

## Supplementary Information for

“Maternal CNVs underlie false-positives in prenatal aneuploidy testing”

Matthew W. Snyder\*<sup>1</sup>, LaVone E. Simmons\*<sup>2</sup>, Jacob O. Kitzman<sup>†1</sup>, Bradley P. Coe<sup>1</sup>,  
Jessica M. Henson<sup>2</sup>, Riza M. Daza<sup>1</sup>, Evan E. Eichler<sup>1,3</sup>, Jay Shendure<sup>#1</sup>, Hilary S.  
Gammill<sup>#2,4</sup>

<sup>1</sup>Department of Genome Sciences, University of Washington, Seattle, WA

<sup>2</sup>Division of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology,  
University of Washington, Seattle, WA

<sup>3</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA

<sup>4</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

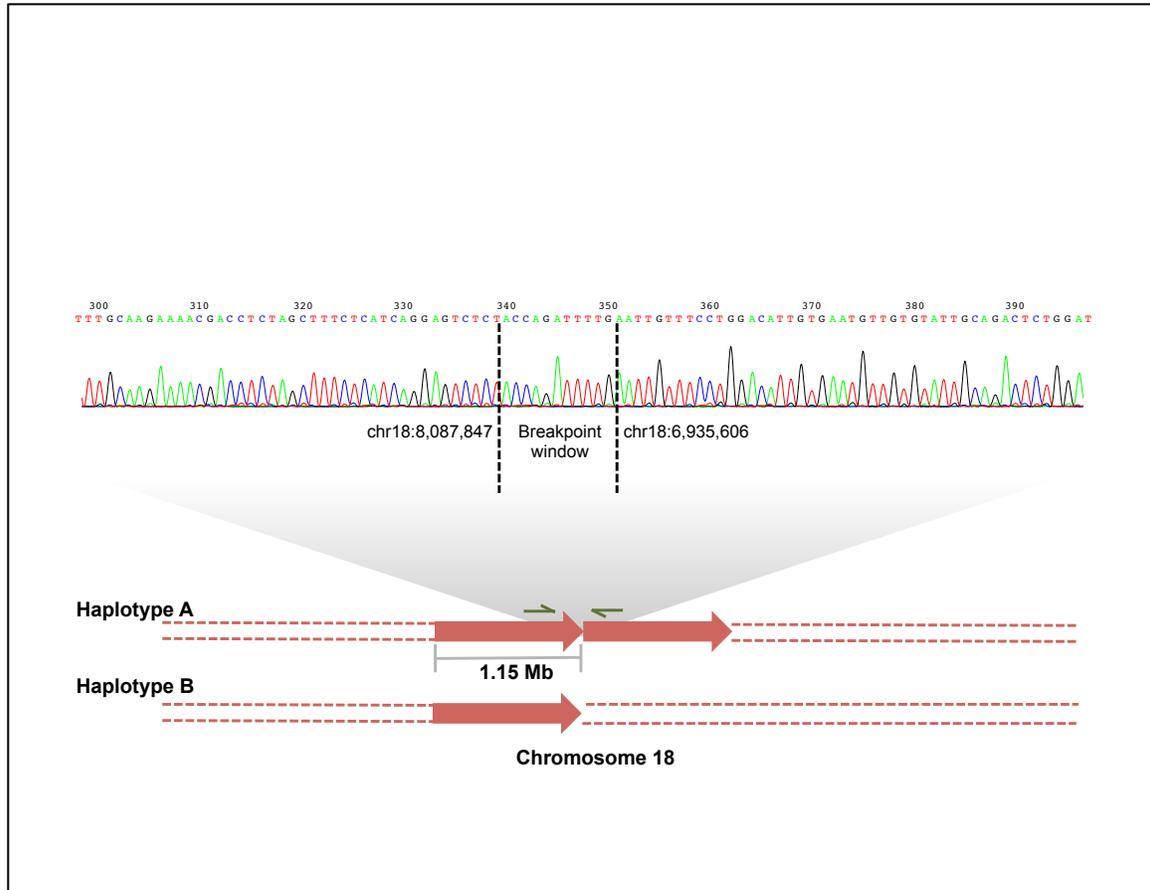
\*These authors contributed equally to this work

#To whom correspondence should be addressed

†Current affiliation: Department of Human Genetics and Department of Computational  
Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI

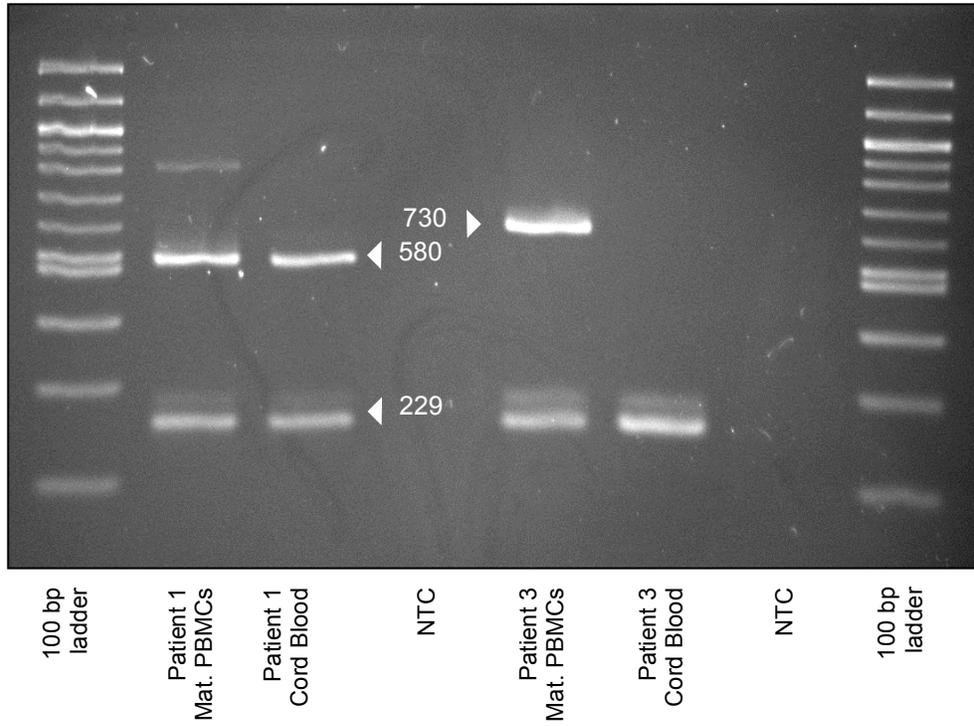
## Contents

Figure S1:	Structure and validation of maternal duplication in Patient 1	2
Figure S2:	Structure and validation of maternal duplication in Patient 3	3
Figure S3:	Validation of CNVs with multiplex PCR	4
Figure S4:	Population frequency and estimated impact on false-positive test rates of maternal CNVs with less stringent filtering	5
Table S1:	Calculation of standard deviations in number of reads derived from chromosomes 13, 18, and 21	6
Table S2:	Control CNV Cohorts and Call Sources	7
Table S3:	Demographic and Pregnancy Characteristics of Study Subjects	8
Table S4:	Breakpoints and PCR primers for detection of maternal CNVs	9
Table S5:	The role of CNV size and fetal fraction in false-positive NIPT results	10
Supplementary Methods		12
Supplementary References		14

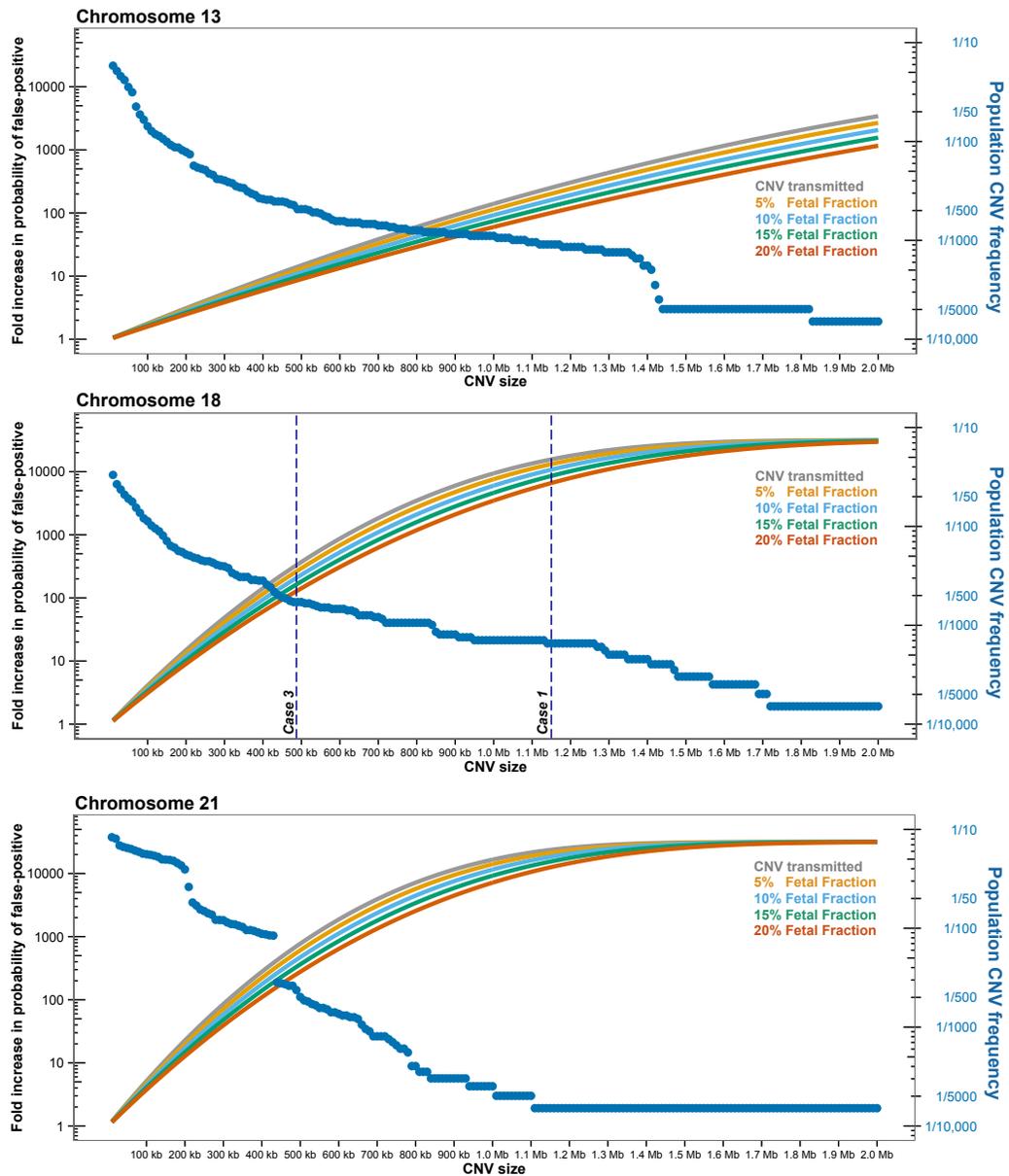


**Figure S1. Structure and validation of maternal duplication in Patient 1.** A schematic of 18p11.31 and 18p11.23 is shown at bottom, with red block arrows representing copies of the 1.15Mb duplicated region. Green arrows represent placement of primers for PCR and Sanger sequencing of the breakpoint window (meaning the window to which the breakpoint has been narrowed; the precise breakpoint cannot be determined with single nucleotide resolution due to microhomology). A partial Sanger sequencing trace of the PCR amplicon using the forward primer is shown.





**Figure S3. Validation of CNVs with multiplex PCR.** PCR primers were designed to yield a product of the expected size in the event of a tandem duplication in the region of interest. Patient 1 demonstrated the expected ~730 bp product in maternal PBMC DNA; the infant inherited the CNV. Patient 3 demonstrated the expected ~580 bp product in maternal PBMC DNA; the infant did not inherit the CNV. Additional PCR primers for chromosome 18 were designed to yield a 229 bp product for all samples as a positive control. PCR products were purified and Sanger sequenced for breakpoint confirmation (**Figures S1 & S2**). NTC: no-template control.



**Figure S4. Population frequency and estimated impact on false-positive test rates of maternal CNVs with less stringent filtering.** The burden of non-pathogenic CNVs on chromosomes 13, 18, and 21 in a cohort of 19,584 individuals predominantly of European ancestry is displayed for a range of CNV sizes (blue, right vertical axis), as in Figure 2. Here, unlike in Figure 2, reference panel CNVs were not filtered for overlap with unique genomic regions. This less conservative approach accounts for the uncertainty of reference panel CNV breakpoints and unspecified coordinates of unique genomic regions used by NIPT providers. CNV frequencies in each size bin refer to CNVs of the given size or larger. For each size bin, the estimated fold increase of the probability of a false-positive test is shown for a range of fetal fractions (gray and colored lines, left vertical axis). The sizes of the CNVs carried by Patient 1 and Patient 3 are highlighted (top, dashed vertical lines).

Uniquely mapped reads	Chromosome	CV	Uniquely mappable sequence (bp)	Proportion on this chrom of all mappable sequence	Mean number of reads from pool of 26.2M mapped	1 SD in diploid samples
26,200,000	13	0.0045	80,382,000	0.03543	928,190	4176.86
	18	0.0023	62,670,000	0.02762	723,644	1664.38
	21	0.0044	27,966,000	0.01233	322,929	1420.89

**Table S1. Calculation of standard deviations (SD) in number of reads derived from chromosomes 13, 18, and 21.** From a fixed number of uniquely mapped reads, the mean number of reads derived from each chromosome is calculated based on the amount of unique sequence on that chromosome. In combination with this mean, the coefficient of variation (CV) for each chromosome (ref. 19) determines the standard deviation in truly diploid pregnancies.

Cohort	Array Platform	Number of Samples	Description	Raw Data Source	CNV Call Source
<b>HGDP</b>	HumanHap650Yv3_A	983	The HGDP consists of 1064 individuals sampled from 51 different world populations. N=983 after sample quality control.	PMID:18292342	dbVar: nstd54
<b>NINDS (Coriell 550K)</b>	HumanHap550v3_A	441	Genotype data from NINDS were derived from two sets of neurological disease controls totaling 790 people and consist of individuals of European descent with no family history of or any first-degree relative with amyotrophic lateral sclerosis, ataxia, autism, brain aneurysm, dystonia, Parkinson disease, or schizophrenia.	dbGaP Accession: phs000089	dbVar: nstd54
<b>NINDS (317K+240K)</b>	Illumina 317K+240K	227	Genotype data from NINDS were derived from two sets of neurological disease controls totaling 790 people and consist of individuals of European descent with no family history of or any first-degree relative with amyotrophic lateral sclerosis, ataxia, autism, brain aneurysm, dystonia, Parkinson disease, or schizophrenia.	dbGaP Accession: phs000089	dbVar: nstd54
<b>PARC (CAP and PRINCE)</b>	Illumina 317K	936	The PARC samples are a subset of the cohorts used in two statin trials, CAP and PRINCE and consist of 960 middle-age (40-70 years) individuals of European descent living in the United States with moderately high levels of total cholesterol.	PMIDs: 11434828, 16516587	dbVar: nstd54
<b>London (Parents)</b>	Illumina 550K	760	The London samples represent parents of asthmatic children from Mexico City.	PMID: 19714205	dbVar: nstd54
<b>PARC2 (CAP2)</b>	Human610-Quadv1_B	232	The PARC samples are a subset of the cohorts used in two statin trials, CAP2 and PRINCE2, and consist of middle-age (40-70 years) individuals of European descent living in the United States with moderately high levels of total cholesterol	PMIDs: 11434828, 16516587	dbVar: nstd54
<b>PARC2 (PRINCE2)</b>	Illumina 610K Quad	534	The PARC samples are a subset of the cohorts used in two statin trials, CAP2 and PRINCE2, and consist of middle-age (40-70 years) individuals of European descent living in the United States with moderately high levels of total cholesterol.	PMIDs: 11434828, 16516587	dbVar: nstd54
<b>FHCRC</b>	Human610-Quadv1_B	1430	The FHCRC set are part of an ongoing Genome-wide Association Study to Identify Genetic Components of Hip Fracture in the Women's Health Initiative. Samples represent post-menopausal (50-79 years) female controls for pancreatic cancer, colon cancer, and cases and controls for a hip fracture study.	FHCRC	dbVar: nstd54
<b>inChianti</b>	HumanHap550v3_a	695	Population-based study of older persons living in the Chianti geographic area.	<a href="http://www.inchiantistudy.net/">http://www.inchiantistudy.net/</a>	dbVar: nstd54
<b>WTCCC2(NBS)</b>	Custom Illumina 1.2M	2090	UK Blood Service Control Group (blood donors, age range 18-69 years). Custom Illumina 1.2M Data.	<a href="http://www.wtccc.org.uk/">http://www.wtccc.org.uk/</a>	dbVar: nstd54
<b>ARIC</b>	SNP6	8733	The Atherosclerosis Risk in Communities (ARIC) Cohort Component samples are from a prospective epidemiologic study conducted in four U.S. communities, designed to investigate the etiology and natural history of atherosclerosis, the etiology of clinical atherosclerotic diseases, and variation in cardiovascular risk factors, medical care and disease by race, gender, location, and date.	dbGaP Accession: phs000090	Affymetrix GTC 4.1 + Filtering
<b>WTCCC2(58C)</b>	SNP6	2523	1958 British Birth Cohort	<a href="http://www.wtccc.org.uk/">http://www.wtccc.org.uk/</a>	Affymetrix GTC 4.1 + Filtering

**Table S2. Control CNV cohorts and call sources.**

	Maternal Age	Ethnicity	Gravidity and Parity*	BMI	Indication for cfDNA testing	Gestational age of cfDNA testing and result	Diagnostic Testing and Clinical Outcome†
<b>Patient 1</b>	36	Caucasian	G1P0	28	Maternal age 35 years or older at delivery	18 weeks: Trisomy 18	20 week ultrasound: normal fetal anatomy. Amniocentesis at 20 weeks: 46,XY. Term delivery, normal newborn exam.
<b>Patient 2</b>	25	Caucasian	G2P1	65	Fetal ultrasonographic findings indicating an increased risk of aneuploidy	25 weeks: Trisomy 18	20 week ultrasound: Echogenic intracardiac focus, otherwise normal fetal anatomy. Declined amniocentesis. 31 and 36 week ultrasounds with normal interval growth. Term delivery, normal newborn exam. No infant karyotype performed.
<b>Patient 3</b>	34	Caucasian	G2P0	25	Maternal age 35 years or older at delivery	12 weeks: Trisomy 18	12 week, 16 week and 20 week ultrasounds with normal fetal anatomy. Declined amniocentesis. Term delivery, normal newborn exam. No infant karyotype performed.
<b>Patient 4</b>	38	Caucasian	G4P0	24	Maternal age 35 years or older at delivery	12 weeks: Trisomy 13	13 week, 16 week and 20 week ultrasounds with normal fetal anatomy. Amniocentesis at 16 weeks: 46,XX. 29 week ultrasound with normal growth. Term delivery, normal newborn exam.

**Table S3. Demographic and Pregnancy Characteristics of Study Subjects.**

\*Gravidity and Parity status at time of study enrollment.

† Note: though classification of neonatal outcomes according to newborn examinations has been traditionally accepted<sup>5</sup> and was considered a normal outcome in our study, a newborn examination does have limitations. Specifically, a normal newborn examination rules out a diagnosis of Trisomy 18 but does not provide an evaluation for the presence of a small supernumerary marker chromosome, which could be present in the setting of a normal phenotype<sup>22,23</sup>.

	Breakpoints	Band	CNV Size	PCR primers
<b>Patient 1</b>	Chr 18: 6,935,598 - 8,087,852	18p11.31 – 18p11.23	1.15 Mb	5'-TGACCACTTTCAGCATGCCA-3' 5'-GCTTGGAAGAAGACTCAGTGGA-3'
<b>Patient 3</b>	Chr 18: 6,351,540 -6,839,310	18p11.31	487 kb	5'-AGGGACTTTCTACTTGAGAAGCA-3' 5'-CCTTCTTGGCAGGGGAAAT-3'
<b>Figure 1 amplification positive control</b>		n /a	n/a	5'- TCGAAGTGTGCTTTCCTGA-3' 5'- ACATTTTCCAGAGGCCGACA-3'

**Table S4. Breakpoints and PCR primers for detection of maternal CNVs.** Breakpoints are in hg19 coordinates. The chromosome 18 amplification positive control, used in Figure S3, is expected to produce a 229 bp amplicon.

Chrom.	False-positive rate †	CNV transmitted			CNV not transmitted		
		Z score increase (Fold change in FP)	CNV size	Population prevalence	Fetal Fraction	CNV size	Population prevalence
13	3 / 1,914 (0.16%)	1.0 (43-fold)	723 kb	8 / 19,584 (0.041%)	5%	762 kb	7 / 19,584 (0.036%)
					10%	804 kb	7 / 19,584 (0.036%)
					15%	851 kb	7 / 19,584 (0.036%)
					20%	904 kb	6 / 19,584 (0.031%)
		2.0 (718-fold)	1.45 Mb	1 / 19,584 (0.005%)	5%	1.52 Mb	1 / 19,584 (0.005%)
					10%	1.61 Mb	1 / 19,584 (0.005%)
					15%	1.70 Mb	1 / 19,584 (0.005%)
					20%	1.81 Mb	1 / 19,584 (0.005%)
18	3 / 1,905 (0.16%)	1.0 (43-fold)	288 kb	58 / 19,584 (0.296%)	5%	303 kb	58 / 19,584 (0.296%)
					10%	320 kb	51 / 19,584 (0.260%)
					15%	339 kb	45 / 19,548 (0.230%)
					20%	360 kb	45 / 19,548 (0.230%)
		2.0 (718-fold)	577 kb	14 / 19,548 (0.071%)	5%	607 kb	13 / 19,548 (0.066%)
					10%	641 kb	10 / 19,584 (0.051%)
					15%	678 kb	9 / 19,548 (0.046%)
					20%	721 kb	5 / 19,584 (0.026%)
21	6 / 1,909 (0.31%)	1.0 (43-fold)	246 kb	24 / 19,584 (0.123%)	5%	259 kb	22 / 19,584 (0.112%)
					10%	273 kb	17 / 19,584 (0.087%)
					15%	290 kb	16 / 19,584 (0.082%)
					20%	308 kb	15 / 19,584 (0.077%)
		2.0 (718-fold)	492 kb	8 / 19,584 (0.041%)	5%	518 kb	5 / 19,584 (0.026%)
					10%	547 kb	4 / 19,584 (0.020%)
					15%	579 kb	3 / 19,584 (0.015%)
					20%	615 kb	3 / 19,584 (0.015%)

**Table S5. The role of CNV size and fetal fraction in false-positive NIPT results.** Z score increases refer to the number of normalized units the sampling distribution is shifted relative to the underlying reference distribution by a maternal CNV. The CNV size depends on the transmission of the CNV and, for CNVs that are not transmitted, on the fetal fraction of cfDNA in

the maternal plasma. The population burden of such CNVs, as estimated from a cohort of 19,584 controls of predominantly European ancestry, is given for each fetal fraction. All calculations assume that the pregnancy is diploid, and are based on 26.2 million reads per sample. FP: false-positive NIPT result for the given chromosome.

<sup>†</sup>Estimated false-positive rates as reported by Bianchi and colleagues<sup>5</sup>. The rate for chromosome 13 includes the 899 patients for whom standard screening was available as well as the 1,015 patients for whom only cfDNA-based results were available.

## **Supplementary Methods**

### *Maternal plasma library preparation and sequencing*

Maternal plasma was collected by standard methods and stored in 1 mL aliquots at -80°C until use. Circulating cfDNA was purified from plasma with the QIAamp Circulating Nucleic Acid kit (Qiagen). DNA yield was measured with a Qubit fluorometer (Invitrogen). Sequencing libraries were prepared with the ThruPLEX-FD kit (Rubicon Genomics). Library amplification was monitored by real-time PCR to avoid over-amplification. All libraries were sequenced on HiSeq 2000 instruments (Illumina) using paired-end 101 bp reads with an index read of 9 bp.

### *Alignment*

Reads were mapped to the 1000 Genomes human reference genome sequence including decoy sequences (hs37d5) with BWA v0.7.3a<sup>24</sup>. PCR duplicate read pairs were removed using the Picard toolkit (<http://picard.sourceforge.net/>).

### *Identification of large CNVs*

Read depth from shotgun sequencing of maternal cfDNA was calculated in non-overlapping genomic windows of varying sizes, each containing 10,000 singly unique k-mers (SUNKs), as described previously<sup>25</sup>. GC-correction was applied separately to each window. Candidate large CNVs were first identified by visual inspection of read-depth profiles on relevant chromosomes and validated by PCR and Sanger sequencing.

### *Modeling: Fold increase in probability of false-positive results and population CNV burden*

The Illumina Verifi® test (from which NIPT results for all four subjects were derived) reportedly uses an average of 26.2 million reads per sample<sup>16</sup> and retains only reads uniquely assignable to a single genomic origin. Per-chromosome coefficients of variation<sup>19</sup> were used to estimate the standard deviation of the number of reads mapping to each chromosome  $i$  using the formula

$$CV_i = \frac{sd_i}{\mu_i}$$

where  $\mu_i$  was estimated from the average total number of reads (26.2M) for each sample and the total length of uniquely mappable sequence on each chromosome (**Table S4**). Given  $sd_i$ , the minimum number of additional reads required to reach a  $k$  unit shift in z-score (i.e.,  $k$  standard deviations above the mean) is  $k \times sd_i$ . For a given copy-number gain (i.e., duplication of sequence) of size  $s$  on autosomal chromosome  $i$ ,  $C_{s,i}$  represents the proportion of unique sequence on  $i$  duplicated by the CNV. The number of additional reads expected to result from this duplication is

$$\mu_i \times C_{s,i} \div 2 \times [1.0 - (FF \times I)]$$

where  $FF$  represents the fetal fraction of the sample and  $I$  is an indicator variable equal to 0 if the CNV is inherited by the fetus and 1 if the CNV is not inherited. The CNV size required to reach an expected z-score of  $k$  for a given autosome can be calculated by setting this expression equal to  $k \times sd_i$  and solving this equation for  $C_{s,i}$ . **Table S3** provides details of this calculation for representative fetal fractions.

Again given  $sd_i$ , the minimum deficit of reads required to reach a  $-k$  unit shift in z-score (i.e.,  $k$  standard deviations below the mean) is  $k \times sd_i$ . For a given copy-number loss (i.e., deletion of sequence) of size  $s$  on the X chromosome,  $C_{s,X}$  represents the proportion of unique sequence on the X chromosome deleted by the CNV. The deficit of reads expected to result from this deletion is

$$\mu_X \times C_{s,X} \div 2 \times [1.0 - (FF \times I)]$$

if the fetal karyotype is 46,XX, where  $FF$  represents the fetal fraction of the sample and  $I$  is an indicator variable equal to 0 if the CNV is inherited by the fetus and 1 if the CNV is not inherited. The CNV size, in this case a deletion of sequence, required to reach an expected z-score of  $-k$  for the X chromosome can be calculated by setting this expression equal to  $k \times sd_i$  and solving this equation for  $C_{s,X}$ . If the fetal karyotype is 46,XY, the expectation of the number of reads derived

from the X chromosome may be adjusted downward by an additional factor of  $[1 - (0.5 \times FF)]$  to account for the different sex chromosome compositions of the maternal and fetal compartments of the cfDNA. This calculation assumes that the presence of absence of reads derived from the Y chromosome is not relevant to the z-score calculation or inference of fetal karyotype. **Table S4** provides details of this calculation.

The fold increase in the probability of false-positive results was determined for each CNV size by first finding the value of  $k$  associated with a CNV of that size on a particular chromosome (as detailed above), and then calculating

$$\frac{\Pr(Z > (4.0 - k))}{\Pr(Z > 4.0)}$$

for  $0 < k < 4$ .

### Supplementary References

22. Eckmann-Scholz C, Tonnie H, Liehr T, et al. Normal prenatal ultrasound findings in a case with de novo mosaic small supernumerary marker chromosome 18 - how to counsel? *J Matern Fetal Neonatal Med* 2012;25:200–202.
23. Marle N, Martinet D, Aboura A, et al. Molecular characterization of 39 de novo sSMC: contribution to prognosis and genetic counselling, a prospective study. *Clin Genet* 2014;85:233-244.
24. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009;25:1754-60.
25. Sudmant PH, Kitzman JO, Antonacci F, et al. Diversity of human copy number variation and multicopy genes. *Science* 2010;330(6004):641–646.