

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 May 2007 (24.05.2007)

PCT

(10) International Publication Number
WO 2007/057647 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/GB2006/004221

(22) International Filing Date:
13 November 2006 (13.11.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0523276.4 15 November 2005 (15.11.2005) GB

(71) Applicant (for all designated States except US): **LONDON BRIDGE FERTILITY, GYNAECOLOGY AND GENETICS CENTRE LTD** [GB/GB]; One St Thomas' Street, London, Greater London SE1 9RY (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **HANDYSIDE, Alan** [GB/GB]; One St Thomas' Street, London, Greater London SE1 9RY (GB).

(74) Agents: **KREMER, Simon** et al.; Mewburn Ellis LLP, York House, 23 Kingsway, London Greater London WC2B 6HP (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CHROMOSOMAL ANALYSIS BY MOLECULAR KARYOTYPING

(57) Abstract: The invention provides a method of karyotyping (for example for the detection of trisomy) a target cell to detect chromosomal imbalance therein, the method comprising: (a) interrogating closely adjacent biallelic SNPs across the chromosome of the target cell (b) comparing the result at (a) with the SNP haplotype of paternal and maternal chromosomes to assemble a notional haplotype of target cell chromosomes of paternal origin and of maternal origin (c) assessing the notional SNP haplotype of target cell chromosomes of paternal origin and of maternal origin to detect aneuploidy of the chromosome in the target cell. Also provided are related computer-implemented embodiments and systems.



WO 2007/057647 A1

CHROMOSOMAL ANALYSIS BY MOLECULAR KARYOTYPING**TECHNICAL FIELD**

5 The present invention relates generally to methods and materials for use in detecting abnormalities of the number of whole chromosomes or chromosome regions (aneuploidy). It has particular utility for prenatal diagnosis, either before pregnancy is established in gametes and cells taken from early embryos or later in pregnancy in samples of cells from the placenta or fetus.

10

BACKGROUND ART

In normal meiosis the precursor cells of the sperm or ova must multiply and then reduce the number of chromosomes to one half set in each gamete in two specialised meiotic divisions. During the early stages of meiosis following DNA replication, the four duplicated chromatids of a homologous pair align closely along their length and may exchange segments, resulting in non-recombinant (no exchange) and recombinant chromosomes and generating genetic variation. The resultant gametes, therefore, each contain a single chromosome which is either a recombinant of both homologous chromosomes or is non-recombinant and identical to one of the parental chromosomes. This is shown in Figure 1(a).

20

Aneuploidy is defined as an abnormal number of whole chromosomes or parts of chromosomes causing a genetic imbalance which may be lethal at early stages of development, cause miscarriage in later pregnancy or result in a viable but abnormal pregnancy. The most frequent and clinically significant aneuploidies involve single chromosomes (strictly 'aneusomy') in which there are either three ('trisomy') or only one ('monosomy') instead of the normal pair of chromosomes.

25

30 In early development, aneuploidy can arise through abnormal chromosomal segregation following replication and cell division either (1) during the two meiotic divisions which normally result in a haploid, half set, in each gamete before fertilisation, or (2) during the early divisions of the cleavage stage fertilised embryo. Figure 1(b) shows the effects of non-disjunction ie failure of replicated chromosomes to separate during division, which is a common cause of abnormal chromosome segregation, during the two meiotic divisions.

35

The aim of molecular karyotyping is to identify numerical or structural chromosomal abnormalities and in particular to identify any imbalance instead of the normal two copies of a chromosome or chromosome segment.

5

Currently there are two basic approaches to molecular karyotyping:

10

The first is to use molecular genetic markers, often highly polymorphic short tandem repeats (STRs), for each of the parental chromosomes. Where there is a different repeat on each of the parental chromosomes the STR marker is fully informative i.e. capable of identifying the presence of each chromosome (at that position). By use of a number of STR markers, the confidence in the method can be improved. An example of the use of STRs in trisomy analysis is given by Findlay *et al.* (1998) *Journal of Assisted Reproduction and Genetics* Vol 15, No 5: 1998: 266-275.

15

20

The second approach, comparative genomic hybridisation (CGH), involves fluorescent labelling of test and normal genomic DNA and comparison of quantitative differences in chromosome specific sequences by hybridising either to metaphase chromosomes or DNA clones on a microarray (array CGH). Generally test and reference DNA is labelled with different fluorochromes and hybridised to the DNA microarray with a series of cloned target DNAs, for example, BAC clones, derived from particular chromosomal regions. Such 'chips' are tailored to prenatal diagnosis and other diagnostic applications by including BACs informative, for example, for common deletion syndromes as well as aneuploidy and unbalanced translocations.

25

Given the importance of karyotyping, it can be seen that novel methods and materials relating to molecular karyotyping would provide a contribution to the art.

DISCLOSURE OF THE INVENTION

30

The present invention discloses that chromosomal analysis by molecular karyotyping, (for example for the detection of trisomy) can be performed by use of genome wide biallelic marker analysis (e.g. biallelic single nucleotide polymorphisms (SNPs)) which are distributed throughout the genome, and which can be readily detected using existing technologies.

35

This finding is unexpected for several reasons, principally because *a priori* it would be assumed that a biallelic marker (which provides only binary information at a given position on the chromosome) could not positively identify the presence of three or more different
5 chromosomes.

Nevertheless, as described below, the invention provides that high density analysis of closely adjacent SNPs is capable of positively identifying, *inter alia*, the presence of two
10 chromosomes derived from one parent and that based on well established assumptions about the frequency and spacing of recombination events between parental chromosomes during meiosis, this will allow accurate detection of trisomy.

Furthermore, the parental origin of the error is identified in each case which is not possible by some karyotyping methods.

15 The methods of the invention therefore do not depend on quantitation of chromosome specific sequences, as used in some currently available methods but rather compare the haplotypes of the test sample with the known haplotypes of the parents. When combined with existing methods for whole genome amplification, the methods of the invention are
20 particularly useful where only relatively small numbers of sample cells are available for analysis.

Thus in one aspect the invention provides a method of karyotyping a target cell to detect chromosomal imbalance therein, the method comprising:

- 25 (a) interrogating closely adjacent biallelic SNPs across the chromosome of the target cell
(b) comparing the result at (a) with the SNP haplotype of paternal and maternal chromosomes to assemble a notional haplotype of target cell chromosomes of paternal origin and of maternal origin
(c) assessing the notional SNP haplotype of target cell chromosomes of paternal origin and
30 of maternal origin to detect aneuploidy of the chromosome in the target cell

As set out below, the method can also be used to assess chromosomal recombination, where it is desired to do so.

35 The target cell will be one of a sexually reproducing, diploid, species, with a genome in

which biallelic SNPs occur at sufficient density to provide a notional haplotype. Preferably the cell will be an avian, reptilian, or mammalian cell. More preferably the cell is a human or non-human mammalian cell. The non-human mammal may, for instance be a primate.

5 In one embodiment the cell is human at least 2, 3, 4, 5, 6 or all of the human chromosomes selected from the following group are assessed: X, Y, 22, 21, 18, 16 and 13. Imbalances in any of these chromosomes may be associated with viable but abnormal pregnancies.

10 Preferably a total of at least 10, 15 or 20 chromosomes are assessed. In one embodiment the entire genome of the target cell (e.g. all 24 human chromosomes) is assessed.

As discussed below, SNPs can be interrogated using conventional techniques. This may be preceded by one or more conventional amplification steps. Preferably, the frequency of the less frequent allele of the biallelic markers in the present maps is at least 10, 20, or
15 30% (i.e. a heterozygosity rate of at least 0.18, 0.32 or 0.42).

The result of the SNP interrogation will depend on the nucleotides found at a polymorphic locus on all the copies of the given chromosome in the target cell (i.e. normally 2 copies, but may be 0, 1, or 3 where there are chromosomal abnormalities, or in the case of the X
20 chromosome, 1 or 3 copies in a female or 0 or 2 copies in a male, and in the case of the Y chromosome, 2 copies in a male embryo). Unless context demands otherwise, where "a" or "the" chromosome is referred to herein in respect of SNP genotyping, this refers to typing all the copies of that chromosome which are present in the target cell.

25 The next step, which is the assembly of the notional haplotype, will now be discussed in more detail.

Assembling a notional haplotype

30 To detect and characterise the presence of chromosomes of paternal or maternal origin the next step is to assemble a notional haplotype of each chromosome. It is termed 'notional' herein since it is inferred rather than determined directly, and in certain embodiments the notional haplotype of say, a given chromosome of paternal original, may be characterised in subsequent steps of the method as arising from two copies of that chromosome of paternal
35 original (in cases of trisomy, for example).

This notional haplotype may be assembled using particular sub-sets of SNPs as follows:

5 Assuming random SNP alleles for each parental chromosome, there are 16 different combinations of the four parental alleles for each SNP (Table 1).

10 Based on the haplotypes (i.e. the sequence of SNP alleles) of each parental chromosome, eight of these combinations can be predicted to result in genotypes in the test DNA that positively identify the presence of one out of the four parental chromosomes at that position ('informative') and four others will, dependent on results, either identify a pair of chromosomes one from each parent ('informative') or identify two possible combinations of parental chromosomes, out of the four possible pair-wise combinations ('semi-informative').

15 Therefore in one embodiment of the invention the notional SNP haplotype of the target cell chromosomes of paternal origin and of maternal origin is assembled using:

- (i) informative SNP alleles that positively identify which one of the four paternal and maternal chromosomes, a chromosome in the target cell has originated from, or positively identify which paternal chromosome and which maternal chromosome a pair of chromosomes in the target cell have originated from, and optionally
- 20 (ii) semi-informative SNP alleles that positively identify which of two possible combinations of the four possible pair-wise combinations of paternal and maternal chromosomes, a pair of chromosomes in the target cell has originated from.

Characterising target cell chromosomal origin

25

In one embodiment of the invention step (c) is performed as follows:

- (c1) assessing the notional SNP haplotype of target cell chromosomes of paternal origin and of maternal origin and thereby assigning each chromosome as absent, non-recombinant, recombinant, or present in 2 or more copies,
- 30 (c2) deducing aneuploidy of the chromosome in the target cell wherein step (c1) indicates an imbalance of chromosomes of paternal origin and of maternal origin,

35 Once the notional haplotype is compiled, the relevant chromosome may be assigned as non-recombinant or recombinant as follows:

Non-recombinant chromosome: the SNP alleles will be identical to one of the parental chromosomes along the whole length of the chromosome. Therefore wherever there is an informative combination of SNP genotypes in the parents for that particular chromosome, the results will be positive whereas, and equally significantly, at SNPs informative for the other chromosome the results will be negative. Semi-informative SNPs will give results consistent with the presence of that particular parental chromosome (see Figure 2(a)).

Thus in one embodiment a chromosome in the target cell is identified as non-recombinant wherein the results of its notional SNP haplotype are consistent with:

- (i) its SNP alleles being identical to the SNP alleles of one of the two paternal chromosomes or one of the two maternal chromosomes along the length of the chromosome, and
- (ii) an absence of the SNP alleles of the alternative of the two paternal or maternal chromosomes.

Recombinant chromosomes: with recombinant chromosomes, the pattern will be as for non-recombinant chromosomes except that there may be one or more alternating segments of both that parents chromosomes resulting from single, double or higher order recombination between the original parental chromosomes during the first meiotic division (see Figure 2 (a)).

In performing the present invention, consideration is given to the fact that where successive informative or semi-informative SNPs indicate a switch from identifying one parental chromosome to the other (an apparent crossover event) this could be because of (1) an actual crossover during meiosis (i.e. normal recombination, as referred to above), (2) the presence of a second parental chromosome (trisomy, as discussed hereinafter), or (3) a SNP genotyping error.

Considering SNP genotyping error, as genome wide SNP genotyping methods are designed to reduce errors to a minimum, in any given instance a crossover event will be the most likely alternative since there is normally at least one crossover per chromosome arm. A long succession of informative and semi-informative SNP results consistent with that chromosome and not the other parental chromosome would therefore suggest normal recombination (see Figure 2).

Therefore in one embodiment, a chromosome in the target cell is identified as recombinant wherein the results of its notional SNP haplotype correspond to SNP alleles of both of the two paternal chromosomes or two maternal chromosomes in one or more alternating segments consistent with normal recombination between the two chromosomes.

It is known that normal recombination will depend on (1) average, sex and chromosome specific data for the number of recombinations, and (2) interference between chiasmata preventing multiple recombination events over short distances.

Therefore in one embodiment, consistency with normal recombination is assessed based on the statistical likelihood of normal recombination between particular adjacent, informative SNP alleles of the two paternal chromosomes or two maternal chromosomes during the first meiotic division. Preferably the statistical likelihood is assessed based on one or more of the following criteria:

- (i) the average number of recombination events for the specific paternal or maternal chromosome,
- (ii) the position of the apparent recombination events on each chromosome arm relative to each other, the centromere and the telomere ie the ends of the chromosome arm involved

Multiple chromosomes: if successive informative and semi-informative SNPs alternate repeatedly and/or apparently randomly between the two parental chromosomes, it is highly likely that the test DNA is trisomic rather than a series of double crossover or recombination events (Figure 3; Figure 4).

This is because the pattern of normal recombination is non-random and specifically the presence of one crossover physically inhibits another crossover nearby, a phenomenon known as crossover interference (Broman and Weber, 2000).

With two non-recombinant chromosomes from one parent the SNP notional haplotype result will alternate all along the chromosome. With other combinations of non-recombinant and recombinant chromosomes, the two parental haplotypes will be detectable for a segment of the chromosome which shows the repetitive, apparently random, alternation.

In terms of the frequencies of 'normal' alternating segments, based on a large experimental

data set, Broman and Weber (2000) propose that the probability of a double crossover between two non-recombinant informative polymorphisms can be estimated according to the formula:

5 $p = (0.0114d - 0.0154)^4$

where p is the probability of a double crossover in an interval d measured as genetic distance in centiMorgans (cM) between non-recombinant loci.

10 The probability that one SNP, indicating the presence of the other parental chromosome (or >1 informative SNP with no intervening contradictory informative SNPs) is the result of a double crossover is therefore defined by the probability between adjacent flanking SNPs informative for that chromosome (Figure 2).

15 Thus in one embodiment the statistical likelihood of a double crossover between two SNP alleles is calculated according to the formula:

$p = (0.0114d - 0.0154)^4$

20 where p is the probability of a double crossover in an interval d measured as genetic distance in centiMorgans (cM) between the SNP alleles.

As an example of the operation of this formula, for an average spacing of 0.32cM (as with the Affymetrix GeneChip 10K system) and n SNPs:

25

N	d (cM)	p
10	3.2	$1.6 \cdot 10^{-7}$
50	16	$8.35 \cdot 10^{-4}$
100	32	0.015
200	64	0.254

30 Thus in most cases, the probability of a pattern alternating between the haplotypes of both chromosomes from one parent at successive informative and semi-informative SNPs will be very low particularly where the density of SNPs analysed is high and generally much lower than the possibility of genotyping error.

The probability of trisomy is further increased with the number and extent of this alternating pattern which will depend on the number and position of true crossover events on both of the chromosomes.

5

In addition to the number of apparent crossovers in the notional haplotypes, several other assumptions about the number and distribution of crossovers across the genome (Lynn et al, 2004) may be used in assigning the chromosome as likely recombinant or not:

- 10
- Direct counts of the number of chiasmata indicate that the average total number in males is 50.6 (Hulten, 1974) and in females 70 (Hulten and Tease, 2003; Tease and Hulten, 2004).
 - Average number of crossovers on individual chromosomes.
 - Distribution of crossovers on individual chromosomes.

15

Density and nature of SNPs

Across the whole genome when analysing, for example, 10K SNPs, despite the high accuracy rate, one or more random genotyping errors causing isolated individual positive results at informative SNPs for the second parental chromosome may occur.

20

Therefore in preferred embodiments a threshold number of positive and negative informative and semi-informative SNPs is set.

25 In one embodiment at least 5000 and/or 2500 informative and/or semi-informative SNP alleles, respectively, distributed across the whole genome are used to assemble the notional SNP haplotype of target cell chromosomes of paternal origin and of maternal origin. However for individual chromosomes a less number may be sufficient – this can be assessed by those skilled in the art according to the preferred method of typing and the accuracy associated with it and with any optional method of amplification employed.

30

In one embodiment the average distance between the interrogated SNPs is less than 0.1, 0.2, 0.3, 0.4 or 0.5 kb.

35 In one embodiment the average distance between the interrogated SNPs is less than 0.1,

0.2, 0.3, 0.4 or 0.5 cM

Because of allele dropout (ADO), ie the random failure to amplify one of the parental alleles, when amplifying the DNA from single or small numbers of cells for application in preimplantation genetic diagnosis, SNP genotype analysis may preferably be based in whole or in part on those results giving a heterozygous result at an informative or semi-informative SNP. Thus in one embodiment at least 2500 heterozygous informative SNP alleles ("AB" in Table 1) are used to assemble the notional SNP haplotype of target cell chromosomes of paternal origin and of maternal origin.

Karyotyping

As noted above, aneuploidy of the chromosome in the target cell is detected wherein the notional haplotype indicates an imbalance of chromosomes of paternal origin and of maternal origin. Details of the detection strategies of different aneuploidies are as follows:

In one embodiment where the notional SNP haplotype of target cell chromosomes indicates the presence of one chromosome of paternal origin and one chromosome of maternal origin and the cell is deduced to be normal diploid in respect of the relevant chromosome.

Nullsomy: In one embodiment, where the notional SNP haplotype of target cell chromosomes indicates an absence of any chromosome or chromosome segment of paternal origin and maternal origin, the cell is deduced to be nullsomic for the relevant chromosome or chromosome segment.

Monosomy: Here, there will be only one chromosome from one parent, but it can be either non-recombinant or recombinant. Monosomy will therefore be detected in two ways (1) apparent homozygosity (either 'AA' or 'BB' and not 'AB') for all SNPs along the chromosome, and (2) identity to the haplotype for one of the parental chromosomes (non-recombinant) or an alternating pattern between the two haplotypes from one parent (consistent with normal recombination between the two chromosomes). Thus in one embodiment, where the notional SNP haplotype of the target cell chromosomes indicates an absence of a chromosome or chromosome segment of either paternal origin or maternal origin but not both, the cell is deduced to be monosomic for the relevant chromosome or chromosome segment.

Monosomies may be detected whether they arise before or after fertilisation.

Trisomy: As discussed above, where the notional SNP haplotype of target cell
5 chromosomes indicates the presence of both of the two paternal chromosomes or two
maternal chromosomes in a pattern and/or frequency inconsistent with normal
recombination between the two chromosomes, the cell is deduced to be trisomic for all or
part of the relevant chromosome or chromosome segment.

10 Specifically the method is adapted to detect trisomy where paired chromosomes in each of
the paternal or maternal cells differ (which is commonly the case), and where the two
chromosomes of paternal or maternal origin in the target cell differ over all or part of the
chromosome (which would apply to the majority of trisomies i.e. most of those arising during
meiosis – see Figure 1(b)) and informative and semi-informative SNPs appear and are
15 interrogated in the regions which differ (which would typically apply where sufficient density
of SNPs are assessed).

It therefore provides a useful tool in detecting aneuploidy in these common situations.

Having described certain embodiments of the invention above, some particular modes of
20 operation will now be discussed.

Combined and multiple detection strategies

If desired, the invention may be combined with one or more other karyotyping strategies.

25 For example a further step may include quantitation of alleles to increase the accuracy and
resolution of trisomy detection i.e. in one embodiment the method further comprises
confirming the deduction by quantifying the SNPs across the chromosome of the target cell
(Meng et al., 2005).

30 In one embodiment the method further comprises diagnosing the presence of an inherited
genetic disease in the target cell by comparing the notional SNP haplotype of the target cell
with the SNP alleles of the paternal chromosomes and the maternal chromosomes and one
or more affected siblings to diagnose the disease in the target cell by linkage. Linkage is a
35 method in which instead of detecting a disease-causing gene mutation itself, one or more

informative markers such as STRs or SNPs, either close to or within the affected gene, are used to track the affected copy of the gene by comparison with the markers inherited by an affected child (Abou-Sleiman et al, 2002). With genome wide SNP analysis of the target cell genotype as discussed above, multiple closely linked SNPs flanking the affected gene may
5 be analysed permitting highly accurate linkage analysis.

In one embodiment the method further comprises diagnosing the presence of a susceptibility to a common disease or cancer in the target cell by comparing the notional SNP haplotype with a haplotype known to be associated with said disease. Such
10 associations are being increasingly established, for example via the "International HapMap Consortium" which is mapping genome wide variation in SNP haplotypes in the human population, is to facilitate disease association studies (International HapMap Consortium, 2005). The associations do not *per se* form part of the invention, but the combination of this haplotype analysis with the karyotyping method described herein forms one aspect of
15 the invention.

Paternal and maternal cells and chromosomal haplotypes

In one embodiment paternal and maternal cells are provided from blood or buccal cavity
20

Analysis of SNPs from related individuals across at least one generation allows the identification of a haplotype for each chromosome in positions where the alleles are different. Specifically, the haplotype of SNP alleles can be ascertained by analysing the DNA of each parent and comparing the results with a haploid gamete, child or parent or a
25 combination of these. Those skilled in the art are aware of numerous algorithms and related software programmes that allow the haplotypes to be inferred from analysis of diploid individuals e.g. PHASE (Stephens and Donnelly, 2003) and SIMHAP (www.genepi.com.au/simhap).

30 In one embodiment SNP haplotype of paternal and maternal chromosomes is derived from analysis of the SNP haplotype of cells removed from sibling fertilized embryos following in vitro fertilisation (IVF) following whole genome amplification.

In one embodiment SNP haplotype of paternal and maternal chromosomes is derived from
35 analysis of multiple single parental haploid gametes following whole genome amplification.

Where two chromosomes or chromosome segments from one parent are shown to be identical, the method will not be applicable for that chromosome or segment and alternative methods should be used.

5

Target cells

In one embodiment the target cell has been provided from a mammalian embryo which has resulted from IVF. In one embodiment the embryo is a pre-implantation embryo (see e.g. Handyside et al, 2004).

10

Where the invention is applied to animals such as livestock, the embryo may be recovered from the uterus.

15

In one embodiment the target cell(s) have been provided from a fetus

In one embodiment a number equal to, or at least, 1, 2, 3, 4, or 5 cells are provided

20

In one embodiment SNP interrogation is preceded by whole genome amplification

In one embodiment the whole genome amplification employs isothermal Multiple Displacement Amplification (MDA) which permits whole genome amplification using the bacteriophage phi29 polymerase for amplification from small numbers of cells (see e.g. Handyside et al, 2004).

25

Interrogation of SNPs

Preferred markers are biallelic SNPs, which occur throughout the genome (~10 million per genome, wherein the SNP is defined as >1% variation between individuals in a population).

30

Various methods for large scale single nucleotide polymorphism (SNP) analysis exist (see Syvanen, 2005, especially Table 1). These include SNPstream (Bell, P.A. *et al.*

SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques Suppl.*, 70-72, 74, 76-77 (2002)); Genorama, APEX (Kurg, A. *et*

35

al. Arrayed primer extension: solid-phase four-colour DNA resequencing and mutation

detection technology. *Genet. Test.* 4, 1-7 (2000)); GeneChip 100K (Matsuzaki, H. *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat. Methods* 1, 109-111 (2004)); Perlegen wafers (Hinds, D.A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072-1079 (2005)); Molecular Inversion Probes (Hardenbol, P. *et al.* Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 15, 269-275 (2005)); GoldenGate Assay (Fan, J.B. *et al.* Highly parallel SNP genotyping. *Cold Spring Harb. Symp. On Quant. Biol.* LXVII, 69-78 (2003)). Other methods include the Illumina "BeadArray".

A preferred embodiment employs the Affymetrix GeneChip™ 10K Microarray is designed to analyse 10,000 distributed at an average distance of 0.2Kb across each of 22 chromosomes (see Matsuzaki, H. *et al.* Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res.* 14, 414-425 (2004))

In the case of oligonucleotide chips, the oligonucleotides that can be bonded to a chip according to the invention will be capable of distinguishing biallelic SNPs across the genome. Preferred are 25 nucleotide-long oligonucleotides.

Thus in one embodiment the SNPs are interrogated on a "gene" or "oligonucleotide" chip or microarray. As is well known in the art these are miniaturized vehicles, in most cases made of glass or silicon, on whose surface oligonucleotides of known sequence are immobilized in an ordered grid of high density.

Another preferred embodiment employs the Illumina's "infinium"™ Human-1 BeadChip. This system may enable whole-genome genotyping of over 100,000 SNP markers, 70% of which are located in exons or within 10kb of transcripts (see e.g. *Pharmacogenomics* (2005) 6(7), 777-782). The system is based on the random assembly of derivatized microscopic beads approximately 3 µm in size) into wells of a patterned substrate, and may permit specified combinations of SNPs to be interrogated.

Systems

Preferably a system for use in the present invention would comprises means for SNP

interrogation plus a programmed storage device or medium for causing a computer to analyse the resulting data. The SNP interrogation data could be stored for later analysis or analysed 'on the fly' – as used herein the term "database" covers both types of data source.

5

Preferred means for SNP interrogation would be an oligonucleotide chip which would interrogate at least the preferred chromosomes at the appropriate density discussed above. Preferably it would include the whole genome.

10 Thus preferred means for SNP interrogation would include:

(i) A high density of biallelic SNPs on chromosomes frequently associated with miscarriage or viable abnormal pregnancies (X, Y, 22, 21, 18, 16, 13). The means may interrogate a full polymorphic set on these chromosomes.

15

(ii) SNPs which are highly heterozygous in the general population increasing their informativeness,

(iii) Relatively increased density in all the known microdeletion syndrome regions,

(iv) Relatively increased density in regions associated with common single gene defects,

(v) Known SNP 'haplotags' associated with predisposition to common complex diseases.

20 The present invention may be implemented with a computer. Typically this would include a central processing unit (CPU) connected by a system bus or other connecting means to a communication interface, system memory (RAM), non-volatile memory (ROM), and one or more other storage devices such as a hard disk drive, a diskette drive, and a CD ROM drive.

25

The computer also includes a display device, such as a printer, CRT monitor or an LCD display, and an input device, such as a keyboard, mouse, pen, touch-screen, or voice activation system. The input device may receive data directly from the means for SNP interrogation via an interface (as for example with the Affymetrix system).

30

The computer stores and executes various programs such as an operating system and application programs.

35

The computer-usable medium would cause the computer to analyse haplotypes and perform molecular karyotyping to assign parental origin along the length of each

chromosome, and report on aneuploidy where this was detected. The medium may for example be selected from the group consisting of a hard disk a floppy disk, Random Access Memory, Read Only Memory and Electrically Erasable Programmable Read Only Memory.

5

Thus the invention provides a computer-usable medium having computer-readable program code or instructions stored thereon (i.e. a programmed storage device) for causing a computer to execute a method to determine aneuploidy or chromosomal recombination in a target cell, the method being any one of those discussed herein.

10

Preferably the method comprises:

- (a) accessing a database comprising genotype data obtained from a plurality of closely adjacent biallelic SNP loci present in a chromosome of the target cell,
- (b) accessing a database comprising SNP haplotype data of the corresponding paternal and maternal chromosomes (i.e. 'P1', 'P2', 'M1', 'M2'),
- (c) comparing target cell SNP data from the database of step (a) with SNP haplotype data from the database of step (b) to assemble a notional haplotype of regions of the target cell chromosomes of paternal origin and of maternal origin,
- (d) assessing the notional SNP haplotype of target cell chromosomes of paternal origin and of maternal origin to detect aneuploidy or chromosomal recombination of the chromosome in the target cell.

15
20

Optionally, each SNP locus of the 'x' SNPs of the database in step (b) is assigned a value 'n' in accordance with which of the 16 combinations of four parental SNP alleles is present at that locus, and wherein step (c) comprises assembling a notional haplotype at that locus by comparing

25

- (i) the genotype data for the biallelic SNP at that locus from the database of step (a) and,
- (ii) the value 'n' at that locus from the database of step (b) with,
- (iii) a chromosomal origin table,

30

and thereby assigning the SNP locus of the target cell chromosomes as originating from a paternal or maternal chromosome.

By 'chromosomal origin table' is meant a reference set of data by which the chromosomal origin (e.g. 'P1', 'P2', 'M1', or 'M2') can be assigned at that locus based on the values at (i) and (ii). This may correspond to that given in Table 1, last column.

35

Preferably the notional SNP haplotype of regions of the target cell chromosomes of paternal origin and of maternal origin is assembled using a subset of the 'x' SNP loci from the database in step (b), which subset consists of:

- 5 (i) informative SNP alleles that positively identify which one of the four paternal and maternal chromosomes, a chromosome in the target cell has originated from, or positively identify which paternal chromosome and which maternal chromosome a pair of chromosomes in the target cell have originated from, and optionally
- 10 (ii) semi-informative SNP alleles that positively identify which of two possible combinations of the four possible pair-wise combinations of paternal and maternal chromosomes, a pair of chromosomes in the target cell has originated from.

Optionally the subset may consist of heterozygous informative SNP alleles.

- 15 Optionally the method may comprises storing the notional haplotype result obtained for each SNP locus of the 'x' SNPs, or a subset thereof.

20 An example of software (Excel Visual Basic for Applications (VBA) code I) is listed in Appendix 1, and this identifies the different combinations of parental alleles at each SNP and assigns parental origin at informative SNP loci. Figure 4 shows resulting notional haplotypes and the actual target cell haplotypes as determined independently.

The program may identify the chromosome in the target cell as non-recombinant wherein the results of its notional SNP haplotype are consistent with:

- 25 (i) its SNP alleles being identical to the SNP alleles of one of the two paternal chromosomes or one of the two maternal chromosomes along the length of the chromosome, and
- (ii) an absence of the SNP alleles of the alternative of the two paternal or maternal chromosomes.

30 The program may identify the chromosome in the target cell as recombinant wherein the results of its notional SNP haplotype correspond to SNP alleles of both of the two paternal chromosomes or two maternal chromosomes in one or more alternating segments consistent with normal recombination between the two chromosomes.

35

The program may identify the chromosome in the target cell as trisomic for the chromosome where the notional SNP haplotype of the target cell chromosome indicates the presence of both of the two paternal chromosomes or two maternal chromosomes in a pattern and/or frequency inconsistent with normal recombination between the two
5 chromosomes

The program may statistically analyze the likelihood of normal recombination between the SNP loci based on one or more of the following criteria:

- 10 (i) a database the average number of recombination events for the specific paternal or maternal chromosome,
- (ii) the position of the apparent recombination events on each chromosome arm relative to each other, the centromere and the telomere

Optionally the program may calculate a numerical measure of probability of, for example,
15 trisomy based on this frequency and pattern data.

The program may identify the chromosome in the target cell as nullsomic for the chromosome where the notional SNP haplotype of target cell chromosome indicates an absence of the chromosome or a segment thereof of both paternal origin and maternal
20 origin.

The program may identify the chromosome in the target cell as monosomic for the chromosome where the notional SNP haplotype of the target cell chromosome indicates an absence of the chromosome or a segment thereof of paternal origin and maternal origin but
25 not both.

Optionally a threshold number of positive and negative informative and optionally semi-informative SNPs is set, and a karyotype is determined only when this number is exceeded.

30 The invention also provides a computer programmed to execute a method as described above.

The invention will now be further described with reference to the following non-limiting Figures and Examples. Other embodiments of the invention will occur to those skilled in
35 the art in the light of these.

The disclosure of all references cited herein, inasmuch as it may be used by those skilled in the art to carry out the invention, is hereby specifically incorporated herein by cross-reference.

5

FIGURES

Figure 1(a): schematic showing normal meiosis I and II

10 Figure 1(b): schematic showing non-disjunction during meiosis leading to aneuploidy

Figure 2: schematic representation of SNP genotype analysis for a chromosome pair in test DNA from a fetus or embryo. Each row represents a SNP locus and each column, the haplotypes for the two paternal chromosomes (P1 and P2) and two maternal chromosomes (M1 and M2). A positive result at an informative SNP is shaded. A negative result at an informative SNP is marked with ?PX (?MX omitted for clarity). In Figure 2(a) the segment marked A of the paternal copy of the chromosome is identified as a double recombinant because multiple consecutive informative SNPs positively identify the haplotype of P1 and others informative for P2 are negative. The length of the chromosomal segment and number of crossovers both for this chromosome and overall would also be taken into account and because of crossover interference would normally extend over a greater number of SNPs than in this diagrammatic representation. In Figure 2(b) The maternal chromosome is non-recombinant in this region and only SNPs informative for M1 are positive.

25

Figure 3: Schematic representation of SNP genotype analysis for a chromosome pair in test DNA from a fetus or embryo as described in Figure 2. In this example, the segment marked B is assigned as trisomic because of the positive result for multiple alternating SNPs informative for both chromosomes. The probability of each individual positive result for P1 being a double recombinant between flanking positive results for P2 (marked A) is very small given the average genetic distance between SNPs when using high density SNP analysis.

30

Figure 4: Parental and test offspring genotype data analysed using the Excel VBA code in Appendix 1 illustrating how informative combinations of biallelic SNPs allow identification

35

the parental origin along a chromosome.

Figure 5: Flowchart illustrating the present invention with multiple detection strategies.

5 **EXAMPLES**

Example 1- Theoretical background

10 During meiosis and the formation of gametes, homologous chromosomes pair and recombine resulting in four chromosomes, which on average will consist of two non-recombinant and two recombinant chromosomes. Each of the resulting chromosomes therefore now has a unique SNP haplotype.

15 Following fertilisation, each embryo has a unique combination of haplotypes from the non-recombinant or recombinant chromosomes segregated to the two gametes from the two parents. In euploid embryos, with the normal pairs of each chromosome, each chromosome will have a distinct haplotype and the parental origin of each chromosome will be identifiable from the non-recombinant or unique recombinant haplotype. Similarly, trisomy and monosomy will also be detectable.

20 Table 1 below shows how SNPs can be classified as informative, semi-informative, or non-informative.

25 Table 1 The 16 combinations of parental SNP haplotypes based on a random distribution of alleles. Informative combinations identify a parental haplotype irrespective of the result. Informative/semi-informative combinations of alleles identify one of both parents chromosomes or two possible combinations depending on the genotype of the DNA being tested.

Informativity	#	P1	P2	M1	M2	Test genotype
Non-informative	1	A	A	A	A	AA
	2	B	B	B	B	BB

	3	A	A	B	B	AB
	4	B	B	A	A	AB
Informative (all results identify presence of 1 parental chromosome)	5	A	B	B	B	AB=P1 BB=P2
	6	B	A	B	B	AB=P2 BB=P1
	7	B	B	A	B	AB=M1 BB=M2
	8	B	B	B	A	AB=M2 BB=M1
	9	B	A	A	A	AB=P1 AA=P2
	10	A	B	A	A	AB=P2 AA=P1
	11	A	A	B	A	AB=M1 AA=M2
	12	A	A	A	B	AB=M2 AA=M1
Informative/ semi- informative (2/3 possible results identify a pair of parental chromosomes and 1/3 results could be either of two combinations)	13	A	B	A	B	AA=P1M1 BB=P2M2 AB=P1M2 or P2M1
	14	B	A	B	A	AA=P2M2 BB=P1M1 AB=P2M1 or P1M2
	15	A	B	B	A	AA=P1M2 BB=P2M1 AB=P1M1 or P2M2
	16	B	A	A	B	AA=P2M1 BB=P1M2 AB=P2M2 or P1M1

Example 2 – Informativity of SNPs and ADO

5 In some embodiments when DNA is amplified from single cells, for example, for preimplantation genetic diagnosis (PGD), one of the parental alleles may fail to amplify at random resulting in allele dropout (ADO). Table 2 below demonstrates that where ADO occurs at informative SNPs, half of these will be detected because the apparent test genotype is not possible and therefore the true heterozygous result ("AB") can be inferred.

10

Table 2: Effect of allele dropout (ADO) at informative SNPs

Maternal (M) chr		Paternal (P) chr		Test genotype and chromosome identified
1	2	1	2	
B	A	A	A	AA = M2 AB = M1 (BB = AB*)
A	B	A	A	AA = M1 AB = M2 (BB = AB*)
A	A	B	A	AA = P2 AB = P1 (BB = AB*)
A	A	A	B	AA = P1 AB = P2 (BB = AB*)
A	B	B	B	BB = M2 AB = M1 (AA = AB*)
B	A	B	B	BB = M1 AB = M2 (AA = AB*)
B	B	A	B	BB = P2 AB = P1 (AA = AB*)
B	B	B	A	BB = P1 AB = P2 (AA = AB*)

15 *For this combination of parental SNPs, the test genotype cannot have two copies of this allele indicating allele dropout (ADO) i.e. failure to amplify one of the parental alleles at random. The test genotype can therefore be assumed to be AB. This approach increases the power of the test when the invention is used in single cell applications such as preimplantation genetic screening.

20

Example 3 - Combined SNP quantitative and sequence based analysis

If relative *quantitation* of each SNP allele is achieved, the normal disomic genotype combinations, "AA", "AB" and "BB", are supplemented by "A" and "B" in monosomy and

“AAA”, “BBB”, “AAB” and “ABB”.

Table 3 demonstrates the extra information that is available by combining genotyping and quantitation of SNPs. While these possible combination of SNP alleles genotyping may be uninformative, quantitation would identify the chromosome as trisomic even though the parental origin is unknown in the first two examples.

Table 3 Combined genotyping and quantitation of SNPs

Maternal (M) chr		Paternal (P) chr		Test genotype and chromosome identified
1	2	1	2	
A	A	A	A	AAA = Trisomic M1 M2 or P1 P2
B	B	B	B	BBB = Trisomic M1 M2 or P1 P2
A	A	B	B	AAB = M1 M2 P1 or P2 ABB = M1 or M2 P1 P2
B	B	A	A	AAB = M1 or M2 P1 P2 BBA = M1 M2 P1 or P2
A	B	A	B	AAB = M1 P1 + M2 or P2 ABB = M2 P2 + M1 or P1
B	A	A	B	AAB = M2 P1 + M1 or P2 ABB = M1 P2 + M2 or P1
A	B	B	A	AAB = M1 P2 + M2 or P1 BBA = M2 P1 + M1 or P2
B	A	B	A	AAB = M2 P2 + M1 or P1 BBA = M1 P1 + M2 or P2

10

Example 4 – use of multiple displacement amplification (MDA)

A Microsoft Excel VBA macro (SNP analysis (1)) to analyse these combinations of SNPs and test genotypes is set out below in Appendix 1.

15

The results are shown in Figure 4. The data is test genotype analysis of SNPs across chromosome 21. Maternal and paternal haplotypes have been created based on an individual genotyped using the Affymetrix GeneChip 10K microarray. At some SNPs the

alleles could not be reliably identified ("No call"). Otherwise, for illustration purposes, the test genotype is assumed to be 100% accurate and is based on the actual test haplotypes shown. The numbered parental haplotype combinations are as defined in Table 1. In this example only fully informative SNPs combinations (5-12) have been used to identify the test haplotypes. In addition, semi-informative combinations would be used and the information further analysed to distinguish double recombination from trisomy or genotyping errors. Two examples are shown - one normal disomic and one trisomic for chromosome 21. The black circles indicate the critical SNP results that indicate this test sample is trisomic i.e. they demonstrate the presence of alternating segments which are not consistent with normal recombination between the two chromosomes.

Example 5 - Prenatal diagnosis following CVS or amniocentesis

Current methods of analysis for chromosomes include karyotyping, fluorescent in situ hybridisation (FISH) for a restricted number of chromosome pairs or quantitative fluorescent PCR. SNP profiling according to the present invention combines detection of aneuploidy, deletions and unbalanced translocations.

Current methods for detection of single gene defects combine mutation detection (requiring identification of the mutation) and informative linked markers in some cases requiring prior linkage analysis and test development. Because of the density of SNPs analysable linkage to combinations of SNPs will also be possible in many cases.

Example 6 - Preimplantation genetic diagnosis following IVF

Preimplantation genetic diagnosis (PGD) requires analysis of single or small numbers of cells removed from each early embryo and, if possible, within 36-72h, so that embryos identified as unaffected can be transferred without cryopreservation.

Current methods for aneuploidy include sequential FISH for analysis of 9 chromosomes, comparative genomic hybridisation of all chromosomes requiring cryopreservation and multiplex fluorescent PCR. Each reciprocal translocation and each type of Robertsonian translocation require the development of a specific strategy. Current methods for single gene detection include mutation detection and linkage analysis by multiplex fluorescent PCR.

	Advantages
Prenatal diagnosis	<ul style="list-style-type: none"> • Universal screen for aneuploidy (possible exceptions) • High resolution detection of unbalanced translocations and common deletions
Preimplantation genetic diagnosis	<ul style="list-style-type: none"> • Universal screen for aneuploidy (including translocations) combined with single gene defect linkage detection • No requirement for test development • Identifies parental origin of aneuploidy

Table 4 Advantages of the present invention in prenatal diagnosis and preimplantation genetic diagnosis.

5

References

Handyside et al (2004) Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease. Mol Hum Reprod 10, 767-772.

10

Syvanen, AC (2005) Toward genome wide SNP genotyping. Nature Genetics 37, S5-S10.

15

Broman, K.W. and Weber, J.L. (2000) Characterisation of human crossover interference. Am J Hum Genet 66, 1911-1926.

Hulten, maternal (1974) Chiasma distribution at diakinesis in the normal human male. Hereditas 76, 55-78.

Hulten MA and Tease C (2003) Genetic maps: direct meiotic analysis. In: Cooper DN (ed) Encyclopaedia of the Human Genome Nature Publishing Group, London

- 5 Lynn, A, Ashley, T and Hassold, T (2004) Variation in human meiotic recombination. *Annu Rev Genomics Hum Genet* 5, 317-349.

Tease C and Hulten MA (2004) Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. *Cytogenet*
10 *Genome Res* 107, 208-215.

Meng H, Hager K and Gruen JR (2005) Detection of Turner syndrome using high-throughput quantitative genotyping. *J Clin Endocrinol Metab* 90, 3419-3422.

- 15 International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437, 1299-1320.

Abou-Sleiman PM, Apessos A, Harper JC, Serhal P, Winston RM and Delhanty JD (2002) First application of preimplantation genetic diagnosis to neurofibromatosis type 2 (NF2).
20 *Prenatal Diagnosis* 22, 519-524.

Stephens M and Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73, 1162-1169.

Appendix 1

Microsoft Excel VBA macro: SNP analysis (1)

5

```
Private Sub CommandButton1_Click()
```

```
For i = 1 To X (no. of SNPs analysed)
```

```
10 If Cells(i, 1) = "A" And Cells(i, 2) = "A" And Cells(i, 3) = "A" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 1
```

```
If Cells(i, 1) = "B" And Cells(i, 2) = "B" And Cells(i, 3) = "B" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 2
```

```
If Cells(i, 1) = "A" And Cells(i, 2) = "A" And Cells(i, 3) = "B" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 3
```

```
If Cells(i, 1) = "B" And Cells(i, 2) = "B" And Cells(i, 3) = "A" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 4
```

```
If Cells(i, 1) = "A" And Cells(i, 2) = "B" And Cells(i, 3) = "A" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 5
```

```
15 If Cells(i, 1) = "A" And Cells(i, 2) = "B" And Cells(i, 3) = "B" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 6
```

```
If Cells(i, 1) = "B" And Cells(i, 2) = "A" And Cells(i, 3) = "A" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 7
```

```
If Cells(i, 1) = "B" And Cells(i, 2) = "A" And Cells(i, 3) = "B" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 8
```

```
If Cells(i, 1) = "A" And Cells(i, 2) = "A" And Cells(i, 3) = "A" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 9
```

```
If Cells(i, 1) = "B" And Cells(i, 2) = "B" And Cells(i, 3) = "A" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 10
```

```
20 If Cells(i, 1) = "A" And Cells(i, 2) = "A" And Cells(i, 3) = "B" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 11
```

```
If Cells(i, 1) = "B" And Cells(i, 2) = "B" And Cells(i, 3) = "B" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 12
```

```
If Cells(i, 1) = "A" And Cells(i, 2) = "B" And Cells(i, 3) = "A" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 13
If Cells(i, 1) = "B" And Cells(i, 2) = "A" And Cells(i, 3) = "A" And Cells(i, 4) = "B" Then Cells(i, 5).Value = 14
If Cells(i, 1) = "A" And Cells(i, 2) = "B" And Cells(i, 3) = "B" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 15
If Cells(i, 1) = "B" And Cells(i, 2) = "A" And Cells(i, 3) = "B" And Cells(i, 4) = "A" Then Cells(i, 5).Value = 16
```

```
5 Next i
```

```
For i = 1 To X
```

```
If Cells(i, 5) = 5 And Cells(i, 6) = "AB" Then Cells(i, 9).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 5 And Cells(i, 6) = "BB" Then Cells(i, 9).Interior.Color = RGB(200, 0, 0)
```

```
10 If Cells(i, 5) = 5 And Cells(i, 6) = "AA" Then Cells(i, 8).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 6 And Cells(i, 6) = "AB" Then Cells(i, 8).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 6 And Cells(i, 6) = "AA" Then Cells(i, 8).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 6 And Cells(i, 6) = "BB" Then Cells(i, 9).Interior.Color = RGB(200, 0, 0)
```

```
15
```

```
If Cells(i, 5) = 7 And Cells(i, 6) = "AB" Then Cells(i, 8).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 7 And Cells(i, 6) = "BB" Then Cells(i, 8).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 7 And Cells(i, 6) = "AA" Then Cells(i, 9).Interior.Color = RGB(200, 0, 0)
```

```
20 If Cells(i, 5) = 8 And Cells(i, 6) = "AB" Then Cells(i, 9).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 8 And Cells(i, 6) = "AA" Then Cells(i, 9).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 8 And Cells(i, 6) = "BB" Then Cells(i, 8).Interior.Color = RGB(200, 0, 0)
```

```
If Cells(i, 5) = 9 And Cells(i, 6) = "AB" Then Cells(i, 11).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 9 And Cells(i, 6) = "BB" Then Cells(i, 11).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 9 And Cells(i, 6) = "AA" Then Cells(i, 10).Interior.Color = RGB(0, 0, 200)

If Cells(i, 5) = 10 And Cells(i, 6) = "AB" Then Cells(i, 10).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 10 And Cells(i, 6) = "AA" Then Cells(i, 10).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 10 And Cells(i, 6) = "BB" Then Cells(i, 11).Interior.Color = RGB(0, 0, 200)

10 If Cells(i, 5) = 11 And Cells(i, 6) = "AB" Then Cells(i, 10).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 11 And Cells(i, 6) = "BB" Then Cells(i, 10).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 11 And Cells(i, 6) = "AA" Then Cells(i, 11).Interior.Color = RGB(0, 0, 200)

15 If Cells(i, 5) = 12 And Cells(i, 6) = "AB" Then Cells(i, 11).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 12 And Cells(i, 6) = "AA" Then Cells(i, 11).Interior.Color = RGB(0, 0, 200)
If Cells(i, 5) = 12 And Cells(i, 6) = "BB" Then Cells(i, 10).Interior.Color = RGB(0, 0, 200)
```

Next i

End Sub

Claims

1. A method of karyotyping a target cell to detect chromosomal imbalance therein, the
5 method comprising:
(a) interrogating closely adjacent biallelic SNPs across the chromosome of the target cell
(b) comparing the result at (a) with the SNP haplotype of paternal and maternal
chromosomes to assemble a notional haplotype of target cell chromosomes of paternal
origin and of maternal origin
10 (c) assessing the notional SNP haplotype of target cell chromosomes of paternal origin and
of maternal origin to detect aneuploidy of the chromosome in the target cell.
2. A method as claimed in claim 1 wherein the target cell is a human cell.
- 15 3. A method as claimed in claim 2 wherein:
(i) at least 2, 3, 4, 5, 6, or 7 of the chromosomes selected from the group consisting of: X,
Y, 22, 21, 18, 16 and 13 are interrogated, or
(ii) all 24 chromosomes are interrogated.
- 20 4. A method as claimed in any one of the preceding claims wherein the notional SNP
haplotype of the target cell chromosomes of paternal origin and of maternal origin is
assembled in step (b) using:
(i) informative SNP alleles that positively identify which one of the four paternal and
maternal chromosomes, a chromosome in the target cell has originated from, or positively
25 identify which paternal chromosome and which maternal chromosome a pair of
chromosomes in the target cell have originated from, and optionally
(ii) semi-informative SNP alleles that positively identify which of two possible combinations
of the four possible pair-wise combinations of paternal and maternal chromosomes, a pair
of chromosomes in the target cell has originated from.
30
5. A method as claimed in claim 4 wherein a threshold number of positive and
negative informative and semi-informative SNPs are interrogated.
6. A method as claimed in claim 4 or claim 5 wherein a threshold number of
35 heterozygous informative SNP alleles are interrogated.

7. A method as claimed in any one of the preceding claims wherein aneuploidy of the chromosome in the target cell is detected in step (c) by:

(c1) assessing the notional SNP haplotype of target cell chromosomes of paternal origin and of maternal origin and thereby assigning each chromosome as recombinant or non-recombinant and/or present in 0, 1, or more copies,

(c2) deducing aneuploidy of the chromosome in the target cell wherein step (c1) indicates an imbalance of chromosomes of paternal origin and of maternal origin.

8. A method as claimed in any one of the preceding claims wherein a chromosome in the target cell is identified as non-recombinant wherein the results of its notional SNP haplotype are consistent with:

(i) its SNP alleles being identical to the SNP alleles of one of the two paternal chromosomes or one of the two maternal chromosomes along the length of the chromosome, and

(ii) an absence of the SNP alleles of the alternative of the two paternal or maternal chromosomes.

9. A method as claimed in any one of the preceding claims wherein a chromosome in the target cell is identified as recombinant wherein the results of its notional SNP haplotype correspond to SNP alleles of both of the two paternal chromosomes or two maternal chromosomes in one or more alternating segments consistent with normal recombination between the two chromosomes.

10. A method as claimed in any one of the preceding claims wherein the notional SNP haplotype of target cell chromosomes indicates the presence of both of the two paternal chromosomes or two maternal chromosomes in a pattern and/or frequency inconsistent with normal recombination between the two chromosomes, the cell is deduced to be trisomic for all or part of the relevant chromosome or chromosome segment.

11. A method as claimed in claim 9 or claim 10 wherein consistency with normal recombination is assessed based on the statistical likelihood of normal recombination between particular adjacent, informative SNP alleles of the two paternal chromosomes or two maternal chromosomes during the first meiotic division.

12. A method as claimed in claim 11 wherein the statistical likelihood is assessed based on one or more of the following criteria:

- (i) the average number of recombination events for the specific paternal or maternal chromosome,
(ii) distance between apparent recombination events on each chromosome arm, and their position relative to each other, the centromere and the telomere ie the ends of the chromosome arm involved.

13. A method as claimed in claim 12 wherein the statistical likelihood (p) of two recombination events between SNP alleles separated by a distance of d centiMorgans (cM) is calculated according to the formula:

$$p = (0.0114d - 0.0154)^4$$

14. A method as claimed in any one of claims 1 to 9 wherein the notional SNP haplotype of target cell chromosomes indicates the presence of one chromosome of paternal origin and one chromosome of maternal origin and the cell is deduced to be normal diploid in respect of the relevant chromosome.

15. A method as claimed in any one of claims 1 to 9 wherein the notional SNP haplotype of target cell chromosomes indicates an absence of any chromosome or chromosome segment of paternal origin and maternal origin, the cell is deduced to be nullsomic for the relevant chromosome or chromosome segment.

16. A method as claimed in any one of claims 1 to 9 wherein the notional SNP haplotype of the target cell chromosomes indicates an absence of a chromosome or chromosome segment of either paternal origin or maternal origin but not both, the cell is deduced to be monosomic for the relevant chromosome or chromosome segment.

17. A method as claimed in any one of the preceding claims wherein the average distance between the interrogated SNPs is less than 0.1, 0.2, 0.3, 0.4 or 0.5 kb.

18. A method as claimed in any one of the preceding claims wherein the average distance between the interrogated SNPs is less than 0.1, 0.2, 0.3, 0.4 or 0.5 cM.

19. A method as claimed in any one of the preceding claims further comprising the step of quantitation of interrogated SNP alleles.

20. A method as claimed in any one of the preceding claims further comprising diagnosing the presence of an inherited genetic disease in the target cell by comparing the notional SNP haplotype of the target cell with the SNP alleles of the paternal chromosomes and the maternal chromosomes and one or more affected siblings to diagnose the disease
5 in the target cell by linkage.
21. A method as claimed in any one of the preceding claims further comprising diagnosing susceptibility to a common disease or cancer in the target cell by comparing the notional SNP haplotype with a haplotype known to be associated with said disease.
10
22. A method as claimed in any one of the preceding claims wherein the SNP haplotype of paternal and maternal chromosomes is derived from analysis of the SNP haplotype of cells removed from sibling fertilized embryos following in vitro fertilisation (IVF) following whole genome amplification.
15
23. A method as claimed in any one of claims 1 to 21 wherein the SNP haplotype of paternal and maternal chromosomes is derived from analysis of multiple single parental haploid gametes following whole genome amplification.
- 20 24. A method as claimed in any one of the preceding claims wherein the target cell has been provided from a mammalian embryo which has optionally resulted from IVF.
25. A method as claimed in claim 24 wherein the embryo is a pre-implantation embryo.
- 25 26. A method as claimed in any one of claims 1 to 23 wherein the target cell is provided from a foetus.
27. A method as claimed in any one of the preceding claims wherein a number equal to, or at least, 1, 2, 3, 4, or 5 target cells are provided and tested.
30
28. A method as claimed in any one of the preceding claims wherein SNP interrogation is preceded by whole genome amplification.
29. A method as claimed in claim 28 wherein the whole genome amplification employs
35 isothermal Multiple Displacement Amplification (MDA).

30. A method as claimed in any one of the preceding claims wherein the SNP interrogation is performed by means of an oligonucleotide chip.

5 31. A computer-usable medium having computer-readable program code stored thereon for causing a computer to execute a method to determine aneuploidy or chromosomal recombination in a target cell, which method is the method of any of claims 1 to 30.

10 32. A computer-usable medium having computer-readable program code stored thereon for causing a computer to execute a method to determine aneuploidy or chromosomal recombination in a target cell, which method comprises:
(a) accessing a database comprising genotype data obtained from a plurality of closely adjacent biallelic SNP loci present in a chromosome of the target cell,
(b) accessing a database comprising SNP haplotype data of the corresponding paternal
15 and maternal chromosomes,
(c) comparing target cell SNP data from the database of step (a) with SNP haplotype data from the database of step (b) to assemble a notional haplotype of regions of the target cell chromosomes of paternal origin and of maternal origin,
(d) assessing the notional SNP haplotype of target cell chromosomes of paternal origin
20 and of maternal origin to determine aneuploidy or chromosomal recombination of the chromosome in the target cell.

25 33. A computer-usable medium as claimed in claim 32 wherein each SNP locus of the 'x' SNPs of the database in step (b) is assigned a value 'n' in accordance with which of the 16 combinations of four parental SNP alleles is present at that locus, and wherein step (c) comprises assembling a notional haplotype at that locus by comparing:
(i) the genotype data for the biallelic SNP at that locus from the database of step (a) and,
(ii) the value 'n' at that locus from the database of step (b) with,
(iii) a chromosomal origin table,
30 and thereby assigning the locus of the target cell chromosomes as originating from a paternal or maternal chromosome.

35 34. A computer-usable medium as claimed in claim 32 or claim 33 wherein the notional SNP haplotype of regions of the target cell chromosomes of paternal origin and of maternal origin is assembled using a subset of the 'x' SNP loci from the database in step (b), which subset consists of:

(i) informative SNP alleles that positively identify which one of the four paternal and maternal chromosomes, a chromosome in the target cell has originated from, or positively identify which paternal chromosome and which maternal chromosome a pair of chromosomes in the target cell have originated from, and optionally

5 (ii) semi-informative SNP alleles that positively identify which of two possible combinations of the four possible pair-wise combinations of paternal and maternal chromosomes, a pair of chromosomes in the target cell has originated from.

10 35. A computer-usable medium as claimed in claim 34 wherein the subset consists of heterozygous informative SNP alleles.

36. A computer-usable medium as claimed in claim 34 or claim 35 wherein a threshold number of positive and negative informative and optionally semi-informative SNPs is set, and a karyotype is determined only when this number is exceeded.

15 37. A computer-usable medium as claimed in any one of claims 32 to 36 wherein the notional haplotype result obtained for each SNP locus of the 'x' SNPs, or a subset thereof, is stored and optionally displayed by an output device of said computer.

20 38. A computer-usable medium as claimed in any one of claims 32 to 37 wherein the chromosome in the target cell is identified as non-recombinant wherein the results of its notional SNP haplotype are consistent with:

(i) its SNP alleles being identical to the SNP alleles of one of the two paternal chromosomes or one of the two maternal chromosomes along the length of the chromosome, and

25 (ii) an absence of the SNP alleles of the alternative of the two paternal or maternal chromosomes.

30 39. A computer-usable medium as claimed in any one of claims 32 to 38 wherein the chromosome in the target cell is identified as recombinant wherein the results of its notional SNP haplotype correspond to SNP alleles of both of the two paternal chromosomes or two maternal chromosomes in one or more alternating segments consistent with normal recombination between the two chromosomes.

35 40. A computer-usable medium as claimed in any one of claims 32 to 39 wherein the chromosome in the target cell is identified as trisomic for the chromosome where the

notional SNP haplotype of the target cell chromosome indicates the presence of both of the two paternal chromosomes or two maternal chromosomes in a pattern and/or frequency inconsistent with normal recombination between the two chromosomes.

5 41. A computer-usable medium as claimed in claim 39 or claim 40 wherein the program statistically analyses the likelihood of normal recombination between the SNP loci based on one or more of the following criteria:

(i) comparison of the apparent recombination event with a database of the average number of recombination events for the specific paternal or maternal chromosome,

10 (ii) the position of the apparent recombination events on each chromosome arm relative to each other, the centromere and the telomere.

42. A computer-usable medium as claimed in any one of claims 32 to 41 wherein the chromosome in the target cell is identified as nullsomic for the chromosome where the
15 notional SNP haplotype of target cell chromosome indicates an absence of the chromosome or a segment thereof of both paternal origin and maternal origin.

43. A computer-usable medium as claimed in any one of claims 32 to 42 wherein the chromosome in the target cell is identified as monosomic for the chromosome where the
20 notional SNP haplotype of the target cell chromosome indicates an absence of the chromosome or a segment thereof of paternal origin and maternal origin but not both.

44. A computer programmed to execute a method as claimed in any of claims 1 to 30.

25 45. A computer programmed with the computer-usable medium as claimed in any of claims 31 to 43.

46. A system for karyotyping a target cell to detect chromosomal imbalance therein, the system comprising:

30 (i) means for interrogating closely adjacent biallelic SNPs across the chromosome of the target cell,

(ii) a computer-usable medium as claimed in any one of claims 31 to 43.

47. A system for karyotyping a target cell to detect chromosomal imbalance therein, the
35 system comprising:

(i) means for interrogating closely adjacent biallelic SNPs across the chromosome of the

target cell,

(ii) a computer as claimed in claim 44 or claim 45, optionally interfaced with said means.

48. A system as claimed in claim 46 or claim 47 wherein the means for SNP
5 interrogation is an oligonucleotide chip.

49. A system as claimed in any one of claims 46 to 48 wherein the means for SNP
interrogation is adapted to interrogate:

- 10 (i) a high density of biallelic SNPs on chromosomes frequently associated with miscarriage
or viable abnormal pregnancies,
(ii) a high density of biallelic SNPs in regions associated with microdeletion syndrome,
(ii) a high density of biallelic SNPs in regions associated with common single gene defects,
(v) SNP haplotypes associated with predisposition to common complex diseases.

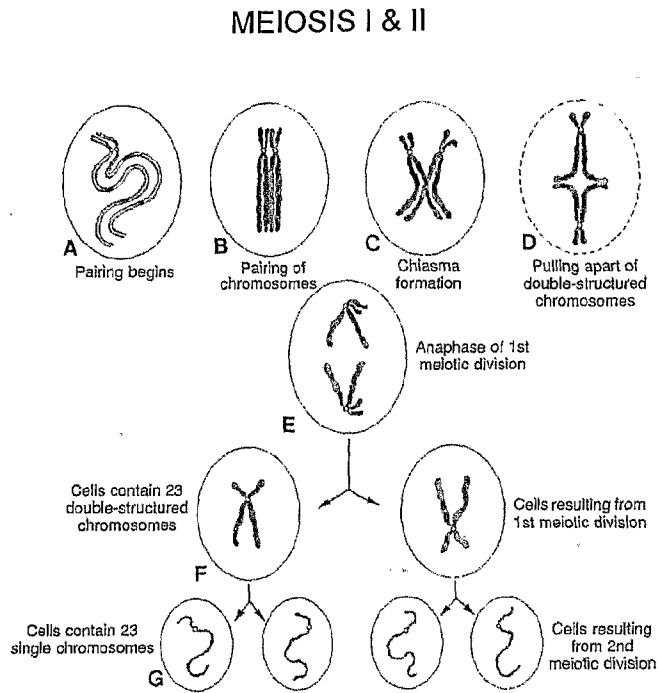


Figure 1a

NON-DISJUNCTION RESULTING IN ANEUPLOIDY

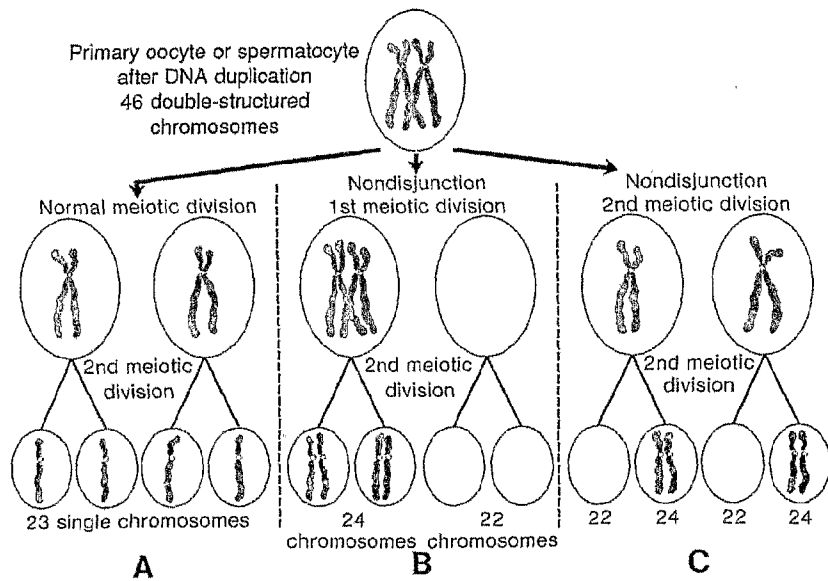


Figure 1b

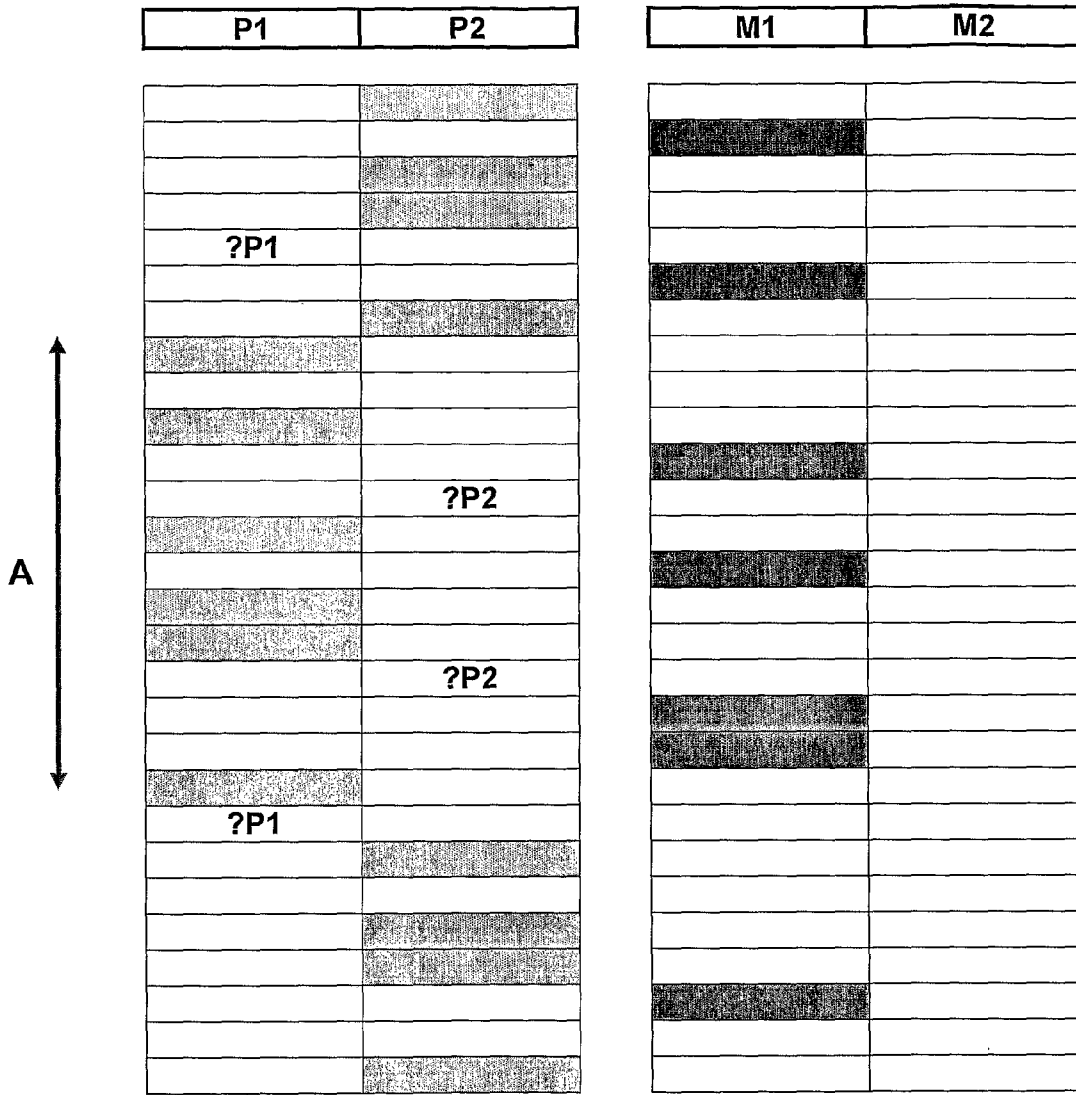


Figure 2a and 2b

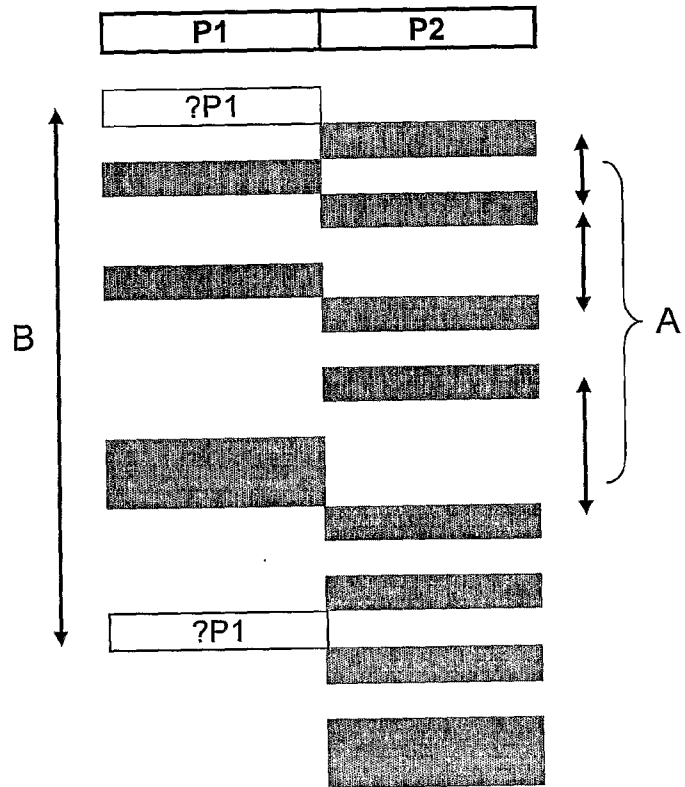
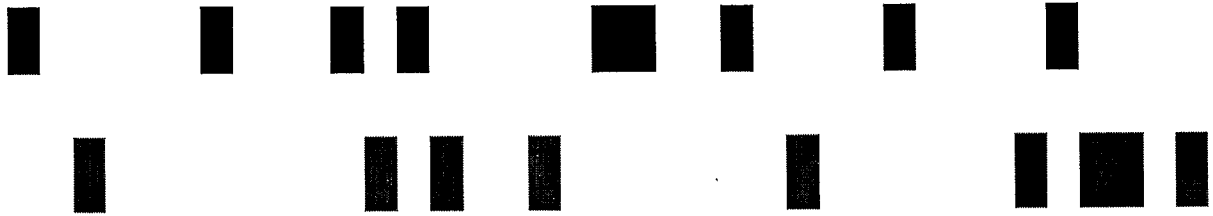
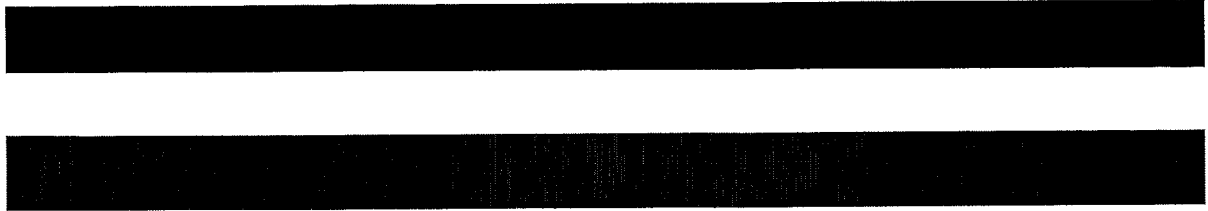


Figure 3

Figure 4 (Continued)



AB	BB	AB	AB	BB	AB	AB	AA	NoCall	AA	BB	AB	AB	AB	BB	AB	AA	AA	BB	BB	BB	AB	AB	AA	AB	BB	AA	AB	NoCall	BB	AB	AA	AB	AB	AB	BB	AA
12	13	5	14	2	3	12	16		16	10	5	9	8	2	15	7	1	10	10	2	3	12	1	5	2	16	9		2	15	7	12	5	5	13	7
A	B	A	B	B	B	A	A	B	A	B	A	B	B	B	A	A	A	A	B	B	B	B	A	A	A	A	B	B	B	A	A	A	A	A	B	A
B	A	A	A	B	B	B	B	B	B	A	A	A	B	B	B	A	A	A	A	A	B	B	B	A	A	B	B	A	B	B	A	A	B	A	A	A
B	B	B	A	B	A	B	A	A	A	B	B	A	A	B	B	A	A	B	B	B	A	B	A	A	B	B	A	A	A	B	B	A	B	B	B	A
B	A	A	B	B	A	B	B	NoCall	B	B	A	A	B	B	A	A	B	B	B	B	A	B	A	A	B	B	A	NoCall	B	A	B	B	A	A	A	B

Figure 4 (Continued)



AB
AB
AA
AA
NoCall
AA
AB
AA
BB
AB
AB
AA
BB
AB
AB
AB
AB
AB
AB
BB
AB
BB
AA
AB
AB

9 7 5 15 5 6 1 10 3 9 5 10 16 16 6 6 12 9 12 9 10 9 16 2 16 2 5 6 9

B A A A B A A B A B A A B A A B B A B A B A B A B A B A B A B A B A B

A A A B A B A A B A A B A A B A B B B A B A A A B B A B B A B A B A B A

A A B B B B B A B A B A A B B A A B B B A B A A B A A B B A B B A B B A

A B A A
NoCall
A A A B A A A B B B A B A B A B A B A B B B A A A

Figure 4 (Continued)



AB AB AA AB AB BB AB AB BB AA BB AB AB AB AA AB AA BB AB AA AB AB AB AB AB AA BB BB BB AA AB AB AA AA

13 14 15 16 13 2 9 4 8 5 2 3 9 7 13 15 9 15 8 9 1 13 4 9 6 3 9 12 11 8 10 10 5 12 9 5 5

B B A A B B B A B A B B B A B A B A B B B B A A B B A A B B A A B B A A B B A A B A A

A A B B A B A A B A B A B B A A A B A A B B A A A A A B B A A B B A A B B A A B A A

B A B A B B A B A B B A A A B B A B A A A B B A B A A A B A A B A A B A A B B A A B B A A B B

A B A B A B A B B A B A B A A B A A A A B A A A A B A A A A B A A B B B A B A A A A

Figure 4 (Continued)



AB	NoCall	AA	AB	AB	AB	AA	BB	AA	BB	AB	AB	AB	BB	AA	AB	AB	BB	BB	AA	AB	AA	AB	BB	AA	AB	AB	AA	AA	NoCall	AA	AB	AA	BB	AB	AB	
6	5	12	16	3	11	14	1	10	16	12	16	8	5	12	13	14	8	1	9	5	3	2	15	16	9	7	5	15	5	6	1	10	3	9		
B	A	A	A	B	A	B	A	B	A	A	A	B	A	A	B	B	B	A	A	B	A	B	B	A	A	B	A	A	A	B	A	B	B	B	B	
B	A	B	B	A	A	A	A	B	A	B	B	A	A	B	A	A	B	A	A	A	A	B	B	B	A	A	A	A	B	A	B	B	B	B	B	B
B	A	B	B	A	A	A	A	B	A	B	B	A	A	B	B	B	A	A	A	A	B	B	B	A	A	A	A	B	B	B	B	B	B	B	B	B
A	NoCall	A	B	B	A	A	A	B	B	B	B	B	A	B	A	B	B	A	A	A	B	B	B	A	B	A	A	A	NoCall	A	A	A	B	A	A	

Figure 4 (Continued)



AA
BB
AB
AB
AB
AB
AB
AB
AB
AB
BB
AB
AB
BB
AB
BB
AA
AB
AB

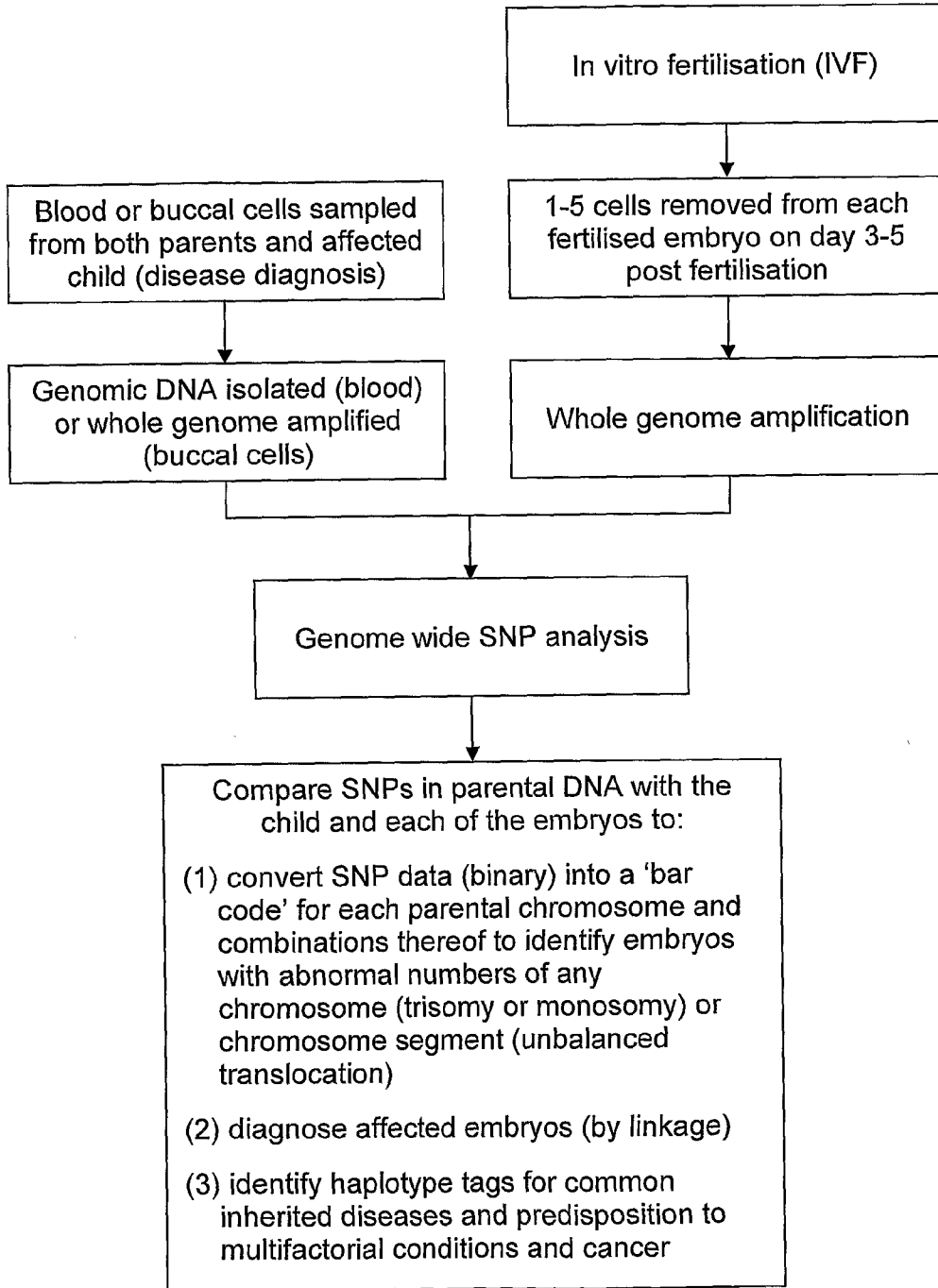
5 10 16 16 6 6 12 9 12 9 10 9 16 2 16 2 5 6 9

A B A A B B A B A B B A B A B B A B A B B

A A B B B B B A B A A A B B B B A B A

B B A A B B B A B A B A A B A B B B B A

A B B B A A B A B A B A B B B B A A A



INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2006/004221

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SLATER HOWARD R ET AL: "High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 77, no. 5, 16 September 2005 (2005-09-16), pages 709-726, XP002423358 ISSN: 0002-9297 High-throughput approach for detecting chromosomal abnormalities using a high density oligonucleotide array the whole document ----- -/--	1-49

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed
- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *8* document member of the same patent family

Date of the actual completion of the international search 15 March 2007	Date of mailing of the international search report 04/04/2007
---	---

Name and mailing address of the ISA/ European Patent Office, P.E. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Bort, Sylvia
---	---

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2006/004221

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>PAPERNA T ET AL: "Genes for the CPE Receptor (CPETR1) and the Human Homolog of RVP1 (CPETR2) Are Localized within the Williams-Beuren Syndrome Deletion" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 54, no. 3, 15 December 1998 (1998-12-15), pages 453-459, XP004457222 ISSN: 0888-7543 Detection of haploinsufficiency of genes in the chromosomal region 7q11.23 using SNP the whole document</p>	1-49
A	<p>FAN JIAN-BING ET AL: "Paternal origins of complete hydatidiform moles proven by whole genome single-nucleotide polymorphism haplotyping" GENOMICS, vol. 79, no. 1, January 2002 (2002-01), pages 58-62, XP002423543 ISSN: 0888-7543 Paternal origins of complete hydatidiform moles (diploid tumours) using a high density SNP oligonucleotide array abstract</p>	
A	<p>LE CAIGNEC C ET AL: "Detection of genomic imbalances by array based comparative genomic hybridisation in fetuses with multiple malformations" JOURNAL OF MEDICAL GENETICS, vol. 42, no. 2, February 2005 (2005-02), pages 121-128, XP002423544 ISSN: 0022-2593 Detection of genomic imbalances using array based comparative genomic hybridisation and quantitative multiplex PCR in fetuses with multiple malformation the whole document</p>	