(12) UK Patent Application (19) GB (11) 2559437 (13) A

(43) Date of A Publication 08.08.2018

(21) Application No: 1711528.8

(22) Date of Filing: 18.07.2017

(71) Applicant(s):
Congenica Ltd.
Merlin Place, Milton Road, Cambridge, CB4 0DP, United Kingdom

(72) Inventor(s):
Nick Lench
Matt Hurles
John McGonigle
Alan Martin
Suzie Drury

(74) Agent and/or Address for Service:
Basck Ltd
16 Saxon Road, CAMBRIDGE, Cambridgeshire, CB5 8HS, United Kingdom

(51) INT CL:
G06F 19/20 (2011.01)

(56) Documents Cited:
WO 2007/113490 A1 **WO 1995/017524 A1**
JP 2004126857 A

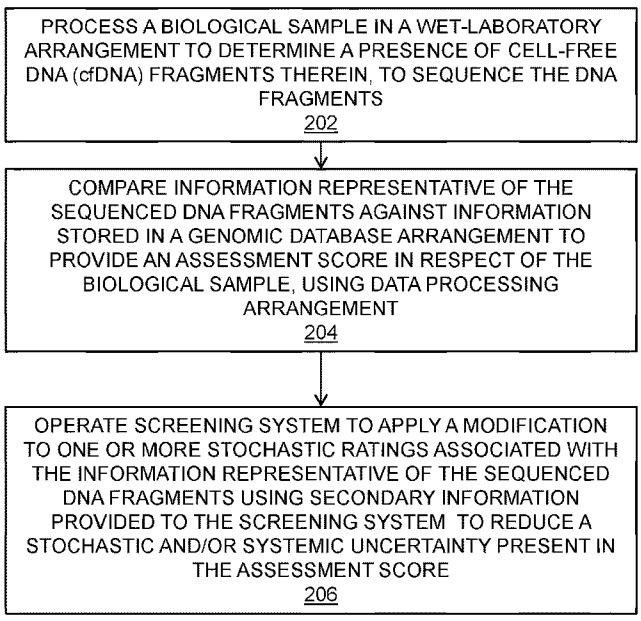
(58) Field of Search:
INT CL **G06F**
Other: **Online: WPI, Epodoc, Internet, XPESP, XPIEE, XPIPCOM, XPI3E, XPMISC, XPLNCS, XPRD, Patent Fulltext**

(54) Title of the Invention: **Prenatal screening and diagnostic system and method**
Abstract Title: **Genomic sequence testing using database to reduce uncertainty**

(57) A screening system to process a biological sample in a wet-laboratory arrangement to determine a presence of cell-free DNA (cfDNA) fragments therein, to sequence the DNA fragments, and to use a data processing arrangement to compare information representative of the sequenced DNA fragments against information stored in a genomic database arrangement to provide an assessment score in respect of the biological sample. The screening system is operable to apply a modification to one or more uncertainty metrics associated with the information representative of the sequenced DNA fragments using secondary information provided to the screening system to reduce a stochastic and/or systemic uncertainty present in the assessment score. This may distinguish between cell-free DNA fragments of maternal, placental or fetal origin, using genome locality scores for mutations, sequence error scores for a given nucleic acid base, patient or mosaicism. This allows weighting to modify the confidence of a call. Biological samples containing cfDNA fragments may be extracted from a pregnant woman non-invasively.

200

FIG. 2



GB 2559437 A

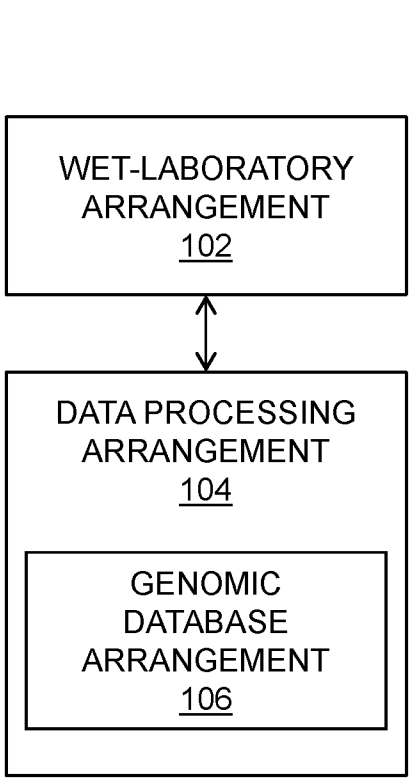

100
WET-LABORATORY ARRANGEMENT
102
DATA PROCESSING ARRANGEMENT
104
GENOMIC DATABASE ARRANGEMENT
106

FIG. 1

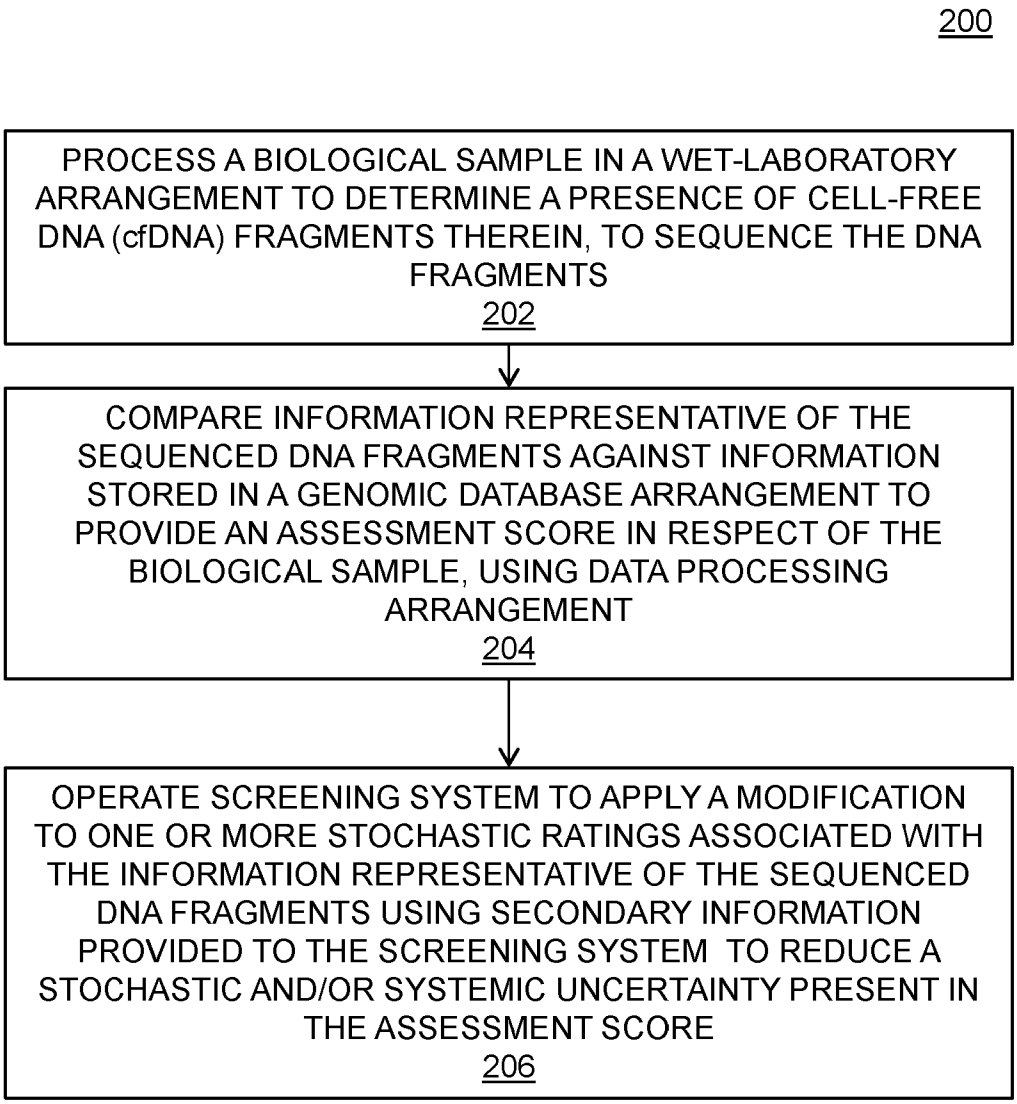

200
PROCESS A BIOLOGICAL SAMPLE IN A WET-LABORATORY ARRANGEMENT TO DETERMINE A PRESENCE OF CELL-FREE DNA (cfDNA) FRAGMENTS THEREIN, TO SEQUENCE THE DNA FRAGMENTS
202
COMPARE INFORMATION REPRESENTATIVE OF THE SEQUENCED DNA FRAGMENTS AGAINST INFORMATION STORED IN A GENOMIC DATABASE ARRANGEMENT TO PROVIDE AN ASSESSMENT SCORE IN RESPECT OF THE BIOLOGICAL SAMPLE, USING DATA PROCESSING ARRANGEMENT
204
OPERATE SCREENING SYSTEM TO APPLY A MODIFICATION TO ONE OR MORE STOCHASTIC RATINGS ASSOCIATED WITH THE INFORMATION REPRESENTATIVE OF THE SEQUENCED DNA FRAGMENTS USING SECONDARY INFORMATION PROVIDED TO THE SCREENING SYSTEM TO REDUCE A STOCHASTIC AND/OR SYSTEMIC UNCERTAINTY PRESENT IN THE ASSESSMENT SCORE
206

FIG. 2

PRENATAL SCREENING AND DIAGNOSTIC SYSTEM AND METHOD

TECHNICAL FIELD

The present disclosure relates to screening systems, for example prenatal screening systems that are operable to process maternal blood in order to determine fetal (foetal) characteristics. Moreover, the present disclosure concerns methods of using aforementioned screening systems, for example methods of using aforementioned screening systems for processing maternal blood in order to determine fetal (foetal) characteristics. Additionally, the present disclosure is concerned with computer program products comprising a non-transitory computer-readable storage medium having computer-readable instructions stored thereon, the computer-readable instructions being executable by a computerized device comprising processing hardware to execute the aforesaid methods.

BACKGROUND

Zygote formation and associated subsequent fetal (foetal) development is a complex biological process that does not always occur without various defects arising. It is of great societal benefit that such defects are detected reliably, for example as early as possible, during fetal growth. For detecting aforementioned defects, it is contemporary practice to provide antenatal or prenatal screening to pregnant women to prevent or treat potential health problems that may occur during pregnancy. Such problems may affect both a given mother and/or a fetus of the given mother, and the problems may be determined by lifestyle, environmental factors or genetically; for example, exposure of the given mother during pregnancy to ionizing radiation (for example to nuclear radiation arising from a nuclear accident, for example as occurred at Fukushima Dai'ichi and Chernobyl) can cause

fetal abnormalities. However, of particular importance are fetal abnormalities that are genetic in origin; these abnormalities may be caused by mutations inherited from one or both parents of a given fetus, or may arise "*de novo*", namely stochastically. Such mutations can range extensively from changes in single nucleotides (namely 'rungs' of a DNA spiral of a given chromosome) to the presence of additional whole chromosomes within a genome of the given fetus. Of particular clinical significance are the chromosomal disorders known as aneuploidies that occur when there is an abnormal number of chromosomes (for example, Down's Syndrome); for example, the human genome normally has 46 chromosomes, but problems can arise for the given fetus when its genome has 45 or 47 chromosomes. Many chromosomal disorders are fatal or result in multiple congenital anomalies for a given newborn child who may be affected by such disorders.

Prenatal screening for fetal chromosomal abnormalities during pregnancy is widely available through public and private healthcare providers. This prenatal screening is normally carried out during a first trimester of a given pregnancy (namely, an initial 10 to 14 weeks after conception) and typically involves a task of performing a maternal blood test combined with an ultrasound scan of a fetus; this task is known as the "*Combined Test*". Levels, namely concentrations, of human chorionic gonadotrophin (hCG) and pregnancy-associated plasma protein (PAPP-A) are measured from the maternal blood sample, along with performing a nuchal translucency (NT) scan; once other medical factors have been taken into account (e.g. maternal age), finally a risk-score is provided at a conclusion of the task. If a pregnancy is categorised as high-risk, an invasive diagnostic procedure (for example, chorionic villus sampling, amniocentesis, cordocentesis) is offered to the mother to confirm or rule out Down's syndrome (trisomy chromosome 21 - T21), Edwards's syndrome (trisomy chromosome 18 -

T18) and Patau syndrome (trisomy chromosome 13 - T13). Pregnant women are also offered a second ultrasound scan at 18 to 21 weeks after conception to check for structural fetal anomalies such as cardiac malformations, brain malformations and skeletal abnormalities. This second scan can be used to direct antenatal treatments, identify anomalies that require early intervention following delivery or enable follow-on diagnostic testing and pregnancy management. Invasive tests such as chorionic villus sampling, amniocentesis and cordocentesis carry a 1% chance of miscarriage of a fetus.

During recent years, non-invasive techniques (without involving a risk of miscarriage) have been developed for diagnosing fetal chromosomal anomalies, wherein the non-invasive techniques rely on a presence of circulating cell-free fetal DNA (referred hereinafter as '*cffDNA*') in the mother's blood. The testing of cell-free fetal DNA (cffDNA) has now entered routine clinical practice for non-invasive prenatal testing (NIPT) for aneuploidy (T21, T18, T13) and has a potential for being employed over a broad range of applications as a replacement for the aforementioned Combined Test. The number of anomalies that can be tested by non-invasive prenatal testing (NIPT) are increasing as methods are developed for the identification of sub-chromosomal rearrangements, such as 22q11.2/DiGeorge syndrome and other microdeletion syndromes; '*microdeletion*' relates to an absence of short sequences in a DNA molecule that would normally be present in the DNA molecule, for example a given chromosome. However, the false positive rate for these anomalies is considered to be too high to offer non-invasive prenatal testing (NIPT) on a screening basis and is only offered if there is an accompanying clinical indication such as a congenital heart defect. Thus, NIPT is classified as *testing* rather than *diagnosis* as the cffDNA which is measured is derived from the placenta rather than the fetus, meaning that false positives can

occur due to confined placental mosaicism. For such a reason of testing purposes only, it is recommended that positive NIPT results are confirmed by an invasive amniocentesis, as aforementioned.

Non-invasive prenatal diagnosis (NIPD), namely a form of test, is classified as a diagnostic assay with no need for a subsequent invasive assay to confirm results provided by the diagnostic assay. The use of NIPD is more limited than NIPT, and is used for fetuses at risk of single gene disorders (namely, inherited and *de novo* mutations) or who present with a suspicion of a genetic disorder on performing fetal ultrasound imaging of a given fetus.

Accurately, reconstructing a given fetus' genetic information from circulating cffDNA present in a sample of corresponding maternal blood is an exceptionally challenging task. This task is challenging, at least in part, because cell-free fetal (cffDNA) occurs only as small fragments in the sample of maternal blood, and represents only a small fraction of the total cell-free DNA (cfDNA) present in plasma present in the sample of maternal blood. The aforementioned cffDNA is actually derived from a placenta of the mother, not from the given fetus (foetus), and thus can display upon sequencing different genetic information to that of the given fetus; such different genetic information can arise from confined placental mosaicism that complicates a process of predicting genetic information of the given fetus, namely *'child's genetics'*. Additionally, cffDNA is non-uniformly distributed across a given human genome, and as half of the child's genetics closely resemble that of the mother, from whom it was inherited, a considerable difficulty arises in practice when estimating an extent to which there is identified cffDNA coverage across the given genome. Added to this difficulty, there arise systematic and methodological difficulties that not all regions of the given genome are equally easy to sequence and accurately called, and that errors can be introduced to an underlying determined

sequence during associated library preparation and template amplification stages. Thus, the accurate reconstruction of a child's genetic profile over a large area of its genome is associated with considerable uncertainty. Yet despite this difficulty, being able to reconstruct the child's genetics is exactly what is required to be able to predict successfully a risk to the child of inheriting, or acquiring *de novo*, a genetic disorder.

Known approaches to try to address this challenge of reconstructing a given child's genetic profile have focused on a combination of a rules-based approach, combined with statistical techniques to determine on a per variant basis whether or not a variant call is:

(a) real; and

(b) of fetal (foetal) origin.

These approaches use knowledge of parental genotypes, to assess a likelihood of a given observed allele frequency, given a read depth that is achieved at that locus, and to make some decision on whether or not to accept or reject this call.

Relative Haplotype Dosage (RHDO) analysis has been used in a situation where a given father is homozygous for a site and a corresponding given mother is heterozygous (Lam *et al.*, 2012). RHDO analysis is performed on a per locus basis, and involves determining whether or not the number of sequenced reads in respect of the two alleles favours one allele or another (Chiu *et al.*, 2008; Lo *et al.*, 2010). Such a favouring is determined by whether or not there is balance or imbalance in the proportions of reads in respect of a particular allele. If a given child were heterozygous, it would be expected to observe allelic balance, as the given child would have the exact same ratio as the mother in whose blood the reads occur (New *et al.*, 2014; Xiong *et al.*, 2015). However, if the child were homozygous for a given allele, then it would be expected that this data would represent itself as an

allelic imbalance, with a large proportion of reads favouring a given site (Xiong *et al.*, 2015). The expectation of the degree of imbalance is dependent upon the fetal (foetal) fraction. From a prediction of allelic balance or imbalance, the given child's genotype at a given site can then be scored. Such an analysis is carried out in a step-wise manner for each variant on a chromosome from those occurring near a start of the chromosome to those occurring near an end of the chromosome in the order they occur (Lam *et al.*, 2012).

More recently, Hidden Markov models (HMM) have also been used to deal with a potential for one or more *de novo* mutations at any given site. In such a case, maternal inheritance of the foetus is inferred by HMM. Classically, a HMM has three parameters: a latent state, an emission probability and a transition probability. In such an approach using HMM, the allele inherited from the mother is determined from two factors:

(i) the maternal inheritance of a previous variant in order along the chromosome (latent state) and;

(ii) the SNP type (emission probability) (Kitzman *et al.*, 2012).

There is therefore employed a model that also accounts for natural haplotype switching events such as genetic recombination (namely transition probability). A proprietary Viterbi™ algorithm employs a recursive algorithm that searches for a given sequence with a maximum associated probability; the proprietary Viterbi™ algorithm is susceptible to being used to generate a most probable latent state sequence (Chan and Jiang, 2015). Altogether, the maternal inheritance of the fetus (foetus) is susceptible to being deduced. A similar method has be used for estimation of paternal inheritance of the foetus (Chan *et al.*, 2016).

In addition, Chan *et al*, 2016 have used a high coverage base filter method, whereby there is determined a likelihood that each individual base has been

called (namely is identified) accurately, using a strict threshold for the number of times a base must have been observed before it is accepted as a '*true*' call. This method filters out much of the erroneous variation allowing for a much more confident estimation of real from non-real variance, but requires incredibly high sequence depth for this method to be tenable (Chan *et al.*, 2016); achieving a high sequence depth is both costly and time-consuming to achieve. By this method, in combination with those described above, they were able to recapture a large proportion of the variants in the child that had occurred *de novo*.

In order to filter at the level of the variant, a dynamic cut off range has been used in order to attempt to identify *de novo* variations. This method was developed to distinguish between *de novo* mutations present in the foetus and sequencing errors. This is achieved by calculating a probability of a same given variant being observed as many times as a current variant purely due to sequencing error and applying a cut off relative to this probability (Chan *et al.*, 2016).

Therefore, in light of the foregoing discussion, there exist problems associated with conventional pre-natal screening systems.

SUMMARY

The present disclosure seeks to provide an improved screening system that is capable of providing a lower occurrence of false-positive and false-negatives when the screening system is employed for providing a prenatal screening service.

Moreover, the present disclosure seeks to provide an improved method of using a screening system that is capable of providing a lower occurrence of false-positive and false-negatives when the screening system is employed for providing a prenatal screening service.

In a first aspect, embodiments of the present disclosure provide a screening system that is operable:

(i) to process a biological sample in a wet-laboratory arrangement to determine a presence of cell-free DNA (cfDNA) fragments therein, to sequence the DNA fragments; and

(ii) to use a data processing arrangement to compare information representative of the sequenced DNA fragments against information stored in a genomic database to provide an assessment score in respect of the biological sample,

characterized in that the screening system is operable to apply a modification to one or more uncertainty automated screening arrangement **100** metrics associated with the information representative of the sequenced DNA fragments using secondary information provided to the screening system to reduce a stochastic and/or systemic uncertainty present in the assessment score.

The present disclosure is of advantage in that it provides a personalized non-invasive system and method of identifying genetic abnormalities in a fetus. Moreover, the system disclosed herein is advantageous as it provides no increased risk of miscarriage and has a higher accuracy with false negative and false positive results prevention.

Embodiments of the disclosure are advantageous in terms of providing a rapid, simple, low cost, patient specific and highly efficient method and system of screening that takes into consideration external factors. Moreover, the method and system is helpful in increasing confidence in a screening test.

Optionally, the screening system is operable to distinguish between cell-free DNA fragments of maternal origin and cell-free DNA fragments of placental and/or fetal origin.

Optionally, the screening system is operable to employ at least one of following scores when computing the modification in the data processing arrangement:

(a) a genome locality score, wherein the genome locality score includes a likelihood of mutation within a region;

(b) a sequence error score, wherein the sequence error score includes a likelihood of a given nucleic acid base is a result of PCR infidelity during template amplification and/or a miscall during a sequencing process;

(c) a patient modifier score, wherein the patient modifier score includes details from external sources; and

(d) a mosaicism detection score, wherein the mosaicism detection score includes a likelihood of variants occurring in a region of imbalanced maternal genotype.

Optionally, the likelihood of mutation within the region is calculated on a basis of frequencies of change susceptible to occur to the region and/or frequencies of calling spurious variants in the region.

Optionally, the sequence error score is calculated using a maternal genetic sequence.

More optionally, the information for external sources includes at least information received from abnormality scans.

Yet more optionally, the screening system is operable to convert the genome locality score into a weight for a particular locus.

Optionally, the screening system is operable to apply the sequence error score as a weight and to modify confidence in a base call.

Optionally, the screening system is operable to convert the details from external sources in to a weight.

More optionally, the screening system is operable to combine the genome locality score, the sequence error score, the patient modifier score and/or mosaicism detection score to modify the confidence of a call.

Optionally, the biological sample containing cfDNA fragments therein is extracted from a pregnant woman in a non-invasive manner.

In a second aspect, embodiments of the present disclosure provide a method of using a screening system that is operable:

(i) to process a biological sample in a wet-laboratory arrangement to determine a presence of cell-free DNA (cfDNA) fragments therein, to sequence the DNA fragments; and

(ii) to use a data processing arrangement to compare information representative of the sequenced DNA fragments against information stored in a genomic database arrangement to provide an assessment score in respect of the biological sample,

characterized in that the method includes operating the screening system to apply a modification to one or more uncertainty metrics associated with the information representative of the sequenced DNA fragments using secondary information provided to the screening system to reduce a stochastic and/or systemic uncertainty present in the assessment score.

Optionally, the method includes distinguishing between cell-free DNA fragments of maternal origin from cell-free DNA fragments of placental and/or fetal origin.

Optionally, the method includes employing at least one of following scores when computing the modification in the data processing arrangement:

(a) a genome locality score, wherein the genome locality score includes a likelihood of mutation within a region;

(b) a sequence error score, wherein the sequence error score includes a likelihood of a given nucleic acid base is a result of PCR infidelity during template amplification and/or a miscall during the sequencing process;

(c) a patient modifier score, wherein the patient modifier score includes details from external sources; and

(d) a mosaicism detection score, wherein the mosaicism detection score includes a likelihood of variants occurring in a region of imbalanced maternal genotype.

Optionally, the method includes calculating the likelihood of mutation within the region on the basis of frequencies of change is to occur to the region and/or frequencies of calling spurious variants in the region.

Optionally, the method includes calculating the sequence error score using maternal genetic sequence.

Optionally, the method includes receiving information from abnormality scans.

More optionally, the method includes converting the genome locality score into a weight for a particular locus.

Yet more optionally, the method includes applying the sequence error score as a weight and to modify confidence in a base call.

Optionally, the method includes converting the details from external sources in to a weight.

More optionally, the method includes combining the genome locality score, the sequence error score, the patient modifier score and/or mosaicism detection score to modify the confidence of a call.

Optionally, the method includes extracting the biological sample containing cfDNA fragments therein from a pregnant woman in a non-invasive manner.

In a third aspect, embodiments of the present disclosure provide a computer program product comprising a non-transitory computer-readable storage medium having computer-readable instructions stored thereon, the computer-readable instructions being executable by a computerized device comprising processing hardware to execute the aforementioned method.

Additional aspects, advantages, features and objects of the present disclosure would be made apparent from the drawings and the detailed description of the illustrative embodiments construed in conjunction with the appended claims that follow.

It will be appreciated that features of the present disclosure are susceptible to being combined in various combinations without departing from the scope of the present disclosure as defined by the appended claims.

A better understanding of the present invention may be obtained through the following examples which are set forth to illustrate, but are not to be construed as limiting the present invention.

BRIEF DESCRIPTION OF DRAWINGS

Embodiments of the present disclosure will be more fully understood from examples described herein below and the accompanying drawings, which is given by way of illustration only, and thus are not limitative of the present invention, and wherein:

FIG. 1 is a schematic illustration of a screening system, pursuant to the present disclosure; and

FIG. 2 is an illustration of steps of a method of using the screening system, pursuant to the present disclosure.

In the accompanying diagrams, an underlined number is employed to represent an item over which the underlined number is positioned or an item

to which the underlined number is adjacent. A non-underlined number relates to an item identified by a line linking the non-underlined number to the item. When a number is non-underlined and accompanied by an associated arrow, the non-underlined number is used to identify a general item at which the arrow is pointing.

LIST OF ABBREVIATIONS

Abbreviation	**Meaning**
PCR	Polymerase Chain Reaction
cfDNA	Cell-Free DNA

DEFINITIONS

As used herein, the following terms shall have the following meanings:

As used herein, the term '*data processing arrangement*' refers to a process and/or system that can be embodied in software that determines the biological significance of acquired data (i.e., the ultimate results of an assay). For example, a data processing arrangement can determine the amount of each nucleotide sequence species based upon the data collected. A data processing arrangement also may control an instrument and/or a data collection system based upon results determined. A data processing and a data collection arrangement often are integrated and provide feedback to operate data acquisition by the instrument, and hence provide assay-based judging methods provided herein.

As used herein, the term '*database*' refers to a nucleic acid databases known in the art including, for example, GenBank, dbEST, dbSTS, EMBL (European Molecular Biology Laboratory) and DDBJ (DNA Databank of Japan). BLAST or

similar tools can be used to search the identified sequences against a sequence database.

As used herein, the term *'genetic information'* refers to information related nucleic acids, altered nucleotide sequence, chromosomes, segments of chromosomes, polymorphic regions, translocated regions, the like or combinations of the foregoing. Furthermore, the nucleic acids may include, but are not limited to, DNA, cDNA, RNA, mRNA, tRNA and rRNA. Moreover, the genetic information may include information related to mutations, copy number variations, transversions, translocations, inversion, deletions, aneuploidy, partial aneuploidy, polyploidy, chromosomal instability, chromosomal structure alterations, gene fusions, chromosome fusions, gene truncations, gene amplification, gene duplications, chromosomal lesions, DNA lesions, abnormal changes in nucleic acid chemical modifications, abnormal changes in epigenetic patterns, abnormal changes in nucleic acid methylation infection or cancer.

As used herein, the term *'cell-free DNA'* refers to DNA that is not within a cell. In one embodiment, cell free DNA includes DNA circulating in blood. In another embodiment, cell free DNA includes DNA existing outside a cell. In yet another embodiment, cell free DNA includes DNA existing outside a cell as well as DNA present in a blood sample after such blood sample has undergone partial or gentle cell lysing.

As used herein, the terms *'biological sample'* refers to the sample obtained from a female who is pregnant, the sample may include, but is not limited to, plasma, serum, peripheral blood and urine. Typically, the sample is a maternal plasma sample, although other tissue sources that contain both maternal and fetal DNA are optionally used. Maternal plasma can be obtained from a peripheral whole blood sample from a pregnant woman and the plasma can be obtained by standard methods. A volume of 3 ml to 5 ml

of plasma is sufficient to provide suitable DNA material for analysis. The cell free DNA can be extracted from the sample using standard techniques, non-limiting examples of which include a Qiasymphony protocol (Qiagen) suitable for free fetal DNA isolation or any other automated or manual extraction method suitable for cell free DNA isolation.

As used herein, the term '*biological characteristics*' refers to the genetic variations, abnormalities, irregularities or mutations which range extensively from changes in single nucleotides to the presence of additional whole chromosomes or abnormal number of chromosomes. The chromosomal abnormality is a structural abnormality, including, but not limited to, copy number changes including microdeletions and microduplications, insertions, translocations, inversions and small-size mutations including point mutations and mutational signatures.

As used herein, the term '*wet-laboratory arrangement*' refers to a facility, clinic and/or a setup of: instruments, equipment and/or devices used for extraction, collection, processing and/or analysis of body fluid samples; instruments, equipment and/or devices used for extraction, collection, processing and/or analysis of genetic material; instruments, equipment and/or devices used for amplification, enrichment and/or processing of genetic material received from the body fluid samples; instruments, equipment and/or devices used for extraction and/or analysis of the genetic information received from the amplified genetic material. Herein the instruments, equipment and/or devices may include but not limited to centrifuge, ELISA, spectrophotometer, PCR, RT- PCR, High-Throughput-Screening (HTS) system, Microarray system, Ultrasound, genetic analyser, deoxyribonucleic acid (DNA) sequencer and SNP analyser. The wet-laboratory arrangement is operable to monitor and/or scan a fetus. Herein, the wet-laboratory arrangement may include equipment, instruments and/or

devices for scanning the fetus; examples include ultrasound scan, presymptomatic genetic testing and/or combined tests.

As used herein, *'polymerase chain reaction (PCR)'* is a technique used in molecular biology to amplify a single copy or a few copies of a segment of DNA by several orders of magnitude, thereby generating potentially thousands of millions of copies of a particular given DNA sequence.

As used herein, *'bridge amplification'* or *'amplification'* is employed in massively parallel sequencing for DNA sequencing purposes using a concept of massively parallel processing, wherein use is made of miniaturized and parallelized platforms for sequencing of 1 million to 43 billion short reads (50 to 400 nucleic acid bases each) per instrument run.

DETAILED DESCRIPTION

The practice of the embodiments described in further detail below will employ, unless otherwise indicated, conventional methods of diagnostics, molecular biology, cell biology, biochemistry and immunology within the skill of the art. Such techniques are explained fully in the literature, for example in published scientific academic research literature relating to genetics.

It is appreciated that certain features of the invention, which are for clarity described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely various features of the invention, which are for brevity, described in the context of a single embodiment, may also be provided separately and/or in any suitable sub-combination.

The following detailed description illustrates embodiments of the present disclosure and ways in which they can be implemented. Although some modes of carrying out the present disclosure has been disclosed, those

skilled in the art would recognize that other embodiments for carrying out or practicing the present disclosure are also possible.

In overview, embodiments of the present disclosure are concerned with a screening system as illustrated by **100** in FIG. 1. The screening system **100** includes a wet-laboratory arrangement **102**, wherein the wet-laboratory arrangement **102** includes apparatus such as biological sample collection apparatus, centrifuges, PCR rapid gene sequencing apparatus and similar apparatuses. Furthermore, the screening system **100** is operable to process a biological sample in the wet-laboratory arrangement to determine a presence of DNA (namely, cfDNA) fragments therein, and to sequence the DNA fragments.

In operation, the biological sample is obtained from a person, for example a pregnant mother. Optionally, the biological sample is a blood sample or a tissue sample. Optionally, with regard to the pregnant mother, the biological sample is a non-invasive sample, wherein collection of sample does not have an associated risk of miscarriage therewith. However, optionally, the biological sample is supplemented with an invasive sample if required, for example amniotic fluid, placental tissue and so forth. Furthermore, the biological sample includes plasma that includes, as a component part thereof, a mixture of cell-free DNA (cfDNA) fragments. Specifically, the cell-free DNA (cfDNA) may comprise a portion derived from the pregnant mother, from the placenta of the pregnant mother and/or from the fetus. Moreover, the portion of the cell-free DNA (cfDNA) derived from the pregnant mother and/or the fetus is referred as cell-free fetal DNA (cffDNA).

Furthermore, the wet-laboratory arrangement **102** sequences DNA fragments to determine the presence of cell-free DNA (cfDNA). Specifically, DNA fragments present in plasma are amplified and sequenced to generate

information representative of sequenced DNA fragments. Optionally, the information representative of sequenced DNA fragments comprises a large amount of nucleic acid-base sequence information. Subsequently, the nucleic acid-base sequence information is processed in the data processing arrangement **104.**

In an embodiment, the screening system **100** is operable to distinguish between cell-free DNA (cfDNA) fragments of maternal origin and cell-free DNA (cfDNA) fragments of placental and/or fetal origin. Specifically, the wet-laboratory arrangement **102** may isolate the cell-free DNA (cfDNA) fragments of placental and/or fetal origin from the cell-free DNA (cfDNA) fragments of maternal origin present in the biological sample. Moreover, the cell-free DNA (cfDNA) fragments of placental and/or fetal origin are analysed in a data processing arrangement **104**.

Moreover, the screening system **100** further includes the data processing arrangement **104**, including a genomic database arrangement **106**, for receiving information representative of sequenced DNA fragments from the wet-laboratory arrangement **102**. Optionally, the data processing arrangement provides feedback data to the wet-laboratory arrangement **102** for controlling various tests performed thereat. Furthermore, the genomic database arrangement **106** stores information comprising genomic mapping data and research data analysing structure, location and sequencing of human genes, and clinical effects of mutations and their co-relation with biological sequences and structures.

Furthermore, the data processing arrangement **104** is operable to compare information representative of the sequenced DNA fragments against information stored in a genomic database arrangement **106** to provide an assessment score in respect of the biological sample. Optionally, the data processing arrangement may compute results from screening tests

implemented upon the biological sample processed by the wet-laboratory arrangement **102**. For example, the wet-laboratory arrangement **102** may provide a prenatal screening service, but is not limited thereto. More optionally, the data processing arrangement **104** may compare sequencing of DNA fragments against information stored in the genomic database arrangement **106** to assess a risk of a genetic disorder in the compared DNA fragments. Specifically, cell free-DNA fragments are compared against information stored in the genomic database arrangement **106**. In an exemplary embodiment, the information representative of the sequenced DNA fragments may comprise a sequential arrangement of 'P–Q–R–S–P–Q–R–S' DNA base pairs with an anomaly 'P–R–Q–S'. In such an embodiment, the data processing arrangement **102** may compare the anomaly against sequential arrangements of DNA stored in the genomic database **106**. Subsequently in the embodiment, the data processing arrangement **104** may assess if the anomaly may or may not cause a genetic disorder. Additionally, the data processing arrangement **104** may compare and provide the assessment score representative of a risk to the fetus of inheriting or acquiring a genetic disorder. It will be appreciated that the DNA base pairs P, Q, R, S represent DNA base pairs adenine, thymine, guanine and cytosine for illustrative purposes only and do not represent the actual arrangement of the DNA base pairs which may be responsible for a specific disease.

It will be appreciated that there is one or more uncertainty metrics, associated with the information representative of the sequenced DNA fragments provided by the wet-laboratory arrangement **102**. Specifically, the one or more of these probabilistic ratings are representative of measurement of stochastic noise during prenatal screening. More specifically, the stochastic noise may increase a risk of a false assessment score (such as, a false negative score or a false positive score) when

computed by the data processing arrangement **104.** Moreover, the risk of a false assessment score contributes towards stochastic and/or systemic uncertainty present in the assessment score. Additionally, a higher stochastic and/or systemic uncertainty reduces confidence in results provided by the screening tests.

The screening system **100** is operable to apply a modification to one or more of the metrics of uncertainty associated with the information representative of the sequenced DNA fragments. Specifically, the modification is applied to one or more uncertainty metrics to reduce a stochastic and/or systemic uncertainty present in the assessment score. More specifically, the modification is applied to one or more stochastic ratings using secondary information provided to the screening system **100**.

In an embodiment, the secondary information provided to the screening may refer to genetic information, environmental conditions, nutritional information related with diet and so forth associated with the person providing the biological sample. Specifically, the secondary information relates to factors that may affect a risk of inheriting a congenital defect.

In an embodiment, the screening system is operable to employ at least one of following scores when computing the modification in the data processing arrangement:

(a) a genome locality score, wherein the genome locality score includes a likelihood of mutation within a region;

(b) a sequence error score, wherein the sequence error score includes a likelihood of a given nucleic acid base is a result of PCR infidelity during template amplification and/or a miscall during a sequencing process;

(c) a patient modifier score, wherein the patient modifier score includes details from external sources; and

(d) a mosaicism detection score, wherein the mosaicism detection score includes a likelihood of variants occurring in a region of imbalanced maternal genotype

In an embodiment, the genome locality score is calculated to include a frequency of identification (namely, calling) of an incorrect genotype. Specifically, the genotype may be in a given area of a genome or chromosome. More specifically, the incorrect genotype may be identified due to issues of the genomic architecture. Examples of issues of the genomic architecture include, but are limited to, one or more occurrences of repetitive sequence, low genetic conservations, gene sequence topologies. Furthermore, genome locality score relates to likelihood of mutation within a region. Specifically, the region may be a region of interest. In an example, the region of interest may be assay capture areas that may extend to all exomes or a whole given genome. Moreover, the data processing arrangement **104** may calculate the genome locality score.

In an embodiment, the likelihood of mutation within the region is calculated on the basis of frequencies of change susceptible to occur in the region and/or frequencies of calling spurious variants in the region. Specifically, the genome locality score takes into estimation, a likelihood of one or more changes to occur in a region. Moreover, frequencies of calling spurious variants in the region are taken into estimation. In an embodiment, the screening system **100** is operable to convert the genome locality score into a weight for a particular locus. Specifically, the weight of the genome locality score is representative of an indication of a potential error in the calling of genotypes in this locality.

In an embodiment, the sequence error score includes a likelihood of an error during amplification and/or sequencing of DNA fragments. Specifically, the DNA fragments are amplified and sequenced to generate information

representative of sequenced DNA fragments. Consequently, polymerase chain reactions (PCR) may be implemented during such amplification and sequencing. Specifically, polymerase chain reactions (PCR) employ a DNA polymerase for accurate replication of DNA fragments. Subsequently, an error in amplification and sequencing (namely, replication) by the DNA polymerase is referred to PCR infidelity. Therefore, the sequence error includes likelihood that a given nucleic acid base is generated due to PCR infidelity during amplification of the template (namely, DNA fragments). Furthermore, an error in the sequencing process is referred as being a miscall during sequencing process. Specifically, the sequencing process includes different nucleic acid base concentrations that may lead to the miscall.

In an embodiment, the sequence error score is calculated using a maternal genetic sequence. Specifically, the DNA fragments employed for amplification and sequencing may be of maternal origin. Furthermore, estimation of sequence error score is relatively less complex using information obtained from the maternal genetic sequence. In an embodiment, the screening system **100** is operable to apply the sequence error score as a weight and to modify a confidence in a base call. Specifically, the weight of the sequence error score is a representative of accuracy of the amplification and sequencing process.

In an embodiment, the patient modifier score is employed when computing the modification in the data processing arrangement **104**. Specifically, the patient modifier score includes variations in a phenotype influenced by factors such as diet, climate, exposure to chemicals or ionizing radiation, illness and so forth. Optionally, the information for external sources includes at least information received from abnormality scans. Specifically, the abnormality scan may be performed during pregnancy to ensure a healthy development of the fetus. More specifically, any anomaly and/or

abnormality is reported to include in the patient modifier score by the screening system **100**.

In an embodiment, the patient modifier score may include expectations of *de novo* mutations in line with paternal age. For example, more de novo mutations may be expected in foetuses with older fathers and thus, may contribute towards a higher patient modifier score. Optionally, the patient modifier score may include dominant-recessive inheritance. For example, a risk of a child inheriting diabetes having parents with recessive gene responsible for diabetes may be more. In an embodiment, the screening system **100** is operable to convert details from the external sources into a weight.

In an embodiment, the mosaicism detection score includes a likelihood of variant occurring in a region of imbalanced maternal genotype. Specifically, the cell-free DNA (cfDNA) fragments of the placental origin may potentially display genetic abnormalities, even when such abnormalities may not exist in the fetus. Therefore, maternal genotypes may provide an indication if there is a real risk of abnormality in the fetus or if it is a false call. In another embodiment, the false call may be due to an imbalance in maternal allele frequencies.

In an embodiment, the screening system **100** is operable to combine the genome locality score, the sequence error score, the patient modifier score and/or mosaicism detection score to modify the confidence of a call. Specifically, the scores take into account the factors that may cause an error in the assessment score. Therefore, when such factors are accounted for in the assessment score, the confidence of a call generated thereby may be positively affected.

In an embodiment, the biological sample containing cfDNA fragments therein is extracted from a pregnant woman in a non-invasive manner. Specifically,

the biological sample extracted from the pregnant woman may be extracted in a non-invasive manner to prevent, or reduce, a risk of miscarriage. Furthermore, examples of non-invasive manners may include techniques which may not involve extraction of sample from the amniotic sac.

In FIG. 2, there is shown a flow chart of a method **200** of using a screening system (such as the screening system **100** of FIG. 1) pursuant to the present disclosure. At a step **202**, the flow chart initiates. At the step **202**, a biological sample is processed in a wet-laboratory arrangement to determine a presence of cell-free DNA (cfDNA) fragments therein, to sequence the DNA fragments. At a step **204**, information representative of the sequenced DNA fragments is compared against information stored in a genomic database arrangement, using a data processing arrangement to provide an assessment score in respect of the biological sample. At a step **206**, the screening system is operated to apply a modification to one or more probabilistically defined uncertainty metrics associated with the information representative of the sequenced DNA fragments using secondary information provided to the screening system to reduce a stochastic and/or systemic uncertainty present in the assessment score.

The steps **202** to **206** are only illustrative and other alternatives can also be provided where one or more steps are added, one or more steps are removed, or one or more steps are provided in a different sequence without departing from the scope of the claims herein. In an embodiment, the method **200** includes distinguishing between cell-free DNA fragments of maternal origin from cell-free DNA fragments of placental and/or fetal origin.

In another embodiment, the method **200** includes employing at least one of following scores when computing the modification in the data processing arrangement:

(a) a genome locality score, wherein the genome locality score includes a likelihood of mutation within a region;

(b) a sequence error score, wherein the sequence error score includes a likelihood of a given nucleic acid base is a result of PCR infidelity during template amplification and/or a miscall during the sequencing process;

(c) a patient modifier score, wherein the patient modifier score includes details from external sources; and

(d) a mosaicism detection score, wherein the mosaicism detection score includes a likelihood of variants occurring in a region of imbalanced maternal genotype.

In yet another embodiment, the method **200** includes calculating the likelihood of a mutation within the region on the basis of frequencies of change within that region and/or frequencies of calling spurious variants in the region. In an embodiment, the method **200** includes calculating the sequence error score using maternal genetic sequence.

In an embodiment, the method **200** includes receiving information from abnormality scans. In another embodiment, the method **200** includes converting the genome locality score into a weight for a particular locus. Optionally, the method **200** includes applying the sequence error score as a weight and to modify confidence in a base call. More optionally, the method **200** includes converting the details from external sources in to a weight.

Optionally, the method **200** includes combining the genome locality score, the sequence error score, the patient modifier score and/or mosaicism detection score to modify the confidence of a given call. Optionally, the method **200** includes extracting the biological sample containing cfDNA fragments therein from a pregnant woman in a non-invasive manner.

Optionally, the aforementioned method **200** of using the screening system is implemented by using a computer program product comprising a non-transitory computer-readable storage medium having computer-readable instructions stored thereon, the computer-readable instructions being executable by a computerized device comprising processing hardware.

Modifications to embodiments described in the foregoing are possible without departing from the scope of the invention as defined by the accompanying claims. Expressions such as "including", "comprising", "incorporating", "consisting of", "have", "is" used to describe and claim the present invention are intended to be construed in a non-exclusive manner, namely allowing for items, components or elements not explicitly described also to be present. Reference to the singular is also to be construed to relate to the plural. Numerals included within parentheses in the accompanying claims are intended to assist understanding of the claims and should not be construed in any way to limit subject matter claimed by these claims.

REFERENCES

[1] Chan, L.L., and Jiang, P. (2015). Bioinformatics analysis of circulating cell-free DNA sequencing data. Clin. Biochem. *48*, pp 962–975.

[2] Chan, K.C.A., Jiang, P., Sun, K., Cheng, Y.K.Y., Tong, Y.K., Cheng, S.H., Wong, A.I.C., Hudecova, I., Leung, T.Y., Chiu, R.W.K., *et al.* (2016). Second generation noninvasive fetal genome analysis reveals *de novo* mutations, single-base parental inheritance, and preferred DNA ends. Proc. Natl. Acad. Sci. U. S. A. 201615800.

[3] Chiu, R.W.K., Chan, K.C.A., Gao, Y., Lau, V.Y.M., Zheng, W., Leung, T.., Foo, C.H.F., Xie, B., Tsui, N.B.Y., Lun, F.M.F., et al. (2008). Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc. Natl. Acad. Sci. U. S. A. *105*, pp 20458–20463.

[4] Kitzman, J.O., Snyder, M.W., Ventura, M., Lewis, A.P., Simmons, L.E., Gammill, H.S., Rubens, C.E., Santillan, D. a., Murray, J.C., Tabor, H.K., et al. (2012). Non-invasive Whole Genome Sequencing of Human Fetus. Sci. Transl. Med. *4*, pp 1–18.

[5] Lam, K.W.G., Jiang, P., Liao, G.J.W., Chan, K.C.A., Leung, T.Y., Chiu, R.W.K., and Lo, Y.M.D. (2012). Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: Application to ??-thalassemia. Clin. Chem. *58*, pp 1467–1475.

[6] Lo, Y.M.D., Chan, K.C.A., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M.F., Zheng, Y.W., Leung, T.Y., Lau, T.K., Cantor, C.R., et al. (2010). Maternal

Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. Sci. Transl. Med. *2*, 61ra91-61ra91.

[7] New, M.I., Tong, Y.K., Yuen, T., Jiang, P., Pina, C., Chan, K.C.A., Khattab, A., Liao, G.J.W., Yau, M., Kim, S.-M., et al. (2014). Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. J. Clin. Endocrinol. Metab. *99*, E1022-30.

[8] Xiong, L., Barrett, A.N., Hua, R., Tan, T.Z., Ho, S.S.Y., Chan, J.K.Y., Zhong, M., and Choolani, M. (2015). Non-invasive prenatal diagnostic testing for β-thalassaemia using cell-free fetal DNA and next generation sequencing. Prenat. Diagn. *35*, pp 258–265.

CLAIMS

We claim:

1. A screening system (100) that is operable:

(i) to process a biological sample in a wet-laboratory arrangement (102) to determine a presence of cell-free DNA (cfDNA) fragments therein, to sequence the DNA fragments; and

(ii) to use a data processing arrangement (104) to compare information representative of the sequenced DNA fragments against information stored in a genomic database arrangement (106) to provide an assessment score in respect of the biological sample,

characterized in that the screening system (100) is operable to apply a modification to one or more uncertainty metrics associated with the information representative of the sequenced DNA fragments using secondary information provided to the screening system (100) to reduce a stochastic and/or systemic uncertainty present in the assessment score.

2. A screening system (100) according to claim 1, characterized in that the screening system (100) is operable to distinguish between cell-free DNA fragments of maternal origin and cell-free DNA fragments of placental and/or fetal origin.

3. A screening system (100) according to claim 1 or 2, characterized in that the screening system (100) is operable to employ at least one of following scores when computing the modification in the data processing arrangement:

(a) a genome locality score, wherein the genome locality score includes a likelihood of a mutation within a region;

(b) a sequence error score, wherein the sequence error score includes a likelihood of a given nucleic acid base being a result of PCR infidelity

during template amplification and/or a miscall during a sequencing process;

(c) a patient modifier score, wherein the patient modifier score includes details from external sources; and

(d) a mosaicism detection score, wherein the mosaicism detection score includes a likelihood of variants occurring in a region of imbalanced maternal genotype.

4. A screening system (100) according to claim 3, characterized in that the likelihood of mutation within the region is calculated on a basis of frequencies of change susceptible to occur to the region and/or frequencies of calling spurious variants in the region.

5. A screening system (100) according to claim 3, characterized in that the sequence error score is calculated using a maternal genetic sequence.

6. A screening system (100) according to claim 3, characterized in that the information for external sources includes at least information received from abnormality scans.

7. A screening system (100) according to any one of the preceding claims 3 to 6, characterized in that the screening system (100) is operable to convert the genome locality score into a weight for a particular locus.

8. A screening system (100) according to any one of the preceding claims 3 to 7, characterized in that the screening system (100) is operable to apply the sequence error score as a weight and to modify confidence in a base call.

9. A screening system (100) according to claim 3, characterized in that the screening system (100) is operable to convert the details from external sources in to a weight.

10. A screening system (100) according to any one of the preceding claims 3 to 8, characterized in that the screening system (100) is operable to combine the genome locality score, the sequence error score, the patient modifier score and/or mosaicism detection score to modify the confidence of a call.

11. A screening system (100) according to any one of the preceding claims, characterized in that the biological sample containing cfDNA fragments therein is extracted from a pregnant woman in a non-invasive manner.

12. A method of using a screening system (100) that is operable:

(i) to process a biological sample in a wet-laboratory arrangement (102) to determine a presence of cell-free DNA (cfDNA) fragments therein, to sequence the DNA fragments; and

(ii) to use a data processing arrangement (104) to compare information representative of the sequenced DNA fragments against information stored in a genomic database arrangement (106) to provide an assessment score in respect of the biological sample,

characterized in that the method includes operating the screening system (100) to apply a modification to one or more uncertainty metrics associated with the information representative of the sequenced DNA fragments using secondary information provided to the screening system (100) to reduce a stochastic and/or systemic uncertainty present in the assessment score.

13. A method of using a screening system (100) according to claim 12, characterized in that the method includes distinguishing between cell-free DNA fragments of maternal origin from cell-free DNA fragments of placental and/or fetal origin.

14. A method of using a screening system (100) according to claim 12 or 13, characterized in that the method includes employing at least one of following scores when computing the modification in the data processing arrangement:

(a) a genome locality score, wherein the genome locality score includes a likelihood of a mutation within a region;

(b) a sequence error score, wherein the sequence error score includes a likelihood of a given nucleic acid base is a result of PCR infidelity during template amplification and/or a miscall during the sequencing process;

(c) a patient modifier score, wherein the patient modifier score includes details from external sources; and

(d) a mosaicism detection score, wherein the mosaicism detection score includes a likelihood of variants occurring in a region of imbalanced maternal genotype.

15. A method of using a screening system (100) according to claim 14, characterized in that the method includes calculating the likelihood of mutation within the region on the basis of frequencies of change susceptible to occur in the region and/or frequencies of calling spurious variants in the region.

16. A method of using a screening system (100) according to claim 14, characterized in that the method includes calculating the sequence error score using maternal genetic sequence.

17. A method of using a screening system (100) according to claim 14, characterized in that the method includes receiving information from abnormality scans.

18. A method of using a screening system (100) according to any one of claims 14 to 17, characterized in that the method includes converting the genome locality score into a weight for a particular locus.

19. A method of using a screening system (100) according to any one of claims 14 to 18, characterized in that the method includes applying the sequence error score as a weight and to modify confidence in a base call.

20. A method of using a screening system (100) according to any one of claims 14, characterized in that the method includes converting the details from external sources in to a weight.

21. A method of using a screening system (100) according to claim 14, characterized in that the method includes combining the genome locality score, the sequence error score, the patient modifier score and/or mosaicism detection score to modify the confidence of a call.

22. A method of using a screening system (100) according to any one of claims 14 to 18, characterized in that the method includes extracting the biological sample containing cfDNA fragments therein from a pregnant woman in a non-invasive manner.

23. A computer program product comprising a non-transitory computer-readable storage medium having computer-readable instructions stored thereon, the computer-readable instructions being executable by a computerized device comprising processing hardware to execute a method as claimed in any of claims 12 to 22.

Intellectual Property Office

Application No:	GB1711528.8	**Examiner:**	Robert Shorthouse
Claims searched:	1-23	**Date of search:**	15 January 2018

Patents Act 1977: Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
A	-	WO2007/113490 A1 (Forensic Science) See abstract
A	-	WO95/17524 A1 (Molecular Tool Inc) See abstract and figure 2
A	-	JP2004126857 A (Foundation for Nara Inst of SC) See WPI abstract accession number 2004-381853 and paragraph 13, claim 5

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC^{X} :

Worldwide search of patent documents classified in the following areas of the IPC

G06F

The following online and other databases have been used in the preparation of this search report

WPI, Epodoc, Internet, XPESP, XPIEE, XPIPCOM, XPI3E, XPMISC, XPLNCS, XPRD, Patent Fulltext

International Classification:

Subclass	Subgroup	Valid From
G06F	0019/20	01/01/2011