Supporting Information

Chiu *et al.* 10.1073/pnas.0810641105

SI Text

Subjects and Sample Collection. Women with singleton euploid and trisomy 21 pregnancies were recruited with informed consent from the Department of Obstetrics and Gynaecology, Prince of Wales Hospital, Hong Kong. Ethical approval was obtained from the Joint New Territories East Cluster–Chinese University of Hong Kong Clinical Research Ethical Committee. Maternal peripheral blood samples were collected into EDTA blood tubes. All maternal blood samples from euploid pregnancies were collected before 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). Among the 14 trisomy 21 pregnancies analyzed, maternal blood was collected from 3 cases before chorionic villus sampling. For the remaining 11 cases of trisomy 21 pregnancies, maternal blood was collected immediately before pregnancy termination at a median of 6 days (range 2–22 days) after chorionic villus sampling or amniocentesis. The gestational ages and clinical setting at the time of maternal blood sampling for these cases are detailed in supporting information (*SI*) Table S5. Full karyotyping was performed in all 28 cases analyzed in the last part of the study to ascertain the fetal karyotypes.

Sample Processing and DNA Extraction. Maternal blood samples were centrifuged at 1,600 × *g* for 10 min at 4 °C, and the plasma portion was harvested and recentrifuged at 16,000 × *g* for 10 min at 4 °C (1). DNA from maternal plasma was extracted with the blood and body fluid protocol of the QIAamp DSP DNA blood mini kit (Qiagen). A total of 5–10 ml maternal plasma was used for DNA extraction. Multiple spin columns were used, with 0.8 ml of plasma being applied per spin column. DNA was eluted from each spin column in 55 μl water. The eluted DNA was pooled and then concentrated by a SpeedVac Concentrator (Thermo) into a final volume of 20 μl per case. The extracted plasma DNA was quantified by real-time PCR using an ABI 7300 Sequence Detector (Applied Biosystems). A β-globin real-time PCR assay was performed as described previously (2). A conversion factor of 6.6 pg of DNA per cell was used to calculate the yield of extracted plasma DNA.

Massively Parallel Genomic Sequencing. For plasma DNA sequencing, 11–50 ng of plasma DNA was used for DNA library construction by the beta Chromatin Immunoprecipitation Sequencing (ChIP-Seq) sample preparation kit (Illumina). The beta ChIP-Seq protocol was used because it was the protocol at the time optimized for low amounts of input DNA. No actual chromatin immunoprecipitation step was undertaken.

Briefly, end repairing of the plasma DNA fragments was performed using T4 DNA polymerase, Klenow DNA polymerase, and T4 polynucleotide kinase. Commercially available adapters (Illumina) were ligated to the DNA fragments after addition of terminal A-residues. The adapter-ligated DNA fragments were then enriched using a 15-cycle PCR with standard primers. The primers are PCR Primer 1.1 and PCR Primer 2.1 from Illumina. For the first experiment described in *Results*, the enriched adapter-ligated DNA fragments in the range of 150–300 bp were size selected using 2% agarose electrophoresis. The selected DNA libraries were then additionally amplified using the same 15-cycle PCR.

After the first experiments, all other experiments in this study followed the protocol with the omission of the last two steps. The adapter-ligated DNA was purified directly using spin columns provided in a QIAquick PCR purification kit (Qiagen). The adapter-ligated DNA libraries were quantified by using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and run on a gel to check for size distribution. The adapter-ligated DNA libraries were hybridized to the surface of sequencing flow cells. Only one lane of a flow cell was used for a library obtained from a single case. DNA clusters were generated using an Illumina cluster station, followed by 36 cycles of sequencing on an Illumina Genome Analyzer. Image processing and base calling were done using the Genome Analysis Pipeline Software (GAPipeline-0.2.2.5) (Illumina).

Basic Local Alignment Search Tool (BLAST) Analysis. For samples 1, 2, and 3, all U0–1-0–0 chrY sequences and 120 sequences on the other chromosomes were randomly picked to confirm the uniqueness of chromosomes by performing BLAST search against the human reference genome database in the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov).

Calculation of *Z*-Scores. The mean and SD of the percentage of representation of each human chromosome except chrY were calculated using the data of the cases carrying a male euploid fetus and were denoted as mean_{chr} and SD_{chr}, respectively. For example, the mean and SD of % chr1 for the 10 euploid male cases were denoted as $\text{mean}_{\text{chr1}}$ and SD_{chr1}, respectively. The *z*-score of a chromosome for a particular case was defined as the difference between the percentage of genomic representation (GR) of that chromosome of that particular case and the mean percentage of representation of that chromosome in the reference data set divided by the SD. For example, the *z*-score of case 1 is calculated as

$$z\ \text{score}_{\text{chr1\_case1}} = (\text{GR}_{\text{chr1\_case1}} - \text{mean}_{\text{chr1}})/\text{SD}_{\text{chr1}},$$

where $z\ \text{score}_{\text{chr1\_case1}}$ represents the *z*-score of chromosome 1 for case 1, $\text{GR}_{\text{chr1\_case1}}$ represents % chr1 for case 1, $\text{mean}_{\text{chr1}}$ represents the mean of % chr1 for the reference population (10 euploid male fetuses in this study), and SD_{chr1} represents the SD of % chr1 for the reference population.

Therefore, the *z*-score of a chromosome represents the number of SDs away from the reference mean percentage of representation of that chromosome. A *z*-score $> \pm 3$ for any chromosome of any case signifies a difference from the 99th percentile of the euploid cases for the same chromosome, i.e., a *P* value of 0.01.

Statistical Analyses. All statistics were performed using SigmaStat v.3.0 (Systat).

1. Chiu RWK, *et al.* (2001) Effects of blood processing protocols on fetal and total DNA quantification in maternal plasma. *Clin Chem* 47:1607–1613.

2. Lo YMD, *et al.* (1998) Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am J Hum Genet* 62:768–775.

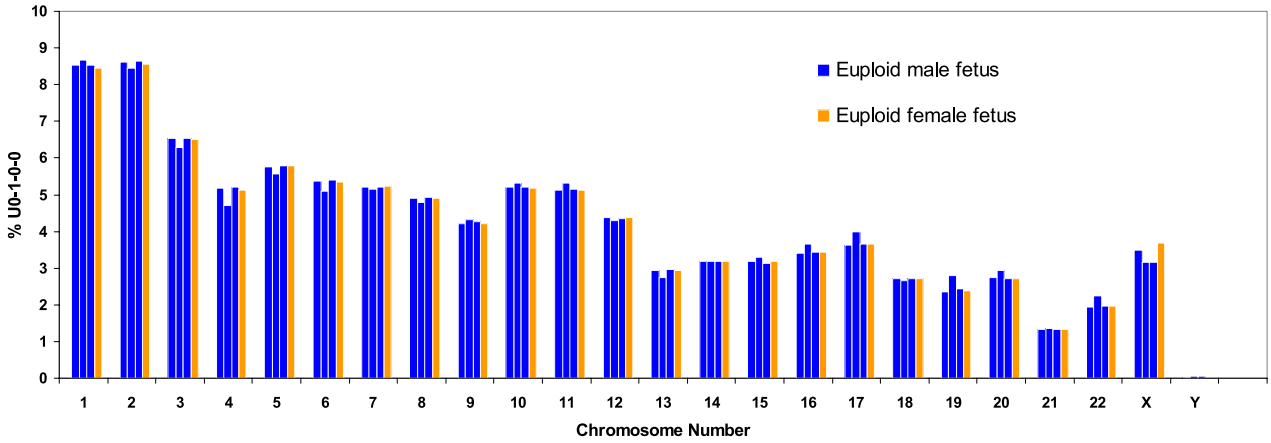


Fig. S1. Bar chart of % U0–1–0–0 sequences per chromosome for four euploid pregnancies.

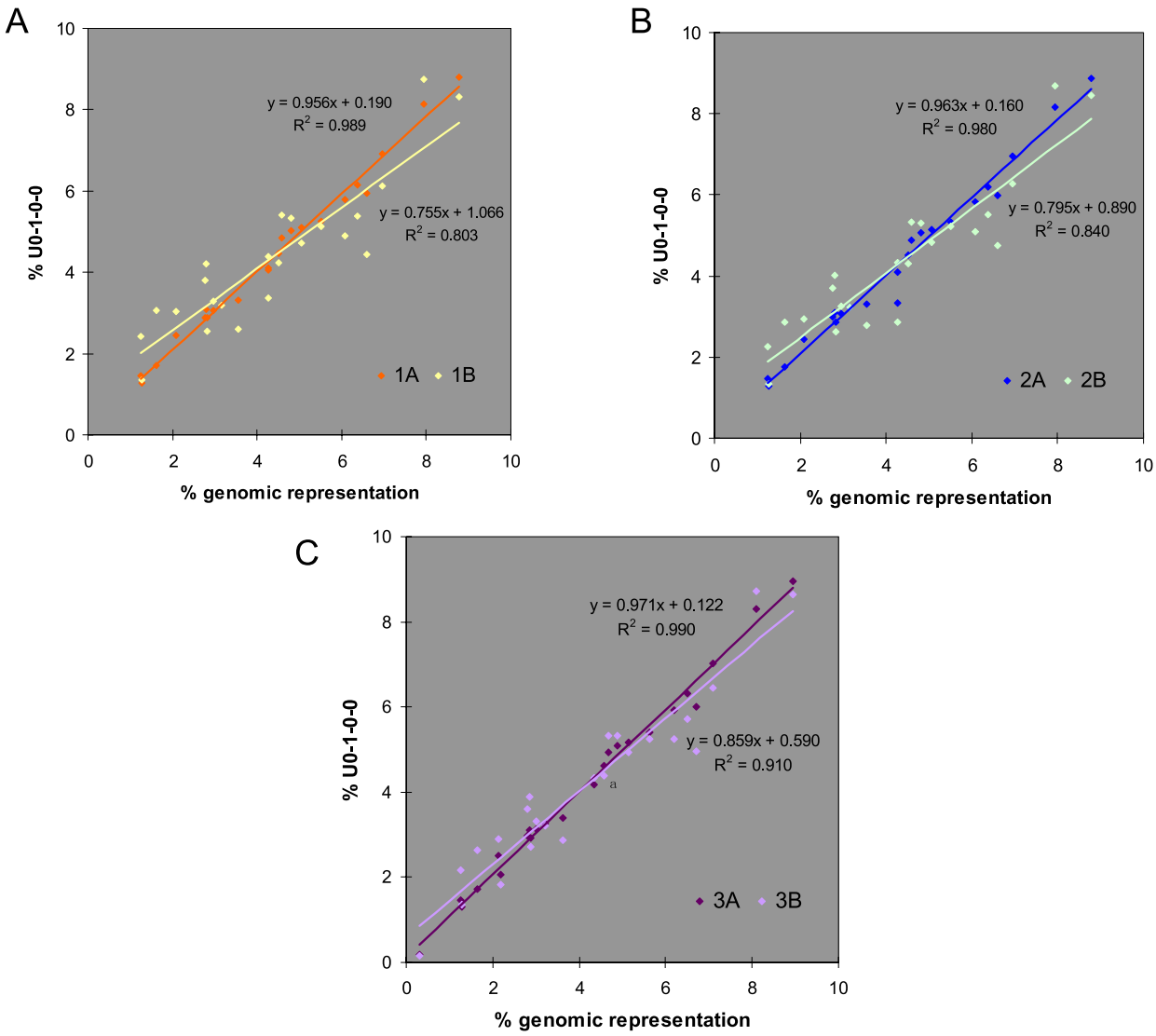


Fig. S2. Linear regression plots of % U0–1–0–0 per chromosome against the expected percentage of genomic representation of each human chromosome for (*A*) a maternal plasma sample involving a female fetus (sample 1), (*B*) a maternal plasma sample involving a male fetus (sample 2), and (*C*) a mixture of plasma from two adult males (sample 3) processed using the new (protocol A) and original (protocol B) protocols. For samples 1A, 1B, 2A, and 2B, the repeat-masked reference haploid female genome was used to calculate the expected genomic representation of each chromosome. For samples 3A and 3B, the repeat-masked reference haploid female genome plus the chromosome Y nucleotide contents (see *Materials and Methods*) was used to calculate the expected genomic representation of each chromosome.

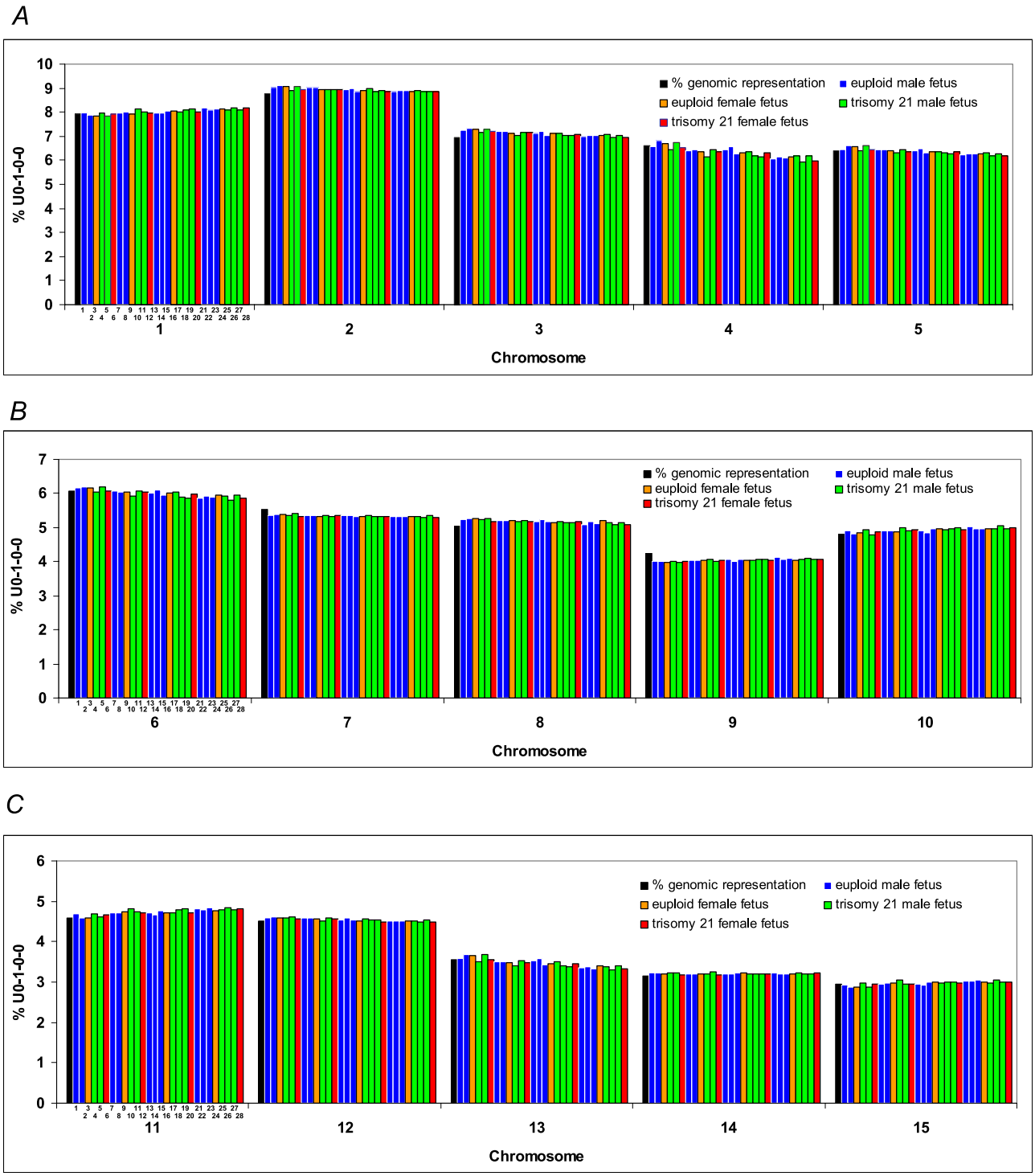
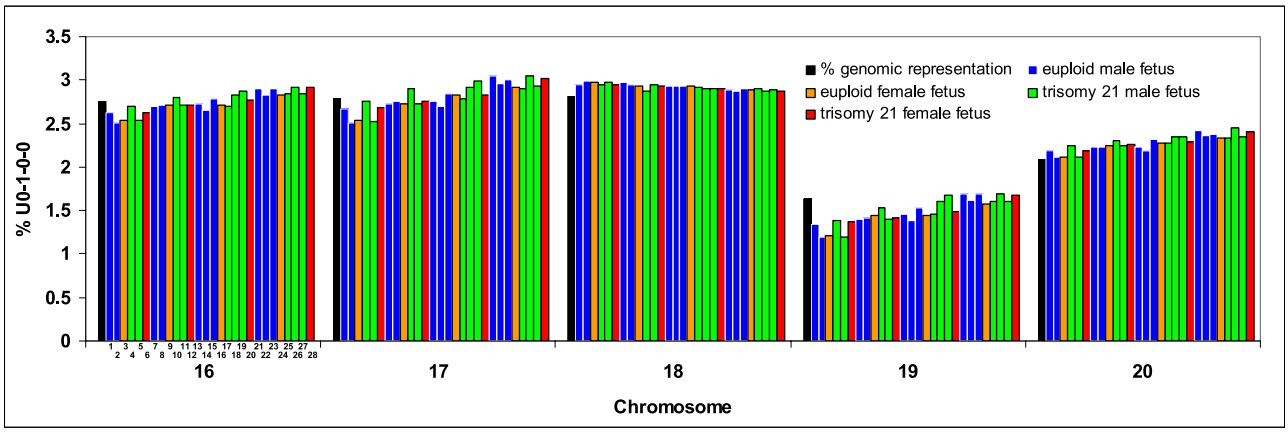
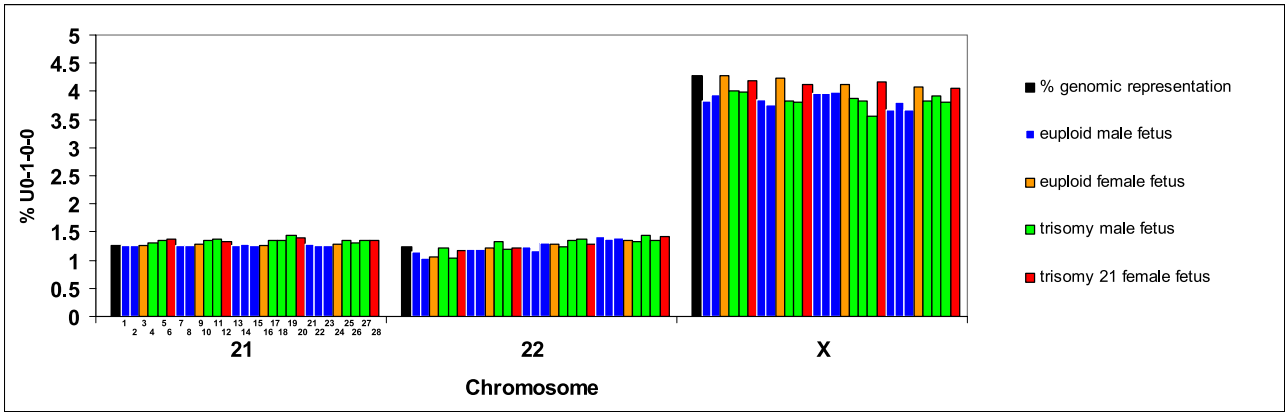


Fig. S3. Bar chart of % U0–1–0–0 sequences per chromosome for 28 maternal plasma samples. The percentage of genomic representation of each chromosome as expected for a repeat-masked reference haploid female genome was plotted for comparison (black bars). The sample numbers correspond to the cases described in Table S5.

D



E



F

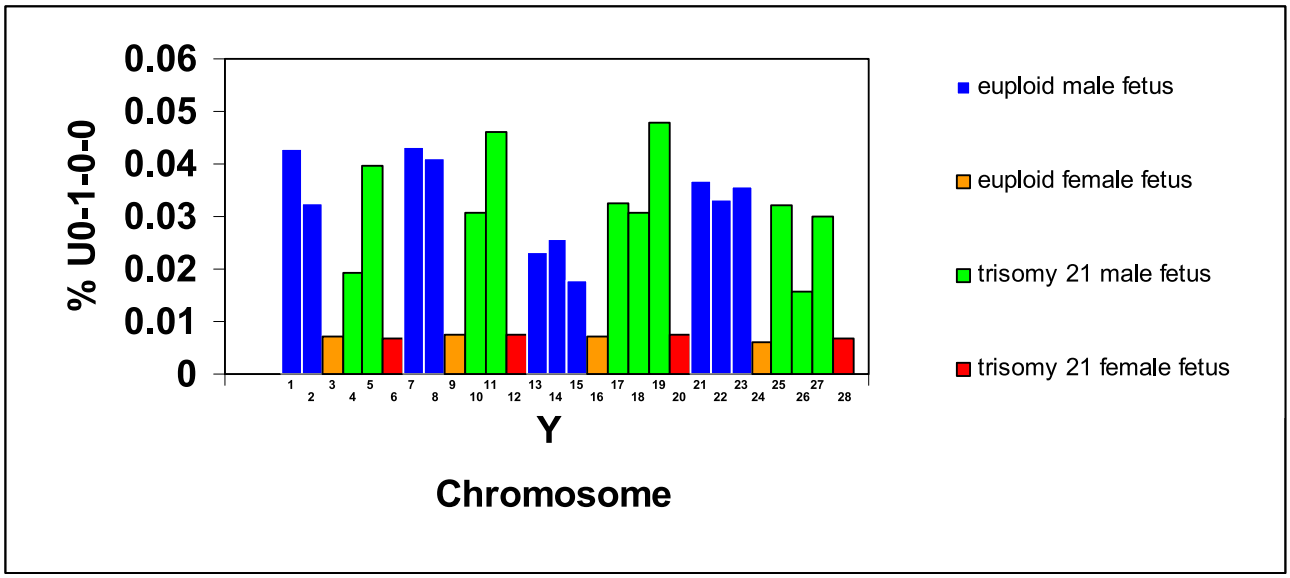


Fig. S3. Continued.

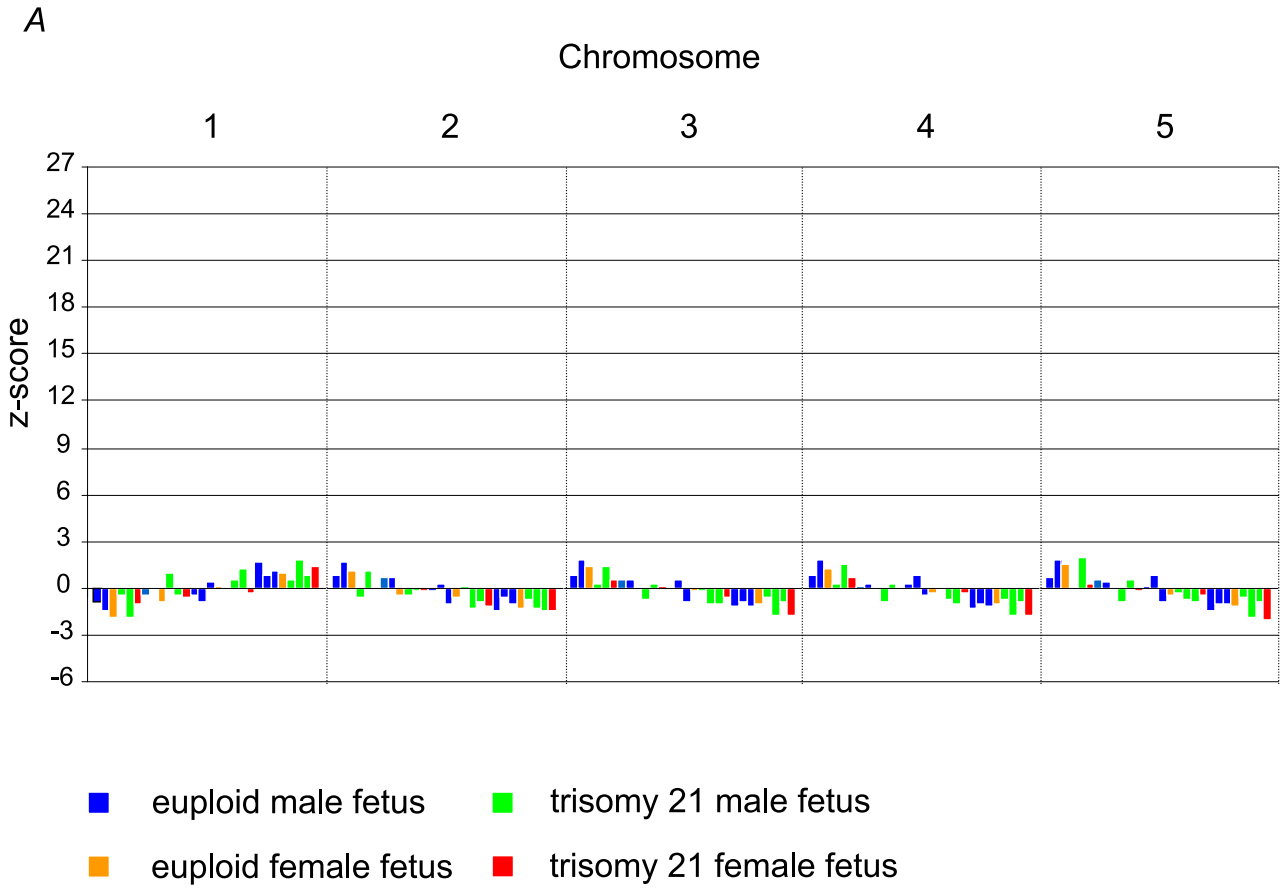


Fig. S4. Plot of *z*-score for each chromosome for 28 maternal plasma samples. Each of the 28 bars shown for each chromosome corresponds to the *z*-scores for each of the 28 maternal plasma samples. Samples 1–28 are shown consecutively from left to right. The sample numbers correspond to the cases described in Table S5.

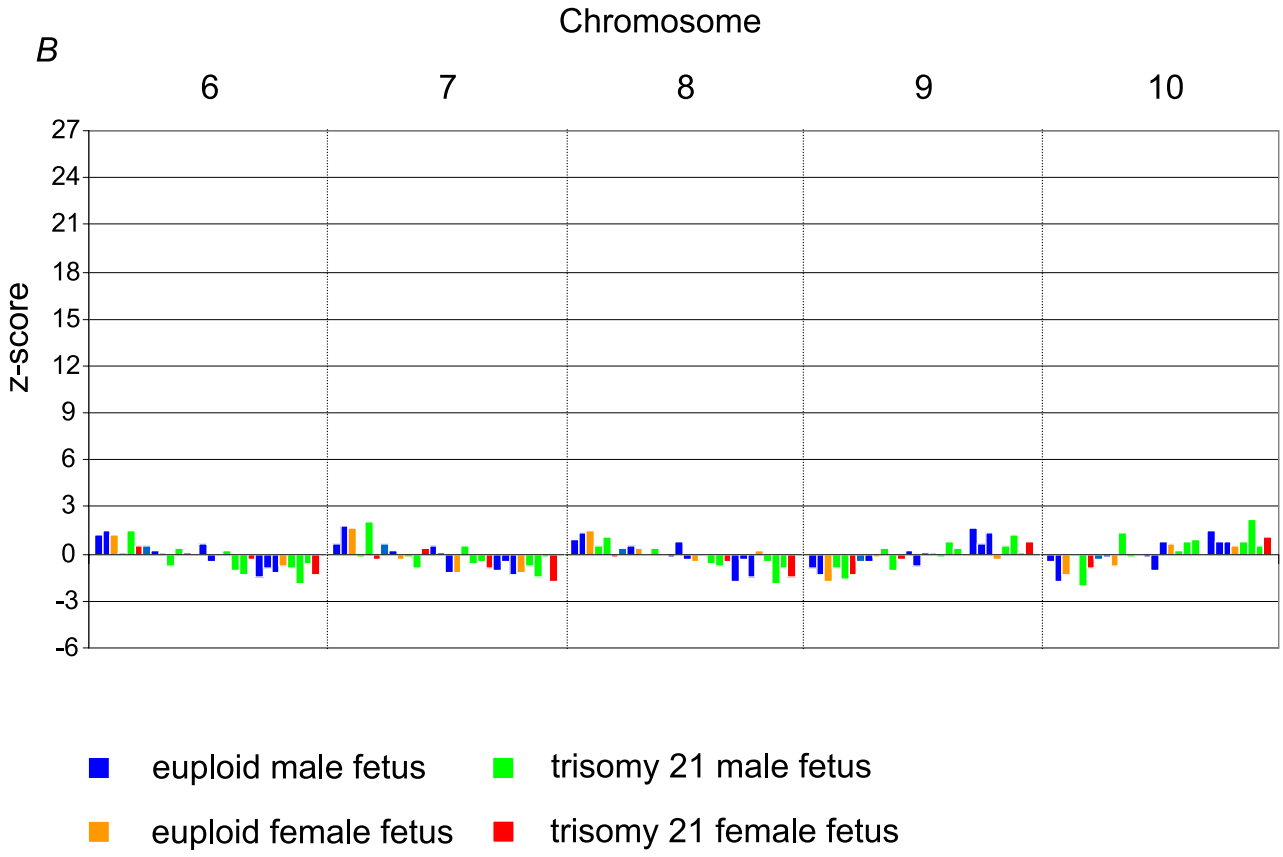


Fig. S4. Continued.

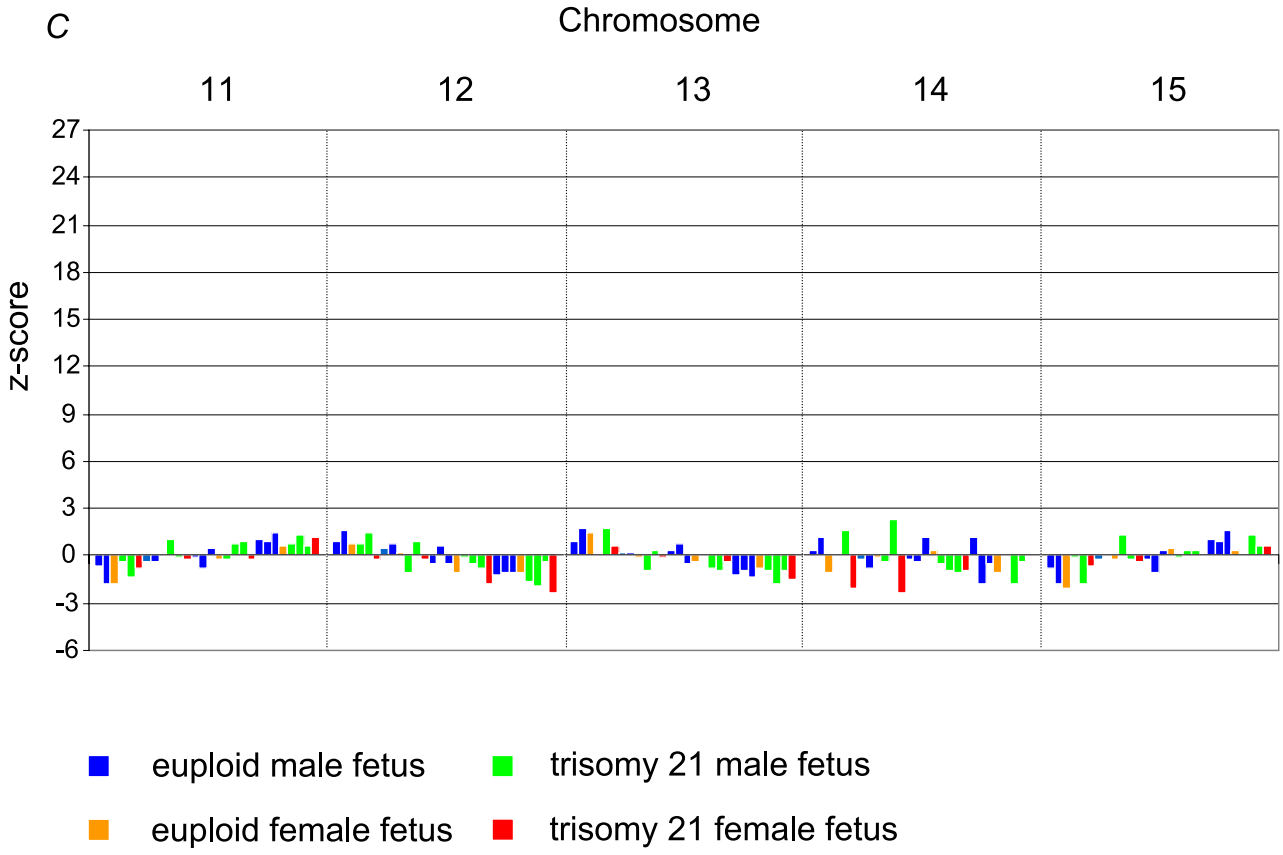


Fig. S4. Continued.

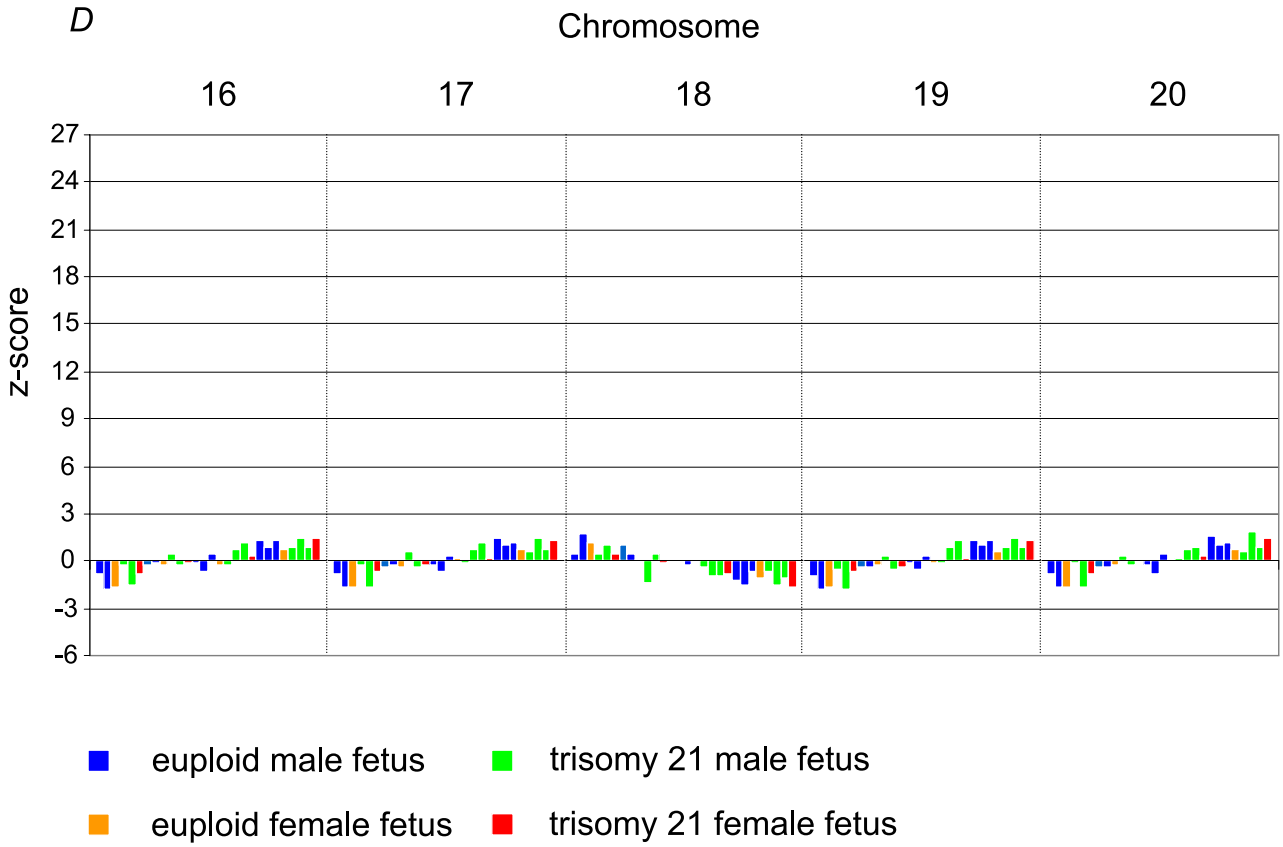


Fig. S4. Continued.

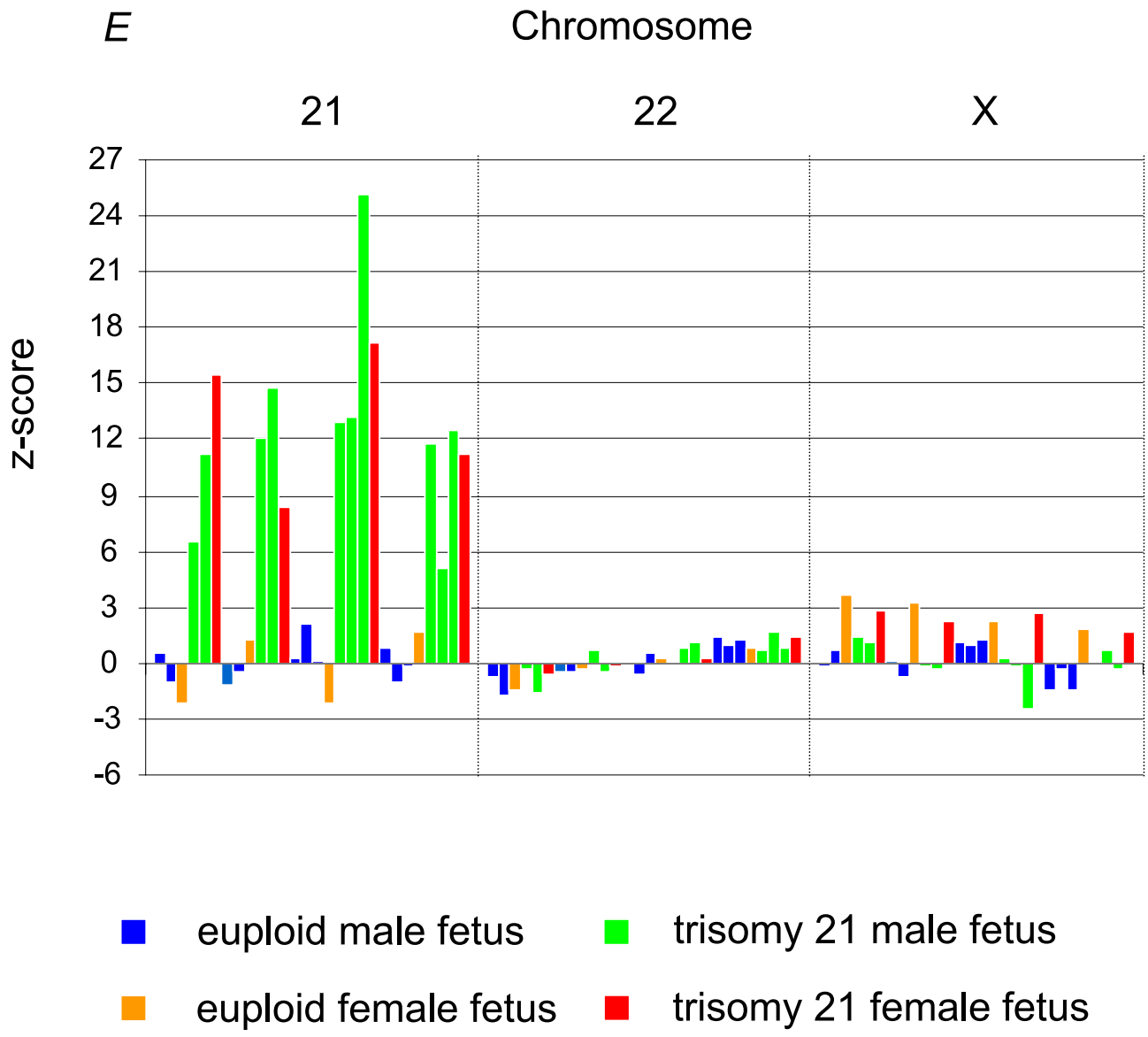


Fig. S4. Continued.

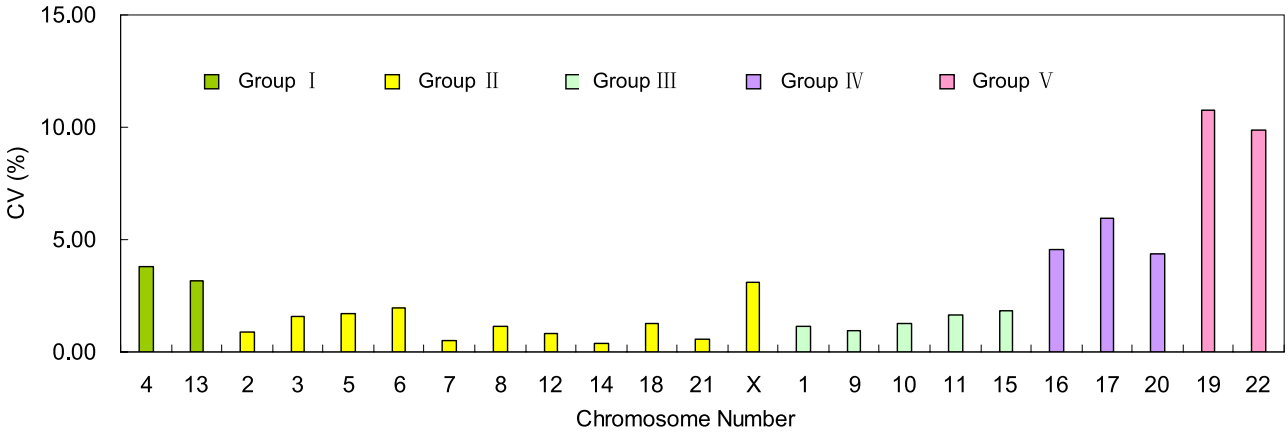


Fig. S5. Plot of coefficient of variation (CV) per chromosome. Chromosomes are grouped according to their GC contents. Group I chromosomes have the lowest GC contents while group V chromosomes have the highest GC contents.

Table S1. Clinical details and sequencing counts of maternal plasma samples analyzed by the first protocol

Sample	Fetal sex	GA (weeks + days)	Input DNA (ng)	Karyotype	Total sequenced counts	Total U0–1–0–0 counts	U0–1–0–0/total sequenced counts (%)	ChrY U0–1–0–0 counts	% chrY[†]
3009	M	17+3	20.9	46XY	10,643,413	1,990,068	18.7	636	0.032
3034	M	17+5	11.3	46XY	9,578,522	1,792,238	18.7	858	0.048
3143	M	17+2	16.8	46XY	9,549,403	1,885,644	19.7	1,054	0.056
3735	F*	41+2	92.6	NA	9,087,258	1,968,465	21.7	177	0.009

M, male; F, female; GA, gestational age; NA, not available.
*Confirmed to be a healthy female fetus at birth.
[†]Percentage of U0–1–0–0 sequences mapped to chromosome Y among all U0–1–0–0 sequences.

Table S2. Clinical details and sequencing counts of maternal plasma samples analyzed by the new and original protocols

Sample	Fetal sex	GA (weeks + days)	Input DNA (ng)	Total sequenced counts	Total U0–1–0–0 counts	U0–1–0–0/total sequenced counts (%)	ChrY U0–1–0–0 counts	% chrY†
1A	F	38 + 4	50	9,194,718	2,096,366	22.8	184	0.009
1B			50	10,030,317	2,012,469	20.1	218	0.011
2A	M	38 + 4	50	8,560,897	2,187,094	25.5	1,444	0.066
2B			50	9,556,038	2,093,089	21.9	1,615	0.077
3A	NA*	NA*	50	8,573,956	2,010,318	23.4	3,523	0.175
3B			50	9,614,592	2,057,491	21.4	3,468	0.169

M, male; F, female; GA, gestational age; NA, not applicable.
*Case 3 was a mixture of plasma samples from two adult males.
†Percentage of U0–1–0–0 sequences mapped to chromosome Y among all U0–1–0–0 sequences.

Table S3. BLAST analysis of U0–1–0–0 sequences aligned to chromosome Y

Sample no.	Sample origin	No. of U0–1–0–0 sequences mapped to chromosome Y	Unique to chromosome Y	
			Counts	%
1A	Maternal plasma from a pregnancy with a female fetus	184	67/184	36.4
1B		218	66/218	30.3
2A	Maternal plasma from a pregnancy with a male fetus	1,444	1,335/1,444	92.5
2B		1,615	1,461/1,615	90.5
3A	Mixture of plasma from two adult males	3,523	3,409/3,523	96.8
3B		3,468	3,325/3,468	95.9

Table S4. BLAST analysis of U0–1–0–0 sequences aligned to the other chromosomes other than chromosome Y

	% of uniqueness		
Chromosome no.	Sample 1A	Sample 2A	Sample 3A
1	100	99.17	100.00
2	100	100.00	100.00
3	100	100.00	100.00
4	100	100.00	100.00
5	100	100.00	99.17
6	100	100.00	100.00
7	100	100.00	100.00
8	100	100.00	100.00
9	100	100.00	100.00
10	100	100.00	100.00
11	100	100.00	99.17
12	100	100.00	100.00
13	100	100.00	100.00
14	100	99.17	100.00
15	100	100.00	100.00
16	100	100.00	100.00
17	100	100.00	100.00
18	100	99.17	100.00
19	100	99.17	100.00
20	100	100.00	100.00
21	100	100.00	100.00
22	100	100.00	99.17
X	100	97.50	97.50

Table S5. Clinical details and sequencing counts of maternal plasma samples from pregnancies with euploid or trisomy 21 fetuses

Sample no.	Case no.	Fetal sex	*Time of blood sampling: GA (weeks + days)	*Time of blood sampling: Setting	Input DNA (ng)	Karyotype	Batch	Flow cell	Total sequenced counts	Total U0–1–0–0 counts	U0–1–0–0/total sequenced counts (%)
1	2972	M	13 + 2	PI	16	46XY	1	A	12,183,087	2,269,126	18.6
2	3245	M	17 + 2	PI	14	46XY	1	A	11,895,455	2,385,658	20.1
3	2987	F	17 + 6	PI	20	46XX	1	A	11,293,242	2,339,280	20.7
4	1519	M	20 + 3	PT (22 days)	29	47XY + 21	1	A	10,403,129	2,331,208	22.4
5	2849	M	14 + 3	PT (2 days)	40	47XY + 21	1	A	11,716,539	2,272,403	19.4
6	2104	F	14 + 3	PT (9 days)	21	47XX + 21	1	A	11,153,644	2,344,612	21.0
7	4181	M	17 + 2	PI	18	46XY	1	B	11,231,617	2,233,436	19.9
8	2791	M	13 + 5	PI	13	46XY	1	B	10,679,744	2,374,594	22.2
9	3518	F	19	PI	19	46XX	1	B	10,299,471	2,285,026	22.2
10	3228	M	14 + 5	PT (6 days)	19	47XY + 21	1	B	10,625,019	2,271,819	21.4
11	3438	M	13 + 1	PT (5 days)	23	47XY + 21	1	B	10,417,954	2,416,779	23.2
12	4022	F	19 + 6	PT (22 days)	49	47XX + 21	1	B	9,064,594	2,310,510	25.5
13	4402	M	12 + 4	PI	20	46XY	2	C	11,782,969	2,769,857	23.5
14	4404	M	17 + 5	PI	20	46XY	2	C	11,203,712	2,743,024	24.5
15	4420	M	13 + 5	PI	20	46XY	2	C	9,820,505	2,590,205	26.4
16	4423	F	18 + 3	PI	20	46XX	2	C	11,118,971	2,697,932	24.3
17	2291	M	12 + 4	PI	17	47XY + 21	2	C	11,576,114	2,500,342	21.6
18	2479	M	17 + 3	PT (5 days)	20	47XY + 21	2	C	10,212,009	2,575,318	25.2
19	2503	M	11 + 6	PI	20	47XY + 21	2	C	10,063,515	2,451,967	24.4
20	2016	F	19 + 3	PT (7 days)	20	47XX + 21	2	C	11,060,628	2,702,828	24.4
21	4421	M	12 + 3	PI	20	46XY	2	D	10,206,122	2,427,082	23.8
22	4422	M	17	PI	20	46XY	2	D	10,776,947	2,610,307	24.2
23	4443	M	11 + 6	PI	20	46XY	2	D	9,661,057	2,396,222	24.8
24	4441	F	11 + 2	PI	20	46XX	2	D	12,203,125	2,718,492	22.3
25	3148	M	12 + 3	PI	20	47XY + 21	2	D	11,550,907	2,622,551	22.7
26	4386	M	13 + 6	PT (8 days)	20	47XY + 21	2	D	9,626,221	2,516,289	26.1
27	4416	M	13	PT (4 days)	20	47XY + 21	2	D	11,211,225	2,612,813	23.3
28	2848	F	12 + 6	PT (4 days)	20	47XX + 21	2	D	9,338,553	2,348,996	25.2

M, male; F, female; GA, gestational age; PI, preinvasive; PT, pretermination.

*The gestational age and clinical setting at the time of maternal blood sampling are indicated. All euploid and 3 trisomy 21 pregnancies (cases 17, 19, and 25) were sampled before amniocentesis or chorionic villus sampling (hence, preinvasive). The remaining 11 trisomy 21 pregnancies were sampled immediately prior to termination of pregnancy with the interval after amniocentesis or chorionic villus sampling shown in parentheses.

Table S6. Coefficient of variation (CV) of measuring the percentage representation of each chromosome

Chromosome	CV (%)	Chromosome	CV (%)	Chromosome	CV (%)
1	1.16	9	0.95	17	5.93
2	0.91	10	1.28	18	1.29
3	1.56	11	1.61	19	10.74
4	3.83	12	0.82	20	4.37
5	1.72	13	3.14	21	0.54
6	1.94	14	0.40	22	9.90
7	0.51	15	1.82	X	3.10
8	1.16	16	4.57		