

Claims

We claim:

1. A method for establishing a relationship between coverage depth and GC content of a chromosome, which method comprises:

5 obtaining sequence information of multiple polynucleotide fragments covering said chromosome and another chromosome from more than one sample;

assigning said fragments to chromosomes based on said sequence information;

calculating coverage depth and GC content of said chromosome based on said sequence information for each sample; and

10 determining the relationship between the coverage depth and GC content of said chromosome.

2. The method of claim 1, wherein the polynucleotide fragments range from about 10 to about 1000 bp in length.

3. The method of claim 2, wherein the polynucleotide fragments range from
15 about 15 to about 500 bp in length.

4. The method of claim 3, wherein the polynucleotide fragments range from about 20 to about 200 bp in length.

5. The method of claim 4, wherein the polynucleotide fragments range from about 25 to about 100 bp in length.

20 6. The method of claim 5, wherein the polynucleotide fragments are about 35 bp in length.

7. The method according to any of claims 1-6, wherein the assignment is by comparing the sequence of the fragments with a reference human genomic sequence.

8. The method of claim 7, wherein the reference human genomic sequence is hg18
25 or hg19.

9. The method of claim 7, wherein the fragments that assign to more than one chromosome are disregarded.

10. The method of claim 7, wherein the fragments that do not assign to any chromosome are disregarded.

30 11. The method according to any of claims 1-10, wherein the coverage depth of the chromosome is the ratio between the number of fragments that assigns to said chromosome and the number of Reference Unique Reads of said chromosome.

12. The method of claim 11, wherein the coverage depth is normalized.

13. The method of claim 12, wherein the normalization is calculated against the coverage of another chromosome.

14. The method of claim 12, wherein the normalization is calculated against the coverage of all other autosomes.

5 15. The method of claim 12, wherein the normalization is calculated against the coverage of all other chromosomes.

16. The method according to any of claims 1-15, wherein the GC content of the chromosome is the average GC content of all fragments that assign to said chromosome.

10 17. The method according to any of claims 1-16, wherein at least 2, 5, 10, 20, 50, 100, 200, 500 or 1000 samples are used.

18. The method according to any of claims 1-17, wherein the chromosome is chromosome 1, 2, ..., 22, X or Y.

19. The method of claim 12, wherein the relationship is in the formula:

$$cr_{i,j} = f(GC_{i,j}) + \varepsilon_{i,j}, j = 1, 2, \dots, 22, X, Y,$$

15 wherein $f(GC_{i,j})$ represents the function for the relationship between coverage depth and the corresponding GC content of sample i, chromosome j, $\varepsilon_{i,j}$ represents the residual of sample i, chromosome j.

20 20. The method according to any of claims 1-19, wherein the relationship between coverage depth and GC content is calculated by local polynomial regression.

21. The method of claim 20, wherein the relationship is a non-strong linear relationship.

22. The method of claim 21, wherein the relationship is determined by loess algorithm.

25 23. The method according to any of claims 19-22, further comprising: calculating fitted coverage depth according to the formula:

$$\hat{cr}_{i,j} = f(GC_{i,j}), j = 1, 2, \dots, 22, X, Y$$

24. The method of claim 23, further comprising: calculating standard variation according to the formula:

$$std_j = \sqrt{\sum_i (cr_{i,j} - \hat{cr}_{i,j})^2 / (ns - 1)}, j = 1, 2, \dots, 22, X, Y,$$

30 wherein ns represents the number of reference samples.

25. The method of claim 24, further comprising:
calculating student t-statistic according to the formula:

$$t_{i,j} = (cr_{i,j} - \hat{cr}_{i,j}) / std_j, j = 1, 2, \dots, 22, X, Y.$$

26. The method according to any of claims 1-25, wherein the samples are from
5 pregnant female subjects.

27. The method according to any of claims 1-25, wherein the samples are from male
subjects.

28. The method according to any of claims 1-27, wherein the samples are peripheral
blood samples.

10 29. The method according to any of claims 1-28, wherein the sequence information is
obtained by parallel genomic sequencing.

30. A method to determine a fetal genetic abnormality, which method comprises:

a) obtaining sequence information of multiple polynucleotide fragments
from a sample;

15 b) assigning said fragments to chromosomes based on said sequence information;

c) calculating coverage depth and GC content of a chromosome based on said
sequence information;

d) calculating fitted coverage depth of said chromosome using said GC content of
said chromosome and an established relationship between coverage depth and GC content of said
20 chromosome; and

e) comparing said fitted coverage depth to said coverage depth of said chromosome,
wherein a difference between them indicates fetal genetic abnormality.

31. The method of claim 30, wherein the established relationship between coverage
depth and GC content is obtained by the method of claims 1-29.

25 32. The method according to any of claims 30 or 31, wherein the genetic abnormality
is a chromosomal abnormality.

33. The method of claim 32, wherein the genetic abnormality is aneuploidy.

34. The method according to any of claims 30-33, further comprising:

f) determining the fetal gender.

30 35. The method of claim 34, wherein the fetal gender is determined according to the
formula:

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 cr.a_{i,x} + \beta_2 cr.a_{i,y},$$

wherein $cr_{i,x}$ and $cr_{i,y}$ are normalized relative coverage of X and Y chromosomes, respectively.

36. The method according to any of claims 34 or 35, further comprising:

g) estimating the fetal fraction.

37. The method of claim 36, wherein the fetal fraction is calculated according to the formula:

$$fy_i = (cr_{i,y} - \hat{cr}_{i,yf}) / (\hat{cr}_{i,ym} - \hat{cr}_{i,yf}),$$

wherein $\hat{cr}_{i,yf} = f(GC_{i,yf})$ is the fitted coverage depth calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of samples from pregnant women with a female fetus, and $\hat{cr}_{i,ym} = f(GC_{i,ym})$ refers to the fitted coverage depth calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of male subjects.

38. The method of claim 36, wherein the fetal fraction is calculated according to the formula:

$$fx_i = (cr_{i,x} - \hat{cr}_{i,xf}) / (\hat{cr}_{i,xm} - \hat{cr}_{i,xf}),$$

wherein $\hat{cr}_{i,xf} = f(GC_{i,xf})$ is the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from pregnant women with a female fetus, $\hat{cr}_{i,xm} = f(GC_{i,xm})$ refers to the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from male subjects.

39. The method of claim 36, wherein the fetal fraction is calculated according to the formula:

$$fxy_i = \arg \min_{\varepsilon \in (0,1)} \left(\frac{(\hat{cr}_{i,yf} \cdot (1-\varepsilon) + \hat{cr}_{i,ym} \cdot \varepsilon - cr_{i,x})^2}{(\hat{\sigma}_{X,f} \cdot (1-\varepsilon))^2 + (\hat{\sigma}_{X,m} \cdot \varepsilon)^2} + \frac{(\hat{cr}_{i,yf} \cdot (1-\varepsilon) + \hat{cr}_{i,ym} \cdot \varepsilon - cr_{i,x})^2}{(\hat{\sigma}_{Y,f} \cdot (1-\varepsilon))^2 + (\hat{\sigma}_{Y,m} \cdot \varepsilon)^2} \right),$$

wherein $\hat{cr}_{i,xf} = f(GC_{i,xf})$ is the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from pregnant women with a female fetus, $\hat{cr}_{i,yf} = f(GC_{i,yf})$ refers to the fitted coverage depth calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of samples from pregnant women with a female fetus, $\hat{cr}_{i,xm} = f(GC_{i,xm})$ refers to the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from male subjects, and $\hat{cr}_{i,ym} = f(GC_{i,ym})$ refers to the fitted coverage depth

calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of male subjects.

40. The method according to any of claims 30-39, wherein the comparison of said fitted coverage depth to said coverage depth of the chromosome is conducted by a statistical hypothesis test.

41. The method of claim 40, wherein one hypothesis is that the fetus is euploid (H0) and the other hypothesis is that the fetus is aneuploid (H1).

42. The method of claim 41, wherein the student t-statistic is calculated for both hypotheses.

43. The method of claim 42, wherein the student t-statistic is calculated for H0 and H1 according to formula: $t1_{i,j} = (cr_{i,j} - \hat{cr}_{i,j}) / std_j$ and $t2_{i,j} = (cr_{i,j} - \hat{cr}_{i,j}(1 + f_{xy,i,j} / 2)) / std_j$, respectively, wherein f_{xy} is fetal fraction.

44. The method of claim 43, wherein the log likelihood ratio of t1 and t2 is calculated according to formula: $L_{i,j} = \log(p(t1_{i,j}, degree | D)) / \log(p(t2_{i,j}, degree | T))$, wherein L_{ij} is Log likelihood ratio, wherein degree refers to a t distribution degree, D refers to Diploidy, T refers to Trisomy, and $p(t1_{i,j}, degree | *)$, $*$ =D,T represents conditional probability density given a t distribution degree.

45. The method of claim 33, wherein the fetal aneuploidy is a disorder for an autosome selected from the group consisting of trisomy 13, 18 and 21.

46. The method of claim 33, wherein the fetal aneuploidy is a disorder for a sex chromosome selected from the group consisting of XO, XXX, XXY and XYY.

47. The method of claim 46, wherein the fetal gender is female, and the student t-statistic is calculated according to formula: $t1_{i,x} = (cr_{i,x} - \hat{cr}_{i,x_f}) / std_{x_f}$, wherein $\hat{cr}_{i,x_f} = f(GC_{i,x_f})$ is the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from pregnant women with a female fetus.

48. The method of claim 47, wherein $|t1| > 3.13$ indicates the fetus is XXX or XO.

49. The method of claim 47, wherein $|t1| > 5$ indicates the fetus is XXX or XO.

50. The method of claim 46, wherein the fetal gender is male, and the student t-statistic is calculated according to formula: $t2_i = (cr_{i,x} - (1 - f_{y_i} / 2) \cdot \hat{cr}_{i,x_f}) / std_{x_f}$, wherein $\hat{cr}_{i,x_f} = f(GC_{i,x_f})$ is the fitted coverage depth calculated from the relationship of the

chromosome X coverage depth and corresponding GC content of samples from pregnant women with a female fetus.

51. The method of claim 50, wherein $|t1| > 3.13$ indicates the fetus may be XXY or XYY.

5 52. The method of claim 50, wherein $|t1| > 5$ indicates the fetus is XXY or XYY.

53. A method to determine a fetal genetic abnormality, which method comprises:

a) obtaining sequence information of multiple polynucleotide fragments covering a chromosome and another chromosome from more than one normal samples;

b) assigning said fragments to chromosomes based on said sequence information;

10 c) calculating coverage depth and GC content of said chromosome based on said sequence information from said normal samples;

d) determining the relationship between the coverage depth and GC content of said chromosome;

15 e) obtaining sequence information of multiple polynucleotide fragments from a biological sample;

f) assigning said fragments to chromosomes based on said sequence information from said biological sample;

g) calculating coverage depth and GC content of said chromosome based on said sequence information from said biological sample;

20 h) calculating fitted coverage depth of said chromosome using said GC content of said chromosome and said relationship between coverage depth and GC content of said chromosome; and

i) comparing said fitted coverage depth to said coverage depth of said chromosome, wherein a difference between them indicates fetal genetic abnormality.

25 54. A computer readable medium comprising a plurality of instructions for performing prenatal diagnosis of a fetal genetic abnormality, which comprises the steps of:

a) receiving sequence information of multiple polynucleotide fragments from a sample;

30 b) assigning said polynucleotide fragments to chromosomes based on said sequence information;

c) calculating coverage depth and GC content of a chromosome based on said sequence information;

35 d) calculating fitted coverage depth of said chromosome using said GC content of said chromosome and an established relationship between coverage depth and GC content of said chromosome; and

e) comparing said fitted coverage depth to said coverage depth of said chromosome, wherein a difference between them indicates genetic abnormality.

55. The computer readable medium of claim 54, wherein the comparison in step e) is conducted by a statistical hypothesis test.

5 56. The computer readable medium of claim 55, wherein one hypothesis is that the fetus is normal (H0) and the other hypothesis is that the fetus is abnormal (H1).

57. The computer readable medium of claim 56, wherein the student t-statistic is calculated for both hypotheses.

10 58. The computer readable medium according to any of claims 54-57, wherein the established relationship between coverage depth and GC content is obtained by the method of claims 1-29.

59. The computer readable medium according to any of claims 54-58, further comprising:

f) determining the fetal gender.

15 60. The computer readable medium of claim 59, further comprising:

g) estimating the fetal fraction.

61. The computer readable medium according to any of claims 54-60, wherein the fetal genetic abnormality is a chromosomal abnormality.

20 62. The computer readable medium of claim 61, wherein the fetal genetic abnormality is aneuploidy.

63. The computer readable medium of claim 62, wherein the aneuploidy is a disorder for an autosome selected from the group consisting of trisomy 13, 18 and 21.

64. The computer readable medium of claim 62, wherein the aneuploidy is a disorder for a sex chromosome selected from the group consisting of XO, XXX, XXY and XYY.

25 65. A system for determining fetal genetic abnormality, which method comprises:

a) means for obtaining sequence information of multiple polynucleotide fragments from a sample; and

b) a computer readable medium according to claims 54-64.

66. The system of claim 65, further comprising:

30 c) a biological sample obtained from a pregnant female subject, wherein the biological sample includes multiple polynucleotide fragments.

67. The system according to any of claim 65 or 66, wherein the means for obtaining sequence information is parallel genomic sequencing.

Claims

1. A computer-implemented method to determine a fetal genetic abnormality which is a chromosomal aneuploidy, which method comprises:
 - (a) obtaining sequence information of multiple polynucleotide fragments from a sample, said sample being a peripheral blood sample derived from a pregnant female and containing both maternal and fetal DNA;
 - (b) assigning said fragments to chromosomes based on said sequence information by comparing said fragments to the reference unique reads of the same size for each of said chromosomes, wherein reference unique reads are fragments of a chromosome that have a unique sequence which can be unambiguously assigned to a single chromosomal location based on a reference genomic sequence;
 - (c) determining coverage depth and GC content of a chromosome based on the sequence information for those fragments that assigned to reference unique reads of said chromosome in step (b),
wherein coverage depth is the ratio between the number of fragments assigned uniquely to said chromosome and the number of reference unique reads for said chromosome of the same fragment size based on said reference human genomic sequence;
 - (d) determining fitted coverage depth of said chromosome using said GC content of said chromosome and established relationship between coverage depth and GC content for said chromosome in the absence of aneuploidy, wherein said established relationship has been determined by a method comprising the steps of:
 - (i) obtaining sequence information of multiple polynucleotide fragments covering said chromosome from a plurality of euploid peripheral blood samples containing genomic DNA, wherein the fragment size is the same as the fragment size of the multiple polynucleotide fragments of step (a) above;
 - (ii) assigning said fragments to chromosomes based on said sequence information as in step (b) above;
 - (iii) determining coverage depth and GC content of said chromosome based on said sequence information for each euploid sample as in step (c) above;and

- (iv) using the coverage depth and GC content determined for each sample in step (iii) to determine the relationship between the coverage depth and GC content of said chromosome in the absence of aneuploidy; and
- (e) comparing said fitted coverage depth to the coverage depth of said chromosome determined in step (c), wherein a difference between them indicates fetal chromosomal aneuploidy.

2. The method of claim 1, wherein step (a) further comprises obtaining sequence information of multiple polynucleotide fragments from multiple different samples and the coverage depth is normalised to take account of differences in total number of sequence reads obtained for different samples, for example against the average coverage depth of another chromosome, preferably against the average coverage depth of all other autosomes or against the average coverage depth of all other chromosomes.

3. The method according to claim 1 or claim 2, wherein the GC content of the chromosome is determined as the average GC content of all the fragments that assign to said chromosome for the purpose of step (c).

4. The method according to claim 2 further comprising determining the fetal gender, for example according to the formula

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 cr.a_{i,x} + \beta_2 cr.a_{i,y}$$

wherein $cr.a_{i,x}$ and $cr.a_{i,y}$ are relative coverage depth of the X and Y chromosomes of sample i , respectively.

5. The method according to claim 2 further comprising estimating the fetal fraction, wherein the fetal fraction is calculated relying on the coverage depth of chromosome X and / or Y determined as in step (c) of claim 1 according to a formula selected from:

(i)
$$fy_i = (cr_{i,Y} - \hat{c}r_{i,Yf}) / (\hat{c}r_{i,Ym} - \hat{c}r_{i,Yf}),$$

wherein $cr_{i,Y}$ is the coverage depth of chromosome Y of sample i , $\hat{c}r_{i,Yf} = f(GC_{i,Yf})$ is the fitted coverage depth calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of samples from pregnant women with a female fetus, and $\hat{c}r_{i,Ym} = f(GC_{i,Ym})$ refers to the fitted coverage depth calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of male subjects; or

$$(ii) \quad \hat{f}x_i = (cr_{i,X} - \hat{c}r_{i,Xf}) / (\hat{c}r_{i,Xm} - \hat{c}r_{i,Xf}),$$

wherein $cr_{i,X}$ is the coverage depth of chromosome X of sample i , $\hat{c}r_{i,Xf} = f(GC_{i,Xf})$ is the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from pregnant women with a female fetus, $\hat{c}r_{i,Xm} = f(GC_{i,Xm})$ refers to the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from male subjects; or

$$(iii) \quad \hat{f}xy_i = \arg \min_{\varepsilon \in (0,1)} \left(\frac{(\hat{c}r_{i,Yf} \cdot (1-\varepsilon) + \hat{c}r_{i,Ym} \cdot \varepsilon - cr_{i,X})^2}{(\hat{\sigma}_{X,f} \cdot (1-\varepsilon))^2 + (\hat{\sigma}_{X,m} \cdot \varepsilon)^2} + \frac{(\hat{c}r_{i,Yf} \cdot (1-\varepsilon) + \hat{c}r_{i,Ym} \cdot \varepsilon - cr_{i,X})^2}{(\hat{\sigma}_{Y,f} \cdot (1-\varepsilon))^2 + (\hat{\sigma}_{Y,m} \cdot \varepsilon)^2} \right),$$

wherein $\hat{c}r_{i,Xf} = f(GC_{i,Xf})$ is the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from pregnant women with a female fetus, $\hat{c}r_{i,Yf} = f(GC_{i,Yf})$ refers to the fitted coverage depth calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of samples from pregnant women with a female fetus, $\hat{c}r_{i,Xm} = f(GC_{i,Xm})$ refers to the fitted coverage depth calculated from the relationship of the chromosome X coverage depth and corresponding GC content of samples from male subjects, and $\hat{c}r_{i,Ym} = f(GC_{i,Ym})$ refers to the fitted coverage depth calculated from the relationship of the chromosome Y coverage depth and corresponding GC content of male subjects.

6. The method according to claim 2, wherein the comparison of said fitted coverage depth to the coverage depth of the chromosome determined in step (c) is conducted by a statistical hypothesis test, wherein one hypothesis is that the fetus is euploid (H0) and the other hypothesis is that the fetus exhibits aneuploidy for said chromosome (H1).

7. The method of claim 6, wherein the student t-statistic is calculated for both hypotheses.

8. The method of claim 7, wherein the student t-statistic is calculated for H0 and H1 according to the formulae

$$t1_{i,j} = (cr_{i,j} - \hat{c}r_{i,j}) / std_j \text{ and}$$

$$t_{i,j} = (cr_{i,j} - \hat{cr}_{i,j}(1 + fxy_{i,j}/2)) / std_j$$

respectively, wherein fxy is the fetal fraction,

i is the sample index,

j is the chromosome number,

$$cr_{i,j} = f(GC_{i,j}) + \varepsilon_{i,j}, j = 1, 2, \dots, 22, X, Y$$

wherein $f(GC_{i,j})$ represents the function for the relationship between normalized coverage depth and the corresponding GC content of sample i , chromosome j , $\varepsilon_{i,j}$ represents the residual of sample i , chromosome j ,

$$\hat{cr}_{i,j} = f(GC_{i,j}), j = 1, 2, \dots, 22, X, Y$$

and represents the fitted coverage depth, and

std_j is the standard variation according to the formula

$$std_j = \sqrt{\sum_i (cr_{i,j} - \hat{cr}_{i,j})^2 / (ns - 1)}, j = 1, 2, \dots, 22, X, Y$$

wherein ns represents the number of reference samples.

9. The method of claim 8, wherein the log likelihood ratio of $t1$ and $t2$ is calculated according to formula: $L_{i,j} = \log(p(t1_{i,j}, \text{degree} | D)) / \log(p(t2_{i,j}, \text{degree} | T))$, wherein $L_{i,j}$ is Log likelihood ratio, wherein degree refers to a t distribution degree, D refers to Diploidy, T refers to Trisomy, and $p(t1_{i,j}, \text{degree} | *)$, $*$ = D, T represents conditional probability density given a t distribution degree, if the ratio is larger than 1 the fetus being inferred to exhibit trisomy for said chromosome.

10. The method according to any one of claims 1 to 9 for use in determining a fetal autosome aneuploidy.

11. The method of claim 10, wherein the fetal aneuploidy is selected from the group consisting of trisomy 13, 18 and 21.

12. The method according to claim 4 for use in determining a sex chromosome aneuploidy, such as a sex-chromosome aneuploidy selected from the group consisting of XO, XXX, XXY and XYY.

13. A computer readable medium comprising a plurality of instructions adapted to perform the method of any one of claims 1 to 12 when said instructions are supplemented with sequence information obtained as in step (a) of claim 1.

14. A system comprising means adapted for carrying out the method of any one of claims 1 to 12.