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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 April 2010 (22.04.2010)

PCT

(10) International Publication Number
WO 2010/044923 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/10* (2006.01)
C12Q 1/02 (2006.01)

(21) International Application Number:
PCT/US2009/047765

(22) International Filing Date:
18 June 2009 (18.06.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/136,992 21 October 2008 (21.10.2008) US

(71) Applicant *(for all designated States except US)*: **MOREHOUSE SCHOOL OF MEDICINE** [US/US]; 720 Westview Drive, Atlanta, GA 30310 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SONG, Qing** [CN/US]; 2360 Elon Way, Decatur, GA 30033 (US). **MA, Li** [CN/US]; 1556 Woodlake Dr., Atlanta, GA 30329 (US). **XIAO, Yan** [CA/US]; 5035 Avala Park Ln, Norcross, GA 30092 (US).

(74) Agent: **WANG, Ping, M. D.**; Morris, Manning & Martin, LLP, 1333 H Street, N.W., Suite 820, Washington, DC 20005 (US).

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

[Continued on next page]

(54) Title: METHODS FOR HAPLOTYPE DETERMINATION BY HAPLODISSECTION

(57) Abstract: A method for molecular haplotyping of a subject is disclosed. The method comprises: randomly selecting a set of chromosomes in each of a plurality of lyzed diploid cells of the subject, collecting the selected chromosomes from said plurality of cells into a plurality of sample tubes, wherein each sample tube contains chromosomes selected from one or more cells, genotyping genomic DNA in each sample tube, and determining haplotype of the alleles based on allele nucleotide sequence information and corresponding nucleotide signal intensities from genotyping data. Other methods for molecular haplotyping using single cell lysate or single cell microdissection are also disclosed.

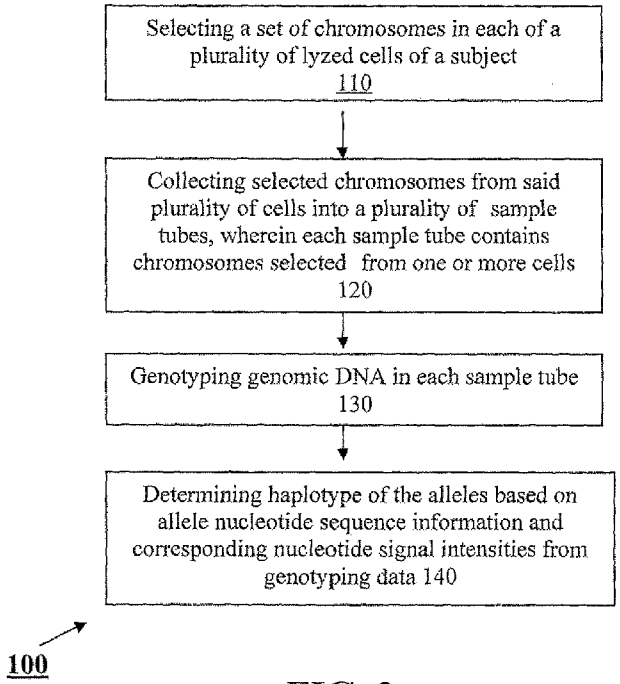


FIG. 2

WO 2010/044923 A1

ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, 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 (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *upon request of the applicant, before the expiration of the time limit referred to in Article 21(2)(a)*

TITLE

METHODS FOR HAPLOTYPE DETERMINATION BY HAPLODISSECTION

[001] This application claims priority from U.S. Provisional Application Serial No. 61/136,992, filed October 21, 2008. The entirety of that provisional application is incorporated herein by reference.

FIELD

[002] The present invention generally relates to the fields of genetics, molecular and cell biology and, in particular, relates to methods for haplotype determination.

BACKGROUND

[003] Normal human somatic cells are diploid (*i.e.,* having two copies of genome: a paternal set of chromosomes and a maternal set of chromosomes in each nucleus). Within each individual, these two sets of chromosomes have different nucleotide sequences (single-nucleotide polymorphism (SNP)) at multiple loci. Conventional genotyping assays analyze a mixture of these two sets of chromosomes, which leads to uncertainty and complexity. For example, for any two SNP loci that are both heterozygous, there will be four possible haplotypes between these two SNPs. However, since the phase information was erased when doing the single SNP genotyping using the conventional platforms, none of these four possible haplotypes can be eliminated. One way to solve this problem is to find a reliable method to re-establish or retract the phase information. Another way is to extract the phase information before doing genotyping.

[004] The skilled artisans in this field used the various statistical algorithms to re-establish the phase information. These algorithms include Clark's algorithm, expectation-maximization (EM) algorithm, coalescence-based algorithms (pseudo-Gibbs sampler and perfect/imperfect phylogeny), and partition-ligation algorithms implemented by a fully Bayesian model (Haplotype) or by EM (PLEM) (Liu N, et al., *Advances in Genetics*, 60: 325-405, 2008). Statistical configuration of haplotypes based on unphased genotype data usually gives a large number of uncertain haplotypes, which significantly reduces the power in genetic applications. In addition, it is still controversial as to whether the configured haplotypes should be treated as objective observations of genotypes and phenotypes in these studies. While genotypes from family members can often help to determine the haplotypes, haplotype inference from family data is often limited by uninformative or missing data. Moreover, late-age onset for most of the common human diseases can preclude collection of DNA samples from previous generations. Therefore, these methods are not suitable for the molecular diagnosis in personalized medicine in the future.

[005] In parallel, some researchers developed experimental methods to extract the phase information in the genomic DNA samples before genotyping. These methods are all based on the physical separation of two homologous genomic DNAs before genotyping. The challenge is how to separate two almost identical copies of chromosomes in diploid cells. Several strategies/technologies have been developed for separating diploid samples into their haploid components, such as 1) Long-range allele-specific genomic PCR (Michalatos-Beloin S, et al., *Nucleic Acids Res* 24: 4841-4843, 1996; and Yu CE, et al., *Genomics* 84: 600-612, 2004); 2) Haplotype-Specific Extraction (HSE) (Nagy M, et al., *Tissue Antigens* 69: 176-180, 2007); 3) Generation of somatic haploid cells, such as GMP conversion (Douglas JA, et al.,

Nat Genet 28: 361-364, 2001); 4) Polony (Mitra et al., *Proc Natl Acad Sci* US A100: 5926-5931, 2003; Zhang K, et al., *Nat Genet* 38:382-387, 2006); 5) Clone-based systematic haplotype (CSH) (Burgtorf C, et al., Genome Res13: 2717-2724, 2003); 6) single molecule dilution (SMD) (Ding C, et al., Proc Natl Acad Sci U S A 100: 7449-7453, 2003); and 7) Sperm typing.

[006] Long-range Allele-Specific Genomic PCR uses specifically designed PCR primers to selectively amplify the target region from only one of the sister chromosomes. Selective amplification is achieved by designing a primer that will match/mismatch to one of the alleles at the 3'-end of the primer. Thus, the primer cannot amplify the unmatched chromosomal DNA template efficiently. Genotyping will be done subsequently on the amplification products. Because these PCR products are obtained from only one of the chromosomes, the alleles of different SNPs along these PCR products reveal the haplotype.

[007] In this method, the maximal distance of the genetic markers in the haplotype is determined by the maximal length that PCR can reach and the chromosome integrity in DNA preparation. Therefore, the haplotype length is restricted by the PCR capacity, which is about 40 kb for long PCR. This method is often technically challenging and requires extensive optimization of PCR conditions for every primer pair to improve the amplification efficiency of long PCR. Different combinations of several primer pairs and buffers are usually recommended to optimize PCR condition. However, this method is not applicable to high throughput analysis of haplotypes.

[008] Haplotype-Specific Extraction (HSE) uses specifically designed probes to selectively capture the fragments from only one of the sister chromosomes. Selective binding is achieved by designing a probe that specifically recognizes one allele of a SNP. If an

individual is a heterozygote, when this probe is added into the denatured genomic DNA samples, the probe will seek and bind only to the genomic DNA fragments containing its target allele. Therefore, the probe-bound DNA fragments are captured by immobilized magnetic beads and the unbound DNA fragments with the other allele of this SNP will be washed away. Now the genomic DNA in diploid state is reduced to haploid state and ready for all subsequent analysis including genotyping/haplotype. Because distinct polymorphic differences always exist between two parental chromosomes, HSE can distinguish and separate the two parental copies for any chromosomal segments.

[009] In this method, the maximal distance of the genetic markers in haplotype is determined by the chromosome integrity in DNA preparation and the DNA denaturation. This method can resolve haplotypes within a distance of <50 kb so far. If molecular haplotypes over extended distances are needed, multiplexed haploseparations have to be carried out.

[010] GMP Conversion Technology is built upon constructions of cell hybrids from viable human cells (typically lymphocytes or fibroblasts) and a rodent cell line. Because these hybrid cells retain only a subset of human chromosomes, they can be either null, monosomic or disomic for each pair of human chromosomes. Those monosomic cells are haploid for the corresponding chromosomes and ready for subsequent genotyping assays for determination of haplotype.

[011] In this technology, cells are electrofused and then propagated under a selective condition, for example, using the HPRT1/HAT (hypoxanthine, aminopterin, and thymidine) system. After 2-4 weeks of growth, fused clones are harvested, and DNA is prepared for analysis. The monosomic clones can be identified by genotyping a few, highly polymorphic markers per chromosome, which minimally requires a single heterozygous genotype.

Nonetheless, there are still some technical challenges on conversion-based haplotyping, including low DNA concentrations, preferential amplification, and insertions or deletions of chromosomal segments (Douglas JA et al., *Nat Genet* 28: 361-364, 2001).

[012] It has been observed that whole chromosomes rather than chromosomal fragments are generally retained in the hybrid cells (*Supra* Douglas 2001). Therefore, this method does not have any restrictions on the distance of SNPs in a haplotype. The application of GMP Conversion Technology is restricted to a very limited number of subjects and chromosomal regions because of the inefficiencies and variations in fusion and selection conditions. Numerous cell lines are required for each individual. Conversion-based haplotyping is still very time-consuming and very costly.

[013] Polony Technology uses a polyacrylamide gel to work on an in situ single molecule of chromosomal DNA. In this technology, genomic DNA from an individual is first diluted to a very low concentration, and then mixed with acrylamide and spread onto a glass microscope slide to form a thin DNA-containing polyacrylamide gel. Because the DNA concentration is so low, the DNA molecules are well separated from each other. An in-gel PCR is then performed directly on this gel, with 2 pairs of PCR primers to amplify two loci of the SNPs of interest from a single DNA molecule. Because the acrylamide matrix restricts the diffusion of linear DNA molecules, PCR products accumulate around their amplification template forming two overlapping PCR colonies (polony). The genotypes of these two SNPs are determined in situ by single-base extension (SBE) assay separately for these two SNPs and the gels are read by a laser scanner. After overlaying the two SBE images, the alleles observed on the same spot indicate the allele combination (haplotype) of these two SNPs of this patient sample.

[014] The maximal haplotype length of Polony is determined by the DNA fragmentation or degradation before, during and after the acrylamide polymerization. It is reported that this method has measured the haplotype as long as 45 kb so far (Mitra, et al., *PNAS USA* 100: 5926-5931, 2003; Zhang K, et al., *Nat Genet* 38:382-387, 2006).

[015] There are several inherent caveats in the Polony method. One major limitation of Polony haplotyping is that it is not efficient for scaling up the number of SNPs. But it is often desirable to haplotype a large number (100-10,000) of SNPs along a chromosome. Second, the DNA molecules may overlap in the gel. Therefore, the DNA concentration and plating condition is critical. Third, the PCR coamplification efficiency is low (4-15% for samples from buccal swabs, 15-34% for samples from the other collection methods). The coamplification efficiency is related to the presence of ungelled acrylamide in the Polony gel during thermal cycling and DNA fragmentation or degradation. Technical optimization (such as degas and polymerization condition) may be required. Lastly, this technology requires metaphase cells.

[016] Clone-based Systematic Haplotyping (CSH) uses fosmid/cosmid cloning to isolate a single copy from diploid chromosomes. Because each vector molecule can hold only one insert molecule, each colony derived from successful vector-insert ligation will hold only a haploid chromosomal segment. By screening the colony library, the clones that contain the target chromosomal segments will be obtained for subsequent haplotyping analysis. Because the vector cannot successfully accept inserts with a very large size beyond their maximal cloning capacities, CSH can separate a haploid fragment of ~50 kilobases. In addition, this method is very time-consuming and costly.

[017] Single Molecule Dilution (SMD) is built upon the idea that a single molecule is certainly a haploid fragment because diploid chromosomes are a pair of copies and require two DNA molecules to constitute a diploid. To obtain a single molecule in each reaction tube, genomic DNA samples are diluted to an extremely low concentration. We have known that each diploid genome of human is ~6.7 pg, so if a tube contains only ~3.3 pg of genomic DNA, it must have single molecules for some chromosomal regions because the DNA amount is not sufficient for every chromosomal region to have two copies in that tube. This very low DNA concentration is achieved by serial dilutions. After serial dilution, for any given chromosomal segment, each tube may contain no DNA, one molecule of DNA for that region, or two molecules of DNA for that region. The tiny amount of DNA samples in these tubes is then amplified and genotyped; allele drop-out at previously identified heterozygous SNP loci of this individual is used to screening out the "single-molecule" tubes for further experiments. The caveat of this method is that it relies on statistical isolation of single DNA molecules, so there is no experimental guarantee for its success.

[018] In this method, due to frequent shearing in serial dilutions, genomic DNA is broken down. The maximal distance is so far reported to be ~24 kb in haplotyping distance (Ding C, et al., *PNAS USA* 100: 7449-7453, 2003).

[019] Sperm Typing is built upon the fact that a sperm is a product of meiosis and only contains a haploid genome. Despite sperm being haploid, sperm haplotypes are not simply equal to the donor's haplotype. The sperm haploid genome is not any one of parental chromosomes of this individual. However, by genotyping several sperms from one individual and then analyzing the haplotype data from these sperms, the haplotypes of this individual can

be inferred. Therefore, sperm typing is different from the above molecular haplotyping methods because it is not a direct haplotyping.

[020] Different sperms have gone through different crossing over events in meiotic recombination, so sperms from the same individual will have different haplotypes. In crossing over, two chromatids exchange their distal arms of chromosomes; usually this distal end of the chromosomes are exchanged only once in humans, sometimes twice or more times. Therefore, it is possible to infer the haplotypes of the original patient from a number of sperm under the assumption that only one crossing over event occurred in the studied sperms. However, since sperm typing is limited to male only, the procedure is tedious and costly, and the haplotypes are inferred results, not direct observations; sperm typing is not widely used for molecular haplotyping.

[021] In summary, the currently available experimental methods for chromosome separation often cause the chromosome breakdown so they cannot obtain the long-range haplotypes. In addition, they are extremely time-consuming and labour-intensive, so they are not practically feasible in researcher laboratories and clinics. There still exists a need for a haplotyping method that can be performed quickly at low cost.

SUMMARY

[022] A method for molecular haplotyping of a subject is disclosed. The method comprises: randomly selecting a set of chromosomes in each of a plurality of lyzed diploid cells of the subject, collecting the selected chromosomes from said plurality of cells into a plurality of sample tubes, wherein each sample tube contains chromosomes selected from one or more cells, genotyping genomic DNA in each sample tube, and determining haplotype of

alleles based on allele nucleotide sequence information and corresponding nucleotide signal intensities from genotyping data.

[023] In one embodiment, the step of determining haplotype of alleles includes the steps of: extracting allele nucleotide sequence information and corresponding nucleotide signal intensities from genotyping data, calculating the nucleotide signal intensity ratio of two alleles (allelic intensity ratio) for each heterozygous locus, and determining haplotype of the alleles.

[024] In another embodiment, the step of calculating allelic intensity ratio includes the steps of: calculating relative ratio of nucleotides A, C, G, and T at homozygous loci, determining a k value for each nucleotide to adjust their signal intensities to the same level, adjusting nucleotide signal intensities at heterozygous loci using the k value, and calculating the allelic intensity ratio at the heterozygous locus.

[025] In another embodiment, the determining step further comprises the steps of sorting the order of alleles at each locus by allelic intensity ratio, keeping the higher-intensity-allele on a first column and the lower-intensity-allele on a second column, determining whether there is breakpoint in each chromosome, if there is no breakpoint in a chromosome, forming one haplotype with alleles in the first column, and another haplotype with alleles in the second column, if there is a breakpoint in a chromosome, using results from other chromosome collection tubes to bridge over the breakpoint.

[026] In a related embodiment, the cells are peripheral blood lymphocytes.

[027] In another related embodiment, chromosomes from 2-10 randomly selected cells are collected into a sample tube and a total of 4-8 sample tubes are collected.

[028] In another related embodiment, the genotyping step comprises amplifying genomic DNA.

[029] Also disclosed is another method for molecular haplotyping of a subject. The method comprises: isolating one or more single diploid cells from the subject, lysing each isolated single diploid cell to generate one or more single cell lysate, dividing each single cell lysate into two equal aliquots, genotyping genomic DNA in each aliquot, creating a catalogue of genotyping data from all aliquots, and determining chromosome haplotype of the subject based on the catalogue.

[030] In a related embodiment, the isolating step includes isolating 4-12 single diploid cells from the subject.

[031] In another related embodiment, the isolating step includes isolating 6-10 single diploid cells from the subject.

[032] In another related embodiment, the isolating step includes isolating 8 single diploid cells from the subject.

[033] Also disclosed is another method for molecular haplotyping of a subject. The method comprises: isolating a single diploid cell from the subject, lysing and staining the isolated single diploid cell to display chromosomes, collecting a set of chromosomes from the single cell by laser microdissection, genotyping genomic DNA in the collected chromosomes, genotyping genomic DNA from one or more intact diploid cells of the same subject, determining haplotype of a chromosome in the collect set of chromosomes, wherein said chromosome is present in a haploid form in said collected set of chromosomes.

[034] In a related embodiment, multiple single diploid cells are isolated and lysed. Multiple sets of chromosomes are collected; each set is collected from a different single cell of the same subject. The number of collected sets is large enough so that each chromosome in the genome of the subject is present in the haploid form at a probability greater than 99%.

[035] In a related embodiment, the subject is an eukaryotic organism.

[036] In a related embodiment, the eukaryotic organism is an animal or a plant.

[037] In a further related embodiment, the animal is a mammal.

[038] Another aspect of the present invention relates to a computer-readable medium having computer-executable instructions for performing the methods described above.

[039] Another aspect of the present invention relates to an assay kit for HaploDissection. In one embodiment, the assay kit contains reagents for cell collection and cytogenetic staining, and reagents for genomic DNA amplification and genome genotyping. In another embodiment, the kit further includes a computer readable medium having computer-executable instructions for determining haplotype based on genotyping data.

BRIEF DESCRIPTION OF THE DRAWINGS

[040] Figure 1 is a schematic illustrating the general principle of an embodiment of the HaploDissection method by introducing imbalance into chromosome ratio.

[041] Figure 2 is a flow chart showing an embodiment of the HaploDissection method.

[042] Figure 3 is a picture showing chromosome collection using a Leica AS LMD computer-directed laser microdissection. Chromosomes in the collection area were collected for haplotyping.

[043] Figure 4 is a flow chart showing an embodiment of a method for determining haplotype-based genotyping data.

[044] Figure 5 is a flow chart showing another embodiment of the HaploDissection method using single cell lysate.

[045] Figure 6 is a diagram illustrating the principle of single cell haplotype split.

[046] Figure 7 is a diagram illustrating the principle of haplotyping with a HaploDissection method using single cell dissection.

[047] Figure 8 is a flow diagram showing the steps of single cell dissecting method and some haplotyping results. Haplotypes are shown by their parental origins (Fa, father; Mo, mother).

DETAILED DESCRIPTION

[048] The practice of the embodiments described in further detail below will employ, unless otherwise indicated, conventional methods of genetics, genomics molecular biology, cell biology, diagnostics and bioinformatics within the skill of the art. Such techniques are explained fully in the literature. All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

[049] One aspect of the present invention relates to a method for molecular haplotyping of a subject. This method is referred to hereinafter as the "HaploDissection" method. As described in more detail below, the new method overcomes the bottleneck on the haplotype length, and has no limitations on SNP numbers or sample numbers. This invention meets the needs of accurate haplotypes in genetic studies, genomic studies and epigenomic studies, especially the genome-wide association studies (GWAS), long-cis-regulatory

interactions for gene expression, and chromatin remodelling studies. Accurate haplotypes are required for interpretation and translation of these results into clinical practice.

[050] One embodiment of the HaploDissection method is illustrated in Figure 1. Briefly, the HaploDissection method maintains the phase information of DNA samples from a subject that are subjected to genotyping, but this maintenance is not based on the separation of two chromosome copies or isolation of a single copy. The method simply introduces an imbalance into the steady 1:1 ratio on the quantities of two parental chromosomes by harvesting a relatively small number of chromosomes into each sample tube. Thus, while the genotype/allele information is still reserved in the DNA samples and applicable to hi-throughput genotyping platforms, the phase information is recorded into the quantitative ratio between two alleles. These relative ratios of alleles are actually one of the outputs from all of those genotyping platforms, but they are usually ignored in the genotyping interpretation. The HaploDissection method will read the output information on both allele readings and allele intensity readings. The genotyping information and the phase information are then analyzed by a specially designed algorithm to determine the haplotype of the subject. In one embodiment, the genotyping information and the phase information are then analyzed by a special designed software called "HapReader."

[051] Because the HaploDissection method protects the chromosome integrity while introducing the quantitative allele imbalance, the haplotypes obtained from this method will be at the entire chromosome range, or unlimited by distance.

[052] An embodiment of the HaploDissection method is shown in Figure 2. In this embodiment, the method 100 includes: selecting (110) a set of chromosomes in each of a plurality of lyzed cells of a subject;, collecting (120) selected chromosomes from said

plurality of cells into a plurality of sample tubes, wherein each sample tube contains chromosomes selected from one or more cells; genotyping (130) genomic DNA in each sample tube; and determining (140) haplotype of the alleles based on allele nucleotide sequence information and corresponding nucleotide signal intensities from genotyping data.

[053] The chromosomes may be selected from any type of cells. In one embodiment, the cells are peripheral blood lymphocytes isolated from a blood sample of the subject. Methods for isolating peripheral blood lymphocytes are well-known in the art. In one embodiment, the isolated peripheral blood lymphocytes are cultured in a growth medium until they start proliferation. Growth media capable of inducing proliferation are well known in the art. In one embodiment, the growth medium is RPMI1640 containing 15% FBS and 100 unit/ml Penicillin/Streptomycin. Mitogens such as phytohemagglutinin (PHA) may be added to the culture medium to stimulate cell proliferation, and a mitotic inhibitor such as colcemid is added to arrest the cells in metaphase. The proliferating peripheral blood lymphocytes are then harvested, lyzed, and stained to display chromosomes using well-known cytogenetics procedures. A set of chromosomes, typically around half of the chromosomes in a lyzed cell, is randomly selected and collected for further analysis. Figure 3 shows an example of selecting chromosomes from a single cell for collection using computer-directed laser microdissection. Chromosomes in the marked area were collected. As noted earlier, the chromosomes are randomly selected for collection.

[054] Randomly selected chromosomes from randomly selected cells are collected into a plurality of sample tubes. Each tube contains chromosomes collected from a plurality of cells. In one embodiment, each tube contains chromosomes collected from 2-20, preferably 2-10, randomly selected cells. A total of 2-12, preferably 4-8, and sample tubes

were collected for each subject. The selected chromosomes are collected using technologies that preserve the chromosome integrity. In one embodiment, the chromosomes are collected using computer-directed laser microdissection.

[055] In the next step, genotyping is performed on collected chromosomes in each sample tube. In one embodiment, the collected DNA is amplified by PCR using methods for unbiased whole genome amplification (WGA). The amplified DNA is then subjected to whole genome genotyping. For each subject, 2-4 tubes of samples will be subjected to the genotyping to ensure high genome coverage and achieve duplications for accuracy. In one embodiment, a genomic DNA sample is included for whole genome genotyping.

[056] The output dataset from genotyping is analyzed to determine the haplotype of the subject using a method that integrates the sequencing information with the phase information, which is reflected on the intensity of sequencing signals at each loci. Briefly, the allele nucleotide sequence information and corresponding nucleotide signal intensities are extracted from the genotyping data obtained from the collected chromosomes in each sample tube. The nucleotide signal intensity ratio of two alleles (allelic intensity ratio) for each locus of a chromosome is calculated and is used for and determining the haplotype of the alleles.

[057] Figure 4 is a flow chart showing an embodiment of a method for determining allele haplotype based on the nucleotide sequence information and corresponding nucleotide signal intensities of the alleles. In this embodiment, the method 200 includes: extracting (210) allele nucleotide sequence information and corresponding signal intensities from genotyping data, calculating (220) relative ratio of nucleotides A, C, G, and T at homozygous loci, determining (230) a k value for each nucleotides to adjust their signal intensities to the same level for a given particular experiment, adjusting (240) nucleotide signal intensities at

heterozygous loci using the k value; calculating (250) the signal intensity ratio of two alleles (allelic intensity ratio) for each locus; sorting (260) the order of alleles at each locus by allelic intensity ratio, keeping the higher-intensity-allele on a first column and the lower-intensity-allele on a second column, determining (270) whether there is breakpoint in each chromosome, if there is no breakpoint in a chromosome, forming (280) one haplotype with alleles in the first column forming the haplotype, and another haplotype with alleles in the second column, if there is a breakpoint in a chromosome, using (290) results from other chromosome collection tubes to bridge over the breakpoint.

[058] In one embodiment, the analysis is performed using a "HapReader" software that is specifically developed for the HaploDissection technology. The software is described in more detail in the EXAMPLES.

[059] Another embodiment of the HaploDissection method is based on the separation of haploid genome from a single cell lysate. In all of the current molecular haplotyping methods, the reduction from diploid to haploid is achieved from an uncertain and large number of chromosome copies. The method of the present invention is built upon the fact that each somatic cell has exactly two copies of each chromosome. This exact number provides a very simple way to separate the two chromosomes. Briefly, the method separates chromosomes on a single cell basis. This new starting point makes the separation much easier than previous inventions because there are only two copies of each chromosome at the starting point. In addition, the method overcomes a major drawback of the other methods -- the short haplotype distance. Therefore, this new method opens the door to a simple and effective way to obtain long-distance haplotypes.

[060] As shown in Figure 5, the method 300 includes the steps of isolating (310) one or more single diploid cells from the subject; lysing (320) each single diploid cell from the subject to generate one or more single cell lysate; dividing (330) each single cell lysate into two equal aliquots; genotyping (340) genomic DNA in each aliquot; creating (350) a catalogue of genotyping data from all aliquots; and determining (360) chromosome haplotype of the subject based on the catalogue.

[061] The diploid cells can be buccal cells, lymphocytes or any other cell types from the subject. In one embodiment, human buccal cells are collected from a subject. Buccal cells are the cells on the inner lining of the mouth or cheek. They are routinely shed and replaced by new cells. As the old cells die, they accumulate in the saliva in the mouth and can easily be collected by a simple procedure using mouthwash. Buccal cells can be easily collected by swabs, cytobrushes, mouthwash, and treated cards, such as FTA or IsoCode cards.

[062] Next, single cells are isolated and kept in individual tubes. This can be done by any method that can isolate single cells while preserving genomic DNA in cells. Examples of such method include, but are not limited to, laser microdissection or flow cytometry.

[063] In one embodiment, isolation of single cells is carried out using laser microdissection or any other methods such as cell sortings. Laser microdissection is a micromanipulation procedure that allows cutting off precisely the cells of interest from tissue samples or smears under direct microscopic visualization by a laser beam. The region of interest is marked on the computer monitor and then cut out by computer control. The single

cell in the collection tube can be immediately checked by an inspection mode under the microscope to ensure the successful isolations.

[064] The staining protocol used during cell isolation should not interfere with the subsequent DNA amplification and allele determination. Preferred staining methods do not include any fixation step and are not based on the use of aggressive chemical agents. In one embodiment, the cells are stained with Papanicolau. In another embodiment, the cells are stained with Hematoxylin and eosin (HE).

[065] Single cells collected by microdissection are then subjected to cell lysis. Many techniques are available for cell disruption, including physical and detergent-based methods. The technique chosen for the disruption of cells must take into consideration the compatibility with the intended downstream applications – genotyping and haplotyping. Therefore, the cell lysis method should not attach the DNA molecules aggressively and break down the chromosomes into small pieces. Any genomic DNA preservative, effective, simple, and low-cost methods can be selected. The origin of the cells or tissues should also be considered with choosing cell lysis protocols.

[066] Both physical lysis methods and detergent-based lysis methods may be used for cell disruption. Preferred cell lysis method includes hypotonic lysis and proteinase K lysis.

[067] Next, the single cell lysate is divided equally into two tubes. To ensure that a haploid copy for any given chromosomes is collected, multiple single cell lysates and corresponding splits are collected. In one embodiment, 4-12 single cell lysates are collected. In other embodiment, 6-10 single cell lysates are collected. In yet other embodiment, 8 single cell lysates are collected.

[068] As shown in Figure 6, a diploid cell contains two copies of each chromosome (one copy from father, one copy from mother). When a solution containing two copies of a chromosome is equally aliquoted into two tubes, the two copies may both go to tube 1 or tube 2, or they may each enter a different tube. The pattern of the chromosome presence in these two splitting tubes can be easily monitored. If one tube does not contain this chromosome, the other tube must contain both copies of the chromosome. If both tubes contain this chromosome, then they must contain one copy in each.

[069] For one split operation, the probability of obtaining a haploid copy for any given chromosomes is:

$$\textit{Success Probability} = \textit{Failure Probability} = \tfrac{1}{4} + \tfrac{1}{4} = 0.50.$$

[070] If n single cells are collected, and the splits describe above are performed n times (one for each single cell), then the probability that none of these tube pairs has a haploid copy for a given chromosome (all tubes are either diploid or aploid with regard to this given chromosome) is:

$$\textit{Failure Probability} = 1/2^n.$$

$$\textit{Success Probability} = 1 - 1/2^n.$$

[071] Therefore, if 8 single cells are collected (*i.e.*, n=8) and split into 16 aliquots, the likelihood of obtaining a haploid copy of any particular chromosome in these splits is:

$$1 - 1/2^8 = 0.9961.$$

[072] This means that there is a 99.61% chance that at least one of these 16 split tubes will contain a haploid copy of the target chromosome. Accordingly, if the sample size is 1000 human individuals, with 8 cells collected from each individual, in the first round of

split operation, 996 individuals will successfully obtain a haploid copy of any chromosomes for molecular haplotyping.

[073] After the split, one tube may contain a haploid copy for one chromosome; however, it may contain two copies of another chromosome. If one tube contains a haploid copy of chromosome A, two copies of chromosome B and no chromosome C, this tube can still be perfectly used for subsequent analysis (such as haplotype determination) on chromosome A. The presence of two copies of chromosome B and the absence of chromosome C will not interfere with the results on chromosome A.

[074] There is an extremely rare case in which the haplotypes from one single somatic cell do not represent the haplotypes of the same individual. This rare case is mitotic crossover which occurs in somatic cells. It is known that mitotic crossover may occur in some asexually reproducing fungi and in human cancer cells. Therefore, it is necessary to take cautions and obtain multiple cells for haplotyping a subject with cancer. In fact, this case can be easily detected by the single cell split strategy.

[075] Specifically, if a cell from an individual contains more than two copies of a certain chromosome, the presence of that chromosome in both split tubes will not be an indication of haploid chromosome in each tube. For example, if there are 3 copies of a chromosome, when both tubes contain this chromosome, one tube will have one copy, the other tube will have two copies. The tube with two copies will show heterozygous genotypes at some polymorphic sites. Therefore, our method may detect the copy number polymorphisms.

[076] The genomic DNA in each tube is then amplified for genotyping. Any methods of unbiased whole genome amplification (WGA) may be employed. Unlike polymerase chain reaction (PCR), which aims to amplify a specific sequence, WGA aims to amplify the entire genome without preference. Comprehensive WGA requires faithful replication of 3 billion bases without the loss or distortion of any particular loci or alleles.

[077] Examples of such methods include, but are not limited to, multiple displacement amplification (MDA, GE Healthcare GenomiPhi and Qiagen Repli-g), primer extension preamplification (PEP), improved primer extension preamplification (iPEP), degenerate-oligonucleotide-primed PCR (DOP, Sigma GenomePlex). The current WGA methods in the market are different on (i) amplification power and yield; (ii) fidelity; (iii) amplification product length; (iv)scalability and (v) the ability to amplify small amounts of starting material, including single cells. For example, Repli-g and GenomiPhi yield products around 10 kb in size, whereas the Sigma GenomePlex yields products around several hundred base pairs. Because the distance that the molecular haplotyping method of the present invention can resolve does not rely on the length of amplification product, the length feature of WGA method is not a critical feature for the present invention. Instead, the amplification power and potential allele bias and locus bias are critical to the present invention.

[078] The feasibility of amplifying from haploid chromosomes has been previously well-demonstrated by the genetic research using human sperms. The ability to genotype single sperm was first reported in 1988 (Li HH, et al., *Nature* 335: 414-417, 1988). Now genotyping on the DNA samples from single sperm cells (haploid) has been widely used by forensic scientists (Di Martino D, et al., *Forensic Sci Int* 146 Suppl: S151-153, 2004). It has been shown that up to 10.4 pg of DNA can be amplified by TaqGold DNA polymerase for a

reliable STR (short tandem repeats, a type of genetic polymorphisms in parallel to SNP) profile. In addition to single sperm cells, amplification of single lymphocytes (diploid) or single blastomeres (diploid) using the WGA-multiple displacement amplification (MDA) method has been successfully carried out.

[079] A common concern of WGA is the allelic drop-outs at heterozygous loci, which is consequence of a biased uneven amplification among the two alleles at one polymorphic locus. The allele drop-out will make a heterozygous individual to display as a homozygous individual in genotyping. Accordingly, when a homozygous genotype was observed, it is still theoretically possible that this is false-homozygous genotype; it may come from allele drop-out in WGA. When using the cell-split method for molecular haplotyping, because there is a single copy of haploid, if there is no locus bias at a particular locus, the allele reading in the subsequent genotyping assays will represent the allele on that corresponding haplotype. Therefore, there will be no concern necessary to distinguish allelic dropout and true homozygotes.

[080] As discussed above, if 16 split tubes from 8 single cells are collected from one individual, there is a 99.6% that some tubes contain a haploid copy of any given chromosome. Other tubes may contain diploid copies or no copy (aploid) of the same chromosome. Each tube may contain haploid copies of certain chromosomes, diploid copies of other chromosome(s), and no copy of yet other chromosome(s). Therefore, it is necessary to create a catalogue for each tube about its content for subsequent analysis, including haplotype determination.

[081] The cataloguing can be done by any methods that can detect the DNA existence. For example, PCR can be used to detect the existence of DNA fragments. If the PCR is designed to cover a sufficient number of regions that representative all of the chromosomes of genome, then it can be used to create a genome wide catalogue. PCR can be also designed to cover only the target genome region of a research project. Besides, whole genome tiling array can be also used to create this catalogue, but in a more systematic and high-throughput pattern.

[082] If both tubes of a split pair contain chromosome A, then both tubes have a haploid copy of chromosome A for haplotype determination. The tubes with a haploid copy of any particular chromosomes will be selected based on this catalogue for subsequent analysis. The samples from this procedure can be used as regular DNA samples and directly subjected to various high-throughput genotyping assays. In haplotype determination, the genotypes from haploid samples will be compared with the diploid samples from the same individual as quality control. Any false-haploid tubes will be easily detected by this comparison.

[083] Another embodiment of the HaploDissection method allows for determination of haplotype of certain chromosomes from a single cell. The method comprises: isolating a single diploid cell from the subject, lysing and staining the isolated single diploid cell to display chromosomes, collecting a set of chromosomes from the single cell by laser microdissection, genotyping genomic DNA in the collected chromosomes, genotyping genomic DNA from another intact single cell from the same subject, determining haplotype of a chromosome in the collect set of chromosomes, wherein said chromosome is present in the haploid form in said collected set of chromosomes.

[084] As shown in Figure 7, when a cell is cut along the dotted line and the right half of the cell is collected, the collected chromosomes include only a single copy of chromosomes 2, 3, and 5 (haploid) two copies of chromosome 1 (diploid) and no copy of chromosome 4 (aploid). Thus, the genotype calls from conventional genotyping platforms with this half cell will directly return the haplotypes for chromosomes 2, 3, and 5, whereas it will be still diploid genotypes for chromosome 1, and no genotype calls for chromosome 4.

[085] As discussed earlier, for each given chromosome, there is a 50% chance that a haploid copy is collected if the collected a set of chromosomes contain about half of the total number of chromosomes in the single diploid cell. The probability increases significantly when multiple sets of chromosomes are collected from multiple single cells. For example, there is a greater than 99.6% chance that a haploid copy of a given chromosome is collected if 8 sets of chromosomes are collected from 8 single cells and each collected set of chromosomes contain about half of the total number of chromosomes in the single cell. Therefore, the haplotyping of the complete set of chromosomes from a subject may be achieved with high probability (greater than 99.6%) using 8 half cells.

[086] Accordingly, in a related embodiment, multiple single diploid cells are isolated and lysed. Multiple sets of chromosomes are collected; each set is collected from a different single cell. The number of collected sets is large enough so that each chromosome in the genome of the subject is present in the haploid form at a probability greater than 99%.

[087] The HaploDissection method described above applies to any eukaryotic organism. In certain embodiments, the subject is an animal or a plant. In other embodiments, the subject is a mammal. In yet other embodiments, the subject is a human.

[088] Another aspect of the present invention relates to a computer-readable medium having computer-executable instructions for performing the methods of the present invention.

[089] Another aspect of the present invention relates to an assay kit for HaploDissection. In one embodiment, the assay kit contains reagents for cell collection, cell lysis and, optionally, cytogenetic staining, as well as reagents for genomic DNA amplification and genome genotyping. In another embodiment, the kit further includes a computer readable medium having computer-executable instructions for determining haplotype based on genotyping data. In another embodiment, the kit further includes an instrument specifically designed for collecting a set of chromosomes from single or a few cells for haplotype reading and chromosome biology examinations.

[090] In one embodiment, the HaploDissection methods are used in prenatal diagnostics to detect the important genotype defects of the fetus. It has been known that some fetus cells are circulating in the maternal peripheral blood. Therefore, the fetus cells can be collected from the pregnant maternal blood. These cells can be subjected to the haplotype analysis using the procedure described above. Because it is usually the haplotypes (the combination of alleles of different genotypes) that cause the diseases, the prenatal diagnosis by haplotype determination will be more accurate than genotype determination. The single cell nature of the present invention provides the feasibility of haplotype determination with fetus cells in the mother's blood.

[091] In another embodiment, the HaploDissection methods are used in personalized medicine. Personalized medicine is the practice that doctors customize treatment based on a person's specific genetic variations. For instance, two people who take

the same anti-hypertension medication may have very different responses. One may have severe, even life-threatening side effects, while the other experiences few if any side effects and seems to sail through the treatment. The reason why the two people have such drastically different reactions to the same medication resides in their genes. People inherit variations in their genes, and even slight variations can have a profound effect on which subtypes of the same disease the person has and how the person responds to certain medications.

[092] In a personalized medicine, the current process in a clinic began to change. Before a patient takes a single dose of medication, the patient may have a blood test done to determine genetic variations. The test may show that patient's variation which is likely to have an adverse effect on the particular medication. The doctors can determine the drug prescriptions and doses to match the patient's genetics. Therefore, the unique genetic profile can help doctors to personalize treatments of patients, improve the drug development, and reduce healthcare costs.

[093] It has been widely accepted now that multi-SNP haplotypes are more accurate to represent a person's genotype than single-SNP genotype. However, there is no simple, cheap and high-throughput experimental method to directly read the haplotypes. The statistical haplotype configuration causes many ambiguities. This technical bottleneck is not only limiting the efforts to discover the genetic basis underlying the common diseases, it is also limiting the application of genetic tests in clinical practice. The HaploDissection methods may solve this technical bottleneck.

[094] For example, before a patient takes any medication, a few cells will be collected from his mouth and haplotypes on those disease mutations will be determined by using the present methods. Doctors will prescribe a drug with a certain dose to match a patient's unique genetic profile to personalize the treatment.

[095] In yet another embodiment, the HaploDissection methods are used in forensic testing. True haplotyping provides a greater precision than single SNP genotyping in forensic studies, in any case of sexual assault or other crimes, as well as paternity testing. In many cases of forensic tests, the available amount of a specimen is usually quite limited. Because of the single cell nature of our invention and true haplotype result out of this technology, the HaploDissection method will increase both sensitivity and precision.

[096] The present invention is further illustrated by the following examples, which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and Tables, are incorporated herein by reference.

EXAMPLE 1: PREPARATION OF CELLS

I. Sample Collection

[097] Collect blood from human individuals and isolate the lymphocytes.

II. Cell Culture

[098] Culture the lymphocytes in RPMI1640 medium containing 15% FBS and 100 unit/ml Penicillin/Streptomycin.

III. Cell Lysis

[099] 1. At the proliferation stage, add phytohemagglutinin to the cell culture.

[0100] 2. Harvest cells 48 hours after phytohemagglutinin (PHA) treatment.

[0101] 3. Add Ethidium Bromide (16.7 ug/ml) and Act-D (6.7ug/ml) into cells.

[0102] 4. Incubate at 37 °C for 0.5 hour. After 0.5 hour, add colcemid (0.083ug/ml) into cells and incubate at 37 °C for 1 hour.

[0103] 5. Centrifuge at 1000 rpm for 10 min.

[0104] 6. Aspirate all but 0.3 ml supernatant, gently resuspended cell pellet. Add pre-warmed 0.075mol/L KCl vortex gently to make sure KCl is mixed well with the pellet.

[0105] 7. Leave at 37 °C for 20 min and room temperature for additional 5 min. Centrifuge at 1000 rpm for 10 min. And remove the supernatant.

[0106] 8. Add cold fixative (methanol: acetic acid, 3:1), gently mix by inverting tubes.

[0107] 9. After fixed and centrifuged for three times, cells were dropped on slide and geimesa staining for 20 minutes.

[0108] 10. Air dry slide for 20 min in hood.

IV. Chromosome Isolation

[0109] 1. Turn on the laser, microscope (Leica, ASLMD) and computer. Put PCR tubes in collectors on holders.

[0110] 2. Put slide on supporter.

[0111] 3. Go to the computer screen. Click on LEICA ADMINISTRATOR to open program.

[0112] 4. Set objective 10X to find cells and then switch to 40X.

[0113] 5. Collect chromosomes: cut and pick up no more than 30 chromosomes at random from each cell. Collect 4-8 samples. Each sample contains chromosomes from about 7~11 cells.

[0114] 6. Exit LEICA ADMINISTRATOR program.

[0115] 7. Turn off the computer, microscope and laser by order.

EXAMPLE 2: WHOLE-GENOME DNA AMPLIFICATION

Single Cell Lysis and Fragmentation

[0116] 1. Isolate a single cell into a PCR-ready vessel using laser capture micro-dissection, cell sorting, or other method. If sorted, the buffer should be of low ionic strength, such as Tris EDTA (TE) buffer, and in the minimal sort volume.

[0117] 2. Add a sufficient volume of water to the single cell sample for a final volume of 9 mL.

[0118] 3. Prepare a working Lysis and Fragmentation Buffer Solution by adding 2 mL of Proteinase K Solution into 32 mL of the 10 Single Cell Lysis & Fragmentation Buffer and vortex thoroughly.

[0119] 4. Add 1 mL of the freshly prepared Proteinase K Solution-10′ Single Cell Lysis & Fragmentation Buffer to the single cell sample and mix thoroughly.

[0120] 5. Incubate DNA mix at 50 °C for 1 hour, then heat to 99 °C for EXACTLY four minutes. Note that the incubation is very time-sensitive and any deviation may alter results. Cool on ice. Spin down sample prior to proceeding to Library Preparation.

Library Preparation

[0121] 6. Add 2 mL of 1 Single Cell Library Preparation Buffer to each sample.

[0122] 7. Add 1 mL of Library Stabilization Solution.

[0123] 8. Mix thoroughly and place in thermal cycler at 95°C for 2 minutes.

[0124] 9. Cool the sample on ice, consolidate the sample by centrifugation, and replace on ice.

[0125] 10. Add 1 mL of Library Preparation Enzyme, mix thoroughly, and centrifuge briefly.

[0126] 11. Place sample in a thermal cycler and incubate as follows:

[0127] 16 °C for 20 minutes; 24 °C for 20 minutes; 37 °C for 20 minutes; 75 °C for 5 minutes; and 4 °C hold.

[0128] 12. Remove samples from thermal cycler and centrifuge briefly. Samples may be amplified immediately or stored at –20 °C for three days.

Amplification

[0129] 13. Add the following reagents to the entire 14 Ml reaction:

[0130] 7.5 mL of 10′ Amplification Master Mix; 48.5 mL of Nuclease-Free Water; and 5.0 mL of WGA DNA polymerase.

[0131] 14. Mix thoroughly, centrifuge briefly, and begin thermocycling. The following profile has been optimized for a PE 9700 or equivalent thermal cycler:

[0132] Initial Denaturation at 95 °C for 3 minutes.

[0133] Perform 35 cycles as follows:

[0134] Denature at 94 °C for 30 seconds; anneal/extend at 65 °C for 5 minutes; and hold at 4 °C.

[0135] After cycling is complete, maintain the reactions at 4 °C or store at –20 °C until ready for analysis or purification. The stability of WGA DNA is equivalent to genomic.

[0136] DNA stored under the same conditions.

EXAMPLE 3: WHOLE GENOME GENOTYPING

[0137] The amplified DNA will be subjected to Illumina high-throughput whole genome genotyping such as Hap3000K and others. For each person, 2-4 tubes of samples will be subjected to the genotyping to ensure high genome coverage and achieve duplications for accuracy. A genomic DNA sample will be included for whole genome genotyping. This step can be done also by using other high-throughput genotyping platforms.

EXAMPLE 4: HAPLOTYPE DETERMINATION

[0138] Whole-genome genotyping data was obtained from Illumina HumanCNV370-Duo BeadChip. This BeadChip content covers over 370,000 markers using the Infinium® Assay. After scanning, all the data uploaded into BeadStudio and analyzed using the BeadStudio Genotyping Module, version 3. After stringent filtering with removal of SNPs with missing genotype, the remaining SNPs were available for analysis. The values of theta, R, X and Y in the Illumina genotyping output are used to determine the relative ratio of two alleles of each SNP. Haplotypes are constructed by the allele ratio along the chromosomes.

[0139] The haplotype construction is done by using the software "HapReader," which is specifically developed for this technology. The essential procedure and algorithm is below:

[0140] 1) The analysis will be done at an individual level, person by person. There will be no combination of datasheet from different individuals.

[0141] 2) Each person will have 3-5 genotyping datasheets, one if from the genomic DNA, and the others are from the chromosome collection tubes. Extract the allele calls and their corresponding signal intensities from each Illumia output data sheet.

[0142] 3) Based on the datasheet from the genomic DNA sample, select the homozygous loci for this person. Calculate the relative ratio of the averages of the A, C, G, and T at these loci. Determine a k value for A, C, G, T to adjust their intensities to the same level for a given particular experiment.

[0143] 4) Using these k values, adjust the heterozygous loci.

[0144] 5) For each locus, calculate the ratio of the two alleles.

[0145] 6) For each locus, sort the order of those two alleles by their allelic intensity ratio. Sort all loci by the same way, and keep the higher-intensity-allele on column-A and the lower-intensity-allele on column-B.

[0146] 7) Examine and compare the ratio values along the chromosome to determine if there is breakpoint in each chromosome.

[0147] 8) If not at step 7), the alleles in column-A form haplotype, and the alleles on column-B will form another haplotype for this person. If yes at step 7), use the results from other chromosome collection tubes to bridge over this breakpoint.

[0148] One aspect of the present invention lies under this step. Collect chromosomes into PCR tubes by microdissection using a Leica ASLMD Laser microdissection system. In this step, not all chromosomes are collected from one cell; instead, only part (around half) of the chromosomes is collected from any single lyzed cell (Figure 2). The selection of chromosomes on any lyzed cells is random. This random collection is respected from 5-11 randomly-selected cells. All of the chromosomes from these 5~11 microdissections are collected into one tube. 4-8 tubes are collected for each person. In this step, the chromosome integrity is preserved by selecting the laser cutting line on the computer, so chromosome integrity is ensured.

[0149] Amplify the collect DNA sample in each PCR tube by using the Sigma GenomePlex WGA-4 kit for 20-24 cycles. In fact, this step can be done with any methods of unbiased whole genome amplification (WGA). These methods include, but are not limited to, multiple displacement amplification (MDA, GE Healthcase GenomiPhi and Qiagen Repli-g), primer extension preamplification (PEP), improved primer extension preamplification (iPEP), and degenerate-oligonucleotide-primed PCR (DOP, Sigma GenomePlex). Repli-g and GenomiPhi yield products around 10 kb in size, whereas the Sigma GenomePlex yields products around several hundred base pairs.

EXAMPLE 5: DETERMINATION OF HAPLOTYPE USING SINGLE CELL LYSATE METHOD

[0150] A Leica AS LDM Laser Microdissection system (Leica Microsystems, Germany) is used to isolate single cells from fresh cytobrush-swab buccal cells from human individuals. Briefly, buccal cells are smeared on a foiled slide (a rectangular UV-sliceable piece of foil fixed at the margins to a normal microscope slide), air-dried for 5 min and then stained very briefly. The section is reviewed under microscope, single cells are selected and cut-off from the foiled slide with the Laser Microdissection system. Single cells are collected into tubes with 10-ul of cell lysis buffer in each tube. The single cell lysate in each tube is then divided equally into two tubes. The genomic DNA in each tube is amplified WGA for genotyping. A genome wide catalogue is created using the genotyping data. Haplotype determination is carried out using the catalogue.

EXAMPLE 6: DETERMINATION OF HAPLOTYPE USING SINGLE CELL DISSECTION METHOD

[0151] *Chromosome Microdissection*: Lymphocytes were cultured in RPMI1640 medium containing 15% FBS and 100 unit/ml Penicillin/Streptomycin. Cells were stimulated by phytohemagglutinin (PHA) for 48 hours, followed by addition of Ethidium Bromide (16.7 ug/ml) and actinomycin D (6.7ug/ml) and incubation at 37 °C for 30 min. Colcemid (0.083ug/ml) was added into cells and incubate at 37 °C for 1 hour. Cells were collected by centrifugation at 1,000 rpm for 10 min, resuspended, incubated in pre-warmed 0.075mol/L KCl at 37 °C for 20 min, and then, at room temperature for 5 min. After fixation with cold fixative (methanol: acetic acid, 3:1), cells were dropped onto slide to break the nuclei followed by Giemsa staining for 20 min. Laser Microdissection Microscope (ASLMD, Leica, Germany) was used to collect half of the chromosomes of one cell.

[0152] *Whole Genome Amplification (WGA)*: The collected chromosomes were amplified by the Sigma GenomePlex WGA4 kit following the manufacturer's protocol. Briefly, the sample was incubated in the Lysis and Fragment Buffer at 50°C for 1 hour, and then heated to 99 °C for 4 min. Then the Single Cell Library Preparation Buffer and Library Stabilization Solution was added into the sample followed by an incubation at 95°C for 2 min. Library was prepared with the following cycles: 16°C for 20 min, 24 °C for 20 min, 37 °C for 20 min, and 75 °C for 5 min. DNA was amplified by an initial denaturation at 95 °C for 3 min followed by 35 cycles of 94°C/30 sec and 65°C/5 min. Amplified DNA was purified by QIAquick PCR purification kit.

[0153] *Genotyping:* The Illumina HumanCNV370-Quad BeadChip was used for genotyping. This BeadChip contains ~370,000 markers including SNPs and copy number variation (CNV) markers. Three independent microdissected samples and one genomic samples extracted with a Qiagen kit were subjected to genotyping experiments. After scanning, the data was uploaded into the BeadStudio and analyzed using the BeadStudio Genotyping Module version 3. No-call threshold was set at default (0.15).

[0154] *Data analysis:* Unphased genotypes of GM10847 and his parents (GM and GM) were retrieved from the International HapMap Project database (Phase 2 Public Release #22, and Phase 3 Public Release #1, Phase 2+3 Release #27). The unphased genotypes of GM10847 were also retrieved from the Illumina database. Haplotypes of GM10847 was computationally reconstructed with his parental genotypes by determining the parental origin of each allele following the Mendelian Law of Inheritance. In the data analysis, only those heterozygous loci of GM10847 were subjected to haplotype determination. The homozygous loci were removed because they do not have the haplotyping issue (phase-known). Allele calls with both allele intensities below 1,000 in the Illumina genotyping output were removed. Genome-wide RepeatMasker detection was retrieved from the UCSC Genome Browser (Human 2006 March Assembly). All data integration was performed with SAS9.1.

[0155] The haplotyping method was tested with the individual GM10847 recruited in the HapMap project (The International HapMap Consortium 2003) by three independent experiments. Following the procedures described above, we the genotype calls of microdissected samples were compared with genotype calls of genomic DNA as well as of data downloaded from the International HapMap Project database (Phase 2 Public Release #22). The monosomic, disomic and null states of each chromosome in each sample were

indicated by whether the chromosome-wide heterozygous calls were converted to homozygous calls in the microdissection samples (Figure 8). It was found that that sample 1 successfully haplotyped chromosomes 2, 4, 6, 15, 16, 17, 18, and 20; sample 2 haplotyped chromosomes 1q, 3, 4, 5, 10, 16, 17, 18, 20 and 21; sample 3 haplotyped chromosomes 3, 7, 9, and 20. Totally 24,481 heterozygous loci were phased.

[0156] The accuracy of this method was determined by replications and comparison with haplotypes resolved from unphase genotypes (HapMap Phase2 Rel#22) using the trio structure under Mendelian Law of Inheritance (Hodge SE, et al., Nat Genet. 21(4):360-1; 1999). Among those 24,245 SNP loci that were successfully phased by our 7DDNA haplotyping method, 464 SNP loci were not covered by the HapMap Phase 2 genotype data, 4,744 SNPs do not have unambiguous haplotypes from the HapMap genotypes due to all-three-heterozygote, and 142 SNPs were not phased due to missing data in HapMap2. So we compared 18,895 SNP loci between our haplotypes and HapMap2 derived haplotypes. There were 18,625 SNPs (98.57%) that showed consistent allele phase as compared with haplotypes resolved by HapMap trio structure. Among those 270 discordant SNP loci, 45 SNP loci were due to the HapMap Phase 2 genotyping error as compared with phase 3 genotype calls, and 103 loci were on various repeats as detected by RepeatMasker. The other discordance may be potentially ascribed to whole genome amplification errors, genotyping errors, or un-annotated segmental duplications besides those identified by RepeatMasker. It was further determined the accuracy directly by 2,089 replications, among which 2,065 SNPs showed consistent result, none of them had inconsistent haplotype, and 24 SNP loci had diploid allele calls in one of the duplicates although the entire chromosome showed discordance (Table 1), with an estimate of 98.85% as the accuracy rate.

Table 1. An estimate of accuracy rate by data reproductivity.

Chr	Total repeated SNPs	Consistent	Inconsistent	Accuracy%
Chr3	517	516	1	99.81
Chr4	565	554	11	98.05
Chr16	189	186	3	98.41
Chr17	217	215	2	99.08
Chr18	212	208	4	98.11
Chr20	389	386	3	99.23
Total	2,089	2,065	24	98.85

[0157] This haplotyping method does not have apparent limitations on the phasing distance, total SNP number, and marker types. The procedure is simple and inexpensive; it does not require a complicated optimization of experimental condition and the cost is close to conventional high throughput genotyping assays. In addition, there is no apparent barrier for this approach to be amendable to automation if chromosome microdissection is automated. The method may be further improved by using better WGA method for single-cell DNA, newer versions of high-throughput genotyping chips or deep sequencers, and more specific chromosome staining, such as chromosome painting of particular chromosomes.

[0158] The above description is for the purpose of teaching the person of ordinary skill in the art how to practice the present invention, and it is not intended to detail all those obvious modifications and variations of it which will become apparent to the skilled worker upon reading the description. It is intended, however, that all such obvious modifications and variations be included within the scope of the present invention, which is defined by the following claims. The embodiments are intended to cover the components and steps in any sequence which is effective to meet the objectives there intended, unless the context specifically indicates the contrary.

WHAT IS CLAIMED IS:

1. A method for molecular haplotyping of a subject, comprising:

randomly selecting a set of chromosomes in each of a plurality of lyzed diploid cells of the subject;

collecting the selected chromosomes from said plurality of cells into a plurality of sample tubes, wherein each sample tube contains chromosomes selected from one or more cells;

genotyping genomic DNA in each sample tube; and

determining haplotype of alleles based on allele nucleotide sequence information and corresponding nucleotide signal intensities from genotyping data.

2. The method of Claim 1, wherein said determining step comprises:

extracting allele nucleotide sequence information and corresponding nucleotide signal intensities from genotyping data;

calculating the nucleotide signal intensity ratio of two alleles (allelic intensity ratio) for each heterozygous locus; and

determining haplotype of the alleles.

3. The method of Claim 2, wherein said calculating step comprises:

calculating relative ratio of nucleotides A, C, G, and T at homozygous loci;

determining a k value for each nucleotide to adjust their signal intensities to the same level;

adjusting nuleotide signal intensities at heterozygous loci using the k value; and

calculating the allelic intensity ratio at the heterozygous locus.

4. The method of Claim 3, further comprises:

sorting the order of alleles at each locus by allelic intensity ratio;

keeping the higher-intensity-allele on a first column and the lower-intensity-allele on a second column; and

determining whether there is breakpoint in each chromosome,

if there is no breakpoint in a chromosome, forming one haplotype with alleles in the first column and another haplotype with alleles in the second column,

if there is a breakpoint in a chromosome, using results from other chromosome collection tubes to bridge over the breakpoint.

5. The method of Claim 1, wherein chromosomes from 2-10 randomly selected cells are collected into each sample tube.

6. The method of Claim 5, wherein a total of 4-8 sample tubes are collected.

7. The method of Claim 1, wherein said genotyping step comprises amplifying genomic DNA.

8. The method of Claim 1, wherein said subject is a mammal.

9. The method of Claim 8, wherein said mammal is a human.

10. A method for molecular haplotyping of a subject, comprising:

isolating one or more single diploid cells from the subject;

lysing each isolated single diploid cell from the subject to generate one or more single cell lysate;

dividing each single cell lysate into two equal aliquots;

genotyping genomic DNA in each aliquot;

creating a catalogue of genotyping data from all aliquots; and

determining chromosome haplotype of the subject based on the catalogue.

11. The method of Claim 10, wherein said isolating step includes isolating 4-12 single diploid cells from the subject.

12. The method of Claim 11, wherein said isolating step includes isolating 6-10 single diploid cells from the subject.

13. The method of Claim 12, wherein said isolating step includes isolating 8 single diploid cells from the subject.

14. The method of Claim 10, wherein said genotyping step comprises amplifying genomic DNA.

15. The method of Claim 10, wherein said subject is a mammal.

16. The method of Claim 15, wherein said mammal is a human.

17. A method for molecular haplotyping of a subject, comprising:

isolating a single diploid cell from the subject;

lysing and staining the isolated single diploid cell to display chromosomes;

collecting a set of chromosomes from the single cell by laser microdissection;

genotyping genomic DNA in the collected chromosomes;

genotyping genomic DNA from one or more intact diploid cells of the same subject; and determining haplotype of a chromosome in the collect set of chromosomes, wherein said chromosome is present in the haploid form in said collected set of chromosomes.

18. The method of Claim 17, wherein said isolating step, the lysing step, and the collecting step are repeated one or more times to determine haplotype for each chromosome of the subject.

19. The method of Claim 18, wherein said isolating step, the lysing step, and the collecting step are repeated enough times so that each chromosome in the genome of the subject is present in the haploid form at a probability greater than 99%.

20. The method of Claim 17, wherein said genotyping step comprises amplifying genomic DNA.

21. The method of Claim 17, wherein said subject is a mammal.

22. The method of Claim 21, wherein said mammal is a human.

23. An assay kit for HaploDissection, comprising:

reagents for cell collection and cytogenetic staining; and

reagents for genomic DNA amplification and genome genotyping.

24. The assay kit of Claim 23, further comprises an instruction manual.

25. The assay kit of Claim 23, further comprises a computer readable medium having computer-executable instructions for determining haplotype based on genotyping data.

26. A computer-readable medium having computer-executable instructions for performing the method of Claim 1.

27. A computer-readable medium having computer-executable instructions for performing the method of Claim 10.

28. A computer-readable medium having computer-executable instructions for performing the method of Claim 17.

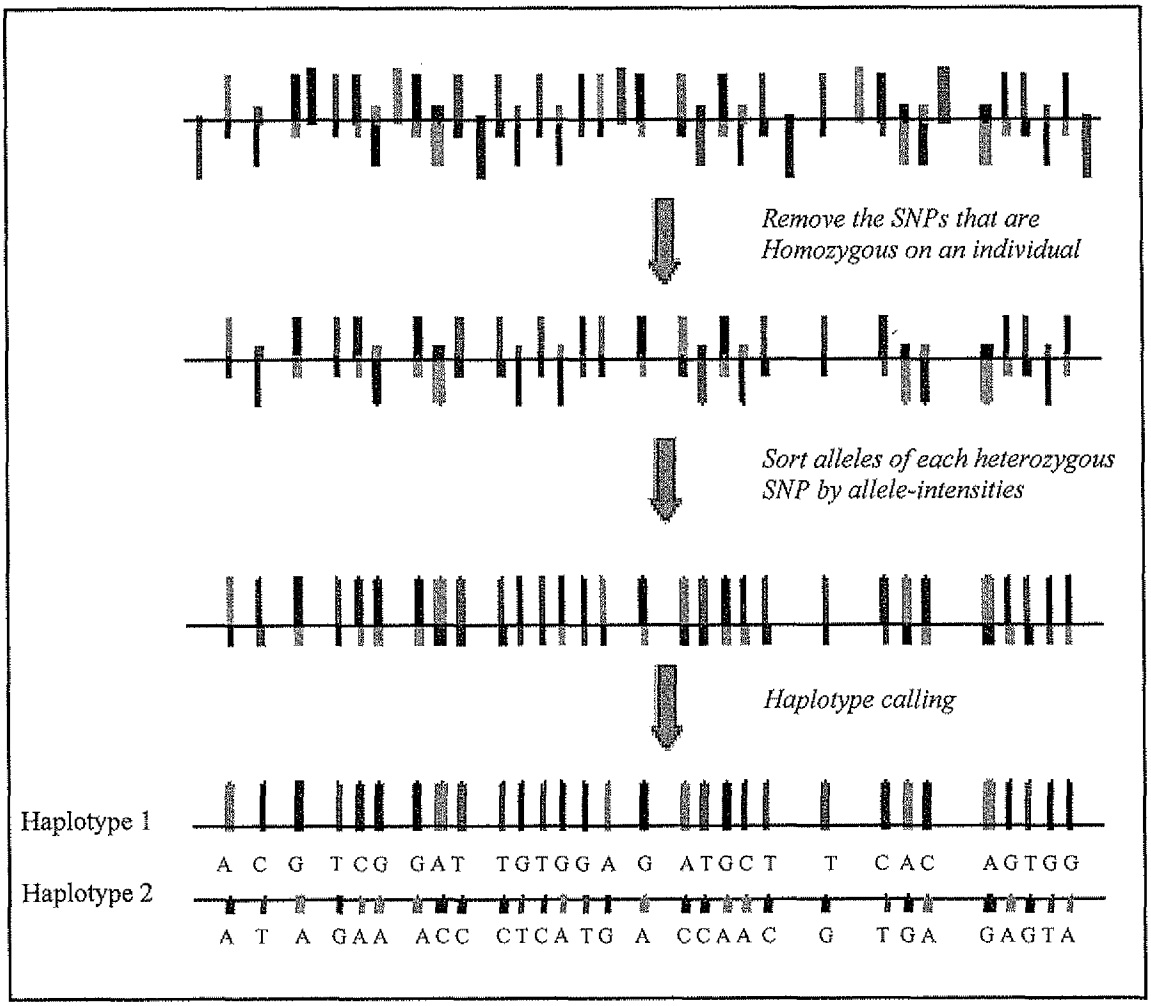

Remove the SNPs that are
Homozygous on an individual
Sort alleles of each heterozygous
SNP by allele-intensities
Haplotype calling
Haplotype 1
A C G T CG GAT TGTGG A G ATGC T T CAC AGTGG
Haplotype 2
A T A GAA ACC CTCA TG A CCAA C G TGA GAGTA

FIG. 1

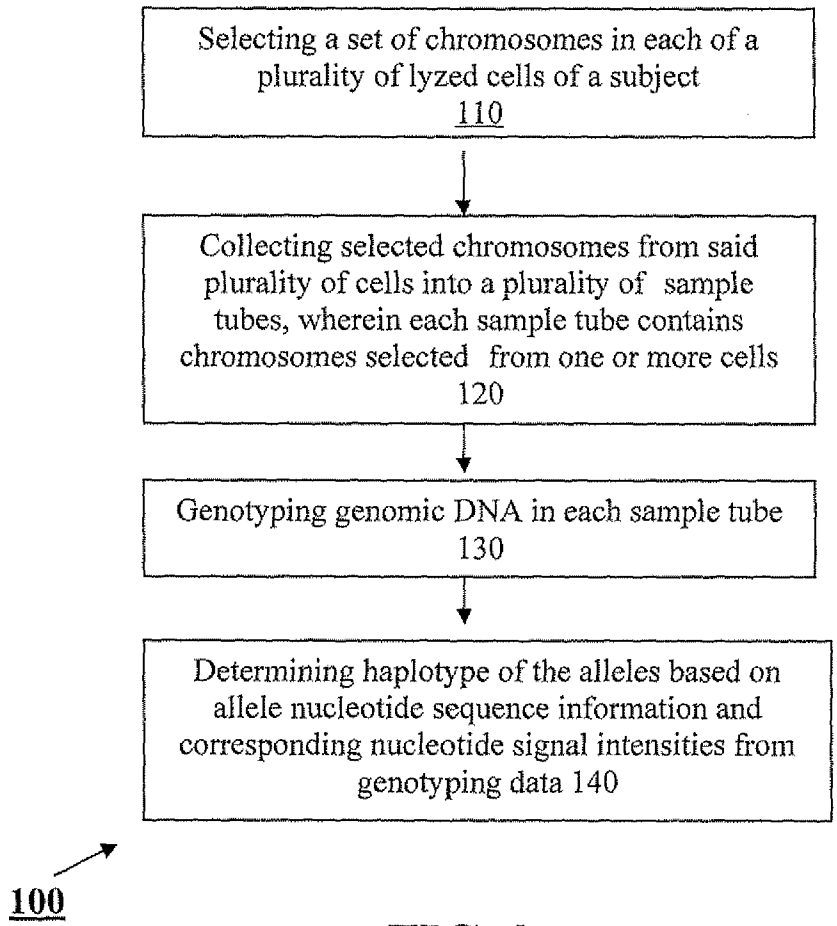

Selecting a set of chromosomes in each of a plurality of lyzed cells of a subject
110
Collecting selected chromosomes from said plurality of cells into a plurality of sample tubes, wherein each sample tube contains chromosomes selected from one or more cells
120
Genotyping genomic DNA in each sample tube
130
Determining haplotype of the alleles based on allele nucleotide sequence information and corresponding nucleotide signal intensities from genotyping data 140
100

FIG. 2

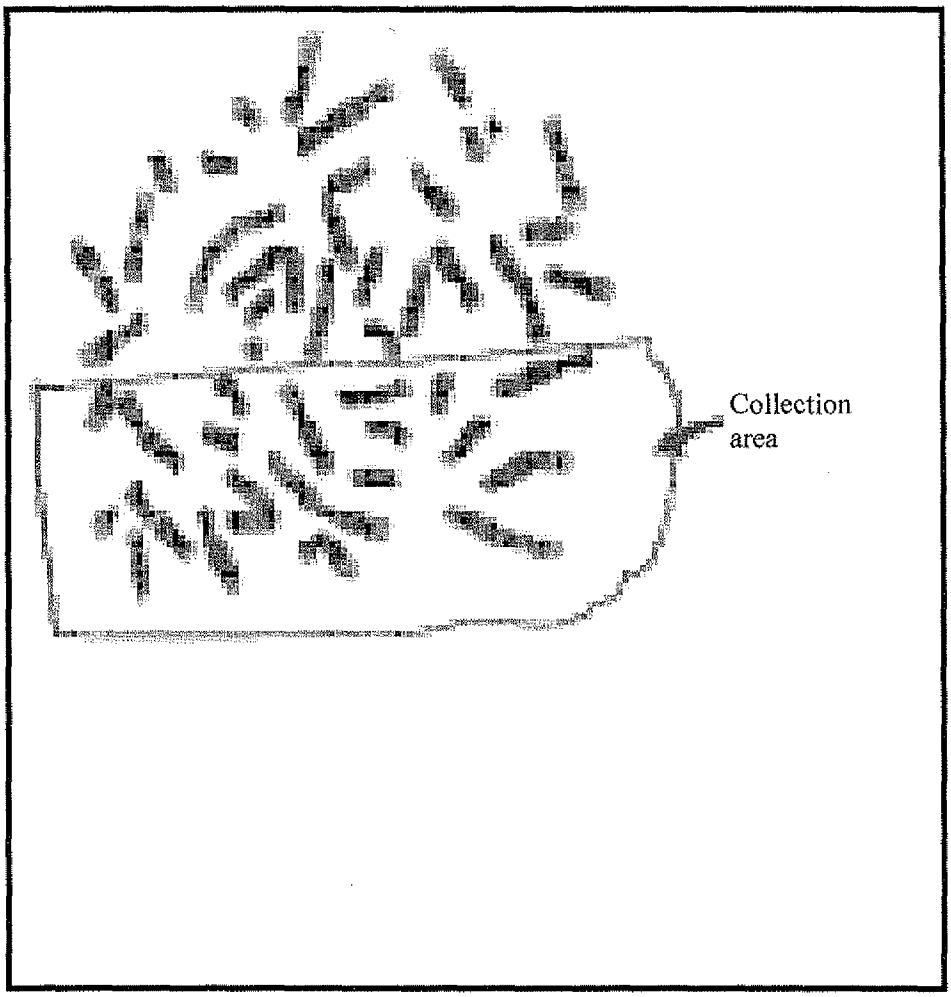


FIG. 3

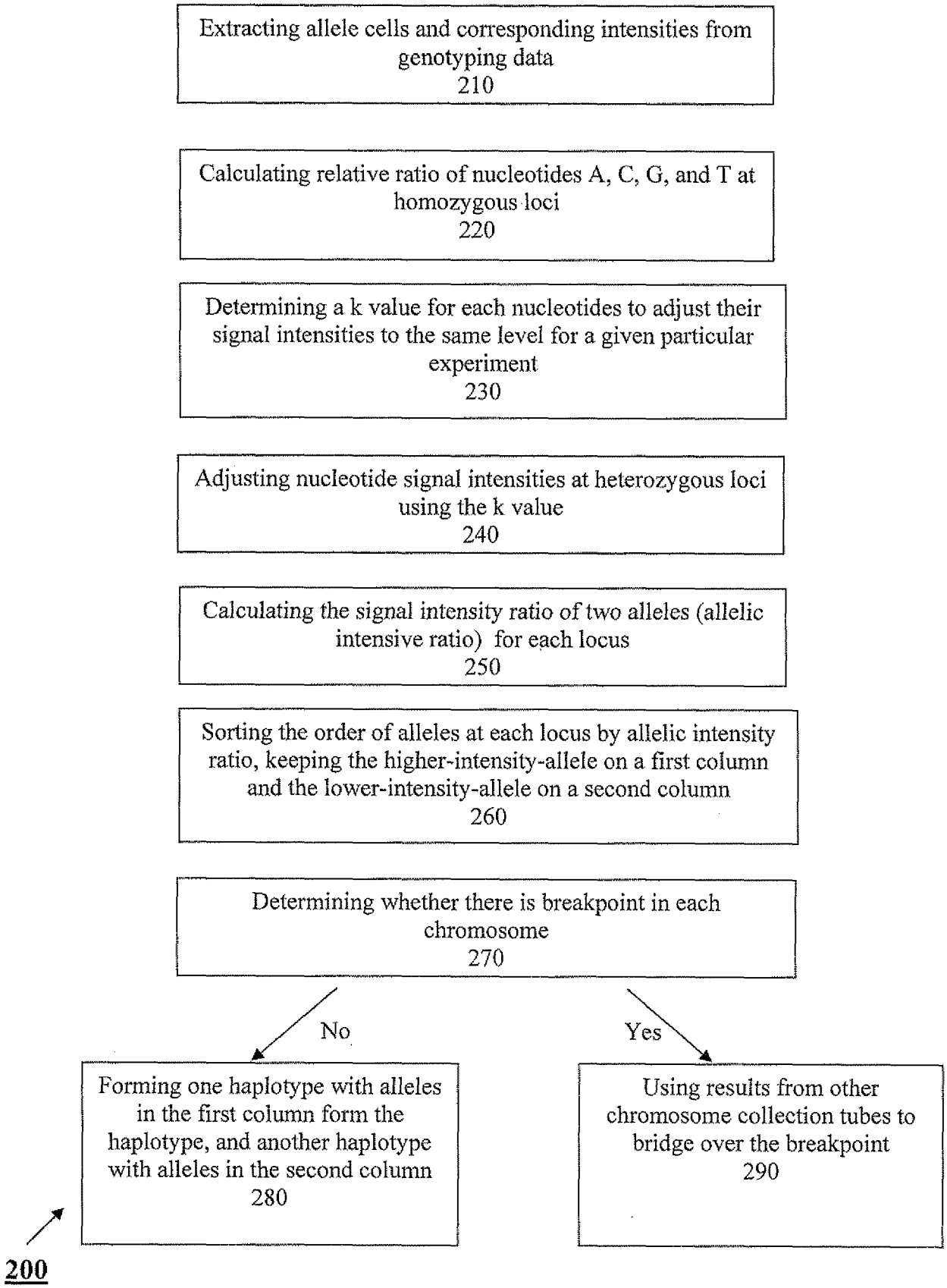

Extracting allele cells and corresponding intensities from genotyping data
210
Calculating relative ratio of nucleotides A, C, G, and T at homozygous loci
220
Determining a k value for each nucleotides to adjust their signal intensities to the same level for a given particular experiment
230
Adjusting nucleotide signal intensities at heterozygous loci using the k value
240
Calculating the signal intensity ratio of two alleles (allelic intensive ratio) for each locus
250
Sorting the order of alleles at each locus by allelic intensity ratio, keeping the higher-intensity-allele on a first column and the lower-intensity-allele on a second column
260
Determining whether there is breakpoint in each chromosome
270
No
Yes
Forming one haplotype with alleles in the first column form the haplotype, and another haplotype with alleles in the second column
280
Using results from other chromosome collection tubes to bridge over the breakpoint
290
200

FIG. 4

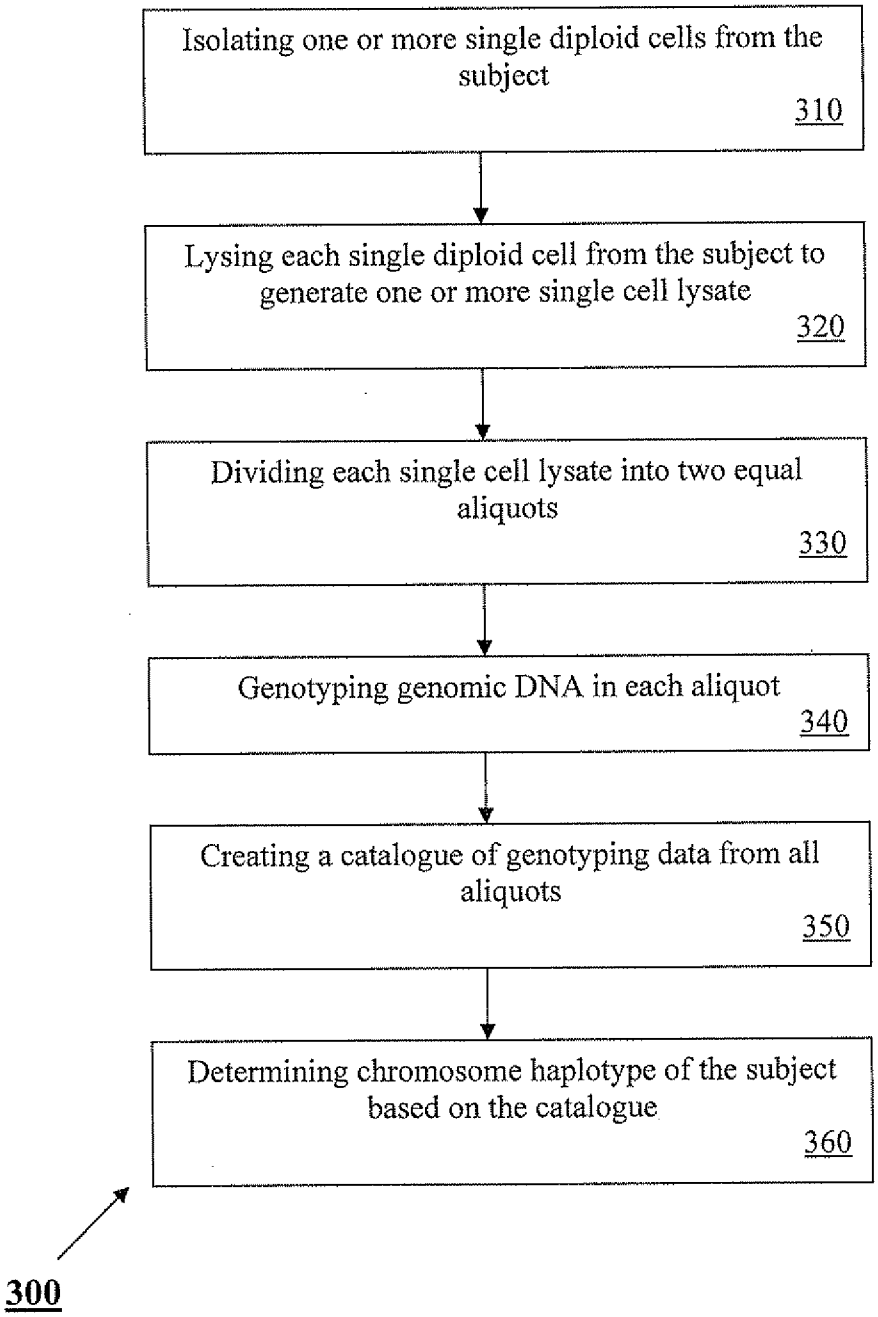

Isolating one or more single diploid cells from the subject
310
Lysing each single diploid cell from the subject to generate one or more single cell lysate
320
Dividing each single cell lysate into two equal aliquots
330
Genotyping genomic DNA in each aliquot
340
Creating a catalogue of genotyping data from all aliquots
350
Determining chromosome haplotype of the subject based on the catalogue
360
300

FIG. 5

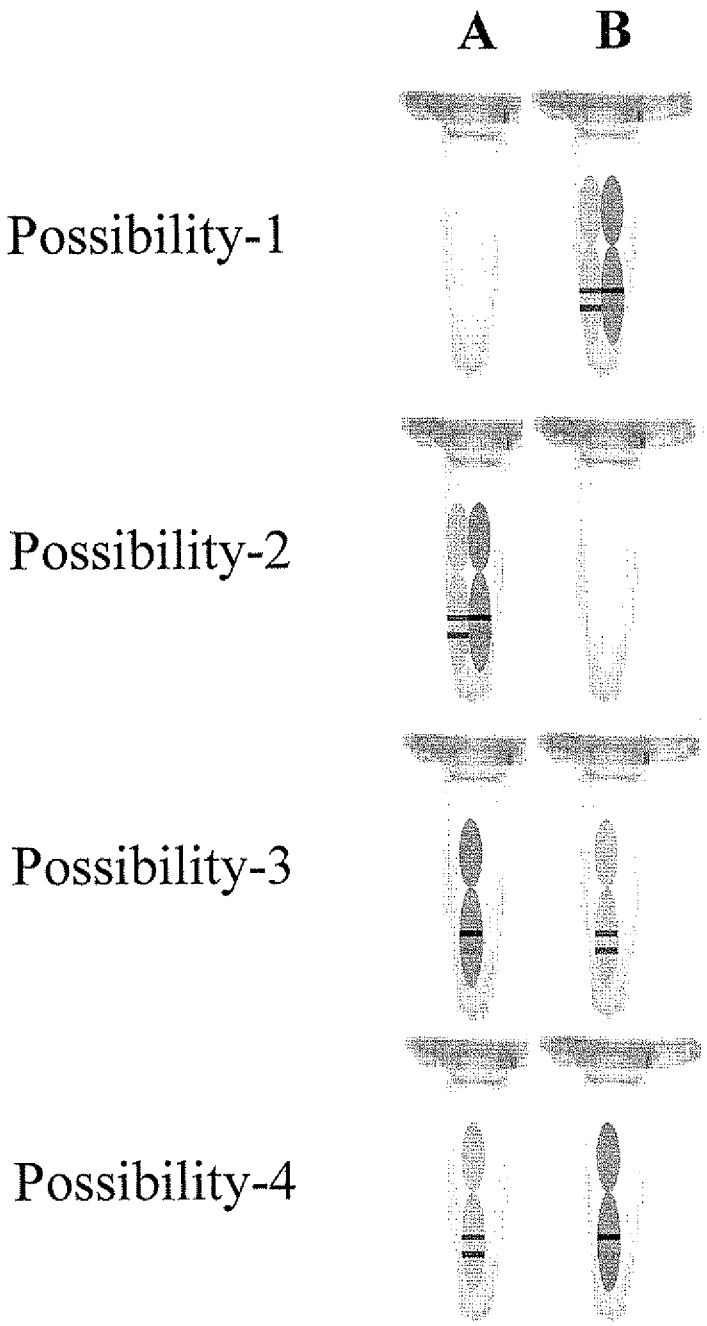

A
B
Possibility-1
Possibility-2
Possibility-3
Possibility-4

FIG. 6

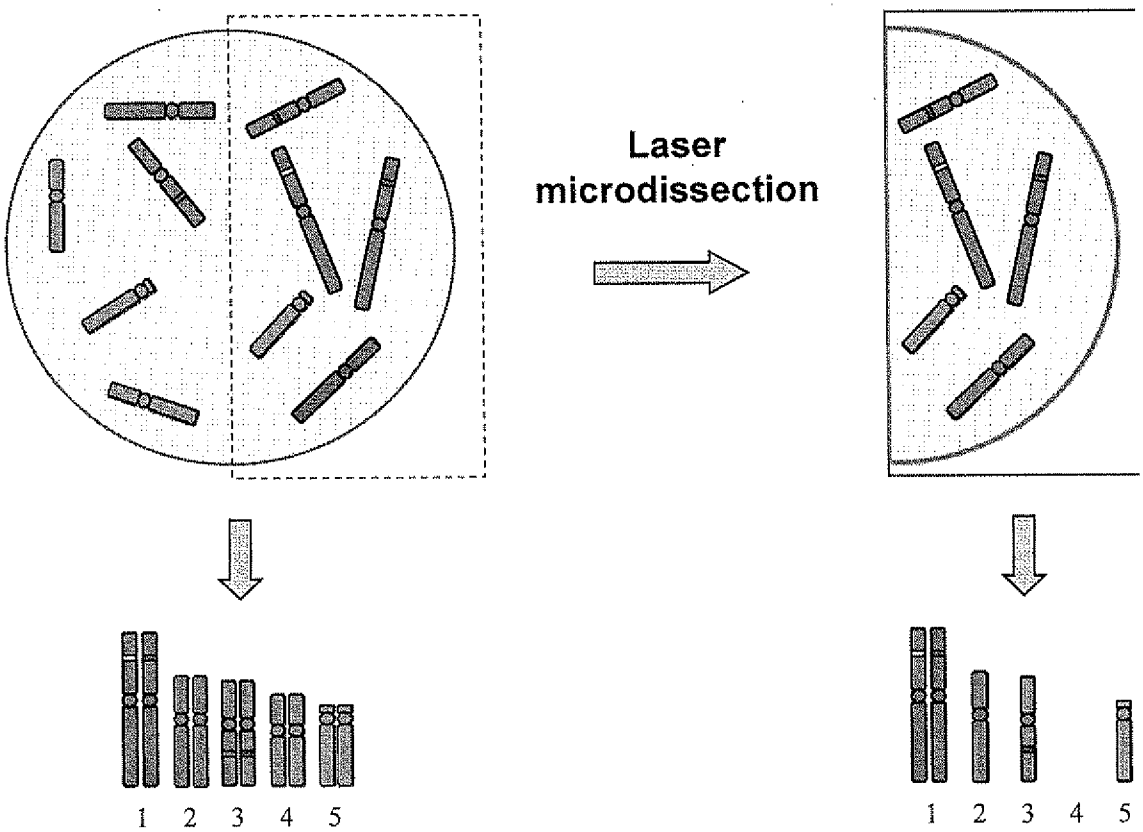


FIG. 7

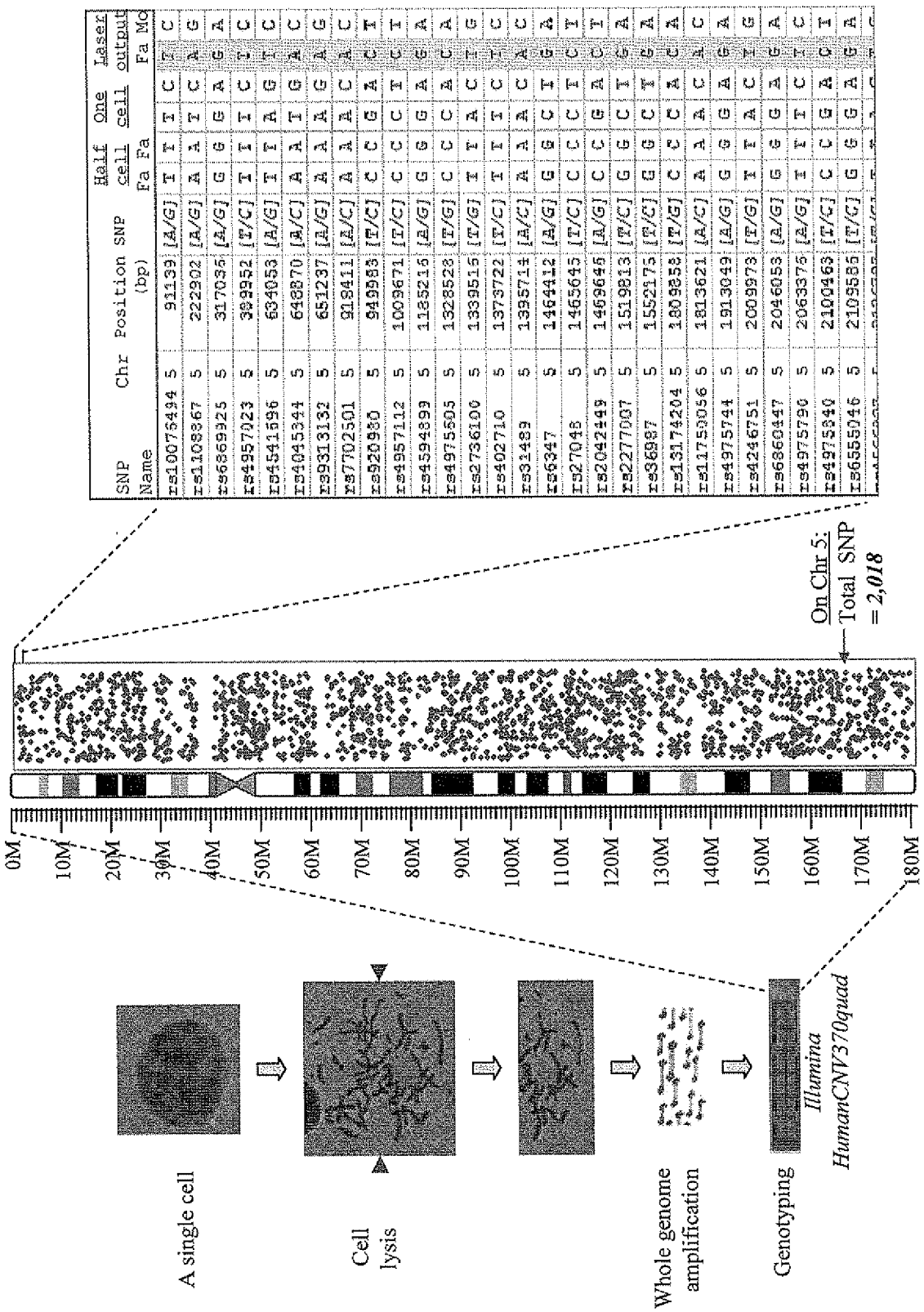

SNP Name
Chr
Position (bp)
SNP
Half cell Fa
Half cell Fa
One cell
Laser output Fa
Mo
rs10075494 5 91139 [A/G] T T T C T C
rs1108367 5 222902 [A/G] A A T C A G
rs6869925 5 317036 [A/G] G G G A G A
rs4957023 5 389952 [T/C] T T T C T C
rs4541696 5 634053 [A/G] T T A G T C
rs4045344 5 648870 [A/C] A A T G A C
rs9313132 5 651237 [A/G] A A A G A G
rs7702501 5 918411 [A/C] A A A C A C
rs520930 5 949983 [T/C] C C G A C T
rs4957112 5 1009671 [T/C] C C C T C T
rs4594899 5 1185216 [A/G] G G G A G A
rs4975505 5 1328523 [T/G] C C C A C A
rs2736100 5 1339515 [T/G] T T A C T G
rs402710 5 1373722 [T/C] T T T C T C
rs31489 5 1395714 [A/C] A A A C A C
rs6347 5 1464412 [A/G] G G C T G A
rs27048 5 1465645 [T/C] C C C T C T
rs2042449 5 1469646 [A/G] C C G A C T
rs2277007 5 1519813 [T/C] G G C T G A
rs36987 5 1552175 [T/C] G G C T G A
rs13174204 5 1809858 [T/G] C C C A C A
rs11750056 5 1813621 [A/C] A A A C A C
rs4975744 5 1913049 [A/G] G G G A G A
rs4246751 5 2009973 [T/G] T T A C T G
rs6860447 5 2046053 [A/G] G G G A G A
rs4975790 5 2063375 [A/G] T T T C T C
rs4975840 5 2100463 [T/C] C C G A C T
rs6555046 5 2109585 [T/G] G G G A G A
On Chr 5:
Total SNP
= 2,018
0M
10M
20M
30M
40M
50M
60M
70M
80M
90M
100M
110M
120M
130M
140M
150M
160M
170M
180M
A single cell
Cell lysis
Whole genome amplification
Genotyping
Illumina HumanCNV370quad

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2009/047765

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68(2006.01)i, C12Q 1/02(2006.01)i, C12N 15/10(2006.01)i

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 1/68; C12P 19/34; G01N 33/48; G01N 33/50

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

Korean utility models and applications for utility models
Japanese utility models and applications for utility models
(Chinese Patents and application for patent)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS, WPI, USPTO, PAJ, NCBI, INSPECT "haplotyping, diploid, lysing, aliquot, genomic DNA, laser microdissection, chromosome, et al."

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007-062486 A1 (SIMONS HAPLOMICS LIMITED et al.) 07 June 2007 See claims 1-20; abstract	10-22
X	US 2003-0027150 A1 (DAVID A. KATZ) 06 February 2003 See claims 1-8, 14-16, and 19; abstract	10-22
X	WO 2008-148161 A1 (SIMONS HAPLOMICS LIMITED et al.) 11 December 2008 See claims 1-11; page 9; abstract	10-22
A	US 2003-0059808 A1 (JUN S. LIU et al.) 27 March 2003 See the whole document	10-22
A	US 2004-0224331 A1 (CANTOR CHARLES R. et al.) 11 November 2004 See the whole document	10-22

☐ 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 application or patent 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 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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 FEBRUARY 2010 (18.02.2010)	**24 FEBRUARY 2010 (24.02.2010)**
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140	Authorized officer KIM Kee-Yeun Telephone No. 82-42-481-8387

Form PCT/ISA/210 (second sheet) (July 2008)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2009/047765

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-9, 23-28
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

see the extra sheet

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required addtional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2009/047765

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007-062486 A1	07.06.2007	CA 2632859-A1	07.06.2007
		CN 101365944 A	11.02.2009
		JP 2009-517054 A	30.04.2009
		US 2009-0280484 A1	12.11.2009
US 2003-0027150 A1	06.02.2003	WO 03-014399 A1	20.02.2003
WO 2008-148161 A1	11.12.2008	WO 2008-148161 A1	11.12.2008
US 2003-0059808 A1	27.03.2003	US 6920398 B2	19.07.2005
US 2004-0224331 A1	11.11.2004	None	

Box No.II

Claims 1-9 are directed at or refer to nucleotide signal intensities form genotyping data without limiting the signal intensities and are considered too broad and so inadequately supported by the description that no meaningful search could be made.

Claims 23-25 are directed at reagents without limiting the reagents and are considered too broad and so inadequately supported by the description that no meaningful search could be made.

Claims 26-28 are directed at a computer readable medium is considered too broad and so inadequately supported by the description that no meaningful search could be made.