Claims

1. A method to identify one or more pharmacogenomic biomarkers, which method comprises:

a) isolating DNA from archived clinical samples of at least two patients exhibiting different values in a relevant phenotype;

b) amplifying said isolated DNA;

c) obtaining high-density genotyping data of said amplified DNA; and

d) performing association analysis based on said genotyping data and said different values in said relevant phenotype,

wherein said pharmacogenomic biomarker(s) are identified.

2. The method of claim 1, wherein the archived clinical samples are selected from the group consisting of plasma samples, serum samples, dried blood spots, urine samples, tissue samples, tumor cells and buccal swabs.

3. The method of claim 2, wherein the archived clinical samples are plasma samples.

4. The method according to any one of claims 1-3, wherein the isolated DNA is suboptimal genomic DNA.

5. The method according to any one of claims 1-4, wherein the amplification is whole-genome amplification (WGA), and the resulting DNA is whole-genome amplified DNA (wgaDNA).

6. The method according to any one of claims 1-5, wherein the high-density genotyping is whole-genome genotyping.

7. The method according to any one of claims 1-6, wherein the high-density genotyping is by using single nucleotide polymorphisms (SNPs).

699572000140

8. The method of claim 7, wherein about 1,000-5,000,000 or more SNPs are used.

9. The method of claim 8, wherein about 1,000,000 SNPs are used.

10. The method according to any one of claims 1-9, wherein the high-density genotyping is array based, bead based, or high-throughput sequencing based.

11. The method according to any one of claims 1-10, wherein the genotyping data is obtained by using a genome-wide genotype calling algorithm.

12. The method of claim 11, further comprising:

e) adjusting the call rate cut-off value of the genome-wide genotype calling algorithm.

13. The method of claim 12, wherein step d) and step e) are repeated multiple times to include and/or exclude samples.

14. The method of claim 13, wherein an optimal inclusion criterion is identified.

15. The method according to any one of claims 11-14, wherein the genotype calls are made by using a call rate cut-off that is lower than a typical call rate cut-off used for whole-genome genotyping of high quality genomic DNA.

16. The method of claim 12, wherein the call rate cut-off used is about 50-95%.

17. The method of claim 13, wherein the call rate cut-off used is about 80-90%.

18. The method of claim 14, wherein the call rate cut-off used is about 90%.

19. The method according to any one of claims 11-18, wherein the genotype calls are generated with the Affymetrix Genotyping Console™ software.

20. The method according to any one of claims 1-19, wherein the genotype calls are generated using the BRLMM algorithm.

21. The method according to any one of claims 1-20, wherein the genotype calls are made using an imputation algorithm.

22. The method of claim 21, wherein the HapMap is used for the imputation algorithm.

23. The method according to any one of claims 1-22, wherein the association analysis is a genome-wide association study (GWAS).

24. The method according to any one of claims 1-23, wherein the association analysis is performed by calculating an associated p-value of each SNP with the relevant phenotype.

25. The method of claim 24, wherein the calculation is based on an allele frequency and/or genotype-based test.

26. The method according to any one of claims 1-25, wherein the relevant phenotype is a categorical trait, a quantitative trait or another relevant phenotype.

27. The method according to any one of claims 1-26, wherein the archived clinical samples are from about 2-1,000 or more patients.

28. The method according to any one of claims 1-27, further comprising performing association analysis based on additional genotyping data using the identified pharmacogenomic biomarkers.

29. The method of claim 28, wherein about 1-500 or more of the identified pharmacogenomic biomarkers are used for the additional genotyping.

30. The method according to claim 28 or 29, wherein some or all of the archived clinical samples from step a) and/or additional clinical samples are used for the additional genotyping.

31. The method according to any one of claims 28-30, wherein the additional genotyping data is obtained by using a verification genotype calling algorithm.

32. The method of claim 31, further comprising adjusting the call rate cut-off value of the verification genotype calling algorithm.

33. The method of claim 32, wherein genotyping and adjusting the call rate cut-off are repeated multiple times to include and/or exclude samples.

34. The method of claim 33, wherein an optimal inclusion criterion is identified.

35. The method according to any one of claims 28-34, further comprising comparing the additional genotyping data obtained by using the verification genotype calling algorithm to the genotyping data obtained by using the genome-wide genotype calling algorithm.

36. The method according to any one of claims 28-35, wherein a subset of the pharmacogenomic biomarkers from step d) is identified.

37. The method of claim 36, wherein the method is used for retrospective study of archived clinical samples from a previously conducted clinical trial.

38. The method according to claim 36 or 37, wherein the method is used for *de novo* identification of a pharmacogenomic biomarker.

39. A pharmacogenomic biomarker identified by the method according to any one of claims 36-38.

40. A group of pharmacogenomic biomarkers identified by the method according to any one of claims 36-38.

41. The pharmacogenomic biomarker according to claim 39 or 40, wherein the biomarker is one or more SNPs.

42. The pharmacogenomic biomarker according to any one of claims 39-41, for use to identify one or more additional pharmacogenomic biomarkers.

43. The pharmacogenomic biomarker according to any one of claims 39-41, for use to develop a companion diagnostic test.

44. A companion diagnostic test using the pharmacogenomic biomarker according to any one of claims 39-41.

45. A method of prognosticating responsiveness of a subject to a treatment using the companion diagnostic test of claim 44.

46. A method of identifying a novel drug target using the pharmacogenomic biomarker according to any one of claims 39-41.

47. A kit comprising a reagent for assessing the pharmacogenomic biomarker identified by the method according to any one of claims 36-38.

48. The kit of claim 47, further comprising instructions for using the pharmacogenomic biomarker to conduct a companion diagnostic test.

49. The kit of claim 47, wherein the reagent is used to detect a polynucleotide and/or polypeptide molecule.

50. A genotyping method using suboptimal genomic DNA samples, which method comprises:

a) receiving sequence information of said suboptimal genomic DNA samples;

b) optimizing an inclusion criterion based on said sequence information; and

c) calculating genotypes based on said sequence information and said optimized inclusion criterion.

51. The method of claim 50, wherein the optimization is repeated multiple times to include and/or exclude samples.

52. The method of claim 51, wherein an optimal inclusion criterion is identified.

53. The method according to any one of claims 50-52, wherein the genotyping data is obtained by using a genome-wide genotype calling algorithm and/or a verification genotype calling algorithm.

54. The method of claim 53, wherein the inclusion criterion is the call rate cut-off value of the genotype calling algorithm.

55. The method of claim 54, wherein the genotype calls are made by using a call rate cut-off that is lower than a typical call rate cut-off used for whole-genome genotyping of high quality genomic DNA.

56. The method of claim 55, wherein the call rate cut-off used is about 50-95%.

57. The method of claim 56, wherein the call rate cut-off used is about 80-90%.

58. The method of claim 57, wherein the call rate cut-off used is about 90%.

59. The method according to any one of claims 50-58, wherein the genotyping data are obtained by using multiple genotyping platforms.

60. The method of claim 59, wherein the genotyping data from multiple genotyping platforms are compared for the optimization.

61. A method to perform association analysis using the genotyping method according to any one of claims 50-60.

62. The method of claim 61, wherein the association analysis is repeated multiple times for the optimization.

63. A computer readable medium comprising a plurality of instructions for a genotyping method using suboptimal genomic DNA samples, which comprises the steps of:

a) receiving sequence information of said suboptimal genomic DNA samples;

b) optimizing an inclusion criterion based on said sequence information using the method according to any one of claims 50-62; and

c) calculating genotypes based on said sequence information and said optimized inclusion criterion.

64. A method to conduct GWAS using suboptimal genomic DNA.

65. The method of claim 64, wherein the suboptimal genomic DNA is from archived samples.

66. The method according to claim 64 or 65, wherein the suboptimal genomic DNA is from plasma samples.

67. The method according to any one of claims 64-66, wherein multiple genotyping platforms are used.

68. The method of claim 67, wherein the same samples are used for the multiple genotyping platforms.

699572000140

69. The method of claim 67, wherein different samples are used for the multiple genotyping platforms.

70. The method according to any one of claims 64-69, further comprising a sample providing high-quality genomic DNA.

71. The method according to any one of claims 64-70, using the association analysis method according to claim 61 or 62.

72. A method to identify one or more pharmacogenomic biomarkers using the method according to any one of claims 64-71.

## Claims

1. A method to identify one or more pharmacogenomic biomarkers, which method comprises:

a) isolating DNA from archived clinical samples of at least two patients exhibiting different values in a relevant phenotype;

b) amplifying said isolated DNA;

c) obtaining high-density genotyping data of said amplified DNA; and

d) performing association analysis based on said genotyping data and said different values in said relevant phenotype,

wherein said pharmacogenomic biomarker(s) are identified.

2. The method of claim 1, wherein the archived clinical samples are selected from the group consisting of plasma samples, serum samples, dried blood spots, urine samples, tissue samples, tumor cells and buccal swabs.

3. The method of claim 2, wherein the archived clinical samples are plasma samples.

4. The method according to any one of claims 1-3, wherein the isolated DNA is suboptimal genomic DNA.

5. The method according to any one of claims 1-4, wherein the amplification is whole-genome amplification (WGA), and the resulting DNA is whole-genome amplified DNA (wgaDNA).

6. The method according to any one of claims 1-5, wherein the high-density genotyping is whole-genome genotyping.

7. The method according to any one of claims 1-6, wherein the high-density genotyping is by using single nucleotide polymorphisms (SNPs).

8. The method according to any one of claims 1-7, wherein the high-density genotyping is array based, bead based, or high-throughput sequencing based.

9. The method according to any one of claims 1-8, wherein the genotyping data is obtained by using a genome-wide genotype calling algorithm, further comprising:

adjusting the call rate cut-off value of the genome-wide genotype calling algorithm.

10. The method of claim 9, wherein the genotype calls are made by using a call rate cut-off that is lower than a typical call rate cut-off used for whole-genome genotyping of high quality genomic DNA.

11. The method of claim 9, wherein the call rate cut-off used is about 50-95%.

12. The method according to any one of claims 1-11, further comprising performing association analysis based on additional genotyping data using the identified pharmacogenomic biomarkers, wherein some or all of the archived clinical samples from step a) and/or additional clinical samples are used for the additional genotyping.

13. The method of claim 12, further comprising comparing the additional genotyping data obtained by using the verification genotype calling algorithm to the genotyping data obtained by using the genome-wide genotype calling algorithm.

14. The method according to claim 12 or 13, wherein a subset of the pharmacogenomic biomarkers from step d) is identified, wherein the method is used for retrospective study of archived clinical samples from a previously conducted clinical trial.

15. A pharmacogenomic biomarker identified by the method of claim 14, for use to develop a companion diagnostic test.

16. A companion diagnostic test using a pharmacogenomic biomarker identified by the method of claim 14.

17. A genotyping method using suboptimal genomic DNA samples, which method comprises:

a) receiving sequence information of said suboptimal genomic DNA samples;

b) optimizing an inclusion criterion based on said sequence information; and

c) calculating genotypes based on said sequence information and said optimized inclusion criterion.

18. The method of claim 17, wherein the optimization is repeated multiple times to include and/or exclude samples.

19. A computer readable medium comprising a plurality of instructions for a genotyping method using suboptimal genomic DNA samples, which comprises the steps of:

a) receiving sequence information of said suboptimal genomic DNA samples;

b) optimizing an inclusion criterion based on said sequence information using the method according to claim 17 or 18; and

c) calculating genotypes based on said sequence information and said optimized inclusion criterion.

20. A method to conduct GWAS using suboptimal genomic DNA.