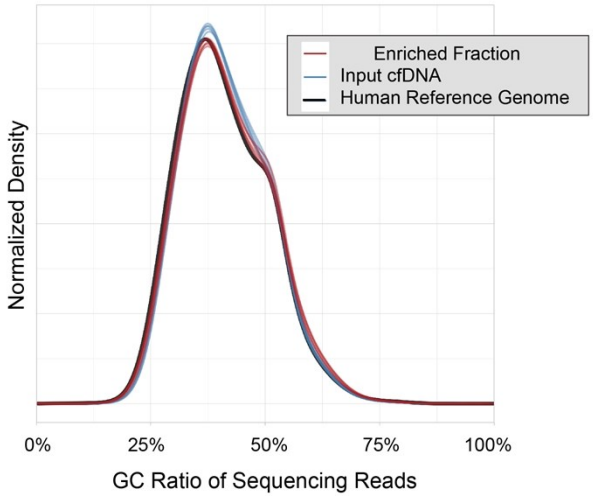
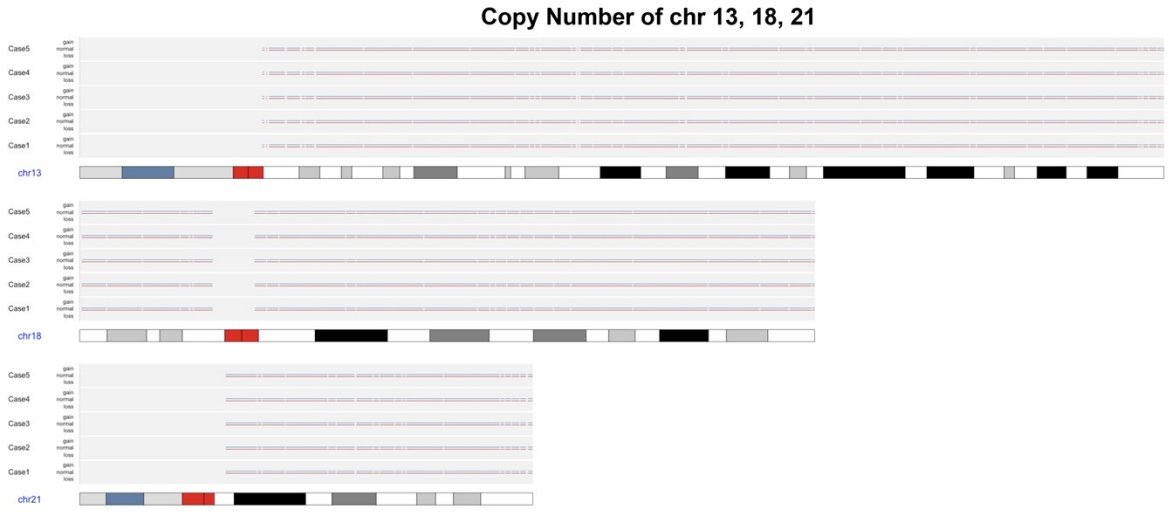
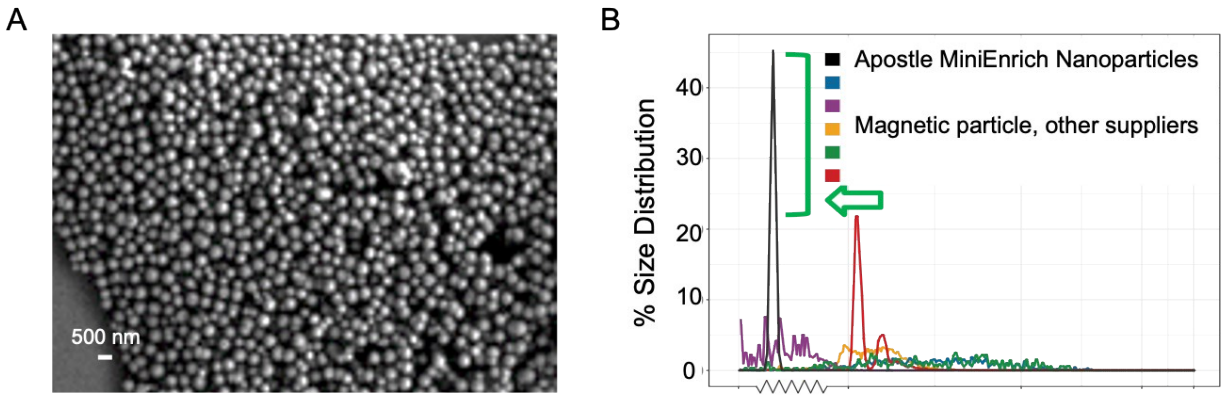
Electronic Supplementary Material (ESI) for Analyst.



Supplementary Figure 1. No evident GC bias introduced by the MiniEnrich workflow. The figure illustrates the GC content distributions of sequencing reads for 5 samples before (blue) and after (red) the enrichment. The theoretical GC content distribution, calculated from 200 million sampled reads from the human reference genome (hg19, autosomes only), is added to the figure as reference (black). The mean GC ratios of cfDNA samples before and after the enrichment are 41.7% and 42.2%, respectively, while the GC ratio of the human reference genome is 41.0%.



Supplementary Figure 2. No evident copy number of Chr 13, 18, 21 bias introduced by the MiniEnrich workflow. The copy number of autosomal chromosomes was estimated using the QDNAseq pipeline. The human genome without problematic regions was first divided into non-overlapping 50 kb bins, and the number of properly paired reads in each bin was counted. These counts were further corrected for both sequencing mappability and GC content using a two-dimensional locally weighted scatterplot smoothing (LOWESS) method. Pearson's correlations calculated between each paired copy number profiles showed high concordance (median Pearson correlation = 0.97), indicating no evident copy number alternations exist during the fetal DNA enrichment procedure.



Supplementary Figure 3. Size distribution of MiniEnrich magnetic nanoparticles. The SEM and DLS measurements show that nanoparticles developed here, name as MiniEnrich Substrates A and B (black), have a decreased particle size as well as uniform size distribution with minimized doublets compared to other magnetic particles (blue, purple, yellow, green, and red), which enables excellent suspension in solution, fast kinetics, and consistent size selection for the high resolution and yield of MiniEnrich workflow. DLS was performed using Zetasizer nano s90 following the manufacture's instruction.

Supplementary Table 1. An overview of additional 14 maternal plasma samples determined as male fetus with gestational weeks and ChrY percentage fold change after MiniEnrich workflow determined by ddPCR.

Case Number	Gestation	Fetus	ChrY% fold change after MiniEnrich
1	$31^{2}/_{7}$ weeks	Male	2.1
2	37 weeks	Male	5.3
3	$38\ ^{1}/_{7}$ weeks	Male	2
4	$16\ ^{2}/_{7}$ weeks	Male	1.8
5	$17\ ^{2}/_{7}$ weeks	Male	5.9
6	16 weeks	Male	5.3
7	16 weeks	Male	2.1
8	$17\ ^{5}/_{7}$ weeks	Male	1.8
9	$16\ ^{5}/_{7}$ weeks	Male	2.9
10	$16\ ^{4}/_{7}$ weeks	Male	1.1
11	$19\ ^{4}/_{7}$ weeks	Male	3.2
12	$16\ ^{3}/_{7}$ weeks	Male	2.6
13	$31\ ^{4}/_{7}$ weeks	Male	4.6
14	$18\ ^{6}/_{7}$ weeks	Male	1.5

Supplementary Table 2. An overview of 5 maternal plasma samples determined as male fetus with gestational weeks.

Case Number	Gestation	Fetus
1	16 weeks	Male
2	28 weeks	Male
3	37 weeks	Male
4	26 weeks	Male
5	25 weeks	Male

Materials and Methods

Clinical Samples

We obtained plasma samples from 10 pregnant women from Department of Pediatric, University of Chinese Academy of Sciences-Shenzhen Hospital (Shenzhen, China) with informed consent. The plasma of these 10 pregnant women were screened for male fetus and then 5 of them were used for the WGS-based validation study. We obtained additional 14 plasma samples from pregnant women carrying male fetus from Longgang District Maternity and Child Healthcare Hospital (Shenzhen, China) with informed consent and their plasma were used for the ddPCR-based validation study. Blood samples were collected in Apostle MiniMax® cfDNA Blood Collection Tubes (Apostle, United States) to protect the cfDNA from degradation and prevent genomic DNA contamination. Blood samples were collected in Apostle MiniMax® cfDNA Blood Collection Tubes (Apostle, United States) to protect the cfDNA from degradation and prevent genomic DNA contamination.

DNA Isolation

Cell-free plasma was obtained within 7 days of blood collection using two-step centrifugation: 1) 1,800×g for 10 min; 2) 11,800×g for 15 min. The cfDNA was isolated using the Apostle MiniMax® High Efficiency cfDNA Isolation Kit (Standard Edition) (Apostle, United States) from 6 mL of plasma collected, following the manufacture's protocol. The fragment size of cfDNA after the isolation were assessed using a Bioanalyzer 2100 (Agilent High Sensitivity DNA Kit, Agilent) and the concentration quantified with a Qubit Fluorometric Quantitation (Qubit™ dsDNA HS Assay kit, Thermo Fisher Scientific).

Verification of Fetal Sex by TaqMan-based real-time qPCR

To select cfDNA samples from pregnancies carrying a male fetus, the sex of fetus was determined by TaqMan-based real-time qPCR on cfDNA samples isolated from maternal plasma. For each reaction in PCR tubes (Applied Biosystems Cat 4358293) , the qPCR reaction mixtures (10 μL total volume) contained 5 μL TaqMan™ Fast Advanced Master Mix (Thermo Fisher), 0.5 μL each of the forward and reverse primers and TaqMan-MGB probe (VIC fluorescent dye, 10 μM), 1 μL DNA sample (0.5 ng/μL), and 2.5 μL RNase/DNase-free water. The concentration of the primers and probe (Sangon Biotech) used in the qPCR assays was 0.5 μM. The sequences of primers and probes of chromosome Y are as following: Y-forward: CATTCTCAAGCAAAACATGG, Y-reverse: CAGCAGTAGAGCAGTCAG, and Y-probe: CGTTGACTACTTGCCCT. The sex-determining region Y gene (SRY, NCBI sequence accession number: NG_011751.1) is Y-chromosome specific, and the primers and probe were designed accordingly, with amplicon of 80 bp length.[1] The thermal cycling was performed on a StepOnePlus Real-Time PCR System (Thermo Fisher) as follows: 95°C for 10 min (1 cycle); then 39 cycles of 60°C for 1 min and 95°C for 15 s; 60°C for 2 min, and finally held indefinitely at 10°C. The ramp rate was 2°C /sec in all steps. Samples with Ct values < 38 were determined to be male fetuses, and negative controls were not amplified until Ct values > 40 or undetected. All PCR reaction was performed with three technical replicates, same concentration and volume of sample loading. Ct value was determined by the delta Ct with correction of ROX reference dye. Raw data analysis and Quantitation were done using StepOnePlus™ Software v2.3. In total, 5 of 10 CAS samples were determined to have a male fetus and were used for the enrichment process. The isolated cfDNA from each male fetus sample was divided into two halves, one was stored at -20°C as the input sample (before enrichment) before sequencing, the other part was processed by MiniEnrich workflow. Sample information (gestational time, fetal sex) is described in **Supplementary Table 2**.

In silico the specificity of PCR was assessed by BLAST (NCBI), and the result indicated that the primer pairs were specific to input template to get amplicon product (＜150 bp) as no other targets were found in selected database (RefSeq Representative Genome Database, Organism limited to Homo sapiens). Amplicon length (80 bp) was confirmed using USCS in silico PCR tool, and Chromosomal location of the amplicon is Chromosome Y:2787794-2787873 (GRCh38: CM000686.2), shown in Genome Browser of USCS. Primers and probe were designed from extra-exon region of SRY. For primers/probe specificity and inhibition testing, SRY-fragment standard dilution series with concentration at 50000, 10000, 2000, 400, 80 copies/ul were used as templates in TaqMan-based real-time qPCR calibration curve (Formula: Ct=-3.216*log(conc) + 28.554). The PCR efficiency was 1.04618 (Formula: PCR efficiency=10^(-1/slope) - 1), and r^2 of the calibration curve was 0.9996. The linear dynamic range of the qPCR system was at least three orders of magnitude (100~100,000).

Enrichment of short DNA fragments

A 20 bp step ladder (P1598, Sigma-Alderich) was diluted (1:80) using TE solution, and then treated with 1.2× volume ratio of Carboxyl-beads (Substrate A, Apostle, US). After binding for 5 min, the Carboxyl-beads with long DNA fragments was separated from the supernatant. Then, the supernatant was mixed with 1.2x volume ratio of Hydroxyl-beads (Substrate B, Apostle, US). The solution was then vortexed for 5 seconds. The tube was briefly centrifuged to collect the beads solution to the bottom and then Isopropanol was added to a final concentration of 55%. After brief vortexing the mixture was shaken at moderate-high speed for 10 minutes. The supernatant was carefully removed and the Hydroxyl-beads containing captured short DNA fragments were separated. The short and long DNA fragments on both beads were then washed with 80% Ethanol

and eluted. The elution volume was the same as input volume. Alternatively a commercial cfDNA standard (HD780, Horizon Discovery) was diluted to 1 ng/μL using elution solution, and then treated with the 1.2, 1.4, 1.6, and 1.8×Carboxyl-beads (Substrate A), followed by the same Hydroxyl-bead (Substrate B) treatment as described above. The size of the input and enriched fragments was measured using a Bioanalyzer 2100 (Agilent High Sensitivity DNA Kit, Agilent), and the percentage of cfDNA <150 bp was calculated using the Agilent 2100 Bioanalyzer software.

For the enrichment of cfDNA isolated from clinical maternal plasma samples, 1.6×Carboxyl-beads was chosen to optimize the enrichment of < 150 bp cfDNA fragments. The isolated cfDNA was mixed with 1.6×volume Carboxyl-bead solution (Substrate A) (1.6 ratio to cfDNA input volume) thoroughly and the mixture was incubated at room temperature for 5 minutes. After Carboxyl-beads separated by magnet, the supernatant was then mixed with 1.2x volume ratio of Hydroxyl-beads (Substrate B, Apostle, US). The solution was then vortexed for 5 seconds. The tube was briefly centrifuged to collect the beads solution to the bottom and then Isopropanol was added to a final concentration of 55%. After brief vortexing the mixture was shaken at moderate-high speed for 10 minutes. The supernatant was carefully removed and the Hydroxyl-beads containing captured short DNA fragments were washed twice with 80% ethanol. Then elution solution was added to the Hydroxyl-beads to recover the short fragments in the original amount of input volume. Enriched samples were analyzed using a Bioanalyzer 2100 (Agilent) and quantified by Qubit Fluorometric Quantitation (Thermo Fisher) before sequencing.

Sequencing and Computational analyses

The WGS was performed on 10 samples (5 before the enrichment and 5 after the enrichment) by BerryGenomics (Shenzhen, China) on an Illumina NoveSeq® platform. Between 3-5 ng of cfDNA

before and after enrichment was used to prepare the cfDNA libraries, following by 2×75 paired-end WGS with an average sequencing depth of 11.9×. Pair-ended reads were aligned to human reference genome GRCh37 using BWA v0.7.10 and de-duplicated using Picard v2.20.7. Reads which mapped to multiple positions were removed. DEFRAG, a well-established Y-chromosome-based fetal fraction estimation tool,[2, 3] was utilized to estimate the fetal fraction for these samples.

The copy number of autosomal chromosomes was estimated using the QDNAseq pipeline.[4] The human genome without problematic regions was first divided into non-overlapping 50 kb bins, and the number of properly paired reads in each bin was counted. These counts were further corrected for both sequencing mappability and GC content using a two-dimensional locally weighted scatterplot smoothing (LOWESS) method. Pearson's correlations calculated between each paired copy number profiles showed high concordance (median Pearson correlation = 0.97), indicating no evident copy number alternations exist during the fetal DNA enrichment procedure.

Droplet Digital PCR on ChrY percentage

ddPCR was performed on 14 samples from Department of Pediatric, University of Chinese Academy of Sciences-Shenzhen Hospital. The ddPCR assays used in this study were optimized and validated in the ddPCR system by Thermo Fisher Scientific (Waltham, MA, USA). Primers and probes designed for conserved regions (known to have single copy) of XIST (Xq13) and SRY (Yp11) was used to identify X and Y chromosome, respectively. Sequences of primers and probes for ddPCR assays are: X-Forward: AGTCACACCCACTTGTT, X-Reverse: GGGCCATACAATCTGTTG, X-Probe: CTCTGACCTGGTAGCAC; Y-Forward: CATTCTCAAGCAAAACATGG, Y-Reverse: CAGCAGTAGAGCAGTCAG, Y-Probe: CGTTGACTACTTGCCCT. The ddPCR reaction mixtures were prepared as 14.5 μL total

volumes, containing 7.25 μL QuantStudio™ 3D Digital PCR Master Mix (Thermo Fisher Scientific), 0.725 μL each of the forward and reverse primers and probes (10 μM), 1.45 μL DNA sample, and 3.625 μL RNase/DNase-free water. The 14.5 μL of reaction mixture was loaded into a QuantStudio™ 3D Digital PCR 20K Chip v2 (Thermo Fisher Scientific), and reactions were partitioned into a chip (median of approximately 20,000 wells per chip) using the QuantStudio™ 3D Digital PCR Sample Loading Blade (Thermo Fisher Scientific) according to the Thermo Fisher Scientific protocol. The chip was sealed by QuantStudio™ 3D Digital PCR Chip Lid v2 using Immersion Fluid (Thermo Fisher Scientific). Then, PCR was performed with a ProFlex™ 2× Flat PCR System (Thermo Fisher Scientific). The thermal cycling conditions were as follows: 96℃ for 10 min (1 cycle); then 39 cycles of 60℃ for 1 min and 98℃ for 30 s; 60℃ for 2 min, and finally held indefinitely at 10℃. The ramp rate was 2℃/sec in all steps. After thermal cycling, wells were analyzed for positive and negative signals using the QuantStudio™ 3D Analysis Suite™ Software. Y chromosome copy proportion of sex chromosomes (chrY%) was calculated as follows: chrY% = copy number of Chr Y/(copy number of Chr X+ copy number of Chr Y). Samples information (gestational time, fetus sex, X chromosome and Y chromosome copies, chrY %) is listed in **Supplementary Table 1**.

Reference

1. E. Butler and R. Li, *J. Forensic. Investig.* 2014, **2(3)**, 10.

2. D. M. Beek, R. Straver, M. M. Weiss, E. M. J. Boon, K. H. Amsterdam, C. B. M. Oudejans, M. J. T. Reinders and E. A. Sistermans, *Prenat. diagn.* 2017, **37(8)**, 769-773.

3. M. S. Hestand, M. Bessem, P. Rijn, R. X. Menezes, D. Sie, I. Bakker, E. M. J. Boon, E. A. Sistermans and M. M. Weiss, *Eur. J. Hum. Genet.* 2019, **27(2)**, 198-202.

4. I. Scheinin, D. Sie, H. Bengtsson, M. A. Wiel, A. B. Olshen, H. F. Thuijl, H. F. Essen, P. P. Eijk, F. Rustenburg, G. A. Meijer, J. C. Reijneveld, P. Wesseling, D. Pinkel, D. G. Albertson and B. Ylstra, *Genome Res.* 2014, **24(12)**, 2022-2032.