Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3143728 A1 2020/12/24

(21) 3 143 728

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2020/06/19
(87) Date publication PCT/PCT Publication Date: 2020/12/24
(85) Entrée phase nationale/National Entry: 2021/12/15
(86) N° demande PCT/PCT Application No.: US 2020/038824
(87) N° publication PCT/PCT Publication No.: 2020/257717
(30) Priorité/Priority: 2019/06/21 (US62/865,130)

(51) Cl.Int./Int.Cl. *G16B 20/20* (2019.01), *C12Q 1/68* (2018.01), *G16B 20/10* (2019.01), *G16B 30/00* (2019.01)

(71) Demandeur/Applicant:
COOPERSURGICAL, INC., US

(72) Inventeurs/Inventors:
BURKE, JOHN, US;
RHEES, BRIAN, US;
BLAZEK, JOSHUA DAVID, US;
LARGE, MICHAEL JON, US

(74) Agent: RIDOUT & MAYBEE LLP

(54) Titre : SYSTEME ET PROCEDE DE DETERMINATION DES RELATIONS GENETIQUES ENTRE UN FOURNISSEUR DE SPERME, UNE FOURNISSEUSE D'OVOCYTES ET LE CONCEPTUS RESPECTIF

(54) Title: SYSTEM AND METHOD FOR DETERMINING GENETIC RELATIONSHIPS BETWEEN A SPERM PROVIDER, OOCYTE PROVIDER, AND THE RESPECTIVE CONCEPTUS

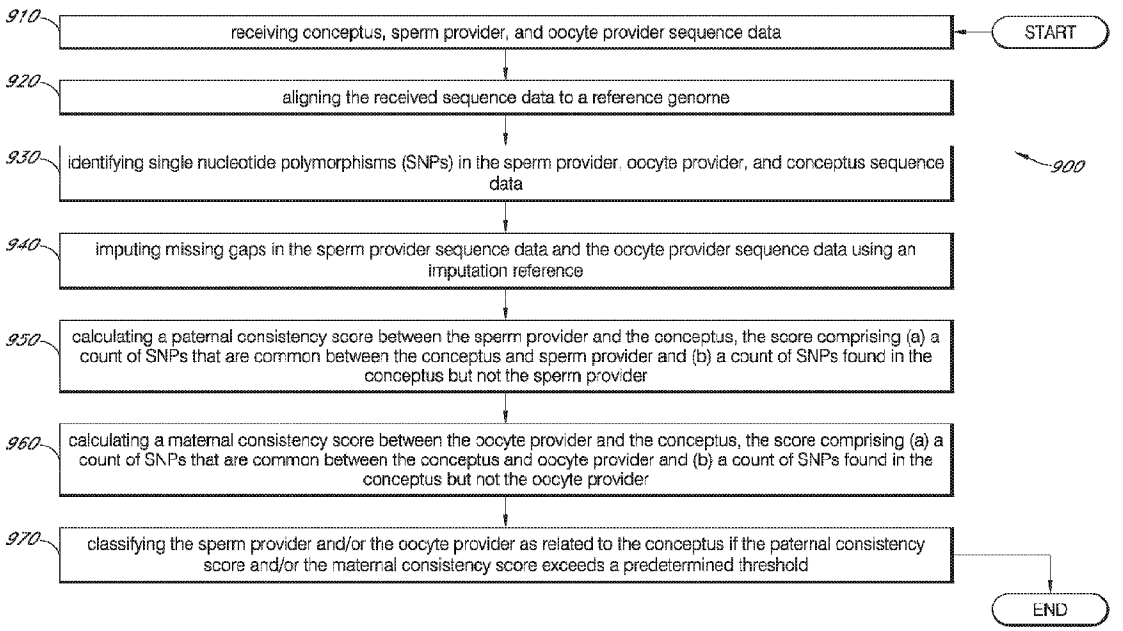


FIG. 9

(57) Abrégé/Abstract:
A method for determining the genetic relationship of a conceptus with sperm and oocyte providers is provided, comprising receiving conceptus, sperm provider, and oocyte provider sequence data; aligning the received sequence data to a reference genome; identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data; imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference; calculating a paternal consistency score between the sperm provider and the conceptus; calculating a maternal consistency score between the oocyte provider and conceptus; and classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds the predetermined threshold.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2020 (24.12.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/257717 A1

(51) International Patent Classification:
G16B 20/10 (2019.01) *G16B 20/20* (2019.01)
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/US2020/038824

(22) International Filing Date:
19 June 2020 (19.06.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/865,130 21 June 2019 (21.06.2019) US

(71) Applicant: **COOPERSURGICAL, INC.** [US/US]; 75 Corporate Drive, Trumbull, Connecticut 06611 (US).

(72) Inventors: **BURKE, John**; 9090 Double Diamond Parkway, Reno, Nevada 89521 (US). **RHEES, Brian**; 9090 Double Diamond Parkway, Reno, Nevada 89521 (US). **BLAZEK, Joshua David**; 1200 Binz Street, Suite 1180, Houston, Texas 77004 (US). **LARGE, Michael Jon**; 1200 Binz Street, Suite 1180, Houston, Texas 77004 (US).

(74) Agent: **NOVAK, Jason J.** et al.; HAYNES and BOONE, LLP, 2323 Victory Avenue, Suite 700, Dallas, California 75219 (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,

(54) Title: SYSTEM AND METHOD FOR DETERMINING GENETIC RELATIONSHIPS BETWEEN A SPERM PROVIDER, OOCYTE PROVIDER, AND THE RESPECTIVE CONCEPTUS

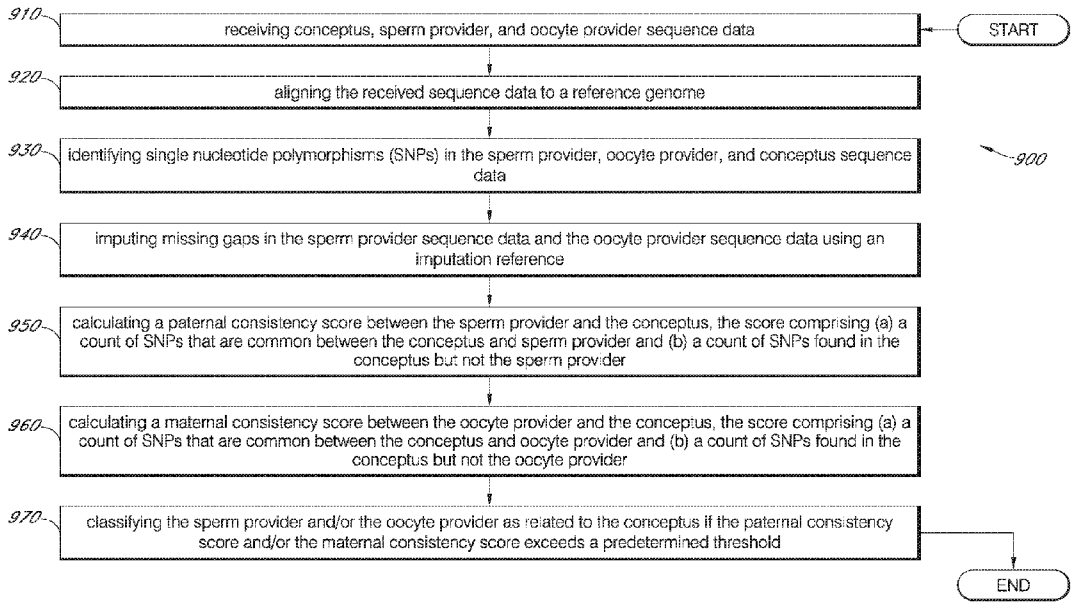


FIG. 9

(57) Abstract: A method for determining the genetic relationship of a conceptus with sperm and oocyte providers is provided, comprising receiving conceptus, sperm provider, and oocyte provider sequence data; aligning the received sequence data to a reference genome; identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data; imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference; calculating a paternal consistency score between the sperm provider and the conceptus; calculating a maternal consistency score between the oocyte provider and conceptus; and classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds the predetermined threshold.

WO 2020/257717 A1

[Continued on next page]

WO 2020/257717 A1

OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

SYSTEM AND METHOD FOR DETERMINING GENETIC RELATIONSHIPS BETWEEN A SPERM PROVIDER, OOCYTE PROVIDER, AND THE RESPECTIVE CONCEPTUS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Patent Application 62/865,130 filed June 21, 2019, which is incorporated herein by reference in its entirety.

INCORPORATION BY REFERENCE

[0002] The disclosures of any patents, patent applications and publications cited herein are incorporated herein by reference in their entirety.

FIELD

[0003] The embodiments provided herein are generally related to systems and methods for analysis of genomic nucleic acids and classification of genomic features. Included among embodiments provided herein are systems and methods relating to determining relatedness between conceptus, oocyte and sperm.

BACKGROUND

[0004] In vitro fertilization (IVF) is an assisted reproductive technology that has become increasingly popular for women of advanced maternal age, couples with difficulties conceiving and as a means for facilitating gestational surrogacy. The process of fertilization involves extracting eggs, retrieving a sperm sample, and then manually combining an egg and sperm in a laboratory setting. The embryo(s) is then implanted in the host uterus to carry the embryo to term. A fourth specimen is produced when embryos are biopsied or cell-free DNA is collected for genetic analysis. Many IVF facilities see thousands of patients a year, producing tens of thousands of embryos. Chain of custody is critical, as a failure to ensure such will result in a couple giving birth to a child that is not theirs. The most common form of genetic analysis done on a preimplantation conceptus, low-pass Next Generation Sequencing (NGS) is not currently able to determine if a conceptus is genetically related to a respective set of parents.

[0005] Current IVF chain of custody concerns are addressed primarily through extensive human supervision, and in some instances, electronic systems that read barcodes or RFID tags. However,

human supervision is error prone, and patient or specimen labeling is fallible. To genetically identify whether a respective sperm contributor, egg contributor, and resultant preimplantation conceptus are related, which can be referred to as Preimplantation Trio Linkage (PTL), SNP arrays can be used. However, this technology is nearing end of life obsolescence and is steadily being replaced by sequencing methods. Moreover, the market lacks options for conducting PTL using low-pass sequencing methods.

[0006] As such, there is a need for more rapid, less costly systems and methods to account for consanguinity detection and sample misidentification so genetic relationships can be properly identified.

SUMMARY

[0007] Provided herein are methods and systems for analysis of genomic nucleic acids and classification of genomic features, including, for the purposes of determining genetic relationships, such as, for example, PTL.

[0008] In accordance with various embodiments, a method is provided for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider. The method can comprise receiving conceptus, sperm provider, and oocyte provider sequence data; aligning the received sequence data to a reference genome; identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data; imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference; calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider; calculating a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

[0009] In accordance with various embodiments, a non-transitory computer-readable medium storing computer instructions is provided for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider. The method can comprise receiving conceptus, sperm

provider, and oocyte provider sequence data; aligning the received sequence data to a reference genome; identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data; imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference; calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider; calculating a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

[0010] In accordance with various embodiments, a system is provided for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider. The system can comprise a data store for receiving conceptus, sperm provider, and oocyte provider sequence data, a computing device communicatively connected to the data store, and a display communicatively connected to the computing device and configured to display a report containing the classified relatedness to the conceptus. The computing device can comprise an alignment engine configured to align the received sequence data to a reference genome; a SNP identification engine configure to identify single nucleotide polymorphisms (SNPs) in the sperm provider sequence data, oocyte provider sequence data, and the conceptus sequence data; an imputation engine configured to impute missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference; and a relatedness engine. The relatedness engine can be configured to calculate a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider; calculate a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and classify the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] For a more complete understanding of the principles disclosed herein, and the advantages thereof, reference is now made to the following descriptions taken in conjunction with the accompanying drawings, in which:

[0012] FIG. 1 is an exemplary flowchart showing a process flow for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, in accordance with various embodiments.

[0013] FIG. 2 is an exemplary flowchart showing a process flow for classifying genetic abnormalities of an offspring, e.g., conceptus or embryo, as either being maternal or paternal in origin, in accordance with various embodiments.

[0014] FIG. 3 is a decision tree diagram depicting a logical flow for detecting and/or classifying the parental origin of identified genetic abnormalities, in accordance with various embodiments.

[0015] FIG. 4 is a bar graph depicting the ratio of the variant alleles of the mother that matched variant alleles of a monosomic embryo out of the total variant alleles analyzed per chromosome (blue bars) compared to the ratio of the variant alleles of the father that matched variant alleles of the embryo out of the total variant alleles analyzed per chromosome (orange bars), in accordance with various embodiments.

[0016] FIGS. 5A and 5B present the results of comparison of the number of variant alleles analyzed of a trisomic embryo that match paternal variant alleles per chromosome to the number of variant alleles analyzed of the embryo that match maternal variant alleles per chromosome, in accordance with various embodiments. FIG. 5A shows a graph of the number of embryo variant alleles shared with paternal source (OvP) vs. the number of embryo variant alleles shared with maternal source (OvM) for each of the 23 chromosomes (blue dots), in accordance with various embodiments. The dotted diagonal line represents the points on the graph at which the number of embryo variant alleles shared with the maternal source would be equal to the number of embryo variant alleles shared with the paternal source for each chromosome. FIG. 5B is another graphic presentation of the results shown in FIG. 6A showing the ratio of counts of shared alleles with mother to counts of shared alleles with father per chromosome, in accordance with various embodiments.

[0017] FIG. 6 is a graph showing the ratio (y-axis) of the number of embryo variant alleles shared with maternal source (OvM) to the number of embryo variant alleles shared with paternal source (OvP) with respect to location on chromosome 6 (x-axis in base pairs), in accordance with various embodiments.

[0018] FIG. 7 shows a graph of the number of embryo variant alleles shared with paternal source (OvP) vs. the number of embryo variant alleles shared with maternal source (OvM) for each of the 23 chromosomes (blue dots), in accordance with various embodiments. The dotted diagonal line represents the points on the graph at which the number of embryo variant alleles shared with the maternal source would be equal to the number of embryo variant alleles shared with the paternal source for each chromosome.

[0019] FIG. 8 is a schematic diagram of a system for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, in accordance with various embodiments.

[0020] FIG. 9 is an exemplary flowchart showing a method for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, in accordance with various embodiments.

[0021] FIG. 10 shows a graph of parental consistency and novel scores on the initial $N = 14$ family data set, in accordance with various embodiments.

[0022] FIG. 11 shows a parental matching cluster plot, in accordance with various embodiments.

[0023] FIG. 12 shows a parental matching cluster plot, in accordance with various embodiments.

[0024] FIG. 13 shows a graph of relative maternal and paternal novelty and likeness scores in determining the genetic relationship of a conceptus, in accordance with various embodiments.

[0025] FIG. 14 shows a graph of relative maternal and paternal novelty and likeness scores in determining the genetic relationship of a conceptus, in accordance with various embodiments.

[0026] FIG. 15 is a block diagram illustrating a computer system for use in performing methods provided herein, in accordance with various embodiments.

[0027] It is to be understood that the figures are not necessarily drawn to scale, nor are the objects in the figures necessarily drawn to scale in relationship to one another. The figures are depictions that are intended to bring clarity and understanding to various embodiments of apparatuses,

systems, and methods disclosed herein. Wherever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts. Moreover, it should be appreciated that the drawings are not intended to limit the scope of the present teachings in any way.

[0028] The above-identified figures are provided by way of representation and not limitation. The figures may show simplified or partial views, and the dimensions of elements in the figures may be exaggerated or otherwise not in proportion. In addition, as the terms "on," "attached to," "connected to," "coupled to," or similar words are used herein, one element (e.g., a material, a layer, a substrate, etc.) can be "on," "attached to," "connected to," or "coupled to" another element regardless of whether the one element is directly on, attached to, connected to, or coupled to the other element or there are one or more intervening elements between the one element and the other element. In addition, where reference is made to a list of elements (e.g., elements a, b, c), such reference is intended to include any one of the listed elements by itself, any combination of less than all of the listed elements, and/or a combination of all of the listed elements. Section divisions in the specification are for ease of review only and do not limit any combination of elements discussed.

DETAILED DESCRIPTION

[0029] Provided herein are methods and systems for analysis of genomic nucleic acids and classification of genomic features, including, for example, genetic abnormalities. In some embodiments, the methods and systems are used in the determining the genetic relationship of a conceptus with a sperm provider and oocyte provide.

[0030] It should be noted that throughout the description, any reference to conceptus also can include an embryo. As such, both terms can be used interchangeably and these use of one term versus the other in no way limits the various embodiments herein to one type of sample versus the other. The sample analyzed is, in its broadest terms, the product of conception. As such, the terms conceptus or embryo can be used interchangeably with fetus, miscarriage, child, etc., with the use of any such term in no way limiting the various embodiments herein to one type of sample versus the other.

[0031] The same can be said with reference to the mother and father (or parents) of the sample analyzed. As the terms mother and father can have a social construct, it is important to understand the breadth attributed to such categories. In accordance with various embodiments, the use of

terms mother, father, or parents is generic as to known identify and, as such, are not to be interpreted as meaning that the parentage of the sample analyzed is known. As will be detailed herein, various embodiments will be discussed in which pattern of inheritance or genetic relationship, for example, is determined. In either case, the parentage can be known or can be unknown yet determined via the various embodiments herein. Further, the term mother can refer to maternal DNA, presumed maternal DNA, eggs/oocytes of known and unknown genetic relationship, egg donors, egg or gestational carriers, blood, and so on. Likewise, the term father can refer to paternal DNA, presumed paternal DNA, sperm, sperm donors, blood, and so on.

[0032] Unless otherwise defined, scientific and technical terms used in connection with the present teachings described herein shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures utilized in connection with, and techniques of, cell and tissue culture, molecular biology, and protein and oligo- or polynucleotide chemistry and hybridization described herein are those well-known and commonly used in the art. Standard techniques are used, for example, for nucleic acid purification and preparation, chemical analysis, recombinant nucleic acid, and oligonucleotide synthesis. Enzymatic reactions and purification techniques are performed according to manufacturer's specifications or as commonly accomplished in the art or as described herein. Standard molecular biological techniques and procedures described herein are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the instant specification. See, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual (Third ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 2000). The nomenclatures utilized in connection with, and the laboratory procedures and standard techniques described herein are those well-known and commonly used in the art.

[0033] A “polynucleotide”, “nucleic acid”, or “oligonucleotide” refers to a linear polymer of nucleosides (including deoxyribonucleosides, ribonucleosides, or analogs thereof) joined by internucleosidic linkages. Typically, a polynucleotide comprises at least three nucleosides. Usually oligonucleotides range in size from a few monomeric units, e.g. 3-4, to several hundreds of monomeric units. Whenever a polynucleotide such as an oligonucleotide is represented by a sequence of letters, such as “ATGCCTG,” it will be understood that the nucleotides are in 5′->3′ order from left to right and that “A” denotes deoxyadenosine, “C” denotes deoxycytidine, “G” denotes

deoxyguanosine, and "T" denotes thymidine, unless otherwise noted. The letters A, C, G, and T may be used to refer to the bases themselves, to nucleosides, or to nucleotides comprising the bases, as is standard in the art.

[0034] DNA (deoxyribonucleic acid) is a chain of nucleotides containing 4 types of nucleotides; A (adenine), T (thymine), C (cytosine), and G (guanine), and RNA (ribonucleic acid) is comprised of 4 types of nucleotides; A, U (uracil), G, and C. Certain pairs of nucleotides specifically bind to one another in a complementary fashion (called complementary base pairing). That is, adenine (A) pairs with thymine (T) (in the case of RNA, however, adenine (A) pairs with uracil (U)), and cytosine (C) pairs with guanine (G). When a first nucleic acid strand binds to a second nucleic acid strand made up of nucleotides that are complementary to those in the first strand, the two strands bind to form a double strand. As used herein, "nucleic acid sequencing data," "nucleic acid sequencing information," "nucleic acid sequence," "genomic sequence," "genetic sequence," or "fragment sequence," or "nucleic acid sequencing read" denotes any information or data that is indicative of the order of the nucleotide bases (e.g., adenine, guanine, cytosine, and thymine/uracil) in a molecule (e.g., whole genome, whole transcriptome, exome, oligonucleotide, polynucleotide, fragment, etc.) of DNA or RNA. It should be understood that the present teachings contemplate sequence information obtained using all available varieties of techniques, platforms or technologies, including, but not limited to: capillary electrophoresis, microarrays, ligation-based systems, polymerase-based systems, hybridization-based systems, direct or indirect nucleotide identification systems, pyrosequencing, ion- or pH-based detection systems, electronic signature-based systems, etc.

[0035] As used herein, the term "cell" is used interchangeably with the term "biological cell." Non-limiting examples of biological cells include eukaryotic cells, plant cells, animal cells, such as mammalian cells, reptilian cells, avian cells, fish cells or the like, prokaryotic cells, bacterial cells, fungal cells, protozoan cells, or the like, cells dissociated from a tissue, such as muscle, cartilage, fat, skin, liver, lung, neural tissue, and the like, immunological cells, such as T cells, B cells, natural killer cells, macrophages, and the like, embryos (e.g., zygotes), oocytes, ova, sperm cells, hybridomas, cultured cells, cells from a cell line, cancer cells, infected cells, transfected and/or transformed cells, reporter cells and the like. A mammalian cell can be, for example, from a human, mouse, rat, horse, goat, sheep, cow, primate or the like.

[0036] A genome is the genetic material of a cell or organism, including animals, such as mammals, e.g., humans and comprises nucleic acids, such as DNA. In humans, total DNA includes, for example, genes, noncoding DNA and mitochondrial DNA. The human genome typically contains 23 pairs of linear chromosomes: 22 pairs of autosomal chromosomes (autosomes) plus the sex-determining X and Y chromosomes. The 23 pairs of chromosomes include one copy from each parent. The DNA that makes up the chromosomes is referred to as chromosomal DNA and is present in the nucleus of human cells (nuclear DNA). Mitochondrial DNA is located in mitochondria as a circular chromosome, is inherited from only the female parent, and is often referred to as the mitochondrial genome as compared to the nuclear genome of DNA located in the nucleus.

[0037] As used herein, the phrase "genomic feature" refers to a defined or specified genome element or region. In some instances, the genome element or region can have some annotated structure and/or function (e.g., a chromosome, a gene, protein coding sequence, mRNA, tRNA, rRNA, repeat sequence, inverted repeat, miRNA, siRNA, etc.) or be a genetic/genomic variant (e.g., single nucleotide polymorphism/variant, insertion/deletion sequence, copy number variation, inversion, etc.) which denotes one or more nucleotides, genome regions, genes or a grouping of genome regions or genes (in DNA or RNA) that have undergone changes as referenced against a particular species or sub-populations within a particular species due to, for example, mutations, recombination/crossover or genetic drift.

[0038] Ploidy refers to the number of sets (designated as *n*) of homologous chromosomes in the genome of a cell or organism. For example, a cell or organism having one set of chromosomes is referred to as monoploid. A cell or organism having two sets of homologous chromosomes (2n) is referred to as diploid. Polyploidy is the condition in which cells, e.g., an embryo, or organisms possess more than two complete haploid sets of chromosomes. Haploid refers to cells that have half of the usual complete set of somatic cell chromosomes of an organism. For example, gametes, or reproductive (sex) cells, such as ova and sperm cells in humans, are haploid. Fusion of haploid gametes during fertilization yields a diploid zygote containing one set of homologous chromosomes from the female gamete and one set of homologous chromosomes from the male gamete. A human embryo with a normal number of autosomes (22) and a single sex chromosome pair (XX or XY) is referred to as a euploid embryo. Thus, for humans, the euploid condition is diploid. In various embodiments, the phrase "all chromosomes" can include all autosomes and sex chromosomes. In various embodiments, the phrase "all chromosomes" does not include sex chromosomes.

[0039] The term "allele" refers to alternative forms of a gene. In humans or other diploid organisms, there are two alleles at each genetic locus. Alleles are inherited from each parent: one allele is inherited from the mother and one allele is inherited from the father. A pair of alleles represents the genotype of a gene. If the two alleles at a particular locus are identical, the genotype is referred to as homozygous. If there are differences in the two alleles at a particular locus, the genotype is referred to as heterozygous.

[0040] The term "haplotype" refers to a set, or combination, of variations, or polymorphisms, in a chromosome that tend to co-segregate due to proximity in the chromosome. Haplotypes can be described with respect to combinations of variations in a single gene, multiple genes or in sequences between genes. Because of the closeness of the variations in a haplotype, there tends to be little to no recombination or crossover of the locations in which the variations occur and they tend to pass through generations and be inherited together.

[0041] As used herein, the phrase "genetic abnormality" refers to a change in a genome relative to a normal, wild-type or reference genome. Generally, genetic abnormalities include chromosomal abnormalities and gene defects. Typically, gene defects include alterations including, but not limited to, single base mutations, substitutions, insertions and deletions and copy number variations. Chromosomal abnormalities include alterations in chromosome number or structure, e.g., duplication and deletion, such as a repeat or loss of a region of a chromosome, inversion and translocation. A common chromosomal abnormality is referred to as aneuploidy which is an abnormal chromosome number due to an extra or missing chromosome. For example, monosomy in a human is an abnormality characterized by a chromosome with a copy loss (only one copy instead of the normal two copies). Trisomy in a human is an abnormality characterized by a chromosome copy gain (three copies instead of the normal two copies). An embryo with an abnormal number of chromosomes is referred to as an aneuploid embryo. Most aneuploidies are of maternal origin and result from errors in segregation during meiosis. Thus, meiotic aneuploidies will occur in all cells of an embryo. However, mitotic errors are also common in human preimplantation embryos and can result in mitotic aneuploidies and chromosomally mosaic embryos having multiple populations of cells (e.g., some cells being aneuploid and some being euploid). Polyploidy in a human cell is an abnormality in which the cell, e.g., in an embryo, possesses more than two complete sets of chromosomes. Examples of polyploidy include triploidy (3n) and tetraploidy (4n). Polyploidy in humans can occur in several forms that result in having either balanced sex chromosomes (undetectable by current CNV methods) or unbalance sex chromosomes (detectable by CNV

methods). A balanced-sex polyploidy in humans contains 3 or more complete copies of the genome in which each copy contains only X chromosomes (e.g., 69:XXX or 92:XXXX) or contains and equivalent number of X and Y chromosomes (e.g., 92:XXYY). An unbalanced sex polyploidy in humans contains 3 or more complete copies of the genome in which at least one copy contains a Y chromosome (e.g., 69:XXY, 69:XYY) and does not contain an equivalent copy number of X and Y chromosomes. Chromosomal abnormalities can have a number of different effects on cells and organisms, including miscarriages and genetic disorders and diseases.

[0042] In general, genomic variants can be identified using a variety of techniques, including, but not limited to: array-based methods (e.g., DNA microarrays, etc.), real-time/digital/quantitative PCR instrument methods and whole or targeted nucleic acid sequencing systems (e.g., NGS systems, capillary electrophoresis systems, etc.). With nucleic acid sequencing, resolution or coverage can be at one or more levels and is some cases is available at single base resolution.

[0043] As used herein, the phrase "pattern of inheritance" refers to the manner of transmission of a genomic feature, such as, for example, aneuploidy, in the genome of a cell, embryo or organism from parent cells or organisms such as diploid cells and organisms. For example, in humans, the offspring, e.g., embryo, receives one gene allele from each parent (one maternal and one paternal) which then make up the two alleles in the diploid cells of the offspring. A pattern of inheritance of a particular allele or genomic feature in an offspring or embryo defines which parent transmitted the genomic feature to the offspring. The parent from whom the genomic feature was transmitted to the offspring or embryo is referred to as the parent of origin.

[0044] As used herein, "offspring" refers to the product of the union of gametes (e.g., female and male germ cells) and includes, but is not limited to, e.g., a blastomere, a zygote, an embryo, fetus, neonate or child. Offspring DNA can be obtained from any source, including, for example, a blastomere biopsy, a trophectoderm biopsy, an inner cell mass biopsy, a blastocoel biopsy, embryo spent media, cfDNA, products of conception, chorionic villus samples and/or amniocentesis.

[0045] As used herein, "parent" or "genetic parent" refers to a contributor of a gamete to an offspring and includes, for example, egg and sperm donors so long as the gamete DNA originates from the donor.

[0046] The phrase "mosaic embryo" denotes embryos containing two or more cytogenetically distinct cell lines. For example, a mosaic embryo can contain cell lines with different types of

aneuploidy or a mixture of euploid and genetically abnormal cells containing DNA with genetic variants that may be deleterious to the viability of the embryo during pregnancy.

[0047] The phrase "next generation sequencing" (NGS) refers to sequencing technologies having increased throughput as compared to traditional Sanger- and capillary electrophoresis-based approaches, for example with the ability to generate hundreds of thousands of relatively small sequence reads at a time. Some examples of next generation sequencing techniques include, but are not limited to, sequencing by synthesis, sequencing by ligation, and sequencing by hybridization. More specifically, the MISEQ, HISEQ and NEXTSEQ Systems of Illumina and the Personal Genome Machine (PGM), Ion Torrent, and SOLiD Sequencing System of Life Technologies Corp, provide massively parallel sequencing of whole or targeted genomes. The SOLiD System and associated workflows, protocols, chemistries, etc. are described in more detail in PCT Publication No. WO 2006/084132, entitled "Reagents, Methods, and Libraries for Bead-Based Sequencing," international filing date Feb. 1, 2006, U.S. patent application Ser. No. 12/873,190, entitled "Low-Volume Sequencing System and Method of Use," filed on Aug. 31, 2010, and U.S. patent application Ser. No. 12/873,132, entitled "Fast-Indexing Filter Wheel and Method of Use," filed on Aug. 31, 2010, the entirety of each of these applications being incorporated herein by reference thereto.

[0048] The phrase "sequencing run" refers to any step or portion of a sequencing process performed to determine some information relating to at least one biomolecule (e.g., nucleic acid molecule).

[0049] The term "read" with reference to nucleic acid sequencing refers to the sequence of nucleotides determined for a nucleic acid fragment that has been subjected to sequencing, such as, for example, NGS. Reads can be any a sequence of any number of nucleotides which defines the read length.

[0050] The phrase "sequencing coverage" or "sequence coverage," used interchangeably herein, generally refers to the relation between sequence reads and a reference, such as, for example, the whole genome of cells or organisms, one locus in a genome or one nucleotide position in the genome. Coverage can be described in several forms (see, e.g., Sims et al. (2014) *Nature Reviews Genetics 15*:121-132). For example, coverage can refer to how much of the genome is being sequenced at the base pair level and can be calculated as NL/G in which N is the number of reads, L is the average read length, and G is the length, or number of bases, of the genome (the reference). For example, if a reference genome is 1000 Mbp and 100 million reads of an average length of 100

bp are sequenced, the redundancy of coverage would be 10x. Such coverage can be expressed as a "fold" such as 1X, 2X, 3X, etc. (or 1, 2, 3, etc. times coverage). Coverage can also refer to the redundancy of sequencing relative to a reference nucleic acid to describe how often a reference sequence is covered by reads, e.g., the number of times a single base at any given locus is read during sequencing. Thus, there may be some bases which are not covered and have a depth of 0 and some bases that are covered and have a depth of anywhere between, for example, 1 and 50. Redundancy of coverage provides an indication of the reliability of the sequence data and is also referred to as coverage depth. Redundancy of coverage can be described with respect to "raw" reads that have not been aligned to a reference or to aligned (e.g., mapped) reads. Coverage can also be considered in terms of the percentage of a reference (e.g., a genome) covered by reads. For example, if a reference genome is 10 Mbp and the sequence read data maps to 8 Mbp of the reference, the percentage of coverage would be 80%. Sequence coverage can also be described in terms of breadth of coverage which refers to the percentage of bases of a reference that are sequenced a given number of times at a certain depth.

[0051] As used herein, the phrase "low coverage" with respect to nucleic acid sequencing refers to sequencing coverage of less than about 10X, or about 0.001x to about 10x, or about 0.002x to about 0.2x, or about 0.01x to about 0.05x.

[0052] As used herein, the phrase "low depth" with respect to nucleic acid sequencing refers to sequencing depth of less than about 10X, or about 0.1x to about 10x, or about 0.2x to about 5x, or about 0.5x to about 2x.

[0053] The term "resolution" with reference to genomic sequence nucleic acid sequence refers to the quality, or accuracy, and extent of the genomic nucleic acid sequence (e.g., sequence of the entire genome or a particular region or locus of the genome) obtained through nucleic acid sequencing of a cell(s), e.g., an embryo, or organism. The resolution of genomic nucleic acid sequence is primarily determined by the coverage and depth of the sequencing process and involves consideration of the number of unique bases that are read during sequencing and the number of times any one base is read during sequencing. The phrases "low resolution sequence" or "low resolution sequence data" or "sparse sequence data," which are used interchangeably herein, with reference to genomic nucleic acid sequence of a cell(s), e.g., an embryo, or organism, refer to the nucleotide base sequence information of genomic nucleic acid that is obtained through low-coverage and low-depth sequencing methods.

Nucleic Acid Sequence Data Generation

[0054] Some embodiments of the methods and systems provided herein for the analysis of genomic nucleic acids and classification of genomic features include analysis of nucleotide sequences of the genome of cells and/or organisms. In some embodiments, the methods and systems provided herein include analysis of sequences obtained from whole genome sequencing of a cell(s) and/or organism(s). In some embodiments, the methods and systems provided herein include analysis of sequences of the whole genome of a cell(s) and/or organism(s). Nucleic acid sequence data can be obtained using a variety of methods described herein and/or know in the art. In one example, sequences of genomic nucleic acid of cells, for example cells of an embryo, may be obtained from next-generation sequencing (NGS) of DNA samples extracted from the cells. NGS, also known as second-generation sequencing, is based on high-throughput, massively parallel sequencing technologies that involve sequencing of millions of nucleotides generated by nucleic acid amplification of samples of DNA (e.g., extracted from embryos) in parallel (see, e.g., Kulski (2016) "Next-Generation Sequencing – An Overview of the History, Tools and 'Omic' Applications," in Next Generation Sequencing – Advances, Applications and Challenges, J. Kulski ed., London: Intech Open, pages 3-60).

[0055] Nucleic acid samples to be sequenced by NGS are obtained in a variety of ways, depending on the source of the sample. For example, human nucleic acids may readily be obtained via cheek brush swabs to collect cells from which nucleic acids are then extracted. In order to obtain optimum amounts of DNA for sequencing from embryos (for example, for pre-implantation genetic screening), cells (e.g., 5-7 cells) commonly are collected through trophectoderm biopsy during the blastocyst stage. Fetal nucleic acids can be obtained, for example, from products or conception, chorionic villus samples and/or amniocentesis. DNA samples require processing, including, for example, fragmentation, amplification and adapter ligation prior to sequencing via NGS. Manipulations of the nucleic acids in such processing may introduce artifacts (e.g., GC bias associated with polymerase chain reaction (PCR) amplification), into the amplified sequences and limit the size of sequence reads. NGS methods and systems are thus associated with error rates that may differ between systems. Additionally, software used in conjunction with identifying bases in a sequence read (e.g., base-calling) can affect the accuracy of sequence data from NGS sequencing. Such artifacts and limitations can make it difficult to sequence and map long repetitive regions of a genome and identify polymorphic alleles and aneuploidy in genomes. For example, because about 40% of the human genome is comprised of repeat DNA elements, shorter single reads of identical

sequence that align to a repeat element in a reference genome often cannot be accurately mapped to a particular region of the genome. One way to address and possibly reduce some of the effects of errors and/or incompleteness in sequence determination is by increasing sequencing coverage and/or depth. However, increases in sequencing coverage are associated with increased sequencing times and costs. Paired-end sequencing can also be utilized, which increases accuracy in placement of sequence reads, e.g., in long repetitive regions, when mapping sequences to a genome or reference, and increases resolution of structural rearrangements such as gene deletions, insertions and inversions. For example, in some embodiments of methods provided herein, use of data obtained from paired-end NGS of nucleic acids from embryos increased read mapping by an average of 15%. Paired-end sequencing methods are known in the art and/or described herein and involve determining the sequence of a nucleic acid fragment in both directions (i.e., one read from one end of the fragment and a second read from the opposite end of the fragment). Paired-end sequencing also effectively increases sequencing coverage redundancy by doubling the number of reads and particularly increases coverage in difficult genomic regions.

[0056] Three main types of nucleic acid NGS sequencing are commonly performed: whole genome sequencing (WGS), whole exome sequencing (WES) and targeted sequencing (TS). Whole genome sequencing is a comprehensive method in which the entire genome, as opposed to a portion thereof, is sequenced. WES is sequencing of the protein-coding regions (exons) of the genome which accounts for only about less than 2% of the genome. Targeted sequencing is a high-depth coverage type of sequencing in which a limited number of specific regions of the genome are sequenced. Targeted sequencing is typically performed using amplicon-based enrichment, in which specific primers are used to amplify only particular regions of interest (targeted amplification) from which nucleic acid libraries are prepared for sequencing or capture-based methods in which fragmented nucleic acids are hybridized to capture oligonucleotides to isolate regions of interest for sequencing.

Nucleic Acid Sequence Analysis

[0057] In some embodiments of the methods and systems provided herein for the analysis of genomic nucleic acids and classification of genomic features, the sequences of nucleic acids obtained from cells, e.g., embryo cells, or organisms are used to reconstruct the genome (or portions of it) of the cells/organisms using methods of genomic mapping. Typically, genomic mapping involves matching sequences to a reference genome (e.g., a human genome) in a process

referred to as alignment. Examples of human reference genomes that may be used in mapping processes include releases from the Genome Reference Consortium such as GRCh37 (hg19) released in 2009 and GRCh38 (hg38) released in 2013 (see, e.g., https://genome.ucsc.edu/cgi-bin/hgGateway?db=hg19 https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39). Through alignment, sequence reads are assigned to genomic loci typically using computer programs to carry out the matching of sequences. Numerous alignment programs are publicly available and include Bowtie (see, e.g., http://bowtie-bio.sourceforge.net/manual.shtml) and BWA (see, e.g., http://bio-bwa.sourceforge.net/). Sequences that have been processed (for example to remove PCR duplicates and low-quality sequences) and matched to a locus are often referred to as aligned and/or mapped sequences or aligned and/or mapped reads.

[0058] In mapping of sequence reads to a genomic reference, it is possible to detect and/or identify single nucleotide variants (SNV). Single nucleotide variants are the result of variation in the genome at a single nucleotide position. Several different NGS analysis programs for SNV detection (e.g., variant calling software) are publicly available, known in the art and/or described herein (e.g., including but not limited to GATK (see, e.g., https://gatk.broadinstitute.org/) and deepvariant (see, e.g., Poplin et al (2018) Nature Biotech. 36:983-987). Briefly, after alignment, the bcftools software (open source) is used to generate a pileup of all bases identified with a minimum coverage (e.g., 1) and minimum depth (e.g., 1) and generate a genotype call from the bam file generated during alignment. Detection and identification of genomic features, such as chromosomal abnormalities, e.g., aneuploidies, through genome mapping of sequences from sample nucleic acids of cells or organisms presents particular challenges, particularly when sequence data is obtained from low-coverage and/or low-resolution sequencing methods. The major challenges in this approach are derived from the concept that NGS methods are prone to introducing errors into the sequencing read during read generation. With error rates anywhere between 1:100 and 1:10,000, depending on the sequencing platform methodology, identifying the difference between a variant and sequencing error at low-coverage and/or low-depth sequencing provides a unique and difficult informatics challenge. Computer programs and systems are known in the art and/or described herein for increasing the ease and/or accuracy of interpretation of sequence data in identifying certain genomic features. For example, systems and methods for automated detection of chromosomal abnormalities including segmental duplications/deletions, mosaic features, aneuploidy and polyploidy with unbalanced sex chromosomes are described in U.S. Patent Application Publication No. 2020/0111573 which is incorporated in its entirety by reference herein.

Such methods include de-noising/normalization (to de-noise raw sequence reads and normalize genomic sequence information to correct for locus effects) and machine learning and artificial intelligence to interpret (or decode) locus scores into karyograms. For example, after sequencing is completed, the raw sequence data is demultiplexed (attributed to a given sample), reads are aligned to a reference genome such as, e.g., HG19, and the total number of reads in each 1-million base pair bin is counted. This data is normalized based on GC content and depth and tested against a baseline generated from samples of known outcome. Statistical deviations from a copy number of 2 are then reported (if present, if not = euploid) as aneuploidy. Using this method, meiotic aneuploids and mitotic aneuploidy can be distinguished from each other based on the CNV metric. Based on the deviations from normal, a karyotype is generated with the total number of chromosomes present, any aneuploidies present, and the mosaic level (if applicable) of those aneuploidies.

[0059] It should also be noted that both the term SNV and SNP (single-nucleotide polymorphism) are used in accordance with various embodiments. Though both terms may be distinguishable to those of ordinary skill in the art (SNPs being well characterized SNVs), the terms can be used interchangeably in accordance with various embodiments herein. Thus, the use of either term should be inclusive of both terms as it applies to the process for analyzing received sequencing data.

[0060] Artifacts, variations in coverage and errors that can occur in NGS also present challenges in use of sequence data to accurately classify particular genomic features, such as in assessing the pattern of inheritance of a genomic feature and determining, predicting/inferring parental origin of a genomic feature, or determining genetic relationships of a conceptus to a sperm provider an oocyte provider For example, challenges of using low-coverage and/or low-depth (e.g., low-resolution) sequence data that can arise in analysis of the pattern of inheritance of a genomic feature, e.g., aneuploidy, include obtaining an extremely low number-to-no high-quality, high-depth SNPs that would meet the standard American College of Medical Genetics requirements (see, e.g., Richards et al. (2015) Genetics in Medicine 17:405-423) for accurate calling leading to a lack of overlap in the sequences of offspring (e.g., an embryo) and parent(s) (e.g, mother and/or father) and preventing accurate assessment of phasing/haplotype determination. Increasing sequence coverage, although with its associated decreases in efficiency and increases in cost, may improve NGS sequence data such that it can be used in determining patterns of inheritance of a genomic feature or the genetic relationships to a conceptus. Other methods for meeting the challenges associated with, for example, assessing pattern of inheritance to determine parental origin of a

genomic feature, e.g., aneuploidy, include use of microarrays and fluorescence in situ hybridization (FISH) which are not high-throughput technologies as is NGS and are also more costly and require longer times to carry out analyses.

[0061] Provided herein are improved, efficient, rapid, and cost-effective methods and systems for detecting and/or identifying one or more, or a plurality, of genomic features of a cell(s), e.g., cells of an embryo, or an organism(s) and/or assessing, classifying or determining a pattern of inheritance of one or more, or a plurality, of genomic features to determine, predict and/or infer the parental origin or lineage of the genomic feature as being maternal or paternal, or determining genetic relationships of a conceptus to a sperm provider an oocyte provider. In some embodiments of methods provided herein, relatively low-coverage and/or low-depth (e.g., low-resolution) sequence data are used to identify genomic features, such as, for example, chromosomal abnormalities, of cells, e.g., cells of an embryo, or organism and assess the pattern of inheritance of the genomic feature to determine, predict and/or infer the parental origin of the genomic feature as being maternal or paternal, or determine genetic relationships of a conceptus to a sperm provider an oocyte provider. In various embodiments, the methods, and systems incorporating the methods, use nucleic acid sequence data (e.g., low-resolution sequence data) obtained from low-coverage and/or low-depth whole genome sequencing of nucleic acids samples of the total genomic nucleic acids or total DNA (or total nuclear DNA) of a cell(s) as opposed to being obtained from sequencing of only a portion or pre-determined specific targeted regions of a genome. In various embodiments, for example, the methods and systems can be used to determine genetic relationships of a conceptus to a sperm provider an oocyte provider, or classify the pattern of inheritance of genomic features in embryos, including, for example, embryos generated through IVF, prior to implantation. In some embodiments, methods and systems provided herein can be used to determine genetic relationships of a conceptus to a sperm provider an oocyte provider, or classify the pattern of inheritance of a chromosomal abnormality, such as aneuploidy, polyploidy and/or chromosomal segmental gains and/or losses (e.g., segmental gains and/or losses of about 20 megabases or greater in size), in an embryo, such as an IVF embryo, prior to implantation.

[0062] FIG. 1 is a diagrammatic representation of the workflow 100 of an exemplary method for detecting and/or identifying a genomic feature and/or assessing, classifying, determining, predicting and/or inferring patterns of inheritance and/or parental origin of one or more genomic features of genomic nucleic acids of a cell(s), e.g., an embryo, or an organism, and to determine genetic relationships of a conceptus to a sperm provider an oocyte provider. In some embodiments, the

method is used to detect or identify a chromosomal abnormality (e.g., polyploidy, such as balanced sex-chromosome polyploidy) and/or assess, classify, determine, predict and/or infer patterns of inheritance and/or parental origin of one or more chromosomal abnormalities. In some embodiments, the chromosomal abnormality is aneuploidy, such as, e.g., meiotic aneuploidy, and/or polyploidy (e.g., balanced-sex chromosome polyploidy.

[0063] As shown in steps 101 and 102 of FIG. 1, sequence reads (data) from sequencing of nucleic acids of an offspring, e.g., embryo, and the assumed maternal and paternal contributors (mother and father, or parents) are received and aligned to a reference (e.g., human) genome, using any suitable alignment software program with the proper settings, for mapping. In some embodiments of the method 100, the sequence reads are obtained from low-coverage and/or low-depth (e.g., low-resolution) sequencing of offspring and parent genomic nucleic acids, e.g., low-coverage and/or low-depth (e.g., resolution) whole genome sequencing of the nucleic acids. In step 103, the sequence data are analyzed a to detect and/or identify SNVs and chromosomal abnormalities, such as chromosomal copy number variation (CNV) and structural copy number variations (e.g., aneuploidy, unbalanced-sex polyploidy and segmental chromosome gains and losses), of the offspring and to detect and/or identify SNVs of the parents. The analysis in step 103 for detection/identification of chromosomal abnormalities is referred to as “CNV” analysis herein. A karyogram containing the final copy number analysis result is generated for the offspring and it, and the SNV data files for the offspring and parents are received in step 104A and 104B. In some embodiments, the method begins at step 104A and 104B with receiving of a karyogram and SNV data for the offspring and SNV data for the parents which can be based on nucleic acid sequence information obtained through low-coverage and/or low-depth (e.g., low-resolution) sequencing of offspring and parent genomic nucleic acids, e.g., low-coverage and/or low-depth (e.g., resolution) whole genome sequencing of the nucleic acids [. In some instances, the sequence reads used in generating the karyogram and SNV data are obtained through low-coverage and/or low-depth (e.g., low-resolution) paired-end sequencing methods which can provide a magnitude of order increase in data. Methods for analysis and interpretation of mapped reads are known in the art and/or described herein. For example, in some embodiments, the karyogram and SNV data are generated in a process that includes analysis of mapped reads conducted as described in U.S. Patent Application Publication No. 2020/011573. For example, using artificial intelligence (AI)/machine learning (ML), copy number variations in a de-noised sample genomic sequence dataset may be

identified when a frequency of genomic sequence reads aligned to a chromosomal position deviates from a frequency threshold. The karyotype of the embryo is then determined from this analysis.

[0064] The single nucleotide variants (SNVs) identified in the nucleic acids from the offspring and both gamete contributors (genetic parents) in this exemplary method are used in predicting or inferring alleles and/or haplotypes in the embryo and parents. If more than 1% of a population does not carry the same nucleotide at a specific position in the genome, the SNV is often referred to as a single nucleotide polymorphism (SNP). An SNV is typically a more generic term for less well-characterized loci. There are about 10 million or more SNPs located throughout the human genome, on average every 200 bp. Although some SNPs may be associated with traits or disorders, most have no known function. No two individuals (except identical twins) have the same pattern of SNPs which exist as major and minor isoforms within a given population. In one embodiment of the methods and systems provided herein for detecting and/or identifying a genomic feature (e.g., polyploidy, such as balanced-sex chromosome polyploidy) and/or assessing, classifying, determining, predicting and/or inferring patterns of inheritance and/or parental origin of one or more genomic features (e.g., aneuploidy and/or polyploidy), genotype imputation (step 105 of FIG. 1) is used to construct and infer predicted alleles and/or haplotypes for all chromosomes of the genome of an offspring (e.g., embryo) and its respective gamete contributors or genetic parents based on analysis of the SNV identified in sequences of the genomes.

[0065] Low-coverage and/or low-depth (e.g., low resolution) sequencing yields sparse data with missing datapoints and thus provides a probabilistic representation of genotypes (genotype likelihoods). Genotype imputation methods are used to refine genotype likelihoods and fill in gaps due to sparsely mapped sequence reads resulting from low-coverage and/or low-depth (e.g., low resolution) sequencing of offspring and parent genomic nucleic acids where the amount of overlap of sequence information between the embryo and parents may be limited. Imputation is a statistical inference of missing genotypes and haplotypic phase whereby alleles or haplotypes can be inferred for all chromosomes of the genome of an offspring (e.g., embryo) and one or both parents. Genotype imputation involves phasing genotypes at genotyped SNV positions from the sequence data and a reference panel of haplotypes of fully phased individuals with completely determined alleles, followed by matching haplotypes which match in the genotyped positions. Publicly available reference panels of known haplotypes (e.g., human haplotypes) include the Haplotype Reference Consortium dataset (see, e.g., https://www.ebi.ac.uk/ega/studies/EGAS00001001710), which contains phased SNV genotypes

coded in the human genome assembly GRCh37, and the 1000 Genomes Project (see, e.g., https://www.internationalgenome.org/). Methods for genotype imputation are known in the art and include, for example, Beagle (see, e.g., Browning et al. (2018) Am J Hum Genet 103(3):338-348, Browning and Browning (2007) Am J Hum Genet 81:1084-1097 and http://faculty.washington.edu/browning/beagle/beagle.html#download), MACH (see, e.g., Li et al. (2009) Ann Rev Genom Hum Genet 10:387-406, Li et al. (2010) Genet Epidemiol 34:816-834 and http://csg.sph.umich.edu/abecasis/MACH/tour/imputation.html).

[0066] In various embodiments of the method of FIG. 1, imputing of missing variant data and phasing of chromosomal haplotypes is performed using Beagle version 5.0. The Beagle program does not make any assumptions about relatedness between genomic samples. Although this may lead to a relative reduction in accuracy of recovery of variants and proper phasing for genomic samples with known relationships compared to other methods, one advantage is that in the event of unrelated samples (e.g. incorrectly labeled parent/offspring sample), Beagle will not attempt to enforce a relationship on its analysis. Additionally, Beagle version 5 is relatively quick compared to some other programs. Imputation and phasing performance of imputation may be limited by the availability of relevant pre-phased haplotype data in the database. In some embodiments of the method of FIG. 1, the imputation process is performed using the Phase 3 1000 Genomes haplotype reference database containing haplotypes from 2,504 subjects. Augmentation of reference databases with additional sequence variant data may improve haplotype inference for a specific target population. The incorporation of imputation in some embodiments of methods and systems provided herein yields a fully phased and complete set of alleles and/or haplotypes predicted for one or both parents and an offspring, e.g., embryo. This information is then used in determining ratios of maternal and/or paternal contribution of alleles to the embryo which are used in detecting and/or identifying a genomic feature and/or assessing, inferring or determining patterns of inheritance (POI) of genomic features, e.g., aneuploidy and polyploidy.

[0067] The fully phased and complete set of inferred alleles and/or haplotypes of variants (SNV-containing sequences) determined in step 105 of the method of FIG. 1 is used to classify genetic abnormalities of an offspring (i.e., abnormalities that were previously identified in step 104A) as either being maternal or paternal in origin in step 106A or step 107 of the method depicted in FIG. 1. As shown in FIG. 1, if the karyogram for an offspring (e.g., embryo) indicates that aneuploidy is present in the embryo, then the method proceeds to step 106A for analysis and determination of pattern of inheritance. If the karyogram for an offspring indicates that aneuploidy and/or

unbalanced-sex polyploidy is not present in the offspring, the method proceeds to step 106B to first determine if the offspring exhibits a genome-wide dosage imbalance and is polyploid and then to step 107 for a determination of the pattern of inheritance of the polyploidy. A limitation of Beagle version 5.0 imputation program is that it assumes that all chromosomes analyzed are diploid and returns imputed data accordingly. As such haploid data is generally coded as homozygous diploid. Triploid (and more complex) haplotypes are similarly imputed as diploid. In such cases, the pattern of inheritance analysis described herein is not based on a presumed correct set of inferred haplotypes but on the frequencies of SNV-containing variants identified in the low-coverage and/or low-depth (e.g., low resolution) sequencing data and their dose relationship to parents, as described herein. Specifically, phased and imputed data are treated as pseudo chromosomes and, if truly aneuploid, are expected to be identical and to match either maternal or paternal genotypes, not both, at least for informative variants. The workflow of an exemplary method 200 used in one embodiment of the pattern of inheritance (POI) determination step 106A or step 107 is depicted in FIG. 2.

[0068] In the classification of an embryo abnormality (e.g., aneuploidy and/or polyploidy) as being of maternal or paternal origin in step 106A or step 107 of the embodiment of the method of FIG. 1, the measure of relatedness of offspring, e.g., embryo, genomic DNA to both the mother and father is calculated. In one embodiment, the calculation involves the counting of the total variants identified in each defined genomic bin (defined region of the genome, e.g., 1 million bases) in the offspring, e.g., embryo, shared with the mother or with the father, and converting the counts into a single relatedness value per genomic region of a user-defined size. Prior to counting, the input allele/haplotype variant sequence data are filtered (see step 201 of FIG. 2). Input data is expected to be imperfect given that low pass sequence data, especially from certain sequencers, generally includes a large number of sequencing artifacts. Therefore, the first step in the counting process is to filter detected variants. Specifically, in some embodiments, the following sequences are identified and excluded from counting: (1) sequences of sites with any missing alleles between the offspring, mother and father (trio) (inability to infer relatedness in such cases), (2) sequences of sites with constant alleles across a trio (non-informativeness), (3) sequences of sites with a novel allele within one individual in the trio (non-informativeness and likelihood of sequencing artifact), and (4) sequences of sites that are consistent with euploid inheritance only (non-informative for aneuploid relationship estimation).

[0069] Filtering can also include excluding SNPs that are inconsistent with Mendelian inheritance. In this embodiment of the method of FIG. 2, the input variants are filtered indirectly (i.e., by comparing relationships between offspring and parents). This method presumes correct trios and does not necessarily filter the correct sites in the event of an incorrectly specified parent.

[0070] After the variants are filtered (generally down to about 10% of the input observed phased/imputed SNV-containing variants), the remaining variants are distributed into bins of a user-defined size by genomic regions within individual chromosomes (e.g., 1 million bases) (step 202 of FIG. 2).

[0071] In step 203 of FIG. 2, variants within each defined bin are counted and two values are generated for each bin. The first value is the number of variants phased together between the offspring and maternal component and the second value is the number of variants phased together between the offspring and the paternal component. To determine inheritance for individual chromosome aneuploidy only the bins known to be aneuploidy (based on the bins associated with the aneuploid CNV call) are assessed for that particular aneuploidy. These include a set of counts not reported in the results file such as: (a) OvM (i.e. "offspring variants shared with maternal source") which is a count of alternate (i.e., non-reference) alleles shared between the embryo and the mother, and (b) OvP (i.e. "offspring variants shared with paternal source"), which is a count of alternate alleles shared between the embryo and the paternal sample. Reference alleles match the reference genome and generally it is expected that the majority of alleles match the reference genome. An alternate allele is one that for any given locus there is a variant that does not match the reference genome. Because a great majority of the bases/allele match the reference genome for the mother, father and offspring, the reference alleles are less informative. Only alternate alleles are used because reference alleles are generally non-informative.

[0072] The output of the counting functions of step 203 is a set of OvM and OvP counts, per bin, across all chromosomes in the specified analysis (only the specific chromosome(s) of the aneuploidy for embryos having an aneuploid karyotype or all chromosomes for an embryo having a karyotype that is not indicative of aneuploidy but correlates with polyploidy). In step 204 of FIG. 2, these counts are converted into a single relatedness value, referred to as a dose measure, per bin. Two statistics are calculated within each bin: (1) L2RAT which is $log_2\left(\frac{OvM+s}{OvP+s}\right)$, where OvM and OvP are the counts, and *s* is the SMOOTH controllable element (provided either in a configuration file or via command line argument). The smoothing parameter *s* is used to prevent infinite values in the case

of bins with zero counts. Values between 1 and 30 are reasonable (larger values shrink the estimate), and (2) DIFF which is defined as OvM – OvP. In some embodiments L2RAT is the default dose measure and is more robust to outliers and can be shrunk as desired. Both estimates are centered at zero (when the offspring shows no stronger relationship to mother or father, and both take positive values when the relationship between offspring is stronger with the mother than the father, and negative values otherwise. In this embodiment, there is an assumption that the dose statistics should be centered at zero for euploid inheritance and that departures from zero beyond that expected by chance indicate stronger relationship (i.e. a greater number of inherited variants) from either the mother or father. This assumption is false in cases where the embryo sample is contaminated with maternal material. In such cases, the dose statistics will be shifted in the positive direction for all chromosomes and therefore may falsely appear polyploid.

[0073] In step 205 of FIG. 2, a one sample t-statistic is calculated on the specified dose measure, e.g., L2RAT. The data set for the calculated test statistic is the set of estimated dose values for all bins in the region of interest. For example, if a 10 megabase region was specified and bin size was 1 megabase, then the t-test would be based on 10 L2RAT estimates, one from each bin. In step 205, a test statistic is generated for every provided region of interest (e.g. genome overall, all specified chromosomes, and any specified segmental regions). The number of bins, the mean and standard deviation of the dose statistic, and the p-value and a custom confidence metric are also provided in this step. Choice of both bin size and minimum analyzed region size are user controllable. Because inferred haplotypes are large, small bin sizes are not optimal as neighboring regions are correlated and subsequently the counts in small adjacent bins are likely to be correlated as well. This could lead to t-statistics with a larger variance than desired (i.e. they could produce larger t-test values, both positive and negative, than warranted).

[0074] In step 206 of FIG. 2, each genetic abnormality, e.g., aneuploidy, is assigned a parental origin using the t-statistic and user-controllable classification/detection thresholds. The parental origin determination is based on the number of parental-offspring allele matches for all chromosomes (genome level association, e.g., in the case of polyploid offspring) and at the individual chromosome level (in the case of aneuploid offspring). In the case where a sample is not polyploid, known chromosomal aneuploidies are classified for parental origin based on the full chromosome test statistic. Thus, at the meiotic aneuploidy level (both whole and partial chromosomes) only the alleles present in that chromosome correlated with the aneuploidy are considered when making the comparison of likeness. The number of bins in the test will depend on both the bin size and the

chromosome length (the analysis will have greater statistical power to classify aneuploidies for longer chromosomes). The output for such a test will either be: "maternal," "paternal," or "not classified." An output of "not classified" is returned when the test statistic fails to exceed a user controllable threshold (CLASS_THRESHOLD_ANEUPLOID). Although these categories are fixed, the specific messages returned are user controllable. In the case where an embryo sample is not polyploid, known segmental aneuploidies (that meet a user controllable size requirement) are classified based on the regional test statistic. The number of bins in the test will depend on both the bin size and the region size (there is greater statistical power to classify aneuploidies for longer regions). The output for such a test will either be: "maternal," "paternal," or "not classified." An output of "not classified" is returned when the test statistic fails to exceed the user controllable threshold (CLASS_THRESHOLD_ANEUPLOID). Although these categories are fixed, the specific messages returned are user controllable.

[0075] Returning to FIG. 1, in step 106B, the maternal and paternal contributions to the entire genome (i.e., all chromosomes) of the offspring, e.g., embryo, are estimated based on the number of shared variant alleles and are used in detecting any previously undetected polyploidy or to confirm polyploidy detected and identified in the karyograms for the offspring. Step 106B is directly conducted after step 105 for any offspring that were not identified as aneuploid through the earlier karyotyping process and that either do, or do not, show evidence of polyploidy in the CNV data, with the exception of diploid male embryos. If the offspring was identified as a diploid male (46,XY) it is not assessed for polyploidy or pattern of inheritance. Additionally, all non-male aneuploid offspring evaluated for pattern of inheritance in step 106A of FIG. 1 undergo analysis in step 106B of the maternal and paternal contributions to the entire genome (i.e., all chromosomes) of the offspring to determine if there is any undetected polyploidy. This step is performed only where the offspring sample is not previously reported as "polyploid" and where the sample is also not male. The output for such a test will either be: "maternal," "paternal," or "not detected." An output of "not detected" is returned when the test statistic fails to exceed a user controllable threshold (DETECT_THRESHOLD_POLYPLOIDY). Although these categories are fixed, the specific messages returned are user controllable. A higher threshold for polyploidy detection is set than for classification of parental origin of polyploidy, even for full genome analysis. In the case of a known (as identified in the initial karyograms) non-euploid offspring sample, parental origin of polyploidy is classified based on the full genome test statistic in step 107 of FIG. 1. The number of bins in the test will depend on the bin size. The output for such a test will either be: "maternal," "paternal," or

"not classified." An output of "not classified" is returned when the test statistic fails to exceed the user controllable threshold (CLASS_THRESHOLD_POLYPLOID). Although these categories are fixed, the specific messages returned are user controllable.

[0076] FIG. 3 is a decision tree diagram depicting a logical flow for conducting some embodiments of the methods for detecting and/or classifying the parental origin of genetic abnormalities provided herein. This diagram depicts, for some embodiments of the method, the process of determining the final result for any specimen being tested by the embodiment and is referred to as the Parent of Origin (POO) pipeline. Testing for an offspring, e.g., embryo, nucleic acid sample proceeds down the decision path as follows. If a sample is a non-polyploid female (XX), it is assessed by the method to determine if the genome dosage significantly deviates from expected, and, if so, it is characterized as polyploid (maternal or paternal) and reported. If a sample is not a non-polyploid XX and is called polyploid (XXY or XYY), then it enters the pipeline for characterization of the pattern of inheritance (POI) for polyploidy and is then reported. If a sample is neither a non-polyploid XX or a polyploid XXY or XYY and is called haploid by CNV analysis, it enters the pipeline for assessment of genome-wide dosage imbalance and is reported. All other samples that do not meet the criteria listed previously are assessed for full chromosome aneuploidies and segmental chromosome aneuploidies. If these are mosaic aneuploidies as determined by CNV analysis, they are not interrogated by the method. If they are whole copy changes (meiotic aneuploidy), the method attempts to classify the regions of interest as maternal, paternal or not classified and the result is reported as such.

[0077] Turning to step 108 of FIG. 1, in accordance with various embodiments, the methods discussed above can be extended to address and overcome the issues of consanguinity and sample misidentification to determine the genetic relationship of a conceptus with a sperm provider and oocyte provider by performing a relatedness analysis. In general, the issue of sample misidentification can be addressed by some general steps. These steps include, for example, 1) imputing full sample genomes using spare sequencing data (discussed in some detail above), 2) calculating relationship metrics between samples and presumed parents (discussed in some detail above), and 3) generating maternal and paternal relationship scores (discussed in some detail above and in more detail below), and 4) comparing scores to thresholds to classify results as either consistent with stated parentage or potentially inconsistent.

[0078] Regarding (1), given a set of sparse sequence data files (particularly when sequenced at low-coverage/depth), a pre-filter step be done by comparing the acquired sequencing data against a reference set of known variants in the human genome (e.g., from the 1000 genome project). This filter can remove variants that may be due to sequencing artifacts. Filtered data files can then undergo imputation. Again, given that low pass sequence data generates numerous sequencing artifacts and imperfections, these are originally included in the imputed data. Therefore, a further filter of detected variants can be performed prior to counting. This second filter can filter, for example, sites with any missing alleles across a trio (inability to infer relatedness in such cases) and sites with constant alleles across a trio (non-informativeness), while retaining sites with a novel allele in the embryo to measure parental consistency (i.e. to confirm relatedness of samples).

[0079] After imputation, and pre-count filtering, variants are counted to measure inheritance consistency and novelty. Specifically, and for example, all filtered variants can be counted, novel variants can be counted if either allele of an embryo (or conceptus) genotype is not inheritable from either parent (e.g., presumed parent), either via euploid or aneuploid transmission, and consistent variants (consistent with stated inheritance) can be counted if at least one allele not shared with the mother is shared with the father, or at least one allele not shared with the father is shared with the mother. Maternally consistent variants can be counted if at least one allele in the offspring is shared with the mother. Paternally consistent variants can be counted if at least one allele in the offspring is shared with the father. Novelty then can be calculated as a ratio of novel variants to total (filtered) variants.

[0080] In accordance with various embodiments, and in accordance with step 108, parental consistency scores can be determined by calculating an overall (parental) score using the count results described above for maternal and paternal consistency. Comparing maternal and paternal consistency can be obtained by taking a log (base 2) of the ratio of individual parent scores.

[0081] Metrics for parental consistency can be calculated while also incorporating novelty maternal and paternal scores using the following formula

$$Score_i = (N_{c_i} \times N_n)/(N_{SNP} \times N_C)$$

where *i* is the parent (either presumed mother or father), N_{c_i} is the number of SNPs consistent between offspring and parent *i*, N_n is the number of novel SNPs (i.e. SNPs in embryo not seen in either parent), N_{SNP} is the number of SNPs observed in the trio, and N_C is the number of SNPs consistent with either the designated maternal or paternal samples.

[0082] As will be discussed and illustrated in the Examples below, these calculations can be used to defined thresholds to compare scores against to determine the generic relationships between presumed parents and the conceptus under consideration.

[0083] In various embodiments, methods are provided for assessing, classifying, determining, predicting and/or inferring the genetic relationship of a conceptus with a sperm provider and oocyte provider. The methods can be implemented via computer software or hardware. The methods can also be implemented on a computing device/system that can include a combination of engines for assessing, classifying, determining, predicting and/or inferring the genetic relationship of a conceptus with a sperm provider and oocyte provider. In various embodiments, the computing device/system can be communicatively connected to one or more of a data source, sample analyzer, and display device via a direct connection or through an internet connection.

[0084] FIG. 8 is a schematic diagram of a system 800 for the genetic relationship of a conceptus with a sperm provider and oocyte provider, in accordance with various embodiments. System 800 can include a data store 810, a computing device 830 and a display 880. System 800 can also include a sample analyzer 890.

[0085] The sample analyzer 890 can be communicatively connected to the data store 810 by way of a serial bus (if both form an integrated instrument platform 812) or by way of a network connection (if both are distributed/separate devices). The sample analyzer 890 can be configured to analyze samples from a conceptus 820, a oocyte 822 and sperm 824. Sample analyzer can be a sequencing instrument, such as a next generation sequencing instrument, configured to sequence samples to collect sequencing data for further analysis. In various embodiments, the sequencing data can then be stored in the data store 810 for subsequent processing. In various embodiments, the sequencing datasets can be fed to the computing device 830 in real-time. In various embodiments, the sequencing datasets can also be stored in the data store 810 prior to processing. In various embodiments, the sequencing datasets can also be fed to the computing device 830 in real-time.

[0086] The data store 810 can be communicatively connected to the computing device 830. In various embodiments, the computing device 830 can be communicatively connected to the data store 810 via a network connection that can be either a "hardwired" physical network connection (e.g., Internet, LAN, WAN, VPN, etc.) or a wireless network connection (e.g., Wi-Fi, WLAN, etc.). In various embodiments, the computing device 830 can be a workstation, mainframe computer, distributed computing node (part of a "cloud computing" or distributed networking system), personal computer, mobile device, etc.

[0087] Data store 810 can be configured to receive conceptus, sperm provider, and oocyte provider sequence data. In various embodiments, the conceptus is a preimplantation conceptus. In various embodiments, at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing. The low-coverage sequencing can be between about 0.001 and 10x. The low-coverage sequencing can be between about 0.01 and 0.5x. The low-coverage sequencing can be between about 0.25 and 0.2x.

[0088] Computing device 830 can further include an alignment engine 840, a single nucleotide polymorphism identification engine (SNP identification engine) 850, an imputation engine 860 and a relatedness engine 870. As stated above, computing device 830 can be communicatively connected to data store 810.

[0089] Alignment engine 840 can be configured to align the received sequence data to a reference genome. Engine 840 can also be configured to identify a region of interest in the aligned conceptus sequence data, and also identify SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data. The region of interest can be genome wide. The region of interest can be an observed copy number variation.

[0090] SNP identification engine 850 can be configured to identify single nucleotide polymorphisms (SNPs) in the sperm provider sequence data, oocyte provider sequence data, and the conceptus sequence data.

[0091] Imputation engine 860 can be configured to impute missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference. The imputation reference can include at least 1000 genomes.

[0092] Imputation engine 860 can be further configured to filter at least one of the conceptus, sperm provider, and oocyte provider sequencing data to remove sequencing artifacts. The filtering

can include excluding SNPs that are not included in a reference list of known SNPs. The reference list can include at least 1000 genomes. The filtering can comprise excluding sequences of sites with any missing alleles between conceptus, sperm and oocyte. The filtering can comprise excluding sequences of sites with constant alleles across between conceptus, sperm and oocyte. The filtering can comprise excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte. Filtering can also include excluding SNPs that are inconsistent with Mendelian inheritance.

[0093] Relatedness engine 870 can be configured to calculate a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider. Relatedness engine 870 can be configured to calculate a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider. Relatedness engine 870 can be configured to classify the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

[0094] Alignment engine can be further configured to identify a region of interest in the aligned conceptus sequence data, and the relatedness engine 870 can be configured to count the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a corresponding region on the oocyte provider sequence data to determine a maternal contribution value, count the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value, and classify a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm. In various embodiments, the region of interest is the entire genome, and the relatedness engine 870 can be configured to count SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid. When the conceptus is polyploid, the relatedness engine 870 can be configured to classify a pattern of inheritance for the polyploid as maternal or paternal based on the relative contribution values between oocyte and sperm.

[0095] After genetic relationship of a conceptus with a sperm provider and oocyte provider has been determined, it can be displayed as a result or summary on a display or client terminal 880 that is communicatively connected to the computing device 830. In various embodiments, display 880 can be a thin client computing device. In various embodiments, display 880 can be a personal computing device having a web browser (e.g., INTERNET EXPLORER™, FIREFOX™, SAFARI™, etc.) that can be used to control the operation of the region of interest engine (ROI engine) 840, the single nucleotide polymorphism identification engine (SNP identification engine) 850, the imputation engine 860, and the pattern of inheritance engine (POI engine) 870.

[0096] It should be appreciated that the various engines can be combined or collapsed into a single engine, component or module, depending on the requirements of the particular application or system architecture. In various embodiments the region of interest engine (ROI engine) 840, the single nucleotide polymorphism identification engine (SNP identification engine) 850, the imputation engine 860, and the pattern of inheritance engine (POI engine) 870 can comprise additional engines or components as needed by the particular application or system architecture.

[0097] FIG. 9 is an exemplary flowchart showing a method 900 for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, in accordance with various embodiments.

[0098] In step 910, conceptus, sperm provider, and oocyte provider sequence data is received. The conceptus can be a preimplantation conceptus. In various embodiments, at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing. The low-coverage sequencing can be between about 0.001 and 10x. The low-coverage sequencing can be between about 0.01 and 0.5x. The low-coverage sequencing can be between about 0.25 and 0.2x.

[0099] In step 920, the received sequence data is aligned to a reference genome.

[00100] In step 930, single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data is identified.

[00101] In step 940, missing gaps in the sperm provider sequence data and the oocyte provider sequence data are imputed using an imputation reference. The imputation reference can comprise at least 1000 genomes.

[00102] In various embodiments, the method can also include filtering at least one of the sperm provider, oocyte provider, and conceptus sequence data to remove sequencing artifacts. The filtering can include excluding SNPs that are not included in a reference list of known SNPs. The reference list can include about 1000 known genomes. The filtering can comprise excluding sequences of sites with any missing alleles between the conceptus, sperm and oocyte. The filtering can comprise excluding sequences of sites with constant alleles across between the conceptus, sperm and oocyte. The filtering can comprise excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte. Filtering can also include excluding SNPs that are inconsistent with Mendelian inheritance.

[00103] In step 950, a paternal consistency score between the sperm provider and the conceptus is calculated, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider.

[00104] In step 960, a maternal consistency score between the oocyte provider and conceptus is calculated, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider.

[00105] In step 970, the sperm provider and/or the oocyte provider are classified as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

[00106] In various embodiments, the method can further comprise identifying a region of interest in the aligned conceptus sequence data, and identifying SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data. The region of interest can be genome wide. The region of interest can be a copy number variation.

[00107] In various embodiments, the method can further comprise identifying a region of interest in the aligned conceptus sequence data. The method can comprise counting the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a corresponding region on the oocyte provider sequence data to determine a maternal contribution value. The method can comprise counting the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for

the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value. The method can comprise classifying a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm.

[00108] In various embodiments, when the region of interest is the entire genome, the method can further include counting SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid. When the conceptus is polyploid, a pattern of inheritance for the polyploid can be classified as maternal or paternal based on the relative contribution values between oocyte and sperm.

EXAMPLES

Example 1 – Classification of Parental Origin of Monosomy

[00109] Nucleic acids extracted from a human embryo having a known karyotype 42;XY;-14;-15;-19;-21 and nucleic acid samples from both parents were sequenced using a NextSeq sequencing system (Illumina) at 0.1X coverage. Sequence reads were aligned using the Bowtie2 alignment program and mapped to a human reference genome (HG19). The total number of reads in each 1-million base pair bin were counted. The data were normalized based on GC content and depth and tested against a baseline generated from samples of known outcome. Statistical deviations from a copy number of 2 were reported (if present, if not = euploid) as aneuploidy. A karyotype of 42;XY;-14;-15;-19;-21 was determined. SNVs in the sequencing data were identified as defined by the method described here within. Imputing of missing variant data and phasing of chromosomal haplotypes was performed using the SNV data from the embryo and each parent and the Beagle version 5.0 imputation program using the Phase 3 1000 Genomes haplotype database as a reference panel. The measure of relatedness of embryo genomic DNA to both the mother and father was calculated as described herein by counting variants in the embryo shared with the mother and with the father and converting the counts into a single relatedness value per genomic region of a user-defined size.

[00110] The results of the analysis are depicted in the bar graph shown in FIG. 4. For each of the 22 autosomal chromosomes and the X chromosome (x-axis of graph is chromosome number),

FIG. 4 shows the ratio of the variant alleles in that chromosome that matched the mother (blue bars) and the ratio of variant alleles that matched the father (orange bars) (the y-axis is the ratio of embryo variant alleles that matched with either given parent). A comparison of the ratio of matching variant alleles shows that on average for most of the chromosomes (other than the X chromosome of this XY male embryo), the ratio of matching alleles for the mother and father are fairly similar. However, for chromosomes 14, 15, 19, and 21, the ratio of matching alleles for the father is higher than the ratio of matching alleles for the mother. Furthermore, the overall proportion of counts (the ratio of matching alleles to total variant alleles) of the number of informative alleles matching the father is higher than those matching the mother. These results indicate that the four chromosomal deletions are maternal in origin and are indicative of a maternal pattern of inheritance of monosomy.

Example 2 – Classification of Parental Origin of Trisomy

[00111] Nucleic acids extracted from a human embryo having a known karyotype 47;XX;+16 and nucleic acid samples from both parents were sequenced and analyzed as described in Example 1. FIG. 5A shows a graph of the number of embryo variant alleles shared with paternal source (OvP) vs. the number of embryo variant alleles shared with maternal source (OvM) for each of the 23 chromosomes (blue dots). The dotted diagonal line represents the points on the graph at which the number of embryo variant alleles shared with the maternal source would be equal to the number of embryo variant alleles shared with the paternal source for each chromosome. Dots located above the diagonal line represent chromosomes for which there were more variant alleles shared between the embryo and father than between the embryo and mother. Dots located below the diagonal line represent chromosomes for which there were more variant alleles shared between the embryo and mother than between the embryo and father. As shown in the graph in FIG. 5A, there are more dots below the diagonal line. Furthermore, the dot farthest from the diagonal, which represents counts for shared alleles for chromosome 16, has the greatest ratio (almost 2:1) of counts of shared alleles with mother to shared alleles with father. These results indicate that the additional chromosome 16 in the embryo is of maternal origin and are indicative of a maternal pattern of inheritance of trisomy. FIG. 5B is another graphic presentation of the results shown in FIG. 5A showing the ratio of counts of shared alleles with mother to counts of shared alleles with father per chromosome.

Example 3 – Classification of Parental Origin of a Segmental Deletion of a Chromosome

[00112] Nucleic acids extracted from a human embryo having a known karyotype 46;XY;del(6)(q25.1-qter) ;mos33.0% del(6)(pter-q25.1) and nucleic acid samples from both parents were sequenced and analyzed as described in Example 1. The size of the deletion at the p arm of chromosome 6 in the embryo is approximately 20 million bases. The number of variant alleles analyzed for chromosome 6 of the embryo that matched chromosome 6 variant alleles of the mother and of the father were counted in this example. FIG. 6 shows the ratio (y-axis) of the number of embryo variant alleles shared with maternal source (OvM) to the number of embryo variant alleles shared with paternal source (OvP) counts with respect to location on chromosome 6 (x-axis in base pairs). The position of the q arm of the chromosome begins at about 62,000,000 bp and the position of q25.1-ter is located at about 150,000,000 bp. As shown in the graph, the ratio of OvM/OvP from the position at 150,000,000 bp and higher is almost entirely greater than 1.0. These results indicate that the segmental deletion of chromosome 6 is paternal in origin and are indicative of a paternal pattern of inheritance of the deletion.

Example 4 – Classification of Parental Origin of Polyploidy

[00113] Nucleic acids extracted from a human embryo having a known karyotype 68;XXY;mos28.2% -19 and nucleic acid samples from both parents were sequenced and analyzed as described in Example 1. FIG. 7 shows a graph of the number of embryo variant alleles shared with paternal source (OvP) vs. the number of embryo variant alleles shared with maternal source (OvM) for each of the 23 chromosomes (blue dots). The dotted diagonal line represents the points on the graph at which the number of embryo variant alleles shared with the maternal source would be equal to the number of embryo variant alleles shared with the paternal source for each chromosome. Dots located above the diagonal line represent chromosomes for which there were more variant alleles shared between the embryo and father than between the embryo and mother. Dots located below the diagonal line represent chromosomes for which there were more variant alleles shared between the embryo and mother than between the embryo and father. As shown in the graph in FIG. 7, all dots are well below the diagonal line. The global shift towards higher maternal contribution of alleles shown in FIG. 8 supports a finding of polyploidy that is likely maternal in origin.

[00114] In another example, CNV analysis (conducted using methods described herein) of nucleic acids from a trophectoderm biopsy of a human embryo that were sequenced at 0.1x coverage determined that the embryo was a female and had a loss of chromosome 8. FIG. 10A shows the results of the CNV analysis as a graph of chromosome copy number (CN) on the y-axis for each chromosome (listed on the x-axis) illustrating the deviation from a CN of 2 for chromosome 8. The nucleic acids of the embryo and both parents (sequenced at 0.1x) were analyzed essentially as described in Example 1 and using methods described herein to assess ploidy and pattern of inheritance. The results of the analysis are depicted in the form of a chromosome dose chart in FIG. 10B with chromosome number listed vertically and maternal dose measure on the x-axis. Chromosome dose is a relative measure of the relatedness of the maternal vs paternal samples to the embryo, expressed on a log 2 scale. Positive values indicate a stronger relationship between the embryo and the mother; negative values indicate a stronger relationship between the embryo and the father. As shown in FIG. 10B, the results identified a genome-wide maternal dosage imbalance in the embryo indicating the embryo has additional genetic matter from the mother. The loss of chromosome 8, while not called because of preference given to genome-wide gains, appears to be maternal (blue box in FIG. 10B) as the dosage chart suggests 1 copy of chromosome 8 from the mother and 1 copy of chromosome 8 from the father is present. Based on these results, the embryo is reported as 68,XXX; -8 with the origin of triploidy being maternal.

Example 5 – Classification of Parental Origin of Polyploidy

[00115] To address the utility and performance of metrics on non-consanguineous family data, results for parental consistency and novel scores on the initial $N = 14$ family data set are shown in the FIG. 10, which illustrates a graph of example parental consistency and novel scores on the initial $N = 14$ family data set, in accordance with various embodiments.

[00116] Results are classified by family structure with “Ref” referencing original unpermuted families; “MM-Maternal” (i.e., “mismatch maternal”) referencing a family where the mother was deliberately chosen to be a mother from an unrelated family; “MM-Paternal” (i.e., “mismatch paternal”) referencing a family where the father was deliberately chosen to be the father from an unrelated family, and “MM-Parents” (i.e., “mismatched parents”) referencing a family where embryos were assigned to parents from an unrelated family.

[00117] The approach of utilizing parental consistency and novelty scores, as shown in FIG. 10, successfully discriminates results as each class cleanly separates in distinct clusters. It is also observable that the clusters are not normally distributed and show significant outliers (e.g., deviations in the reference samples along the vertical axis). This can be due to chromosomal abnormalities in the embryos, embryos more genetically distinct at the population level, or sequencing or other laboratory sources of variation.

[00118] In addition, while the clusters are distinct, the best observed separation can be seen along the diagonals. As such, and as described above, transformed parental scores were calculated that also incorporate measures of embryo novelty.

[00119] The resulting plot of this calculation is illustrated by the parental matching cluster plot of FIG. 11. This figure shows four distinct clusters for no mismatch, both parents mismatched, paternal mismatch and maternal mismatch. This figure also displays potential simple thresholds (horizontal and vertical lines) that could be used for classifying the parentage of samples.

[00120] In addressing the issue of the effect of consanguineous families on classification, the initial expectation was that, in consanguineous families, there are fewer variants, given that the paternal and maternal samples look more similar than they would otherwise. As such, the parental consistency component of the scores could be more variability and the embryo novelty component could be smaller. To test this expectation, a set of 13 families were evaluated. Results for these families, both their correct scores and their scores under similar permuted scenarios, excluding samples failing sequencing QC, is illustrated by the parental matching cluster plot of FIG. 12. The results observable on the plot generally show the same clustering pattern as was originally observed (compare FIGS. 11 and 12).

Example 6 – Verification and Validation

[00121] The systems and method discussed herein, in accordance with various embodiments, was tested against gold standard for validation purposes, with FIGS. 13 and 14 illustrating the results of such test.

[00122] FIG. 13 shows a graph of relative maternal and paternal scores in determining the genetic relationship of a conceptus, in accordance with various embodiments. In particular, the

graph uses training data to test established thresholds for accuracy in isolating inheritance clusters. The test subjects were both observed polyploidy and non-polyploidy. A table of thresholds is provided by below Table 1.

TABLE 1

	Paternal Score	Maternal Score
Correct Parents	<0.155	<0.155
Unexpected Paternal Inheritance	<0.155	>0.155
Unexpected Maternal Inheritance	>0.155	<0.155
Unexpected Parental Inheritance	>0.155	>0.155

[00123] The graph produced clusters, similar to FIGS. 11-12 of Example 5, showing four distinct clusters for no mismatch, both parents mismatched, paternal mismatch and maternal mismatch.

[00124] Similar to FIG. 13, FIG. 14 shows a graph of relative maternal and paternal novelty and likeness scores in determining the genetic relationship of a conceptus, in accordance with various embodiments. In particular, the graph illustrates the accuracy of the various embodiments herein versus the gold standard SNP array technology discussed earlier. As detailed on the figure, the accuracy for classifying genetic relationships between conceptus and presumed parents was between 99.4 and 100% of gold standard. Sensitivity and specificity metrics were also exceedingly high, with only one false positive and zero false negatives reported.

Computer Implemented System

[00125] In various embodiments, the methods for determining pattern of inheritance in an embryo for a region of interest can be implemented via computer software or hardware. That is, as depicted in FIG. 8, the methods disclosed herein can be implemented on a computing device 830 that includes an alignment engine 840, a single nucleotide polymorphism identification engine (SNP

identification engine) 850, an imputation engine 860, and a relatedness engine 870. In various embodiments, the computing device 830 can be communicatively connected to a data store 810 and a display device 880 via a direct connection or through an internet connection.

[00126] It should be appreciated that the various engines depicted in FIG. 8 can be combined or collapsed into a single engine, component or module, depending on the requirements of the particular application or system architecture. Moreover, in various embodiments, the alignment engine 840, the SNP identification engine 850, the imputation engine 860, and the relatedness engine 870 can comprise additional engines or components as needed by the particular application or system architecture.

[00127] FIG. 15 is a block diagram illustrating a computer system 1500 upon which embodiments of the present teachings may be implemented. In various embodiments of the present teachings, computer system 1500 can include a bus 1502 or other communication mechanism for communicating information and a processor 1504 coupled with bus 1502 for processing information. In various embodiments, computer system 1500 can also include a memory, which can be a random-access memory (RAM) 1506 or other dynamic storage device, coupled to bus 1502 for determining instructions to be executed by processor 1504. Memory can also be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 1504. In various embodiments, computer system 1500 can further include a read only memory (ROM) 1508 or other static storage device coupled to bus 1502 for storing static information and instructions for processor 1504. A storage device 1510, such as a magnetic disk or optical disk, can be provided and coupled to bus 1502 for storing information and instructions.

[00128] In various embodiments, computer system 1500 can be coupled via bus 1502 to a display 1512, such as a cathode ray tube (CRT) or liquid crystal display (LCD), for displaying information to a computer user. An input device 1514, including alphanumeric and other keys, can be coupled to bus 1502 for communication of information and command selections to processor 1504. Another type of user input device is a cursor control 1516, such as a mouse, a trackball or cursor direction keys for communicating direction information and command selections to processor 1504 and for controlling cursor movement on display 1512. This input device 1514 typically has two degrees of freedom in two axes, a first axis (i.e., x) and a second axis (i.e., y), that

allows the device to specify positions in a plane. However, it should be understood that input devices 1514 allowing for 3-dimensional (x, y and z) cursor movement are also contemplated herein.

[00129] Consistent with certain implementations of the present teachings, results can be provided by computer system 1500 in response to processor 1504 executing one or more sequences of one or more instructions contained in memory 1506. Such instructions can be read into memory 1506 from another computer-readable medium or computer-readable storage medium, such as storage device 1510. Execution of the sequences of instructions contained in memory 1506 can cause processor 1504 to perform the processes described herein. Alternatively, hard-wired circuitry can be used in place of or in combination with software instructions to implement the present teachings. Thus, implementations of the present teachings are not limited to any specific combination of hardware circuitry and software.

[00130] The term "computer-readable medium" (e.g., data store, data storage, etc.) or "computer-readable storage medium" as used herein refers to any media that participates in providing instructions to processor 1554 for execution. Such a medium can take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Examples of non-volatile media can include, but are not limited to, dynamic memory, such as memory 1206. Examples of transmission media can include, but are not limited to, coaxial cables, copper wire, and fiber optics, including the wires that comprise bus 1502.

[00131] Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, PROM, and EPROM, a FLASH-EPROM, another memory chip or cartridge, or any other tangible medium from which a computer can read.

[00132] In addition to computer-readable medium, instructions or data can be provided as signals on transmission media included in a communications apparatus or system to provide sequences of one or more instructions to processor 1504 of computer system 1500 for execution. For example, a communication apparatus may include a transceiver having signals indicative of instructions and data. The instructions and data are configured to cause one or more processors to implement the functions outlined in the disclosure herein. Representative examples of data communications transmission connections can include, but are not limited to, telephone modem

connections, wide area networks (WAN), local area networks (LAN), infrared data connections, NFC connections, etc.

[00133] It should be appreciated that the methodologies described herein, flow charts, diagrams and accompanying disclosure can be implemented using computer system 1500 as a standalone device or on a distributed network or shared computer processing resources such as a cloud computing network.

[00134] The methodologies described herein may be implemented by various means depending upon the application. For example, these methodologies may be implemented in hardware, firmware, software, or any combination thereof. For a hardware implementation, the processing unit may be implemented within one or more application specific integrated circuits (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), processors, controllers, micro-controllers, microprocessors, electronic devices, other electronic units designed to perform the functions described herein, or a combination thereof.

[00135] In various embodiments, the methods of the present teachings may be implemented as firmware and/or a software program and applications written in conventional programming languages such as C, C++, Python, etc. If implemented as firmware and/or software, the embodiments described herein can be implemented on a non-transitory computer-readable medium in which a program is stored for causing a computer to perform the methods described above. It should be understood that the various engines described herein can be provided on a computer system, such as computer system 1500, whereby processor 1504 would execute the analyses and determinations provided by these engines, subject to instructions provided by any one of, or a combination of, memory components 1506/1508/1510 and user input provided via input device 1514.

[00136] While the present teachings are described in conjunction with various embodiments, it is not intended that the present teachings be limited to such embodiments. On the contrary, the present teachings encompass various alternatives, modifications, and equivalents, as will be appreciated by those of skill in the art.

[00137] In describing the various embodiments, the specification may have presented a method and/or process as a particular sequence of steps. However, to the extent that the method

or process does not rely on the particular order of steps set forth herein, the method or process should not be limited to the particular sequence of steps described, and one skilled in the art can readily appreciate that the sequences may be varied and still remain within the spirit and scope of the various embodiments.

Recitation of Embodiments

[00138] Embodiment 1: A method for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, comprising:

receiving conceptus, sperm provider, and oocyte provider sequence data;

aligning the received sequence data to a reference genome;

identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data;

imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference;

calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider;

calculating a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and

classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

[00139] Embodiment 2: The method of Embodiment 1, wherein the conceptus is a preimplantation conceptus.

[00140] Embodiment 3: The method of Embodiments 1 and 2, further comprising identifying a region of interest in the aligned conceptus sequence data, and identifying SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data.

[00141] Embodiment 4: The method of Embodiment 3, wherein the region of interest is genome wide.

[00142] Embodiment 5: The method of Claim 3, wherein the region of interest is a copy number variation.

[00143] Embodiment 6: The method of any one of Embodiments 1 to 5, wherein at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing.

[00144] Embodiment 7: The method of Embodiment 6, wherein the low-coverage sequencing is between about 0.001 and 10x.

[00145] Embodiment 8: The method of Embodiment 6, wherein the low-coverage sequencing is between about 0.01 and 0.5x.

[00146] Embodiment 9: The method of Embodiment 6, wherein the low-coverage sequencing is between about 0.25 and 0.2x.

[00147] Embodiment 10: The method of any one of Embodiments 1 to 9, wherein the imputation reference comprises at least 1000 genomes.

[00148] Embodiment 11: The method of any one of Embodiments 1 to 10, further comprising filtering at least one of the conceptus, sperm provider, and oocyte provider sequencing data to remove sequencing artifacts.

[00149] Embodiment 12: The method of Embodiment 11, wherein the filtering includes excluding SNPs that are not included in a reference list of known SNPs.

[00150] Embodiment 13: The method of Embodiment 12, wherein the reference list includes about 1000 known genomes.

[00151] Embodiment 14: The method of Embodiment 11, wherein the filtering comprises excluding SNPs that are inconsistent with Mendelian inheritance.

[00152] Embodiment 15: The method of Embodiment 11, wherein the filtering comprises excluding sequences of sites with any missing alleles between the conceptus, sperm and oocyte.

[00153] Embodiment 16: The method of Embodiment 11, wherein the filtering comprises excluding sequences of sites with constant alleles across between the conceptus, sperm and oocyte.

[00154] Embodiment 17: The method of Embodiment 11, wherein the filtering comprises excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte.

[00155] Embodiment 18: A method of any one of Embodiments 1 to 17, further comprising:

identifying a region of interest in the aligned conceptus sequence data,

counting the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a corresponding region on the oocyte provider sequence data to determine a maternal contribution value;

counting the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value; and

classifying a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm.

[00156] Embodiment 19: The method of Embodiment 18, wherein the region of interest is the entire genome, the method further comprising counting SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid.

[00157] Embodiment 20: The method of Embodiment 19, wherein for a conceptus that is polyploid, the method further comprising classifying a pattern of inheritance for the polyploid as maternal or paternal based on the relative contribution values between oocyte and sperm.

[00158] Embodiment 21: A non-transitory computer-readable medium storing computer instructions for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, comprising:

receiving conceptus, sperm provider, and oocyte provider sequence data;

aligning the received sequence data to a reference genome;

identifying single nucleotide polymorphisms (SNPs) in the sperm provider sequence data, oocyte provider sequence data, and the conceptus sequence data;

imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference;

calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider;

calculating a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and

classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

[00159] Embodiment 22: The method of Embodiment 21, wherein the conceptus is a preimplantation conceptus.

[00160] Embodiment 23: The method of Embodiments 21 and 22, further comprising identifying a region of interest in the aligned conceptus sequence data, and identifying SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data.

[00161] Embodiment 24: The method of Embodiment 23, wherein the region of interest is genome wide.

[00162] Embodiment 25: The method of Embodiment 23, wherein the region of interest is a copy number variation.

[00163] Embodiment 26: The method of any one of Embodiments 21 to 25, wherein at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing.

[00164] Embodiment 27: The method of Embodiment 26, wherein the low-coverage sequencing is between about 0.001 and 10x.

[00165] Embodiment 28: The method of Embodiment 26, wherein the low-coverage sequencing is between about 0.01 and 0.5x.

[00166] Embodiment 29: The method of Embodiment 26, wherein the low-coverage sequencing is between about 0.25 and 0.2x.

[00167] Embodiment 30: The method of any one of Embodiments 21 to 29, wherein the imputation reference comprises at least 1000 genomes.

[00168] Embodiment 31: The method of any one of Embodiments 21 to 30, further comprising filtering at least one of the conceptus, sperm provider, and oocyte provider sequencing data to remove sequencing artifacts.

[00169] Embodiment 32: The method of Embodiment 31, wherein the filtering includes excluding SNPs that are not included in a reference list of known SNPs.

[00170] Embodiment 33: The method of Embodiment 32, wherein the reference list includes about 1000 known genomes.

[00171] Embodiment 34: The method of Embodiment 31, wherein the filtering comprises excluding SNPs that are inconsistent with Mendelian inheritance.

[00172] Embodiment 35: The method of Embodiment 31, wherein the filtering comprises excluding sequences of sites with any missing alleles between the conceptus, sperm and oocyte.

[00173] Embodiment 36: The method of Embodiment 31, wherein the filtering comprises excluding sequences of sites with constant alleles across between the conceptus, sperm and oocyte.

[00174] Embodiment 37: The method of Embodiment 31, wherein the filtering comprises excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte.

[00175] Embodiment 38: The method of any one of Claims 21 to 37, further comprising:

identifying a region of interest in the aligned conceptus sequence data,

counting the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a corresponding region on the oocyte provider sequence data to determine a maternal contribution value;

counting the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value; and

classifying a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm.

[00176] Embodiment 39: The method of Embodiment 38, wherein the region of interest is the entire genome, the method further comprising counting SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid.

[00177] Embodiment 40: The method of Embodiment 39, wherein for a conceptus that is polyploid, the method further comprising classifying a pattern of inheritance for the polyploid as maternal or paternal based on the relative contribution values between oocyte and sperm.

[00178] Embodiment 41: A system for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, comprising:

a data store for receiving conceptus, sperm provider, and oocyte provider sequence data;

a computing device communicatively connected to the data store, the computing device comprising

an alignment engine configured to align the received sequence data to a reference genome;

a SNP identification engine configure to identify single nucleotide polymorphisms (SNPs) in the sperm provider sequence data, oocyte provider sequence data, and the conceptus sequence data;

an imputation engine configured to impute missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference; and

a relatedness engine configured to:

calculate a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider;

calculate a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and

classify the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold;

and

a display communicatively connected to the computing device and configured to display a report containing the classified relatedness to the conceptus.

[00179] Embodiment 42: The system of Embodiment 41, wherein the conceptus is a preimplantation conceptus.

[00180] Embodiment 43: The system of Embodiments 41 or 42, the alignment engine further configured to identify a region of interest in the aligned conceptus sequence data, and identify SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data.

[00181] Embodiment 44: The system of Embodiment 43, wherein the region of interest is genome wide.

[00182] Embodiment 45: The system of Embodiment 43, wherein the region of interest is a copy number variation.

[00183] Embodiment 46: The system of any one of Embodiments 41 to 45, wherein at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing.

[00184] Embodiment 47: The system of Embodiment 46, wherein the low-coverage sequencing is between about 0.001 and 10x.

[00185] Embodiment 48: The system of Embodiment 46, wherein the low-coverage sequencing is between about 0.01 and 0.5x.

[00186] Embodiment 49: The system of Embodiment 46, wherein the low-coverage sequencing is between about 0.25 and 0.2x.

[00187] Embodiment 50: The system of any one of Embodiments 41 to 49, wherein the imputation reference comprises at least 1000 genomes.

[00188] Embodiment 51: The system of any one of Embodiments 41 to 50, the imputation engine further configured to filter at least one of the conceptus, sperm provider, and oocyte provider sequencing data to remove sequencing artifacts.

[00189] Embodiment 52: The system of Embodiment 51, wherein the filtering includes excluding SNPs that are not included in a reference list of known SNPs.

[00190] Embodiment 53: The system of Embodiment 52, wherein the reference list includes about 1000 known genomes.

[00191] Embodiment 54: The system of Embodiment 51, wherein the filtering comprises excluding SNPs that are inconsistent with Mendelian inheritance.

[00192] Embodiment 55: The system of Embodiment 51, wherein the filtering comprises excluding sequences of sites with any missing alleles between the conceptus, sperm and oocyte.

[00193] Embodiment 56: The system of Claim Embodiment 51, wherein the filtering comprises excluding sequences of sites with constant alleles across between the conceptus, sperm and oocyte.

[00194] Embodiment 57: The system of Embodiment 51, wherein the filtering comprises excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte.

[00195] Embodiment 58: The system of any one of Embodiments 41 to 47,

the alignment engine further configured to identify a region of interest in the aligned conceptus sequence data, and

the relatedness engine further configured to

count the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a corresponding region on the oocyte provider sequence data to determine a maternal contribution value;

count the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value; and

classify a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm.

[00196] Embodiment 59: The system of Embodiment 58, wherein the region of interest is the entire genome, the relatedness engine configured to count SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid.

[00197] Embodiment 60: The system of Embodiment 59, wherein for a conceptus that is polyploid, the relatedness engine configured to classify a pattern of inheritance for the polyploid as maternal or paternal based on the relative contribution values between oocyte and sperm.

What is claimed:

1. A method for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, comprising:

receiving conceptus, sperm provider, and oocyte provider sequence data;

aligning the received sequence data to a reference genome;

identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data;

imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference;

calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider;

calculating a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and

classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

2. The method of Claim 1, wherein the conceptus is a preimplantation conceptus.

3. The method of Claim 1, further comprising identifying a region of interest in the aligned conceptus sequence data, and identifying SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data.

4. The method of Claim 3, wherein the region of interest is genome wide.

5. The method of Claim 3, wherein the region of interest is a copy number variation.

6. The method of Claim 1, wherein at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing.

7. The method of Claim 6, wherein the low-coverage sequencing is between about 0.001 and 10x.

8. The method of Claim 6, wherein the low-coverage sequencing is between about 0.01 and 0.5x.

9. The method of Claim 6, wherein the low-coverage sequencing is between about 0.25 and 0.2x.

10. The method of Claim 1, wherein the imputation reference comprises at least 1000 genomes.

11. The method of Claim 1, further comprising filtering at least one of the conceptus, sperm provider, and oocyte provider sequencing data to remove sequencing artifacts.

12. The method of Claim 11, wherein the filtering includes excluding SNPs that are not included in a reference list of known SNPs.

13. The method of Claim 12, wherein the reference list includes about 1000 known genomes.

14. The method of Claim 11, wherein the filtering comprises excluding SNPs that are inconsistent with Mendelian inheritance.

15. The method of Claim 11, wherein the filtering comprises excluding sequences of sites with any missing alleles between the conceptus, sperm and oocyte.

16. The method of Claim 11, wherein the filtering comprises excluding sequences of sites with constant alleles across between the conceptus, sperm and oocyte.

17. The method of Claim 11, wherein the filtering comprises excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte.

18. The method of Claim 1, further comprising:

identifying a region of interest in the aligned conceptus sequence data,

counting the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a

corresponding region on the oocyte provider sequence data to determine a maternal contribution value;

counting the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value; and

classifying a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm.

19. The method of Claim 18, wherein the region of interest is the entire genome, the method further comprising counting SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid.

20. The method of Claim 19, wherein for a conceptus that is polyploid, the method further comprising classifying a pattern of inheritance for the polyploid as maternal or paternal based on the relative contribution values between oocyte and sperm.

21. A non-transitory computer-readable medium storing computer instructions for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, comprising:

receiving conceptus, sperm provider, and oocyte provider sequence data;

aligning the received sequence data to a reference genome;

identifying single nucleotide polymorphisms (SNPs) in the sperm provider sequence data, oocyte provider sequence data, and the conceptus sequence data;

imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference;

calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider;

calculating a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and

classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold.

22. The method of Claim 21, wherein the conceptus is a preimplantation conceptus.

23. The method of Claim 21, further comprising identifying a region of interest in the aligned conceptus sequence data, and identifying SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data.

24. The method of Claim 23, wherein the region of interest is genome wide.

25. The method of Claim 23, wherein the region of interest is a copy number variation.

26. The method of Claim 21, wherein at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing.

27. The method of Claim 26, wherein the low-coverage sequencing is between about 0.001 and 10x.

28. The method of Claim 26, wherein the low-coverage sequencing is between about 0.01 and 0.5x.

29. The method of Claim 26, wherein the low-coverage sequencing is between about 0.25 and 0.2x.

30. The method of Claim 21, wherein the imputation reference comprises at least 1000 genomes.

31. The method of Claim 21, further comprising filtering at least one of the conceptus, sperm provider, and oocyte provider sequencing data to remove sequencing artifacts.

32. The method of Claim 31, wherein the filtering includes excluding SNPs that are not included in a reference list of known SNPs.

33. The method of Claim 32, wherein the reference list includes about 1000 known genomes.

34. The method of Claim 31, wherein the filtering comprises excluding SNPs that are inconsistent with Mendelian inheritance.

35. The method of Claim 31, wherein the filtering comprises excluding sequences of sites with any missing alleles between the conceptus, sperm and oocyte.

36. The method of Claim 31, wherein the filtering comprises excluding sequences of sites with constant alleles across between the conceptus, sperm and oocyte.

37. The method of Claim 31, wherein the filtering comprises excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte.

38. The method of Claim 21, further comprising:

identifying a region of interest in the aligned conceptus sequence data,

counting the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a corresponding region on the oocyte provider sequence data to determine a maternal contribution value;

counting the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value; and

classifying a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm.

39. The method of Claim 38, wherein the region of interest is the entire genome, the method further comprising counting SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid.

40. The method of Claim 39, wherein for a conceptus that is polyploid, the method further comprising classifying a pattern of inheritance for the polyploid as maternal or paternal based on the relative contribution values between oocyte and sperm.

41. A system for determining the genetic relationship of a conceptus with a sperm provider and oocyte provider, comprising:

a data store for receiving conceptus, sperm provider, and oocyte provider sequence data;

a computing device communicatively connected to the data store, the computing device comprising

an alignment engine configured to align the received sequence data to a reference genome;

a SNP identification engine configure to identify single nucleotide polymorphisms (SNPs) in the sperm provider sequence data, oocyte provider sequence data, and the conceptus sequence data;

an imputation engine configured to impute missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference; and

a relatedness engine configured to:

calculate a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider;

calculate a maternal consistency score between the oocyte provider and conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider; and

classify the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold;

and

a display communicatively connected to the computing device and configured to display a report containing the classified relatedness to the conceptus.

42. The system of Claim 41, wherein the conceptus is a preimplantation conceptus.

43. The system of Claim 41, the alignment engine further configured to identify a region of interest in the aligned conceptus sequence data, and identify SNPs in the sperm provider, oocyte provider, and the identified region of interest in the conceptus sequence data.

44. The system of Claim 43, wherein the region of interest is genome wide.

45. The system of Claim 43, wherein the region of interest is a copy number variation.

46. The system of Claim 41, wherein at least one of the conceptus, sperm provider, and oocyte provider sequence data is acquired by low-coverage sequencing.

47. The system of Claim 46, wherein the low-coverage sequencing is between about 0.001 and 10x.

48. The system of Claim 46, wherein the low-coverage sequencing is between about 0.01 and 0.5x.

49. The system of Claim 46, wherein the low-coverage sequencing is between about 0.25 and 0.2x.

50. The system of Claim 41, wherein the imputation reference comprises at least 1000 genomes.

51. The system of Claim 41, the imputation engine further configured to filter at least one of the conceptus, sperm provider, and oocyte provider sequencing data to remove sequencing artifacts.

52. The system of Claim 51, wherein the filtering includes excluding SNPs that are not included in a reference list of known SNPs.

53. The system of Claim 52, wherein the reference list includes about 1000 known genomes.

54. The system of Claim 51, wherein the filtering comprises excluding SNPs that are inconsistent with Mendelian inheritance.

55. The system of Claim 51, wherein the filtering comprises excluding sequences of sites with any missing alleles between the conceptus, sperm and oocyte.

56. The system of Claim 51, wherein the filtering comprises excluding sequences of sites with constant alleles across between the conceptus, sperm and oocyte.

57. The system of Claim 51, wherein the filtering comprises excluding sequences of sites with a novel allele within one of the conceptus, sperm and oocyte.

58. The system of Claim 41,

the alignment engine further configured to identify a region of interest in the aligned conceptus sequence data, and

the relatedness engine further configured to

count the number of SNPs that are common between the conceptus and the oocyte in the identified region of interest for the conceptus sequence data and a corresponding region on the oocyte provider sequence data to determine a maternal contribution value;

count the number of SNPs that are common between the conceptus and the sperm in the identified region of interest for the conceptus sequence data and a corresponding region on the sperm provider sequence data to determine a paternal contribution value; and

classify a pattern of inheritance for the conceptus as maternal or paternal based on the relative contribution values between oocyte and sperm.

59. The system of Claim 58, wherein the region of interest is the entire genome, the relatedness engine configured to count SNPs across the entire genome to determine the maternal and paternal contribution values and determine if the conceptus is polyploid.

60. The system of Claim 59, wherein for a conceptus that is polyploid, the relatedness engine configured to classify a pattern of inheritance for the polyploid as maternal or paternal based on the relative contribution values between oocyte and sperm.

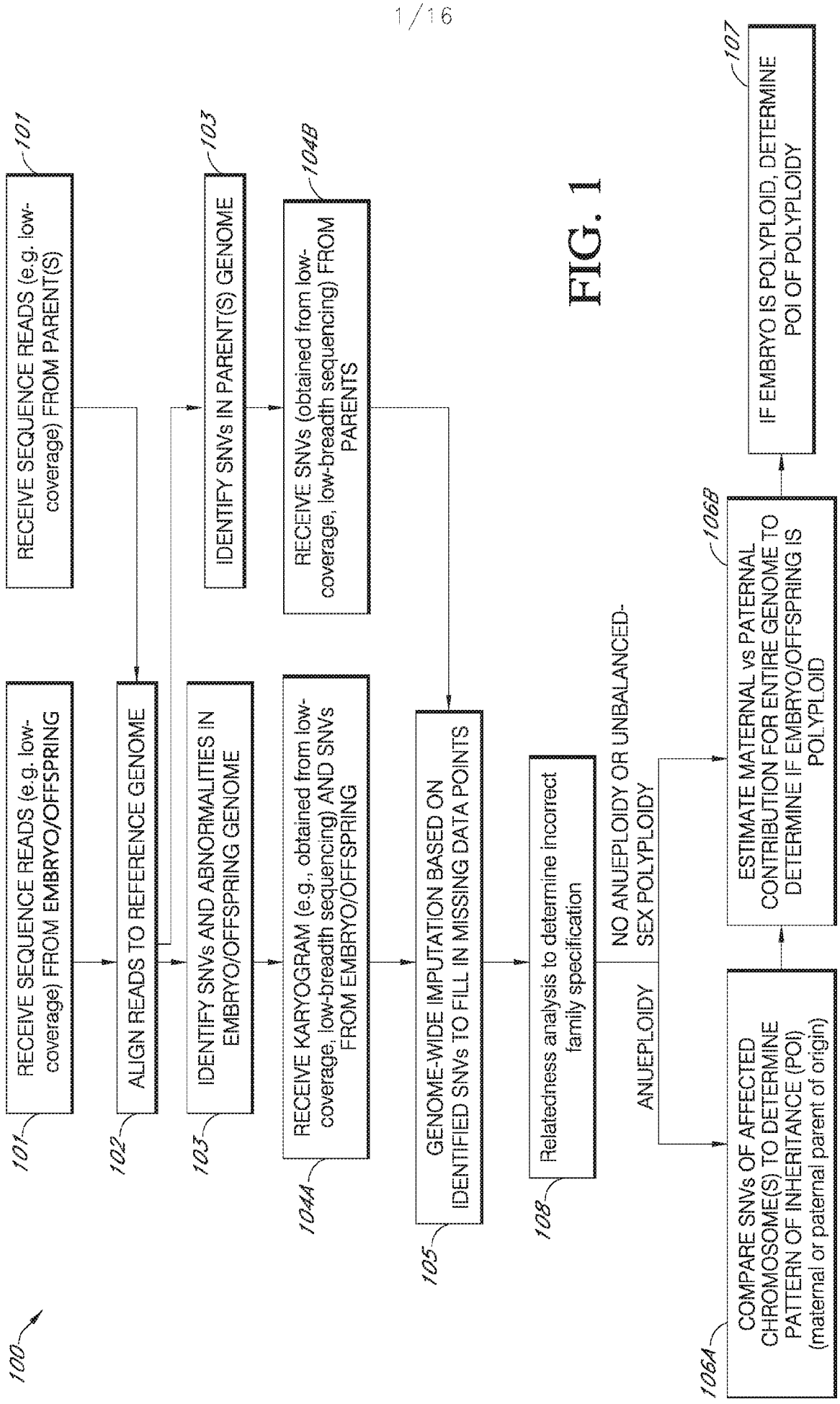

100
101
RECEIVE SEQUENCE READS (e.g. low-coverage) FROM EMBRYO/OFFSPRING
101
RECEIVE SEQUENCE READS (e.g. low-coverage) FROM PARENT(S)
102
ALIGN READS TO REFERENCE GENOME
103
IDENTIFY SNVs AND ABNORMALITIES IN EMBRYO/OFFSPRING GENOME
103
IDENTIFY SNVs IN PARENT(S) GENOME
104A
RECEIVE KARYOGRAM (e.g., obtained from low-coverage, low-breadth sequencing) AND SNVs FROM EMBRYO/OFFSPRING
104B
RECEIVE SNVs (obtained from low-coverage, low-breadth sequencing) FROM PARENTS
105
GENOME-WIDE IMPUTATION BASED ON IDENTIFIED SNVs TO FILL IN MISSING DATA POINTS
108
Relatedness analysis to determine incorrect family specification
ANUEPLOIDY
NO ANUEPLOIDY OR UNBALANCED-SEX POLYPLOIDY
106A
COMPARE SNVs OF AFFECTED CHROMOSOME(S) TO DETERMINE PATTERN OF INHERITANCE (POI) (maternal or paternal parent of origin)
106B
ESTIMATE MATERNAL vs PATERNAL CONTRIBUTION FOR ENTIRE GENOME TO DETERMINE IF EMBRYO/OFFSPRING IS POLYPLOID
107
IF EMBRYO IS POLYPLOID, DETERMINE POI OF POLYPLOIDY

FIG. 1

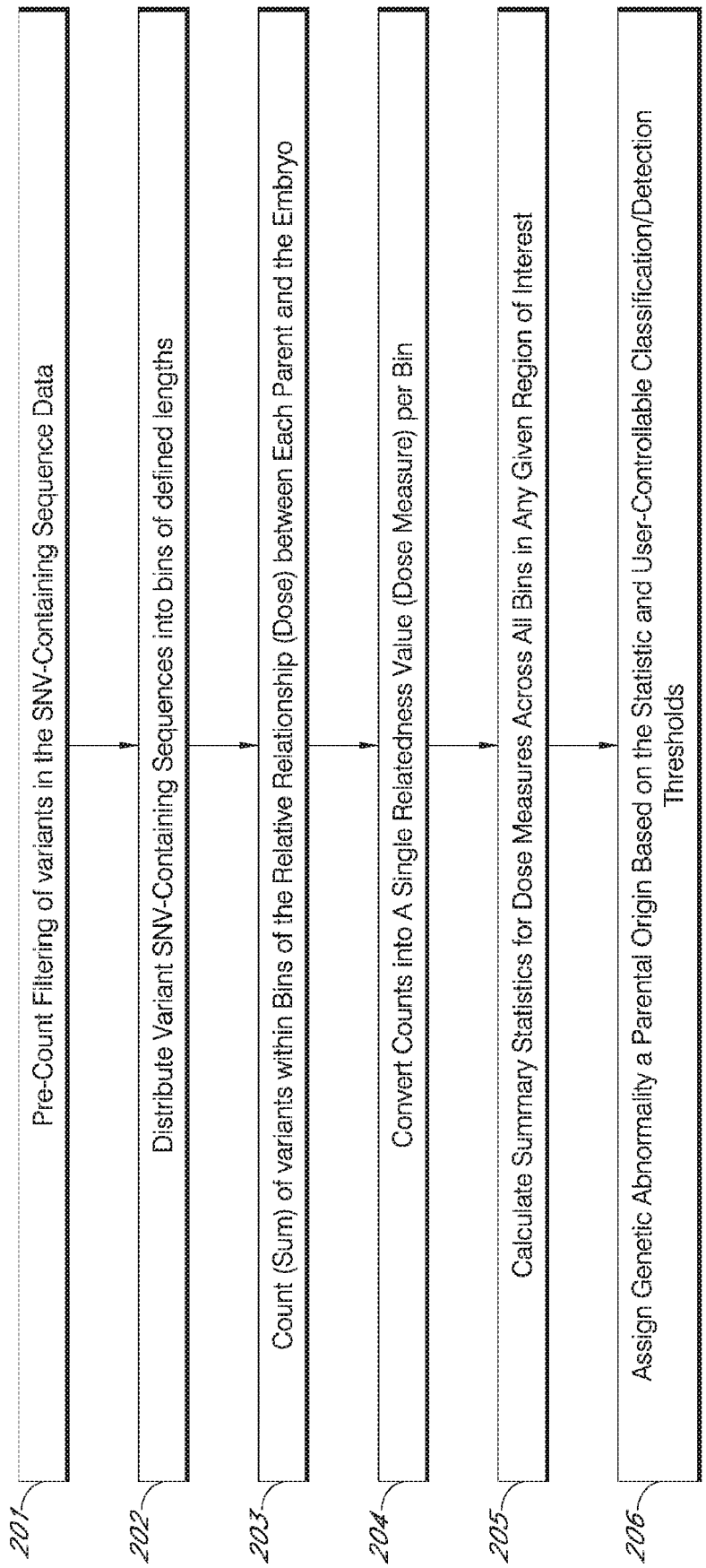

200
Pre-Count Filtering of variants in the SNV-Containing Sequence Data
Distribute Variant SNV-Containing Sequences into bins of defined lengths
Count (Sum) of variants within Bins of the Relative Relationship (Dose) between Each Parent and the Embryo
Convert Counts into A Single Relatedness Value (Dose Measure) per Bin
Calculate Summary Statistics for Dose Measures Across All Bins in Any Given Region of Interest
Assign Genetic Abnormality a Parental Origin Based on the Statistic and User-Controllable Classification/Detection Thresholds
201
202
203
204
205
206

FIG. 2

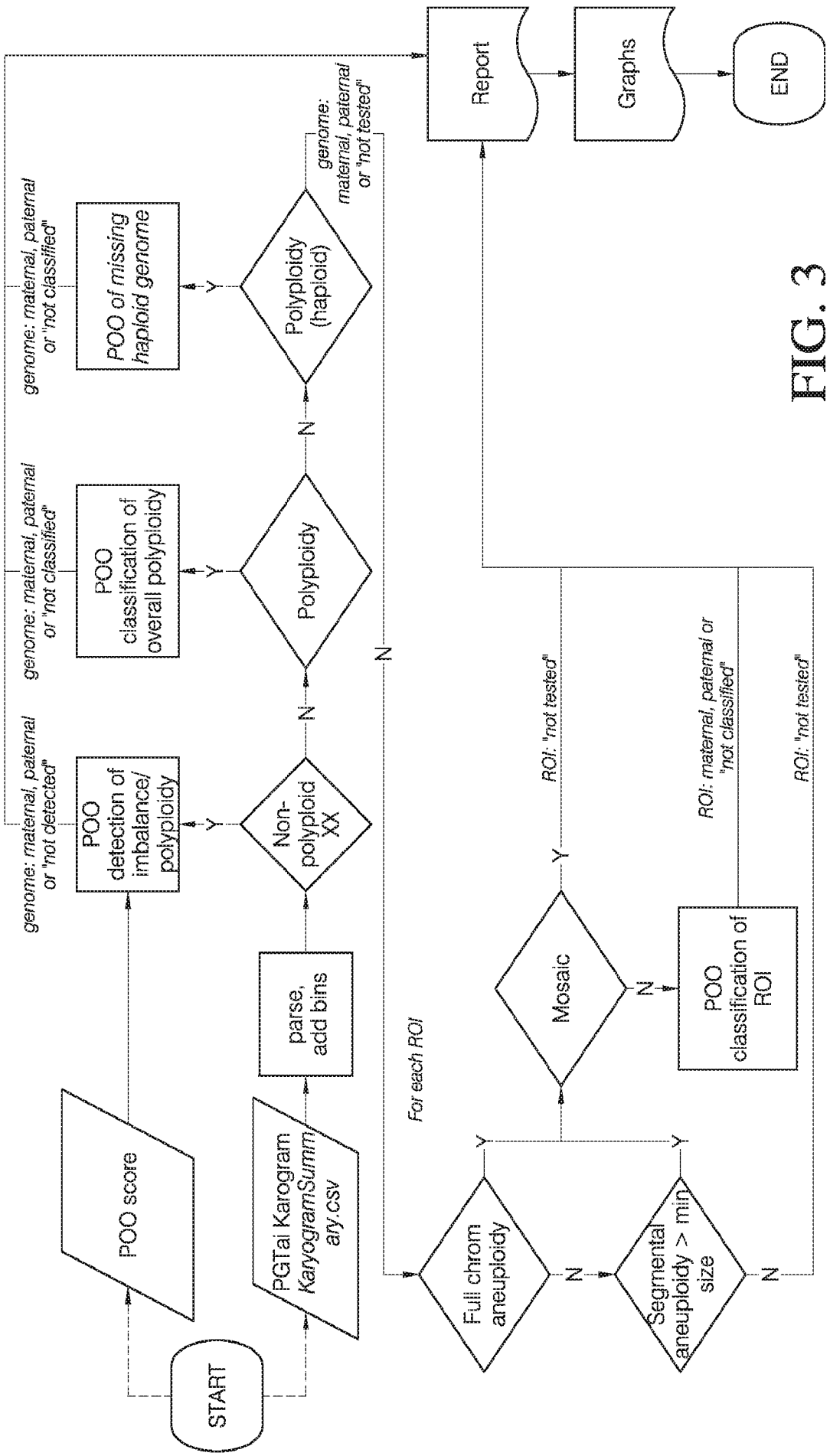

START
POO score
PGTai Karogram KaryogramSummary.csv
parse, add bins
For each ROI
Full chrom aneuploidy
Segmental aneuploidy > min size
Mosaic
POO classification of ROI
ROI: "not tested"
ROI: maternal, paternal or "not classified"
ROI: "not tested"
Y
N
Non-polyploid XX
POO detection of imbalance/ polyploidy
genome: maternal, paternal or "not detected"
Polyploidy
POO classification of overall polyploidy
genome: maternal, paternal or "not classified"
Polyploidy (haploid)
POO of missing haploid genome
genome: maternal, paternal or "not classified"
genome: maternal, paternal or "not tested"
Report
Graphs
END

FIG. 3

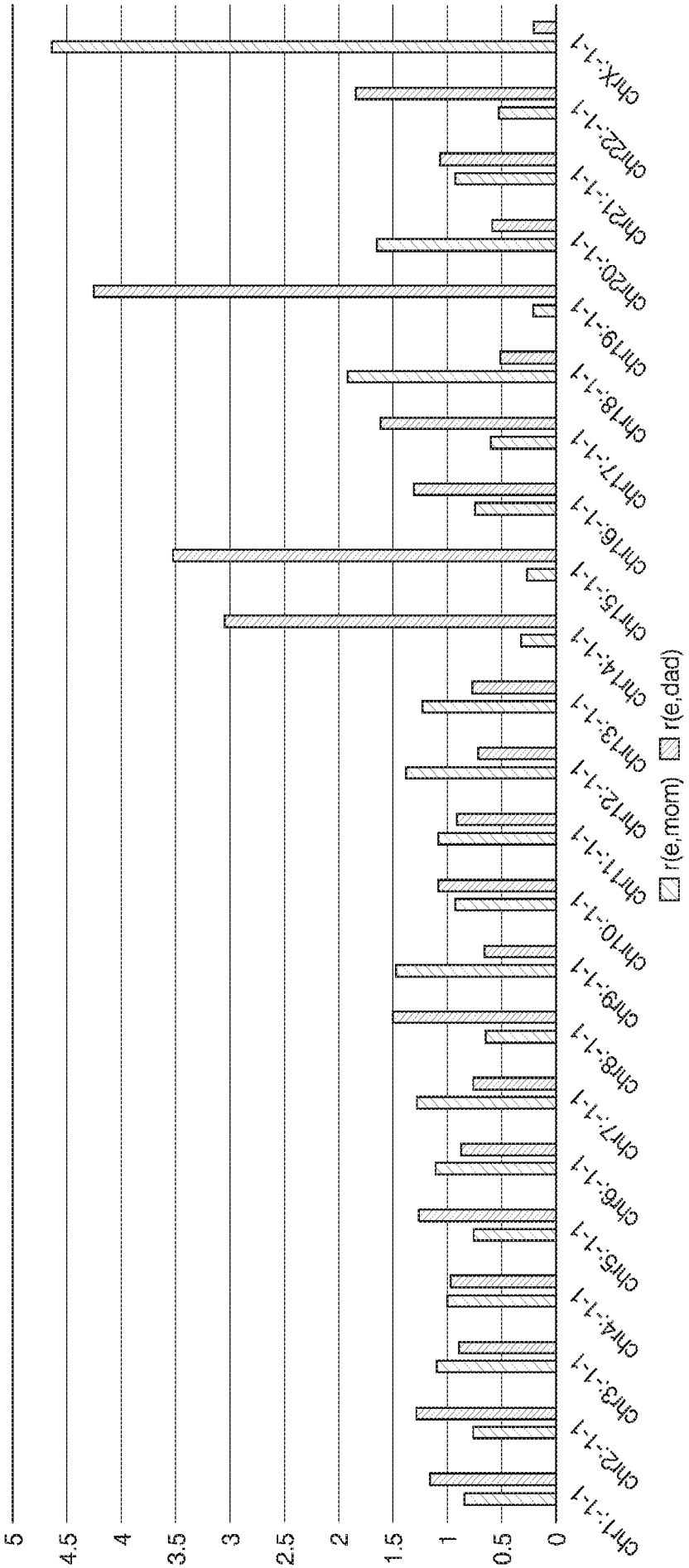

chrX:-1-1
chr22:-1-1
chr21:-1-1
chr20:-1-1
chr19:-1-1
chr18:-1-1
chr17:-1-1
chr16:-1-1
chr15:-1-1
chr14:-1-1
chr13:-1-1
chr12:-1-1
chr11:-1-1
chr10:-1-1
chr9:-1-1
chr8:-1-1
chr7:-1-1
chr6:-1-1
chr5:-1-1
chr4:-1-1
chr3:-1-1
chr2:-1-1
chr1:-1-1
r(e,mom)
r(e,dad)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0

FIG. 4

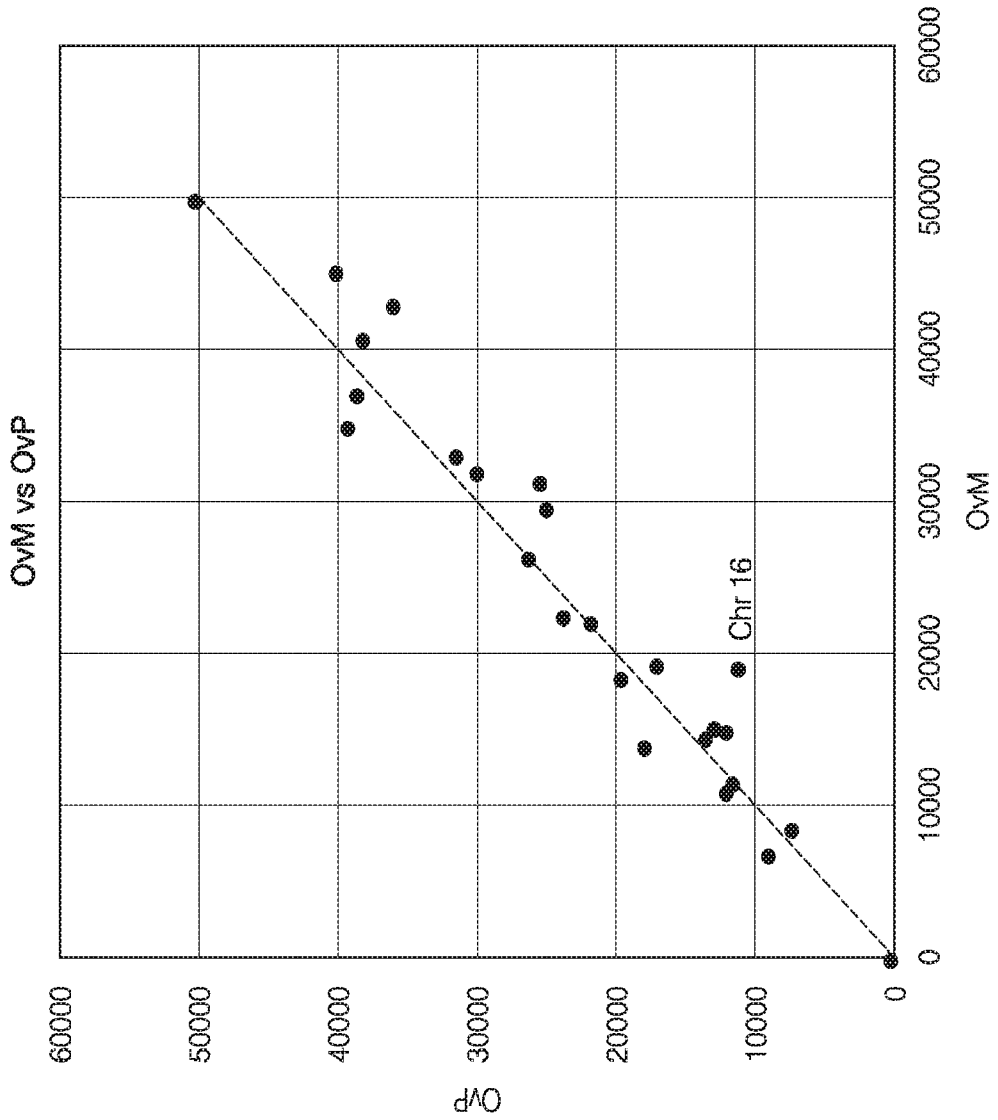

OvM vs OvP
Chr 16
OvP
OvM
0
10000
20000
30000
40000
50000
60000

FIG. 5A

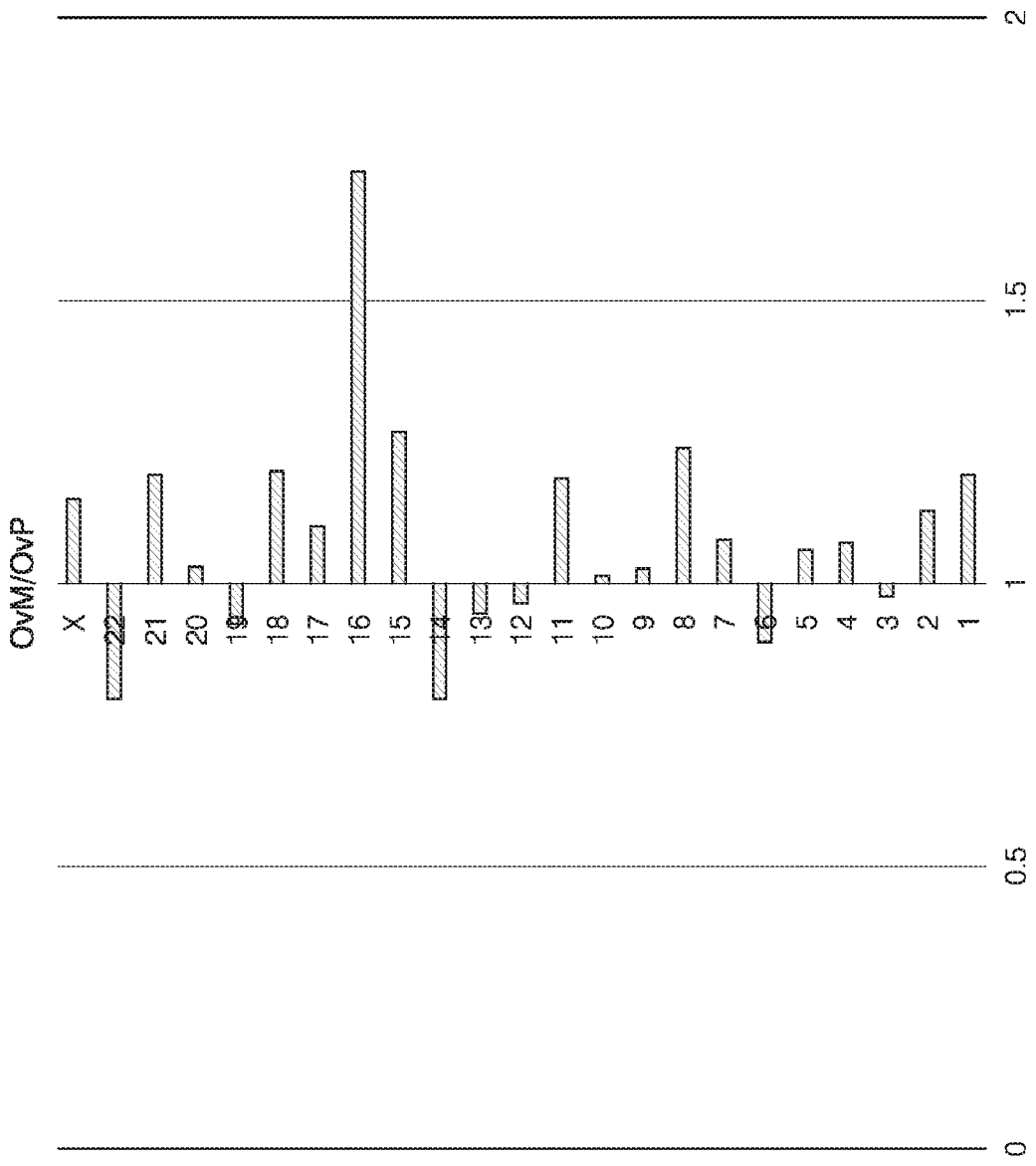

OvM/OvP
X
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0.5
1
1.5
2
Maternal/Paternal Dose Measure

FIG. 5B

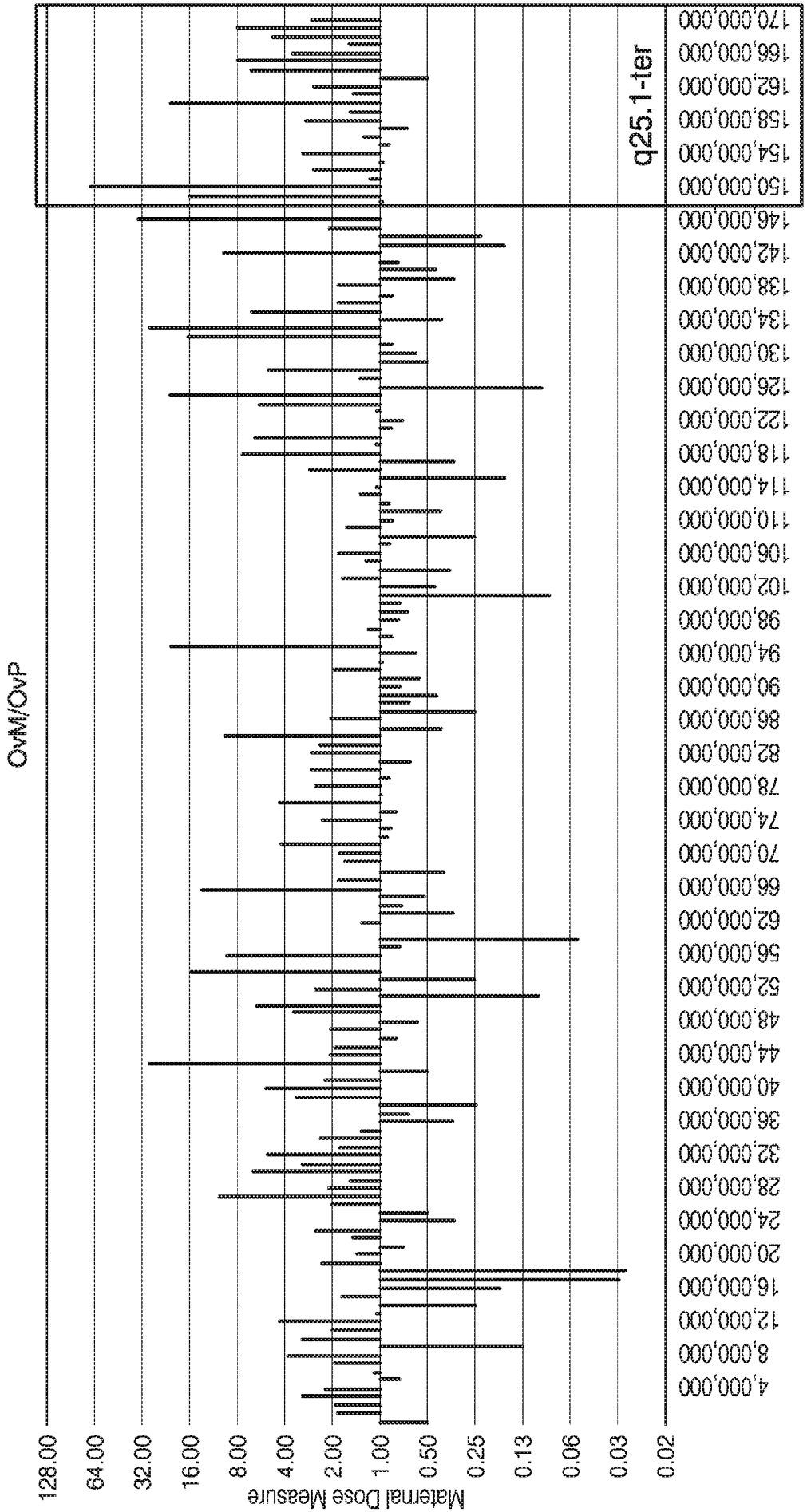

OvM/OvP
q25.1-ter
170,000,000
166,000,000
162,000,000
158,000,000
154,000,000
150,000,000
146,000,000
142,000,000
138,000,000
134,000,000
130,000,000
126,000,000
122,000,000
118,000,000
114,000,000
110,000,000
106,000,000
102,000,000
98,000,000
94,000,000
90,000,000
86,000,000
82,000,000
78,000,000
74,000,000
70,000,000
66,000,000
62,000,000
56,000,000
52,000,000
48,000,000
44,000,000
40,000,000
36,000,000
32,000,000
28,000,000
24,000,000
20,000,000
16,000,000
12,000,000
8,000,000
4,000,000
128.00
64.00
32.00
16.00
8.00
4.00
2.00
1.00
0.50
0.25
0.13
0.06
0.03
0.02
Maternal Dose Measure

FIG. 6

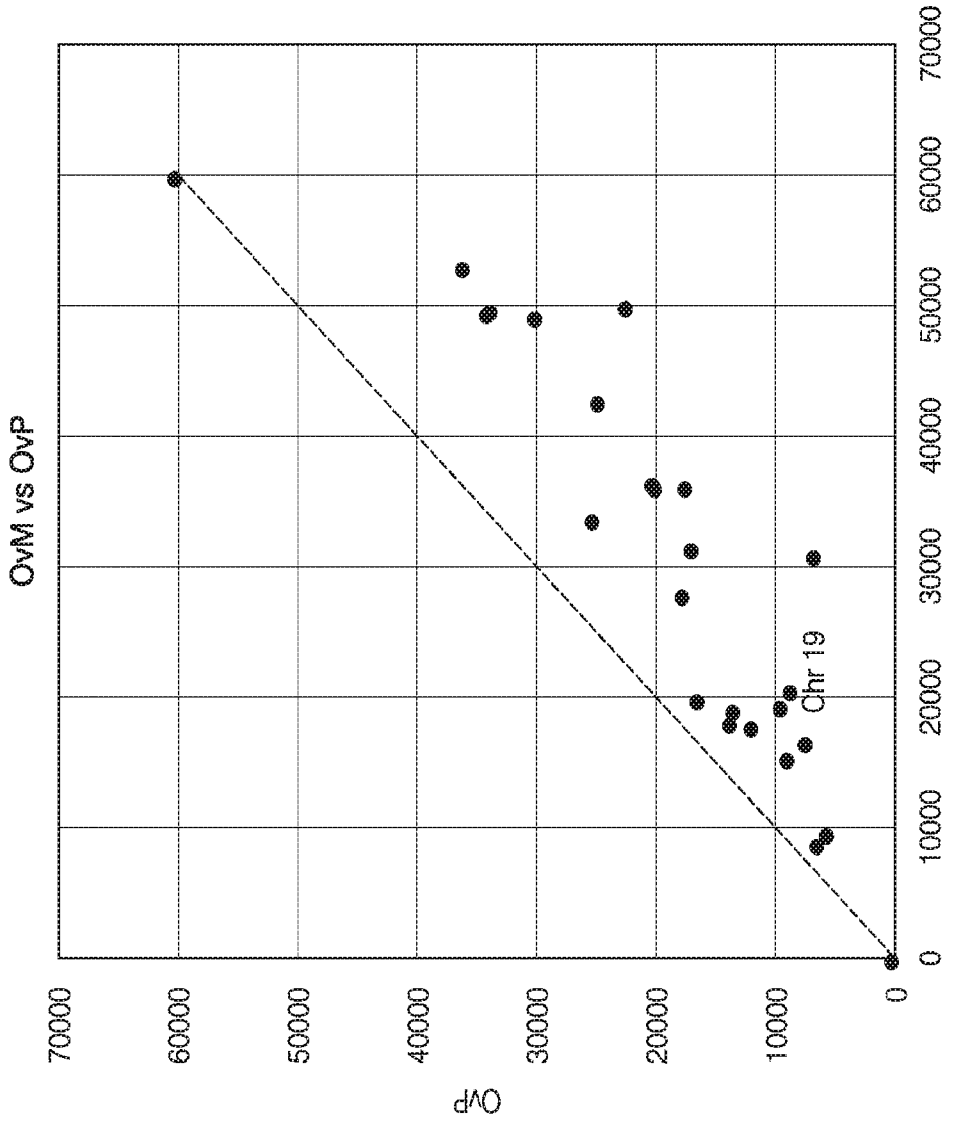

OvM vs OvP
OvP
OvM
0
10000
20000
30000
40000
50000
60000
70000
Chr 19

FIG. 7

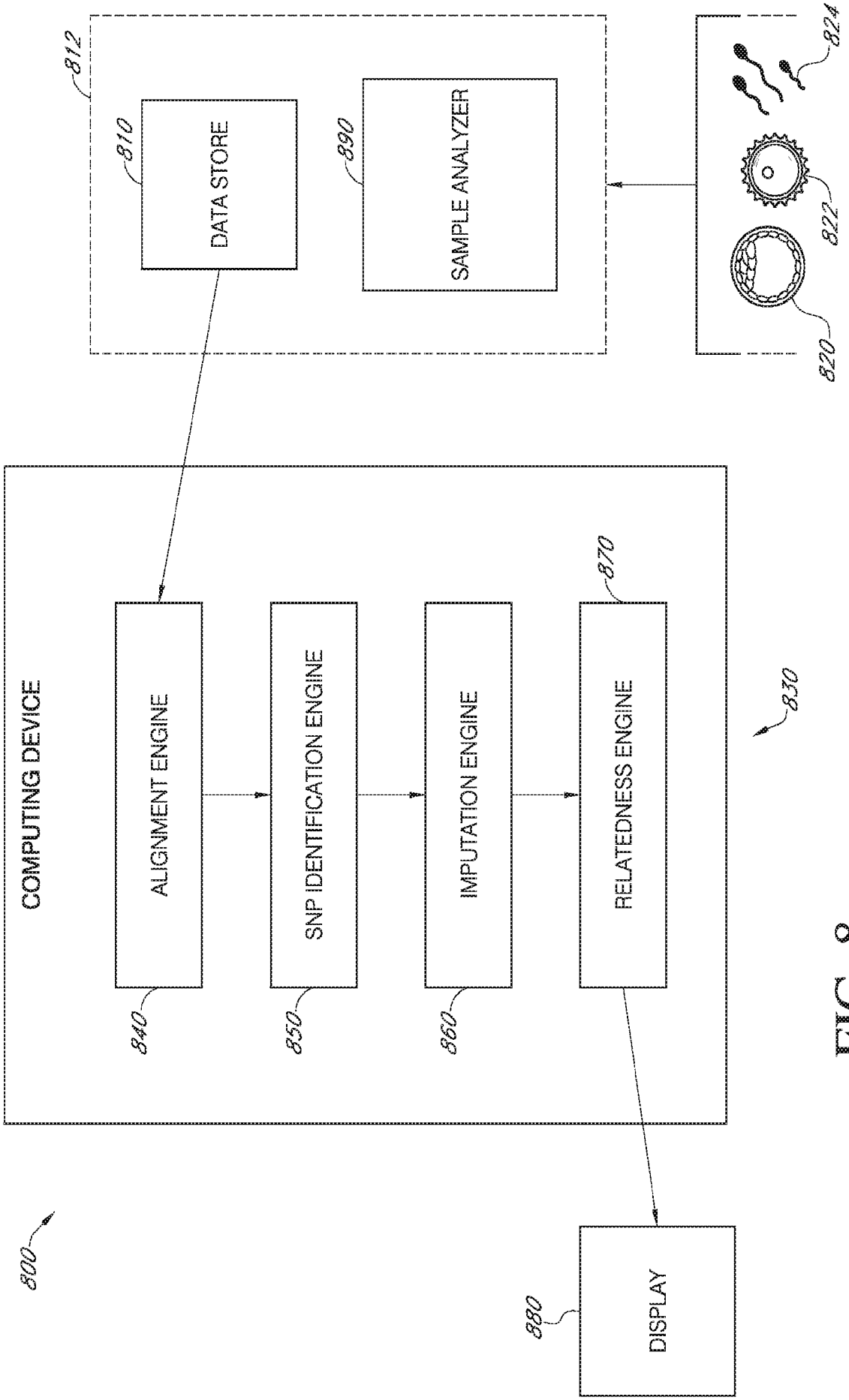

812
810
DATA STORE
890
SAMPLE ANALYZER
824
822
820
COMPUTING DEVICE
830
ALIGNMENT ENGINE
840
SNP IDENTIFICATION ENGINE
850
IMPUTATION ENGINE
860
870
RELATEDNESS ENGINE
800
DISPLAY
880

FIG. 8

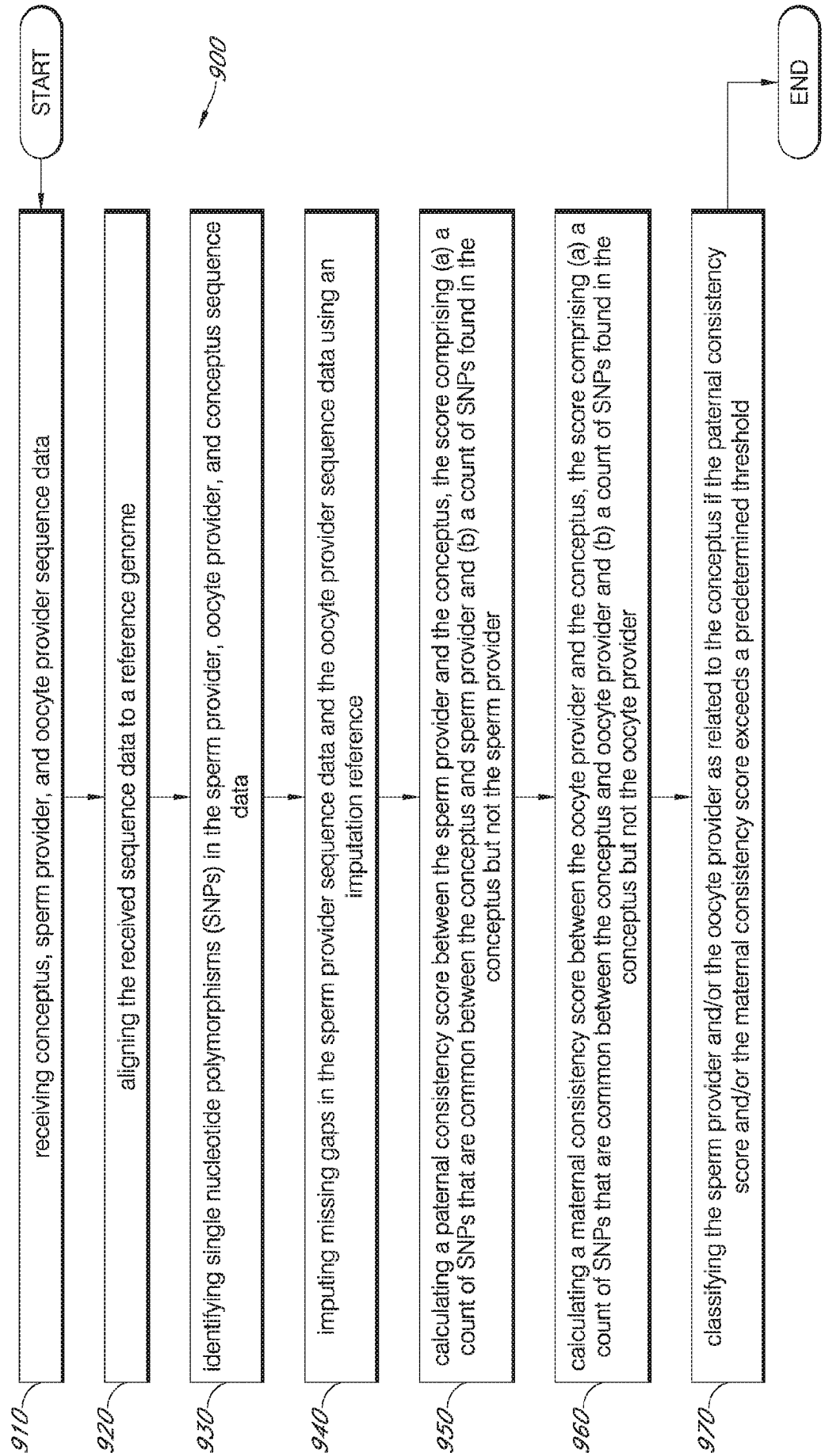

START
900
receiving conceptus, sperm provider, and oocyte provider sequence data
910
aligning the received sequence data to a reference genome
920
identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data
930
imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference
940
calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider
950
calculating a maternal consistency score between the oocyte provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider
960
classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold
970
END

FIG. 9

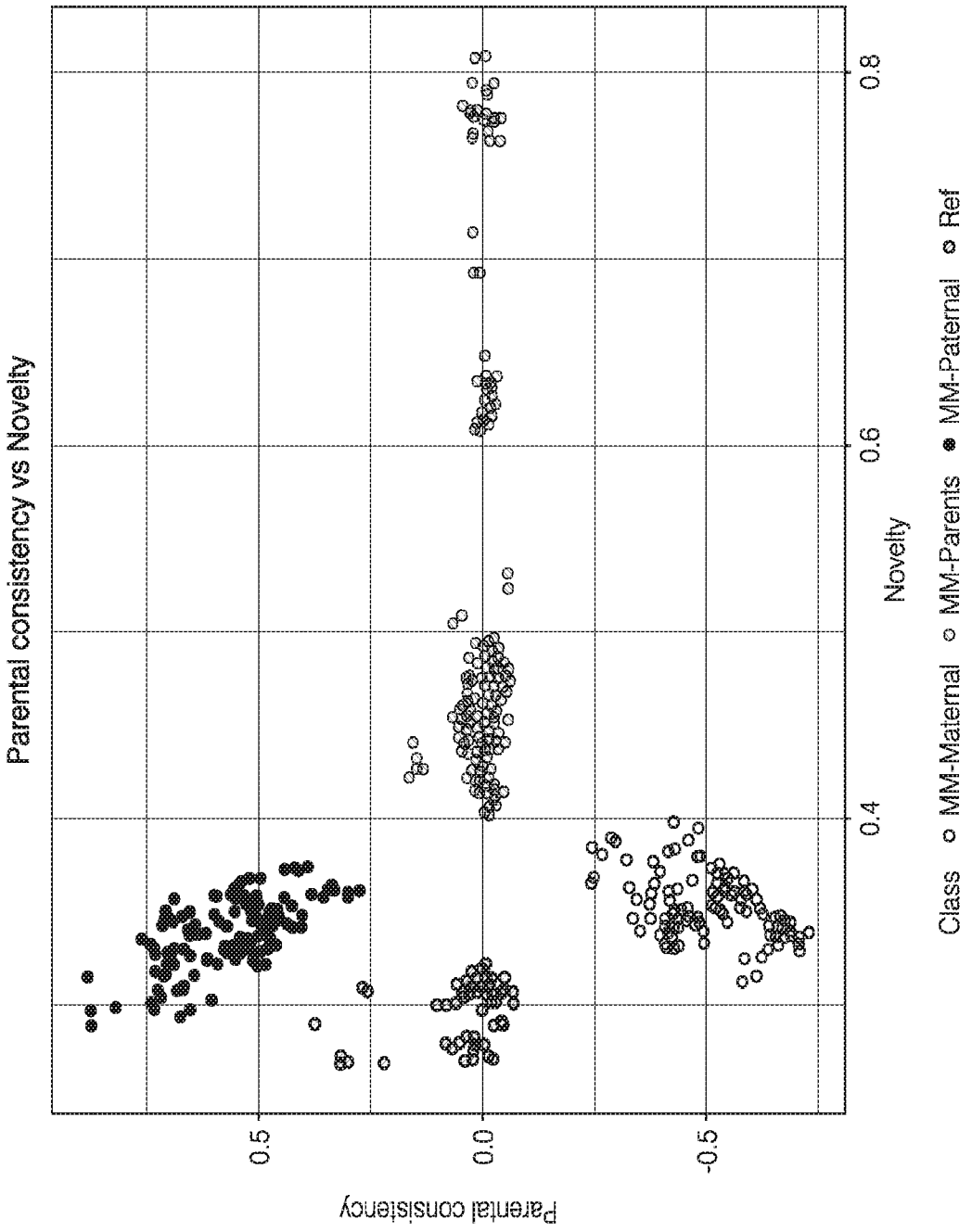

Parental consistency vs Novelty
Parental consistency
0.5
0.0
-0.5
Novelty
0.4
0.6
0.8
Class ○ MM-Maternal ○ MM-Parents ● MM-Paternal ○ Ref

FIG. 10

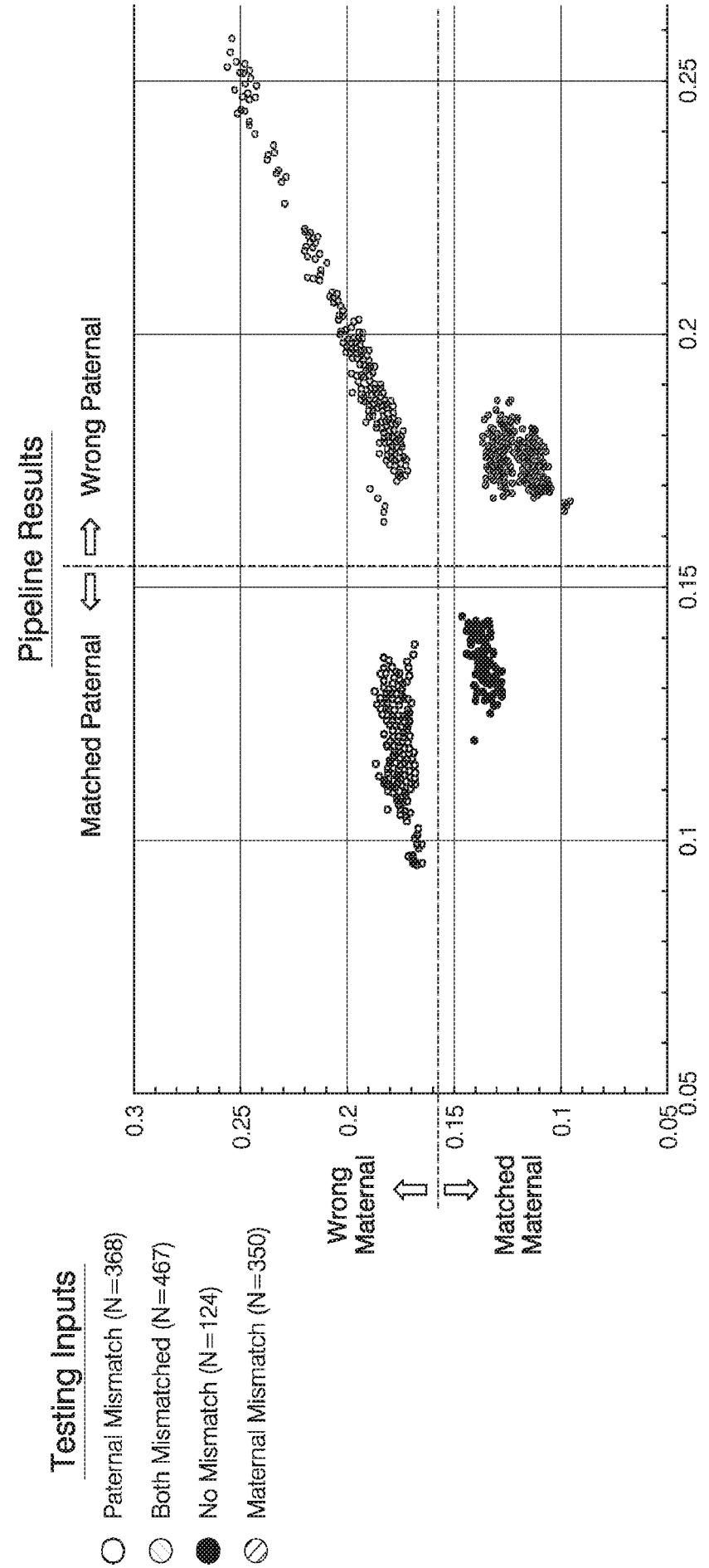

Pipeline Results
Matched Paternal
Wrong Paternal
Testing Inputs
Paternal Mismatch (N=368)
Both Mismatched (N=467)
No Mismatch (N=124)
Maternal Mismatch (N=350)
Wrong Maternal
Matched Maternal
0.3
0.25
0.2
0.15
0.1
0.05
0.05
0.1
0.15
0.2
0.25

FIG. 11

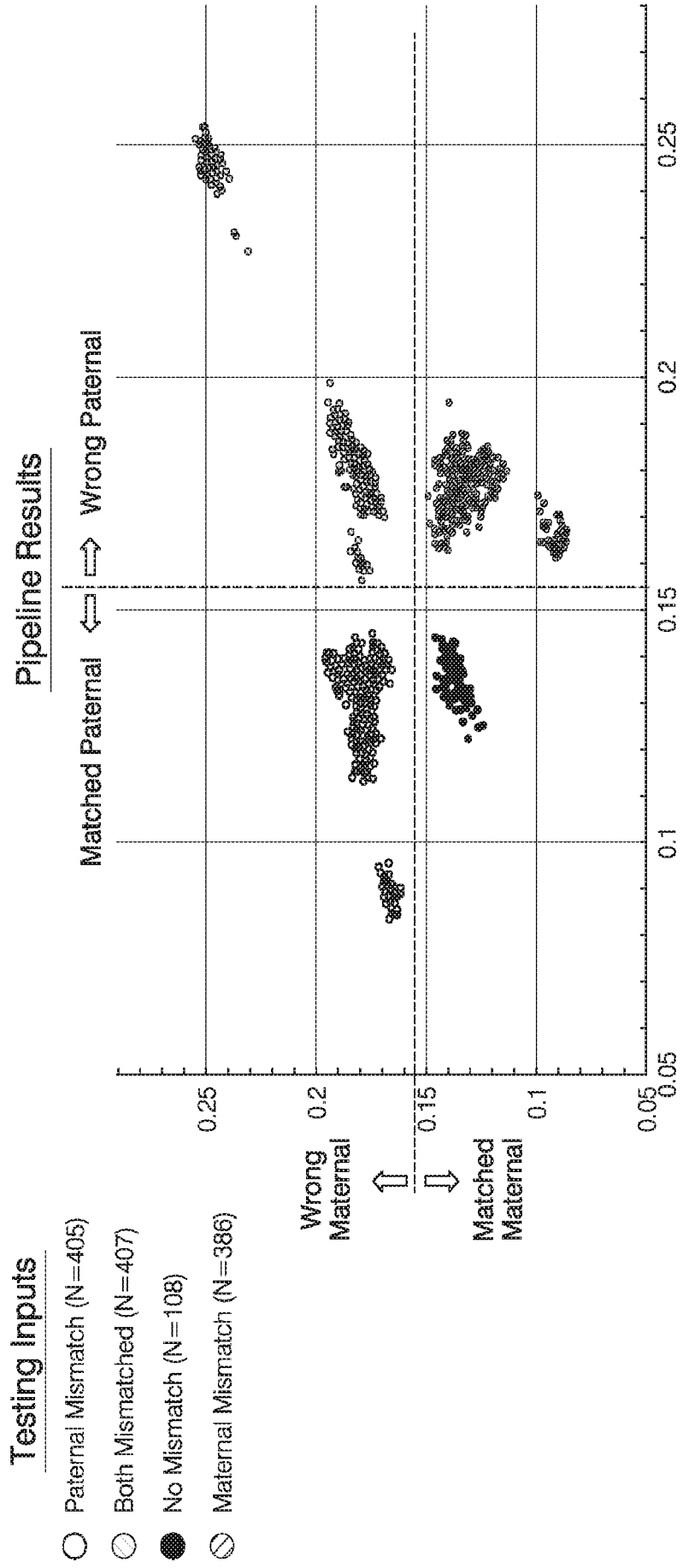

Pipeline Results
Matched Paternal
Wrong Paternal
Wrong Maternal
Matched Maternal
0.25
0.2
0.15
0.1
0.05
Testing Inputs
Paternal Mismatch (N=405)
Both Mismatched (N=407)
No Mismatch (N=108)
Maternal Mismatch (N=386)

FIG. 12

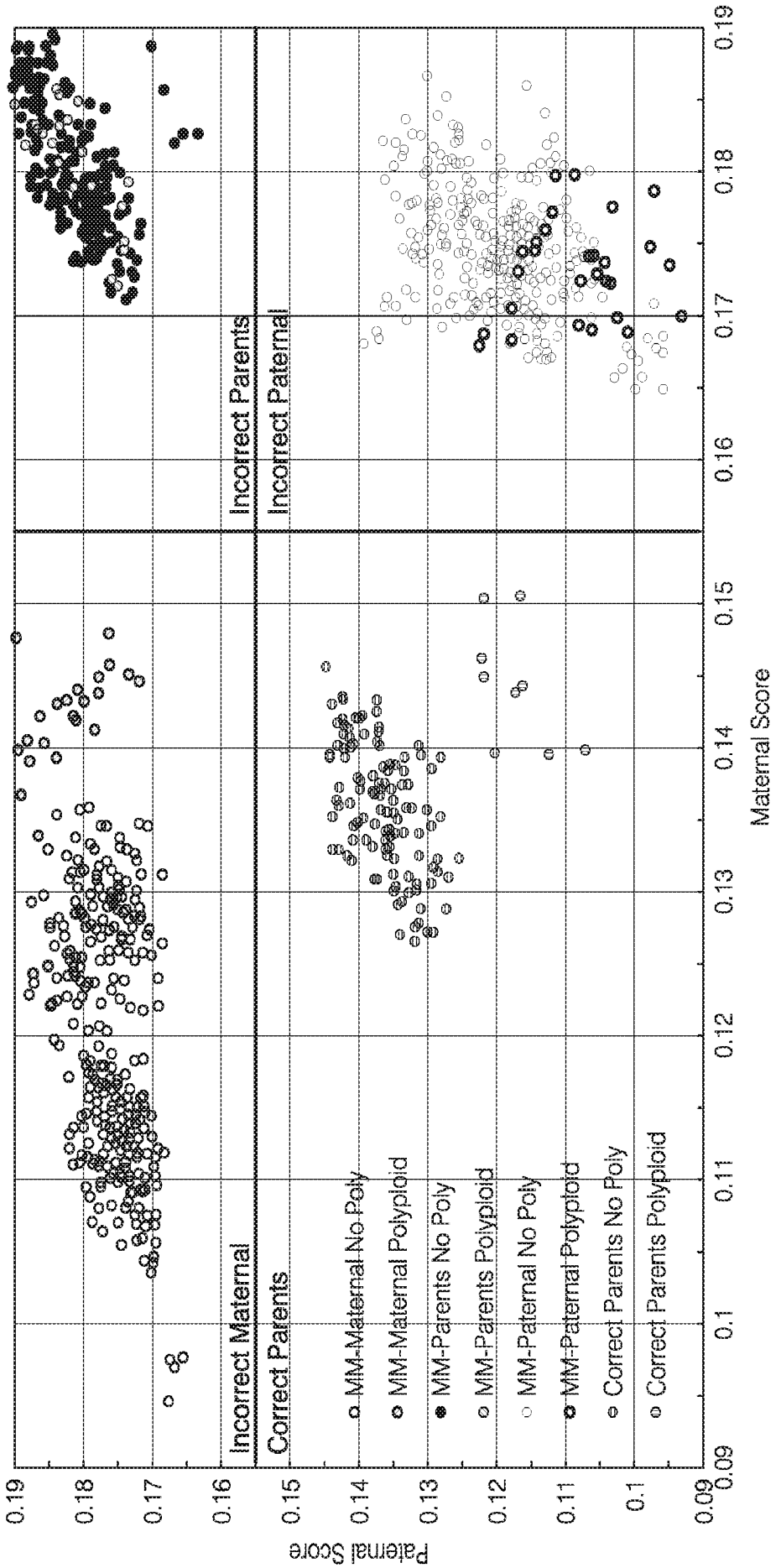

Incorrect Parents
Incorrect Paternal
Incorrect Maternal
Correct Parents
MM-Maternal No Poly
MM-Maternal Polyploid
MM-Parents No Poly
MM-Parents Polyploid
MM-Paternal No Poly
MM-Paternal Polyploid
Correct Parents No Poly
Correct Parents Polyploid
Maternal Score
Paternal Score
0.09
0.1
0.11
0.12
0.13
0.14
0.15
0.16
0.17
0.18
0.19

FIG. 13

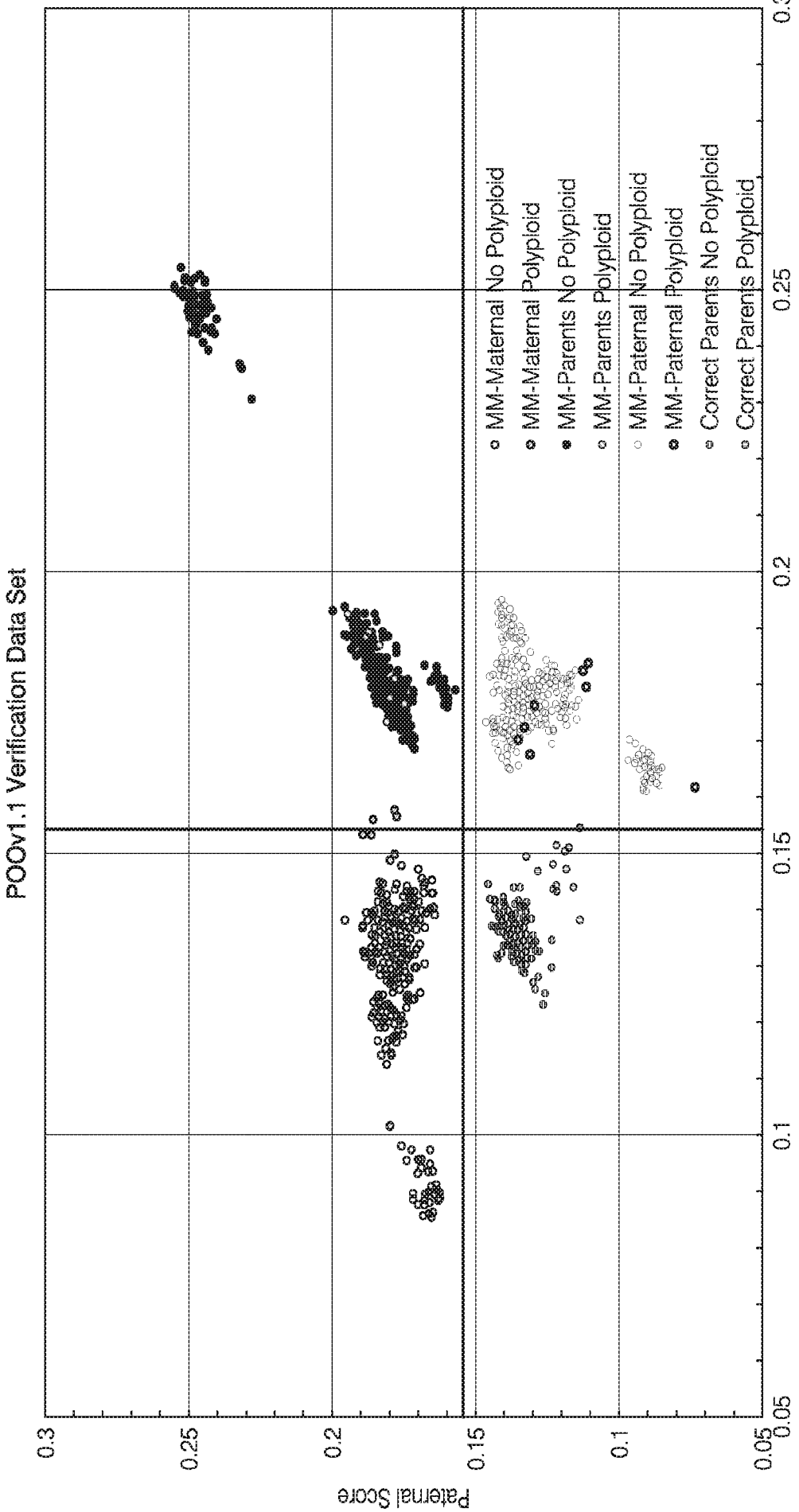

POOv1.1 Verification Data Set
Paternal Score
Maternal Score
0.3
0.25
0.2
0.15
0.1
0.05
MM-Maternal No Polyploid
MM-Maternal Polyploid
MM-Parents No Polyploid
MM-Parents Polyploid
MM-Paternal No Polyploid
MM-Paternal Polyploid
Correct Parents No Polyploid
Correct Parents Polyploid

FIG. 14

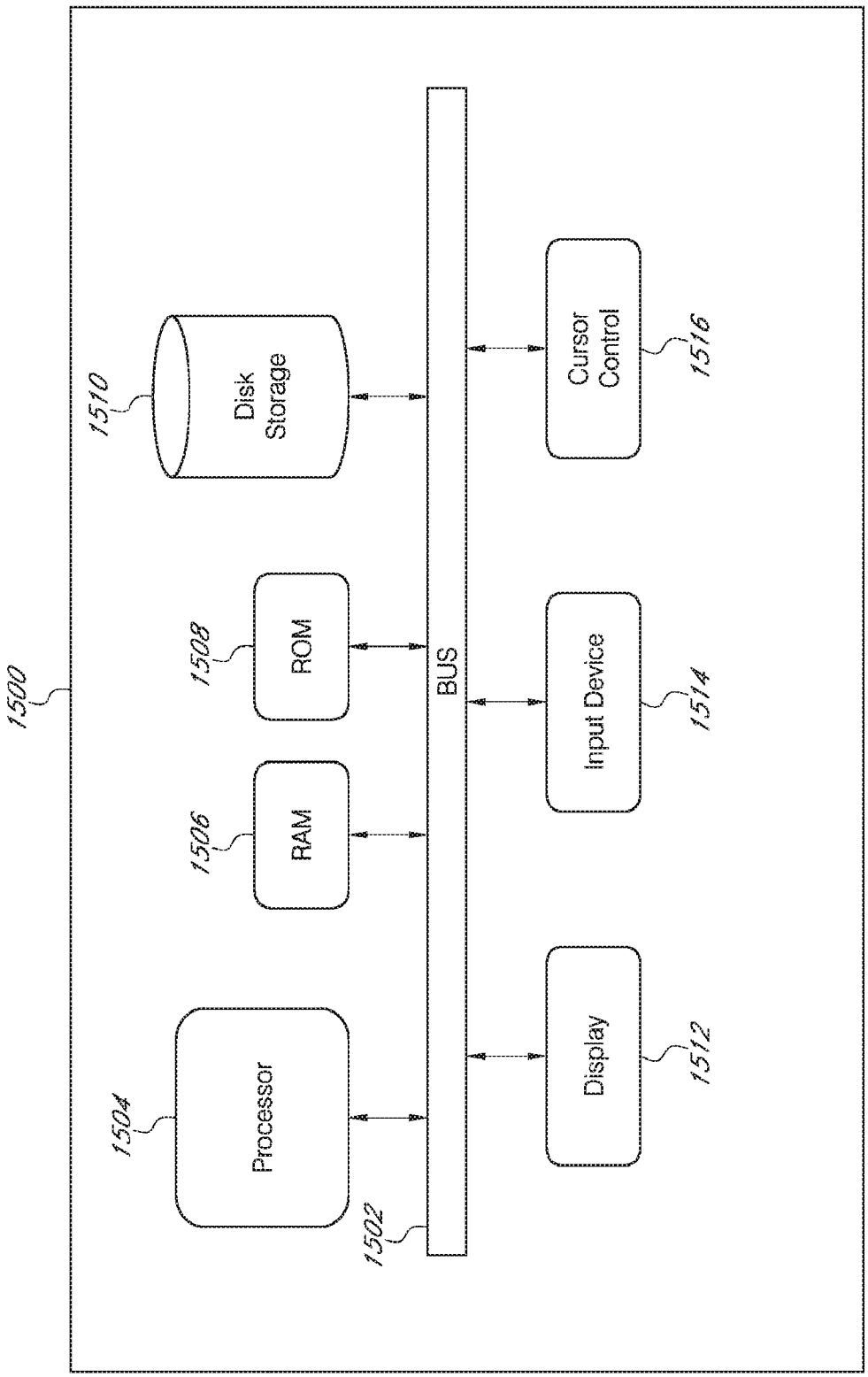

1500
1502
BUS
1504
Processor
1506
RAM
1508
ROM
1510
Disk Storage
1512
Display
1514
Input Device
1516
Cursor Control

FIG. 15

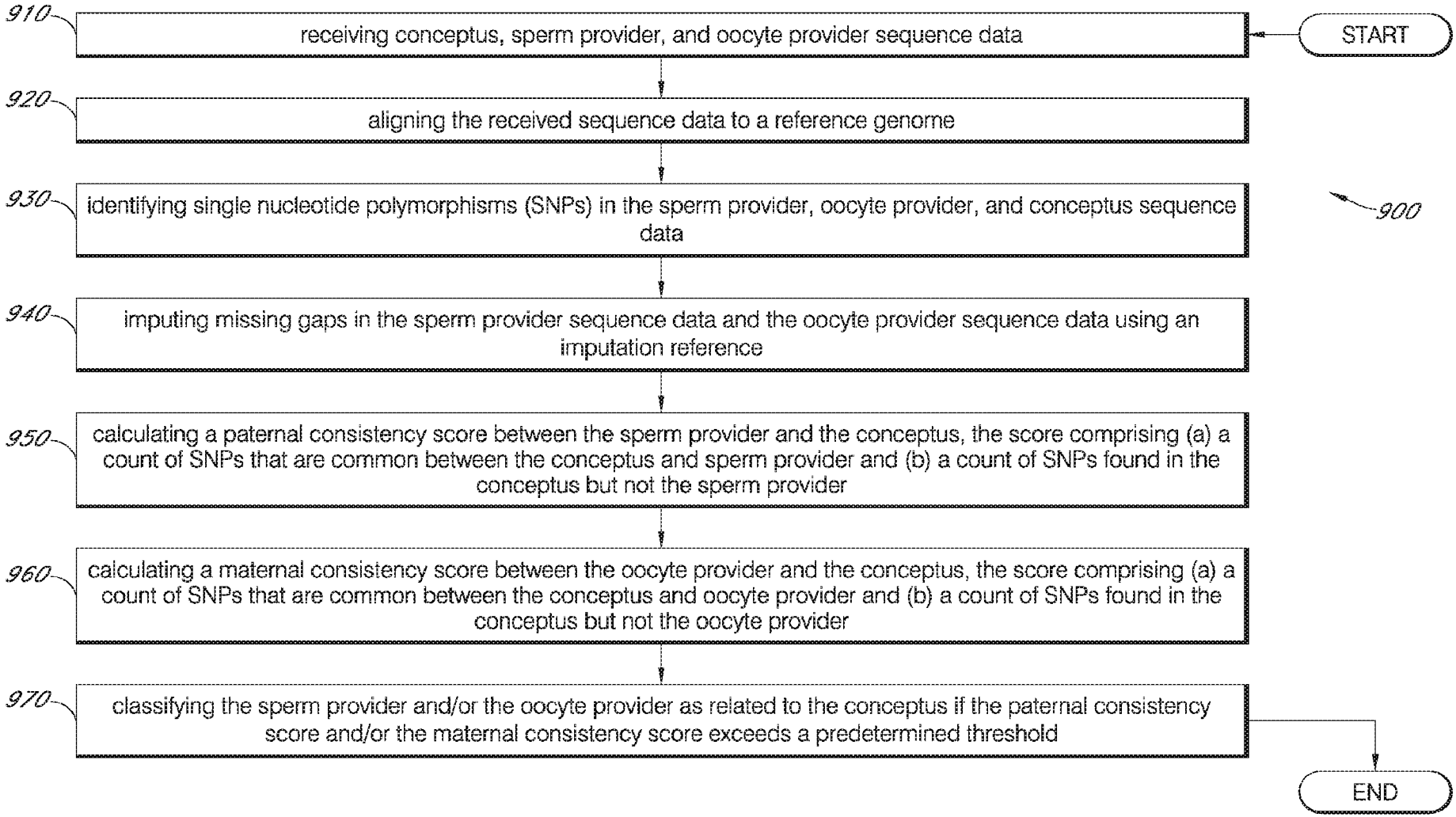

910
receiving conceptus, sperm provider, and oocyte provider sequence data
START
920
aligning the received sequence data to a reference genome
930
identifying single nucleotide polymorphisms (SNPs) in the sperm provider, oocyte provider, and conceptus sequence data
900
940
imputing missing gaps in the sperm provider sequence data and the oocyte provider sequence data using an imputation reference
950
calculating a paternal consistency score between the sperm provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and sperm provider and (b) a count of SNPs found in the conceptus but not the sperm provider
960
calculating a maternal consistency score between the oocyte provider and the conceptus, the score comprising (a) a count of SNPs that are common between the conceptus and oocyte provider and (b) a count of SNPs found in the conceptus but not the oocyte provider
970
classifying the sperm provider and/or the oocyte provider as related to the conceptus if the paternal consistency score and/or the maternal consistency score exceeds a predetermined threshold
END

FIG. 9