and plasma DNA, e.g., the degraded nature and low concentrations of plasma DNA.

Current target-enrichment strategies include molecular inversion probes *(12)*, microdroplet-based PCR *(13)*, on-array capture *(14)*, and in-solution capture *(15)*. Molecular inversion probes, which consist of a common linker flanked by target-specific sequences, anneal to the target sequences and become circularized by ligase *(12)*. Digestion removes the noncircularized background DNA, and the PCR amplifies the circularized target DNA via universal primers to a common linker. The unstable performance of such probes in differentiating heterozygous alleles complicates its application for quantitative analysis *(12)*. Microdroplet-based PCR (e.g., the RainDance platform) uses up to 1.5 million separate PCRs to amplify targeted sequences in parallel, but it requires a relatively high DNA input (7.5 μg per sample) *(13)*. The rationale of on-array capture is to hybridize a DNA library to immobilized probes. Nontargeted DNA fragments are removed by washing, and targeted fragments are harvested by elution *(14)*. A vast excess of DNA in the library (20 μg per sample) over that of the probes is typically required to ensure complete hybridization *(14)*. Considering the low concentrations of plasma DNA *(3)*, microdroplet-based PCR and on-array capture might not be ideal for plasma DNA–based applications. Therefore, in this study we adopted the recently described in-solution capture strategy for targeted sequencing of plasma DNA, especially in view of the relatively small amounts of DNA required for this approach *(15)*.

Materials and Methods

STUDY PARTICIPANTS AND SAMPLE COLLECTION

With informed consent, we recruited 12 pregnant women with singleton female fetuses [first trimester (12–13 weeks): PW257, PW279, PW280, and PW338; second trimester (16–19 weeks): M3667, M4108, M5166, and M6114; third trimester (38 weeks): M6011, M6028, M6029, and M6043] from the Department of Obstetrics and Gynaecology, Prince of Wales Hospital, Hong Kong. The study was approved by the clinical research ethics committee of the institution. Maternal peripheral blood samples were collected into EDTA-containing blood tubes prior to invasive obstetrics procedures (before chorionic villus sampling in the first trimester, before amniocentesis in the second trimester, and before elective cesarean section in the third trimester). Fetal genomic DNA samples were obtained from chorionic villi for the first-trimester cases and from the placental samples after delivery for the second- and third-trimester cases.

SAMPLE PROCESSING AND DNA EXTRACTION

Maternal peripheral blood samples (6–10 mL) were centrifuged at 1600*g* for 10 min at 4 °C *(16)*; the plasma portion (2.4–4.8 mL) was recentrifuged at 16 000*g* for 10 min at 4 °C. We removed any residual plasma from the blood cell portion by recentrifugation at 2500*g* for 5 min. Plasma DNA was extracted with the DSP DNA Blood Mini Kit (Qiagen), as described previously *(7)*. DNA was extracted from chorionic villi and peripheral blood cells with the QIAamp DNA Blood Mini Kit (Qiagen) according to the manufacturer's blood protocol. DNA was extracted from placental tissues with the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's tissue protocol. We quantified the extracted plasma DNA by real-time PCR with an ABI 7300 Sequence Detector (Applied Biosystems). A β-globin real-time PCR assay was performed as described previously *(3)*. A conversion factor of 6.6 pg of DNA per cell was used to calculate the amount of the extracted plasma DNA. Genomic DNA extracted from chorionic villi, peripheral blood cells, and placental tissues was quantified with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies/Thermo Scientific).

MICROARRAY GENOTYPING

Maternal genomic DNA extracted from buffy coat and fetal genomic DNA extracted from chorionic villi and placental tissues were genotyped with the Genome-Wide Human SNP Array 6.0 (Affymetrix).

Foresight EX1009
Foresight v Personalis

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

PLASMA DNA LIBRARY PREPARATION

The plasma DNA molecules were already fragmented in nature, so no additional fragmentation step was required. For DNA library construction, we used the Paired-End DNA Sample Preparation Kit (Illumina) and 5–30 ng plasma DNA for each case. Because the library-preparation section in Agilent Technologies' SureSelect protocol is designed primarily for genomic DNA, we replaced it with Illumina's protocol of chromatin immunoprecipitation sequencing to prepare the plasma DNA library. This protocol is better suited for small amounts of input DNA. The adapter-ligated DNA was purified directly with the spin columns provided in the QIAquick PCR Purification Kit (Qiagen) without further size selection. A 15-cycle PCR and standard Illumina primers (PCR Primer PE 1.0 and PCR Primer PE 2.0) were then used to amplify the adapter-ligated DNA. We quantified the DNA libraries with a NanoDrop ND-1000 spectrophotometer and used the DNA 1000 Kit with a 2100 Bioanalyzer (Agilent) to check the size distribution of the libraries. We generated 0.6–1 μg of an amplified plasma DNA library for each sample, with an approximate mean size of 290 bp.

TARGET ENRICHMENT

The SureSelect Human X Chromosome Kit (Agilent) covered 85% of the exons on human chromosome X (chrX). For all 12 cases in this study, we incubated 500 ng of the amplified plasma DNA library with the capture probes for 24 h at 65 °C, in accordance with the manufacturer's instructions. After hybridization, we selected captured targets by pulling down the biotinylated probe/target hybrids with streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin; Invitrogen) and purified the targets with the MinElute PCR Purification Kit (Qiagen). Finally, we enriched the targeted-DNA libraries by 12-cycle PCR amplification with the SureSelect GA PCR primers (Agilent). The PCR products were purified by the QIAquick PCR Purification Kit.

SEQUENCING AND ALIGNMENT

Samples from the 3 trimesters were sequenced on 3 separate flow cells. In each flow cell, 4 pairs of libraries with and without target enrichment were loaded onto 8 separate lanes and sequenced on a Genome Analyzer IIx (Illumina) with a paired-end format of 36 bp × 2. All 36-bp sequenced reads were aligned to the unmasked human reference genome (Hg18) (http://genome.ucsc.edu) with the aid of SOAPaligner/soap2 (http://soap.genomics.org.cn). Two mismatches were allowed for single-nucleotide polymorphism (SNP) calling. The range of fragment sizes of paired-end reads was defined as 40–600 bp.

We defined a SNP as informative when the mother was homozygous and the fetus was heterozygous. For example, if the maternal genotype of a SNP was AA and the fetal genotype was AT, we defined this SNP as informative. In this case, the A allele was shared by the mother and the fetus, whereas the T allele was fetus specific and paternally inherited. On the basis of these criteria, we identified informative SNPs by using the maternal and fetal genotyping information by the Genome-Wide Human SNP Array 6.0. Detailed annotation (e.g., coordinates and alleles) of each informative SNP was obtained at the same time. According to these informative SNP coordinates, we performed SNP calling for the maternal plasma DNA and counted the number of times each allele (e.g., shared allele and fetus-specific allele) was sequenced by paired-end reads that had passed the quality filters (with default parameters) in Illumina's pipeline. Forward and reverse reads were used independently for SNP calling. An allele was scored as a fetus specific when it was observed one or more times.

Results

EFFICIENCY OF TARGET ENRICHMENT

On average, 14.5 million paired-end reads passed the Illumina pipeline quality filters for each sample. We uniquely mapped 79.9% of paired-end reads in nonenriched samples and 85.8% paired-end reads in target-enriched samples to the reference human genome (Hg18). The higher mapped rate of target-enriched samples might be due to the increased proportion of the targeted exons, which were more likely than other sequences to be mapped uniquely to the reference genome. The latter included repeat sequences. Duplicated paired-end reads (i.e., reads with identical sequences and start–end coordinates) were considered clones of the same original plasma DNA template. All but one of the duplicated reads were filtered, leaving only 1 copy for subsequent bioinformatics analysis. Such duplicated reads accounted for 17.6% of the mapped reads among the target-enriched samples and 0.3% of the nonenriched samples. The higher proportion of duplicated reads might be due to the additional 12-cycle PCR in the target-enrichment protocol (Table 1).

We calculated the genomic representations (GRs) of all 22 autosomes and chrX by dividing the number of nonduplicate paired-end reads mapped to each chromosome by the total number of nonduplicate paired-end reads. The expected GR for each chromosome was calculated for the reference genome as previously described *(7)*. The GRs of all chromosomes in the nonenriched samples were comparable to the expected values. In the target-enriched samples, the GRs for the 22 autosomes were 50% lower than the expected values, and the GR for

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

Foresight EX1009
Foresight v Personalis

Table 1. Sequencing data of samples with and without target enrichment.

FC[a]	Lane	Sample	Target enrichment	Raw pairs, $\times 10^{6}$[b]	Mapped pairs, $\times 10^{6}$[c]	Mapped rate, %[d]	Nonduplicate pairs, $\times 10^{6}$[e]	Nonduplicate rate, %[f]	Targeted region depth[g]	Enrichment, -fold[h]
1	1	PW257	No	17.5	14.0	80.0	14.0	99.6	0.41	NA[i]
	2	PW279	No	17.3	13.7	79.2	13.6	99.3	0.40	NA
	3	PW280	No	17.1	13.4	78.4	13.4	99.5	0.39	NA
	4	PW338	No	17.0	13.8	81.2	13.7	99.3	0.40	NA
	5	PW257	Yes	14.0	12.2	87.1	9.86	80.8	105	255
	6	PW279	Yes	16.7	14.4	86.2	11.3	78.5	114	287
	7	PW280	Yes	13.2	11.3	85.6	9.29	82.2	96.2	245
	8	PW338	Yes	16.2	14.0	86.4	11.3	80.7	113	280
2	1	M3667	No	23.7	19.9	84.0	19.9	99.7	0.56	NA
	2	M4108	No	24.4	20.4	83.6	20.3	99.5	0.57	NA
	3	M5166	No	26.4	22.2	84.1	22.1	99.5	0.61	NA
	4	M6114	No	28.1	22.9	81.5	22.8	99.6	0.61	NA
	5	M3667	Yes	12.6	11.0	87.3	8.89	80.8	99.2	177
	6	M4108	Yes	12.9	11.3	87.6	8.83	78.1	96.1	168
	7	M5166	Yes	12.3	10.7	87.0	8.44	78.9	91.6	149
	8	M6114	Yes	10.7	9.34	87.3	6.51	69.7	60.9	100
3	1	M6011	No	6.77	5.22	77.1	5.22	99.9	0.16	NA
	2	M6028	No	8.00	6.20	77.5	6.19	99.9	0.18	NA
	3	M6029	No	9.46	7.40	78.2	7.39	99.8	0.21	NA
	4	M6043	No	10.6	7.83	74.2	7.82	99.8	0.22	NA
	5	M6011	Yes	5.69	4.66	81.8	4.44	95.3	47.4	296
	6	M6028	Yes	7.53	6.38	84.7	5.71	89.5	52.1	289
	7	M6029	Yes	9.14	7.67	83.9	6.98	91.0	52.7	251
	8	M6043	Yes	11.3	9.56	84.5	8.01	83.7	67.3	306

[a] Samples in flow cell (FC) 1, FC 2, and FC 3 are from the first, second, and third trimesters, respectively.
[b] Raw pairs represent the paired-end reads that have passed the quality filters in the Illumina pipeline.
[c] Mapped pairs represent the sequence reads (read 1 and read 2) that both map uniquely by SOAP2 alignment (with 2 allowed mismatches) to the unmasked Hg18 reference genome.
[d] Mapped rate is the number of mapped pairs divided by the number of raw pairs.
[e] Nonduplicate pairs represent the mapped paired-end reads after filtering of the duplicated reads. Duplicated reads are defined as paired-end reads with the same sequences and start–end coordinates.
[f] The nonduplicate rate is the number of nonduplicate pairs divided by the number of mapped pairs.
[g] The targeted-region depth is the sum of the bases within the targeted region divided by the length of the targeted region.
[h] The -fold enrichment is the enriched targeted-region depth divided by the nonenriched targeted-region depth.
[i] NA, not applicable.

chrX, which contained the targeted region, was approximately 10-fold higher than the expected value (Fig. 1; see Figs. 1 and 2 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol57/issue1).

The depth of sequence coverage represented the mean number of times each base had been sequenced in a particular region. We calculated the sequence depth of the targeted region by dividing the total number of sequenced bases within the targeted region by the length of the targeted region (3.05 Mb). For the regions covered by the enrichment kit, the mean sequence coverage was 0.39 times for the nonenriched samples and 83.0 times for the enriched samples, indicating a mean enrichment of 213-fold (Table 1).

SEQUENCE COVERAGE WITHIN THE TARGETED REGION

We evaluated the sequence coverage across the targeted region to identify potentially underrepresented regions. In this study, we applied a sliding window of 10 kb across the 3.05-Mb targeted region and calculated the mean sequence coverage within each window of

Foresight EX1009
Foresight v Personalis

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

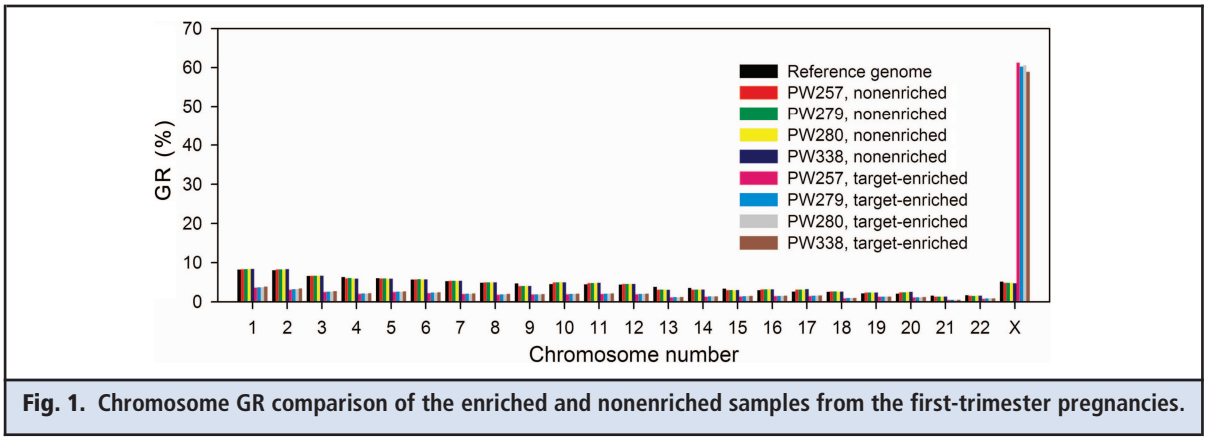

Fig. 1. Chromosome GR comparison of the enriched and nonenriched samples from the first-trimester pregnancies.

each sample, both with and without target enrichment (Fig. 2A; see Figs. 3A and 4A in the online Data Supplement). Consistent with the above-mentioned analysis of target-enrichment efficiency (Table 1), the sequence coverage in target-enriched samples was significantly higher than in nonenriched samples. We also observed a small number of underrepresented regions within the targeted region for all target-enriched samples (Fig. 2A; see Figs. 3A and 4A in the online Data Supplement). Further analysis revealed these underrepresented regions to contain a series of gaps (i.e., sequence coverage = 0) (Fig. 2B; see Figs. 3B and 4B in the online Data Supplement), which accounted for 5% of the targeted region (0.14 Mb). The algorithm used in this study was unable to map 98% of the gaps because they shared homologous sequences with other parts of the genome. In other words, DNA fragments derived from these gaps could not be mapped uniquely back to the reference genome.

We next evaluated the sample-to-sample consistency of sequence coverage within the mappable targeted region. To enable direct comparison of different samples regardless of the variation in target-enrichment efficiency, we normalized the sequence coverage by dividing the observed coverage of each base within the targeted region by the mean depth of the targeted region (Table 1) for the same sample. We then compared the normalized coverages of each base for any 2 target-enriched samples. Different samples were well correlated with respect to the distribution of sequence coverages (r^2 values from 0.72 to 0.93; see Table 1 in the online Data Supplement).

FETUS-SPECIFIC ALLELE DETECTION

The maternal and fetal genotypes for the 12 cases were determined by microarray analysis, and a SNP was considered informative when the mother was homozygous and the fetus was heterozygous, as described above in the Sequencing and Alignment section. According to these criteria, we identified 105 209 to 127 638 informative SNPs throughout the entire genome for the 12 cases; 52–113 informative SNPs fell within the targeted region (Tables 2 and 3). Because the DNA in maternal plasma was a mixture of maternal and fetal DNA, the genotype observed for maternal plasma DNA was a mixture of maternal and fetal genotypes. Therefore, the fetus-specific allele of an informative SNP should be detectable by sequencing maternal plasma DNA, as long as sufficiently deep sequencing was performed. At the sequencing depth of the current study, only 3.5% of the fetus-specific alleles within the targeted region were detected before target enrichment. By comparison, 95.9% of these became detectable after target enrichment. Therefore, target enrichment greatly increased the detection rate of fetus-specific alleles within the targeted region (Table 2).

FETAL DNA PROPORTIONS BEFORE AND AFTER ENRICHMENT

These data have demonstrated the enrichment of sequences within the targeted region. Another question of particular concern was whether maternal and fetal DNA molecules were enriched evenly, i.e., whether their relative proportions were preserved. Thus, we compared the proportions of fetal DNA on the basis of the read counts of all informative SNPs within the targeted region for each sample, with and without enrichment. For an informative SNP, for example, when the maternal genotype was AA and the fetal genotype was AT, the T allele counts should represent half a genome equivalent of fetal DNA. On the other hand, the A allele counts should represent 1 genome equivalent of maternal DNA plus the other half of the genome equivalent of fetal DNA. Thus, the fetal DNA proportion could be calculated according to this equation: Fetal DNA percentage = fetus-specific allele counts × 2/(fetus-specific allele counts + shared allele counts) × 100%.

To accurately determine the fetal DNA proportions in the samples, we first based our calculations on

Foresight EX1009
Foresight v Personalis

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

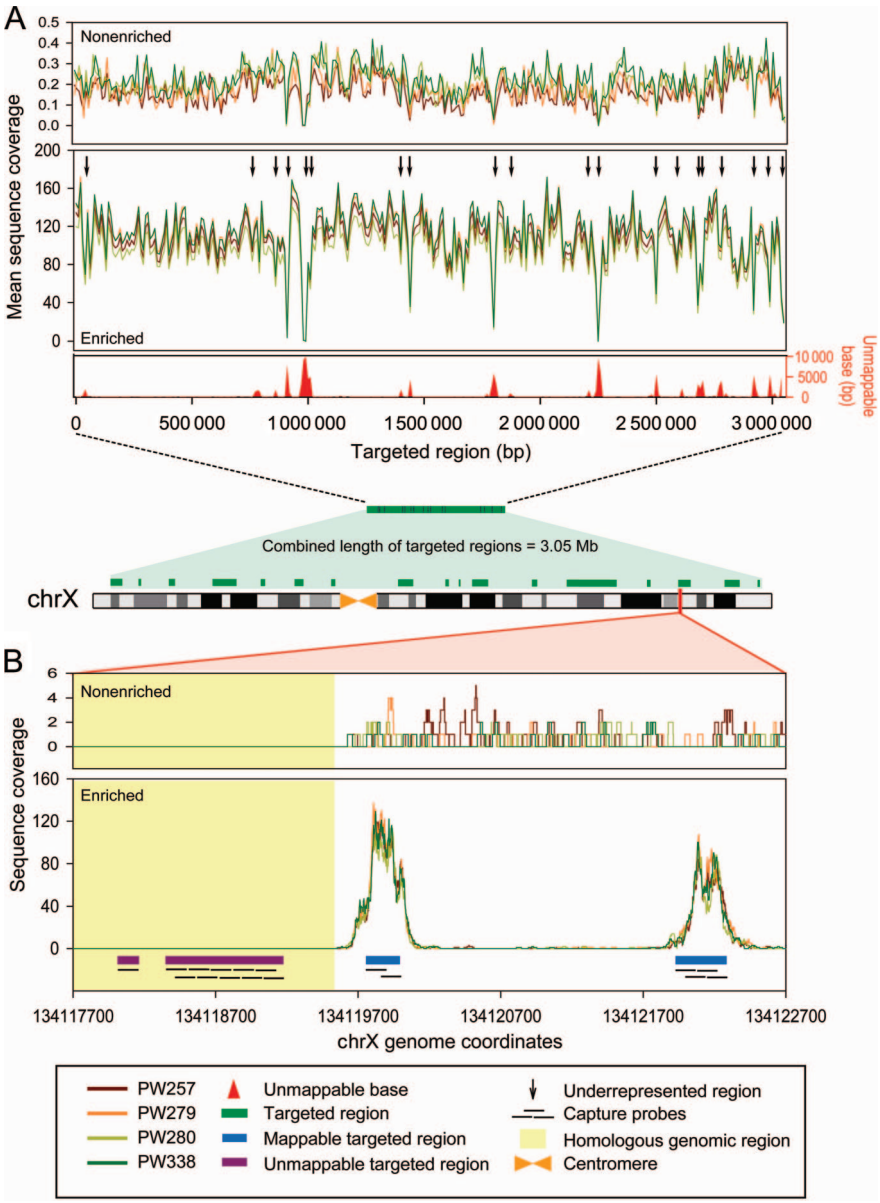

Fig. 2. Sequence coverage within the targeted region in enriched and nonenriched samples from the first trimester of pregnancy.

A schematic representation of chrX is in the middle of the figure. The targeted regions are scattered along the whole of chrX and are denoted by green horizontal bars. The total length of the targeted region is 3.05 Mb. (A), Sequence coverage along the entire targeted region. The solid lines (left *y* axis) represent the mean sequence coverage of 10-kb sliding windows with and without target enrichment, which is calculated by dividing the total number of sequenced bases within each sliding window by the window size (10 kb). The red peaks at the bottom (right *y* axis) represent the number of unmappable bases within each 10-kb sliding window. Note the concordance of the locations of the unmappable bases with the underrepresented regions (arrows). (B), High-resolution view within a 5-kb segment on chrX illustrating the relationship between the nature (i.e., mappable vs unmappable) of the targeted regions and the sequence coverage (per base) with and without target enrichment. Blue horizontal bars represent the mappable targeted regions, and the purple horizontal bars represent the unmappable targeted regions, which fall within the homologous genomic region (yellow area) sharing homologous sequences with other parts of the genome. Black horizontal bars represent the capture probes covering the targeted regions (probe length, 120 bp).

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

Foresight EX1009
Foresight v Personalis

Table 2. Rate of detection of fetus-specific alleles among informative SNPs in the targeted region in enriched and nonenriched samples.

FC[a]	Lane	Sample	Target enrichment	Informative SNPs in targeted region, n	Detectable fetus-specific alleles, n	Fetus-specific allele detection rate, %[b]	Shared allele counts, n	Fetus-specific allele counts, n	Fetal DNA proportion, %[c]
1	1	PW257	No	63	4	6.3	24	5	34.5
	2	PW279	No	54	1	1.9	20	1	9.5
	3	PW280	No	59	0	0.0	25	0	0.0
	4	PW338	No	52	1	1.9	24	1	8.0
	5	PW257	Yes	63	62	98.4	6842	593	16.0
	6	PW279	Yes	54	54	100	6514	729	20.1
	7	PW280	Yes	59	57	96.6	6020	456	14.1
	8	PW338	Yes	52	52	100	7002	398	10.8
2	1	M3667	No	113	6	5.3	54	6	20.0
	2	M4108	No	66	2	3.0	40	2	9.5
	3	M5166	No	67	3	4.5	54	4	13.8
	4	M6114	No	61	2	3.3	46	2	8.3
	5	M3667	Yes	113	109	96.5	11 639	672	10.9
	6	M4108	Yes	66	63	95.5	6357	745	21.0
	7	M5166	Yes	67	65	97.0	6358	384	11.4
	8	M6114	Yes	61	51	83.6	3866	204	10.0
3	1	M6011	No	63	6	9.5	13	6	63.2
	2	M6028	No	61	2	3.3	6	2	50.0
	3	M6029	No	69	2	2.9	11	2	30.8
	4	M6043	No	65	0	0.0	15	0	0.0
	5	M6011	Yes	63	60	95.2	3072	661	35.4
	6	M6028	Yes	61	60	98.4	2570	511	33.2
	7	M6029	Yes	69	66	95.7	3835	575	26.1
	8	M6043	Yes	65	61	93.9	3922	776	33.0

[a] Samples in flow cell (FC) 1, FC 2, and FC 3 are from the first, second, and third trimesters, respectively.
[b] The fetus-specific allele detection rate is the number of detectable fetus-specific alleles divided by the number of informative SNPs in the targeted region.
[c] Fetal DNA proportion (%) = fetus-specific allele counts × 2/(fetus-specific allele counts + shared allele counts) × 100%.

data from all of the informative SNPs in the genome in the runs without target enrichment. On the basis of this calculation, the mean proportions of fetal DNA were 16.08%, 13.68%, and 32.08% for the first-, second-, and third-trimester samples, respectively (Table 3). These results are consistent with data generated with digital PCR *(17)*. If we focus only on the targeted region, however, the range of fetus-specific reads was 0–6 for the samples without target enrichment. Because of the low sequence coverage, inadequate sampling of the fetal DNA molecules would prevent an accurate estimate of the proportion of fetal DNA (Table 2).

Conversely, with target enrichment, we observed a much larger number of fetus-specific allele counts (204–776) and shared allele counts (2570–11 639) within the targeted region. The mean proportions of fetal DNA were 15.25%, 13.33%, and 31.93% for first-, second-, and third-trimester samples, respectively. These findings are consistent with the fetal DNA percentages estimated for the whole-genome data in the nonenriched samples (Fig. 3). These results indicated that maternal and fetal DNA molecules were enriched to similar extents within the targeted region.

Discussion

Our study demonstrated that use of an in-solution capture strategy allowed even enrichment (213-fold) of maternal and fetal DNA molecules in the target region, increasing the representation of the targeted region from 0.1% to 21.3% of the total sequenced reads. These

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

Foresight EX1009
Foresight v Personalis

Table 3. Detection of fetus-specific alleles among informative SNPs in the whole genome in nonenriched samples.

FC[a]	Lane	Sample	Target enrichment	Informative SNPs in whole genome, n	Shared allele counts, n	Fetus-specific allele counts, n	Fetal DNA proportion, %[b]
1	1	PW257	No	109 330	39 972	3850	17.6
	2	PW279	No	107 102	36 970	4809	23.0
	3	PW280	No	109 135	39 258	2773	13.2
	4	PW338	No	105 209	39 531	2186	10.5
2	1	M3667	No	127 638	68 061	4239	11.7
	2	M4108	No	111 105	58 589	6297	19.4
	3	M5166	No	112 253	66 770	4293	12.1
	4	M6114	No	111 924	68 501	4197	11.5
3	1	M6011	No	120 184	15 309	3064	33.4
	2	M6028	No	110 730	16 778	3114	31.3
	3	M6029	No	107 362	19 889	3404	29.2
	4	M6043	No	110 321	21 070	4369	34.4

[a] Samples in flow cell (FC) 1, FC 2, and FC 3 are from the first, second, and third trimesters, respectively.
[b] Fetal DNA proportion (%) = fetus-specific allele counts × 2/(fetus-specific allele counts + shared allele counts) × 100%.

results imply that regions of interest can be sequenced more effectively with less sequencing effort. In previous studies on the NIPD of T21 *(7, 8, 18)*, one had to sequence from the entire genome nonselectively to obtain data for chr21, even though the prime focus of interest was chr21, which accounts for only about 1.5% of the genome. If we had substituted the 3.05-Mb target region in this study with a 1.5-Mb target region on chr21 and another 1.5-Mb target region on a reference chromosome, the representation of chr21 sequence data would presumably increase after target enrichment from approximately 1.5% to 10.7% (21.3%/2), an approximately 7-fold enrichment. A previous bioinformatics study *(18)* estimated that 140 Mb and 240

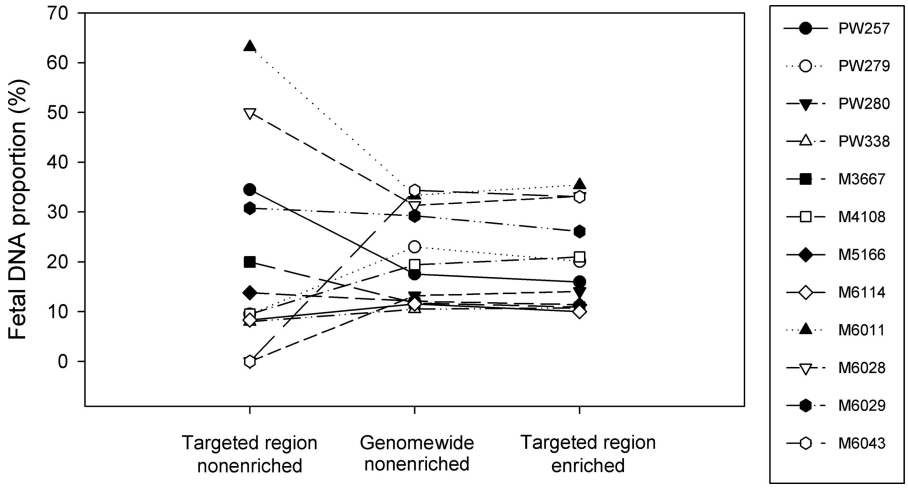

Fig. 3. Estimation of the fetal DNA proportion in enriched and nonenriched samples.
Targeted region nonenriched, fetal DNA proportion estimated by informative SNP counts within the targeted region in the nonenriched samples; genomewide nonenriched, fetal DNA proportion estimated by whole-genome informative SNP counts in nonenriched samples; targeted region enriched, fetal DNA proportion estimated by informative SNP counts within the targeted region in the target-enriched samples.

Foresight EX1009
Foresight v Personalis

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

Mb of sequence data would provide sufficient statistical power for T21 detection for samples with fetal DNA percentages of 10% and 5%, respectively. Applying target enrichment to this model indicates that approximately 86% (six-sevenths) of the sequencing effort could be saved.

Although the 213-fold enrichment of the target is encouraging, an additional consideration is how well the enriched sequences represent the original plasma DNA sample. To investigate the latter issue, one can compare the fetal DNA proportions before and after target enrichment. To estimate the proportion of fetal DNA as close to the "true" value as possible, we used whole-genome informative SNP counts to calculate the fetal DNA proportion before target enrichment. We then compared this value with the fetal DNA proportion obtained after target enrichment by using informative SNP counts within the targeted region. The consistency between these 2 parameters indicated the unbiased enrichment of both fetal and maternal DNA within the targeted region, a finding that strengthens the potential for applying targeted sequencing in NIPD.

Maternal plasma DNA represents a mixture of DNA from 2 individuals of the same species, in which the minor fetal DNA population is swamped by the overwhelming background of maternal DNA. Although it is possible to detect paternally inherited fetal alleles in maternal plasma, the low sequence coverage (0.39 times) before target enrichment makes such an application very expensive to implement. After target enrichment in the present study, the mean sequence coverage increased 83-fold, which allowed 95.9% of the paternally inherited alleles within the targeted region to be detected. The high probability for detecting a fetus-specific allele, if present, suggests that targeted sequencing would be a powerful method for accurately measuring the proportion of fetal DNA in a particular sample. This variable is an important parameter in many diagnostic applications involving fetal DNA in maternal plasma, e.g., for detecting aneuploidy and monogenic disease *(5, 7–9, 17–19)*. Although the detection of paternally inherited sequences is already achievable with other existing methods *(1, 4)*, targeted sequencing has an advantage in that it can capture multiple species of such paternally inherited sequences from multiple genomic regions of interest (see Table 2 in the online Data Supplement).

Because of the potential of targeted sequencing to save costs and increase throughput, this approach may play an important role in the translation of plasma DNA sequencing to molecular diagnostics, with impacts on such fields as NIPD, cancer diagnostics, detection of infectious agents, and transplantation monitoring.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: Y.M.D. Lo, Sequenom.
Stock Ownership: Y.M.D. Lo, Sequenom.
Honoraria: None declared.
Research Funding: Y.M.D. Lo and R.W.K. Chiu, University Grants Committee of the Government of the Hong Kong Special Administrative Region, China, under the Areas of Excellence Scheme (AoE/M-04/06), and Sequenom; Y.M.D. Lo, an Endowed Chair from the Li Ka Shing Foundation.
Expert Testimony: None declared.
Other Remuneration: F.M.F. Lun, Y.W.L. Zheng, K.C.A. Chan, Y.M.D. Lo, and R.W.K. Chiu, patents or filed patent applications on noninvasive prenatal diagnosis.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We thank Lisa Y.S. Chan and Yongjie Jin for performing the sequencing, and Hao Sun, Peiyong Jiang, and Zhang Chen for bioinformatics support.

References

1. Lo YMD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. Presence of fetal DNA in maternal plasma and serum. Lancet 1997;350:485–7.
2. Lo YMD, Chiu RWK. Prenatal diagnosis: progress through plasma nucleic acids. Nat Rev Genet 2007;8:71–7.
3. Lo YMD, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet 1998;62:768–75.
4. Van der Schoot CE, Soussan AA, Koelewijn J, Bonsel G, Paget-Christiaens LG, de Haas M. Non-invasive antenatal RHD typing. Transfus Clin Biol 2006;13:53–7.
5. Chiu RWK, Cantor CR, Lo YMD. Non-invasive prenatal diagnosis by single molecule counting technologies. Trends Genet 2009;25:324–31.
6. Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods 2008;5:16–8.
7. Chiu RWK, Chan KCA, Gao Y, Lau VY, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A 2008;105:20458–63.
8. Chiu RWK, Sun H, Akolekar R, Clouser C, Lee C, McKernan K, et al. Maternal plasma DNA analysis with massively parallel sequencing by ligation for noninvasive prenatal diagnosis of trisomy 21. Clin Chem 2010;56:459–63.
9. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc Natl Acad Sci U S A 2008;105:16 266–71.
10. Ng SB, Turner EH, Robertson PD, Flygare SD,

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

Foresight EX1009
Foresight v Personalis

Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009;461:272–6.
11. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 2010; 329:75–8.
12. Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. Massively parallel exon capture and library-free resequencing across 16 genomes. Nat Methods 2009;6:315–6.
13. Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH, et al. Microdroplet-based PCR enrichment for large-scale targeted sequencing. Nat Biotechnol 2009;27:1025–31.
14. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, et al. Direct selection of human genomic loci by microarray hybridization. Nat Methods 2007;4:903–5.
15. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009;27:182–9.
16. Chiu RWK, Poon LL, Lau TK, Leung TN, Wong EM, Lo YMD. Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. Clin Chem 2001;47:1607–13.
17. Lun FMF, Chiu RWK, Chan KCA, Leung TY, Lau TK, Lo YMD. Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clin Chem 2008;54:1664–72.
18. Chu T, Bunce K, Hogge WA, Peters DG. Statistical model for whole genome sequencing and its application to minimally invasive diagnosis of fetal genetic disease. Bioinformatics 2009;25:1244–50.
19. Lun FMF, Tsui NB, Chan KCA, Leung TY, Lau TK, Charoenkwan P, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. Proc Natl Acad Sci U S A 2008; 105:19 920–5.

Downloaded from https://academic.oup.com/clinchem/article/57/1/92/5621222 by guest on 27 October 2022

Foresight EX1009
Foresight v Personalis