Supporting Information

Fan *et al.* 10.1073/pnas.0808319105

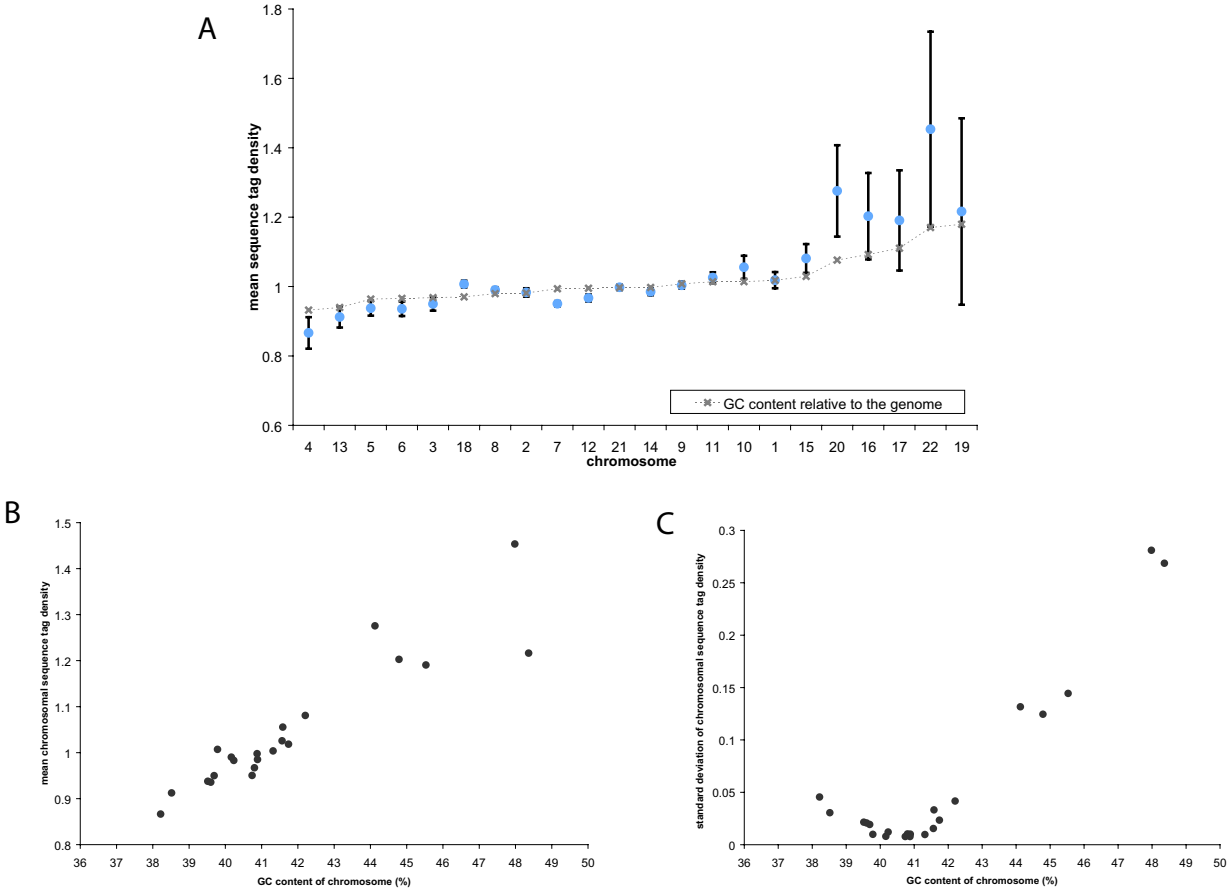

Fig. S1. Chromosomal sequence tag density relates to chromosomal GC content. (*A*) The mean sequence tag density for each chromosome of all samples, including cell-free plasma DNA from pregnant women and male donor, as well as genomic DNA control from male donor, is plotted. Exceptions are chromosomes 13, 18, and 21, where cell-free DNA samples from women carrying aneuploid fetuses are excluded. The error bars represent standard deviation. The chromosomes are ordered by their GC content. GC content of each chromosome relative to the genome-wide value (41%) is also plotted. (*B*) Scatter plot of mean sequence tag density for each chromosome versus GC content of the chromosome. The correlation coefficient is 0.927, and the correlation is statistically significant ($P < 10^{-9}$). (*C*) Scatter plot of the standard deviation of sequence tag density of each chromosome versus GC content of the chromosome. The correlation coefficient between standard deviation of sequence tag density and the absolute deviation of chromosomal GC content from the genome-wide GC content is 0.963, and the correlation is statistically significant ($P < 10^{-12}$).

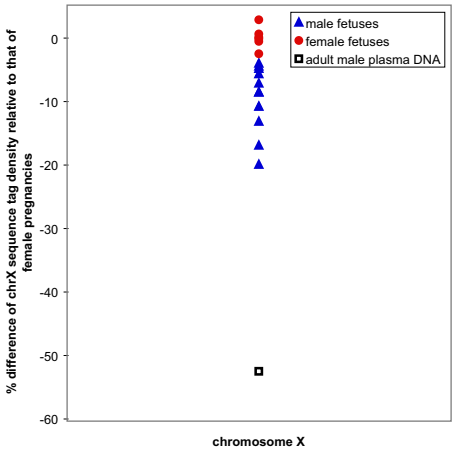

Fig. S2. Percent difference of chromosome X sequence tag density of all samples as compared with the median chromosome X sequence tag density of all female pregnancies ($n = 6$). All male pregnancies ($n = 12$) show underrepresentation of chromosome X.

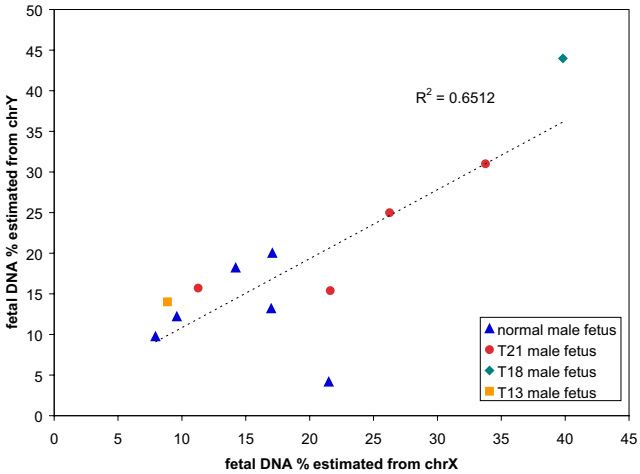

Fig. S3. Comparison of the estimation of fetal DNA fraction for cell-free DNA samples from 12 male pregnancies using sequencing data from chromosomes X and Y. The dashed line represents a simple linear regression line, with a slope of 0.85. The R^2 value represents the square of the correlation coefficient. There is a statistically significant correlation between fetal DNA fraction estimated from chromosomes X and Y ($P = 0.0015$).

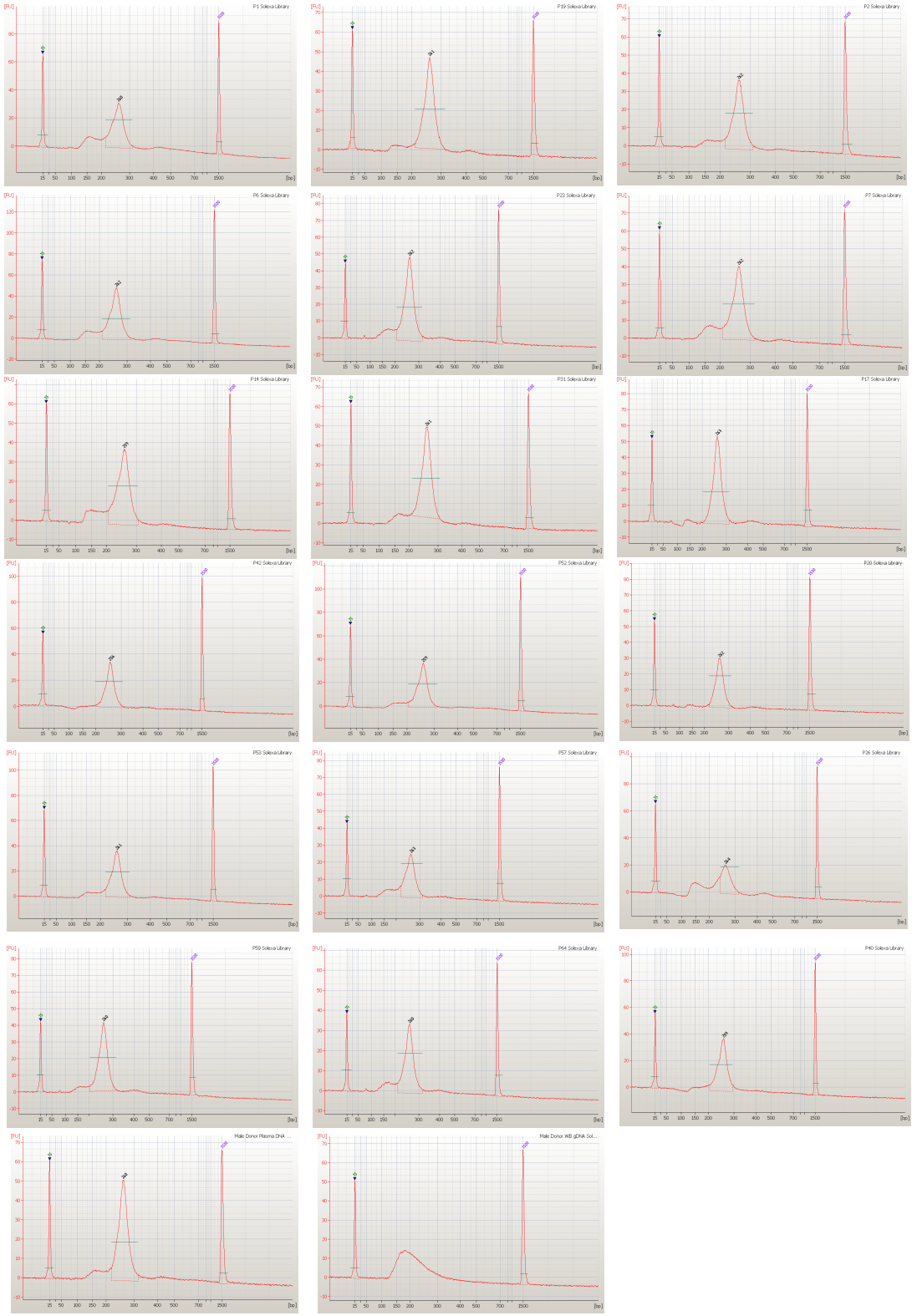

Fig. S4. Electropherograms of Solexa sequencing libraries prepared from cell-free plasma DNA obtained from 18 pregnant women and 1 male donor. Solexa library prepared from sonicated whole-blood genomic DNA from the male donor is also shown (bottommost, center column). For libraries prepared from cell-free DNA, all have peaks at average 261 bp (range: 256–264 bp). The actual peak size of DNA fragments in plasma DNA is ≈169 bp [after removal of Solexa universal adaptor (92 bp)]. This corresponds to the size of a chromatosome. The 15-bp and 1,500-bp peaks are molecular weight markers.

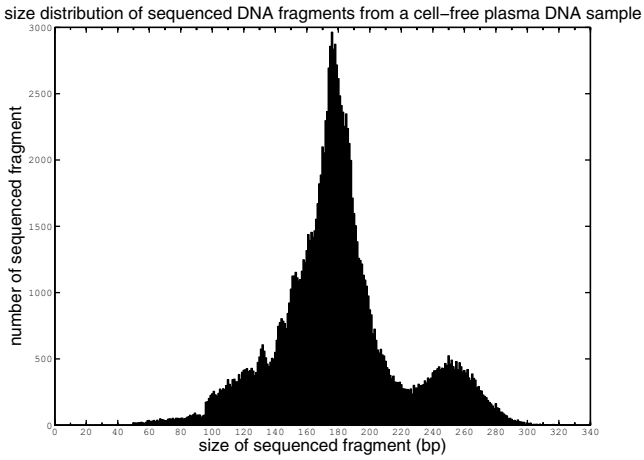

Fig. S5. Length distribution of sequenced fragments from maternal cell-free plasma DNA sample of a normal male pregnancy at 1-bp resolution. Sequencing was done on the 454/Roche platform. Reads that have at least 90% mapping to the human genome with ≥90% accuracy are retained, totaling 144,992 reads. The *y* axis represents the number of reads obtained. The median length is 177 bp, and the mean length is 180 bp.

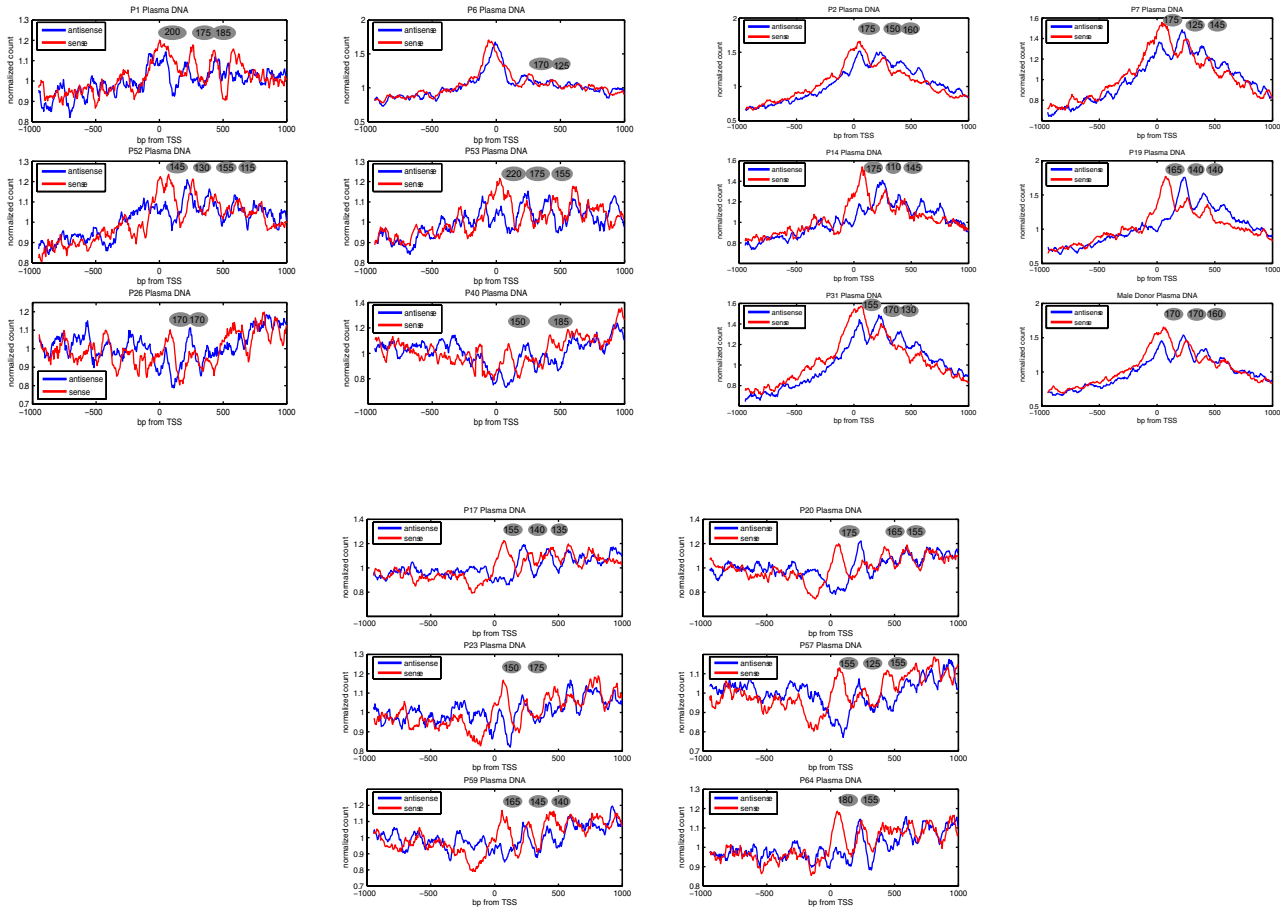

Fig. S6. Distribution of sequence tags around transcription start sites (TSS) of RefSeq genes. These plots are similar to Fig. 4 in the main text. Each plot represents the distribution for each plasma DNA or gDNA sample. See *Materials and Methods* for the construction of these figures. Note that these data are obtained from three different sequencing runs (P1, P6, P52, P53, P26, P40, and P42 were sequenced together; male genomic DNA, male plasma DNA, P2, P7, P14, P19, and P31 were sequenced together; P17, P20, P23, P57, P59, and P64 were sequenced together). (*Upper Left six graphs*) The distribution of tags around TSS for the first batch is shown, except P42, which is shown in the main text. (*Upper Right six graphs*) The distribution of tags around TSS for the second batch is shown, except male donor gDNA, which is shown in the main text. (*Bottom six graphs*) The distribution of tags around TSS for the third batch is shown. The second batch of samples suffers greater GC bias as observed from inter- and intrachromosomal variation. Their distributions around TSS have similar trends with more tags at the TSS. Such trend is not as prominent as in the distributions of samples sequenced in other runs. Nonetheless, at least three well positioned nucleosomes are detectable downstream of transcription start sites for most plasma DNA samples, suggesting that cell-free plasma DNA shares features of nucleosomal DNA, a piece of evidence that this DNA is of apoptotic origin.

Table S1. Sample information and sequencing metrics

Sample	Fetal karyotype	Gestational age, weeks	Vol. of plasma used for sequencing library creation, ml	Amt. of DNA in plasma (cell equivalent per ml of plasma)*	Approximate amt of input DNA for sequencing library construction, ng*	Total no. of sequence tags	No. of sequence tags mapped uniquely to the human genome (hg18) with at most one mismatch	Fetal DNA estimated by digital PCR with SRY assay, % (male fetuses)	Fetal DNA estimated by ChrY sequence tags, % (male fetuses)	Fetal DNA estimated by depletion of ChrX sequence tags, % (male fetuses)	Fetal DNA estimated by addition of trisomic chr sequence tags, % (aneuploid fetuses)	Overall GC content of sequence tags, %
P1 plasma DNA§	47XX+21	35	1.6	761	8.0	8,206,694	4,632,637	–	–	–	35.0	43.65
P2 plasma DNA§	47XY+21	18	1.4	585	5.2	7,751,384	4,313,884	6.4	15.4	21.6	15.5	48.72
P6 plasma DNA§	47XX+21	14	1.6	410	4.3	6,699,183	3,878,383	–	–	–	22.9	44.78
P7 plasma DNA§	47XY+21	18	2.2	266	3.8	8,324,473	4,294,865	9.1	31.0	33.8	28.6	48.07
P14 plasma DNA§	47XX+21	23	3.2	57	1.2	8,924,944	3,603,767	–	–	–	30.5	46.38
P17 plasma DNA§	47XX+21	16	2.3	210	3.2	11,599,833	5,968,932	–	–	–	7.8	44.29
P19 plasma DNA§	46XY	18	3.2	333	7.0	7,305,417	3,280,521	<5.9‡	4.14	21.5	–	50.09
P20 plasma DNA§	47XY+21	18	1.3	408	3.6	11,454,876	6,032,684	10.0	15.7	11.3	11.5	44.02
P23 plasma DNA§	46XY	10	1.6	258	2.7	11,851,612	6,642,795	5.3	12.2	9.6	–	43.80
P26 plasma DNA§	46XY	13	3.0	340	6.7	11,471,297	3,851,477	10.3	18.2	14.2	–	42.51
P31 plasma DNA§	46XY	20	2.2	278	4.0	8,967,562	4,683,777	Missing data‡	13.2	17.0	–	48.27
P40 plasma DNA§	46XY	11	2.6	217	3.7	9,205,197	4,187,561	8.6	20.0	17.1	–	42.65
P42 plasma DNA§	46XY	11	3.0	276	5.5	8,364,774	4,315,527	<4.4‡	9.7	7.9	–	44.14
P52 plasma DNA§	47XY+21	25	1.6	645	6.8	9,192,596	5,126,837	6.3	25.0	26.3	26.4	44.34
P53 plasma DNA§	47XX+21	19	1.6	539	5.7	9,771,887	5,434,222	–	–	–	25.8	44.18
P57 plasma DNA§	47XX+18	23	2.0	199	2.6	15,041,417	7,470,487	–	–	–	23.0	42.89
P59 plasma DNA§	47XY+18	21	2.0	426	5.6	11,910,483	6,684,871	26.4	44.0	39.8	45.1	43.64
P64 plasma DNA§	47XY+13	17	1.8	204	2.4	12,097,478	6,701,148	<4.4‡	14.0	8.9	16.7	44.21
Male donor plasma DNA§	–	–	1.8	485	5.8	6,669,125	3,692,931	–	–	–	–	48.30
Male donor whole-blood genomic DNA§	–	–	–	–	2.1	8,519,495	5,085,412	–	–	–	–	46.53
P25 plasma DNA¶	46XY	11	5.6	132	4.9	242,599	144,992†	–	–	–	–	41.38
P13 plasma DNA§	46XY	18	5.6	77	2.9	4,168,455	2,835,333	9.8	5.7	n/a‖	–	39.60

n/a, not applicable.

*As quantified by digital PCR with EIF2C1 TaqMan Assay, converting from copies to nanograms assuming 6.6 pg per cell equivalent.

†For 454 sequencing, this number represents the number of reads with at least 90% accuracy and 90% coverage when mapped to hg18.

‡Insufficient materials were available for quantifying fetal DNA percentage with digital PCR for these samples (either no samples remained for analysis or there was insufficient sampling).

§Sequenced on Solexa/Illumina platform.

¶Sequenced on 454/Roche platform.

‖Sample P13 was the first to be analyzed by shotgun sequencing. It was a normal fetus, and the chromosome value was clearly disomic. However, there were some irregularities with this sample, and it was not included in further analysis. This sample was sequenced on a different Solexa instrument than the rest of the samples of this study, and it was sequenced in the presence of a number of samples of unknown origin. The phasing values, prephasing values, and frequency cross-talk matrices used for base calling were estimated automatically for the run this sample was sequenced in, whereas the corresponding values for the other runs were estimated from the phiX control lane. The GC content of this sample was lower than the GC bias of the human genome, whereas the rest of the samples are above. It had the lowest number of reads and also the smallest number of reads mapped successfully to the human genome. This sample appeared to be outlier in sequence tag density for most chromosomes, and the fetal DNA fraction calculated from chromosomes X was not well defined. For these reasons, we suspect that the irregularities are due to technical problems with the sequencing process.

Other Supporting Information Files

SI Appendix (PDF)