



(86) Date de dépôt PCT/PCT Filing Date: 2010/12/01
(87) Date publication PCT/PCT Publication Date: 2011/07/28
(45) Date de délivrance/Issue Date: 2018/08/21
(85) Entrée phase nationale/National Entry: 2012/07/04
(86) N° demande PCT/PCT Application No.: US 2010/058612
(87) N° publication PCT/PCT Publication No.: 2011/090558
(30) Priorités/Priorities: 2010/01/19 (US61/296,358);
2010/07/01 (US61/360,837); 2010/10/26 (US61/407,017);
2010/10/26 (US61/455,849)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2018.01),
C12N 15/10 (2006.01), *C12P 19/34* (2006.01),
C12Q 1/6806 (2018.01), *C12Q 1/6809* (2018.01),
C12Q 1/6869 (2018.01), *C40B 30/00* (2006.01),
C40B 40/06 (2006.01), *G06F 19/20* (2011.01)

(72) Inventeurs/Inventors:
QUAKE, STEPHEN, US;
RAVA, RICHARD P., US;
CHINNAPPA, MANJULA, US;
COMSTOCK, DAVID A., US;
HEILEK, GABRIELLE, US

(73) Propriétaire/Owner:
VERINATA HEALTH, INC., US

(74) Agent: BORDEN LADNER GERVAIS LLP

(54) Titre : DETERMINATION SIMULTANEE DE L'ANEUPLOIDIE ET DE LA FRACTION FfTALE
(54) Title: SIMULTANEOUS DETERMINATION OF ANEUPLOIDY AND FETAL FRACTION

(57) **Abrégé/Abstract:**

The invention provides compositions and methods for simultaneously determining the presence or absence of fetal aneuploidy and the relative amount of fetal nucleic acids in a sample obtained from a pregnant female. The method encompasses the use of sequencing technologies and exploits the occurrence of polymorphisms to provide a streamlined noninvasive process applicable to the practice of prenatal diagnostics.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
28 July 2011 (28.07.2011)

PCT

(10) International Publication Number
WO 2011/090558 A1(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12P 19/34* (2006.01)(74) Agents: **SKUBATCH, Maya** et al.; Wilson Sonsini
Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA
94304-1050 (US).(21) International Application Number:
PCT/US2010/058612(22) International Filing Date:
1 December 2010 (01.12.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/296,358 19 January 2010 (19.01.2010) US
61/360,837 1 July 2010 (01.07.2010) US
61/455,849 26 October 2010 (26.10.2010) US
61/407,017 26 October 2010 (26.10.2010) US(71) Applicant (for all designated States except US): **VERI-
NATA HEALTH, INC.** [US/US]; 1531 Industrial Road,
San Carlos, CA 94070 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **QUAKE, Stephen**
[US/US]; 636 Alvarado Row, Redwood City, CA 94305
(US). **RAVA, Richard, P.** [US/US]; 711 Edgewood
Road, Redwood City, CA 94062 (US). **CHINNAPPA,
Manjula** [US/US]; 702 Bounty Drive, Apt. 207, Foster
City, CA 94404 (US). **COMSTOCK, David, A.**
[US/US]; 965 East El Camino Real #938, Sunnyvale, CA
94087 (US). **HEILEK, Gabrielle** [US/US]; 227 Ada Av-
enue #E, Mountain View, CA 94043 (US).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG,
ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))

(54) Title: SIMULTANEOUS DETERMINATION OF ANEUPLOIDY AND FETAL FRACTION

(57) Abstract: The invention provides compositions and methods for simultaneously determining the presence or absence of fetal aneuploidy and the relative amount of fetal nucleic acids in a sample obtained from a pregnant female. The method encompasses the use of sequencing technologies and exploits the occurrence of polymorphisms to provide a streamlined noninvasive process applicable to the practice of prenatal diagnostics.



WO 2011/090558 A1

SIMULTANEOUS DETERMINATION OF ANEUPLOIDY AND FETAL FRACTION

FIELD OF THE INVENTION

[0002] The invention relates generally to the field of diagnostics, and provides a method that is applicable to the practice of noninvasive prenatal diagnostics.

BACKGROUND OF THE INVENTION

[0003] Prenatal diagnosis to determine potential fetal abnormalities provides an opportunity for necessary care and management during pregnancy, the neonatal period and delivery. Imaging techniques such as ultrasonography, magnetic resonance imaging and fetal echocardiography are useful for the identification of structural abnormalities of the fetus. Amniocentesis, chronic villus sampling and fetal blood sampling provide fetal cells and tissues for the analysis of chromosomal, genetic and biochemical abnormalities, but are invasive and pose great risk to the pregnancy.

[0004] The existence of circulating cell-free DNA in maternal blood (Lo *et al.*, *Lancet* 350:485-487 [1997]) is being exploited for developing noninvasive processes that use fetal nucleic acids from a maternal peripheral blood sample to determine fetal chromosomal abnormalities (Fan HC and Quake SR *Anal Chem* 79:7576-7579 [2007]; Fan *et al.*, *Proc Natl Acad Sci* 105:16266-16271 [2008]).

[0005] These methods provide a paradigm shift in prenatal diagnosis, as they could effectively pronounce the end of invasive procedures. However, the sensitivity of fetal aneuploidy determination largely depends on the fetal DNA fraction, which has been determined to be <10% of the total circulating cell-free DNA (cfDNA) (Lo *et al.*, *Am J Hum Genet* 62:768-775 [1998]). Given the relatively low concentration of fetal circulating nucleic acids, false negative results can arise if there is insufficient starting nucleic acid for analysis. Accordingly, assays for the noninvasive determination of fetal DNA fraction have been developed, but typically rely on comparing the amount of fetal-specific locus (such as the SRY locus on chromosome Y in male pregnancies) to that of a locus on any autosome that is common to both the mother and the fetus (Dahllan *et al.*, *Lancet* 369:474-481 [2007]; Li *et al.*, *Clin Chem* 1002-1011 [2004]; Fan *et al.*, *Proc Natl Acad Sci* 105:16266-16271 [2008]). In addition, the assays used for quantifying fetal fraction are performed independently of the assays being developed for determining the presence or absence of aneuploidies in circulating cfDNA.

[0006] Therefore, it would be desirable to provide a prenatal test that affords an internal control to measure the adequacy of input fetal nucleic acids and avoid incorrect diagnoses of fetal chromosomal abnormalities.

The present invention provides compositions and methods that enable the simultaneous determination of fetal fraction and the determination of the presence or absence of aneuploidy from a single diagnostic sequencing

process. The method allows for determining fetal fraction in a gender-independent manner, which relies on quantification of alleles on multiple chromosomes. The noninvasive diagnostic method encompasses the use of next generation sequencing (NGS) technology that can be implemented in a streamlined and cost-effective process to provide noninvasive prenatal diagnoses of fetal aneuploidies with greater confidence.

SUMMARY OF THE INVENTION

[0008] The invention provides compositions and methods for simultaneously determining the presence or absence of fetal aneuploidy and the relative amount of fetal nucleic acids in a sample obtained from a pregnant female. The method encompasses the use of sequencing technologies and exploits the occurrence of polymorphisms to provide a streamlined noninvasive process applicable to the practice of prenatal diagnostics.

[0009] In one embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy.

[0010] In another embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; wherein enriching comprises amplifying a plurality of polymorphic target nucleic acids in a portion of said mixture; (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy.

[0011] In another embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; wherein enriching comprises amplifying a plurality of polymorphic target nucleic acids in a portion of a purified mixture of fetal and maternal nucleic acids; (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy.

[0012] In another embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; wherein enriching comprises combining at least a portion of a first sequencing library of said mixture of fetal and maternal nucleic acid molecules with at least a portion of a second sequencing library of amplified polymorphic target nucleic acids; (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy.

[0013] In one embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; wherein each of the plurality of

polymorphic target nucleic acids comprises at least one single nucleotide polymorphism (SNP); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. In some embodiments, the at least one SNP, is a single SNP selected from each of said plurality of polymorphic target nucleic acids comprises a SNP selected from rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. Alternatively, the at least one SNP is a set of two tandem SNPs selected from sets rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989- rs13047336; rs987980-rs987981; rs4143392- rs4143391; rs1691324- rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102- rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959 –rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672.

[0014] In one embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; wherein each of the plurality of polymorphic target nucleic acids comprises at least one single nucleotide polymorphism (SNP); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. The step of enriching comprises amplifying a plurality of polymorphic target nucleic acids in a portion of said mixture. In some embodiments, the at least one SNP, is a single SNP selected from each of said plurality of polymorphic target nucleic acids comprises a SNP selected from rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. Alternatively, the at least one SNP is a set of two tandem SNPs selected from sets rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989- rs13047336; rs987980- rs987981; rs4143392- rs4143391; rs1691324- rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102- rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959 –rs9980934; rs2833734-

rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672.

[0015] In one embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; wherein each of the plurality of polymorphic target nucleic acids comprises at least one single nucleotide polymorphism (SNP); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. The step of enriching comprises amplifying a plurality of polymorphic target nucleic acids in a portion of a purified mixture of fetal and maternal nucleic acids. In some embodiments, the at least one SNP, is a single SNP selected from rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. Alternatively, the at least one SNP is a set of two tandem SNPs selected from sets rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672.

[0016] In one embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; wherein each of the plurality of polymorphic target nucleic acids comprises at least one single nucleotide polymorphism (SNP); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. The step of enriching comprises combining at least a portion of a first sequencing library of said mixture of fetal and maternal nucleic acid molecules with at least a portion of a second sequencing library of amplified polymorphic target nucleic acids. In some embodiments, the at least one SNP, is a single SNP selected from each of said plurality of polymorphic target nucleic acids comprises a SNP selected from rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684,

rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. Alternatively, the at least one SNP is a set of two tandem SNPs selected from sets rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672.

[0017] In another embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids, wherein each of the plurality of polymorphic target nucleic acids comprises at least one short tandem repeat (STR); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. In some embodiments, the at least one STR is less than about 200 base pairs. In other embodiments, each of said plurality of polymorphic target nucleic acids comprises an STR selected from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 and D1GATA113.

[0018] In another embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids, wherein each of the plurality of polymorphic target nucleic acids comprises at least one short tandem repeat (STR); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. The step of enriching comprises amplifying a plurality of polymorphic target nucleic acids in a portion of the mixture. In some embodiments, the at least one STR is less than about 200 base pairs. In other embodiments, each of said plurality of polymorphic target nucleic acids comprises an STR selected from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 and D1GATA113.

[0019] In another embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a)

enriching said mixture for a plurality of polymorphic target nucleic acids, wherein each of the plurality of polymorphic target nucleic acids comprises at least one short tandem repeat (STR); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. The step of enriching comprises amplifying a plurality of polymorphic target nucleic acids in a portion of a purified mixture of fetal and maternal nucleic acids. In some embodiments, the at least one STR is less than about 200 base pairs. In other embodiments, each of said plurality of polymorphic target nucleic acids comprises an STR selected from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 and D1GATA113.

[0020] In another embodiment, a method is provided for simultaneously determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, the method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids, wherein each of the plurality of polymorphic target nucleic acids comprises at least one short tandem repeat (STR); (b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein sequencing comprises providing a plurality of sequence tags; and (c) based on the sequencing, simultaneously determining the fetal fraction and the presence or absence of the fetal aneuploidy. The step of enriching comprises combining at least a portion of a first sequencing library of said mixture of fetal and maternal nucleic acid molecules with at least a portion of a second sequencing library of amplified polymorphic target nucleic acids. In some embodiments, the at least one STR is less than about 200 base pairs. In other embodiments, each of said plurality of polymorphic target nucleic acids comprises an STR selected from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 and D1GATA113.

[0021] In the embodiments of the method summarized above and described in further detail below, the maternal sample is a biological sample that can be chosen from but is not limited to blood, plasma, serum, urine and saliva. Preferably, the fetal and maternal nucleic acid molecules in the maternal sample are cell-free DNA (cfDNA) molecules. The polymorphic target nucleic acids can be on the same or on different chromosomes.

[0022] In the embodiments of the method summarized above and described in further detail below, the aneuploidy that is determined can be a chromosomal or a partial aneuploidy. In some embodiments, the aneuploidy is a chromosomal aneuploidy that is selected from trisomy 8, trisomy 13, trisomy 15, trisomy 16, trisomy 18, trisomy 21, trisomy 22, monosomy X, and XXX. In some embodiments, determining the aneuploidy comprises calculating a chromosome dose based on the number of said sequence tags for a chromosome of interest and for a normalizing chromosome, and comparing said dose to a threshold value, while determining the fetal fraction comprises identifying at least one informative polymorphic site in said enriched mixture, and calculating the fetal fraction from the amount of fetal and maternal polymorphic sites in said enriched sample.

[0023] In the embodiments of the method summarized above and described in further detail below, sequencing that can be used for the simultaneous determination is performed using next generation (NGS) sequencing. In some embodiments, sequencing is massively parallel sequencing using sequencing-by-synthesis with reversible dye

terminators. In other embodiments, sequencing is sequencing-by-ligation. In yet other embodiments, sequencing is single molecule sequencing. The sequencing of the enriched mixture can further comprise an amplification.

[0024] In another embodiment, a composition comprising at least one set of primers for amplifying at least one SNP in a maternal sample *e.g.* a plasma sample, comprising a mixture of nucleic acid molecules is provided. Nucleic acid molecules can be cfDNA molecules. In one embodiment, the composition comprises at least one set of primers for amplifying at least one SNP selected from rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. In one embodiment, the at least one set of primers is selected from primer sets of SEQ ID NOs:57-112.

[0025] In another embodiment, a composition comprising at least one set of primers for amplifying at least one STR in a maternal sample *e.g.* a plasma sample, comprising a mixture of nucleic acid molecules is provided. Nucleic acid molecules can be cfDNA molecules. In one embodiment, the composition comprises at least one set of primers for amplifying at least one STR selected from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 and D1GATA113. In one embodiment, the at least one set of STR primers is selected from primer sets of SEQ ID NOs:113-196.

[0026] In another embodiment, a kit for preparing a sequencing library for massively parallel sequencing of fetal and maternal nucleic acid molecules in a maternal sample is provided. In some embodiments, the maternal sample is a plasma sample. The kit comprises a composition comprising at least one set of primers for amplifying at least one polymorphic nucleic acid in the mixture of fetal and maternal nucleic acid molecules. The polymorphic nucleic acid sequences each comprise at least one SNP or an STR. Sequences comprising tandem SNPs are encompassed in the kit of the invention. In some embodiments, sequencing is single molecule sequencing. In some embodiments, the massively parallel sequencing is sequencing-by-synthesis with reversible dye terminators. In other embodiments, the massively parallel sequencing is sequencing-by-ligation.

[0027] Preferably, the fetal and maternal nucleic acid molecules are cfDNA molecules. In some embodiments, the maternal sample is a plasma sample. The kit comprises a composition comprising at least one set of primers for amplifying at least one polymorphic nucleic acid comprised in the fetal and maternal nucleic acid molecules. In some embodiments, the polymorphic nucleic acid comprises a SNP. In other embodiment, the polymorphic nucleic acid comprises an STR.

[0028] In one aspect, there is provided a method for simultaneous determining aneuploidy and fetal fraction in a maternal sample comprising a mixture of fetal and maternal nucleic acid molecules, said method comprising: (a) enriching said mixture for a plurality of polymorphic target nucleic acids; (b) massively parallel sequencing at least a portion of the enriched mixture obtained in step (a), wherein said sequencing comprising providing a plurality of sequence reads which are mapped to a genome to identify a plurality of sequence tags; and based on said plurality of sequence tags, simultaneous determining said fetal fraction and said aneuploidy.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0030] **Figure 1** is a flowchart of a method **100** for simultaneously determining the presence or absence of aneuploidy and the fetal fraction in a maternal test sample comprising a mixture of fetal and maternal nucleic acids.

[0031] **Figure 2** is a flowchart of a method **200** for simultaneously determining the presence or absence of fetal aneuploidy and the fetal fraction in a maternal plasma test sample enriched for polymorphic nucleic acids.

[0032] **Figure 3** is a flowchart of a method **300** for simultaneously determining the presence or absence of fetal aneuploidy and the fetal fraction in a maternal purified cfDNA test sample that has been enriched with polymorphic nucleic acids.

[0033] **Figure 4** is a flowchart of a method **400** for simultaneously determining the presence or absence of fetal aneuploidy and the fetal fraction in a sequencing library constructed from fetal and maternal nucleic acids derived from a maternal test sample and enriched with polymorphic nucleic acids.

[0034] **Figure 5** is a flowchart of a method **500** for determining the presence or absence of a copy number variation in a test sample comprising a mixture of nucleic acids.

[0035] **Figure 6** is a bar diagram showing the identification of fetal and maternal polymorphic sequences (SNPs) used to determine fetal fraction in a test sample. The total number of sequence reads (Y-axis) mapped to the SNP sequences identified by rs numbers (X-axis), and the relative level of fetal nucleic acids (*) are shown.

[0036] **Figure 7** illustrates the distribution of the chromosome dose for chromosome 21 determined from sequencing cfDNA extracted from a set of 48 blood samples obtained from human subjects pregnant with male or female fetuses. Chromosome 21 doses for qualified *i.e.* normal for chromosome 21 (O), and trisomy 21 test samples are shown (Δ) for chromosomes 1-12 and X (Figure 7A), and for chromosomes 1-22 and X (Figure 7B).

[0037] **Figure 8** illustrates the distribution of the chromosome dose for chromosome 18 determined from sequencing cfDNA extracted from a set of 48 blood samples obtained from human subjects pregnant with male or female fetuses. Chromosome 18 doses for qualified *i.e.* normal for chromosome 18 (O), and trisomy 18 (Δ) test samples are shown for chromosomes 1-12 and X (Figure 8A), and for chromosomes 1-22 and X (Figure 8B).

[0038] **Figure 9** illustrates the distribution of the chromosome dose for chromosome 13 determined from sequencing cfDNA extracted from a set of 48 blood samples obtained from human subjects pregnant with male or female fetuses. Chromosome 13 doses for qualified *i.e.* normal for chromosome 13 (O), and trisomy 13 (Δ) test samples are shown for chromosomes 1-12 and X (Figure 9A), and for chromosomes 1-22 and X (Figure 9B).

[0039] **Figure 10** illustrates the distribution of the chromosome doses for chromosome X determined from sequencing cfDNA extracted from a set of 48 test blood samples obtained from human subjects pregnant with either male or female fetuses. Chromosome X doses for males (46,XY; (O)), females (46,XX; (Δ)); monosomy X (45,X; (+)), and complex karyotypes (Cplx (X)) samples are shown for chromosomes 1-12 and X (Figure 10A), and for chromosomes 1-22 and X (Figure 10B).

[0040] **Figure 11** illustrates the distribution of the chromosome doses for chromosome Y determined from sequencing cfDNA extracted from a set of 48 test blood samples obtained from human subjects pregnant with either male or female fetuses. Chromosome Y doses for males (46,XY; (Δ)), females (46,XX; (O)); monosomy X (45,X;

(+)), and complex karyotypes (Cplx (X)) samples are shown for chromosomes 1-12 (Figure 11A), and for chromosomes 1-22 (Figure 11B).

[0041] Figure 12 shows the coefficient of variation (CV) for chromosomes 21 ($\square$), 18 (O) and 13 (Δ) that was determined from the chromosome doses of qualified *i.e.* non-affected, samples shown in Figures 7, 8, and 9, respectively.

[0042] Figure 13 shows the coefficient of variation (CV) for chromosomes X ($\square$) and Y (O) that was determined from the chromosome doses of qualified *i.e.* non-affected, samples shown in Figures 10 and 11, respectively.

DETAILED DESCRIPTION OF THE INVENTION

[0043] The invention provides compositions and methods for simultaneously determining the presence or absence of fetal aneuploidy and the relative amount of fetal nucleic acids in a sample obtained from a pregnant female. The method encompasses the use of sequencing technologies *e.g.* next generation sequencing, and exploits the occurrence of polymorphisms to provide a streamlined noninvasive process applicable to the practice of prenatal diagnostics.

[0044] Unless otherwise indicated, the practice of the present invention involves conventional techniques commonly used in molecular biology, microbiology, protein purification, protein engineering, protein and DNA sequencing, and recombinant DNA fields, which are within the skill of the art. Such techniques are known to those of skill in the art and are described in numerous standard texts and reference works.

[0045] Numeric ranges are inclusive of the numbers defining the range. It is intended that every maximum numerical limitation given throughout this specification includes every lower numerical limitation, as if such lower numerical limitations were expressly written herein. Every minimum numerical limitation given throughout this specification will include every higher numerical limitation, as if such higher numerical limitations were expressly written herein. Every numerical range given throughout this specification will include every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein.

[0046] The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the Specification as a whole. Accordingly, as indicated above, the terms defined immediately below are more fully defined by reference to the specification as a whole.

[0047] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Various scientific dictionaries that include the terms included herein are well known and available to those in the art. Although any methods and materials similar or equivalent to those described herein find use in the practice or testing of the present invention, some preferred methods and materials are described. Accordingly, the terms defined immediately below are more fully described by reference to the Specification as a whole. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art.

DEFINITIONS

[0048] As used herein, the singular terms "a", "an," and "the" include the plural reference unless the context clearly indicates otherwise. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation and amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[0049] The term “assessing” herein refers to characterizing the status of a chromosomal aneuploidy by one of three types of calls: “normal”, “affected”, and “no-call”. For example, in the presence of trisomy the “normal” call is determined by the value of a parameter *e.g.* a test chromosome dose that is below a user-defined threshold of reliability, the “affected” call is determined by a parameter *e.g.* a test chromosome dose, that is above a user-defined threshold of reliability, and the “no-call” result is determined by a parameter *e.g.* a test chromosome dose, that lies between the a user-defined thresholds of reliability for making a “normal” or an “affected” call.

[0050] The term “copy number variation” herein refers to variation in the number of copies of a nucleic acid sequence that is 1 kb or larger present in a test sample in comparison with the copy number of the nucleic acid sequence present in a qualified sample. A “copy number variant” refers to the 1 kb or larger sequence of nucleic acid in which copy-number differences are found by comparison of a sequence of interest in test sample with that present in a qualified sample. Copy number variants/variations include deletions, including microdeletions, insertions, including microinsertions, duplications, multiplications, inversions, translocations and complex multi-site variants. CNV encompass chromosomal aneuploidies and partial aneuploidies.

[0051] The term “aneuploidy” herein refers to an imbalance of genetic material caused by a loss or gain of a whole chromosome, or part of a chromosome.

[0052] The term “chromosomal aneuploidy” herein refers to an imbalance of genetic material caused by a loss or gain of a whole chromosome, and includes germline aneuploidy and mosaic aneuploidy.

[0053] The term “partial aneuploidy” herein refers to an imbalance of genetic material caused by a loss or gain of part of a chromosome *e.g.* partial monosomy and partial trisomy, and encompasses imbalances resulting from translocations, deletions and insertions.

[0054] The term “tandem SNPs” herein refers to two or more SNPs that are present within a polymorphic target nucleic acid sequence.

[0055] The terms “polymorphic target nucleic acid”, “polymorphic sequence”, “polymorphic target nucleic acid sequence” and “polymorphic nucleic acid” are used interchangeably herein to refer to a nucleic acid sequence *e.g.* a DNA sequence, that comprises one or more polymorphic sites.

[0056] The term “polymorphic site” herein refers to a single nucleotide polymorphism (SNP), a small-scale multi-base deletion or insertion, a Multi-Nucleotide Polymorphism (MNP) or a Short Tandem Repeat (STR).

[0057] The term “plurality” is used herein in reference to a number of nucleic acid molecules or sequence tags that is sufficient to identify significant differences in copy number variations (*e.g.* chromosome doses) in test samples and qualified samples using in the methods of the invention. In some embodiments, at least about 3×10^6 sequence tags, at least about 5×10^6 sequence tags, at least about 8×10^6 sequence tags, at least about 10×10^6 sequence tags, at least about 15×10^6 sequence tags, at least about 20×10^6 sequence tags, at least about 30×10^6 sequence tags, at least about 40×10^6 sequence tags, or at least about 50×10^6 sequence tags comprising between 20 and 40bp reads are obtained for each test sample.

[0058] The terms “polynucleotide”, “nucleic acid” and “nucleic acid molecules” are used interchangeably and refer to a covalently linked sequence of nucleotides (*i.e.*, ribonucleotides for RNA and deoxyribonucleotides for DNA) in which the 3' position of the pentose of one nucleotide is joined by a phosphodiester group to the 5' position of the pentose of the next, include sequences of any form of nucleic acid, including, but not limited to RNA, DNA and cfDNA molecules. The term “polynucleotide” includes, without limitation, single- and double-stranded polynucleotide.

[0059] The term “portion” when used in reference to the amount of sequence information of fetal and maternal nucleic acid molecules in a biological sample herein refers to the amount of sequence information of fetal and maternal nucleic acid molecules in a biological sample that in sum amount to less than the sequence information of <1 human genome.

[0060] The term “test sample” herein refers to a sample comprising a mixture of nucleic acids comprising at least one nucleic acid sequence whose copy number is suspected of having undergone variation. Nucleic acids present in a test sample are referred to as “test nucleic acids”.

[0061] The term “qualified sample” herein refers to a sample comprising a mixture of nucleic acids that are present in a known copy number to which the nucleic acids in a test sample are compared, and it is a sample that is normal *i.e.* not aneuploid, for the sequence of interest *e.g.* a qualified sample used for identifying a normalizing chromosome for chromosome 21 is a sample that is not a trisomy 21 sample.

[0062] The term “enrich” herein refers to the process of amplifying polymorphic target nucleic acids contained in a portion of a maternal sample, and combining the amplified product with the remainder of the maternal sample from which the portion was removed.

[0063] The term “qualified nucleic acid” is used interchangeably with “qualified sequence” is a sequence against which the amount of a test sequence or test nucleic acid is compared. A qualified sequence is one present in a biological sample preferably at a known representation *i.e.* the amount of a qualified sequence is known. A “qualified sequence of interest” is a qualified sequence for which the amount is known in a qualified sample, and is a sequence that is associated with a difference in sequence representation in an individual with a medical condition.

[0064] The term “sequence of interest” herein refers to a nucleic acid sequence that is associated with a difference in sequence representation in healthy versus diseased individuals. A sequence of interest can be a sequence on a chromosome that is misrepresented *i.e.* over- or under-represented, in a disease or genetic condition. A sequence of interest may also be a portion of a chromosome, or a chromosome. For example, a sequence of interest can be a chromosome that is over-represented in an aneuploidy condition, or a gene encoding a tumor-suppressor that is under-represented in a cancer. Sequences of interest include sequences that are over- or under- represented in the total population, or a subpopulation of cells of a subject. A “qualified sequence of interest” is a sequence of interest in a qualified sample. A “test sequence of interest” is a sequence of interest in a test sample.

[0065] The term “plurality of polymorphic target nucleic acids” herein refers to a number of nucleic acid sequences each comprising at least one polymorphic site *e.g.* one SNP, such that at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40 or more different polymorphic sites are amplified from the polymorphic target nucleic acids to identify and/or quantify fetal alleles present in maternal samples comprising fetal and maternal nucleic acids.

[0066] The term “normalizing sequence” herein refers to a sequence that displays a variability in the number of sequence tags that are mapped to it among samples and sequencing runs that best approximates that of the sequence of interest for which it is used as a normalizing parameter, and/or that can best differentiate an affected sample from one or more unaffected samples. A “normalizing chromosome” is an example of a “normalizing sequence”.

[0067] The term “differentiability” herein refers to the characteristic of a normalizing chromosome that enables to distinguish one or more unaffected *i.e.* normal, samples from one or more affected *i.e.* aneuploid, samples.

[0068] The term “group of chromosomes” herein refers to two or more chromosomes.

[0069] The term “sequence dose” herein refers to a parameter that relates the sequence tag density of a sequence of interest to the tag density of a normalizing sequence. A “test sequence dose” is a parameter that relates the sequence tag density of a sequence of interest *e.g.* chromosome 21, to that of a normalizing sequence *e.g.* chromosome 9,

determined in a test sample. Similarly, a “qualified sequence dose” is a parameter that relates the sequence tag density of a sequence of interest to that of a normalizing sequence determined in a qualified sample.

[0070] The term “sequence tag density” herein refers to the number of sequence reads that are mapped to a reference genome sequence *e.g.* the sequence tag density for chromosome 21 is the number of sequence reads generated by the sequencing method that are mapped to chromosome 21 of the reference genome. The term “sequence tag density ratio” herein refers to the ratio of the number of sequence tags that are mapped to a chromosome of the reference genome *e.g.* chromosome 21, to the length of the reference genome chromosome 21.

[0071] The term “parameter” herein refers to a numerical value that characterizes a quantitative data set and/or a numerical relationship between quantitative data sets. For example, a ratio (or function of a ratio) between the number of sequence tags mapped to a chromosome and the length of the chromosome to which the tags are mapped, is a parameter.

[0072] The terms “threshold value” and “qualified threshold value” herein refer to any number that is calculated using a qualifying data set and serves as a limit of diagnosis of a copy number variation *e.g.* an aneuploidy, in an organism. If a threshold is exceeded by results obtained from practicing the invention, a subject can be diagnosed with a copy number variation *e.g.* trisomy 21.

[0073] The term “read” refers to a DNA sequence of sufficient length (*e.g.*, at least about 30 bp) that can be used to identify a larger sequence or region, *e.g.* that can be aligned and specifically assigned to a chromosome or genomic region or gene.

[0074] The term “sequence tag” is herein used interchangeably with the term “mapped sequence tag” to refer to a sequence read that has been specifically assigned *i.e.* mapped, to a larger sequence *e.g.* a reference genome, by alignment. Mapped sequence tags are uniquely mapped to a reference genome *i.e.* they are assigned to a single location to the reference genome. Tags that can be mapped to more than one location on a reference genome *i.e.* tags that do not map uniquely, are not included in the analysis.

[0075] The terms “aligned”, “alignment”, or “aligning” refer to one or more sequences that are identified as a match in terms of the order of their nucleic acid molecules to a known sequence from a reference genome. Such alignment can be done manually or by a computer algorithm, examples including the Efficient Local Alignment of Nucleotide Data (ELAND) computer program distributed as part of the Illumina Genomics Analysis pipeline. The matching of a sequence read in aligning can be a 100% sequence match or less than 100% (non-perfect match).

[0076] The term “reference genome” refers to any particular known genome sequence, whether partial or complete, of any organism or virus which may be used to reference identified sequences from a subject. For example, a reference genome used for human subjects as well as many other organisms is found at the National Center for Biotechnology Information at www.ncbi.nlm.nih.gov. A “genome” refers to the complete genetic information of an organism or virus, expressed in nucleic acid sequences.

[0077] The term “artificial target sequences genome” herein refers to a grouping of known sequences that encompass alleles of known polymorphic sites. For example, a “SNP reference genome” is an artificial target sequences genome comprising a grouping of sequences that encompass alleles of known SNPs.

[0078] The term “clinically-relevant sequence” herein refers to a nucleic acid sequence that is known or is suspected to be associated or implicated with a genetic or disease condition. Determining the absence or presence of a clinically-relevant sequence can be useful in determining a diagnosis or confirming a diagnosis of a medical condition, or providing a prognosis for the development of a disease.

[0079] The term “derived” when used in the context of a nucleic acid or a mixture of nucleic acids, herein refers to the means whereby the nucleic acid(s) are obtained from the source from which they originate. For example, in one embodiment, a mixture of nucleic acids that is derived from two different genomes means that the nucleic acids *e.g.* cfDNA, were naturally released by cells through naturally occurring processes such as necrosis or apoptosis. In another embodiment, a mixture of nucleic acids that is derived from two different genomes means that the nucleic acids were extracted from two different types of cells from a subject.

[0080] The term “maternal sample” herein refers to a biological sample obtained from a pregnant subject *e.g.* a woman.

[0081] The term “original maternal sample” herein refers to a biological sample obtained from a pregnant subject *e.g.* a woman, who serves as the source from which a portion is removed to amplify polymorphic target nucleic acids. The “original sample” can be any sample obtained from a pregnant subject, and the processed fractions thereof *e.g.* a purified cfDNA sample extracted from a maternal plasma sample.

[0082] The term “biological fluid” herein refers to a liquid taken from a biological source and includes, for example, blood, serum, plasma, sputum, lavage fluid, cerebrospinal fluid, urine, semen, sweat, tears, saliva, and the like. As used herein, the terms “blood,” “plasma” and “serum” expressly encompass fractions or processed portions thereof. Similarly, where a sample is taken from a biopsy, swab, smear, etc., the “sample” expressly encompasses a processed fraction or portion derived from the biopsy, swab, smear, etc.

[0083] The terms “maternal nucleic acids” and “fetal nucleic acids” herein refer to the nucleic acids of a pregnant female subject and the nucleic acids of the fetus being carried by the pregnant female, respectively.

[0084] The term “corresponding to” herein refers to a nucleic acid sequence *e.g.* a gene or a chromosome, that is present in the genome of different subjects, and which does not necessarily have the same sequence in all genomes, but serves to provide the identity rather than the genetic information of a sequence of interest *e.g.* a gene or chromosome.

[0085] The term “substantially cell free” herein refers to preparations of the desired sample from which components that are normally associated with it are removed. For example, a plasma sample is rendered essentially cell free by removing blood cells *e.g.* white blood cells, which are normally associated with it. In some embodiments, substantially free samples are processed to remove cells that would otherwise contribute to the desired genetic material that is to be tested for an aneuploidy.

[0086] As used herein, the term “fetal fraction” refers to the fraction of fetal nucleic acids present in a sample comprising fetal and maternal nucleic acid.

[0087] As used herein the term “chromosome” refers to the heredity-bearing gene carrier of a living cell which is derived from chromatin and which comprises DNA and protein components (especially histones). The conventional internationally recognized individual human genome chromosome numbering system is employed herein.

[0088] As used herein, the term “polynucleotide length” refers to the absolute number of nucleic acid molecules (nucleotides) in a sequence or in a region of a reference genome. The term “chromosome length” refers to the known length of the chromosome given in base pairs *e.g.* provided in the NCBI36/hg18 assembly of the human chromosome found on the world wide web at genome.ucsc.edu/cgi-bin/hgTracks?hgid=167155613&chromInfoPage=

[0089] The term “subject” herein refers to a human subject as well as a non-human subject such as a mammal, an invertebrate, a vertebrate, a fungus, a yeast, a bacteria, and a virus. Although the examples herein concern human cells and the language is primarily directed to human concerns, the concept of this invention is applicable to

genomes from any plant or animal, and is useful in the fields of veterinary medicine, animal sciences, research laboratories and such.

[0090] The term “condition” herein refers to “medical condition” as a broad term that includes all diseases and disorders, but can include injuries and normal health situations, such as pregnancy, that might affect a person's health, benefit from medical assistance, or have implications for medical treatments.

DESCRIPTION

[0091] The method described herein is a sequencing method that enables the simultaneous determination of the fraction of the minor fetal nucleic acid component in a sample comprising a mixture of fetal and maternal nucleic acids. In particular, the method enables the determination of the fraction of cfDNA contributed by a fetus to the mixture of fetal and maternal cfDNA in a maternal sample *e.g.* a plasma sample. The difference between the maternal and fetal fraction is determined by the relative contribution of a polymorphic allele derived from the fetal genome to the contribution of the corresponding polymorphic allele derived from the maternal genome.

Polymorphic sequences can be used in conjunction with clinically-relevant diagnostic tests as a positive control for the presence of cfDNA in order to highlight false-negative or false-positive results stemming from low levels of cfDNA below the identification limit. The method described is useful across a range of gestational ages.

[0092] Exemplary embodiments of the method of the invention are illustrated in Figures 1-4 as follows.

[0093] **Figure 1** provides a flow diagram of one embodiment of method of the invention **100** for simultaneously determining a fetal aneuploidy and the fraction of fetal nucleic acids in a maternal biological sample. In step **110** a test sample comprising a mixture of fetal and maternal nucleic acids is obtained from a subject. The sample is a maternal sample that is obtained from a pregnant female, for example a pregnant woman. Any maternal biological sample can be used a source of fetal and maternal nucleic acids which are contained in cells or that are “cell-free”. In some embodiments, it is advantageous to obtain a maternal sample that comprises cell-free nucleic acids *e.g.* cfDNA. Preferably, the maternal biological sample is a biological fluid sample. A biological fluid includes, as non-limiting examples, blood, plasma, serum, sweat, tears, sputum, urine, sputum, ear flow, lymph, saliva, cerebrospinal fluid, lavages, bone marrow suspension, vaginal flow, transcervical lavage, brain fluid, ascites, milk, secretions of the respiratory, intestinal and genitourinary tracts, and leukopheresis samples. In some embodiments, the biological fluid sample is a sample that is easily obtainable by non-invasive procedures *e.g.* blood, plasma, serum, sweat, tears, sputum, urine, sputum, ear flow, and saliva. In some embodiments, the biological sample is a peripheral blood sample, or the plasma and/or the serum fractions thereof. In another embodiment, the sample is a mixture of two or more biological samples *e.g.* a biological sample can comprise two or more of a biological fluid samples. As used herein, the terms “blood,” “plasma” and “serum” expressly encompass fractions or processed portions thereof. In some embodiments, the biological sample is processed to obtain a sample fraction *e.g.* plasma, that contains the mixture of fetal and maternal nucleic acids. In some embodiments, the mixture of fetal and maternal nucleic acids is further processed from the sample fraction *e.g.* plasma, to obtain a sample comprising a purified mixture of fetal and maternal nucleic acids *e.g.* cfDNA. Cell-free nucleic acids, including cell-free DNA, can be obtained by various methods known in the art from biological samples including but not limited to plasma, serum and urine (Fan *et al.*, Proc Natl Acad Sci 105:16266-16271 [2008]; Koide *et al.*, Prenatal Diagnosis 25:604-607 [2005]; Chen *et al.*, Nature Med. 2: 1033-1035 [1996]; Lo *et al.*, Lancet 350: 485-487 [1997]). To separate cfDNA from cells, fractionation, centrifugation (*e.g.*, density gradient centrifugation), DNA-specific precipitation, or high-throughput cell sorting and/or separation methods can be used. Commercially available kits for manual and automated

separation of cfDNA are available (Roche Diagnostics, Indianapolis, IN, Qiagen, Valencia, CA, Macherey-Nagel, Duren, DE). In some instances, it can be advantageous to fragment the nucleic acid molecules in the nucleic acid sample. Fragmentation can be random, or it can be specific, as achieved, for example, using restriction endonuclease digestion. Methods for random fragmentation are well known in the art, and include, for example, limited DNase digestion, alkali treatment and physical shearing. In one embodiment, sample nucleic acids are obtained from as cfDNA, which is not subjected to fragmentation. In other embodiments, the sample nucleic acids are obtained as genomic DNA, which is subjected to fragmentation into fragments of approximately 500 or more base pairs, and to which NGS methods can be readily applied.

[0094] In step **120 (Figure 1)** the mixture of nucleic acids present in the sample is enriched for polymorphic target nucleic acids each comprising a polymorphic site. In some embodiments, the nucleic acids that are enriched are cfDNA. Target nucleic acids are segments of genetic material that are known to comprise at least one polymorphic site. In some embodiments, the target nucleic acids comprise a SNP. In other embodiments, the target nucleic acid comprises an STR. Enrichment of a mixture of fetal and maternal nucleic acids comprises amplifying target sequences from a portion of nucleic acids contained in the original maternal sample, and combining part or the entire amplified product with the remainder of the original maternal sample. In step **130**, at least a portion of the enriched mixture is sequenced, sequence differences stemming from the polymorphic nature of the target sequences are identified, and the relative contribution of polymorphic sequences derived from the fetal genome *i.e.* the fetal fraction, is determined in step **140**. In some embodiments, the original maternal sample is a biological fluid sample *e.g.* plasma. In other embodiments, the original maternal sample is a processed fraction of plasma comprising purified fetal and maternal cfDNA.

[0095] Polymorphic sites that are contained in the target nucleic acids include without limitation single nucleotide polymorphisms (SNPs), tandem SNPs, small-scale multi-base deletions or insertions, called IN-DELS (also called deletion insertion polymorphisms or DIPs), Multi-Nucleotide Polymorphisms (MNPs) and Short Tandem Repeats (STRs). The polymorphic sites that are encompassed by the method of the invention are located on autosomal chromosomes, thereby enabling the determination of fetal fraction independently of sex of the fetus. Any polymorphic site that can be encompassed by the reads generated by the sequencing methods described herein can be used to determine simultaneously the fetal fraction and the presence or absence of an aneuploidy in a maternal sample.

[0096] In one embodiment, the mixture of fetal and maternal nucleic acids in the sample is enriched for target nucleic acids that comprise at least one SNP. In some embodiments, each target nucleic acid comprises a single *i.e.* one SNP. Target nucleic acid sequences comprising SNPs are available from publicly accessible databases including, but not limited to Human SNP Database at world wide web address wi.mit.edu, NCBI dbSNP Home Page at world wide web address ncbi.nlm.nih.gov, world wide web address lifesciences.perkinelmer.com, Celera Human SNP database at world wide web address celera.com, the SNP Database of the Genome Analysis Group (GAN) at world wide web address gan.iarc.fr. In one embodiment, the SNPs chosen for enriching the fetal and maternal cfDNA are selected from the group of 92 individual identification SNPs (IISNPs) described by Pakstis *et al.* (Pakstis *et al.* Hum Genet 127:315-324 [2010]), which have been shown to have a very small variation in frequency across populations ($F_{st} < 0.06$), and to be highly informative around the world having an average heterozygosity ≥ 0.4 . SNPs that are encompassed by the method of the invention include linked and unlinked SNPs. Each target nucleic acid comprises at least one polymorphic site *e.g.* a single SNP, that differs from that present on another target nucleic acid to generate a panel of polymorphic sites *e.g.* SNPs, that contain a sufficient number of polymorphic sites of

which at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40 or more are informative. For example, a panel of SNPs can be configured to comprise at least one informative SNP. In one embodiment, the SNPs that are targeted for amplification are selected from rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022.

[0097] In other embodiments, each target nucleic acid comprises two or more SNPs *i.e.* each target nucleic acid comprises tandem SNPs. Preferably, each target nucleic acid comprises two tandem SNPs. The tandem SNPs are analyzed as a single unit as short haplotypes, and are provided herein as sets of two SNPs. To identify suitable tandem SNP sequences, the International HapMap Consortium database can be searched (The International HapMap Project, Nature 426:789-796 [2003]). The database is available on the world wide web at hapmap.org. In one embodiment, tandem SNPs that are targeted for amplification are selected from the following sets of tandem pairs of SNPS rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672.

[0098] In another embodiment, the mixture of fetal and maternal nucleic acids in the sample is enriched for target nucleic acids that comprise at least one STR. STR loci are found on almost every chromosome in the genome and may be amplified using a variety of polymerase chain reaction (PCR) primers. Tetranucleotide repeats have been preferred among forensic scientists due to their fidelity in PCR amplification, although some tri- and pentanucleotide repeats are also in use. A comprehensive listing of references, facts and sequence information on STRs, published PCR primers, common multiplex systems, and related population data are compiled in STRBase, which may be accessed via the World Wide Web at ibm4.carb.nist.gov:8800/dna/home.htm. Sequence information from GenBank® (<http://www2.ncbi.nlm.nih.gov/cgi-bin/genbank>) for commonly used STR loci is also accessible through STRBase. The polymorphic nature of tandem repeated DNA sequences that are widespread throughout the human genome have made them important genetic markers for gene mapping studies, linkage analysis, and human identity testing. Because of the high polymorphism of STRs, most individuals will be heterozygous *i.e.* most people will possess two alleles (versions) of each—one inherited from each parent—with a different number of repeats. Therefore, the non-maternally inherited fetal STR sequence will differ in the number of repeats from the maternal sequence. Amplification of these STR sequences will result in two major amplification products corresponding to the maternal alleles (and the maternally inherited fetal allele) and one minor product corresponding

to the non-maternally inherited fetal allele. This technique was first reported in 2000 (Pertl *et al.*, Human Genetics 106:45-49 [2002]) and has subsequently been developed using simultaneous identification of multiple different STR regions using real-time PCR (Liu *et al.*, Acta Obstet Gyn Scand 86:535-541 [2007]). Thus, the fraction of fetal nucleic acid in a maternal sample can also be determined by sequencing polymorphic target nucleic acids comprising STRs, which vary among individuals in the number of tandem repeated units between alleles. In one embodiment, simultaneous determination of aneuploidy and fetal fraction comprises sequencing at least a portion of fetal and maternal nucleic acids present in a maternal sample that has been enriched for polymorphic sequences comprising STRs. Given that the size of fetal cfDNA is between X and Y bp, the polymorphic sequences comprise miniSTR, which can be amplified to generate amplicons that are of lengths about the size of the circulating fetal DNA fragments. The method can use one or a combination of any number of informative miniSTRs to determine the fraction of fetal nucleic acid. For example, any one or a combination of any number of miniSTRs, for example the miniSTRs disclosed in Table 15, can be used. In one embodiment, the fraction of fetal nucleic acid in a maternal sample is performed using a method that includes determining the number of copies of the maternal and fetal nucleic acid present in the maternal sample by amplifying at least one autosomal miniSTR chosen from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, D2S1338, D1S1677, D2S441, D4S2364, D10S1248, D14S1434, D22S1045, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627, and D1GATA113. In another embodiment, the at least one autosomal miniSTR is the group of miniSTRs CSF1PO, FGA, D13S317, D16S539, D18S51, D2S1338, D21S11 and D7S820.

[0099] Enrichment of the sample for the target nucleic acids is accomplished by methods that comprise specifically amplifying target nucleic acid sequences that comprise the polymorphic site. Amplification of the target sequences can be performed by any method that uses PCR or variations of the method including but not limited to asymmetric PCR, helicase-dependent amplification, hot-start PCR, qPCR, solid phase PCR, and touchdown PCR. Alternatively, replication of target nucleic acid sequences can be obtained by enzyme-independent methods *e.g.* chemical solid-phase synthesis using the phosphoramidites. Amplification of the target sequences is accomplished using primer pairs each capable of amplifying a target nucleic acid sequence comprising the polymorphic site *e.g.* SNP, in a multiplex PCR reaction. Multiplex PCR reactions include combining at least 2, at least three, at least 3, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40 or more sets of primers in the same reaction to quantify the amplified target nucleic acids comprising at least two, at least three, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40 or more polymorphic sites in the same sequencing reaction. Any panel of primer sets can be configured to amplify at least one informative polymorphic sequence.

Amplification of SNPs

[00100] A number of nucleic acid primers are already available to amplify DNA fragments containing the SNP polymorphisms and their sequences can be obtained, for example, from the above-identified databases. Additional primers can also be designed, for example, using a method similar to that published by Vieux, E. F., Kwok, P-Y and Miller, R. D. in BioTechniques (June 2002) Vol. 32. Supplement: "SNPs: Discovery of Marker Disease, pp. 28-32. In one embodiment, at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40 or more sets of primers is chosen to amplify a target nucleic

acid comprising at least one informative SNPs in a portion of a mixture of fetal and maternal cfDNA. In one embodiment, the sets of primers comprise forward and reverse primers that encompass at least one informative SNP selected from rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. Exemplary sets of primers that are used to amplify the tandem SNPs provided in Example 5 (Tables 5 and 6) and disclosed herein as SEQ ID NOs:57-112 to amplify a target nucleic acid comprising at least one informative SNP in a portion of a mixture of fetal and maternal cfDNA. In another embodiment, the group of 13 sets of primers SEQ ID NOs:1-26 is used to amplify a target nucleic acid each comprising at least one SNP *e.g.* a single SNP, in a portion of a mixture of fetal and maternal cfDNA.

[00101] In yet another embodiment, at least one set of primers is used to amplify a target nucleic acid each comprising at least one SNP *e.g.* a set of two tandem SNPs, in a portion of a mixture of fetal and maternal cfDNA. In one embodiment, the sets of primers comprise forward and reverse primers that encompass at least one informative tandem SNP selected from rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672.

[00102] The primers used for amplifying the target sequences comprising the tandem SNPs are designed to encompass both SNP sites. Exemplary sets of primers that are used to amplify the tandem SNPs disclosed herein are provided in Example 10 and disclosed as SEQ ID Nnos:197-310.

[00103] Amplification of the target nucleic acids is performed using sequence-specific primers that allow for sequence specific amplification. For example, the PCR primers are designed to discriminate against the amplification of similar genes or paralogs that are on other chromosomes by taking advantage of sequence differences between the target nucleic acid and any paralogs from other chromosomes. The forward or reverse PCR primers are designed to anneal close to the SNP site and to amplify a nucleic acid sequence of sufficient length to be encompassed in the reads generated by massively parallel sequencing methods. In some embodiments, some massively parallel sequencing methods require that nucleic acid sequence have a minimum length (bp) to enable bridging amplification that may optionally be used prior to sequencing. Thus, the PCR primers used for amplifying target nucleic acids are designed to amplify sequences that are of sufficient length to be bridge amplified and to identify SNPs that are encompassed by the sequence reads. In some embodiments, the first of two primers in the primer set comprising the forward and the reverse primer for amplifying the target nucleic acid is designed to identify a single SNP present within a sequence read of about 20bp, about 25bp, about 30bp, about 35bp, about

40bp, about 45bp, about 50bp, about 55bp, about 60bp, about 65bp, about 70bp, about 75bp, about 80bp, about 85bp, about 90bp, about 95bp, about 100bp, about 110bp, about 120bp, about 130, about 140bp, about 150bp, about 200bp, about 250bp, about 300bp, about 350bp, about 400bp, about 450bp, or about 500bp. It is expected that technological advances in massively parallel sequencing technologies will enable single-end reads of greater than 500bp. In one embodiment, one of the PCR primers is designed to amplify SNPs that are encompassed in sequence reads of 36 bp. The second primer is designed to amplify the target nucleic acid as an amplicon of sufficient length to allow for bridge amplification. In one embodiment, the exemplary PCR primers are designed to amplify target nucleic acids that contain a single SNP selected from SNPs rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005 and rs530022. In other embodiments, the forward and reverse primers are each designed for amplifying target nucleic acids each comprising a set of two tandem SNPs, each being present within a sequence read of about 20bp, about 25bp, about 30bp, about 35bp, about 40bp, about 45bp, about 50bp, about 55bp, about 60bp, about 65bp, about 70bp, about 75bp, about 80bp, about 85bp, about 90bp, about 95bp, about 100bp, about 110bp, about 120bp, about 130, about 140bp, about 150bp, about 200bp, about 250bp, about 300bp, about 350bp, about 400bp, about 450bp, or about 500bp. In one embodiment, at least one of the primers is designed to amplify the target nucleic acid comprising a set of two tandem SNPs as an amplicon of sufficient length to allow for bridge amplification.

[00104] The SNPs, single or tandem SNPs, are contained in amplified target nucleic acid amplicons of at least about 100bp, at least about 150bp, at least about 200bp, at least about 250bp, at least about 300bp, at least about 350bp, or at least about 400bp. In one embodiment, target nucleic acids comprising a polymorphic site *e.g.* a SNP, are amplified as amplicons of at least about 110 bp, and that comprise a SNP within 36 bp from the 3' or 5' end of the amplicon. In another embodiment, target nucleic acids comprising two or more polymorphic sites *e.g.* two tandem SNPs, are amplified as amplicons of at least about 110 bp, and that comprise the first SNP within 36 bp from the 3' end of the amplicon, and/or the second SNP within 36 bp from the 5' end of the amplicon.

Amplification of STRs

[00105] A number of nucleic acid primers are already available to amplify DNA fragments containing the STRs and their sequences can be obtained, for example, from the above-identified databases. Various sized PCR amplicons have been used to discern the respective size distributions of circulating fetal and maternal DNA species, and have shown that the fetal DNA molecules in the plasma of pregnant women are generally shorter than maternal DNA molecules (Chan *et al.*, Clin Chem 50:8892 [2004]). Size fractionation of circulating fetal DNA has confirmed that the average length of circulating fetal DNA fragments is <300 bp, while maternal DNA has been estimated to be between about 0.5 and 1 Kb (Li *et al.*, Clin Chem, 50: 1002-1011 [2004]). These findings are consistent with those of Fan *et al.*, who determined using NGS that fetal cfDNA is rarely >340bp (Fan *et al.*, Clin Chem 56:1279-1286 [2010]). The method of the invention encompasses determining the fraction of fetal nucleic acid in a maternal sample that has been enriched with target nucleic acids each comprising one miniSTR comprising quantifying at least one fetal and one maternal allele at a polymorphic miniSTR, which can be amplified to generate amplicons that are of lengths about the size of the circulating fetal DNA fragments.

[00106] In one embodiment, the method comprises determining the number of copies of at least one fetal and at least one maternal allele at least at one polymorphic miniSTR that is amplified to generate amplicons that are less than about 300 bp, less than about 250 bp, less than about 200 bp, less than about 150 bp, less than about 100 bp, or

less than about 50 bp. In another embodiment, the amplicons that are generated by amplifying the miniSTRs are less than about 300 bp. In another embodiment, the amplicons that are generated by amplifying the miniSTRs are less than about 250 bp. In another embodiment, the amplicons that are generated by amplifying the miniSTRs are less than about 200 bp. Amplification of the informative allele includes using miniSTR primers, which allow for the amplification of reduced-size amplicons to discern STR alleles that are less than about 500 bp, less than about 450 bp, less than about 400 bp, less than about 350 bp, less than about 300 base pairs (bp), less than about 250 bp, less than about 200 bp, less than about 150 bp, less than about 100 bp, or less than about 50 bp. The reduced-size amplicons generated using the miniSTR primers are known as miniSTRs that are identified according to the marker name corresponding to the locus to which they have been mapped. In one embodiment, the miniSTR primers include mini STR primers that have permitted the maximum size reduction in amplicon size for all 13 CODIS STR loci in addition to the D2S1338, Penta D, and pentaE found in commercially available STR kits (Butler *et al.*, J Forensic Sci 48:1054-1064 [2003]), miniSTR loci that are unlinked to the CODIS markers as described by Coble and Butler (Coble and Butler, J Forensic Sci 50:43-53 [2005]), and other miniSTRs that have been characterized at NIST. Information regarding the miniSTRs characterized at NIST is accessible via the world wide web at cstl.nist.gov/biotech/strbase/newSTRs.htm. Any one pair or a combination of two or more pairs of miniSTR primers can be used to amplify at least one miniSTR. For example, at least one set of primers is selected from set CSF1PO_F (SEQ ID NO:81) and CSF1PO_R (SEQ ID NO:82), set FGA_F (SEQ ID NO:83) and FGA_R (SEQ ID NO:84), set TH01_F (SEQ ID NO:85) and TH01_R (SEQ ID NO:86), set TPOX_F (SEQ ID NO:87) and TPOX_R (SEQ ID NO:88), set vWA_F (SEQ ID NO:89) and vWA_R (SEQ ID NO:90), set D3S1358_F (SEQ ID NO:91) and D3S1358_R (SEQ ID NO:92), set D5S818_F (SEQ ID NO:93) and D5S818_R (SEQ ID NO:94), set D7S820_F (SEQ ID NO:95) and D7S820_R (SEQ ID NO:96), set D7S820_F (SEQ ID NO:97) and D7S820_R (SEQ ID NO:98), set D13S317_F (SEQ ID NO:99) and D13S317_R (SEQ ID NO:100), set D16S539_F (SEQ ID NO:101) and D16S539_R (SEQ ID NO:102), set D18S51_F (SEQ ID NO:103) and D18S51_R (SEQ ID NO:104), set D21S11_F (SEQ ID NO:105) and D21S11_R (SEQ ID NO:106), set D2S1338_F (SEQ ID NO:107) and D2S1338_R (SEQ ID NO:108), set Penta D_F (SEQ ID NO:109) and Penta D_R (SEQ ID NO:110), set Penta E_F (SEQ ID NO:111) and Penta E_R (SEQ ID NO:112), set (D22S1045_F; SEQ ID NO:113) and D22S1045_F (SEQ ID NO:114), set D20S1082_R (SEQ ID NO:115) and D20S1082_F (SEQ ID NO:116), set D20S482_R (SEQ ID NO:117) and D20S482_F (SEQ ID NO:118), set D18S853_R (SEQ ID NO:119) and D18S853_F (SEQ ID NO:120), set D17S1301_F (SEQ ID NO:121) and D17S1301_R (SEQ ID NO:122), set D17S974_F (SEQ ID NO:123) and D17S974_R (SEQ ID NO:124), set D14S1434_F (SEQ ID NO:125) and D14S1434_R (SEQ ID NO:126), set D12ATA63_F (SEQ ID NO:127) and D12ATA63_R (SEQ ID NO:128), D11S4463_F (SEQ ID NO:129) and D11S4463_R (SEQ ID NO:130), set D10S1435_F (SEQ ID NO:131) and D10S1435_R (SEQ ID NO:132), set D10S1248_F (SEQ ID NO:133) and D10S1248_R (SEQ ID NO:134), set D9S2157_F (SEQ ID NO:135) and D9S2157_R (SEQ ID NO:136), set D9S1122_F (SEQ ID NO:137) and D9S1122_R (SEQ ID NO:138), set D8S1115_F (SEQ ID NO:139) and D8S1115_R (SEQ ID NO:140), set D6S1017_F (SEQ ID NO:141) and D6S1017_R (SEQ ID NO:142), D6S474_F (SEQ ID NO:143) and D6S474_R (SEQ ID NO:144), set D5S2500_F (SEQ ID NO:145) and D5S2500_R (SEQ ID NO:146), set D4S2408_F (SEQ ID NO:147) and D4S2408_R (SEQ ID NO:148), set D4S2364U_F (SEQ ID NO:149) and D4S2364U_R (SEQ ID NO:150), set D3S452_F (SEQ ID NO:151) and D3S452_R (SEQ ID NO:152), set D3S3053_F (SEQ ID NO:153) and D3S3053_R (SEQ ID NO:154), set D2S1776_F (SEQ ID NO:155) and D2S1776_R (SEQ ID NO:156), set D2S441_F (SEQ ID NO:157) and D2S441_R (SEQ ID NO:158), set D1S1677_F (SEQ ID NO:159) and

D1S1677_R (SEQ ID NO:160), set D1S1627_F (SEQ ID NO:161) and D1S1627_R (SEQ ID NO:162), and set D1GATA113_F (SEQ ID NO:163) and D1GATA113_R (SEQ ID NO:164).

[00107] Enrichment of the sample is obtained by amplifying target nucleic acids contained in a portion of the mixture of fetal and maternal nucleic acids in the original sample, and combining at least a portion or all of the amplified product with the remainder of the original unamplified sample. Enrichment comprises amplifying the target nucleic acids that are contained in a portion of biological fluid sample. In one embodiment, the sample that is enriched is the plasma fraction of a blood sample (See **Figure 2**). For example, a portion of an original maternal plasma sample is used for amplifying target nucleic acid sequences. Subsequently, some or all of the amplified product is combined with the remaining unamplified original plasma sample thereby enriching it (see Example 8). In another embodiment, the sample that is enriched is the sample of purified cfDNA that is extracted from plasma (See **Figure 3**). For example, enrichment comprises amplifying the target nucleic acids that are contained in a portion of an original sample of purified mixture of fetal and maternal nucleic acids *e.g.* cfDNA that has been purified from a maternal plasma sample, and subsequently combining some or all of the amplified product with the remaining unamplified original purified sample (see Example 7). In yet another embodiment, the sample that is enriched is a sequencing library sample prepared from a purified mixture of fetal and maternal nucleic acids (see **Figure 4**). For example, enrichment comprises amplifying the target nucleic acids that are contained in a portion of an original sample of purified mixture of fetal and maternal nucleic acids *e.g.* cfDNA that has been purified from a maternal plasma sample, preparing a first sequencing library of unamplified nucleic acid sequences, preparing a second sequencing library of amplified polymorphic target nucleic acids, and subsequently combining some or all of the second sequencing library with some or all of the first sequencing library (see Example 6). The amount of amplified product that is used to enrich the original sample is selected to obtain sufficient sequencing information for determining both the presence or absence of aneuploidy and the fetal fraction from the same sequencing run. At least about 3%, at least about 5%, at least about 7%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30% or more of the total number of sequence tags obtained from sequencing are mapped to determine the fetal fraction.

[00108] In step 130 (**Figure 1**), the enriched mixture of fetal and maternal nucleic acids is sequenced. Sequence information that is needed for the simultaneous determination of aneuploidy and fetal fraction can be obtained using any of the known DNA sequencing methods. In one embodiment, the method described herein employs next generation sequencing technology (NGS) in which clonally amplified DNA templates or single DNA molecules are sequenced in a massively parallel fashion within a flow cell (*e.g.* as described in Volkerding *et al.* Clin Chem 55:641-658 [2009]; Metzker M Nature Rev 11:31-46 [2010]). In addition to high-throughput sequence information, NGS provides digital quantitative information, in that each sequence read is a countable "sequence tag" representing an individual clonal DNA template or a single DNA molecule. This quantification allows NGS to expand the digital concept of counting cell-free DNA molecules (Fan *et al.*, Proc Natl Acad Sci U S A 105:16266-16271 [2008]; Chiu *et al.*, Proc Natl Acad Sci U S A 2008;105:20458-20463 [2008]). The sequencing technologies of NGS include pyrosequencing, sequencing-by-synthesis with reversible dye terminators, sequencing by oligonucleotide probe ligation and real time sequencing.

[00109] Some of the sequencing technologies are available commercially, such as the sequencing-by-hybridization platform from Affymetrix Inc. (Sunnyvale, CA) and the sequencing-by-synthesis platforms from 454 Life Sciences (Bradford, CT), Illumina/Solexa (Hayward, CA) and Helicos Biosciences (Cambridge, MA), and the sequencing-by-ligation platform from Applied Biosystems (Foster City, CA), as described below. In addition to the single

molecule sequencing performed using sequencing-by-synthesis of Helicos Biosciences, other single molecule sequencing technologies are encompassed by the method of the invention and include the SMRT™ technology of Pacific Biosciences, the Ion Torrent™ technology, and nanopore sequencing being developed for example, by Oxford Nanopore Technologies.

[00110] While the automated Sanger method is considered as a ‘first generation’ technology, Sanger sequencing including the automated Sanger sequencing, can also be employed by the method of the invention. Additional sequencing methods that comprise the use of developing nucleic acid imaging technologies *e.g.* atomic force microscopy (AFM) or transmission electron microscopy (TEM), are also encompassed by the method of the invention. Exemplary sequencing technologies are described below.

[00111] In one embodiment, the DNA sequencing technology that is used in the method of the invention is the Helicos True Single Molecule Sequencing (tSMS) (*e.g.* as described in Harris T.D. et al., Science 320:106-109 [2008]). In the tSMS technique, a DNA sample is cleaved into strands of approximately 100 to 200 nucleotides, and a polyA sequence is added to the 3’ end of each DNA strand. Each strand is labeled by the addition of a fluorescently labeled adenosine nucleotide. The DNA strands are then hybridized to a flow cell, which contains millions of oligo-T capture sites that are immobilized to the flow cell surface. The templates can be at a density of about 100 million templates/cm². The flow cell is then loaded into an instrument, *e.g.*, HeliScope™ sequencer, and a laser illuminates the surface of the flow cell, revealing the position of each template. A CCD camera can map the position of the templates on the flow cell surface. The template fluorescent label is then cleaved and washed away. The sequencing reaction begins by introducing a DNA polymerase and a fluorescently labeled nucleotide. The oligo-T nucleic acid serves as a primer. The polymerase incorporates the labeled nucleotides to the primer in a template directed manner. The polymerase and unincorporated nucleotides are removed. The templates that have directed incorporation of the fluorescently labeled nucleotide are discerned by imaging the flow cell surface. After imaging, a cleavage step removes the fluorescent label, and the process is repeated with other fluorescently labeled nucleotides until the desired read length is achieved. Sequence information is collected with each nucleotide addition step.

[00112] In one embodiment, the DNA sequencing technology that is used in the method of the invention is the 454 sequencing (Roche) (*e.g.* as described in Margulies, M. et al. Nature 437:376-380 [2005]). 454 sequencing involves two steps. In the first step, DNA is sheared into fragments of approximately 300-800 base pairs, and the fragments are blunt-ended. Oligonucleotide adaptors are then ligated to the ends of the fragments. The adaptors serve as primers for amplification and sequencing of the fragments. The fragments can be attached to DNA capture beads, *e.g.*, streptavidin-coated beads using, *e.g.*, Adaptor B, which contains 5’-biotin tag. The fragments attached to the beads are PCR amplified within droplets of an oil-water emulsion. The result is multiple copies of clonally amplified DNA fragments on each bead. In the second step, the beads are captured in wells (pico-liter sized). Pyrosequencing is performed on each DNA fragment in parallel. Addition of one or more nucleotides generates a light signal that is recorded by a CCD camera in a sequencing instrument. The signal strength is proportional to the number of nucleotides incorporated. Pyrosequencing makes use of pyrophosphate (PPi) which is released upon nucleotide addition. PPi is converted to ATP by ATP sulfurylase in the presence of adenosine 5’ phosphosulfate. Luciferase uses ATP to convert luciferin to oxyluciferin, and this reaction generates light that is discerned and analyzed.

[00113] In one embodiment, the DNA sequencing technology that is used in the method of the invention is the SOLiD™ technology (Applied Biosystems). In SOLiD™ sequencing-by-ligation, genomic DNA is sheared into

fragments, and adaptors are attached to the 5' and 3' ends of the fragments to generate a fragment library. Alternatively, internal adaptors can be introduced by ligating adaptors to the 5' and 3' ends of the fragments, circularizing the fragments, digesting the circularized fragment to generate an internal adaptor, and attaching adaptors to the 5' and 3' ends of the resulting fragments to generate a mate-paired library. Next, clonal bead populations are prepared in microreactors containing beads, primers, template, and PCR components. Following PCR, the templates are denatured and beads are enriched to separate the beads with extended templates. Templates on the selected beads are subjected to a 3' modification that permits bonding to a glass slide. The sequence can be determined by sequential hybridization and ligation of partially random oligonucleotides with a central determined base (or pair of bases) that is identified by a specific fluorophore. After a color is recorded, the ligated oligonucleotide is cleaved and removed and the process is then repeated.

[00114] In one embodiment, the DNA sequencing technology that is used in the method of the invention is the single molecule, real-time (SMRT™) sequencing technology of Pacific Biosciences. In SMRT sequencing, the continuous incorporation of dye-labeled nucleotides is imaged during DNA synthesis. Single DNA polymerase molecules are attached to the bottom surface of individual zero-mode wavelength identifiers (ZMW identifiers) that obtain sequence information while phospholinked nucleotides are being incorporated into the growing primer strand. A ZMW is a confinement structure which enables observation of incorporation of a single nucleotide by DNA polymerase against the background of fluorescent nucleotides that rapidly diffuse in and out of the ZMW (in microseconds). It takes several milliseconds to incorporate a nucleotide into a growing strand. During this time, the fluorescent label is excited and produces a fluorescent signal, and the fluorescent tag is cleaved off. Identification of the corresponding fluorescence of the dye indicates which base was incorporated. The process is repeated.

[00115] In one embodiment, the DNA sequencing technology that is used in the method of the invention is nanopore sequencing (e.g. as described in Soni GV and Meller A. Clin Chem 53: 1996-2001 [2007]). Nanopore sequencing DNA analysis techniques are being industrially developed by a number of companies, including Oxford Nanopore Technologies (Oxford, United Kingdom). Nanopore sequencing is a single-molecule sequencing technology whereby a single molecule of DNA is sequenced directly as it passes through a nanopore. A nanopore is a small hole, of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential (voltage) across it results in a slight electrical current due to conduction of ions through the nanopore. The amount of current which flows is sensitive to the size and shape of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree, changing the magnitude of the current through the nanopore in different degrees. Thus, this change in the current as the DNA molecule passes through the nanopore represents a reading of the DNA sequence.

[00116] In one embodiment, the DNA sequencing technology that is used in the method of the invention is the chemical-sensitive field effect transistor (chemFET) array (e.g., as described in U.S. Patent Application Publication No. 20090026082). In one example of the technique, DNA molecules can be placed into reaction chambers, and the template molecules can be hybridized to a sequencing primer bound to a polymerase. Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be discerned by a change in current by a chemFET. An array can have multiple chemFET sensors. In another example, single nucleic acids can be attached to beads, and the nucleic acids can be amplified on the bead, and the individual beads can be transferred to individual reaction chambers on a chemFET array, with each chamber having a chemFET sensor, and the nucleic acids can be sequenced.

[00117] In one embodiment, the DNA sequencing technology that is used in the method of the invention is the Halcyon Molecular's method that uses transmission electron microscopy (TEM). The method, termed Individual Molecule Placement Rapid Nano Transfer (IMPRNT), comprises utilizing single atom resolution transmission electron microscope imaging of high-molecular weight (150kb or greater) DNA selectively labeled with heavy atom markers and arranging these molecules on ultra-thin films in ultra-dense (3nm strand-to-strand) parallel arrays with consistent base-to-base spacing. The electron microscope is used to image the molecules on the films to determine the position of the heavy atom markers and to extract base sequence information from the DNA. The method is further described in PCT patent publication WO 2009/046445. The method allows for sequencing complete human genomes in less than ten minutes.

[00118] In one embodiment, the DNA sequencing technology is the Ion Torrent single molecule sequencing, which pairs semiconductor technology with a simple sequencing chemistry to directly translate chemically encoded information (A, C, G, T) into digital information (0, 1) on a semiconductor chip. In nature, when a nucleotide is incorporated into a strand of DNA by a polymerase, a hydrogen ion is released as a byproduct. Ion Torrent uses a high-density array of micro-machined wells to perform this biochemical process in a massively parallel way. Each well holds a different DNA molecule. Beneath the wells is an ion-sensitive layer and beneath that an ion sensor. When a nucleotide, for example a C, is added to a DNA template and is then incorporated into a strand of DNA, a hydrogen ion will be released. The charge from that ion will change the pH of the solution, which can be identified by Ion Torrent's ion sensor. The sequencer—essentially the world's smallest solid-state pH meter—calls the base, going directly from chemical information to digital information. The Ion personal Genome Machine (PGM™) sequencer then sequentially floods the chip with one nucleotide after another. If the next nucleotide that floods the chip is not a match. No voltage change will be recorded and no base will be called. If there are two identical bases on the DNA strand, the voltage will be double, and the chip will record two identical bases called. Direct identification allows recordation of nucleotide incorporation in seconds.

[00119] Other sequencing methods include digital PCR and sequencing by hybridization. Digital polymerase chain reaction (digital PCR or dPCR) can be used to directly identify and quantify nucleic acids in a sample. Digital PCR can be performed in an emulsion. Individual nucleic acids are separated, e.g., in a microfluidic chamber device, and each nucleic acid is individually amplified by PCR. Nucleic acids can be separated such there is an average of approximately 0.5 nucleic acids/well, or not more than one nucleic acid/well. Different probes can be used to distinguish fetal alleles and maternal alleles. Alleles can be enumerated to determine copy number. In sequencing by hybridization, the hybridization comprises contacting the plurality of polynucleotide sequences with a plurality of polynucleotide probes, wherein each of the plurality of polynucleotide probes can be optionally tethered to a substrate. The substrate might be flat surface comprising an array of known nucleotide sequences. The pattern of hybridization to the array can be used to determine the polynucleotide sequences present in the sample. In other embodiments, each probe is tethered to a bead, e.g., a magnetic bead or the like. Hybridization to the beads can be identified and used to identify the plurality of polynucleotide sequences within the sample.

[00120] In one embodiment, the method employs massively parallel sequencing of millions of DNA fragments using Illumina's sequencing-by-synthesis and reversible terminator-based sequencing chemistry (e.g. as described in Bentley *et al.*, Nature 6:53-59 [2009]). Template DNA can be genomic DNA e.g. cfDNA. In some embodiments, genomic DNA from isolated cells is used as the template, and it is fragmented into lengths of several hundred base pairs. In other embodiments, cfDNA is used as the template, and fragmentation is not required as cfDNA exists as short fragments. For example fetal cfDNA circulates in the bloodstream as fragments of <300 bp, and maternal

cfDNA has been estimated to circulate as fragments of between about 0.5 and 1 Kb (Li et al., Clin Chem, 50: 1002-1011 [2004]). Illumina's sequencing technology relies on the attachment of fragmented genomic DNA to a planar, optically transparent surface on which oligonucleotide anchors are bound. Template DNA is end-repaired to generate 5'-phosphorylated blunt ends, and the polymerase activity of Klenow fragment is used to add a single A base to the 3' end of the blunt phosphorylated DNA fragments. This addition prepares the DNA fragments for ligation to oligonucleotide adapters, which have an overhang of a single T base at their 3' end to increase ligation efficiency. The adapter oligonucleotides are complementary to the flow-cell anchors. Under limiting-dilution conditions, adapter-modified, single-stranded template DNA is added to the flow cell and immobilized by hybridization to the anchors. Attached DNA fragments are extended and bridge amplified to create an ultra-high density sequencing flow cell with hundreds of millions of clusters, each containing ~1,000 copies of the same template. In one embodiment, the randomly fragmented genomic DNA *e.g.* cfDNA, is amplified using PCR before it is subjected to cluster amplification. Alternatively, an amplification-free genomic library preparation is used, and the randomly fragmented genomic DNA *e.g.* cfDNA is enriched using the cluster amplification alone (Kozarewa *et al.*, Nature Methods 6:291-295 [2009]). The templates are sequenced using a robust four-color DNA sequencing-by-synthesis technology that employs reversible terminators with removable fluorescent dyes. High-sensitivity fluorescence identification is achieved using laser excitation and total internal reflection optics. Short sequence reads of about 20-40 bp *e.g.* 36 bp, are aligned against a repeat-masked reference genome and genetic differences are called using specially developed data analysis pipeline software. After completion of the first read, the templates can be regenerated in situ to enable a second read from the opposite end of the fragments. Thus, either single-end or paired end sequencing of the DNA fragments is used according to the method. Partial sequencing of DNA fragments present in the sample is performed, and sequence tags comprising reads of predetermined length *e.g.* 36 bp, that are mapped to a known reference genome are counted. In one embodiment, the reference genome sequence is the NCBI36/hg18 sequence, which is available on the world wide web at genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18&hgsid=166260105). Other sources of public sequence information include GenBank, dbEST, dbSTS, EMBL (the European Molecular Biology Laboratory), and the DDBJ (the DNA Databank of Japan). A number of computer algorithms are available for aligning sequences, including without limitation BLAST (Altschul et al., 1990), BLITZ (MPsrch) (Sturrock & Collins, 1993), FASTA (Person & Lipman, 1988), BOWTIE (Langmead *et al.*, Genome Biology 10:R25.1-R25.10 [2009]), or ELAND (Illumina, Inc., San Diego, CA, USA). In one embodiment, one end of the clonally expanded copies of the plasma cfDNA molecules is sequenced and processed by bioinformatic alignment analysis for the Illumina Genome Analyzer, which uses the Efficient Large-Scale Alignment of Nucleotide Databases (ELAND) software.

[00121] The length of the sequence read is associated with the particular sequencing technology. NGS methods provide sequence reads that vary in size from tens to hundreds of base pairs. In some embodiments of the method described herein, the sequence reads are about 20bp, about 25bp, about 30bp, about 35bp, about 40bp, about 45bp, about 50bp, about 55bp, about 60bp, about 65bp, about 70bp, about 75bp, about 80bp, about 85bp, about 90bp, about 95bp, about 100bp, about 110bp, about 120bp, about 130, about 140bp, about 150bp, about 200bp, about 250bp, about 300bp, about 350bp, about 400bp, about 450bp, or about 500bp. It is expected that technological advances will enable single-end reads of greater than 500bp enabling for reads of greater than about 1000bp when paired end reads are generated. In one embodiment, the sequence reads are 36bp. Other sequencing methods that can be employed by the method of the invention include the single molecule sequencing methods that can sequence nucleic acids molecules >5000 bp. The massive quantity of sequence output is transferred by an analysis pipeline

that transforms primary imaging output from the sequencer into strings of bases. A package of integrated algorithms performs the core primary data transformation steps: image analysis, intensity scoring, base calling, and alignment.

[00122] In one embodiment, partial sequencing of DNA fragments present in the sample is performed, and sequence tags comprising reads of predetermined length *e.g.* 36 bp, that map to a known reference genome are counted. Only sequence reads that uniquely align to the reference genome are counted as sequence tags. In one embodiment, the reference genome is the human reference genome NCBI36/hg18 sequence, which is available on the world wide web at genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18&hgsid=166260105). Other sources of public sequence information include GenBank, dbEST, dbSTS, EMBL (the European Molecular Biology Laboratory), and the DDBJ (the DNA Databank of Japan). In another embodiment, the reference genome comprises the human reference genome NCBI36/hg18 sequence and an artificial target sequences genome, which includes the target polymorphic sequences *e.g.* a SNP genome. Mapping of the sequence tags is achieved by comparing the sequence of the tag with the sequence of the reference genome to determine the chromosomal origin of the sequenced nucleic acid (*e.g.* cfDNA) molecule, and specific genetic sequence information is not needed. A number of computer algorithms are available for aligning sequences, including without limitation BLAST (Altschul *et al.*, 1990), BLITZ (MPsrch) (Sturrock & Collins, 1993), FASTA (Person & Lipman, 1988), BOWTIE (Langmead *et al.*, Genome Biology 10:R25.1-R25.10 [2009]), or ELAND (Illumina, Inc., San Diego, CA, USA). In one embodiment, one end of the clonally expanded copies of the plasma cfDNA molecules is sequenced and processed by bioinformatic alignment analysis for the Illumina Genome Analyzer, which uses the Efficient Large-Scale Alignment of Nucleotide Databases (ELAND) software. Analysis of sequencing information for the determination of aneuploidy may allow for a small degree of mismatch (0-2 mismatches per sequence tag) to account for minor polymorphisms that may exist between the reference genome and the genomes in the mixed sample. Analysis of sequencing information for the determination of fetal fraction may allow for a small degree of mismatch depending on the polymorphic sequence. For example, a small degree of mismatch may be allowed if the polymorphic sequence is an STR. In cases when the polymorphic sequence is a SNP, all sequence that match exactly to either of the two alleles at the SNP site are counted first and filtered from the remaining reads, for which a small degree of mismatch may be allowed.

[00123] In step 140, the sequencing information obtained in step 130 is analyzed and the simultaneous determination of the fetal fraction and determination of the presence or absence of aneuploidy is made.

[00124] A plurality of sequence tags are obtained per sample. In some embodiments, at least about 3×10^6 sequence tags, at least about 5×10^6 sequence tags, at least about 8×10^6 sequence tags, at least about 10×10^6 sequence tags, at least about 15×10^6 sequence tags, at least about 20×10^6 sequence tags, at least about 30×10^6 sequence tags, at least about 40×10^6 sequence tags, or at least about 50×10^6 sequence tags comprising between 20 and 40bp reads are obtained from mapping the reads to the reference genome per sample. In one embodiment, all the sequence reads are mapped to all regions of the reference genome. In one embodiment, the tags comprising reads that have been mapped to all regions *e.g.* all chromosomes, of the human reference genome are counted, and the fetal aneuploidy *i.e.* the over- or under-representation of a sequence of interest *e.g.* a chromosome or portion thereof, in the mixed DNA sample is determined, and the tags comprising reads that are mapped to the artificial target sequences genome are counted to determine the fetal fraction. The method does not require differentiation between the maternal and fetal genomes.

Determination of aneuploidy

[00125] The accuracy required for correctly determining whether an aneuploidy is present or absent in a sample, is predicated in part on the variation of the number of sequence tags that map to the reference genome among samples within a sequencing run (inter-chromosomal variability), and the variation of the number of sequence tags that map to the reference genome in different sequencing runs (inter-sequencing variability). For example, the variations can be particularly pronounced for tags that map to GC-rich or GC-poor reference sequences. In one embodiment, the method uses sequencing information to calculate chromosome dose, which intrinsically account for the accrued variability stemming from interchromosomal, inter-sequencing and platform-dependent variability. Chromosome doses are determined from sequencing information *i.e.* the number of sequence tags, for the sequence of interest *e.g.* chromosome 21, and the number of sequence tags for a normalizing sequence. Identification of a normalizing sequence is performed in a set of qualified samples known not to contain an aneuploidy of the sequence of interest. The flow chart provided in Figure 5 shows the process 500 whereby normalizing sequences *e.g.* normalizing chromosomes, are identified, and the presence or absence of an aneuploidy is determined. In step 510, a set of qualified maternal samples is obtained to identify qualified normalizing sequences *e.g.* normalizing chromosomes, and to provide variance values for use in determining statistically meaningful identification of an aneuploidy in test samples. In step 510, a plurality of biological qualified samples are obtained from a plurality of subjects known to comprise cells having a normal copy number for any one sequence of interest *e.g.* a chromosome of interest such as a chromosome associated with an aneuploidy. In one embodiment, the qualified samples are obtained from mothers pregnant with a fetus that has been confirmed using cytogenetic means to have a normal copy number of chromosomes relative to the chromosome of interest. The biological qualified maternal samples may be biological fluid samples *e.g.* plasma samples, or any suitable sample as described above that contains a mixture of fetal and maternal cfDNA molecules.

[00126] In step 520, at least a portion of each of all the qualified nucleic acids contained in the qualified maternal samples are sequenced to generate sequence reads of between 20 and 40bp *e.g.* 36bp, which are aligned to a reference genome, *e.g.* hg18. In some embodiments, the sequence reads comprise about 20bp, about 25bp, about 30bp, about 35bp, about 40bp, about 45bp, about 50bp, about 55bp, about 60bp, about 65bp, about 70bp, about 75bp, about 80bp, about 85bp, about 90bp, about 95bp, about 100bp, about 110bp, about 120bp, about 130, about 140bp, about 150bp, about 200bp, about 250bp, about 300bp, about 350bp, about 400bp, about 450bp, or about 500bp. It is expected that technological advances will enable single-end reads of greater than 500bp enabling for reads of greater than about 1000bp when paired end reads are generated. In one embodiment, the sequence reads comprise 36bp. Sequence reads are aligned to a human reference genome, and the reads that are uniquely mapped to the human reference genome are counted as sequence tags. In one embodiment, at least about 3×10^6 qualified sequence tags, at least about 5×10^6 qualified sequence tags, at least about 8×10^6 qualified sequence tags, at least about 10×10^6 qualified sequence tags, at least about 15×10^6 qualified sequence tags, at least about 20×10^6 qualified sequence tags, at least about 30×10^6 qualified sequence tags, at least about 40×10^6 qualified sequence tags, or at least about 50×10^6 qualified sequence tags comprising between 20 and 40bp reads are obtained from reads that map uniquely to a reference genome.

[00127] In step 530, all the tags obtained from sequencing the nucleic acids in the qualified maternal samples are counted to determine a qualified sequence tag density. In one embodiment the sequence tag density is determined as the number of qualified sequence tags mapped to the sequence of interest on the reference genome. In another embodiment, the qualified sequence tag density is determined as the number of qualified sequence tags mapped to a

sequence of interest normalized to the length of the qualified sequence of interest to which they are mapped. Sequence tag densities that are determined as a ratio of the tag density relative to the length of the sequence of interest are herein referred to as tag density ratios. Normalization to the length of the sequence of interest is not required, and may be included as a step to reduce the number of digits in a number to simplify it for human interpretation. As all qualified sequence tags are mapped and counted in each of the qualified samples, the sequence tag density for a sequence of interest *e.g.* chromosome of interest, in the qualified samples is determined, as are the sequence tag densities for additional sequences from which normalizing sequences *e.g.* chromosomes, are identified subsequently. In one embodiment, the sequence of interest is a chromosome that is associated with a chromosomal aneuploidy *e.g.* chromosome 21, and the qualified normalizing sequence is a chromosome that is not associated with a chromosomal aneuploidy and whose variation in sequence tag density best approximates that of chromosome 21. For example, a qualified normalizing sequence is a sequence that has the smallest variability. In some embodiments, the normalizing sequence is a sequence that best distinguishes one or more qualified, samples from one or more affected samples *i.e.* the normalizing sequence is a sequence that has the greatest differentiability. The level of differentiability can be determined as a statistical difference between the chromosome doses in a population of qualified samples and the chromosome dose(s) in one or more test samples. In another embodiment, the sequence of interest is a segment of a chromosome associated with a partial aneuploidy, *e.g.* a chromosomal deletion or insertion, or unbalanced chromosomal translocation, and the normalizing sequence is a chromosomal segment that is not associated with the partial aneuploidy and whose variation in sequence tag density best approximates that of the chromosome segment associated with the partial aneuploidy.

[00128] In step 540, based on the calculated qualified tag densities, a qualified sequence dose for a sequence of interest is determined as the ratio of the sequence tag density for the sequence of interest and the qualified sequence tag density for additional sequences from which normalizing sequences are identified subsequently. In one embodiment, doses for the chromosome of interest *e.g.* chromosome 21, is determined as a ratio of the sequence tag density of chromosome 21 and the sequence tag density for each of all the remaining chromosomes *i.e.* chromosomes 1-20, chromosome 22, chromosome X, and chromosome Y.

[00129] In step 545, a normalizing sequence *e.g.* a normalizing chromosome, is identified for a sequence of interest *e.g.* chromosome 21, in a qualified sample based on the calculated sequence doses. The method identifies sequences that inherently have similar characteristics and that are prone to similar variations among samples and sequencing runs, and which are useful for determining sequence doses in test samples. In some embodiments, the normalizing sequence is one that best differentiates an affected sample *i.e.* an aneuploid sample, from one or more qualified samples. In other embodiments, a normalizing sequence is a sequence that displays a variability in the number of sequence tags that are mapped to it among samples and sequencing runs that best approximates that of the sequence of interest for which it is used as a normalizing parameter, and/or that can best differentiate an affected sample from one or more unaffected samples.

[00130] In some embodiments, more than one normalizing sequence is identified. For example, the variation *e.g.* coefficient of variation, in chromosome dose for chromosome of interest 21 is least when the sequence tag density of chromosome 14 is used. In other embodiments, two, three, four, five, six, seven, eight or more normalizing sequences are identified for use in determining a sequence dose for a sequence of interest in a test sample.

[00131] In one embodiment, the normalizing sequence for chromosome 21 is selected from chromosome 9, chromosome 1, chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 8, chromosome 10, chromosome 11, chromosome 12, chromosome 13, chromosome 14, chromosome

15, chromosome 16, and chromosome 17. Preferably, the normalizing sequence for chromosome 21 is selected from chromosome 9, chromosome 1, chromosome 2, chromosome 11, chromosome 12, and chromosome 14. Alternatively, the normalizing sequence for chromosome 21 is a group of chromosomes selected from chromosome 9, chromosome 1, chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 8, chromosome 10, chromosome 11, chromosome 12, chromosome 13, chromosome 14, chromosome 15, chromosome 16, and chromosome 17. In other embodiments, the normalizing sequence for chromosome 21 is a group of chromosomes selected from chromosome 9, chromosome 1, chromosome 2, chromosome 11, chromosome 12, and chromosome 14.

[00132] In one embodiment, the normalizing sequence for chromosome 18 is selected from chromosome 8, chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 9, chromosome 10, chromosome 11, chromosome 12, chromosome 13, and chromosome 14. Preferably, the normalizing sequence for chromosome 18 is selected chromosome 8, chromosome 2, chromosome 3, chromosome 5, chromosome 6, chromosome 12, and chromosome 14. Alternatively, the normalizing sequence for chromosome 18 is a group of chromosomes selected from chromosome 8, chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 9, chromosome 10, chromosome 11, chromosome 12, chromosome 13, and chromosome 14. In other embodiments, the normalizing sequence for chromosome 18 is a group of chromosomes selected from chromosome 8, chromosome 2, chromosome 3, chromosome 5, chromosome 6, chromosome 12, and chromosome 14.

[00133] In one embodiment, the normalizing sequence for chromosome X is selected from chromosome 1, chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 8, chromosome 9, chromosome 10, chromosome 11, chromosome 12, chromosome 13, chromosome 14, chromosome 15, and chromosome 16. Preferably, the normalizing sequence for chromosome X is selected from chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, and chromosome 8. Alternatively, the normalizing sequence for chromosome X is a group of chromosomes selected from chromosome 1, chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 8, chromosome 9, chromosome 10, chromosome 11, chromosome 12, chromosome 13, chromosome 14, chromosome 15, and chromosome 16. In other embodiments, the normalizing sequence for chromosome X is a group of chromosomes selected from chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, and chromosome 8.

[00134] In one embodiment, the normalizing sequence for chromosome 13 is a chromosome selected from chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 8, chromosome 9, chromosome 10, chromosome 11, chromosome 12, chromosome 14, chromosome 18, and chromosome 21. Preferably, the normalizing sequence for chromosome 13 is selected from chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, and chromosome 8. In another embodiment, the normalizing sequence for chromosome 13 is a group of chromosomes selected from chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, chromosome 7, chromosome 8, chromosome 9, chromosome 10, chromosome 11, chromosome 12, chromosome 14, chromosome 18, and chromosome 21. In other embodiments, the normalizing sequence for chromosome 13 is a group of chromosomes selected from chromosome 2, chromosome 3, chromosome 4, chromosome 5, chromosome 6, and chromosome 8.

[00135] The variation in chromosome dose for chromosome Y is greater than 30 independently of which normalizing chromosome is used in determining the chromosome Y dose. Therefore, any one chromosome, or a group of two or more chromosomes selected from chromosomes 1-22 and chromosome X can be used as the

normalizing sequence for chromosome Y. In one embodiment, the at least one normalizing chromosome is a group of chromosomes consisting of chromosomes 1-22, and chromosome X. In another embodiment, the at least one normalizing chromosome is a group of chromosomes selected from chromosome 2, chromosome 3, chromosome 4, chromosome 5 and chromosome 6.

[00136] Based on the identification of the normalizing sequence(s) in qualified samples, a sequence dose is determined for a sequence of interest in a test sample comprising a mixture of nucleic acids derived from genomes that differ in one or more sequences of interest.

[00137] In step **515**, a test sample *e.g.* plasma sample, comprising fetal and maternal nucleic acids *e.g.* cfDNA, is obtained from a pregnant subject *e.g.* a pregnant woman, for which the presence or absence of a fetal aneuploidy needs to be determined.

[00138] In step **525**, at least a portion of the test nucleic acids in the test sample is sequenced to generate millions of sequence reads comprising between 20 and 500 bp *e.g.* 36bp. As in step **520**, the reads generated from sequencing the nucleic acids in the test sample are uniquely mapped to a human reference genome and are counted. As described in step **520**, at least about 3×10^6 qualified sequence tags, at least about 5×10^6 qualified sequence tags, at least about 8×10^6 qualified sequence tags, at least about 10×10^6 qualified sequence tags, at least about 15×10^6 qualified sequence tags, at least about 20×10^6 qualified sequence tags, at least about 30×10^6 qualified sequence tags, at least about 40×10^6 qualified sequence tags, or at least about 50×10^6 qualified sequence tags comprising between 20 and 40bp reads are obtained from reads that map uniquely to the human reference genome.

[00139] In step **535**, all the tags obtained from sequencing the nucleic acids in the test samples are counted to determine a test sequence tag density. In one embodiment, the number of test sequence tags mapped to a sequence of interest is normalized to the known length of a sequence of interest to which they are mapped to provide a test sequence tag density. As described for the qualified samples, normalization to the known length of a sequence of interest is not required, and may be included as a step to reduce the number of digits in a number to simplify it for human interpretation. As all the mapped test sequence tags are counted in the test sample, the sequence tag density for a sequence of interest *e.g.* a clinically-relevant sequence such as chromosome 21, in the test samples is determined, as are the sequence tag densities for additional sequences that correspond to at least one normalizing sequence identified in the qualified samples.

[00140] In step **550**, based on the identity of at least one normalizing sequence in the qualified samples, a test sequence dose is determined for a sequence of interest in the test sample. The sequence dose *e.g.* chromosome dose, for a sequence of interest in a test sample is a ratio of the sequence tag density determined for the sequence of interest in the test sample and the sequence tag density of at least one normalizing sequence determined in the test sample, wherein the normalizing sequence in the test sample corresponds to the normalizing sequence identified in the qualified samples for the particular sequence of interest. For example, if the normalizing sequence identified for chromosome 21 in the qualified samples is determined to be chromosome 14, then the test sequence dose for chromosome 21 (sequence of interest) is determined as the ratio of the sequence tag density for chromosome 21 in and the sequence tag density for chromosome 14 each determined in the test sample. Similarly, chromosome doses for chromosomes 13, 18, X, Y, and other chromosomes associated with chromosomal aneuploidies are determined. As described previously, a sequence of interest can be part of a chromosome *e.g.* a chromosome segment. Accordingly, the dose for a chromosome segment can be determined as the ratio of the sequence tag density determined for the segment in the test sample and the sequence tag density for the normalizing chromosome

segment in the test sample, wherein the normalizing segment in the test sample corresponds to the normalizing segment identified in the qualified samples for the particular segment of interest.

[00141] In step 555, threshold values are derived from standard deviation values established for a plurality of qualified sequence doses. Accurate classification depends on the differences between probability distributions for the different classes *i.e.* type of aneuploidy. Preferably, thresholds are chosen from empirical distribution for each type of aneuploidy *e.g.* trisomy 21. Possible threshold values that were established for classifying trisomy 13, trisomy 18, trisomy 21, and monosomy X aneuploidies as described in the Examples, which describe the use of the method for determining chromosomal aneuploidies by sequencing cfDNA extracted from a maternal sample comprising a mixture of fetal and maternal nucleic acids.

[00142] In step 560, the copy number variation of the sequence of interest *e.g.* chromosomal or partial aneuploidy, is determined in the test sample by comparing the test sequence dose for the sequence of interest to at least one threshold value established from the qualified sequence doses.

[00143] In step 560, the calculated dose for a test sequence of interest is compared to that set as the threshold values that are chosen according to a user-defined threshold of reliability to classify the sample as a “normal” an “affected” or a “no call” in step 565. The “no call” samples are samples for which a definitive diagnosis cannot be made with reliability.

[00144] Another embodiment of the invention provides a method for providing prenatal diagnosis of a fetal chromosomal aneuploidy in a biological sample comprising fetal and maternal nucleic acid molecules. The diagnosis is made based on receiving the data from sequencing at least a portion of the mixture of the fetal and maternal nucleic acid molecules derived from a biological test sample *e.g.* a maternal plasma sample, computing from the sequencing data a normalizing chromosome dose for one or more chromosomes of interest, determining a statistically significant difference between the normalizing chromosome dose for the chromosome of interest in the test sample and a threshold value established in a plurality of qualified (normal) samples, and providing the prenatal diagnosis based on the statistical difference. As described in step 565 of the method, a diagnosis of normal or affected is made. A “no call” is provided in the event that the diagnosis for normal or affected cannot be made with confidence.

[00145] Quantification of the number of sequence reads aligning to each chromosome for determining chromosomal aneuploidies can also be achieved by normalizing the median number of sequence tags for a chromosome of interest to the median number of tags for each of the other autosomes (Fan *et al.*, Proc Natl Acad Sci 105:16266-16271 [2008]). Alternatively, the number of unique reads aligning to each chromosome is compared to the total number of reads aligning to all chromosomes to derive a percent genomic representation for each chromosome. A “z score” is generated to represent the difference between the percent genomic representation of the chromosome of interest and the mean percent representation for the same chromosome between a euploid control group, divided by the standard deviation (Chiu *et al.*, Clin Chem 56:459-463 [2010]).

Determination of fetal fraction

[00146] The determination of the fetal fraction is based on the total number of tags that map to the first allele and the total number of tags that map to second allele at an informative polymorphic site *e.g.* a SNP, contained in a reference genome. For example, the reference genome is the human reference genome NCBI36/hg18 sequence, or the reference genome comprises the human reference genome NCBI36/hg18 sequence and an artificial target sequences genome, which includes the target polymorphic sequences. For example, the artificial target genome encompasses polymorphic sequences that comprise SNPs rs560681, rs1109037, rs9866013, rs13182883,

rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. In another example, the artificial genome includes the polymorphic target sequences of SEQ ID NOs:1-56 (see Example 5). In another example, the artificial genome comprises polymorphic sequences that comprise tandem SNPs rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672. In another example, the artificial target genome encompasses polymorphic sequences that comprise STRs selected from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, D2S1338, D1S1677, D2S441, D4S2364, D10S1248, D14S1434, D22S1045, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627, and D1GATA113. The composition of the artificial target sequences genome will vary depending on the polymorphic sequences that are used for determining the fetal fraction. Accordingly, an artificial target sequences genome is not limited to the SNP or STR sequences exemplified herein.

[00147] The informative polymorphic site *e.g.* SNP, is identified by the difference in the allelic sequences and the amount of each of the possible alleles. Fetal cfDNA is present at a concentration that is <10% of the maternal cfDNA. Thus, the presence of a minor contribution of an allele to the mixture of fetal and maternal nucleic acids relative to the major contribution of the maternal allele can be assigned to the fetus. Alleles that are derived from the maternal genome are herein referred to as major alleles, and alleles that are derived from the fetal genome are herein referred to as minor alleles. Alleles that are represented by similar levels of mapped sequence tags represent maternal alleles. The results of an exemplary multiplex amplification of target nucleic acids comprising SNPs and derived from a maternal plasma sample is shown in **Figure 6**. Informative SNPs are discerned from the single nucleotide change at a predetermined polymorphic site, and fetal alleles are discerned by their relative minor contribution to the mixture of fetal and maternal nucleic acids in the sample when compared to the major contribution to the mixture by the maternal nucleic acids. Accordingly, the relative abundance of fetal cfDNA in the maternal sample is determined as a parameter of the total number of unique sequence tags mapped to the target nucleic acid sequence on a reference genome for each of the two alleles of the predetermined polymorphic site. In one embodiment, the fraction of fetal nucleic acids in the mixture of fetal and maternal nucleic acids is calculated for each of the informative allele (allele_x) as follows:

$$\% \text{ fetal fraction allele}_x = ((\sum \text{Fetal sequence tags for allele}_x) / (\sum \text{Maternal sequence tags for allele}_x)) \times 100$$

and fetal fraction for the sample is calculated as the average of the fetal fraction of all of the informative alleles. Optionally, the fraction of fetal nucleic acids in the mixture of fetal and maternal nucleic acids is calculated for each of the informative allele (allele_x) as follows:

$$\% \text{ fetal fraction allele}_x = ((2 \times \sum \text{Fetal sequence tags for allele}_x) / (\sum \text{Maternal sequence tags for allele}_x)) \times 100,$$

to compensate for the presence of 2 fetal alleles, one being masked by the maternal background.

[00148] The percent fetal fraction is calculated for at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40 or more informative alleles. In one embodiment, the fetal fraction is the average fetal fraction determined for at least 3 informative alleles.

[00149] In one embodiment, the step of enriching the mixture of fetal and maternal nucleic acids for polymorphic target nucleic acids comprises amplifying the target nucleic acids in a portion of a test sample *e.g.* a plasma test sample, and combining all or a portion of the amplified product with the remaining plasma test sample. The embodiment of the method **200** is depicted in flowchart provided in **Figure 2**. In step **210**, a test sample *e.g.* a biological fluid sample such as a blood sample, is obtained from a pregnant woman, and in step **220** a portion of the cfDNA contained in the plasma fraction of the blood sample is used for amplifying target nucleic acids comprising polymorphic sites *e.g.* SNPs. In one embodiment, at least about 1%, at least about 1.5%, at least about 2% at least about 10% of the maternal plasma was used to amplify the target nucleic acids. In step **230**, a portion or all of the amplified target nucleic acids is combined with the mixture of fetal and maternal cfDNA present in the maternal sample, and the combined cfDNA and amplified nucleic acids are purified in step **240**, and used for preparing a library that was sequenced in step **250**. The library was prepared from purified cfDNA and comprising at least about 10%, at least about 15%, at least about 20%, at least about 25% , at least about 30% , at least about 35% , at least about 40% , at least about 45%, or at least about 50% amplified product. In step **260**, the data from the sequencing runs is analyzed and the simultaneous determination of the fetal fraction and presence or absence of aneuploidy is made.

[00150] In one embodiment, the step of enriching the mixture of fetal and maternal nucleic acids for polymorphic target nucleic acids comprises a plurality of polymorphic target nucleic acids in a portion of a mixture of fetal and maternal nucleic acids purified from a maternal test sample. In one embodiment, a portion of a mixture of fetal and maternal nucleic acids *e.g.* cfDNA, purified from a maternal plasma sample is used for amplifying polymorphic nucleic acid sequences, and a portion of the amplified product is combined with the unamplified mixture of purified fetal and maternal nucleic acids *e.g.* cfDNA (see **Figure 3**). The embodiment of the method **300** is depicted in flowchart provided in **Figure 3**. In step **310**, a test sample *e.g.* a biological fluid sample such as a blood sample, comprising a mixture of fetal and maternal nucleic acids is obtained from a pregnant woman, and the mixture of fetal and maternal nucleic acids is purified from the plasma fraction in step **320**. As described above, methods for the separation of cell-free DNA from plasma are well-known. In step **330**, a portion of the cfDNA contained in the purified sample is used for amplifying target nucleic acids comprising polymorphic sites *e.g.* SNPs. At least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, or at least about 50% of purified cfDNA is used for amplifying the target nucleic acids. Preferably, amplification of the target sequences can be performed by any method that uses PCR or variations of the method including but not limited to asymmetric PCR, helicase-dependent amplification, hot-start PCR, qPCR, solid phase PCR, and touchdown PCR. In step **340**, a portion *e.g.* at least about 0.01% of the

amplified product is combined with the unamplified purified cfDNA sample, and the mixture of amplified and unamplified fetal and maternal nucleic acids is sequenced in step **350**. In one embodiment, sequencing is performed using any one of the NGS technologies. In step **360**, the data from the sequencing runs is analyzed and the simultaneous determination of the fetal fraction and presence or absence of aneuploidy is made as described in step **140** of the embodiment depicted in Figure 1.

[00151] In another embodiment, the step **120** of enriching the mixture of fetal and maternal nucleic acids for polymorphic target nucleic acids comprises combining at least a portion of a first sequencing library of unamplified fetal and maternal nucleic acid molecules with at least a portion of a second sequencing library of amplified polymorphic target nucleic acids. Thus, the sample that is enriched is the library sample that is prepared for sequencing (**Figure 4**). Enrichment of the library sample for the target nucleic acids is accomplished by methods that comprise specifically amplifying the nucleic acid sequences that comprise the polymorphic site. In step **410**, a test sample *e.g.* a biological fluid sample such as a blood sample, comprising a mixture of fetal and maternal nucleic acids is obtained from a pregnant woman, and the mixture of fetal and maternal nucleic acids is purified from the plasma fraction in step **420**. In step **430**, a portion of the cfDNA contained in the purified sample is used for amplifying target nucleic acids comprising polymorphic sites *e.g.* SNPs. At least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, or at least about 30% of the purified cfDNA is used for amplifying target nucleic acid sequences. Preferably, amplification of the target sequences can be performed by any method that uses PCR or variations of the method including but not limited to asymmetric PCR, helicase-dependent amplification, hot-start PCR, qPCR, solid phase PCR, and touchdown PCR. In step **440**, the amplified target nucleic acids comprising the polymorphic sites *e.g.* SNPs, are used to prepare a target nucleic acid sequencing library. Similarly, the portion of purified unamplified cfDNA is used to prepare a primary sequencing library in step **450**. In step **460**, a portion of the target library is combined with the primary library generated from the unamplified mixture of nucleic acids, and the mixture of fetal and maternal nucleic acids comprised in the two libraries is sequenced in step **470**. The enriched library comprises at least about 5%, at least about 10%, at least about 15%, at least about 20%, or at least about 25% of the target library. In step **480**, the data from the sequencing runs is analyzed and the simultaneous determination of the fetal fraction and presence or absence of aneuploidy is made as described in step **140** of the embodiment depicted in Figure 1.

Determination of aneuploidies for prenatal diagnoses

[00152] Cell-free fetal DNA and RNA circulating in maternal blood can be used for the early non-invasive prenatal diagnosis (NIPD) of an increasing number of genetic conditions, both for pregnancy management and to aid reproductive decision-making. The presence of cell-free DNA circulating in the bloodstream has been known for over 50 years. More recently, presence of small amounts of circulating fetal DNA was discovered in the maternal bloodstream during pregnancy (Lo *et al.*, Lancet 350:485-487 [1997]). Thought to originate from dying placental cells, cell-free fetal DNA (cfDNA) has been shown to consist of short fragments typically fewer than 200 bp in length (Chan *et al.*, Clin Chem 50:88-92 [2004]), which can be discerned as early as 4 weeks gestation (Illanes *et al.*, Early Human Dev 83:563-566 [2007]), and known to be cleared from the maternal circulation within hours of delivery (Lo *et al.*, Am J Hum Genet 64:218-224 [1999]). In addition to cfDNA, fragments of cell-free fetal RNA (cfRNA) can also be discerned in the maternal bloodstream, originating from genes that are transcribed in the fetus or placenta. The extraction and subsequent analysis of these fetal genetic elements from a maternal blood sample offers novel opportunities for NIPD.

[00153] The present method is a polymorphism-independent method that for use in NIPD and that does not require that the fetal cfDNA be distinguished from the maternal cfDNA to enable the determination of a fetal aneuploidy. In some embodiments, the aneuploidy is a complete chromosomal trisomy or monosomy, or a partial trisomy or monosomy. Partial aneuploidies are caused by loss or gain of part of a chromosome, and encompass chromosomal imbalances resulting from unbalanced translocations, unbalanced inversions, deletions and insertions. By far, the most common known aneuploidy compatible with life is trisomy 21 *i.e.* Down Syndrome (DS), which is caused by the presence of part or all of chromosome 21. Rarely, DS can be caused by an inherited or sporadic defect whereby an extra copy of all or part of chromosome 21 becomes attached to another chromosome (usually chromosome 14) to form a single aberrant chromosome. DS is associated with intellectual impairment, severe learning difficulties and excess mortality caused by long-term health problems such as heart disease. Other aneuploidies with known clinical significance include Edward syndrome (trisomy 18) and Patau Syndrome (trisomy 13), which are frequently fatal within the first few months of life. Abnormalities associated with the number of sex chromosomes are also known and include monosomy X *e.g.* Turner syndrome (XO), and triple X syndrome (XXX) in female births and Klinefelter syndrome (XXY) and XYY syndrome in male births, which are all associated with various phenotypes including sterility and reduction in intellectual skills. The method of the invention can be used to diagnose these and other chromosomal abnormalities prenatally.

[00154] According to embodiments of the present invention the trisomy determined by the present invention is selected from trisomy 21 (T21; Down Syndrome), trisomy 18 (T18; Edward's Syndrome), trisomy 16 (T16), trisomy 22 (T22; Cat Eye Syndrome), trisomy 15 (T15; Prader Willi Syndrome), trisomy 13 (T13; Patau Syndrome), trisomy 8 (T8; Warkany Syndrome) and the XXY (Klinefelter Syndrome), XYY, or XXX trisomies. It will be appreciated that various other trisomies and partial trisomies can be determined in fetal cfDNA according to the teachings of the present invention. These include, but not limited to, partial trisomy 1q32-44, trisomy 9 p, trisomy 4 mosaicism, trisomy 17p, partial trisomy 4q26-qter, trisomy 9, partial 2p trisomy, partial trisomy 1q, and/or partial trisomy 6p/monosomy 6q.

[00155] The method of the present invention can be also used to determine chromosomal monosomy X, and partial monosomies such as, monosomy 13, monosomy 15, monosomy 16, monosomy 21, and monosomy 22, which are known to be involved in pregnancy miscarriage. Partial monosomy of chromosomes typically involved in complete aneuploidy can also be determined by the method of the invention. Monosomy 18p is a rare chromosomal disorder in which all or part of the short arm (p) of chromosome 18 is deleted (monosomic). The disorder is typically characterized by short stature, variable degrees of mental retardation, speech delays, malformations of the skull and facial (craniofacial) region, and/or additional physical abnormalities. Associated craniofacial defects may vary greatly in range and severity from case to case. Conditions caused by changes in the structure or number of copies of chromosome 15 include Angelman Syndrome and Prader-Willi Syndrome, which involve a loss of gene activity in the same part of chromosome 15, the 15q11-q13 region. It will be appreciated that several translocations and microdeletions can be asymptomatic in the carrier parent, yet can cause a major genetic disease in the offspring. For example, a healthy mother who carries the 15q11-q13 microdeletion can give birth to a child with Angelman syndrome, a severe neurodegenerative disorder. Thus, the present invention can be used to identify such a deletion in the fetus. Partial monosomy 13q is a rare chromosomal disorder that results when a piece of the long arm (q) of chromosome 13 is missing (monosomic). Infants born with partial monosomy 13q may exhibit low birth weight, malformations of the head and face (craniofacial region), skeletal abnormalities (especially of the hands and feet), and other physical abnormalities. Mental retardation is characteristic of this condition. The mortality rate during

infancy is high among individuals born with this disorder. Almost all cases of partial monosomy 13q occur randomly for no apparent reason (sporadic). 22q11.2 deletion syndrome, also known as DiGeorge syndrome, is a syndrome caused by the deletion of a small piece of chromosome 22. The deletion (22 q11.2) occurs near the middle of the chromosome on the long arm of one of the pair of chromosome. The features of this syndrome vary widely, even among members of the same family, and affect many parts of the body. Characteristic signs and symptoms may include birth defects such as congenital heart disease, defects in the palate, most commonly related to neuromuscular problems with closure (velo-pharyngeal insufficiency), learning disabilities, mild differences in facial features, and recurrent infections. Microdeletions in chromosomal region 22q11.2 are associated with a 20 to 30-fold increased risk of schizophrenia. In one embodiment, the method of the invention is used to determine partial monosomies including but not limited to monosomy 18p, partial monosomy of chromosome 15 (15q11-q13), partial monosomy 13q, and partial monosomy of chromosome 22 can also be determined using the method.

[00156] The method of the invention can be also used to determine any aneuploidy if one of the parents is a known carrier of such abnormality. These include, but not limited to, mosaic for a small supernumerary marker chromosome (SMC); t(11;14)(p15;p13) translocation; unbalanced translocation t(8;11)(p23.2;p15.5); 11q23 microdeletion; Smith-Magenis syndrome 17p11.2 deletion; 22q13.3 deletion; Xp22.3 microdeletion; 10p14 deletion; 20p microdeletion, DiGeorge syndrome [del(22)(q11.2q11.23)], Williams syndrome (7q11.23 and 7q36 deletions); 1p36 deletion; 2p microdeletion; neurofibromatosis type 1 (17q11.2 microdeletion), Yq deletion ; Wolf-Hirschhorn syndrome (WHS, 4p16.3 microdeletion); 1p36.2 microdeletion; 11q14 deletion; 19q13.2 microdeletion; Rubinstein-Taybi (16 p13.3 microdeletion); 7p21 microdeletion; Miller-Dieker syndrome (17p13.3), 17p11.2 deletion; and 2q37 microdeletion.

Compositions and kits

[00157] Compositions comprising primers for amplifying polymorphic sites are provided to enable the quantification of fetal fraction and aneuploidy by sequencing mixtures of fetal and maternal nucleic acids *e.g.* cfDNA, present in a sample. Preferably, the sample is a maternal blood plasma sample. In one embodiment, the composition includes primers for amplifying polymorphic target nucleic acids that each comprise at least one SNP. The at least one SNP is selected from SNPs rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs 4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022. The corresponding sets of primers for amplifying the SNPs are PROVIDED IN Example 3 and disclosed as SEQ ID NOs; 57-112.

[00158] In another embodiment, the composition includes primers for amplifying polymorphic target nucleic acids that each comprise at least one tandem SNP. In one embodiment,, the composition includes primers for amplifying the exemplary tandem SNPs disclosed herein, and the composition comprises the corresponding exemplary primers of SEQ ID NOS:57-112.

[00159] In another embodiment, the composition includes primers for amplifying polymorphic target nucleic acids that each comprise at least one STR. Exemplary STRs include CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 and D1GATA113, which can be amplified by the corresponding sets of primers provided in Example 5 (Tables 5 and 6) and disclosed as SEQ ID NOs; 113-196.

[00160] The compositions of the invention can be included in kits for massively parallel sequencing mixtures of fetal and maternal nucleic acid molecules *e.g.* cfDNA, present in a maternal sample *e.g.* a plasma sample. The kits comprise a composition comprising at least one set of primers for amplifying at least one polymorphic target nucleic acid in said fetal and maternal nucleic acid molecules. Polymorphic nucleic acids can comprise a SNP or an STR. Sequencing methods are NGS methods of single nucleic acid molecules or clonally amplified nucleic acid molecules. The NGS methods are massively parallel sequencing methods including pyrosequencing, sequencing by synthesis with reversible dye terminators, real-time sequencing, or sequencing by oligonucleotide probe ligation.

[00161] The present invention is described in further detail in the following Examples which are not in any way intended to limit the scope of the invention as claimed. The attached Figures are meant to be considered as integral parts of the specification and description of the invention. The following examples are offered to illustrate, but not to limit the claimed invention.

EXPERIMENTAL

Example 1

Sample processing and cfDNA extraction

[00162] Peripheral blood samples were collected from pregnant women in their first or second trimester of pregnancy and who were deemed at risk for fetal aneuploidy. Informed consent was obtained from each participant prior to the blood draw. Blood was collected before amniocentesis or chorionic villus sampling. Karyotype analysis was performed using the chorionic villus or amniocentesis samples to confirm fetal karyotype.

[00163] Peripheral blood drawn from each subject was collected in ACD tubes. One tube of blood sample (approximately 6-9 mL/tube) was transferred into one 15-mL low speed centrifuge tube. Blood was centrifuged at 2640 rpm, 4°C for 10 min using Beckman Allegra 6 R centrifuge and rotor model GA 3.8.

[00164] For cell-free plasma extraction, the upper plasma layer was transferred to a 15-ml high speed centrifuge tube and centrifuged at 16000 x g, 4°C for 10 min using Beckman Coulter Avanti J-E centrifuge, and JA-14 rotor. The two centrifugation steps were performed within 72 h after blood collection. Cell-free plasma comprising cfDNA was stored at -80°C and thawed only once before amplification of plasma cfDNA or for purification of cfDNA.

[00165] Purified cell-free DNA (cfDNA) was extracted from cell-free plasma using the QIAamp Blood DNA Mini kit (Qiagen) essentially according to the manufacturer's instruction. One milliliter of buffer AL and 100 µl of Protease solution were added to 1 ml of plasma. The mixture was incubated for 15 minutes at 56°C. One milliliter of 100% ethanol was added to the plasma digest. The resulting mixture was transferred to QIAamp mini columns that were assembled with VacValves and VacConnectors provided in the QIAvac 24 Plus column assembly (Qiagen). Vacuum was applied to the samples, and the cfDNA retained on the column filters was washed under vacuum with 750µl of buffer AW1, followed by a second wash with 750µl of buffer AW24. The column was centrifuged at 14,000 RPM for 5 minutes to remove any residual buffer from the filter. The cfDNA was eluted with buffer AE by centrifugation at 14,000 RPM, and the concentration determined using Qubit™ Quantitation Platform (Invitrogen).

Example 2

Preparation and sequencing of primary and enriched sequencing libraries

a. Preparation of sequencing libraries

[00166] All sequencing libraries i.e. primary and enriched libraries, were prepared from approximately 2 ng of purified cfDNA that was extracted from maternal plasma. Library preparation was performed using reagents of the NEBNext™ DNA Sample Prep DNA Reagent Set 1 (Part No. E6000L; New England Biolabs, Ipswich, MA), for Illumina® as follows. Because cell-free plasma DNA is fragmented in nature, no further fragmentation by nebulization or sonication was done on the plasma DNA samples. The overhangs of approximately 2 ng purified cfDNA fragments contained in 40µl were converted into phosphorylated blunt ends according to the NEBNext® End Repair Module by incubating in a 1.5ml microfuge tube the cfDNA with 5µl 10X phosphorylation buffer, 2µl deoxynucleotide solution mix (10 mM each dNTP), 1µl of a 1:5 dilution of DNA Polymerase I, 1µl T4 DNA Polymerase and 1µl T4 Polynucleotide Kinase provided in the NEBNext™ DNA Sample Prep DNA Reagent Set 1 for 15 minutes at 20°C. The enzymes were then heat inactivated by incubating the reaction mixture at 75°C for 5 minutes. The mixture was cooled to 4°C, and dA tailing of the blunt-ended DNA was accomplished using 10µl of the dA-tailing master mix containing the Klenow fragment (3' to 5' exo minus) (NEBNext™ DNA Sample Prep DNA Reagent Set 1), and incubating for 15 minutes at 37°C. Subsequently, the Klenow fragment was heat inactivated by incubating the reaction mixture at 75°C for 5 minutes. Following the inactivation of the Klenow fragment, 1µl of a 1:5 dilution of Illumina Genomic Adaptor Oligo Mix (Part No. 1000521; Illumina Inc., Hayward, CA) was used to ligate the Illumina adaptors (Non-Index Y-Adaptors) to the dA-tailed DNA using 4µl of the T4 DNA ligase provided in the NEBNext™ DNA Sample Prep DNA Reagent Set 1, by incubating the reaction mixture for 15 minutes at 25°C. The mixture was cooled to 4°C, and the adaptor-ligated cfDNA was purified from unligated adaptors, adaptor dimers, and other reagents using magnetic beads provided in the Agencourt AMPure XP PCR purification system (Part No. A63881; Beckman Coulter Genomics, Danvers, MA). Eighteen cycles of PCR were performed to selectively enrich adaptor-ligated cfDNA (25µl) using Phusion® High-Fidelity Master Mix (25µl; Finnzymes, Woburn, MA) and Illumina's PCR primers (0.5 µM each) complementary to the adaptors (Part No. 1000537 and 1000537). The adaptor-ligated DNA was subjected to PCR (98°C for 30 seconds; 18 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 30; final extension at 72°C for 5 minutes, and hold at 4°C) using Illumina Genomic PCR Primers (Part Nos. 100537 and 1000538) and the Phusion HF PCR Master Mix provided in the NEBNext™ DNA Sample Prep DNA Reagent Set 1, according to the manufacturer's instructions. The amplified product was purified using the Agencourt AMPure XP PCR purification system (Agencourt Bioscience Corporation, Beverly, MA) according to the manufacturer's instructions available at www.beckmangenomics.com/products/AMPureXPProtocol_000387v001.pdf. The purified amplified product was eluted in 40µl of Qiagen EB Buffer, and the concentration and size distribution of the amplified libraries was analyzed using the Agilent DNA 1000 Kit for the 2100 Bioanalyzer (Agilent technologies Inc., Santa Clara, CA).

b. Sequencing

[00167] Sequencing of library DNA was performed using the Genome Analyzer II (Illumina Inc., San Diego, CA, USA) according to standard manufacturer protocols. Copies of the protocol for whole genome sequencing using Illumina/Solexa technology may be found at BioTechniques.RTM. Protocol Guide 2007 Published December 2006: p 29, and on the world wide web at biotechniques.com/default.asp?page=protocol&subsection=article_display&id=112378. The DNA library was diluted to 1nM and denatured. Library DNA (5pM) was subjected to cluster amplification according to the procedure described in Illumina's

Cluster Station User Guide and Cluster Station Operations Guide, available on the world wide web at illumina.com/systems/genome_analyzer/cluster_station.ilmn. The amplified DNA was sequenced using Illumina's Genome Analyzer II to obtain single-end reads of 36bp. Only about 30 bp of random sequence information are needed to identify a sequence as belonging to a specific human chromosome. Longer sequences can uniquely identify more particular targets. In the present case, a large number of 36 bp reads were obtained, covering approximately 10% of the genome.

Example 3

Analysis of sequencing data for the determination of aneuploidy and fetal fraction

a. Analysis of sequencing data for the determination of aneuploidy

[00168] Upon completion of sequencing of the sample, the Illumina "Sequencer Control Software" transferred image and base call files to a Unix server running the Illumina "Genome Analyzer Pipeline" software version 1.51. The Illumina "Gerald" program was run to align sequences *i.e.* 36bp reads, to the hg18 reference human genome provided by National Center for Biotechnology Information (NCBI36/hg18, available on the world wide web at genome.ucsc.edu/cgi-bin/hgGateway?org=Human&db=hg18&hgsid=166260105). The sequence data generated from the above procedure that uniquely aligned to the genome was read from Gerald output (export.txt files) by a program (c2c.pl) running on a computer running the Linux operating system. Sequence alignments with base mismatches were allowed and included in alignment counts only if they aligned uniquely to the genome. Sequence alignments with identical start and end coordinates (duplicates) were excluded.

[00169] Between about 15 and 25 million 36 bp tags with 2 or less mismatches were mapped uniquely to the human genome for each sample. All mapped tags were counted and included in the calculation of chromosome doses in both test and qualifying samples. Regions extending from base 0 to base 2×10^6 , base 10×10^6 to base 13×10^6 , and base 23×10^6 to the end of chromosome Y, were specifically excluded from the analysis because tags derived from either male or female fetuses map to these regions of the Y-chromosome.

b. Analysis of sequencing data for the determination of fetal fraction

[00170] Concomitant to the analysis for determining aneuploidy, the sequencing data was analyzed to determine the fetal fraction. Following the transfer of the image and base call files to the Unix server running the Illumina "Genome Analyzer Pipeline" software version 1.51 as described in **a.**, the 36bp reads were aligned to a 'SNP genome' using the BOWTIE program. The SNP genome was identified as the grouping of the 30 DNA sequences *i.e.* SEQ ID NOS: 1-30, that encompass the alleles of the 15 SNP disclosed in Table 5 in Example 5. Only reads that mapped uniquely to the SNP genome were used for the analysis of fetal fraction. Reads that matched perfectly to the SNP genome were counted as tags and filtered. Of the remaining reads, only reads having one or two mismatches were counted as tags and included in the analysis. Tags mapped to each of the SNP alleles were counted, and the fetal fraction was determined as described in Example 6.

Example 4

Identification of Normalizing Chromosomes for Determining Aneuploidy

[00171] To identify normalizing chromosomes to be used in determining chromosome doses and subsequent presence or absence of aneuploidy, plasma cfDNA was obtained from peripheral blood of 48 volunteer pregnant as described in Example 1, and sequenced as described in Example 2. The sequencing data provided in this example was obtained from sequencing a library constructed from fetal and maternal cfDNA that had been enriched for target

nucleic acids comprised in a second sequencing library that had been constructed from amplified sequences containing SNPs as described below.

[00172] The total number of sequence tags that were mapped to each chromosome in the reference genome (sequence tag density) was determined. Alternatively, the number of mapped sequence tags may be normalized to the length of the chromosome to generate a sequence tag density ratio. The normalization to chromosome length is not a required step, and can be performed solely to reduce the number of digits in a number to simplify it for human interpretation. Chromosome lengths that can be used to normalize the sequence tags counts can be the lengths provided on the world wide web at genome.ucsc.edu/goldenPath/stats.html#hg18.

[00173] Table 1 provides the computed ratio for chromosomes X, and Y, and autosomes 1-22 in an exemplary cfDNA sample (11351; 46,XY).

TABLE 1
Sequence Tag Density for Chromosomes 1-22, X and Y (n=1; sample 11351, 46 XY)

Chromosome Name	Sequence Tag Density
chr1	1,857,858
chr2	1,910,676
chr3	1,562,572
chr4	1,376,498
chr5	1,383,453
chr6	1,317,821
chr7	1,192,136
chr8	1,162,856
chr9	914,624
chr10	1,112,763
chr11	1,093,028
chr12	1,051,209
chr13	717,684
chr14	710,878
chr15	675,596
chr16	683,529
chr17	647,571
chr18	615,140
chr19	432,191
chr20	557,068
chr21	284,701
chr22	305,365
chrX	1,060,456
chrY	5380

[00174] The resulting sequence tag density for each chromosome was related to the sequence tag density of each of the remaining chromosomes to derive a qualified chromosome dose, which was calculated as the ratio of the sequence tag density for the chromosome of interest *e.g.* chromosome 21, and the sequence tag density of each of the remaining chromosomes *i.e.* chromosomes 1-20, 22 and X. Chromosomes doses were determined for all chromosomes in all samples, and the average doses for chromosomes of interest 13, 18, 21, X and Y in the qualified samples are provided in Table 2, and depicted in Figures 7-11. Figures 7-11 also depict the chromosome doses for the test samples. The chromosome doses for each of the chromosomes of interest in the qualified samples provides a measure of the variation in the total number of mapped sequence tags for each chromosome of interest relative to that of each of the remaining chromosomes. Thus, qualified chromosome doses can identify the chromosome or a group of chromosomes *i.e.* normalizing chromosome, that has a variation among samples that is closest to the variation of the chromosome of interest, and that would serve as ideal sequences for normalizing values for further statistical evaluation. Figures 12 and 13 depict the calculated average chromosome doses determined in a population of qualified samples for chromosomes 13, 18, and 21, and chromosomes X and Y.

[00175] In some instances, the best normalizing chromosome, may not have the least variation, but may have a distribution of qualified doses that best distinguishes a test sample or samples from the qualified samples *i.e.* the best normalizing chromosome may not have the lowest variation, but may have the greatest differentiability. Thus, differentiability accounts for the variation in chromosome dose and the distribution of the doses in the qualified samples.

[00176] Tables 3 and 4 provide the coefficient of variation as the measure of variability, and student t-test values as a measure of differentiability for chromosomes 18, 21, X and Y, wherein the smallest the T-test value, the greatest the differentiability. The differentiability for chromosome 13 was determined as the ratio of difference between the mean chromosome dose in the qualified samples and the dose for chromosome 13 in the only T13 test sample, and the standard deviation of mean of the qualified dose.

[00177] The qualified chromosome doses also serve as the basis for determining threshold values when identifying aneuploidies in test samples as described in the following.

TABLE 2

Qualified Chromosome Dose for Chromosomes 13, 18, 21, X and Y (n=1; sample 11351, 46 XY)

Chromosome	chr 21	chr 18	chr 13	chr X	chrY
chr1	0.153242	0.331102	0.386296	0.570795	0.002896
chr2	0.149005	0.321949	0.375618	0.555016	0.002816
chr3	0.1822	0.393671	0.459297	0.678661	0.003443
chr4	0.20683	0.446888	0.521384	0.770401	0.003908
chr5	0.20579	0.444641	0.518763	0.766528	0.003889
chr6	0.216039	0.466786	0.544599	0.804704	0.004082
chr7	0.238816	0.515998	0.602015	0.889543	0.004513
chr8	0.244829	0.528991	0.617174	0.911941	0.004627
chr9	0.311277	0.672561	0.784677	1.159445	0.005882
chr10	0.255851	0.552804	0.644957	0.952994	0.004835
chr11	0.26047	0.562785	0.656602	0.9702	0.004922

chr12	0.270832	0.585174	0.682722	1.008797	0.005118
chr13	0.396694	0.857118	1	1.477609	0.007496
chr14	0.400492	0.865324	1.009574	1.491755	0.007568
chr15	0.421407	0.910515	1.062298	1.56966	0.007963
chr16	0.416516	0.899947	1.049969	1.551443	0.007871
chr17	0.439644	0.949919	1.108271	1.63759	0.008308
chr18	0.462823	1	1.1667	1.723926	0.008746
chr19	0.658739	1.423306	1.660571	2.453674	0.012448
chr20	0.51107	1.104246	1.288324	1.903638	0.009658
chr21	1	2.160653	2.520834	3.724806	0.018897
chr22	0.93233	2.014442	2.35025	3.472749	0.017618
chrX	0.26847	0.580071	0.676769	1	0.005073
chrY	52.9184	114.3383	133.3985	197.1108	1

TABLE 3**Qualified Chromosome Dose, Variance and Differentiability for chromosomes 21 and 18**

	21 (n=35)				18 (n=40)			
	Avg	Stdev	CV	T Test	Avg	Stdev	CV	T Test
chr1	0.15332	0.002129	1.39	1.06E-10	0.32451	0.008954	2.76	2.74E-03
chr2	0.15106	0.002053	1.36	8.52E-08	0.31984	0.001783	0.56	5.32E-05
chr3	0.18654	0.004402	2.36	8.07E-07	0.39511	0.002364	0.60	1.93E-05
chr4	0.21578	0.011174	5.18	1.47E-04	0.45714	0.014794	3.24	1.37E-03
chr5	0.21068	0.005332	2.53	1.08E-06	0.44626	0.003250	0.73	3.18E-05
chr6	0.22112	0.005453	2.47	1.74E-06	0.46818	0.003434	0.73	2.24E-05
chr7	0.24233	0.002314	0.96	2.39E-08	0.51341	0.005289	1.03	1.24E-04
chr8	0.24975	0.003772	1.51	1.06E-07	0.52898	0.002161	0.41	6.32E-05
chr9	0.31217	0.003050	0.98	1.60E-09	0.66100	0.014413	2.18	8.17E-04
chr10	0.25550	0.003164	1.24	2.42E-11	0.54091	0.013953	2.58	2.26E-03
chr11	0.26053	0.002596	1.00	1.32E-10	0.55158	0.013283	2.41	1.29E-03
chr12	0.27401	0.002061	0.75	1.40E-08	0.58032	0.007198	1.24	1.57E-04
chr13	0.41039	0.017637	4.30	3.09E-05	0.86961	0.021614	2.49	2.36E-04
chr14	0.40482	0.002908	0.72	1.10E-08	0.85732	0.011748	1.37	2.16E-04
chr15	0.41821	0.008238	1.97	1.24E-10	0.88503	0.029199	3.30	5.72E-03
chr16	0.40668	0.021232	5.22	2.91E-05	0.86145	0.056245	6.53	1.04E-01

chr17	0.42591	0.027001	6.34	5.85E-04	0.90135	0.068151	7.56	1.24E-01
chr18	0.46529	0.016239	3.49	8.02E-09				
chr19	0.63003	0.063272	10.04	3.30E-02	1.33522	0.150794	11.29	3.04E-01
chr20	0.49925	0.023907	4.79	1.65E-05	1.05648	0.064440	6.10	7.98E-02
chr21					2.06768	0.087175	4.22	5.10E-05
chr22	0.88726	0.083330	9.39	3.43E-02	1.87509	0.198316	10.58	2.43E-01
chrX	0.27398	0.016109	5.88	1.16E-04	0.58665	0.027280	4.65	7.50E-02

TABLE 4

Qualified Chromosome Dose, Variance and Differentiability for chromosomes 13, X and Y

	13 (n=47)				X (n=20)			
	Avg	Stdev	CV	Diff	Avg	Stdev	CV	T Test
chr1	0.37213	0.018589	5.00	2.41	0.58035	0.02706	4.66	5.68E-05
chr2	0.36707	0.010067	2.74	3.03	0.57260	0.01432	2.50	1.53E-09
chr3	0.45354	0.008121	1.79	3.67	0.70741	0.01126	1.59	9.04E-13
chr4	0.52543	0.005306	1.01	2.39	0.82144	0.01192	1.45	5.86E-16
chr5	0.51228	0.008273	1.61	3.95	0.79921	0.01100	1.38	2.32E-13
chr6	0.53756	0.008901	1.66	3.91	0.83880	0.01261	1.50	3.64E-13
chr7	0.58908	0.018508	3.14	2.83	0.91927	0.02700	2.94	1.86E-08
chr8	0.60695	0.015797	2.60	3.05	0.94675	0.02173	2.30	3.40E-10
chr9	0.75816	0.033107	4.37	2.59	1.18180	0.04827	4.08	9.63E-06
chr10	0.62018	0.029891	4.82	2.56	0.96642	0.04257	4.40	4.55E-05
chr11	0.63248	0.029204	4.62	2.55	0.98643	0.04222	4.28	1.82E-05
chr12	0.66574	0.023047	3.46	2.76	1.03840	0.03301	3.18	1.26E-07
chr13					1.56355	0.01370	0.88	6.33E-17
chr14	0.98358	0.035331	3.59	2.67	1.58114	0.08076	5.11	2.29E-04
chr15	1.01432	0.055806	5.50	2.39	1.53464	0.12719	8.29	2.01E-02
chr16	0.98577	0.085933	8.72	2.17	1.61094	0.14829	9.21	2.68E-02
chr17	1.03217	0.100389	9.73	2.13	1.74904	0.07290	4.17	1.62E-04
chr18	1.13489	0.040058	3.53	2.62	2.38397	0.30515	12.80	1.07E-01
chr19	1.52678	0.203732	13.34	1.98	1.88186	0.14674	7.80	1.56E-02
chr20	1.20919	0.100371	8.30	2.27	3.71853	0.22406	6.03	4.21E-04
chr21	2.38087	0.132418	5.56	2.29	3.35158	0.40246	12.01	8.66E-02

chr22	2.14557	0.271281	12.64	2.13	0.58035	0.02706	4.66	5.68E-05
chrX	0.66883	0.029157	4.36	1.04				
chr2-6	0.46965	0.006987	1.49	4.17				
chr3-6	0.50496	0.005373	1.06	5.16				
	Y (n=25)							
	Avg	Stdev	CV	T Test				
Chr 1-22, X	0.00728	0.00227	31.19	1.30E-13				

[00178] Examples of diagnoses of T21, T13, T18 and Turner syndrome obtained using the normalizing chromosomes, chromosome doses and differentiability for each of the chromosomes of interest are described in Example 6.

Example 5

Selection of autosomal SNPs for the determination of fetal fraction

[00179] A set of 28 autosomal SNPs were selected from a list of 92 SNPs (Pakstis *et al.*, Hum Genet 127:315-324 [2010]), and SNP sequences available from Applied Biosystems on the world wide web at appliedbiosystems.com, and validated for use in multiplexed PCR amplification and for massively parallel sequencing. Primers were designed to hybridize to a sequence close to the SNPs site on the cfDNA to ensure that it be included in the 36 bp read generated from the massively parallel sequencing on the Illumina Analyzer GII, and to generate amplicons of sufficient length to undergo bridge-amplification during cluster formation. Thus, primers were designed to generate amplicons that were at least 110 bp, which when combined with the universal adaptors (Illumina Inc., San Diego, CA) used for cluster amplification, resulted in DNA molecules of at least 200bp. Primer sequences were identified, and primer sets *i.e.* forward and reverse primers, were synthesized by Integrated DNA Technologies (San Diego, CA), and stored as a 1µM solution to be used for amplifying polymorphic target sequences as described in Examples 5-8. Table 5 provides the RefSNP (rs) accession ID numbers, the primers used for amplifying the target cfDNA sequence, and the sequences of the amplicons comprising the possible SNP alleles that would be generated using the primers. The SNPs given in Table 5 were used for the simultaneous amplification of 13 target sequences in a multiplexed assay. The panel provided in Table 5 is an exemplary SNP panel. Fewer or more SNPs can be employed to enrich the fetal and maternal DNA for polymorphic target nucleic acids. Additional SNPs that can be used include the SNPs given in Table 6. The SNP alleles are shown in bold and are underlined. Other SNPs that can be used to determine fetal fraction according to the present method include rs315791, rs3780962, rs1410059, rs279844, rs38882, rs9951171, rs214955, rs6444724, rs2503107, rs1019029, rs1413212, rs1031825, rs891700, rs1005533, rs2831700, rs354439, rs1979255, rs1454361, rs8037429, and rs1490413, which have been analyzed for determining fetal fraction by TaqMan PCR, and are disclosed in US Provisional applications 61/296,358 and 61/360,837.

TABLE 5
SNP Panel for the Determination of Fetal Fraction

SNP ID	Chr	Amplicon: Allele 1	Amplicon: Allele 2	Forward Primer Sequence, name and SEQ ID NO:	Reverse Primer Sequence, name and SEQ ID NO:
rs560681	1	CACATGCACAGCCA GCAACCCTGTCAGC AGGAGTTCCCACCA GTTTCTTTCTGAGAA CATCTGTTTCAGGTTT CTCTCCATCTCT <u>A</u> TT TACTCAGGTCACAG GACCTTGGGG (SEQ ID NO:1)	CACATGCACAGCCA GCAACCCTGTCAGC AGGAGTTCCCACCA GTTTCTTTCTGAGAA CATCTGTTTCAGGTTT CTCTCCATCTCT <u>G</u> TT TACTCAGGTCACAG GACCTTGGGG (SEQ ID NO:2)	CACATGCA CAGCCAGC AACCC (rs560681_C 1_1_F; SEQ ID NO:57)	CCCCAAGGTC CTGTGACCTGA GT (rs560681_C1_1_ R; SEQ ID NO:58)
rs1109037	2	TGAGGAAGTGAGGC TCAGAGGGTAAGAA ACTTTGTCACAGAGC TGGTGGTGAGGGTG GAGATTTTACACTCC CTGCCTCCCACACCA GTTTCTCC <u>A</u> GAGTGG AAAGACTTTCATCTC GCACTGGCA (SEQ ID NO:3)	TGAGGAAGTGAGGC TCAGAGGGTAAGAA ACTTTGTCACAGAGC TGGTGGTGAGGGTG GAGATTTTACACTCC CTGCCTCCCACACCA GTTTCTCC <u>G</u> GAGTGG AAAGACTTTCATCTC GCACTGGCA (SEQ ID NO:4)	TGAGGAAG TGAGGCTC AGAGGGT (rs110937_C 2_1_F ; SEQ ID NO:59)	TGCCAGTGCG AGATGAAAGT CTTT (rs110937_C2_1_ R ; SEQ ID NO:60)
rs9866013	3	GTGCCTTCAGAACCT TTGAGATCTGATTCT ATTTTAAAGCTTCT TAGAAGAGAGATTG CAAAGTGGGTTGTTT CTCTAGCCAGACAG GGCAGG <u>C</u> AAATAGG GGTGGCTGGTGGGA TGGGA (SEQ ID NO:5)	GTGCCTTCAGAACCT TTGAGATCTGATTCT ATTTTAAAGCTTCT TAGAAGAGAGATTG CAAAGTGGGTTGTTT CTCTAGCCAGACAG GGCAGG <u>T</u> AAATAGG GGTGGCTGGTGGGA TGGGA (SEQ ID NO:6)	GTGCCTTC AGAACCTT TGAGATCT GAT (rs9866013_ C3_1_F; SEQ ID NO:61)	TCCCATCCCAC CAGCCACCC (rs9866013_C3_1_ _R; SEQ ID NO:62)
rs13182883	5	AGGTGTGTCTCTCTT TTGTGAGGGGAGGG GTCCCTTCTGGCCTA GTAGAGGGCCTGGC CTGCAGTGAGCATTC	AGGTGTGTCTCTCTT TTGTGAGGGGAGGG GTCCCTTCTGGCCTA GTAGAGGGCCTGGC CTGCAGTGAGCATTC	AGGTGTGT CTCTCTTTT GTGAGGGG (rs13182883_ C5_1_F; SEQ	CCTTTGTCCCA CCTCCCCACC (rs13182883_C5_ 1_R; SEQ ID NO:64)

		AAATCCTC <u>A</u> AGGAA CAGGGTGGGGAGGT GGGACAAAGG (SEQ ID NO:7)	AAATCCTC <u>G</u> AGGAA CAGGGTGGGGAGGT GGGACAAAGG (SEQ ID NO:8)	ID NO:63)	
rs13218440	6	CCTCGCCTACTGTGC TGTTTCTAACCATCA TGCTTTTCCCTGAAT CTCTTGAGTCTTTTT CTGCTGTGGACTGA AACTTGATCCTGAG ATTCACCTCTAGTCC CTCTG <u>A</u> GCAGCCTCC TGGAATACTCAGCT GGGATGG (SEQ ID NO:9)	CCTCGCCTACTGTGC TGTTTCTAACCATCA TGCTTTTCCCTGAAT CTCTTGAGTCTTTTT CTGCTGTGGACTGA AACTTGATCCTGAG ATTCACCTCTAGTCC CTCTG <u>G</u> GCAGCCTCC TGGAATACTCAGCT GGGATGG (SEQ ID NO:10)	CCTCGCCT ACTGTGCT GTTTCTAA CC (rs13218440_ C6_1_F; SEQ ID NO:65)	CCATCCCAGCT GAGTATTCCA GGAG (rs13218440_ C6_1_R; SEQ ID NO:66)
rs7041158	9	AATTGCAATGGTGA GAGGTTGATGGTAA AATCAAACGGA ACTTGTTATTTTGT CATTCTGATGGACT GGAACTGAGGAT TTTCAATTTCCT <u>C</u> TCCAACCAAGAC ACTTCTCACTGG (SEQ ID NO:11)	AATTGCAATGGTGA GAGGTTGATGGTAA AATCAAACGGA ACTTGTTATTTTGT CATTCTGATGGACT GGAACTGAGGAT TTTCAATTTCCT <u>T</u> TCCAACCAAGAC ACTTCTCACTGG (SEQ ID NO:12)	AATTGCAA TGGTGAGA GGTTGATG GT (SEQ ID NO:67)	CCAGTGAGAA GTGTCTTGGGT TGG (SEQ ID NO:68)
rs740598	10	GAAATGCCTTCTCAG GTAATGGAAGGTTA TCCAAATATTTTTCG TAAGTATTTCAAATA GCAATGGCTCGTCTA TGGTTAGTCTC <u>A</u> CAG CCACATTCTCAGAAC TGCTCAAACC (SEQ ID NO:13)	GAAATGCCTTCTCAG GTAATGGAAGGTTA TCCAAATATTTTTCG TAAGTATTTCAAATA GCAATGGCTCGTCTA TGGTTAGTCTC <u>G</u> CAG CCACATTCTCAGAAC TGCTCAAACC (SEQ ID NO:14)	GAAATGCC TTCTCAGG TAATGGAA GGT (SEQ ID NO:69)	GGTTTGAGCA GTTCTGAGAAT GTGGCT (SEQ ID NO:70)
rs10773760	12	ACCCAAAACACTGG AGGGGCCTCTTCTCA TTTTCGGTAGACTGC AAGTGTTAGCCGTC GGGACCAGCTTCTGT CTGGAAGTTCGTCA AATTGCAGTTA <u>A</u> GTC	ACCCAAAACACTGG AGGGGCCTCTTCTCA TTTTCGGTAGACTGC AAGTGTTAGCCGTC GGGACCAGCTTCTGT CTGGAAGTTCGTCA AATTGCAGTTA <u>G</u> GTC	ACCCAAAA CACTGGAG GGGCCT (SEQ ID NO:71)	CCCTTATCTGC TATGTGGCATA CTTGG (SEQ ID NO:72)

		CAAGTATGCCACAT AGCAGATAAGGG (SEQ ID NO:15)	CCAAGTATGCCACA TAGCAGATAAGGG (SEQ ID NO:16)		
rs4530059	14	GCACCAGAATTTAA ACAACGCTGACAAT AAATATGCAGTCGA TGATGACTTCCCAGA GCTCCAGAAGCAAC TCCAGCACAC <u>A</u> GAG AGGCGCTGATGTGC CTGTCAGGTGC (SEQ ID NO:17)	GCACCAGAATTTAA ACAACGCTGACAAT AAATATGCAGTCGA TGATGACTTCCCAGA GCTCCAGAAGCAAC TCCAGCACAC <u>G</u> GAG AGGCGCTGATGTGC CTGTCAGGTGC (SEQ ID NO:18)	GCACCAGA ATTTAAAC AACGCTGA CAA (SEQ ID NO:73)	GCACCTGACA GGCACATCAG CG (SEQ ID NO:74)
rs7205345	16	TGACTGTATACCCCA GGTGCACCCTTGGGT CATCTCTATCATAGA ACTTATCTCACAGAG TATAAGAGCTGATTT CTGTGTCTGCCT <u>C</u> TC ACACTAGACTTCCAC ATCCTTAGTGC (SEQ ID NO:19)	TGACTGTATACCCCA GGTGCACCCTTGGGT CATCTCTATCATAGA ACTTATCTCACAGAG TATAAGAGCTGATTT CTGTGTCTGCCT <u>G</u> TC ACACTAGACTTCCAC ATCCTTAGTGC (SEQ ID NO:20)	TGACTGTA TACCCAG GTGCACCC (SEQ ID NO:75)	GCACTAAGGA TGTGGAAGTCT AGTGTG (SEQ ID NO:76)
rs8078417	17	TGTACGTGGTCACCA GGGGACGCCTGGCG CTGCGAGGGAGGCC CCGAGCCTCGTGCCC CCGTGAAGCTTCAG CTCCCCTCCC <u>C</u> GGCT GTCCTTGAGGCTCTT CTCACACT (SEQ ID NO:21)	TGTACGTGGTCACCA GGGGACGCCTGGCG CTGCGAGGGAGGCC CCGAGCCTCGTGCCC CCGTGAAGCTTCAG CTCCCCTCCC <u>T</u> GGCT GTCCTTGAGGCTCTT CTCACACT (SEQ ID NO:22)	TGTACGTG GTCACCAG GGGACG (SEQ ID NO:77)	AGTGTGAGAA GAGCCTCAAG GACAGC (SEQ ID NO:78)
rs576261	19	CAGTGGACCCTGCT GCACCTTTCCTCCCC TCCCATCAACCTCTT TTGTGCCTCCCCCTC CGTGTACCACCTTCT CTGTCACCA <u>A</u> CCCTG GCCTCACAACCTCTCT CCTTTGCCAC (SEQ	CAGTGGACCCTGCT GCACCTTTCCTCCCC TCCCATCAACCTCTT TTGTGCCTCCCCCTC CGTGTACCACCTTCT CTGTCACCA <u>C</u> CCCTG GCCTCACAACCTCTCT CCTTTGCCAC (SEQ	CAGTGGAC CCTGCTGC ACCTT (SEQ ID NO:79)	GTGGCAAAGG AGAGAGTTGT GAGG (SEQ ID NO:80)

		ID NO:23)	ID NO:24)		
rs2567608	20	CAGTGGCATAGTAG TCCAGGGGCTCCTCC TCAGCACCTCCAGC ACCTTCCAGGAGGC AGCAGCGCAGGCAG AGAACCCGCTGGAA G <u>A</u> ATCGGCGGAAGT TGTCGGAGAGG (SEQ ID NO:25)	CAGTGGCATAGTAG TCCAGGGGCTCCTCC TCAGCACCTCCAGC ACCTTCCAGGAGGC AGCAGCGCAGGCAG AGAACCCGCTGGAA G <u>G</u> ATCGGCGGAAGT TGTCGGAGAGG (SEQ ID NO:26)	CAGTGGCA TAGTAGTC CAGGGGCT (SEQ ID NO:81)	CCTCTCCGACA ACTTCCGCCG (SEQ ID NO:82)

TABLE 6

SNP Panel for the Determination of Fetal Fraction

SNP ID	Chr	Amplicon: Allele 1	Amplicon: Allele 2	Forward Primer Sequence, name and SEQ ID NO:	Reverse Primer Sequence, name and SEQ ID NO:
rs430046	16	AGGTCTGGGGGCC GCTGAATGCCAAGC TGGGAATCTTAAAT GTTAAGGAACAAG GTCATACAATGAAT GGTGTGATGTAAAA GCTTGGGAGGTGAT TT <u>C</u> TGAGGGTAGGT GCTGGGTTTAATGG GAGGA (SEQ ID NO:27)	AGGTCTGGGGGGCCGC TGAATGCCAAGCTGG GAATCTTAAATGTTA AGGAACAAGGTCATA CAATGAATGGTGTGA TGTAAGCTTGGGA GGTGATTTT <u>T</u> GAGGG TAGGTGCTGGGTTTA ATGGGAGGA (SEQ ID NO:28)	AGGTCTGG GGGCCGCT GAAT (rs430046_C 1_1_F; SEQ ID NO:83)	TCCTCCCATTA AACCCAGCAC CT (rs430046_C1_1_ R; SEQ ID NO:84)
rs9951171	18	ACGGTTCTGTCCTG TAGGGGAGAAAAG TCCTCGTTGTTCTT CTGGGATGCAACAT GAGAGAGCAGCAC ACTGAGGCTTTATG G <u>A</u> TTGCCCTGCCAC AAGTGAACAGG (SEQ ID NO:29)	ACGGTTCTGTCCTGT AGGGGAGAAAAGTCC TCGTTGTTCTCTGGG ATGCAACATGAGAGA GCAGCACACTGAGGC TTTATGG <u>G</u> TTGCCCT GCCACAAGTGAACAG G (SEQ ID NO:30)	ACGGTTCT GTCCTGTA GGGGAGA (rs9951171_ C1_1_F; SEQ ID NO:85)	CCTGTTCACTT GTGGCAGGGC A (rs9951171_C1_1_ _R; SEQ ID NO:86)
rs338882	5	GCGCAGTCAGATG GGCGTGCTGGCGTC	GCGCAGTCAGATGGG CGTGCTGGCGTCTGT	GCGCAGTC AGATGGGC	TCCAGCCCTTG TCCCAAACGT

		TGTCTTCTCTCTCTC CTGCTCTCTGGCTT CATTTTTCTCTCCTT CTGTCTCACCTTCT TTCGTGTGCCTGTG CA <u>C</u> ACACACGTTTG GGACAAGGG CTGGA (SEQ ID NO:31)	CTTCTCTCTCTCCTGC TCTCTGGCTTCATTTT TCTCTCCTTCTGTCTC ACCTTCTTTTCGTGTGC CTGTGCA <u>T</u> ACACACG TTTGGGACAAGGG CTGGA (SEQ ID NO:32)	GTGC (rs338882_C 1_1_F; SEQ ID NO:87)	GT (rs338882_C1_1_ R; SEQ ID NO:88)
rs10776839	9	GCCGGACCTGCGA AATCCCAAAATGCC AAACATTCCCGCCT CACATGATCCCAGA GAGAGGGGACCCA GTGTTCCCAGCTTG CAGCTGAGGAGCC CGAG <u>G</u> TTGCCGTCA GATCAGAGCCCCA GTTGCCCG (SEQ ID NO:33)	GCCGGACCTGCGAAA TCCCAAAATGCCAAA CATTCCCGCCTCACA TGATCCCAGAGAGAG GGGACCCAGTGTTC CAGCTTGCAGCTGAG GAGCCCGAG <u>T</u> TTGCC GTCAGATCAGAGCCC CAGTTGCCCG (SEQ ID NO:34)	GCCGGACC TGCGAAAT CCCAA (rs10776839 C1_1_F; SEQ ID NO:89)	CGGGCAACTG GGGCTCTGATC (rs10776839_C1_ 1_R; SEQ ID NO:90)
rs9905977	17	AGCAGCCTCCCTCG ACTAGCTCACACTA CGATAAGGAAAATT CATGAGCTGGTGTC CAAGGAGGGCTGG GTGACTCGTGGCTC AGTCAGC <u>A</u> TCAAG ATTCCTTTCGTCTTT CCCCTCTGCC (SEQ ID NO:35)	AGCAGCCTCCCTCGA CTAGCTCACACTACG ATAAGGAAAATTCAT GAGCTGGTGTCCAAG GAGGGCTGGGTGACT CGTGGCTCAGTCAGC <u>G</u> TCAAGATTCCTTTC GTCTTTCCCCTCTGCC (SEQ ID NO:36)	AGCAGCCT CCCTCGAC TAGCT (rs9905977_ C1_1_F; SEQ ID NO:91)	GGCAGAGGGG AAAGACGAAA GGA (rs9905977_C1_1_ _R; SEQ ID NO:92)
rs1277284	4	TGGCATTGCCTGTA ATATACATAGCCAT GGTTTTTTATAGGC AATTTAAGATGAAT AGCTTCTAAACTAT AGATAAGTTTCATT ACCCAGGAAGCT GAACTATAGCTACT TT <u>A</u> CCCAAAATCAT	TGGCATTGCCTGTAA TATACATAGCCATGG TTTTTTATAGGCAATT TAAGATGAATAGCTT CTAAACTATAGATAA GTTTCATTACCCCAG GAAGCTGAACTATAG CTACTTT <u>C</u> CCCAAAA TCATTAGAATGGTGC	TGGCATTG CCTGTAAT ATACATAG (rs1277284_ C4_1_F; SEQ ID NO:93)	AAGCACCATT CTAATGATTTT GG (rs1277284_C4_1_ _R; SEQ ID NO:94)

		TAGAATGGTGCTT (SEQ ID NO:37)	TT (SEQ ID NO:38)		
rs258684	7	ATGAAGCCTTCCAC CAACTGCCTGTATG ACTCATCTGGGGAC TTCTGCTCTATACT CAAAGTGGCTTAGT CACTGCCAATGTAT TTCCATATGAGGGA CGATGATTACTAAG GAAATATAGAAAC AACAACTGATC (SEQ ID NO:39)	ATGAAGCCTTCCACC AACTGCCTGTATGAC TCATCTGGGGACTTC TGCTCTATACTCAA GTGGCTTAGTCACTG CCAATGTATTTCCAT ATGAGGGACG <u>G</u> TGAT TACTAAGGAAATATA GAAACAACAACCTGAT C (SEQ ID NO:40)	ATGAAGCC TTCCACCA ACTG (rs258684_C 7_1_F; SEQ ID NO:95)	GATCAGTTGTT GTTTCTATATT TCCTT (rs258684_C7_1_ R; SEQ ID NO:96)
rs1347696	8	ACAACAGAATCAG GTGATTGGAGAAA AGATCACAGGCCTA GGCACCCAAGGCTT GAAGGATGAAAGA ATGAAAGATGGAC GGAA <u>C</u> AAAATTAG GACCTTAATTCTTT GTTTCAGTTCAG (SEQ ID NO:41)	ACAACAGAATCAGGT GATTGGAGAAAAGAT CACAGGCCTAGGCAC CCAAGGCTTGAAGGA TGAAAGAATGAAAGA TGGACGGAA <u>G</u> AAAAT TAGGACCTTAATTCTT TGTTTCAGTTCAG (SEQ ID NO:42)	ACAACAGA ATCAGGTG ATTGGA (rs1347696_ C8_4_F; SEQ ID NO:97)	CTGAACTGAA CAAAGAATTA AGGTC (rs1347696_C8_4_ _F; SEQ ID NO:98)
rs508485	11	TTGGGGTAAATTTT CATTGTCATATGTG GAATTTAAATATAC CATCATCTACAAAG AATTCCACAGAGTT AAATATCTTAAGTT AAACACTTAAAATA AGTGTTTGCGTGAT ATTTTGATGAC <u>A</u> GAG TAAACAGAGTCTAA TTCCCACCCC (SEQ ID NO:43)	TTGGGGTAAATTTTC ATTGTCATATGTGGA ATTTAAATATACCAT CATCTACAAAGAATT CCACAGAGTTAAATA TCTTAAGTTAAACAC TTAAAATAAGTGTTT GCGTGATATTTTGAT GAT <u>A</u> GATAAACAGAG TCTAATTCCCACCCC (SEQ ID NO:44)	TTGGGGTA AATTTTCA TTGTCA (rs508485_C 11_1_F; SEQ ID NO:99)	GGGGTGGGAA TTAGACTCTG (rs508485_C11_1_ _R; SEQ ID NO100)
rs9788670	15	TGCAATTCAAATCA GGAAGTATGACCA AAAGACAGAGATC TTTTTTGGATGATC CCTAGCCTAGCAAT	TGCAATTCAAATCAG GAAGTATGACCAAAA GACAGAGATCTTTTT TGGATGATCCCTAGC CTAGCAATGCCTGGC	TGCAATTC AAATCAGG AAGTATG (rs9788670_c 15_2_F; SEQ	GCAACATCGA GGTTTGTGAG (rs9788670_c15_ 2_R; SEQ ID NO:102)

		GCCTGGCAGCCATG CAGGTGCAATGTCA ACCTTAAATAATGT ATTGCAA <u>A</u> CTCAGA GCTGACAAACCTCG ATGTTGC (SEQ ID NO:45)	AGCCATGCAGGTGCA ATGTCAACCTTAAAT AATGTATTGCAAA <u>T</u> CAGAGCTGACAAACC TCGATGTTGC (SEQ ID NO:46)	ID NO:101)	
rs8137254	22	CTGTGCTCTGCGAA TAGCTGCAGAAGTA ACTTGGGGACCCAA AATAAAGCAGAAT GCTAATGTCAAGTC CTGAGAACCAAGC CCTGGGACTCTGGT GCCATTT <u>C</u> GGATTC TCCATGAGCATGGT (SEQ ID NO:47)	CTGTGCTCTGCGAAT AGCTGCAGAAGTAAC TTGGGGACCCAAAAT AAAGCAGAATGCTAA TGTCAAGTCCTGAGA ACCAAGCCCTGGGAC TCTGGTGCCATTT <u>T</u> G GATTCTCCATGAGCA TGGT (SEQ ID NO:48)	CTGTGCTC TGCGAATA GCTG (rs8137254_c 22_2_F: SEQ ID NO:103)	ACCATGCTCAT GGAGAATCC (rs8137254_c22_ 2_R; SEQ ID NO:104)
rs3143	19	TTTTTCCAGCCAAC TCAAGGCCAAAAA AAATTTCTTAATAT AGTTATTATGCGAG GGGAGGGGAAGCA AAGGAGCACAGGT AGTCCACAGAATA <u>A</u> GACACAAGAAAC CTCAAGCTGTG (SEQ ID NO:49)	TTTTTCCAGCCAACTC AAGGCCAAAAAAAAT TTCTTAATATAGTTAT TATGCGAGGGGAGGG GAAGCAAAGGAGCA CAGGTAGTCCACAGA ATAG <u>G</u> ACACAAGAA ACCTCAAGCTGTG (SEQ ID NO:50)	TTTTTCCA GCCAACTC AAGG (rs3143_c19_ 2_F: SEQ ID NO:105)	CACAGCTTGA GGTTTCTTGTG (rs3143_c19_2_R; SEQ ID NO:106)
rs2182957	13	TCTTCTCGTCCCCT AAGCAAACAACAT CCGCTTGCTTCTGT CTGTGTAACCACAG TGAATGGGTGTGCA CGCTTG <u>A</u> TGGGCCT CTGAGCCCCTGTTG CACAAACCAGAAA (SEQ ID NO:51)	TCTTCTCGTCCCCTAA GCAAACAACATCCGC TTGCTTCTGTCTGTGT AACCACAGTGAATGG GTGTGCACGCTTG <u>G</u> T GGGCCTCTGAGCCCC TGTTGCACAAACCAG AAA (SEQ ID NO:52)	TCTTCTCG TCCCCTAA GCAA (rs2182957_c 13_1_F: SEQ ID NO:107)	TTTCTGGTTTG TGCAACAGG (rs2182957_c13_ 1_R; SEQ ID NO:108)
rs3739005	2	CACATGGGGGCATT AAGAATCGCCCAG GGAGGAGGAGGGA GAACGCGTGCTTTT	CACATGGGGGCATTA AGAATCGCCCAGGGA GGAGGAGGGAGAAC GCGTGCTTTTCACATT	CACATGGG GGCATTA GAAT (rs3739005_c	ACATCGATGA GCACAAAAAC AC (rs3739005_c2_2

		CACATTTGCATTTG AATTTTCGAGTTCC CAGGATGTGTTTTT GTGCTCATCGATGT (SEQ ID NO:53)	TGCATTTGAATTTT <u>T</u> G AGTTCCCAGGATGTG TTTTTGTGCTCATCGA TGT (SEQ ID NO:54)	2_2_F; SEQ ID NO:109)	_R; SEQ ID NO:110)
rs530022	1	GGGCTCTGAGGTGT GTGAAATAAAAAC AAATGTCCATGTCT GTCCTTTTATGGCA TTTTGGGACTTTAC ATTTCAAACATTTC AGACATGTATCACA ACACGA <u>A</u> GGAATA ACAGTTCCAGGGAT ATCT (SEQ ID NO:55)	GGGCTCTGAGGTGTG TGAAATAAAAACAAA TGTCCATGTCTGTCCT TTTATGGCATTTTGGG ACTTTACATTTCAAA CATTTTCAGACATGTA TCACAACACGAG <u>G</u> GGA ATAACAGTTCCAGGG ATATCT (SEQ ID NO:56)	GGGCTCTG AGGTGTGT GAAA (rs530022_c1 _2_F; SEQ ID NO:111)	AGATATCCCTG GAACTGTTATT CC (rs530022_c1_2_ R; SEQ ID NO:112)

Example 6

Simultaneous determination of Aneuploidy and Fetal fraction:

Enrichment of fetal and maternal nucleic acids in a cfDNA sequencing library sample

[00180] To enrich the fetal and maternal cfDNA contained in a primary sequencing library constructed using purified fetal and maternal cfDNA, a portion of a purified cfDNA sample was used for amplifying polymorphic target nucleic acid sequences, and for preparing a sequencing library of amplified polymorphic target nucleic acids, which was used to enrich the fetal and maternal nucleic acid sequences comprised in the primary library.

[00181] A primary sequencing library was prepared using purified cfDNA as described in Example 1.

[00182] A target sequencing library was prepared as follows. cfDNA contained in 5µl of purified cfDNA was amplified in a reaction volume of 50µl containing 7.5µl of a 1µM primer mix (Table 5), 10µl of NEB 5X Mastermix and 27 µl water. Thermal cycling was performed with the Gene Amp9700 (Applied Biosystems). Using the following cycling conditions: incubating at 95°C for 1 minute, followed by 30 cycles at 95°C for 20 seconds, 68°C for 1 minute, and 68°C for 30s, which was followed by a final incubation at 68°C for 5 minutes. A final hold at 4°C was added until the samples were removed for combining with the unamplified portion of the purified cfDNA sample. The amplified product was purified using the Agencourt AMPure XP PCR purification system (Part No. A63881; Beckman Coulter Genomics, Danvers, MA). A final hold at 4°C was added until the samples were removed for preparing the target library. The amplified product was analyzed with a 2100 Bioanalyzer (Agilent Technologies, Sunnyvale, CA), and the concentration of amplified product determined. One fifth of the purified amplified product was used to prepare a target sequencing library of amplified polymorphic nucleic acids as described in Example 2. The primary and the target sequencing libraries were each diluted to 10nM, and the target library was combined at a ratio of 1:9 with the sequencing library to provide an enriched sequencing library. Sequencing of the enriched library and analysis of the sequencing data was performed as described in Example 3.

a. Determination of fetal fraction

[00183] Determination of fetal fraction was performed as described in Example 5, and fetal fraction was calculated as described above *i.e.*

[00184] % fetal fraction allele_x = (($\sum$ Fetal sequence tags for allele_x) / ($\sum$ Maternal sequence tags for allele_x)) x 100

TABLE 7**Simultaneous Determination of Aneuploidy and Fetal Fraction:****Determination of Fetal Fraction**

Sample ID (karyotype)	SNP	SNP TAG COUNTS	FETAL FRACTION (%)
11409 (47, XY+21)	rs13182883.1 Chr.5 length=111 allele=A	261	4.41
	rs13182883.2 Chr.5 length=111 allele=G	5918	
	rs740598.1 Chr.10 length=114 allele=A	5545	7.30
	rs740598.2 Chr.10 length=114 allele=G	405	
	rs8078417.1 Chr.17 length=110 allele=C	8189	6.74
	rs8078417.2 Chr.17 length=110 allele=T	121470	
	rs576261.1 Chr.19 length=114 allele=A	58342	7.62
	rs576261.2 Chr.19 length=114 allele=C	4443	
Fetal Fraction (Mean±S.D.) = 6.5±1.5			
Sample ID			
95133 (47, XX+18)	rs1109037.1 Chr.2 length=126 allele=A	12229	2.15
	rs1109037.2 Chr.2 length=126 allele=G	263	
	rs13218440.1 Chr.6 length=139 allele=A	55949	3.09
	rs13218440.2 Chr.6 length=139 allele=G	1729	
	rs7041158.1 Chr.9 length=117 allele=C	7281	4.12
	rs7041158.2 Chr.9 length=117 allele=T	300	

	rs7205345.1 Chr.16 length=116 allele=C	53999	2.14
	rs7205345.2 Chr.16 length=116 allele=G	1154	
Fetal Fraction (Mean±S.D.) = 2.9±0.9			
Sample ID			
51236 (46,XY+13)	rs13218440.1 Chr.6 length=139 allele=A	1119	1.65
	rs13218440.2 Chr.6 length=139 allele=G	67756	
	rs560681.1 Chr.1 length=111 allele=A	14123	5.18
	rs560681.2 Chr.1 length=111 allele=G	732	
	rs7205345.1 Chr.16 length=116 allele=C	18176	1.63
	rs7205345.2 Chr.16 length=116 allele=G	296	
	rs9866013.1 Chr.3 length=121 allele=C	117	2.33
	rs9866013.2 Chr.3 length=121 allele=T	5024	
Fetal Fraction (Mean±S.D.) = 2.7±1.7			
Sample ID			
54430 (45,XO)	rs1109037.1 Chr.2 length=126 allele=A	19841	1.80
	rs1109037.2 Chr.2 length=126 allele=G	357	
	rs9866013.1 Chr.3 length=121 allele=C	12931	3.81
	rs9866013.2 Chr.3 length=121 allele=T	493	
	rs7041158.1 Chr.9 length=117 allele=C	2800	4.25
	rs7041158.2 Chr.9 length=117 allele=T	119	
	rs740598.1 Chr.10 length=114 allele=A	12903	4.87

	rs740598.2 Chr.10 length=114 allele=G	628	4.65
	rs10773760.1 Chr.12 length=128 allele=A	46324	
	rs10773760.2 Chr.12 length=128 allele=G	2154	
Fetal Fraction (Mean±S.D.) = 3.9±1.2			

b. Determination of aneuploidy

[00185] Determination of aneuploidy of chromosomes 21, 13, 18 and X was performed using chromosome doses as described in Example 4. Chromosome 21 dose was determined using chromosome 14 as the normalizing chromosome; chromosome 13 dose was determined using the group of chromosomes 3, 4, 5, and 6 as the normalizing chromosome; chromosome 18 dose was determined using chromosome 8 as the normalizing chromosome; and chromosome X dose was determined using chromosome 4 as the normalizing chromosome. Thresholds were calculated to be 2 standard deviations above and below the mean determined in the qualified samples.

[00186] Table 7 shows the data for the determination of fetal fraction in exemplary samples. Calculated chromosome dose values for chromosomes 21, 18, 13, X and Y in corresponding exemplary test samples are given in Tables 8, 9, 10, 11, and 12, respectively.

Trisomy 21

[00187] Table 8 provides the calculated dose for chromosome 21 in the test sample (11409). Chromosome 14 was used as the normalizing chromosomes. The calculated threshold for the positive diagnosis of T21 aneuploidy was set at 2 standard deviations from the mean of the qualified (normal) samples. A diagnosis for T21 was given based on the chromosome dose in the test sample being greater than the set threshold. All twelve of the T21 samples that were confirmed to be T21 by karyotype were identified in a population of 48 blood samples.

TABLE 8
Chromosome Dose for a T21 aneuploidy

Chromosome	Sequence Tag Density	Chromosome Dose for Chr 21	Threshold
Chr21	264,404	0.439498	0.410634
Chr14	601,605		

Trisomy 18

[00188] Table 9 provides the calculated dose for chromosome 18 in a test sample (95133). Chromosome 8 was used as the normalizing chromosome. In this instance, chromosome 8 had the lowest variability and greatest differentiability. The calculated threshold for the positive diagnosis of T18 aneuploidy was set at greater than 2 standard deviations from the mean of the qualified (non-T18) samples. A diagnosis for T18 was given based on the chromosome dose in the test sample being greater than the set threshold. Eight T18 samples were identified using chromosome doses, and were confirmed to be T18 by karyotyping.

TABLE 9**Chromosome Dose for a T18 aneuploidy**

Chromosome	Sequence Tag Density	Chromosome Dose for Chr 18	Threshold
Chr18	604,291	0.550731	0.533297
Chr8	1,097,253		

Trisomy 13

[00189] Tables 10 and 11 provide the calculated dose for chromosome 13 in a test sample (51236). The calculated threshold for the positive diagnosis of T13 aneuploidy was set at 2 standard deviations from the mean of the qualified (non-T13) samples. The chromosome dose for chromosome 13 provided in Table 10 was calculated using sequence tag density for chromosome 4 as the normalizing chromosome, while the dose given on Table 11 was determined using the average of the sequence tag densities ratios for the group of chromosomes 3, 4, 5, and 6 as the normalizing chromosome. A diagnosis for T13 was given based on the chromosome dose in the test sample being greater than the set threshold. One T13 sample was identified using chromosome doses, and were confirmed to be T13 by karyotyping.

[00190] The data show that the combination of chromosomes 3, 4, 5, and 6 provide a variability (1.06) that is similar than that of chromosome 4 (1.01), demonstrating that a group of chromosomes can be used as the normalizing chromosome to determine chromosome doses and identify aneuploidies.

TABLE 10**Chromosome Dose for a T13 aneuploidy**

Chromosome	Sequence Tag Density	Chromosome Dose for Chr 13	Threshold
Chr13	669,872	0.538140	0.536044
Chr4	1,244,791		

TABLE 11**Chromosome Dose for a T13 aneuploidy**

Chromosome	Sequence Tag Density	Chromosome Dose for Chr 13	Threshold
Chr13	669,872	0.532674	0.515706
Chr3	1,385,881		
Chr4	1,244,791		
Chr5	1,229,257		
Chr6	1,170,331		

Turner Syndrome (monosomy X)

[00191] Three samples having a chromosome dose less than that of the set threshold were identified as having less than one X chromosome. The same samples were determined to have a Y chromosome dose that was less than the set threshold, indicating that the samples did not have a Y chromosome.

[00192] The calculated doses for chromosomes X and Y in the exemplary monosomy X test sample (54430) are given in Table 12. Chromosome 4 was selected as the normalizing chromosome to calculate the dose for chromosome X; and all chromosomes *i.e.* 1-22, and Y, were used as the normalizing chromosomes. The calculated threshold for the positive diagnosis of Turner Syndrome (monosomy X) was set for the X chromosome at < -2 standard deviations from the mean, and for the absence of the Y chromosome at < - 2 standard deviations from the mean for qualified (non-monosomy X) samples.

TABLE 12
Chromosome Dose for a Turner Syndrome (monosomy X)

Chromosome	Sequence Tag Density	Chromosome Dose for Chr X	Threshold
ChrX	904,049	0.777990	0.797603
Chr4	1,162,031		
ChrY	390	0.0004462	0.002737754
Chr (1-22, X) [average]	874,108.1		

[00193] Thus, the method enables the simultaneous determination of chromosomal aneuploidies and fetal fraction by massively parallel sequencing of a maternal sample comprising a mixture of fetal and maternal cfDNA that has been enriched for a plurality of polymorphic sequences each comprising a SNP. In this example, the mixture of fetal and maternal nucleic acids was enriched by combining a portion of a sequencing library that was constructed from amplified fetal and maternal polymorphic sequences with a sequencing library that was constructed from the remaining unamplified original fetal and maternal cfDNA mixture.

Example 7

Simultaneous determination of Aneuploidy and Fetal fraction:

Enrichment of fetal and maternal nucleic acids in a purified cfDNA sample

[00194] To enrich the fetal and maternal cfDNA contained in a purified sample of cfDNA extracted from a maternal plasma sample, a portion of the purified cfDNA was used for amplifying polymorphic target nucleic acid sequences each comprising one SNP chosen from the panel of SNPs given in Table 6.

[00195] Cell-free plasma was obtained from a maternal blood sample, and cfDNA was purified from the plasma sample as described in Example 1. The final concentration was determined to be 92.8pg/μl.

[00196] cfDNA contained in 5μl of purified cfDNA was amplified in a reaction volume of 50μl containing 7.5μl of a 1uM primer mix (Table 5), 10μl of NEB 5X Mastermix and 27 μl water. Thermal cycling was performed with the Gene Amp9700 (Applied Biosystems). Using the following cycling conditions: incubating at 95°C for 1 minute, followed by 30 cycles at 95°C for 20 seconds, 68°C for 1 minute, and 68°C for 30s, which was followed by a final incubation at 68°C for 5 minutes. A final hold at 4°C was added until the samples were removed for combining with the unamplified portion of the purified cfDNA sample. The amplified product was purified using the Agencourt AMPure XP PCR purification system (Part No. A63881; Beckman Coulter Genomics, Danvers, MA), and the concentration quantified using the Nanodrop 2000 (Thermo Scientific, Wilmington, DE). The purified amplification product was diluted 1:10 in water and 0.9 μl (371pg) added to 40μl of purified cfDNA sample to

obtain a 10% spike. The enriched fetal and maternal cfDNA present in the purified cfDNA sample was used for preparing a sequencing library, and was sequenced as described in Example 2.

[00197] Table 13 provides the tag counts obtained for each of chromosomes 21, 18, 13, X and Y *i.e.* sequence tag density, and the tag counts obtained for the informative polymorphic sequences contained in the SNP reference genome *i.e.* SNP tag density. The data show that sequencing information can be obtained from sequencing a single library constructed from a purified maternal cfDNA sample that has been enriched for sequences comprising SNPs to simultaneously determine the presence or absence of aneuploidy and the fetal fraction. In the example given, the data show that the fraction of fetal DNA in plasma sample AFR105 was quantifiable from the sequencing results of five informative SNPs and determined to be 3.84%. Sequence tag densities are provided for chromosomes 21, 13, 18, X and Y. Sample AFR105 was the only sample that was subjected to the protocol of enriching purified cfDNA for amplified polymorphic sequences. Thus, coefficients of variation and tests for differentiability were not provided. However, the example shows that the enrichment protocol provides the requisite tag counts for determining aneuploidy and fetal fraction from a single sequencing process.

TABLE 13

**Simultaneous Determination of Aneuploidy and Fetal Fraction:
Enrichment of fetal and maternal nucleic acids in a purified cfDNA sample**

Aneuploidy					
	Chromosome 21	Chromosome 18	Chromosome 13	Chromosome X	Chromosome Y
Sequence Tag Density	178763	359529	388204	572330	2219
Karyotype	Unaffected	Unaffected	Unaffected	Unaffected	Unaffected
Fetal Fraction					
SNP		SNP TAG DENSITY		FETAL FRACTION (%)	
rs10773760.1 Chr.12 length=128 allele=A		18903		2.81	
rs10773760.2 Chr.12 length=128 allele=G		532			
rs1109037.1 Chr.2 length=126 allele=A		347		5.43	
rs1109037.2 Chr.2 length=126 allele=G		6394			
rs2567608.1 Chr.20 length=110 allele=A		94503		1.74	
rs2567608.2 Chr.20 length=110 allele=G		1649			
rs7041158.1 Chr.9 length=117 allele=C		107		5.61	
rs7041158.2 Chr.9 length=117 allele=T		6			
rs8078417.1 Chr.17 length=110 allele=C		162668		3.61	
rs8078417.2 Chr.17 length=110 allele=T		5877			
Fetal Fraction (Mean±S.D.) = 3.8±1.7					

Example 8**Simultaneous determination of Aneuploidy and Fetal fraction:****Enrichment of fetal and maternal nucleic acids in a plasma sample**

[00198] To enrich the fetal and maternal cfDNA contained in an original plasma sample derived from a pregnant woman, a portion the original plasma sample was used for amplifying polymorphic target nucleic acid sequences each comprising one SNP chosen from the panel of SNPs given in Table 14, and a portion of the amplified product was combined with the remaining original plasma sample.

[00199] cfDNA contained in 15 μ l of cell-free plasma was amplified in a reaction volume of 50 μ l containing 9 μ l of a 1 μ M mixture of primers (15 plexTable 5), 1 μ l of Phusion blood DNA polymerase, 25 μ l of the 2X Phusion blood PCR buffer containing deoxynucleotide triphosphates (dNTPs: dATP, dCTP, dGTP and dTTP). Thermal cycling was performed with the Gene Amp9700 (Applied Biosystems) using the following cycling conditions: incubating at 95°C for 3 minutes, followed by 35 cycles at 95°C for 20 seconds, 55°C for 30s, and 70°C for 1 minute, which was followed by a final incubation at 68°C for 5 minutes. A final hold at 4°C was added until the samples were removed for combining with the unamplified portion of the cell-free plasma. The amplified product was diluted 1:2 with water and analyzed using the Bioanalyzer. An additional 3 μ l of amplified product was diluted with 11.85 μ l of water to obtain a final concentration of 2ng/ μ l. 2.2 μ l of the diluted amplified product was combined with the remaining plasma sample. The enriched fetal and maternal cfDNA present in the plasma sample was purified as described in Example 1, and used for preparing a sequencing library. Sequencing and analysis of the sequencing data was performed as described in Examples 2 and 3.

[00200] The results are given in Table 14. In the example given, the data show that the fraction of fetal DNA in plasma sample SAC2517 was quantifiable from the sequencing results of one informative SNP and determined to be 9.5%. In the example given, sample SAC2517 was shown by karyotyping to be unaffected for aneuploidies of chromosomes 21, 13, 18, X and Y. Sequence tag densities are provided for chromosomes 21, 13, 18, X and Y. Sample SAC2517 was the only sample that was subjected to the protocol of enriching plasma cfDNA for amplified polymorphic sequences. Thus, coefficients of variation and tests for differentiability could not determined. The example demonstrates that enriching the mixture of fetal and maternal cfDNA present in a plasma sample for nucleic acid sequences that comprise at least one informative SNP can be used to provide the requisite sequence and SNP tag counts for determining aneuploidy and fetal fraction from a single sequencing process.

TABLE 14**Simultaneous Determination of Aneuploidy and fetal fraction:****Enrichment of fetal and maternal nucleic acids in a plasma sample**

Aneuploidy					
	Chromosome 21	Chromosome 18	Chromosome 13	Chromosome X	Chromosome Y
Sequence Tag Density	183851	400582	470526	714055	2449
Karyotype	Unaffected	Unaffected	Unaffected	Unaffected	Unaffected
Fetal Fraction					
SNP	TAG COUNTS		FETAL FRACTION (%)		
rs10773760.1 Chr.12 length=128 allele=A	8536		9.49		

rs10773760.2 Chr.12 length=128 allele=G	89924	
---	-------	--

Example 9

Simultaneous Determination of Aneuploidy and Fetal Fraction in maternal samples enriched for polymorphic sequences comprising STRs

[00201] To simultaneously determine the presence or absence of an aneuploidy and the fetal fraction in a mixture of fetal and maternal cfDNA obtained from a maternal sample, the mixture is enriched for polymorphic sequences comprising STRs, sequenced and analyzed. Enrichment can be of a sequencing library as described in Example 6, of a purified cfDNA sample as described in Example 7, or of a plasma sample as described in Example 8. In each case, sequencing information is obtained from sequencing a single library, which enables for simultaneously determining the presence or absence of an aneuploidy and the fetal fraction. STRs that are amplified are chosen from the codis and non-codis STRs disclosed in Table 9, and amplification of the polymorphic STR sequences is obtained using the corresponding sets of primers provided. The STRs of Table 9 have been disclosed previously, and STRs CSF1PO, FGA, D7S820, D13S317, D16S539, D18S51, D21S11, D2S1338 (see Table 5), have been used to determine fetal fraction in plasma cfDNA samples obtained from women pregnant with either male or female fetuses (see US Provisional applications 61/296,358 and 61/360,837). Quantification of the STRs was performed using capillary electrophoresis (see Example 11). Example 11 shows that STRs can be used to determine fetal fraction.

TABLE 15
CODIS and NON-CODIS miniSTRs

STR Locus (Marker Name)	Chromosome Location	Size Range (bp)	GenBank Accession	Primer Sequences (Forward/Reverse)
Codis miniSTR loci*				
CSF1PO	5q33.1	89-129	X14720	ACAGTAACTGCCTTCATAGATAG (CSF1PO_F; SEQ ID NO:113) GTGTCAGACCCTGTTCTAAGTA (CSF1PO_R; SEQ ID NO:114)
FGA	4q31.3	125-281	M64982	AAATAAAATTAGGCATATTTACAAGC (FGA_F; SEQ ID NO:115) GCTGAGTGATTTGTCTGTAATTG(FGA_R; SEQ ID NO:116)
TH01	11p15.5	51-98	D00269	CCTGTTCCCTTATTTCCC(TH01_F; SEQ ID NO:117) GGGAACACAGACTCCATGGTG(TH01_R; SEQ ID NO:118)
TPOX	2p25.3	65-101	M68651	CTTAGGGAACCCTCACTGAATG(TPOX_F ; SEQ ID NO:119) GTCCTTGTCAGCGTTTATTTGC(TPOX_R; SEQ ID NO:120)

vWA	12p13.31	88-148	M25858	AATAATCAGTATGTGACTTGGATTGA(vWA_F; SEQ ID NO:121) ATAGGATGGATGGATAGATGGA(vWA_R; SEQ ID NO:122)
D3S1358	3p21.31	72-120	NT_005997	CAGAGCAAGACCCTGTCTCAT(D3S1358_F; SEQ ID NO:123) TCAACAGAGGCTTGCATGTAT(D3S1358_R; SEQ ID NO:124)
D5S818	5q23.2	81-117	AC008512	GGGTGATTTTCCTCTTTGGT(D5S818_F; SEQ ID NO:125) AACATTTGTATCTTTATCTGTATCCTTATTTAT(D5S818_R; SEQ ID NO:126)
D7S820	7q21.11	136-176	AC004848	GAACACTTGTCATAGTTTAGAACGAAC(D7S820_F; SEQ ID NO:127) TCATTGACAGAATTGCACCA(D7S820_R; SEQ ID NO:128)
D8S1179	8q24.13	86-134	AF216671	TTTGTATTTTCATGTGTACATTCGTATC(D7S820_F; SEQ ID NO:129) ACCTATCCTGTAGATTATTTTCACTGTG(D7S820_R; SEQ ID NO:130)
D13S317	13q31.1	88-132	AL353628	TCTGACCCATCTAACGCCTA(D13S317_F; SEQ ID NO:131) CAGACAGAAAGATAGATAGATGATTGA(D13S317_R; SEQ ID NO:132)
D16S539	16q24.1	81-121	AC024591	ATACAGACAGACAGACAGGTG(D16S539_F; SEQ ID NO:133) GCATGTATCTATCATCCATCTCT(D16S539_R; SEQ ID NO:134)
D18S51	18q21.33	113-193	AP001534	TGAGTGACAAATTGAGACCTT(D18S51_F; SEQ ID NO:135) GTCTTACAATAACAGTTGCTACTATT(D18S51_R; SEQ ID NO:136)
D21S11	21q21.1	153-221	AP000433	ATTCCCCAAGTGAATTGC(D21S11_F; SEQ ID NO:137) GGTAGATAGACTGGATAGATAGACGA(D21S11_R; SEQ ID NO:138)
D2S1338	2q35	90-142	AC01036	TGGAAACAGAAATGGCTTGG(D2S1338_F; SEQ ID NO:139) GATTGCAGGAGGGAAGGAAG(D2S1338_R; SEQ ID NO:140)

Penta D	21q22.3	94-167	AP001752	GAGCAAGACACCATCTCAAGAA(Penta D_F; SEQ ID NO:141) GAAATTTTACATTTATGTTTATGATTCTC T(Penta D_R; SEQ ID NO:142)
Penta E	15q26.2	80-175	AC027004	GGCGACTGAGCAAGACTC(Penta E_F; SEQ ID NO:143) GGTTATTAATTGAGAAAACCTCCTTACA(Penta E_R; SEQ ID NO:144)
Non-Codis miniSTR loci*				
D22S1045	22q12.3	82 – 115	AL022314 (17)	ATTTTCCCCGATGATAGTAGTCT(D22S1045_F; SEQ ID NO:145) GCGAATGTATGATTGGCAATATTTT(D22S1045_R; SEQ ID NO:146)
D20S1082	20q13.2	73 – 101	AL158015	ACATGTATCCCAGAACTTAAAGTAAAC(D20S1082_F; SEQ ID NO:147) GCAGAAGGGAAAATTGAAGCTG(D20S1082_R; SEQ ID NO:148)
D20S482	20p13	85 – 126	AL121781 (14)	CAGAGACACCGAACCAATAAGA(D20S482_F; SEQ ID NO:149) GCCACATGAATCAATTCCTATAATAAA(D20S482_R; SEQ ID NO:150)
D18S853	18p11.31	82 – 104	AP005130 (11)	GCACATGTACCCTAAAACCTTAAAAT(D18S853_F; SEQ ID NO:151) GTCAACCAAACTCAACAAGTAGTAA(D18S853_R; SEQ ID NO:152)
D17S1301	17q25.1	114 – 139	AC016888 (12)	AAGATGAAATTGCCATGTAAAAATA(D17S1301_F; SEQ ID NO:153) GTGTGTATAACAAAATTCCTATGATGG(D17S1301_R; SEQ ID NO:154)
D17S974	17p13.1	114 – 139	AC034303 (10)	GCACCCAAAACCTGAATGTCATA(D17S974_F; SEQ ID NO:155) GGTGAGAGTGAGACCCTGTC(D17S974_R; SEQ ID NO:156)
D14S1434	14q32.13	70 – 98	AL121612 (13)	TGTAATAACTCTACGACTGTCTGTCTG(D14S1434_F; SEQ ID NO:157) GAATAGGAGGTGGATGGATGG(D14S1434_R; SEQ ID NO:158)
D12ATA63	12q23.3	76 – 106	AC009771 (13)	GAGCGAGACCCTGTCTCAAG(D12ATA63_F; SEQ ID NO:159) GGAAAAGACATAGGATAGCAATTT(D12

				ATA63_R; SEQ ID NO:160)
D11S4463	11q25	88 – 116	AP002806 (14)	TCTGGATTGATCTGTCTGTCC(D11S4463_F; SEQ ID NO:161) GAATTAAATACCATCTGAGCACTGAA(D11S4463_R; SEQ ID NO:162)
D10S1435	10p15.3	82 – 139	AL354747 (11)	TGTTATAATGCATTGAGTTTTATTCTG(D10S1435_F; SEQ ID NO:163) GCCTGTCTCAAAAATAAAGAGATAGACA(D10S1435_R; SEQ ID NO:164)
D10S1248	10q26.3	79 – 123	AL391869 (13)	TTAATGAATTGAACAAATGAGTGAG(D10S1248_F; SEQ ID NO:165) GCAACTCTGGTTGTATTGTCTTCAT(D10S1248_R; SEQ ID NO:166)
D9S2157	9q34.2	71 – 107	AL162417 (10)	CAAAGCGAGACTCTGTCTCAA(D9S2157_F; SEQ ID NO:167) GAAAATGCTATCCTCTTTGGTATAAAT(D9S2157_R; SEQ ID NO:168)
D9S1122	9q21.2	93 – 125	AL161789 (12)	GGGTATTTCAAGATAACTGTAGATAGG(D9S1122_F; SEQ ID NO:168) GCTTCTGAAAGCTTCTAGTTTACC(D9S1122_R; SEQ ID NO:170)
D8S1115	8p11.21	63 – 96	AC090739 (9)	TCCACATCCTCACCAACAC(D8S1115_F; SEQ ID NO:171) GCCTAGGAAGGCTACTGTCAA(D8S1115_R; SEQ ID NO:172)
D6S1017	6p21.1	81 – 110	AL035588 (10)	CCACCCGTCCATTTAGGC(D6S1017_F; SEQ ID NO:173) GTGAAAAAGTAGATATAATGGTTGGTG(D6S1017_R; SEQ ID NO:174)
D6S474	6q21	107 – 136	AL357514 (17)	GGTTTTCCAAGAGATAGACCAATTA(D6S474_F; SEQ ID NO:175) GTCCTCTCATAAATCCCTACTCATATC(D6S474_R; SEQ ID NO:176)
D5S2500	5q11.2	85 – 126	AC008791 (17)	CTGTTGGTACATAATAGGTAGGTAGGT(D5S2500_F; SEQ ID NO:177) GTCGTGGGCCCCATAAATC(D5S2500_R; SEQ ID NO:178)
D4S2408	4p15.1	85 – 109	AC110763 (9)	AAGGTACATAACAGTTCAATAGAAAGC(D4S2408_F; SEQ ID NO:179) GTGAAATGACTGAAAAATAGTAACCA(D

				4S2408_R; SEQ ID NO:180)
D4S2364	4q22.3	67 – 83	AC022317 (9)	CTAGGAGATCATGTGGGTATGATT(D4S2364U_F; SEQ ID NO:181) GCAGTGAATAAATGAACGAATGGA(D4S2364_R; SEQ ID NO:182)
D3S4529	3p12.1	111 – 139	AC117452 (13)	CCCAAATTACTTGAGCCAAT(D3S452_F; SEQ ID NO:183) GAGACAAAATGAAGAAACAGACAG(D3S452_R; SEQ ID NO:184)
D3S3053	3q26.31	84 – 108	AC069259 (9)	TCTTTGCTCTCATGAATAGATCAGT(D3S3053_F; SEQ ID NO:185) GTTTGTGATAATGAACCCACTCAG(D3S3053_R; SEQ ID NO:186)
D2S1776	2q24.3	127 – 161	AC009475 (11)	TGAACACAGATGTTAAGTGTGTATATG(D2S1776_F; SEQ ID NO:187) GTCTGAGGTGGACAGTTATGAAA(D2S1776_R; SEQ ID NO:188)
D2S441	2p14	78 – 110	AC079112 (12)	CTGTGGCTCATCTATGAAAACCTT(D2S441_F; SEQ ID NO:189) GAAGTGGCTGTGGTGTTATGAT(D2S441_R; SEQ ID NO:190)
D1S1677	1q23.3	81 – 117	AL513307 (15)	TTCTGTTGGTATAGAGCAGTGTTT(D1S1677_F; SEQ ID NO:191) GTGACAGGAAGGACGGAATG(D1S1677_R; SEQ ID NO:192)
D1S1627	1p21.1	81 – 100	AC093119 (13)	CATGAGGTTTGCAAATACTATCTTAAC(D1S1627_F; SEQ ID NO:193) GTTTAAATTTCTCCAAATCTCCA(D1S1627_R; SEQ ID NO:194)
D1GATA113	1p36.23	81 – 105	Z97987 (11)	TCTTAGCCTAGATAGATACTTGCTTCC(D1GATA113_F; SEQ ID NO:195) GTCAACCTTTGAGGCTATAGGAA(D1GATA113_R; SEQ ID NO:196)

*(Butler *et al.*, J Forensic Sci 5:1054-1064; Hill *et al.*, Poster #44- 17th International Symposium on Human Identification – 2006)

[00202] Sequencing of the library enriched for polymorphic STR sequences is performed using a NGS technology *e.g.* sequencing by synthesis. Sequence reads of lengths of at least 100 bp are aligned to a reference genome *e.g.* the human reference genome NCBI36/hg18 sequence, and to an STR genome, and the number of sequence tags and STR tags obtained is used to determine the presence or absence of aneuploidy and the fetal fraction, respectively. The STR reference genome includes the sequences of amplicons amplified from the given primers.

Example 10**Simultaneous Determination of Aneuploidy and Fetal Fraction in maternal samples enriched for polymorphic sequences comprising tandem SNPs**

[00203] To determine simultaneously aneuploidy and fetal fraction in maternal samples comprising fetal and maternal nucleic acids, plasma samples, purified cfDNA samples, and sequencing library samples are enriched for polymorphic target nucleic acid sequences each comprising a pair of tandem SNPs selected from rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672. The primers used for amplifying the target sequences comprising the tandem SNPs are designed to encompass both SNP sites. For example, the forward primer is designed to encompass the first SNP, and the reverse primer is designed to encompass the second of the tandem SNP pair *i.e.* each of the SNP sites in the tandem pair is encompassed within the 36 bp generated by the sequencing method. Paired-end sequencing can be used to identify all sequences encompassing the tandem SNP sites. Exemplary sets of primers that are used to amplify the tandem SNPs disclosed herein are set rs7277033-rs2110153_F: TCCTGGAAACAAAAGTATT (SEQ ID NO:197) and rs7277033-rs2110153_R: AACCTTACAACAAAGCTAGAA (SEQ ID NO:198), set rs2822654-rs1882882_F: ACTAAGCCTTGGGGATCCAG (SEQ ID NO:199) and rs2822654-rs1882882_R: TGCTGTGGAAATACTAAAAGG (SEQ ID NO:200), set rs368657-rs376635_F:CTCCAGAGGTAATCCTGTGA (SEQ ID NO:201) and rs368657-rs376635_R:TGGTGTGAGATGGTATCTAGG (SEQ ID NO:202), rs2822731-rs2822732_F:GTATAATCCATGAATCTTGTTT (SEQ ID NO:203) and rs2822731-rs2822732_R:TTCAAATTGTATATAAGAGAGT (SEQ ID NO:204), rs1475881-rs7275487_F:GCAGGAAAGTTATTTTAAAT (SEQ ID NO:205) and rs1475881-rs7275487_R:TGCTTGAGAAAGCTAACACTT (SEQ ID NO:206), rs1735976-rs2827016_F:CAGTGTTTGGAAATTGTCTG (SEQ ID NO:207) and rs1735976-rs2827016_R:GGCACTGGGAGATTATTGTA (SEQ ID NO:208), rs447349-rs2824097_F:TCCTGTTGTTAAGTACACAT (SEQ ID NO:209) and rs447349-rs2824097_R:GGGCCGTAATTACTTTTG (SEQ ID NO:210), rs418989-rs13047336_F:ACTCAGTAGGCACTTTGTGTC (SEQ ID NO:211) and rs418989-rs13047336_R:TCTTCCACCACACCAATC (SEQ ID NO:212), rs987980-rs987981_F:TGGCTTTTCAAAGGTAAAA (SEQ ID NO:213) and rs987980-rs987981_R:GCAACGTTAACATCTGAATTT (SEQ ID NO:214), rs4143392-rs4143391_F: rs4143392-rs4143391 (SEQ ID NO:215) and rs4143392-rs4143391_R:ATTTTATATGTCATGATCTAAG (SEQ ID NO:216), rs1691324-

rs13050434_F: AGAGATTACAGGTGTGAGC (SEQ ID NO:217) and rs1691324-rs13050434_R: ATGATCCTCAACTGCCTCT (SEQ ID NO:218), rs11909758-rs9980111_F: TGAAACTCAAAAGAGAAAAG (SEQ ID NO:219) and rs11909758-rs9980111_R: ACAGATTTCTACTTAAAATT (SEQ ID NO:220), rs2826842-rs232414_F: TGAAACTCAAAAGAGAAAAG (SEQ ID NO:221) and rs2826842-rs232414_R: ACAGATTTCTACTTAAAATT (SEQ ID NO:222), rs2826842-rs232414_F: GCAAAGGGGTACTCTATGTA (SEQ ID NO:223) and rs2826842-rs232414_R: TATCGGGTCATCTTGTTAAA (SEQ ID NO:224), rs1980969-rs1980970_F: TCTAACAAAGCTCTGTCCAAAA (SEQ ID NO:225) and rs1980969-rs1980970_R: CCACACTGAATAACTGGAACA (SEQ ID NO:226), rs9978999-rs9979175_F: GCAAGCAAGCTCTCTACCTTC (SEQ ID NO:227) and rs9978999-rs9979175_R: TGTTCCTCCAAAATTCACATGC (SEQ ID NO:228), rs1034346-rs12481852_F: ATTTCACTATTCCTTCATTTT (SEQ ID NO:229) and rs1034346-rs12481852_R: TAATTGTTGCACACTAAATTAC (SEQ ID NO:230), rs4817013-rs7277036_F: AAAAAGCCACAGAAATCAGTC (SEQ ID NO:231) and rs4817013-rs7277036_R: TTCTTATATCTCACTGGGCATT (SEQ ID NO:232), rs9981121-rs2829696_F: GGATGGTAGAAGAGAAGAAAGG (SEQ ID NO:233) and rs9981121-rs2829696_R: GGATGGTAGAAGAGAAGAAAGG (SEQ ID NO:234), rs455921-rs2898102_F: TGCAAAGATGCAGAACCAAC (SEQ ID NO:235) and rs455921-rs2898102_R: TTTTGTTCCCTTGTCCTGGCTGA (SEQ ID NO:236), rs2898102-rs458848_F: TGCAAAGATGCAGAACCAAC (SEQ ID NO:237) and rs2898102-rs458848_R: GCCTCCAGCTCTATCCAAGTT (SEQ ID NO:238), rs961301-rs2830208_F: CCTTAATATCTTCCCATGTCCA (SEQ ID NO:239) and rs961301-rs2830208_R: ATTGTTAGTGCCTCTTCTGCTT (SEQ ID NO:240), rs2174536-rs458076_F: GAGAAGTGAGGTCAGCAGCT (SEQ ID NO:241) and rs2174536-rs458076_R: TTTCTAAATTTCCATTGAACAG (SEQ ID NO:242), rs11088023-rs11088024_F: GAAATTGGCAATCTGATTCT (SEQ ID NO:243) and rs11088023-rs11088024_R: CAACTTGTCCTTTATTGATGT (SEQ ID NO:244), rs1011734-rs1011733_F: CTATGTTGATAAAACATTGAAA (SEQ ID NO:245) and rs1011734-rs1011733_R: GCCTGTCTGGAATATAGTTT (SEQ ID NO:246), rs2831244-rs9789838_F: CAGGGCATATAATCTAAGCTGT (SEQ ID NO:247) and rs2831244-rs9789838_R: CAATGACTCTGAGTTGAGCAC (SEQ ID NO:248), rs8132769-rs2831440_F: ACTCTCTCCCTCCCCTCT (SEQ ID NO:249) and rs8132769-rs2831440_R: TATGGCCCCAAACTATTCT (SEQ ID NO:250), rs8134080-rs2831524_F: ACAAGTACTGGGCAGATTGA (SEQ ID NO:251) and rs8134080-rs2831524_R: GCCAGGTTTAGCTTTCAAGT (SEQ ID NO:252), rs4817219-rs4817220_F: TTTTATATCAGGAGAAACACTG (SEQ ID NO:253) and rs4817219-rs4817220_R: CCAGAATTTTGGAGGTTTAAT (SEQ ID NO:254), rs2250911-rs2250997_F: TGTCATTCTCCTTTATCTCCA (SEQ ID NO:255) and rs2250911-rs2250997_R: TTCTTTTGCCTCTCCCAAAG (SEQ ID NO:256), rs2831899-rs2831900_F: ACCCTGGCACAGTGTTGACT (SEQ ID NO:257) and rs2831899-rs2831900_R: TGGGCCTGAGTTGAGAAGAT (SEQ ID NO:258), rs2831902-rs2831903_F: AATTTGTAAGTATGTGCAACG (SEQ ID NO:259) and rs2831902-rs2831903_R: TTTTCCCATTTCCAACCTCT (SEQ ID NO:260), rs11088086-rs2251447_F: AAAAGATGAGACAGGCAGGT (SEQ ID NO:261) and rs11088086-rs2251447_R: ACCCCTGTGAATCTCAAAAT (SEQ ID NO:262), rs2832040-rs11088088_F: GCACTTGCTTCTATTGTTTGT (SEQ ID NO:263) and rs2832040-rs11088088_R:

CCCTTCCTCTCTTCCATTCT (SEQ ID NO:264), rs2832141-rs2246777_F: AGCACTGCAGGTA (SEQ ID NO:265) and rs2832141-rs2246777_R: ACAGATACCAAAGAACTGCAA (SEQ ID NO:266), rs2832959 – rs9980934_F: TGGACACCTTTCAACTTAGA (SEQ ID NO:267) and rs2832959 – rs9980934_R: GAACAGTAATGTTGAACTTTTT (SEQ ID NO:268), rs2833734-rs2833735_F: TCTTGCAAAAAGCTTAGCACA (SEQ ID NO:269) and rs2833734-rs2833735_R: AAAAAGATCTCAAAGGGTCCA (SEQ ID NO:270), rs933121-rs933122_F: GCTTTTGCTGAACATCAAGT (SEQ ID NO:271) and rs933121-rs933122_R: CCTTCCAGCAGCATAGTCT (SEQ ID NO:272), rs2834140-rs12626953_F: AAATCCAGGATGTGCAGT (SEQ ID NO:273) and rs2834140-rs12626953_R: ATGATGAGGTCAGTGGTGT (SEQ ID NO:274), rs2834485-rs3453_F: CATCACAGATCATAGTAAATGG (SEQ ID NO:275) and rs2834485-rs3453_R: AATTATTATTTTGCAGGCAAT (SEQ ID NO:276), rs9974986-rs2834703_F: CATGAGGCAAACACCTTTCC (SEQ ID NO:277) and rs9974986-rs2834703_R: GCTGGACTCAGGATAAAGAACA (SEQ ID NO:278), rs2776266-rs2835001_F: TGGAAGCCTGAGCTGACTAA (SEQ ID NO:279) and rs2776266-rs2835001_R: CCTTCTTTTCCCCCAGAATC (SEQ ID NO:280), rs1984014-rs1984015_F: TAGGAGAACAGAAGATCAGAG (SEQ ID NO:281) and rs1984014-rs1984015_R: AAAGACTATTGCTAAATGCTTG (SEQ ID NO:282), rs7281674-rs2835316_F: TAAGCGTAGGGCTGTGTGTG (SEQ ID NO:283) and rs7281674-rs2835316_R: GGACGGATAGACTCCAGAAGG (SEQ ID NO:284), rs13047304-rs13047322_F: GAATGACCTTGGCACTTTTATCA (SEQ ID NO:285) and rs13047304-rs13047322_R: AAGGATAGAGATATACAGATGAATGGA (SEQ ID NO:286), rs2835735-rs2835736_F: CATGCACCGCGCAAATAC (SEQ ID NO:287) and rs2835735-rs2835736_R: ATGCCTCACCCACAAACAC (SEQ ID NO:288), rs13047608-rs2835826_F: TCCAAGCCCTTCTCACTCAC (SEQ ID NO:289) and rs13047608-rs2835826_R: CTGGGACGGTGACATTTTCT (SEQ ID NO:290), rs2836550-rs2212596_F: CCCAGGAAGAGTGGAAGATT (SEQ ID NO:291) and rs2836550-rs2212596_R: TTAGCTTGCATGTACCTGTGT (SEQ ID NO:292), rs2836660-rs2836661_F: AGCTAGATGGGGTGAATTTT (SEQ ID NO:293) and _R: TGGGCTGAGGGGAGATTC (SEQ ID NO:294), rs465612-rs8131220_F: ATCAAGCTAATTAATGTTATCT (SEQ ID NO:295) and rs465612-rs8131220_R: AATGAATAAGGTCCTCAGAG (SEQ ID NO:296), rs9980072-rs8130031_F: TTTAATCTGATCATTGCCCTA (SEQ ID NO:297) and rs9980072-rs8130031_R: AGCTGTGGGTGACCTTGA (SEQ ID NO:298), rs418359-rs2836926_F: TGTCCCACCATTTGTGTATTA (SEQ ID NO:299) and rs418359-rs2836926_R: TCAGACTTGAAGTCCAGGAT (SEQ ID NO:300), rs7278447-rs7278858_F: GCTTCAGGGGTGTTAGTTTT (SEQ ID NO:301) and rs7278447-rs7278858_R: CTTTGTGAAAAGTCGTCCAG (SEQ ID NO:302), rs385787-rs367001_F: CCATCATGGAAAGCATGG (SEQ ID NO:303) and rs385787-rs367001_R: TCATCTCCATGACTGCACTA (SEQ ID NO:304), rs367001-rs386095_F: GAGATGACGGAGTAGCTCAT (SEQ ID NO:305) and rs367001-rs386095_R: CCCAGCTGCACTGTCTAC (SEQ ID NO:306), rs2837296-rs2837297_F: TCTTGTTC CAATCACAGGAC (SEQ ID NO:307) and rs2837296-rs2837297_R: ATGCTGTTAGCTGAAGCTCT (SEQ ID NO:308), and rs2837381-rs4816672_F: TGAAAGCTCCTAAAGCAGAG (SEQ ID NO:309) and rs2837381-rs4816672_R: TTGAAGAGATGTGCTATCAT (SEQ ID NO:310). Polynucleotide sequences *e.g.* GC clamp sequences, can be included to ensure specific hybridization of AT-rich primers (Ghanta *et al.*, PLOS ONE 5(10): doi10.1371/journal.pone.0013184 [2010], available on the world wide web at plosone.org). An example of a GC

clamp sequence that can be included either 5' of the forward primer or 3' of the reverse primer is

GCCGCCTGCAGCCCGCGCCCCCGTGCCCCCGCCCCGCCGCCGGCCCGGGCGCC (SEQ ID NO:311).

Sample preparation and enrichment of cfDNA sequencing library, a purified cfDNA sample, and a plasma sample is performed according to the method described in Examples 6, 7, and 8, respectively. All sequencing libraries are prepared as described in Example 2a., and sequencing is performed as described in Example 2b and including paired-end sequencing. Analysis of the sequencing data for the determination of fetal aneuploidy is performed as described in Examples 3 and 4. Concomitant to the analysis for determining aneuploidy, the sequencing data is analyzed to determine the fetal fraction as follows. Following the transfer of the image and base call files to the Unix server running the Illumina "Genome Analyzer Pipeline" software version 1.51 as described in Example 3a., the 36bp reads are aligned to a 'tandem SNP genome' using the BOWTIE program. The tandem SNP genome is identified as the grouping of the DNA sequences that encompass the alleles of the 58 tandem SNP pairs disclosed above. Only reads that mapped uniquely to the tandem SNP genome are used for the analysis of fetal fraction. Reads that match perfectly to the tandem SNP genome are counted as tags and filtered. Of the remaining reads, only reads having one or two mismatches are counted as tags and included in the analysis. Tags mapped to each of the tandem SNP alleles are counted, and the fetal fraction is determined essentially as described in Example 6 above but accounting for tags mapped to the two tandem SNP alleles x and y present on each of the amplified polymorphic target nucleic acid sequences that are amplified to enrich the samples *i.e.*

$$\% \text{ fetal fraction allele}_{x+y} = ((\sum \text{Fetal sequence tags for allele}_{x+y}) / (\sum \text{Maternal sequence tags for allele}_{x+y})) \times 100$$

Only informative tandem SNPs are used to determine the fetal fraction.

[00204] Optionally, the fraction of fetal nucleic acids in the mixture of fetal and maternal nucleic acids is calculated for each of the informative allele (allele_{x+y}) as follows:

$$\% \text{ fetal fraction allele}_{x+y} = ((2 \times \sum \text{Fetal sequence tags for allele}_{x+y}) / (\sum \text{Maternal sequence tags for allele}_{x+y})) \times 100,$$

to compensate for the presence of 2 sets of tandem fetal alleles, one being masked by the maternal background.

[00205] The percent fetal fraction is calculated for at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40 or more informative sets of tandem alleles. In one embodiment, the fetal fraction is the average fetal fraction determined for at least 3 informative sets of tandem alleles.

Example 11

Determination of Fetal Fraction by Capillary Electrophoresis of Polymorphic Sequences Comprising STRs

[00206] To determine fetal fraction in maternal samples comprising fetal and maternal cfDNA, peripheral blood samples

[00207] were collected from volunteer pregnant women carrying either male or female fetuses. Peripheral blood samples were obtained and processed to provide purified cfDNA as described in Example 1.

[00208] Ten microliters of cfDNA samples were analyzed using the AmpF1STR® MiniFiler™ PCR amplification kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Briefly, cfDNA contained in 10 µl was amplified in a reaction volume of 25 µl containing 5µL fluorescently labeled primers (AmpF1STR® MiniFiler™ Primer Set), and the AmpF1STR® MiniFiler™ Master Mix, which includes AmpliTaq Gold® DNA

polymerase and associated buffer, salt (1.5 mM MgCl₂), and 200 μM deoxynucleotide triphosphates (dNTPs: dATP, dCTP, dGTP and dTTP). The fluorescently labeled primers are forward primers that are labeled with 6FAM™, VIC™, NED™, and PET™ dyes. Thermal cycling was performed with the Gene Amp9700 (Applied Biosystems) using the following cycling conditions: incubating at 95°C for 10 minutes, followed by 30 cycles at 94°C for 20 seconds, 59°C for 2 minute, and 72°C for 1 minute, which was followed by a final incubation at 60°C for 45 minutes. A final hold at 4°C was added until the samples were removed for analysis. The amplified product was prepared by diluting 1ul of amplified product in 8.7ul Hi-Di™ formamide (Applied Biosystems) and 0.3 μl GeneScan™-500 LIZ_ internal size standard (Applied Biosystems), and analyzed with an ABI PRISM3130xl Genetic Analyzer (Applied Biosystems) using Data Collection HID_G5_POP4 (Applied Biosystems), and a 36-cm capillary array. All genotyping was performed with GeneMapper_ID v3.2 software (Applied Biosystems) using manufacturer provided allelic ladders and bins and panels.

[00209] All genotyping measurement were performed on the Applied Biosystems 3130xl Genetic Analyzer, using a ±0.5-nt “window” around the size obtained for each allele to allow for detection and correct assignment of alleles. Any sample allele whose size was outside the ±0.5-nt window was determined to be OL *i.e.* “Off Ladder”. OL alleles are alleles of a size that is not represented in the AmpF/STR® MiniFiler™ Allelic Ladder or an allele that does not correspond to an allelic ladder, but whose size is just outside a window because of measurement error. The minimum peak height threshold of >50 RFU was set based on validation experiments performed to avoid typing when stochastic effects are likely to interfere with accurate interpretation of mixtures. The calculation of fetal fraction is based on averaging all informative markers. Informative markers are identified by the presence of peaks on the electropherogram that fall within the parameters of preset bins for the STRs that are analyzed.

[00210] Calculations of fetal fraction were performed using the average peak height for major and minor alleles at every STR locus determined from triplicate injections. The rules applied to the calculation are:

[00211] 1. off-ladder allele (OL) data for alleles are not included in the calculation; and

[00212] 2. only peak heights derived from >50 RFU (relative fluorescence units) are included in the calculation

[00213] 3. if only one bin is present the marker is deemed non-informative; and

[00214] 4. if a second bin is called but the peaks of the first and second bins are within 50-70% of their relative fluorescence units (RFU) in peak height, the minority fraction is not measured and the marker is deemed not informative.

[00215] The fraction of the minor allele for any given informative marker is calculated by dividing the peak height of the minor component by the sum of the peak height for the major component, and expressed as a percent was first calculated for each informative locus as

$$\text{fetal fraction} = (\text{peak height of minor allele} / \sum \text{peak height of major allele(s)}) \times 100,$$

[00216] The fetal fraction for a sample comprising two or more informative STRs, would be calculated as the average of the fetal fractions calculated for the two or more informative markers.

[00217] Table 16 provides the data obtained from analyzing cfDNA of a subject pregnant with a male fetus.

TABLE 16

Fetal Fraction Determined in cfDNA of a Pregnant Subject by Analysis of STRs

STR	Allele 1	Allele 2	Allele 3	Allele1 Height	Allele 2 Height	Allele 3 Height	Fetal Fraction	Fetal Fraction (Mean/STR)
AMEL	X	Y		3599	106		2.9	
AMEL	X	Y		3602	110		3.1	
AMEL	X	Y		3652	109		3.0	3.0
CSF1PO	11	12		2870	2730			
CSF1PO	11	12		2924	2762			
CSF1PO	11	12		2953	2786			
D13S317	11	12		2621	2588			
D13S317	11	12		2680	2619			
D13S317	11	12		2717	2659			
D16S539	9	11		1056	1416			
D16S539	9	11		1038	1394			
D16S539	9	11		1072	1437			
D18S51	13	15		2026	1555			
D18S51	13	15		2006	1557			
D18S51	13	15		2050	1578			
D21S11	28	31.2		2450	61		2.5	
D21S11	28	31.2		2472	62		2.5	
D21S11	28	31.2		2508	67		2.7	2.6
D2S1338	20	23		3417	3017			
D2S1338	20	23		3407	3020			
D2S1338	20	23		3493	3055			
D7S820	9	12	13	2373	178	1123	5.1	

D7S820	9	12	13	2411	181	1140	5.1	
D7S820	9	12	13	2441	182	1156	5.1	5.1
FGA	17.2	22	25	68	1140	896	3.3	
FGA	17.2	22	25	68	1144	909	3.1	
FGA	17.2	22	25	68	1151	925	3.3	3.2
Fetal Fraction = 3.5								

[00218] The results show that minSTRs can be used to discern fetal and maternal alleles in cfDNA from a maternal plasma sample. It is expected that the miniSTRs can be used in massively parallel sequencing for the simultaneous determination of aneuploidy and fetal fraction.

CLAIMS:

1. A method for simultaneously determining aneuploidy and fetal fraction in a maternal blood, plasma or serum sample comprising a mixture of fetal and maternal nucleic acid molecules, wherein the mixture of fetal and maternal nucleic acid molecules are cell-free DNA (cfDNA) molecules and said fetal fraction is the fraction of said nucleic acid molecules contributed by a fetus to said mixture of fetal and maternal nucleic acid molecules, said method comprising:

(a) enriching said mixture for a plurality of polymorphic target nucleic acids, wherein each of said target nucleic acids is known to comprise at least one polymorphic site, and wherein said enriching comprises:

(i) specifically amplifying said plurality of target nucleic acids in a portion of said mixture using primer pairs each capable of amplifying a target nucleic acid sequence comprising a polymorphic site in a multiplex PCR reaction to generate a panel of amplified polymorphic sites that contains a sufficient number of polymorphic sites such that at least two are informative polymorphic sites; and

(ii) combining at least a portion or all of the amplified product with at least a portion of the remainder of the unamplified sample from which the portion was removed to thereby form an enriched mixture;

(b) sequencing at least a portion of the enriched mixture obtained in step (a), wherein said sequencing comprises providing a plurality of sequence reads which are mapped to a genome to identify a plurality of mapped sequence tags; and

(c) using said plurality of mapped sequence tags, simultaneously determining said fetal fraction and said aneuploidy, wherein:

(i) determining said fetal fraction comprises:

(1) using the mapped sequence tags to identify at least two informative polymorphic sites in said panel of amplified polymorphic sites, wherein said informative polymorphic sites are identified by the difference in the allelic sequences and the number of sequence tags mapped to each of the possible alleles at each polymorphic site; and

(2) calculating the fetal fraction based on the total number of sequence tags that map to a first allele and the total number of sequence tags that map to a second allele at each of said informative polymorphic sites; and

(ii) determining said aneuploidy comprises quantification of the number of sequence tags aligning to a chromosome of interest, and comparing the results obtained for the chromosome of interest to a threshold value, wherein the threshold value is a number that serves as a limit of diagnosis of an aneuploidy and wherein the presence of an aneuploidy for the chromosome of interest is identified if the threshold value is exceeded by the results obtained for the chromosome of interest.

2. The method of Claim 1, wherein said enriching step (a) comprises:

(i) amplifying a plurality of polymorphic target nucleic acids to generate a panel of polymorphic sites that contains a sufficient number of polymorphic sites such that at least three are informative polymorphic sites;

(ii) amplifying a plurality of polymorphic target nucleic acids to generate a panel of polymorphic sites that contains a sufficient number of polymorphic sites such that at least four are informative polymorphic sites; or

(iii) amplifying a plurality of polymorphic target nucleic acids to generate a panel of polymorphic sites that contains a sufficient number of polymorphic sites such that at least five are informative polymorphic sites.

3. The method of Claim 1 or Claim 2, wherein determining said fetal fraction in step (c) comprises calculating the fetal fraction based on:

(i) the total number of sequence tags that map to a first allele and the total number of sequence tags that map to a second allele for each of at least three informative alleles;

(ii) the total number of sequence tags that map to a first allele and the total number of sequence tags that map to a second allele for each of at least four informative alleles; or

(iii) the total number of sequence tags that map to a first allele and the total number of sequence tags that map to a second allele for each of at least five informative alleles.

4. The method of any one of Claims 1-3, wherein said enriching step (a) comprises amplifying a plurality of polymorphic target nucleic acids in a portion of a purified mixture of fetal and maternal nucleic acids.

5. The method of any one of Claims 1-3, wherein said enriching comprises:

(i) preparing a first sequencing library of said mixture of fetal and maternal nucleic acid molecules;

(ii) preparing a second sequencing library of said amplified polymorphic target nucleic acids; and

(iii) combining at least a portion of said first sequencing library with at least a portion of said second sequencing library.

6. The method of any one of Claims 1-5, wherein said polymorphic target nucleic acids are located on the same chromosome.

7. The method of any one of Claims 1-6, wherein said polymorphic target nucleic acids are located on different chromosomes.

8. The method of any one of Claims 1-7, wherein each of said plurality of polymorphic target nucleic acids comprises at least one single nucleotide polymorphism (SNP).

9. The method of Claim 8, wherein the SNP is a single SNP.

10. The method of Claim 9, wherein said at least one SNP is a single SNP selected from each of said plurality of polymorphic target nucleic acids comprises a SNP selected from

rs560681, rs1109037, rs9866013, rs13182883, rs13218440, rs7041158, rs740598, rs10773760, rs4530059, rs7205345, rs8078417, rs576261, rs2567608, rs430046, rs9951171, rs338882, rs10776839, rs9905977, rs1277284, rs258684, rs1347696, rs508485, rs9788670, rs8137254, rs3143, rs2182957, rs3739005, and rs530022.

11. The method of Claim 8, wherein the SNP is a tandem SNP.

12. The method of Claim 11, wherein said at least one SNP is a tandem SNP selected from sets of tandem SNPs rs7277033-rs2110153; rs2822654-rs1882882; rs368657-rs376635; rs2822731-rs2822732; rs1475881-rs7275487; rs1735976-rs2827016; rs447340-rs2824097; rs418989-rs13047336; rs987980-rs987981; rs4143392-rs4143391; rs1691324-rs13050434; rs11909758-rs9980111; rs2826842-rs232414; rs1980969-rs1980970; rs9978999-rs9979175; rs1034346-rs12481852; rs7509629-rs2828358; rs4817013-rs7277036; rs9981121-rs2829696; rs455921-rs2898102; rs2898102-rs458848; rs961301-rs2830208; rs2174536-rs458076; rs11088023-rs11088024; rs1011734-rs1011733; rs2831244-rs9789838; rs8132769-rs2831440; rs8134080-rs2831524; rs4817219-rs4817220; rs2250911-rs2250997; rs2831899-rs2831900; rs2831902-rs2831903; rs11088086-rs2251447; rs2832040-rs11088088; rs2832141-rs2246777; rs2832959-rs9980934; rs2833734-rs2833735; rs933121-rs933122; rs2834140-rs12626953; rs2834485-rs3453; rs9974986-rs2834703; rs2776266-rs2835001; rs1984014-rs1984015; rs7281674-rs2835316; rs13047304-rs13047322; rs2835545-rs4816551; rs2835735-rs2835736; rs13047608-rs2835826; rs2836550-rs2212596; rs2836660-rs2836661; rs465612-rs8131220; rs9980072-rs8130031; rs418359-rs2836926; rs7278447-rs7278858; rs385787-rs367001; rs367001-rs386095; rs2837296-rs2837297; and rs2837381-rs4816672.

13. The method of any one of Claims 1-7, wherein each of said plurality of polymorphic target nucleic acids comprises at least one short tandem repeat (STR).

14. The method of Claim 13, wherein each of said plurality of polymorphic target nucleic acids is an STR selected from CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, Penta D, Penta E, D22S1045, D20S1082, D20S482, D18S853, D17S1301, D17S974, D14S1434, D12ATA63, D11S4463, D10S1435, D10S1248, D9S2157, D9S1122, D8S1115, D6S1017, D6S474, D5S2500, D4S2408, D4S2364, D3S4529, D3S3053, D2S1776, D2S441, D1S1677, D1S1627 and D1GATA113.

15. The method of Claim 13, wherein said at least one STR is less than about 300 base pairs.

16. The method of any one of Claims 1-15, wherein said sequencing:

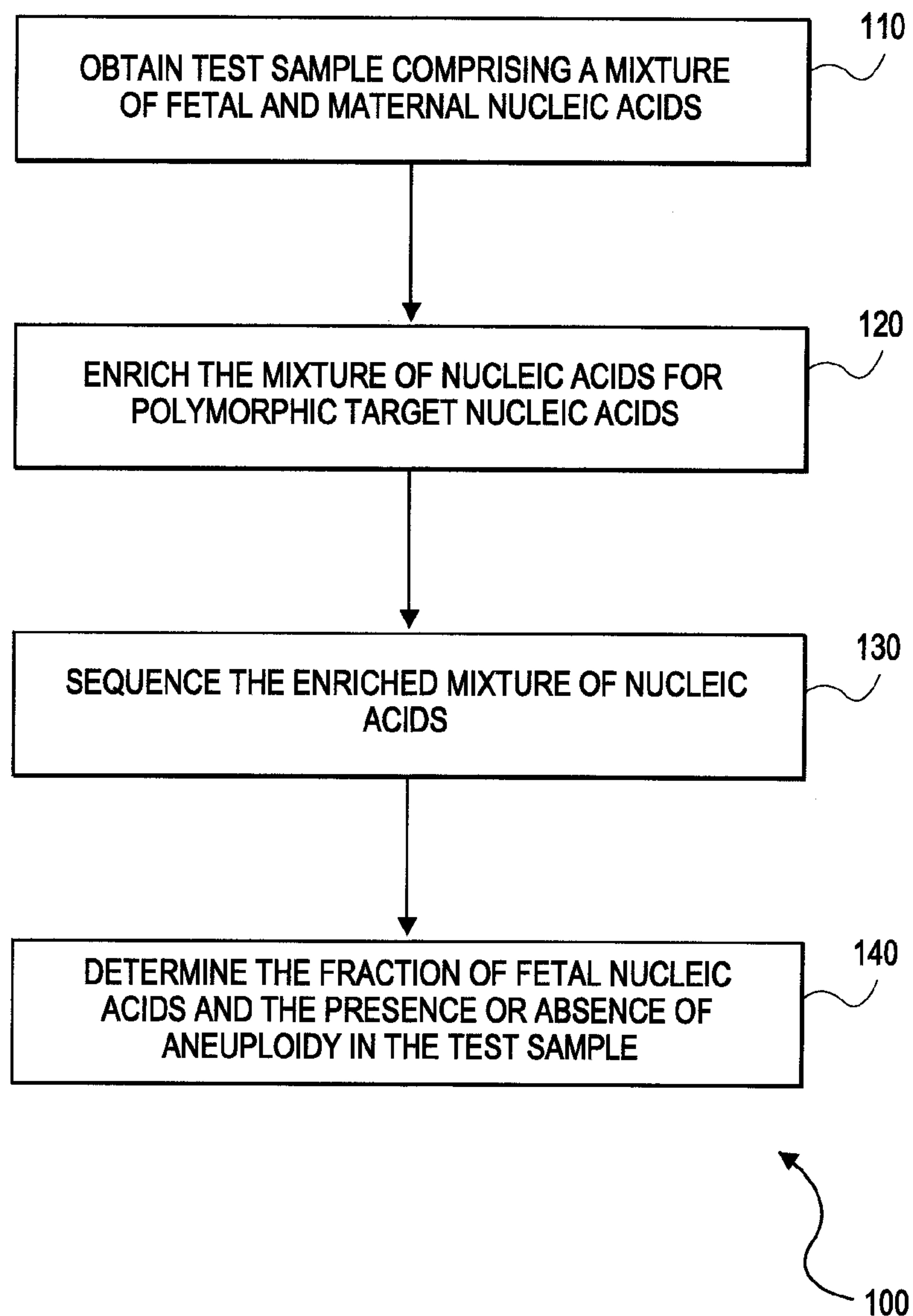
- (i) is next generation sequencing (NGS);
- (ii) is massively parallel sequencing using sequencing-by-synthesis with reversible dye terminators;
- (iii) is sequencing-by-ligation;
- (iv) comprises an amplification; or
- (v) is single molecule sequencing.

17. The method of any one of Claims 1-16, wherein said aneuploidy is a complete chromosomal aneuploidy.

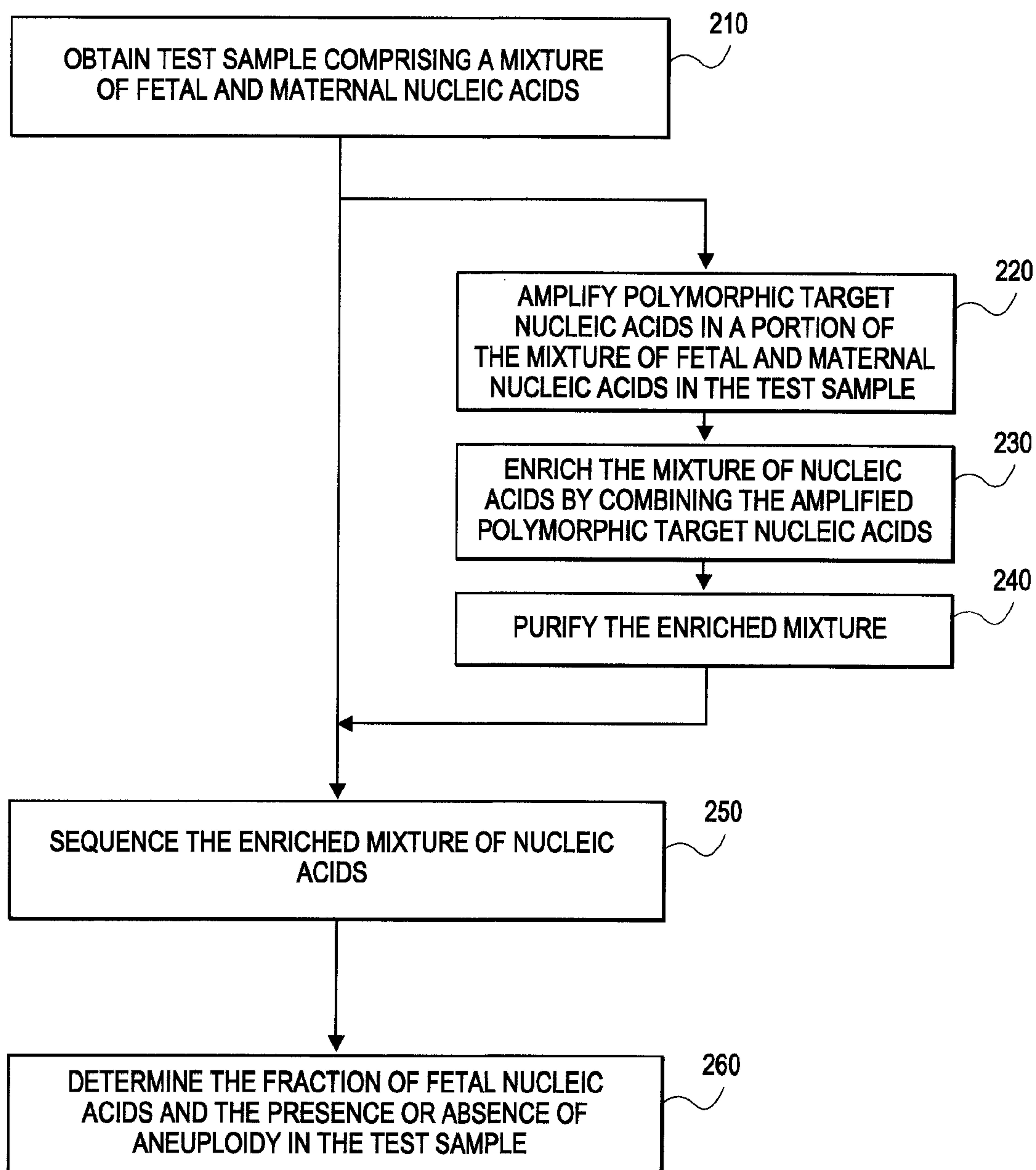
18. The method of claim 17, wherein the complete chromosomal aneuploidy is a chromosomal aneuploidy chosen from trisomy 8, trisomy 13, trisomy 15, trisomy 16, trisomy 18, trisomy 21, trisomy 22, monosomy X, XXX, XXY and XYY

19. The method of any one of Claims 1-16, wherein said aneuploidy is a partial chromosomal aneuploidy.

1/18

**FIG. 1**

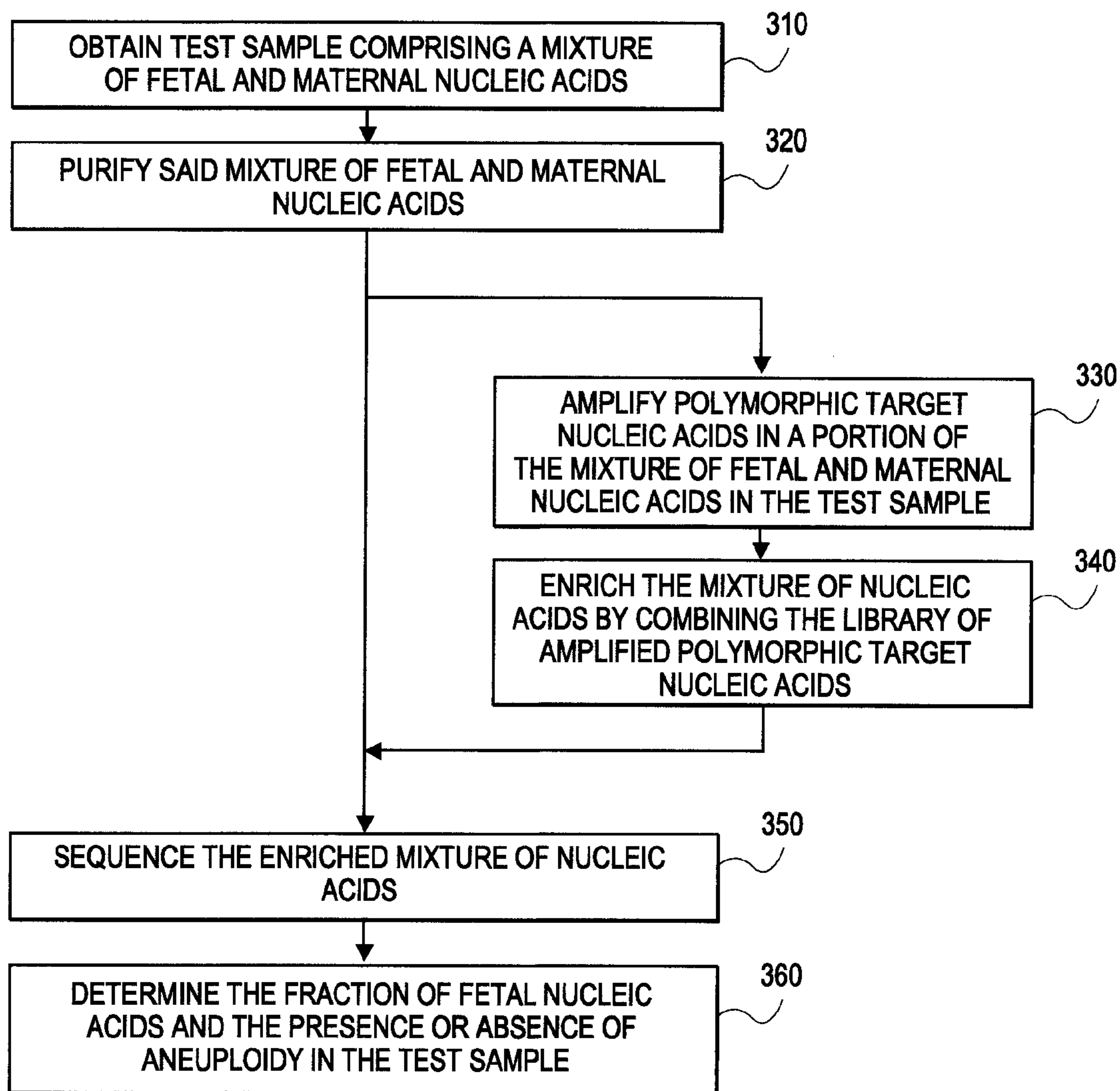
2/18



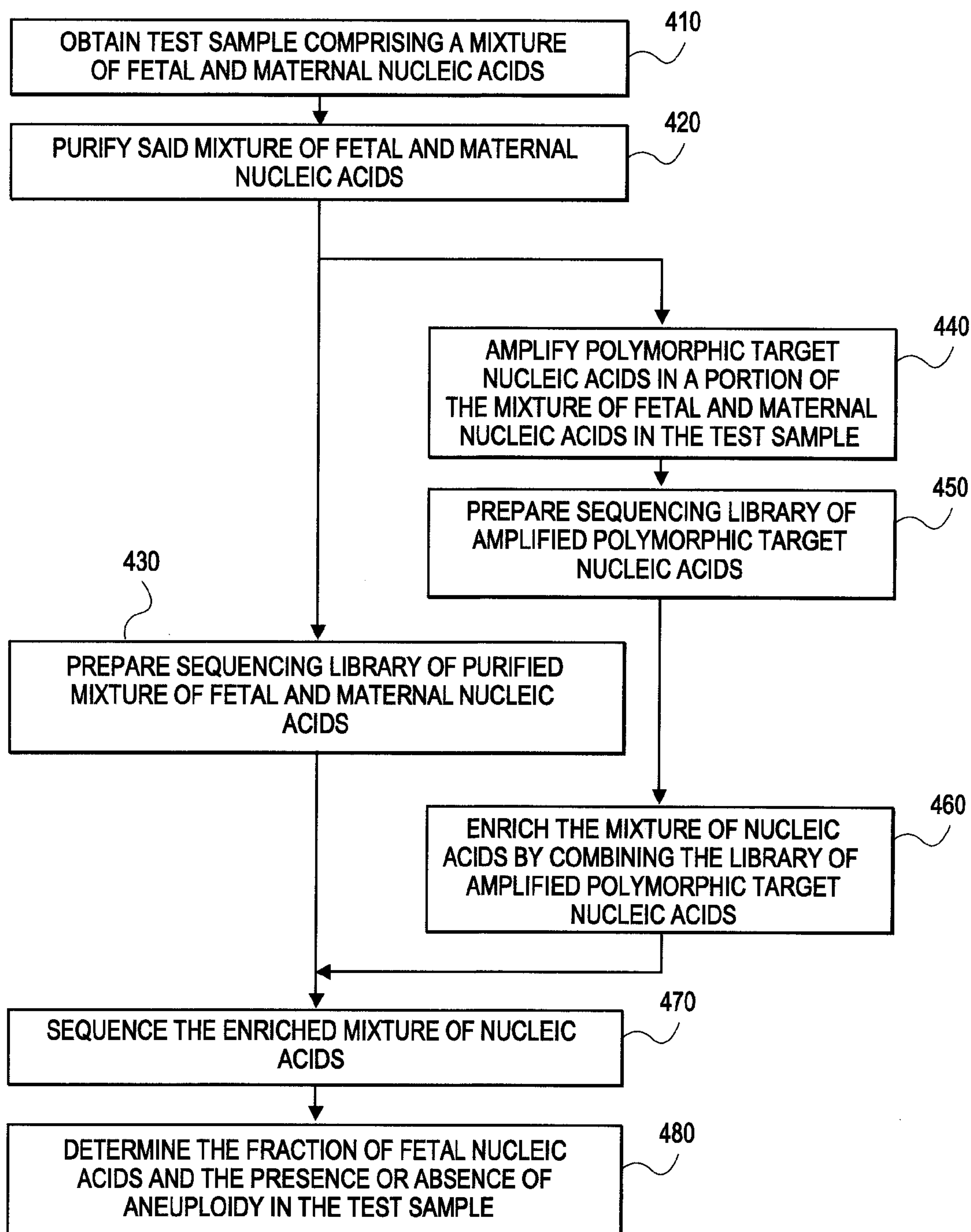
200

FIG. 2

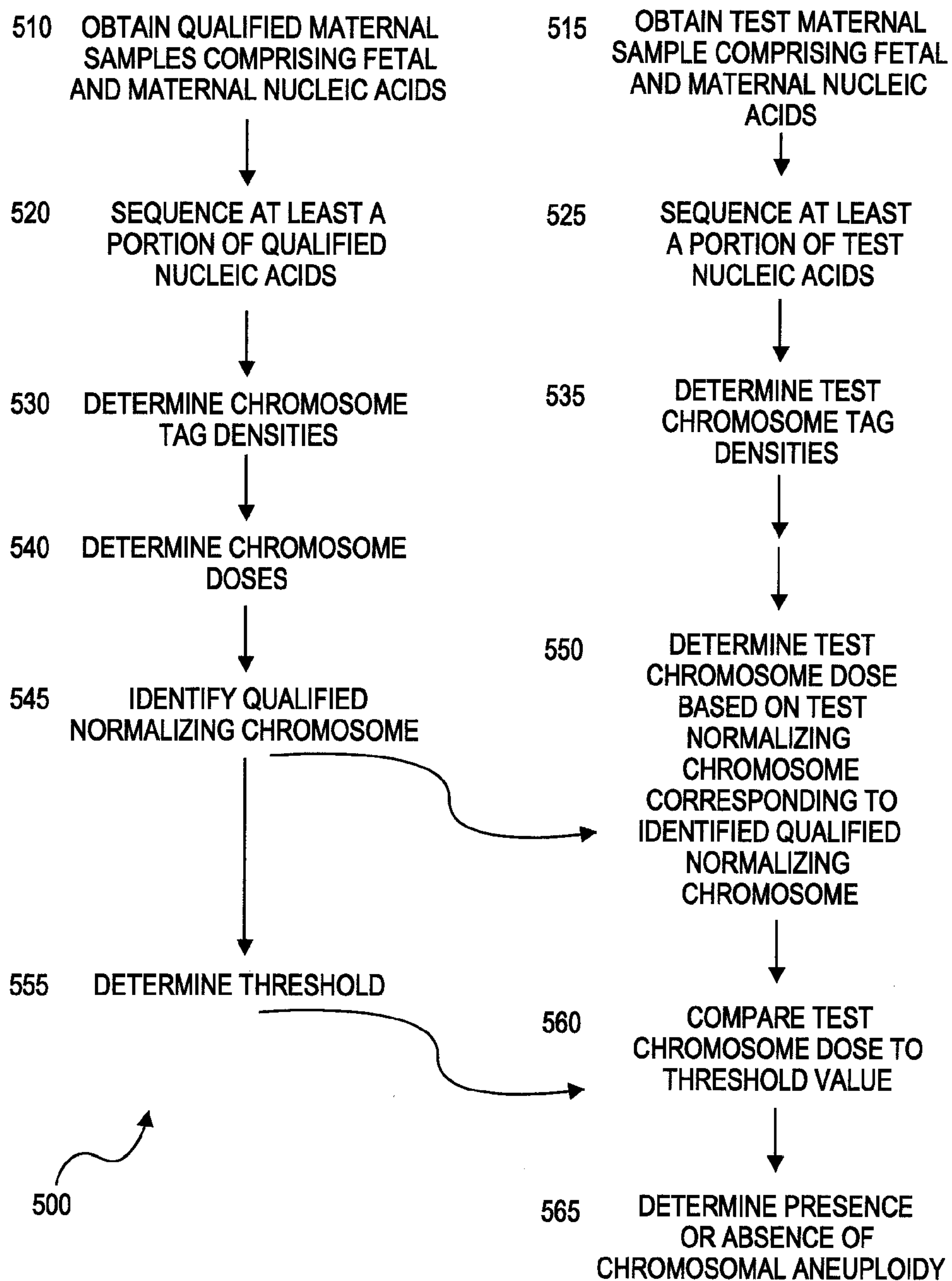
3/18

**FIG. 3**

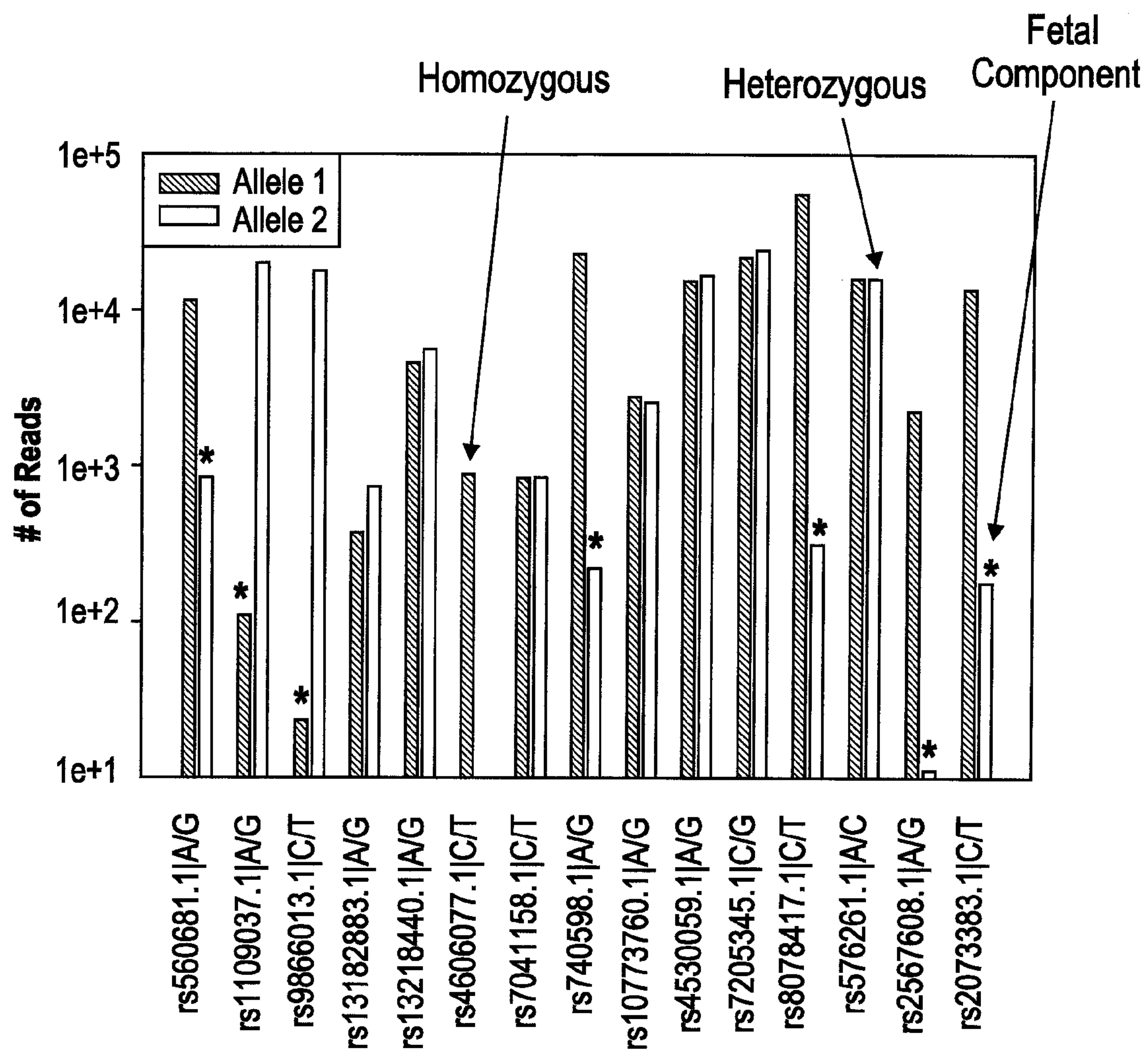
4/18

**FIG. 4**

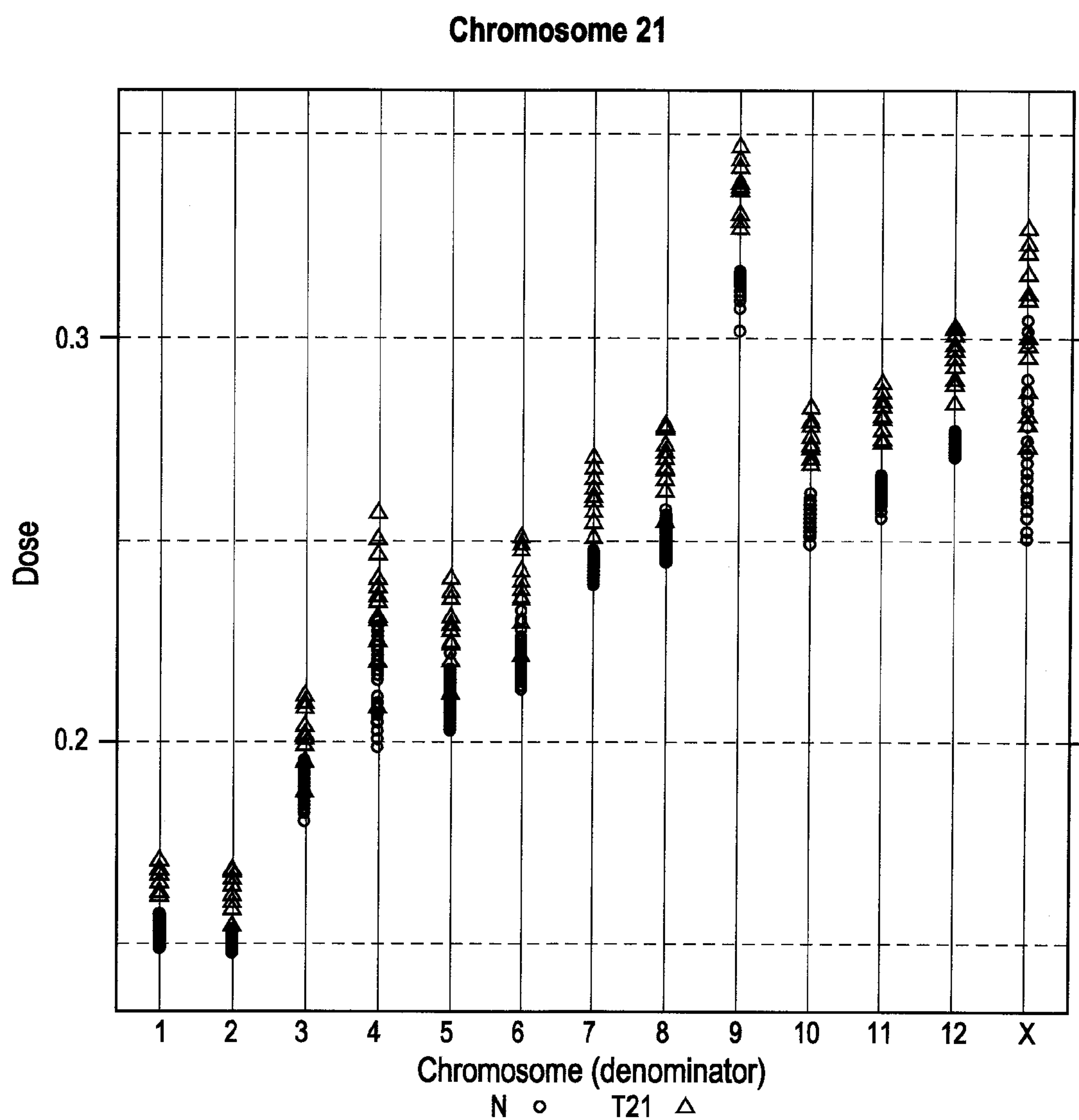
5/18

**FIG. 5**

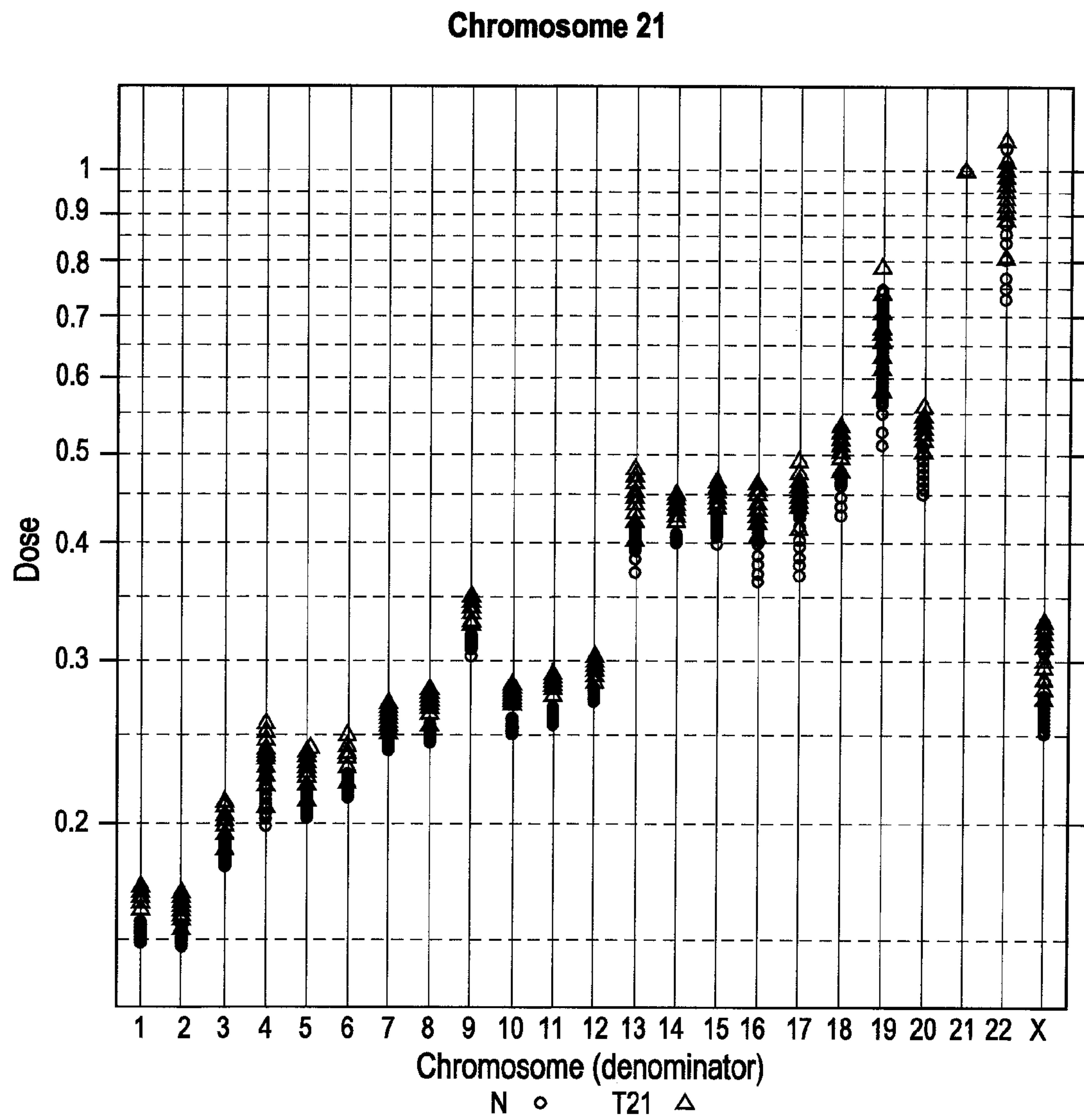
6/18

**FIG. 6**

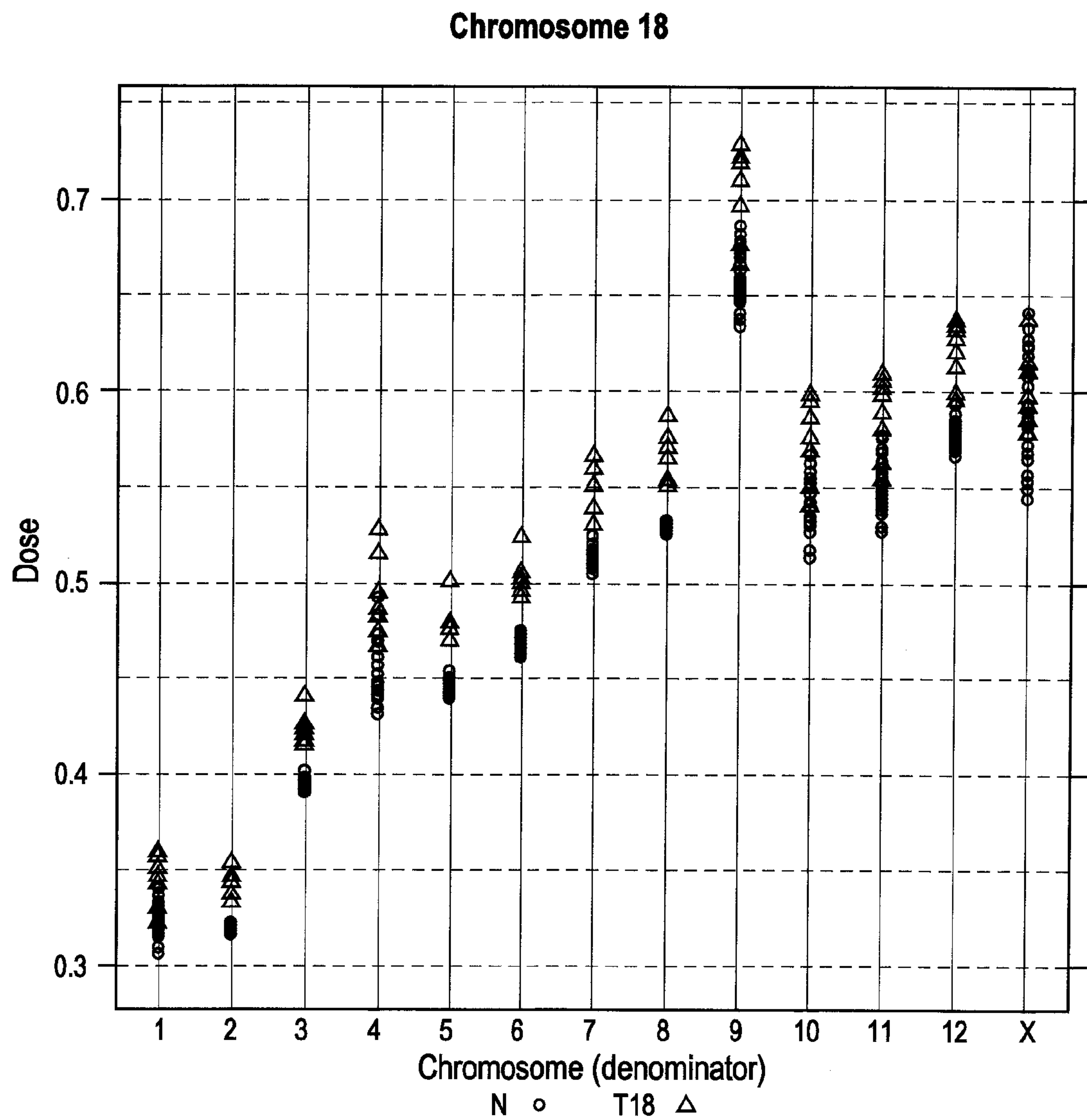
7/18

**FIG. 7A**

8/18

**FIG. 7B**

9/18

**FIG. 8A**

Chromosome 18

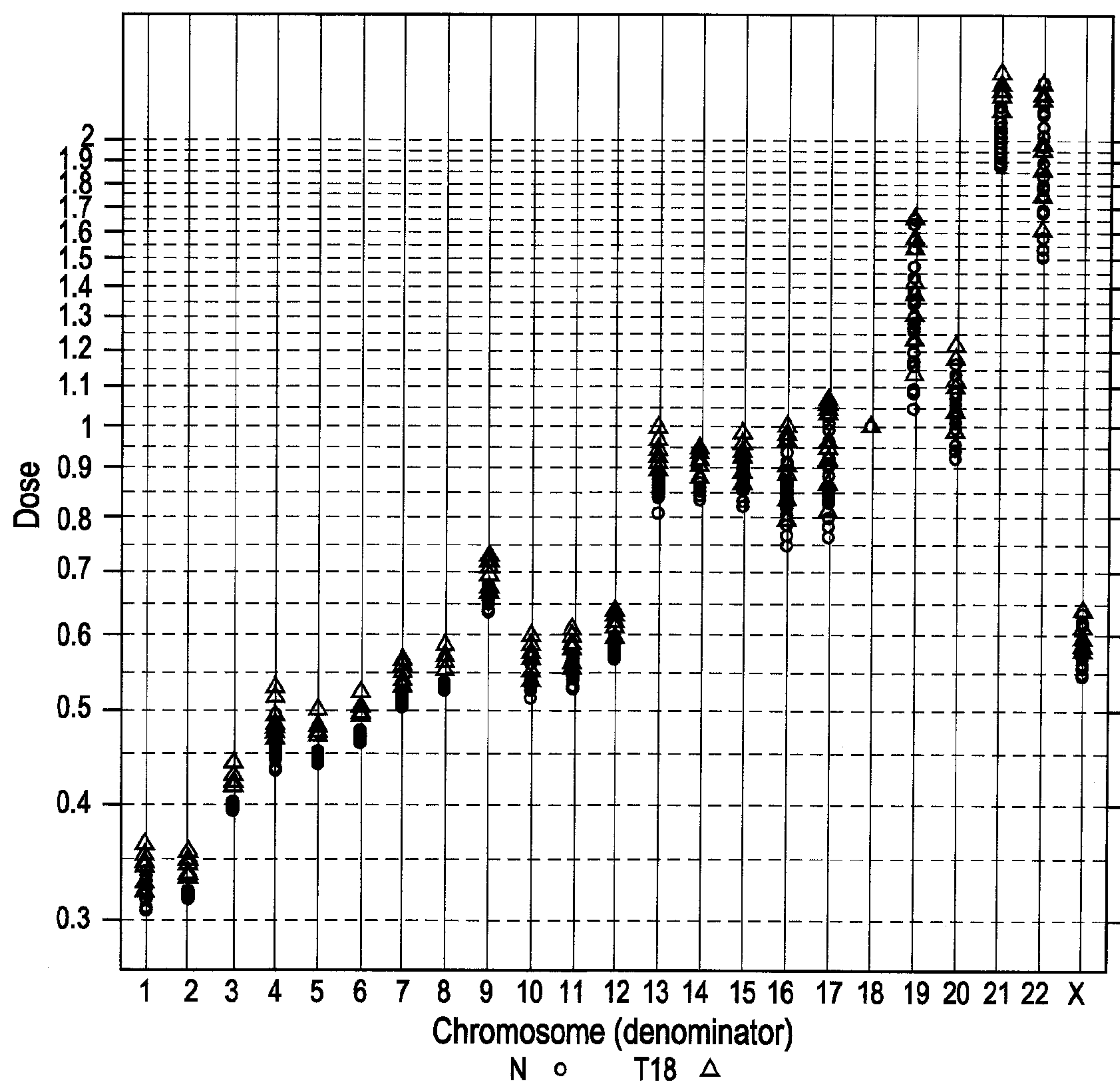
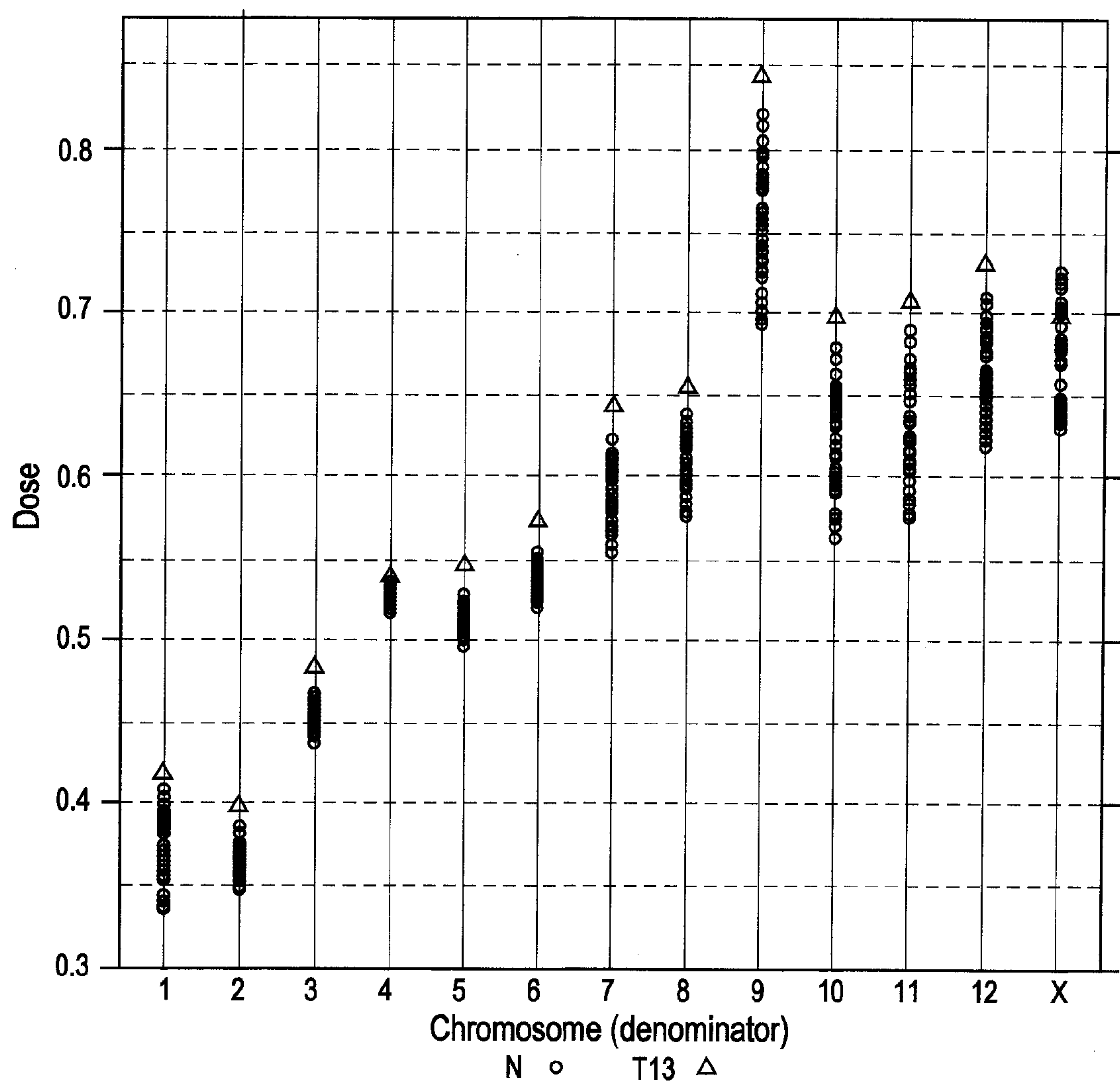


FIG. 8B

11/18

**FIG. 9A**

12/18

Chromosome 13

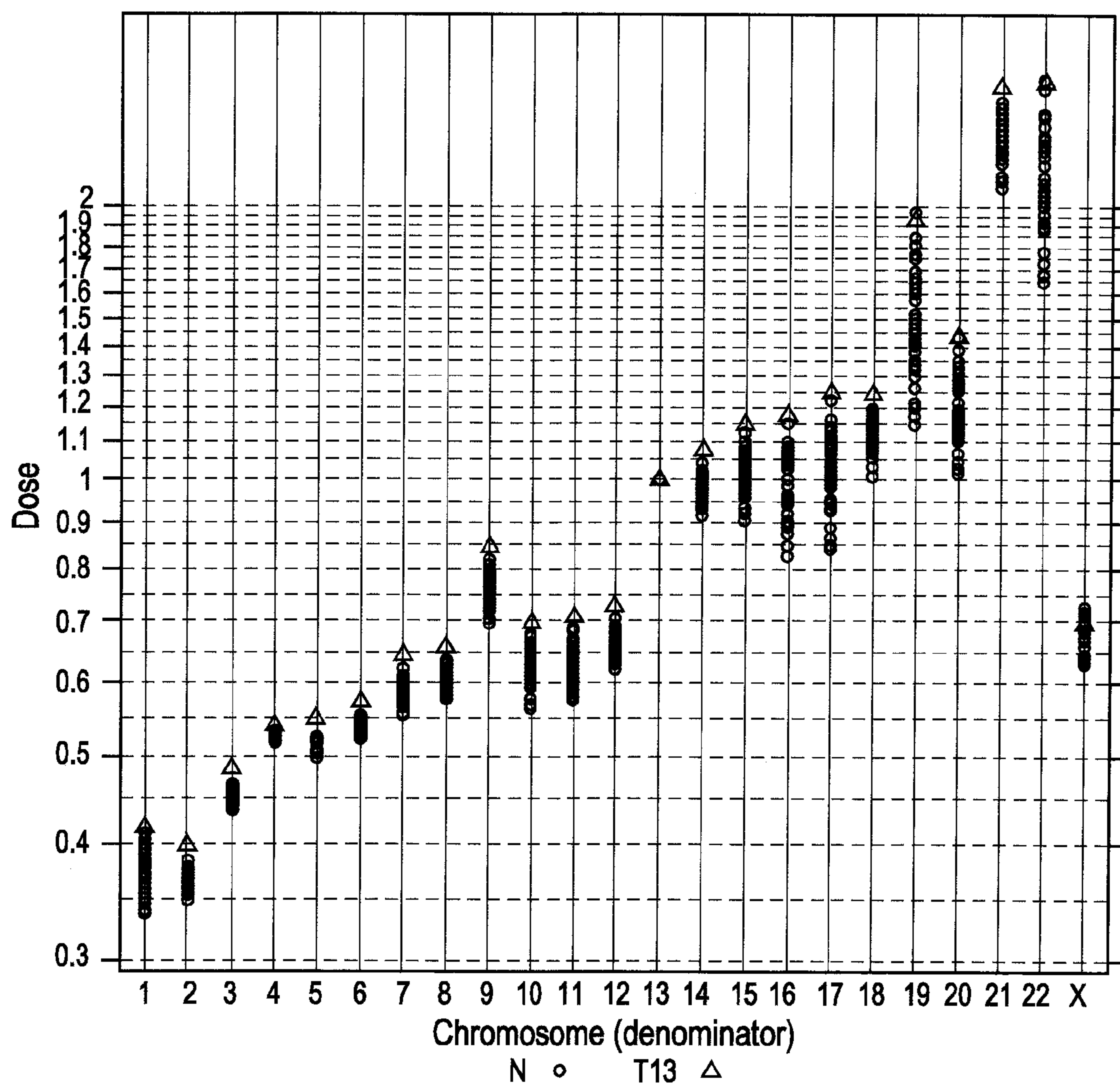
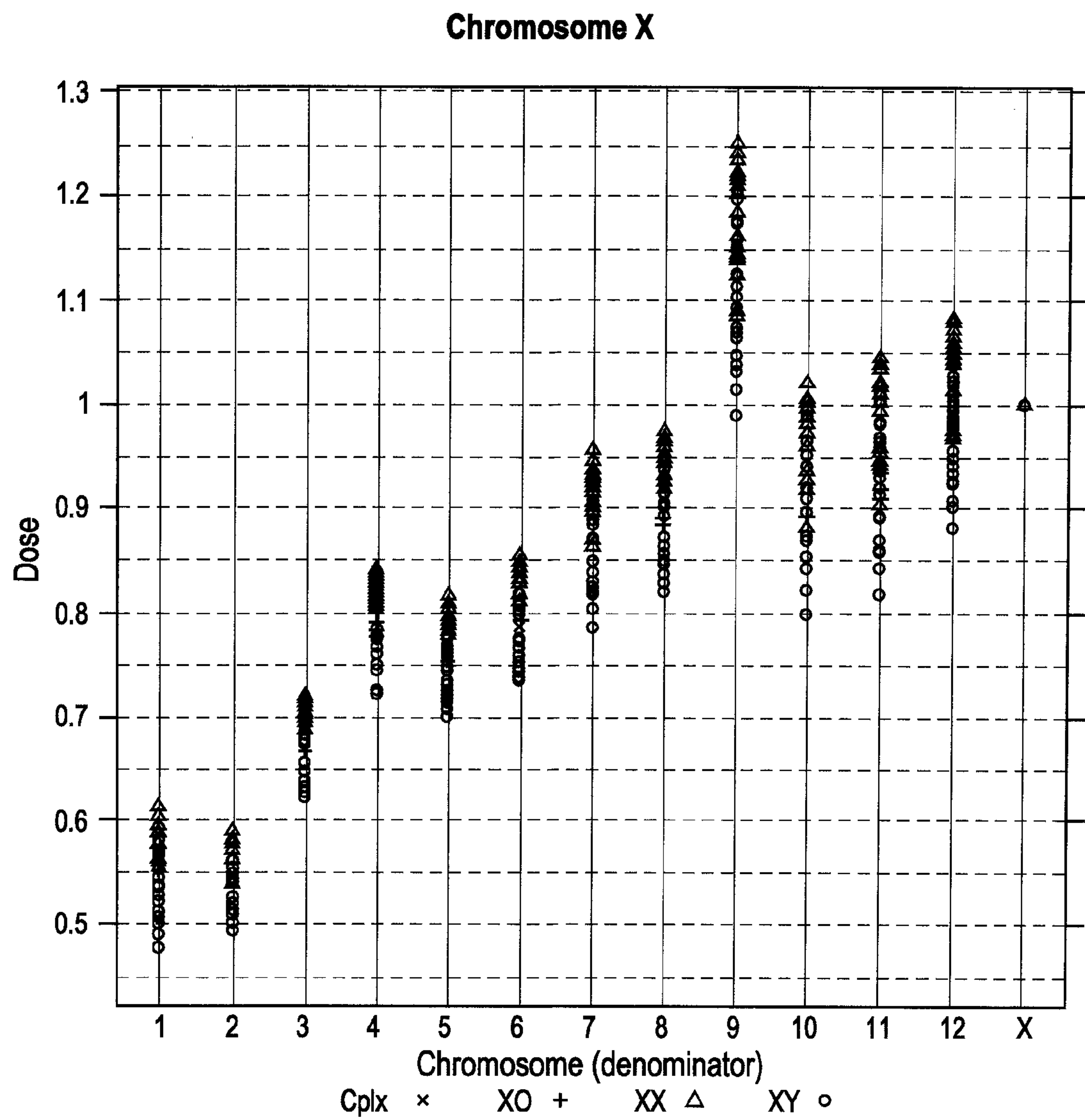
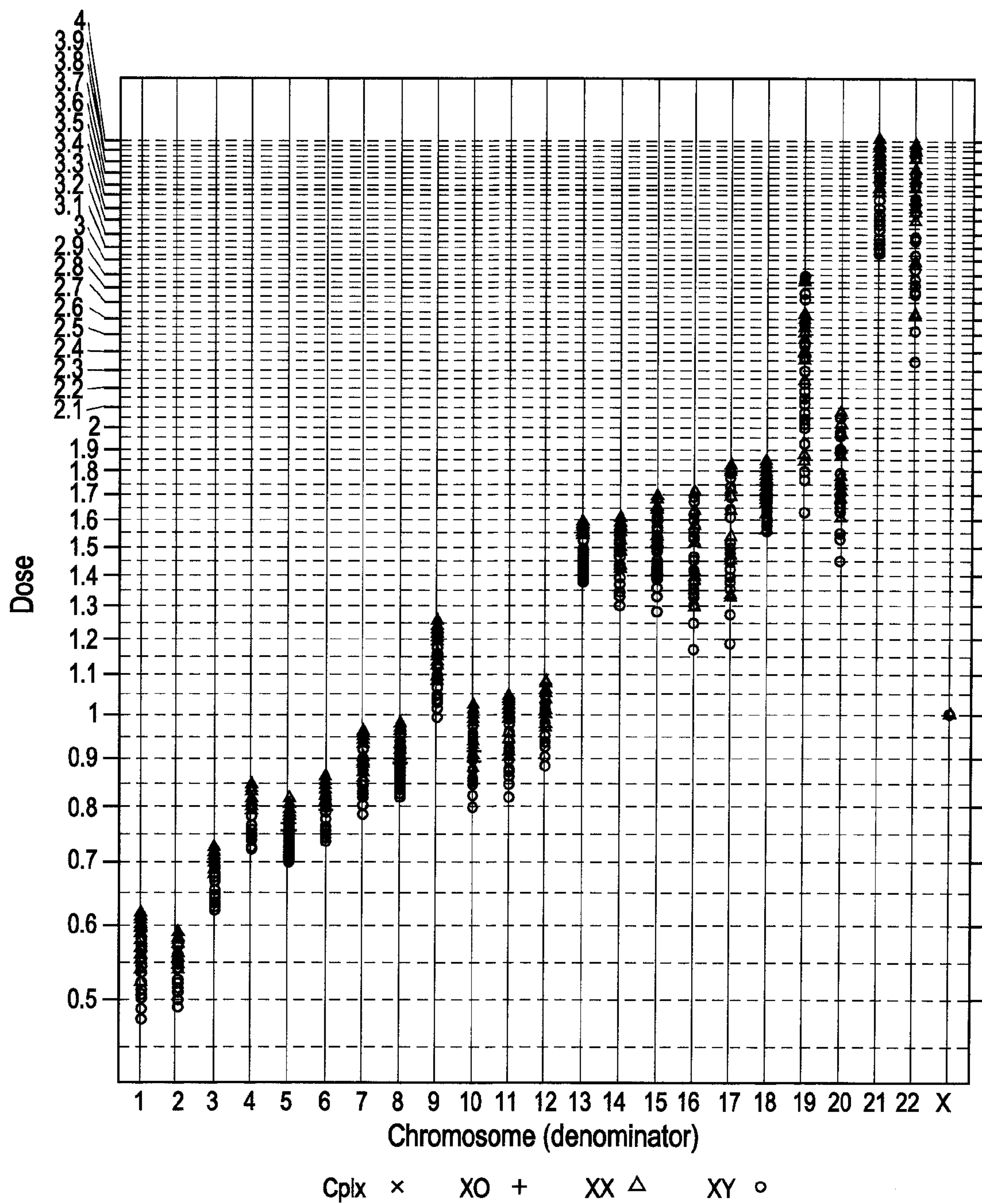


FIG. 9B

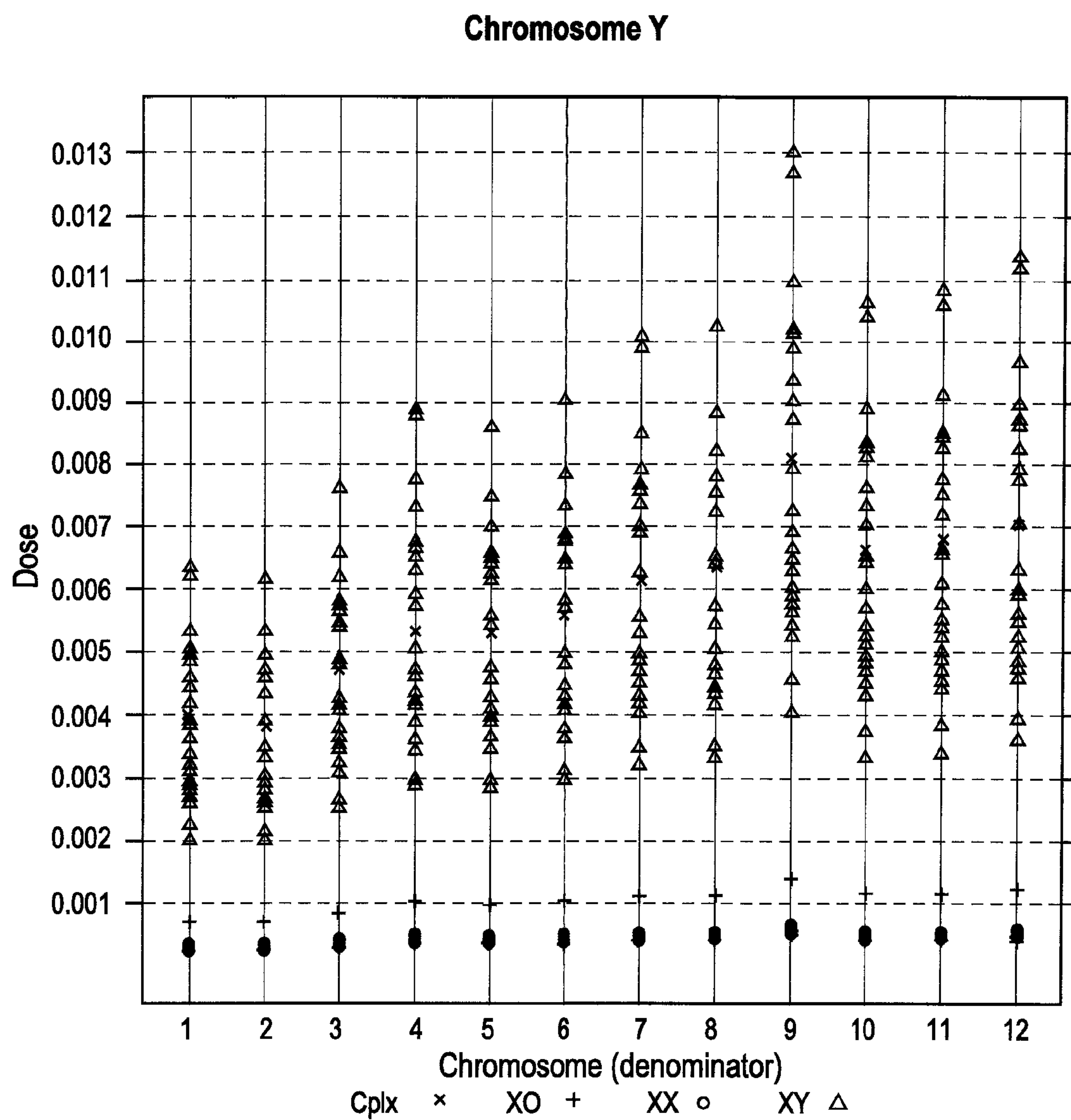
13/18

**FIG. 10A**

14/18

**FIG. 10B**

15/18

**FIG. 11A**

16/18

Chromosome Y

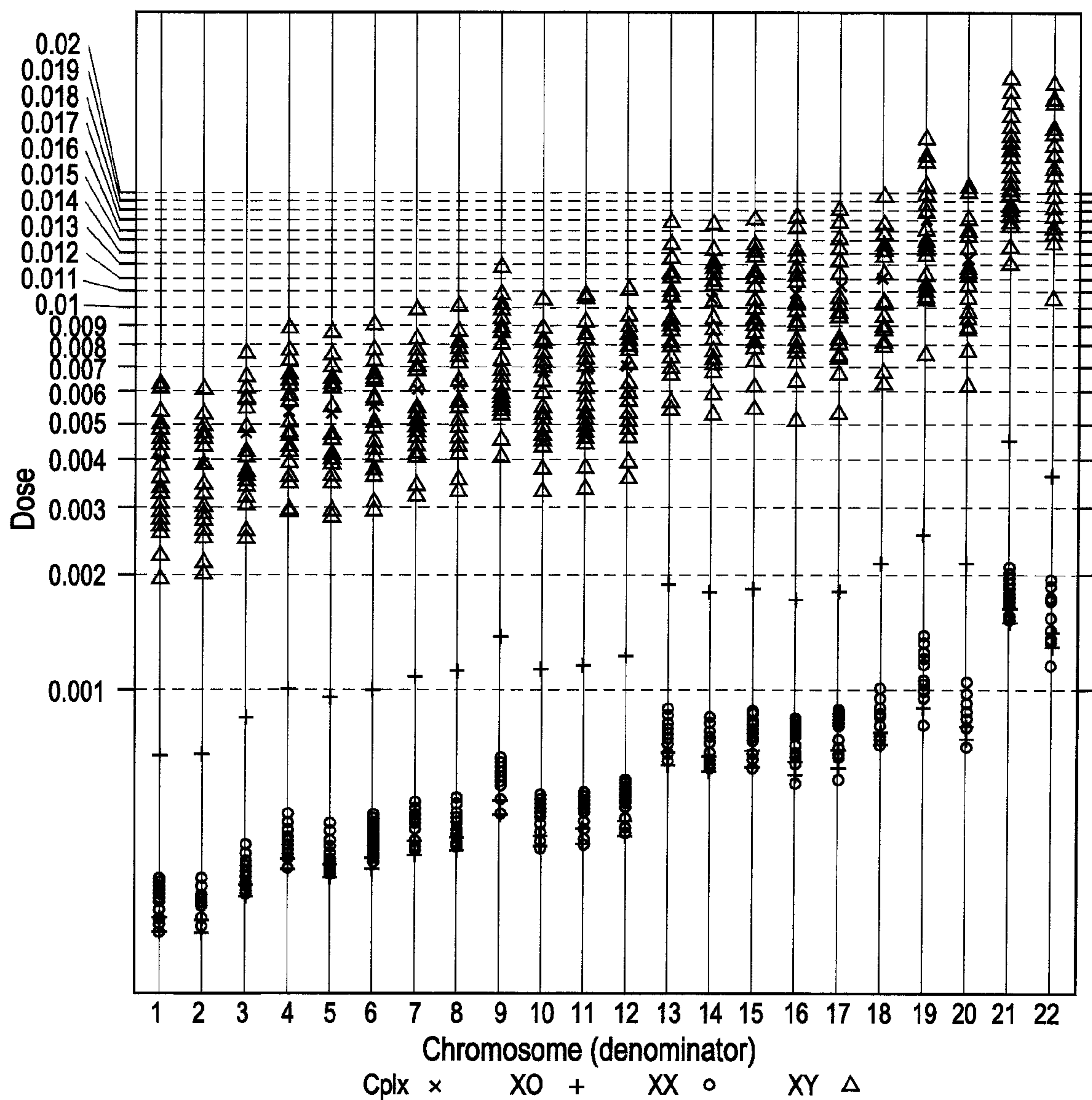
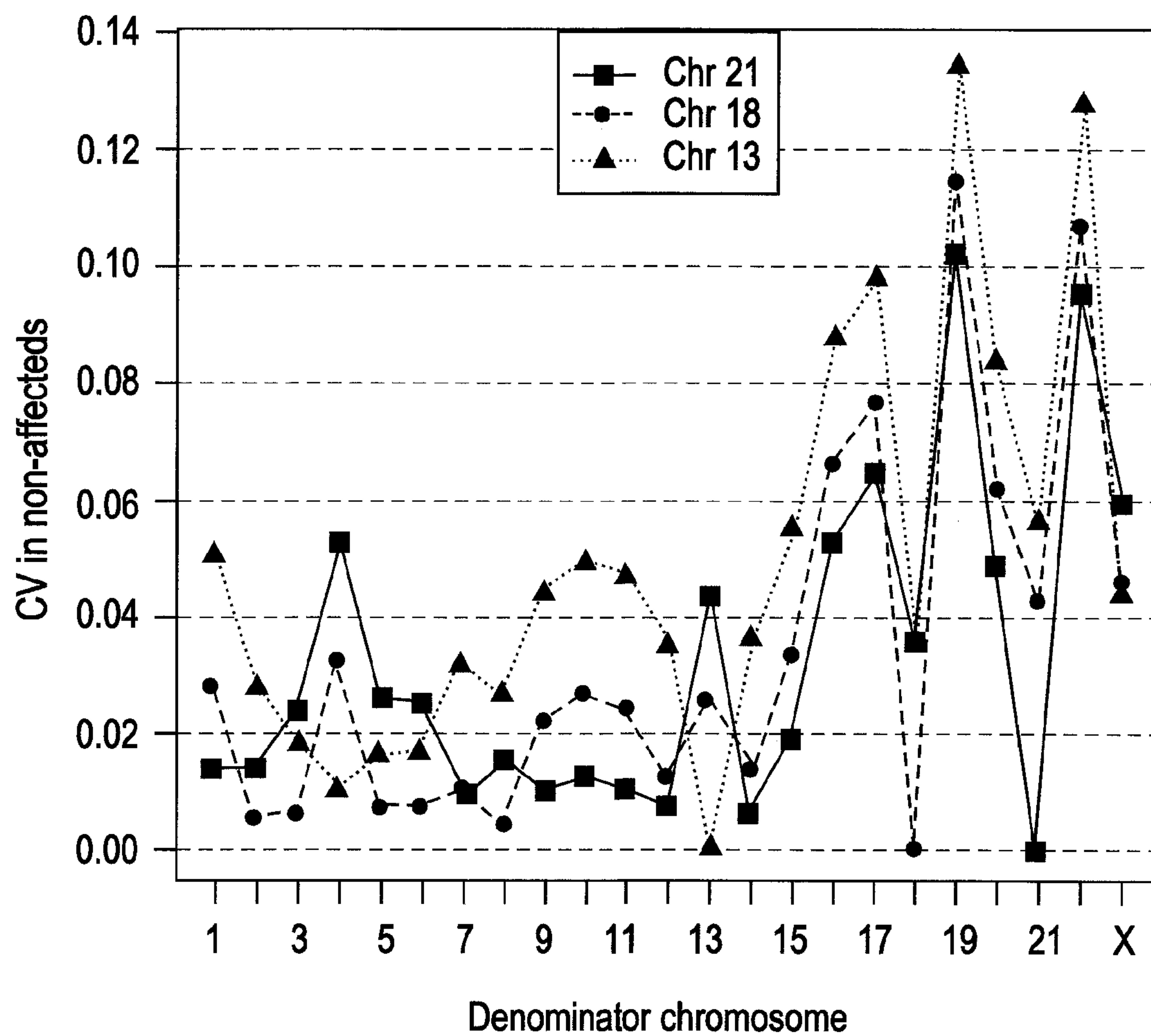
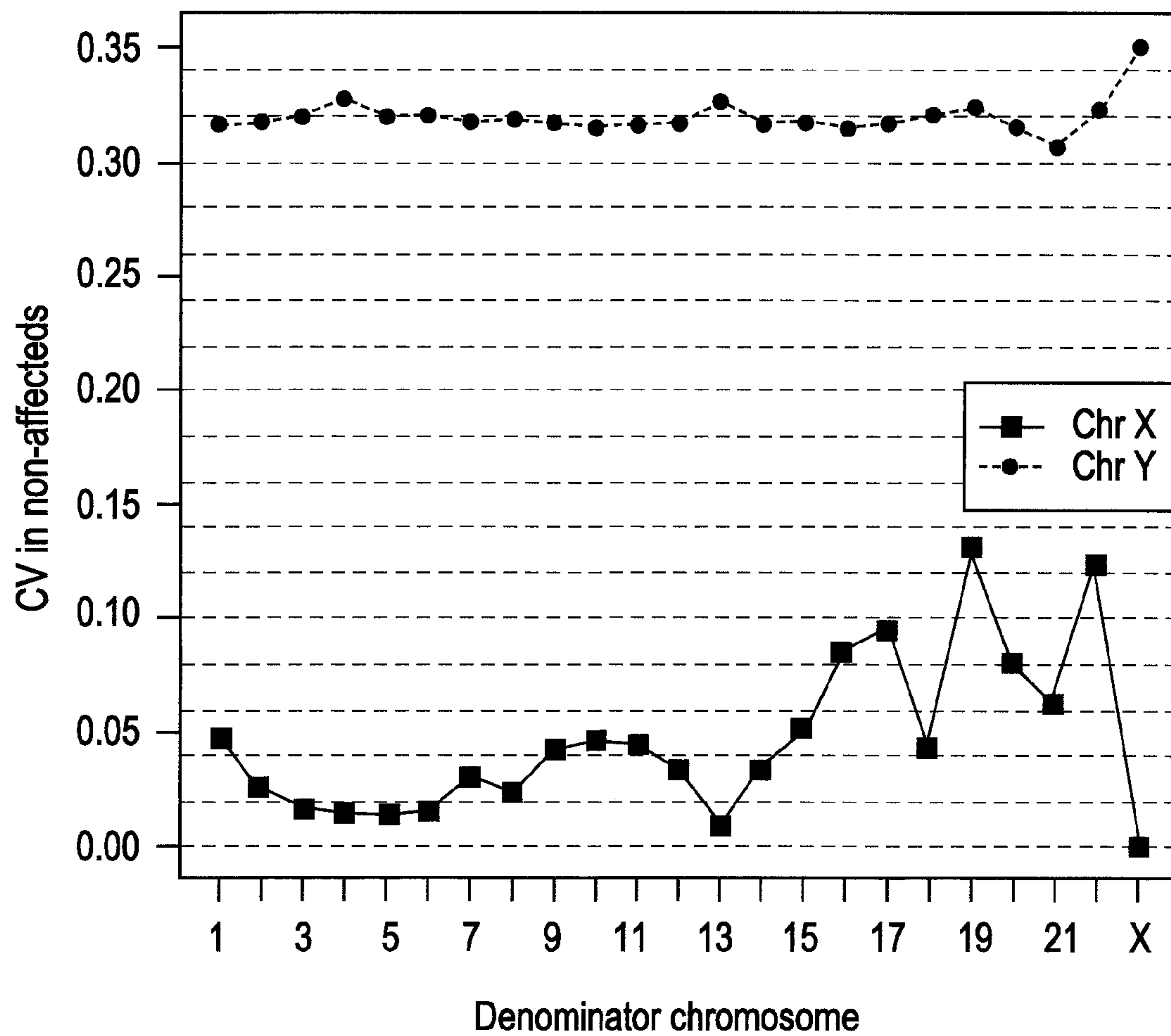


FIG. 11B

17/18

**FIG. 12**

18/18

**FIG. 13**