

biological factors influencing placental RNA levels in maternal plasma. In addition, transcripts that are expressed and released by non-placental tissues in response to pregnancy could not be identified by this method. A sensitive and high-throughput methodology that allows the direct profiling of the maternal plasma transcriptome would therefore be very desirable.

[0004] Similarly, direct plasma RNA profiling may be useful in other scenarios where there is a mixture of RNA molecules from two individuals, such as for organ transplantation. The circulation of transplantation recipients contains nucleic acid molecules from both the donor and recipient. A change in the relative profile of RNA molecules contributed by the donor or the recipient may reveal pathologies in the transplanted organ or recipient, such as graft rejection.

BRIEF SUMMARY

In one aspect, the present invention provides a method of diagnosing a pregnancy-associated disorder using a sample from a female subject pregnant with a fetus, the method comprising:
receiving a plurality of reads, wherein the reads are obtained from an analysis of RNA molecules obtained from the sample, the sample containing a mixture of maternal- and fetal-derived RNA molecules;

identifying, by a computer system, locations of the reads in a reference sequence;

identifying one or more informative loci, each of which is homozygous in a first entity for a corresponding first allele and which is heterozygous in a second entity for the corresponding first allele and a corresponding second allele, wherein the first entity is the pregnant female subject or the fetus, and the second entity is the other one of the pregnant female subject and the fetus;

filtering the one or more informative loci to identify one or more filtered informative loci:

that are located within an expressed region of the reference sequence,
at which at least a first predetermined number of reads in the plurality of reads containing the corresponding first allele are located, and
at which at least a second predetermined number of reads in the plurality of reads containing the corresponding second allele are located,

for each of the filtered informative loci:

determining a first number of reads located at the filtered informative locus and containing the corresponding first allele,

determining a second number of reads located at the filtered informative locus and containing the corresponding second allele,

calculating a first sum of the first numbers and the second sum of the second numbers;

calculating a ratio of the first sum and the second sum, and

comparing the ratio to a cutoff value to determine whether the fetus, mother, or pregnancy has a pregnancy-associated disorder, the cutoff value being determined from one or more samples obtained from pregnant female subjects without the pregnancy-associated disorder.

In another aspect, the present invention provides a method of determining, in a sample from a female subject pregnant with a fetus, a portion of RNA that is of fetal origin, the method comprising:

receiving a plurality of reads, wherein the reads are obtained from an analysis of RNA molecules obtained from the sample, the sample containing a mixture of maternal- and fetal-derived RNA molecules;

identifying, by a computer system, locations of the reads in a reference sequence;

identifying one or more informative maternal loci, each of which is homozygous in the fetus for a corresponding first allele and which is heterozygous in the pregnant female subject for the corresponding first allele and a corresponding second allele;

filtering the one or more informative maternal loci to identify one or more filtered informative maternal loci:

that are located within an expressed region of the reference sequence,

at which at least a first predetermined number of reads in the plurality of reads containing the corresponding first allele are located, and

at which at least a second predetermined number of reads in the plurality of reads containing the corresponding second allele are located,

for each of the filtered informative maternal loci:

determining a first number of reads located at the filtered informative maternal locus and containing the corresponding first allele,

determining a second number of reads located at the filtered informative maternal locus and containing the corresponding second allele,

calculating a sum of the first number and the second number,

calculating a maternal ratio of the second number divided by the sum,

determining a scalar representing a total expression at the filtered informative maternal locus relative to expression of the corresponding second allele in the pregnant female subject,

multiplying the maternal ratio by the scalar to obtain a maternal contribution,

calculating a fetal contribution as one minus the maternal contribution, and determining a portion of RNA in the sample that is of fetal origin, wherein the portion is an average of the fetal contributions for the filtered informative maternal loci.

In yet another aspect, the present invention provides a method of designating a genomic locus as a maternal or fetal marker by analyzing a sample from a female subject pregnant with a fetus, the method comprising:

receiving a plurality of reads, wherein the reads are obtained from an analysis of RNA molecules obtained from the sample, the sample containing a mixture of maternal- and fetal-derived RNA molecules;

identifying, by a computer system, locations of the reads in a reference sequence;

identifying one or more informative loci, each of which is homozygous in a first entity for a corresponding first allele and which is heterozygous in a second entity for the corresponding first allele and a corresponding second allele, wherein the first entity is the pregnant female subject or the fetus, and the second entity is the other one of the pregnant female subject and the fetus;

filtering the one or more informative loci to identify one or more filtered informative loci:

that are located within an expressed region of the reference sequence,

at which at least a first predetermined number of reads in the plurality of reads containing the corresponding first allele are located, and

at which at least a second predetermined number of reads in the plurality of reads containing the corresponding second allele are located,

for each of the filtered informative loci:

determining a first number of reads located at the filtered informative locus and containing the corresponding first allele,

determining a second number of reads located at the filtered informative locus and containing the corresponding second allele,

calculating a ratio of the first number and the second number,

designating the filtered informative locus as a marker for the second entity when the ratio exceeds a cutoff.

In yet another aspect, the present invention provides use of a gene listed in FIG. 21 or FIG. 26 for diagnosing a pregnancy-associated disorder in a female subject pregnant with a fetus by comparing an expression level of the gene with a control value determined from one or more other female subjects, each pregnant with a healthy fetus.

In yet another aspect, the present invention provides a system comprising one or more processors configured to perform any of the above methods.

In yet another aspect, the present invention provides a computer product comprising a computer readable medium storing a plurality of instructions for controlling a processor to perform an operation of any of the methods above.

[0005] Methods, systems, and apparatuses are provided for diagnosing pregnancy-associated disorders, determining allelic ratios, determining maternal or fetal contributions to circulating transcripts, and/or identifying maternal or fetal markers using a sample from a pregnant female subject. In some ~~embodiments~~examples, the sample is blood plasma containing a mixture of maternal- and fetal-derived RNA molecules.

[0006] The RNA molecules are analyzed (e.g., sequenced) to obtain a plurality of reads, and the locations of these reads in a reference sequence are identified (e.g., by sequence alignment). Informative loci are identified that are homozygous for a first allele in either the mother or fetus, and heterozygous for a first allele and a second allele in the other one of the mother and fetus. The informative loci are then filtered and reads located at (e.g., aligning to) the filtered informative loci are further analyzed. In some ~~embodiments~~examples, a ratio of reads corresponding to first alleles and second alleles is calculated and compared to a cutoff to diagnose a pregnancy-associated disorder. In some ~~embodiments~~examples, a portion of RNA in the sample that is of fetal origin is determined using reads located at filtered informative maternal loci. In some ~~embodiments~~examples, a ratio of reads corresponding to

the first allele and the second allele is calculated for an individual filtered informative locus, and the ratio is compared to a cutoff to designate the locus as a maternal or fetal marker.

[0007] The present methods are not limited to prenatal diagnostics and can be applied to any biological sample that contains a mixture of RNA molecules derived from two