



(12) 发明专利申请

(10) 申请公布号 CN 116157868 A

(43) 申请公布日 2023.05.23

(21) 申请号 202180056064.1

(74) 专利代理机构 北京纪凯知识产权代理有限公司 11245

(22) 申请日 2021.08.17

专利代理师 杨林

(30) 优先权数据

63/067,244 2020.08.18 US

63/163,434 2021.03.19 US

(51) Int.Cl.

G16B 20/00 (2006.01)

(85) PCT国际申请进入国家阶段日

2023.02.10

(86) PCT国际申请的申请数据

PCT/US2021/046272 2021.08.17

(87) PCT国际申请的公布数据

W02022/040163 EN 2022.02.24

(71) 申请人 德尔菲诊断公司

地址 美国马里兰州

(72) 发明人 J·凯利 N·C·德拉科波利

S·琼斯

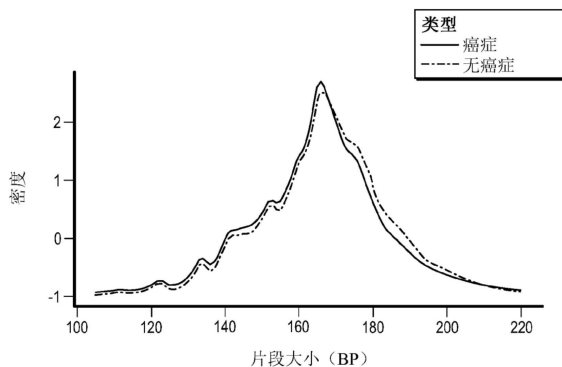
权利要求书8页 说明书15页 附图6页

(54) 发明名称

用于游离DNA片段大小密度以评估癌症的方法和系统

(57) 摘要

本公开提供了利用从患者获得的样本中的游离DNA (cfDNA) 片段大小密度的分析来诊断和预测癌症状态的方法和系统。还提供了包括测序仪的系统,所述测序仪被配置成生成样本的低覆盖度全基因组测序数据集。



平均片段大小密度的比较。

1. 一种确定受试者的癌症状态的方法,其包含:
 - (a) 分析来自受试者的样本中的游离DNA (cfDNA) 片段大小密度的曲线的形状,其中来自所述受试者的cfDNA的所述曲线的所述形状与来自健康受试者的参考样本的cfDNA的曲线的形状的差异指示所述受试者体内癌症;以及
 - (b) 任选地向所述受试者施用癌症治疗。
2. 根据权利要求1所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含将分布的有限混合拟合到片段大小的计数。
3. 根据权利要求2所述的方法,其中将分布的所述有限混合拟合到片段大小的计数包含量化所述样本的组分。
4. 根据权利要求3所述的方法,其中量化所述样本的组包含量化cfDNA片段大小密度的多条曲线。
5. 根据权利要求3所述的方法,其中所述样本包含12种组分。
6. 根据权利要求2所述的方法,其中所述分布包含截断正态分布。
7. 根据权利要求3所述的方法,还包含通过统计参数和对总体混合的贡献来表征所述组分,所述统计参数包含平均值、方差或形状。
8. 根据权利要求7所述的方法,还包含通过确定所述统计参数具有小于或等于1.1的多变量潜在比例缩减来评估收敛。
9. 根据权利要求2所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约10、50、100或105bp和大于约220、250、300、350bp或更大的片段大小。
10. 根据权利要求1所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含使用对给定长度的片段的计数进行多项式回归拟合的系数来量化所述曲线的所述形状。
11. 根据权利要求10所述的方法,还包含将片段的所述计数标准化为具有平均值0和方差1。
12. 根据权利要求10所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于105bp和大于170bp的片段大小。
13. 根据权利要求1所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状还包含:
 - (i) 将来自所述受试者的包含cfDNA片段的样本加工成测序文库;
 - (ii) 对所述测序文库进行低覆盖度全基因组测序,以获得经测序片段;
 - (iii) 将所述经测序片段映射到基因组以获得映射序列的窗口;以及
 - (iv) 分析映射序列的所述窗口以确定cfDNA片段长度。
14. 根据权利要求13所述的方法,其中所述映射序列包含数十至数千个窗口。
15. 根据权利要求13所述的方法,其中所述窗口是非重叠窗口。
16. 根据权利要求13所述的方法,其中所述窗口各自包含约500万个碱基对。
17. 根据权利要求13所述的方法,其中在每个窗口内确定cfDNA片段化图谱。
18. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱包含最大频率的片段大小。
19. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱包含具有变化频率的片段

大小的片段大小分布。

20. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱包含映射序列的所述窗口中的小cfDNA片段与大cfDNA片段的比率。

21. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱包含小cfDNA片段在整个基因组中的窗口中的序列覆盖度。

22. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱包含大cfDNA片段在整个基因组中的窗口中的序列覆盖度。

23. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱包含小cfDNA片段和大cfDNA片段在整个基因组中的窗口中的序列覆盖度。

24. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱在整个基因组上。

25. 根据权利要求17所述的方法,其中所述cfDNA片段化图谱在亚基因组区间上。

26. 一种确定受试者体内DNA-核小体相互作用动力学的方法,其包含使用根据权利要求1至25中任一项所述的方法分析受试者体内游离DNA (cfDNA) 片段大小密度的曲线的形状,其中所述曲线的所述形状指示所述DNA-核小体相互作用动力学。

27. 一种预测受试者的癌症状态的方法,其包含:

(a) 分析从所述受试者获得的样本中的游离DNA (cfDNA) 片段大小密度的曲线的形状;

(b) 将所述样本的cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较;以及

(c) 当所述样本的cfDNA片段大小密度的所述曲线的所述形状不同于所述参考曲线形状时,检测所述受试者体内癌症,从而预测所述受试者的所述癌症状态。

28. 根据权利要求27所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含将分布的有限混合拟合到片段大小的计数。

29. 根据权利要求28所述的方法,其中所述混合包含截断正态分布。

30. 根据权利要求29所述的方法,其中将截断正态分布的所述有限混合拟合到片段大小的计数包含量化所述样本的组分。

31. 根据权利要求30所述的方法,其中量化所述样本的组分包含量化cfDNA片段大小密度的多条曲线的形状的参数。

32. 根据权利要求30所述的方法,其中所述样本包含12种组分。

33. 根据权利要求30所述的方法,还包含通过统计参数和对总体混合的贡献来表征所述组分。

34. 根据权利要求33所述的方法,还包含通过确定所述统计参数具有小于或等于1.1的多变量潜在比例缩减来评估收敛。

35. 根据权利要求28所述的方法,其中分析cfDNA片段大小密度曲线的所述曲线的所述形状包含排除小于约10、50、100或105bp和大于约220、250、300、350bp或更大的片段大小。

36. 根据权利要求27所述的方法,其中所述参考曲线形状是在从健康受试者获得的样本中测量的cfDNA片段大小密度的曲线的形状。

37. 根据权利要求27所述的方法,其中所述癌症选自由以下组成的组:结肠直肠癌、肺癌、乳腺癌、胃癌、胰腺癌、胆管癌和卵巢癌。

38. 根据权利要求27所述的方法,其中比较步骤包含在整个基因组上将cfDNA片段大小

密度的曲线的所述形状与参考曲线形状进行比较。

39. 根据权利要求27所述的方法,其中所述比较步骤包含在亚基因组区间上将cfDNA片段大小密度的曲线的所述形状与参考曲线形状进行比较。

40. 一种治疗受试者体内癌症的方法,其包含:

(a) 检测所述受试者体内癌症,其中所述检测所述受试者体内癌症包含分析从所述受试者获得的样本中的游离DNA(cfDNA)片段大小密度的曲线的形状;将所述样本的cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较;并且

当所述样本的cfDNA片段大小密度的所述曲线的所述形状不同于所述参考曲线形状时,检测所述受试者体内癌症;以及

(b) 向所述受试者施用癌症治疗,

从而治疗所述受试者体内癌症。

41. 根据权利要求40所述的方法,其中所述受试者是人。

42. 根据权利要求40所述的方法,其中所述癌症选自以下组成的组:结肠直肠癌、肺癌、乳腺癌、胃癌、胰腺癌、胆管癌和卵巢癌。

43. 根据权利要求40所述的方法,其中所述癌症治疗选自以下组成的组:手术、辅助化疗、新辅助化疗、放疗、激素疗法、细胞毒性疗法、免疫疗法、过继性T细胞疗法、靶向疗法及其组合。

44. 根据权利要求40所述的方法,其中所述参考曲线形状是在从健康受试者获得的样本中测量的cfDNA片段大小密度的曲线的形状。

45. 根据权利要求40所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含将分布的有限混合拟合到片段大小的计数。

46. 根据权利要求45所述的方法,其中所述混合包含截断正态分布。

47. 根据权利要求46所述的方法,其中将截断正态分布的所述有限混合拟合到片段大小的计数包含量化所述样本的组分。

48. 根据权利要求47所述的方法,其中量化所述样本的组分包含量化cfDNA片段大小密度的多条曲线的形状的参数。

49. 根据权利要求47所述的方法,其中所述样本包含12种组分。

50. 根据权利要求47所述的方法,还包含通过统计参数和对总体混合的贡献来表征所述组分。

51. 根据权利要求50所述的方法,还包含通过确定所述统计参数具有小于或等于1.1的多变量潜在比例缩减来评估收敛。

52. 根据权利要求45所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于105bp和大于220bp的片段大小。

53. 根据权利要求40所述的方法,其中比较步骤包含在整个基因组上将cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较。

54. 根据权利要求40所述的方法,其中所述比较步骤包含在亚基因组区间上将cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较。

55. 一种监测受试者体内癌症的方法,其包含:

(a) 确定所述受试者体内癌症状态,其中通过分析从所述受试者获得的第一样本中的

游离DNA (cfDNA) 片段大小密度的曲线的形状来确定所述癌症状态;将所述第一样本的cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较;并且当所述第一样本的cfDNA片段大小密度的所述曲线的所述形状不同于所述参考曲线形状时,检测所述受试者体内癌症;

(b) 向所述受试者施用癌症治疗;

(c) 确定从所述受试者获得的第二样本的cfDNA片段大小密度的曲线的形状;以及

(d) 将所述第二样本的cfDNA片段大小密度的所述曲线的所述形状与所述第一样本的cfDNA片段大小密度的所述曲线的所述形状和/或与所述参考曲线形状进行比较,

从而监测所述受试者体内癌症。

56. 一种系统,其包含:

(a) 测序仪,所述测序仪被配置成生成样本的低覆盖度全基因组测序数据集;以及

(b) 具有非暂时性计算机可读介质的计算机系统,所述非暂时性计算机可读介质具有指令以:

(i) 处理所述低覆盖度全基因组测序数据集,以产生所述样本的片段大小密度的曲线;

(ii) 将所述样本的片段大小密度的所述曲线拟合到至少两组不同的已建立的统计参数,以产生至少两个建议曲线拟合;

(iii) 显示所述至少两个建议曲线拟合,使得用户能够选择所述至少两个建议曲线拟合中的至少一个来用于进一步处理;以及

(iv) 显示对应于(iii)的所选建议曲线拟合的建议曲线拟合线以及参考曲线拟合,使得能够在所选建议曲线拟合线与所述参考曲线拟合线之间进行比较。

57. 根据权利要求56所述的系统,其中所述计算机系统还包含:

提取单元,所述提取单元可操作以选择合适的分量用于曲线拟合分析;

曲线拟合单元,所述曲线拟合单元可操作以通过使用用户定义的等式来执行正态分布拟合的有限混合或执行多项式回归拟合;

曲线拟合优度分析单元,所述曲线拟合优度分析单元可操作以提供由所述曲线拟合单元生成的拟合质量的指示符;

曲线拟合参数表征单元,所述曲线拟合参数表征单元可操作以将曲线拟合参数与参考值进行比较来分类;以及

表征数据库,所述表征数据库用于存储拟合系数及其表征。

58. 根据权利要求56所述的系统,其中片段大小密度的所述曲线是样本的游离DNA (cfDNA) 片段大小密度的曲线。

59. 根据权利要求56所述的系统,其中拟合片段大小密度的所述曲线包含量化所述曲线的所述形状。

60. 根据权利要求56所述的系统,还包含可操作以显示所述曲线拟合线、所述参考曲线拟合线以及两者的比较的视觉显示装置。

61. 根据权利要求56所述的系统,还包含耦合到所述计算机系统的打印机,以打印显示所述曲线拟合线、所述参考曲线拟合线以及两者的比较的报告。

62. 一种用计算机程序编码的非暂时性计算机可读存储介质,所述程序包含指令,所述指令在由一个或多个处理器执行时使得所述一个或多个处理器执行操作以执行根据权利要求

要求1至55中任一项所述的方法。

63. 一种计算系统,其包含:存储器;和耦合到所述存储器的一个或多个处理器,所述一个或多个处理器被配置成执行操作以执行根据权利要求1至55中任一项所述的方法。

64. 一种确定受试者的癌症状态的方法,其包含:

(a) 分析来自受试者的样本中的二核小体DNA片段的游离DNA (cfDNA) 片段大小密度的曲线的形状,其中来自所述受试者的cfDNA的所述曲线的所述形状与来自健康受试者的参考样本的cfDNA的曲线的形状的差异指示所述受试者体内癌症;以及

(b) 任选地向所述受试者施用癌症治疗。

65. 根据权利要求64所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约230、240、250、260bp和大于约420、430、440、450bp或更大的片段大小。

66. 根据权利要求65所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约260bp和大于约440bp的片段大小。

67. 根据权利要求65所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含针对染色体臂确定长度在约260bp与440bp之间的片段的数目并计算每个臂的二核小体片段的比例。

68. 根据权利要求67所述的方法,还包含计算每个染色体臂中的二核小体片段的数目。

69. 根据权利要求65所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状还包含:

(i) 将来自所述受试者的包含cfDNA片段的样本加工成测序文库;

(ii) 对所述测序文库进行低覆盖度全基因组测序,以获得经测序片段;

(iii) 将所述经测序片段映射到基因组以获得映射序列的窗口;以及

(iv) 分析映射序列的所述窗口以确定cfDNA片段长度。

70. 根据权利要求69所述的方法,其中所述映射序列包含数十至数千个窗口。

71. 根据权利要求69所述的方法,其中所述窗口是非重叠窗口。

72. 根据权利要求69所述的方法,其中所述窗口各自包含约500万个碱基对。

73. 根据权利要求69所述的方法,其中在每个窗口内确定cfDNA片段化图谱。

74. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱包含最大频率的片段大小。

75. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱包含具有变化频率的片段大小的片段大小分布。

76. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱包含映射序列的所述窗口中的小cfDNA片段与大cfDNA片段的比率。

77. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱包含小cfDNA片段在整个基因组中的窗口中的序列覆盖度。

78. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱包含大cfDNA片段在整个基因组中的窗口中的序列覆盖度。

79. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱包含小cfDNA片段和大cfDNA片段在整个基因组中的窗口中的序列覆盖度。

80. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱在整个基因组上。

81. 根据权利要求73所述的方法,其中所述cfDNA片段化图谱在亚基因组区间上。

82. 一种确定受试者体内DNA-核小体相互作用动力学的方法,其包含使用根据权利要求64至81中任一项所述的方法分析受试者体内游离DNA (cfDNA) 片段大小密度的曲线的形状,其中所述曲线的所述形状指示所述DNA-核小体相互作用动力学。

83. 一种预测受试者的癌症状态的方法,其包含:

(a) 分析从所述受试者获得的样本中的二核小体DNA片段的游离DNA (cfDNA) 片段大小密度的曲线的形状;

(b) 将所述样本的cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较;以及

(c) 当所述样本的cfDNA片段大小密度的所述曲线的所述形状不同于所述参考曲线形状时,检测所述受试者体内癌症,从而预测所述受试者的所述癌症状态。

84. 根据权利要求83所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约230、240、250、260bp和大于约420、430、440、450bp或更大的片段大小。

85. 根据权利要求84所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约260bp和大于440bp的片段大小。

86. 根据权利要求84所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含针对染色体臂确定长度在约260bp与440bp之间的片段的数目并计算每个臂的二核小体片段的比例。

87. 根据权利要求86所述的方法,还包含计算每个染色体臂中的二核小体片段的数目。

88. 根据权利要求84所述的方法,其中所述参考曲线形状是在从健康受试者获得的样本中测量的cfDNA片段大小密度的曲线的形状。

89. 根据权利要求84所述的方法,其中所述癌症选自由以下组成的组:结肠直肠癌、肺癌、乳腺癌、胃癌、胰腺癌、胆管癌和卵巢癌。

90. 根据权利要求84所述的方法,其中比较步骤包含在整个基因组上将cfDNA片段大小密度的曲线的所述形状与参考曲线形状进行比较。

91. 根据权利要求84所述的方法,其中所述比较步骤包含在亚基因组区间上将cfDNA片段大小密度的曲线的所述形状与参考曲线形状进行比较。

92. 一种治疗受试者体内癌症的方法,其包含:

(a) 检测所述受试者体内癌症,其中所述检测所述受试者体内癌症包含分析从所述受试者获得的样本中的二核小体DNA片段的游离DNA (cfDNA) 片段大小密度的曲线的形状;将所述样本的cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较;并且当所述样本的cfDNA片段大小密度的所述曲线的所述形状不同于所述参考曲线形状时,检测所述受试者体内癌症;以及

(b) 向所述受试者施用癌症治疗,从而治疗所述受试者体内癌症。

93. 根据权利要求92所述的方法,其中所述受试者是人。

94. 根据权利要求92所述的方法,其中所述癌症选自由以下组成的组:结肠直肠癌、肺癌、乳腺癌、胃癌、胰腺癌、胆管癌和卵巢癌。

95. 根据权利要求92所述的方法,其中所述癌症治疗选自由以下组成的组:手术、辅助

化疗、新辅助化疗、放疗、激素疗法、细胞毒性疗法、免疫疗法、过继性T细胞疗法、靶向疗法及其组合。

96. 根据权利要求92所述的方法,其中所述参考曲线形状是在从健康受试者获得的样本中测量的cfDNA片段大小密度的曲线的形状。

97. 根据权利要求92所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约230、240、250、260bp和大于约420、430、440、450bp或更大的片段大小。

98. 根据权利要求97所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约260bp和大于440bp的片段大小。

99. 根据权利要求97所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含针对染色体臂确定长度在约260bp与440bp之间的片段的数目并计算每个臂的二核小体片段的比例。

100. 根据权利要求99所述的方法,还包含计算每个染色体臂中的二核小体片段的数目。

101. 根据权利要求92所述的方法,其中比较步骤包含在整个基因组上将cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较。

102. 根据权利要求92所述的方法,其中所述比较步骤包含在亚基因组区间上将cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较。

103. 一种监测受试者体内癌症的方法,其包含:

(a) 确定所述受试者体内癌症状态,其中通过分析从所述受试者获得的第一样本中的二核小体DNA片段的游离DNA (cfDNA) 片段大小密度的曲线的形状来确定所述癌症状态;将所述第一样本的cfDNA片段大小密度的所述曲线的所述形状与参考曲线形状进行比较;并且当所述第一样本的cfDNA片段大小密度的所述曲线的所述形状不同于所述参考曲线形状时,检测所述受试者体内癌症;

(b) 向所述受试者施用癌症治疗;

(c) 确定从所述受试者获得的第二样本中的二核小体DNA片段的cfDNA片段大小密度的曲线的形状;以及

(d) 将所述第二样本的cfDNA片段大小密度的所述曲线的所述形状与所述第一样本的cfDNA片段大小密度的所述曲线的所述形状和/或与所述参考曲线形状进行比较,

从而监测所述受试者体内癌症。

104. 根据权利要求102所述的方法,其中分析所述第一样本和所述第二样本中的cfDNA片段大小密度的所述曲线的所述形状包含排除小于约230、240、250、260bp和大于约420、430、440、450bp或更大的片段大小。

105. 根据权利要求103所述的方法,其中分析cfDNA片段大小密度的所述曲线的所述形状包含排除小于约260bp和大于440bp的片段大小。

106. 一种用计算机程序编码的非暂时性计算机可读存储介质,所述程序包含指令,所述指令在由一个或多个处理器执行时使得所述一个或多个处理器执行操作以执行根据权利要求64至104中任一项所述的方法。

107. 一种计算系统,其包含:存储器;和耦合到所述存储器的一个或多个处理器,所述一个或多个处理器被配置成执行操作以执行根据权利要求64至104中任一项所述的方法。

108. 一种系统,其包含:

- (a) 测序仪,所述测序仪被配置成生成样本的低覆盖度全基因组测序数据集;以及
- (b) 根据权利要求105所述的非暂时性计算机可读存储介质或根据权利要求106所述的计算机系统。

用于游离DNA片段大小密度以评估癌症的方法和系统

[0001] 相关申请的交叉引用

[0002] 本申请根据35U.S.C. §119 (e) 要求于2020年8月18日提交的美国临时申请第63/067,244号和于2021年3月19日提交的美国临时申请第63/163,434号的优先权权益。先前申请的公开内容被认为是本申请的公开内容的一部分,并且在此以全文引用的方式并入本申请的公开内容中。

[0003] 发明背景

技术领域

[0004] 本发明总体上涉及遗传分析,并且更具体地涉及用于分析游离DNA片段大小密度以检测和/或评估受试者体内癌症的方法和系统。

背景技术

[0005] 世界范围内人类癌症的发病率和死亡率很大程度上是由于这些疾病的诊断较晚,在这种情况下治疗效果较差。不幸的是,临床证实的可以用于广泛诊断和治疗早期癌症患者的生物标记物并不广泛可得。

[0006] 对游离DNA (cfDNA) 的分析表明此类方法可以为早期诊断提供新的途径。已经表明循环肿瘤DNA (ctDNA) 片段平均比来自非肿瘤细胞的其它cfDNA更短。先前的工作已经探索了将片段分成由与组蛋白核心或接头蛋白结合引起的不同大小的组(例如,短和长,或相互排斥的大小组)并且使用这些片段的计数来量化ctDNA和/或将各个样本分类为存在/不存在肿瘤。然而,先前的研究忽略了片段大小密度的曲线的形状的重要性。

[0007] 照此,需要一种利用对片段大小密度的曲线的形状的分析的癌症检测和/或评估方法,以允许更稳健和可靠地检测受试者体内癌症。

发明内容

[0008] 本公开提供了利用对从患者获得的样本中的cfDNA片段大小密度的曲线的形状的分析的方法和系统。在本文中证明了曲线的形状是对癌症状态的有力预测。

[0009] 照此,在一个实施例中,本发明提供了确定受试者的癌症状态的方法。该方法包括:(a) 分析来自受试者的样本中的cfDNA片段大小密度的曲线的形状,其中来自受试者的cfDNA片段大小密度的曲线的形状与来自健康受试者的参考样本的cfDNA片段大小密度的曲线的形状的差异指示受试者体内癌症;和(b) 任选地向受试者施用癌症治疗。

[0010] 在另一实施例中,本发明提供了确定受试者体内DNA-核小体相互作用动力学的方法,该方法包括使用本发明的方法分析受试者体内cfDNA片段大小密度的曲线的形状。在某些方面,曲线的形状指示DNA-核小体相互作用动力学。

[0011] 在又一实施例中,本发明提供了预测受试者的癌症状态的方法。该方法包括:(a) 分析从受试者获得的样本中的cfDNA片段大小密度的曲线的形状;(b) 将样本的cfDNA片段大小密度的曲线的形状与参考曲线形状进行比较;以及(c) 当样本的cfDNA片段大小密度的

曲线的形状不同于参考曲线形状时,检测受试者体内癌症,从而预测受试者的癌症状态。

[0012] 在另一实施例中,本发明提供了诊断和治疗受试者体内癌症的方法。该方法包括:(a)检测受试者体内癌症,其中所述检测受试者体内癌症包括分析从受试者获得的样本中的cfDNA片段大小密度的曲线的形状,将样本的cfDNA片段大小密度的曲线的形状与参考曲线形状进行比较,以及当样本的cfDNA片段大小密度的曲线的形状不同于参考曲线形状时,检测受试者体内癌症;和(b)向受试者施用癌症治疗,从而治疗受试者体内癌症。

[0013] 在再一实施例中,本发明提供了监测受试者体内癌症的方法。该方法包括:(a)确定受试者体内癌症状态,其中通过分析从受试者获得的第一样本中的cfDNA片段大小密度的曲线的形状来确定癌症状态;将第一样本的cfDNA片段大小密度的曲线的形状与参考曲线形状进行比较;以及当第一样本的cfDNA片段大小密度的曲线的形状不同于参考曲线形状时,检测受试者体内癌症;(b)向受试者施用癌症治疗;(c)确定从受试者获得的第二样本的cfDNA片段大小密度的曲线的形状;以及(d)将第二样本的cfDNA片段大小密度的曲线的形状与第一样本的cfDNA片段大小密度的曲线的形状和/或与参考曲线形状进行比较,从而监测受试者体内癌症。

[0014] 在另一实施例中,本发明提供了用于遗传分析和评估癌症的系统。该系统包括:(a)测序仪,该测序仪被配置成生成样本的低覆盖度全基因组测序数据集;和(b)计算机系统。在各个方面,该计算机系统具有非暂时性计算机可读介质,该非暂时性计算机可读介质具有指令以执行以下各项中的一者或多者:(i)处理低覆盖度全基因组测序数据集,以产生样本的片段大小密度的曲线;(ii)将样本的片段大小密度的曲线拟合到至少两组不同的已建立的统计参数,以产生至少两个建议曲线拟合;(iii)显示该至少两个建议曲线拟合,使得用户能够选择所述至少两个建议曲线拟合中的至少一个来用于进一步处理;以及(iv)显示对应于(iii)的所选建议曲线拟合的建议曲线拟合线以及参考曲线拟合线,使得能够在所选建议曲线拟合线与参考曲线拟合线之间进行比较。

[0015] 在本发明的各个方面,分析cfDNA片段大小密度的曲线的形状包括分析各种片段大小。在一些方面,分析cfDNA片段大小密度的曲线的形状排除小于约10、50、100或105bp和大于约220、250、300、350bp或更大的片段大小。在一些方面,分析cfDNA片段大小密度的曲线的形状排除小于约105bp和大于约170bp的片段大小。在一些方面,分析cfDNA片段大小密度的曲线的形状是关于二核小体DNA片段。在一些方面,分析cfDNA片段大小密度的曲线的形状排除小于约230、240、250、260bp和大于约420、430、440、450bp或更大的片段大小。在一些方面,分析cfDNA片段大小密度的曲线的形状排除小于约260bp和大于约440bp的片段大小。

[0016] 在再一实施例中,本发明提供了用计算机程序编码的非暂时性计算机可读存储介质。该计算机程序包括指令,这些指令在由一个或多个处理器执行时使得该一个或多个处理器执行操作以执行本发明的方法。

[0017] 在又一实施例中,本发明提供了计算系统。该系统包括存储器和耦合到该存储器的一个或多个处理器,其中该一个或多个处理器被配置成执行实施本发明的方法的操作。

[0018] 在又一实施例中,本发明提供了用于遗传分析和评估癌症的系统,该系统包括:(a)测序仪,该测序仪被配置成生成样本的全基因组测序数据集;和(b)本发明的非暂时性计算机可读存储介质和/或计算机系统。

附图说明

[0019] 图1是示出在本发明的一个实施例中使用本公开的方法生成的数据的曲线图。

[0020] 图2是示出在本发明的一个实施例中使用本公开的方法生成的数据的曲线图。

[0021] 图3是示出在本发明的一个实施例中使用本公开的方法生成的数据的曲线图。

[0022] 图4是示出在本发明的一个实施例中使用本公开的方法生成的数据的曲线图。

[0023] 图5是示出在本发明的一个实施例中使用本公开的方法生成的数据的曲线图。

[0024] 图6是示出由二核小体的RSC滑动产生的各种状态的图像。Montel等人(2011.“寡核小体的RSC重塑:原子力显微镜研究(RSC remodeling of oligo-nucleosomes:an atomic force microscopy study)”《核酸研究(Nucleic Acids Research)》39(7):2571-2579.)

[0025] 图7是示出导致不同片段大小的潜在内切核酸酶切割位点的实例的图像。每种状态仅示出了两个位点,但切割可以发生在接头DNA中的多个不同位置处。Montel等人(2011)。

[0026] 图8是示出在本发明的一个实施例中使用本公开的方法生成的数据的曲线图。

[0027] 图9是示出在本发明的一个实施例中使用本公开的方法生成的数据的曲线图。

具体实施方式

[0028] 本发明基于创新的方法和系统,这些方法和系统利用了对来自患者来源样本中的cfDNA的cfDNA片段大小密度的曲线的形状的分析。如本文所讨论的,本发明使用以下各项以两种方法来量化曲线的形状:1)多项式回归;和2)贝叶斯有限混合模型。本发明的方法和系统表示一种用于总结DNA-核小体相互作用动力学的新方法,因为它们与受试者的癌症状态相关。

[0029] 在描述本发明的组合物和方法之前,应当理解本发明不限于所描述的特定方法和系统,因为此类方法和系统可以变化。还应理解,本文所使用的术语仅用于描述具体实施例的目的,并且不旨在为限制性的,因为本发明的范围仅在所附权利要求中受限制。

[0030] 如在本说明书和所附权利要求书中使用的,单数形式“一个(a)”、“一种(an)”和“所述(the)”包含复数指示物,除非上下文另有明确规定。因此,例如,对“所述方法”的提及包含本文所描述的类型的一种或多种方法和/或步骤,在阅读本公开等后,所述一种或多种方法和/或步骤对于本领域的技术人员而言将变得显而易见。

[0031] 除非另外定义,否则本文所用的所有技术和科学术语具有与该发明所属领域中的普通技术人员通常所理解相同的含义。尽管在本发明的实践或测试中可以使用类似于或等同于本文中所描述的方法和材料的任何方法和材料,但现在描述优选的方法和材料。

[0032] 本公开提供了用于分析cfDNA片段大小密度以检测或以其它方式评估癌症的创新的方法和系统。

[0033] 在一个方面,用于开发用于量化片段大小密度的曲线的形状的本发明的方法的数据是基于浅表全基因组序列数据(1-2x覆盖度)。图1示出了平均片段大小密度的比较。在图1中,绘制了通过DNA读段的量归一化的癌症和无癌症个体的平均片段大小密度。

[0034] 如先前研究所指示,平均而言,无癌症个体具有较长的cfDNA片段(平均大小为167.09bp),而患有癌症的个体具有较短的cfDNA片段(平均大小为164.88bp)。然而,除了短

片段的增加之外,图1还示出了这些密度的总体形状的差异。此外,这些差异未由特定峰的周期性(如在先前工作中利用的周期性)捕捉。相反,片段大小密度的这些差异可以表示DNA-核小体相互作用动力学的差异,并且需要仓计数或周期性之外的替代方法来精确建模。本文所述的方法表示一种用于总结DNA-核小体相互作用动力学的新方法,因为它们与血浆供体的癌症状态相关。

[0035] 在各个方面,本公开证明片段大小密度可以被建模为分布的混合,其参数可以预测癌症状态。在一些方面,本公开绘示了使用紧密对应于单个核小体的大小的片段大小(大小小于260bp)。这些通常包括用147bp的DNA包裹的组蛋白八聚体,连同H1组蛋白和接头DNA(20bp),从而得到观察到的167bp大小的主峰。本公开还绘示了使用大于260bp的cfDNA片段,可能包括两个核小体,它们导致中值为334bp的峰。

[0036] 因此,在一个实施例中,本发明提供了确定受试者的癌症状态的方法。该方法包括:(a)分析来自受试者的样本中的cfDNA片段大小密度的曲线的形状,其中来自受试者的cfDNA片段大小密度的曲线的形状与来自健康受试者的参考样本的cfDNA片段大小密度的曲线的形状的差异指示受试者体内癌症;和(b)任选地向受试者施用癌症治疗。

[0037] 在另一实施例中,本发明提供了预测受试者的癌症状态的方法。该方法包括:(a)分析从受试者获得的样本中的cfDNA片段大小密度的曲线的形状;(b)将样本的cfDNA片段大小密度的曲线的形状与参考曲线形状进行比较;以及(c)当样本的cfDNA片段大小密度的曲线的形状不同于参考曲线形状时,检测受试者体内癌症,从而预测受试者的癌症状态。

[0038] 在再一实施例中,本发明提供了治疗受试者体内癌症的方法。该方法包括:(a)检测受试者体内癌症,其中所述检测受试者体内癌症包括分析从受试者获得的样本中的cfDNA片段大小密度的曲线的形状,将样本的cfDNA片段大小密度的曲线的形状与参考曲线形状进行比较,以及当样本的cfDNA片段大小密度的曲线的形状不同于参考曲线形状时,检测受试者体内癌症;和(b)向受试者施用癌症治疗,从而治疗受试者体内癌症。

[0039] 在另一实施例中,本发明提供了监测受试者体内癌症的方法。该方法包括:(a)确定受试者体内癌症状态,其中通过分析从受试者获得的第一样本中的cfDNA片段大小密度的曲线的形状来确定癌症状态;将第一样本的cfDNA片段大小密度的曲线的形状与参考曲线形状进行比较;以及当第一样本的cfDNA片段大小密度的曲线的形状不同于参考曲线形状时,检测受试者体内癌症;(b)向受试者施用癌症治疗;(c)确定从受试者获得的第二样本的cfDNA片段大小密度的曲线的形状;以及(d)将第二样本的cfDNA片段大小密度的曲线的形状与第一样本的cfDNA片段大小密度的曲线的形状和/或与参考曲线形状进行比较,从而监测受试者体内癌症。

[0040] 在各个方面,本发明的方法包括通过将分布的有限混合拟合到片段大小的计数来分析cfDNA片段大小密度的曲线的形状。在某些方面,将分布的有限混合拟合到片段大小的计数包括量化样本的组分,其可以包括量化cfDNA片段大小密度的多条曲线。在一些方面,样本包括约2、3、4、5、6、7、8、9、10、11、12、13、14、15或更多种组分。在一个说明性方面,如实例1中所讨论的,样本包括12种组分。

[0041] 在各个方面,分布包括截断正态分布。

[0042] 在各个方面,方法包括通过统计参数和对总体混合的贡献来表征样本的组分。此类统计参数可以包括但不限于平均值、方差和/或形状以及权重。

[0043] 本发明的方法还可以包括通过确定统计参数具有小于或等于或约为1.5、1.4、1.3、1.2、1.1或1.0的多变量潜在比例缩减因子来评估收敛。

[0044] 在各个方面,分析cfDNA片段大小密度的曲线的形状包括排除小于约10、50、100或105bp和大于约220、250、300、350bp或更大的片段大小的连续范围。在一个方面,分析cfDNA片段大小密度的曲线的形状包含排除小于105bp和大于170bp的片段大小。

[0045] 在某些方面,分析cfDNA片段大小密度的曲线的形状是关于二核小体DNA片段。在一些方面,分析cfDNA片段大小密度的曲线的形状排除小于约230、240、250、260bp和大于约420、430、440、450bp或更大的片段大小。在一个方面,分析cfDNA片段大小密度的曲线的形状排除小于260bp和大于440bp的片段大小。

[0046] 附加地,分析cfDNA片段大小密度的曲线的形状包含使用对给定长度的片段的计数进行多项式回归拟合的系数来量化曲线的形状。如实例1所示,方法还可以包括将片段的计数标准化为具有平均值0和方差1。

[0047] 在各个方面,分析cfDNA片段大小密度的曲线的形状可以包括以下各项中的一者或多者:(i)将来自受试者的包含cfDNA片段的样本加工成测序文库;(ii)对测序文库进行低覆盖度全基因组测序以获得经测序片段;(iii)将经测序片段映射到基因组以获得映射序列的窗口;以及(iv)分析映射序列的窗口以确定cfDNA片段长度。

[0048] 在某些方面,映射序列包括数十至数千个基因组窗口,如10、50、100至1,000、5,000、10,000或更多个窗口。此类窗口可以是非重叠的或重叠的,并且包括约100万、200万、300万、400万、500万、600万、700万、800万、900万或1000万个碱基对。

[0049] 在各个方面,在每个窗口内确定cfDNA片段化图谱。照此,本发明提供了用于确定受试者体内(例如,从受试者获得的样本中的)cfDNA片段化图谱的方法。如本文所用,术语“片段化图谱”、“片段化图式的位置依赖性差异”和“片段大小和以位置依赖性方式在整个基因组中的覆盖度的差异”是等同的并且可以互换使用。

[0050] 在一些方面,通过分析cfDNA片段大小密度的曲线的形状,确定受试者体内cfDNA片段化图谱可以用于鉴定受试者患有癌症。例如,可以对从受试者(例如,从受试者获得的样本)获得的cfDNA片段进行低覆盖度全基因组测序,并且可以将经测序片段映射到参考人类基因组(例如,在非重叠窗口中)并进行评估,以确定cfDNA片段化图谱和所分析的cfDNA片段大小密度的曲线的形状。如本文所述,患有癌症的受试者的cfDNA片段化图谱比健康受试者(例如,未患有癌症的受试者)的cfDNA片段化图谱更不均匀(例如,片段长度更不均匀)。

[0051] 在一些方面,cfDNA片段化图谱包括限定cfDNA片段大小密度的曲线的峰的最大频率的片段大小。在一些方面,cfDNA片段化图谱包括具有变化频率的片段大小的片段大小分布。在一些方面,cfDNA片段化图谱包括映射序列的所述窗口中的小cfDNA片段与大cfDNA片段的比率。在一些方面,cfDNA片段化图谱包括小cfDNA片段在整个基因组中的窗口中的序列覆盖度。在一些方面,cfDNA片段化图谱包括大cfDNA片段在整个基因组中的窗口中的序列覆盖度。在一些方面,cfDNA片段化图谱包括小cfDNA片段和大cfDNA片段在整个基因组中的窗口中的序列覆盖度。在一些方面,cfDNA片段化图谱在整个基因组上。在一些方面,cfDNA片段化图谱在亚基因组区间上。

[0052] cfDNA片段化图谱可以包括一个或多个cfDNA片段化图式。cfDNA片段化图式可以

包括任何适当cfDNA片段化图式。cfDNA片段化图式的实例包括但不限于中值片段大小、片段大小分布、小cfDNA片段与大cfDNA片段的比率和cfDNA片段的覆盖度。在一些方面，cfDNA片段化图式包括两个或更多个(例如，两个、三个或四个)中值片段大小、片段大小分布、小cfDNA片段与大cfDNA片段的比率和cfDNA片段的覆盖度。在一些方面，cfDNA片段化图谱可以是全基因组cfDNA图谱(例如，整个基因组中的窗口中的全基因组cfDNA图谱)。在一些方面，cfDNA片段化图谱可以是靶向区域图谱。靶向区域可以是基因组的任何适当部分(例如，染色体区域)。可以如本文所述确定cfDNA片段化图谱的染色体区域的实例包括但不限于染色体的一部分(例如，2q、4p、5p、6q、7p、8q、9q、10q、11q、12q和/或14q的一部分)和染色体臂(例如，8q、13q、11q和/或3p的染色体臂)。在一些方面，cfDNA片段化图谱可以包括两个或更多个靶向区域图谱。

[0053] 在一些方面，cfDNA片段化图谱可以用于鉴定cfDNA片段长度的变化(例如，改变)。改变可以是全基因组改变或一个或多个靶向区域/基因座的改变。靶区域可以是含有一个或多个癌症特异性改变的任何区域。在一些方面，cfDNA片段化图谱可以用于鉴定(例如，同时鉴定)约10个改变至约500个改变(例如，约25个至约500个、约50个至约500个、约100个至约500个、约200个至约500个、约300个至约500个、约10个至约400个、约10个至约300个、约10个至约200个、约10个至约100个、约10个至约50个、约20个至约400个、约30个至约300个、约40个至约200个、约50个至约100个、约20个至约100个、约25个至约75个、约50个至250个或约100个至约200个改变)。

[0054] 在各个方面，cfDNA片段化图谱可以包括cfDNA片段大小图式。cfDNA片段可以是任何适当大小。例如，在一些方面，cfDNA片段的长度可以为约50个碱基对(bp)至约400bp。如本文所述，患有癌症的受试者的cfDNA片段大小图式可以包含比健康受试者的中值cfDNA片段大小更短的中值cfDNA片段大小。健康受试者(例如，未患有癌症的受试者)可以具有中值cfDNA片段大小为约166.6bp至约167.2bp(例如，约166.9bp)的cfDNA片段大小。在一些方面，患有癌症的受试者可以具有比健康受试者的cfDNA片段大小平均短约1.28bp至约2.49bp(例如，约1.88bp)的cfDNA片段大小。例如，患有癌症的受试者可以具有中值cfDNA片段大小为约164.11bp至约165.92bp(例如，约165.02bp)的cfDNA片段大小。

[0055] 在一些方面，二核小体cfDNA片段的长度可以为约230个碱基对(bp)至约450bp。如本文所述，患有癌症的受试者的二核小体cfDNA片段大小图式可以包含比健康受试者的中值二核小体cfDNA片段大小更短的中值二核小体cfDNA片段大小。在一些方面，如图5所示，很明显，平均而言，无癌症受试者在二核小体范围内具有较长的cfDNA片段(平均大小为334.75bp)，而患有癌症的受试者具有较短的二核小体cfDNA片段(平均大小为329.6bp)。照此，健康受试者(例如，未患有癌症的受试者)可以具有中值cfDNA片段大小为约334.75bp的二核小体cfDNA片段大小。在一些方面，患有癌症的受试者可以具有比健康受试者的二核小体cfDNA片段大小更短的二核小体cfDNA片段大小。例如，患有癌症的受试者可以具有中值cfDNA片段大小为约329.6bp的二核小体cfDNA片段大小。

[0056] cfDNA片段化图谱可以包括cfDNA片段大小分布。如本文所述，患有癌症的受试者可以具有比健康受试者的cfDNA片段大小分布更多变的cfDNA大小分布。在一些方面，大小分布可以在靶向区域内。健康受试者(例如，未患有癌症的受试者)可以具有约1或小于约1的靶向区域cfDNA片段大小分布。在一些方面，患有癌症的受试者可以具有比健康受试者的

靶向区域cfDNA片段大小分布更长(例如,长10、15、20、25、30、35、40、45、50或更多bp,或这些数目之间的任何数目的碱基对)的靶向区域cfDNA片段大小分布。在一些方面,患有癌症的受试者可以具有比健康受试者的靶向区域cfDNA片段大小分布更短(例如,短10、15、20、25、30、35、40、45、50或更多bp,或这些数目之间的任何数目的碱基对)的靶向区域cfDNA片段大小分布。在一些方面,患有癌症的受试者可以具有比健康受试者的靶向区域cfDNA片段大小分布小约47bp至长约30bp的靶向区域cfDNA片段大小分布。在一些方面,患有癌症的受试者可以具有cfDNA片段的长度平均相差10、11、12、13、14、15、15、17、18、19、20或更多bp的靶向区域cfDNA片段大小分布。例如,患有癌症的受试者可以具有cfDNA片段的长度平均相差约13bp的靶向区域cfDNA片段大小分布。在一些方面,大小分布可以是全基因组大小分布。

[0057] cfDNA片段化图谱可以包括小cfDNA片段与大cfDNA片段的比率以及片段比率与参考片段比率的相关性。如本文所用,关于小cfDNA片段与大cfDNA片段的比率,小cfDNA片段的长度可以为约100bp至约150bp。如本文所用,关于小cfDNA片段与大cfDNA片段的比率,大cfDNA片段的长度可以为约151bp至220bp。如本文所述,患有癌症的受试者可以具有比健康受试者低(例如,低2倍、低3倍、低4倍、低5倍、低6倍、低7倍、低8倍、低9倍、低10倍或更多)的片段比率的相关性(例如,cfDNA片段比率与参考DNA片段比率如来自一个或多个健康受试者的DNA片段比率的相关性)。健康受试者(例如,未患有癌症的受试者)可以具有约1(例如,约0.96)的片段比率的相关性(例如,cfDNA片段比率与参考DNA片段比率如来自一个或多个健康受试者的DNA片段比率的相关性)。在一些方面,患有癌症的受试者可以具有比健康受试者的片段比率的相关性(例如,cfDNA片段比率与参考DNA片段比率如来自一个或多个健康受试者的DNA片段比率的相关性)平均低约0.19至约0.30(例如,约0.25)的片段比率的相关性(例如,cfDNA片段比率与参考DNA片段比率如来自一个或多个健康受试者的DNA片段比率的相关性)。

[0058] 目前描述的方法和系统可用于检测、预测、治疗和/或监测受试者体内癌症状态。可以如本文所述评估、监测和/或治疗任何适当受试者,如哺乳动物。可以如本文所述评估、监测和/或治疗的一些哺乳动物的实例包括但不限于人、灵长类,如猴子、狗、猫、马、牛、猪、羊、小鼠和大鼠。例如,可以使用本文所述的方法评估患有或疑似患有癌症的人,并且任选地可以用如本文所述的一种或多种癌症治疗对其进行治疗。

[0059] 可以使用本文所述的方法和系统评估和/或治疗患有或疑似患有任何适当类型的癌症的受试者(例如,通过向受试者施用一种或多种癌症治疗)。癌症可以是任何阶段的癌症。在一些方面,癌症可以是早期癌症。在一些方面,癌症可以是无症状的癌症。在一些方面,癌症可以是残留疾病和/或复发(例如,在手术切除后和/或在癌症疗法后)。癌症可以是任何类型的癌症。可以如本文所述评估、监测和/或治疗的类型的癌症的实例包括但不限于结肠直肠癌、肺癌、乳腺癌、胃癌、胰腺癌、胆管癌、头颈癌、肾癌、骨癌、脑癌、造血细胞癌和卵巢癌。

[0060] 当治疗患有或疑似患有如本文所述的癌症的受试者时,可以向受试者施用一种或多种癌症治疗。癌症治疗可以是任何适当癌症治疗。可以以任何适当频率(例如,在数天至数周的时间段内一次或多次)向受试者施用本文所述的一种或多种癌症治疗。癌症治疗的实例包括但不限于辅助化疗、新辅助化疗、放疗、激素疗法、细胞毒性疗法、免疫疗法、过继

性T细胞疗法(例如,嵌合抗原受体和/或具有野生型或经修饰T细胞受体的T细胞)、靶向疗法,如施用激酶抑制剂(例如,靶向特定遗传病变如易位或突变的激酶抑制剂)(例如,激酶抑制剂、抗体、双特异性抗体)、信号转导抑制剂、双特异性抗体或抗体片段(例如,BiTE)、单克隆抗体、免疫检查点抑制剂、手术(例如,手术切除),或上述各项的任何组合。在一些方面,癌症治疗可以降低癌症的严重性、减轻癌症的症状,和/或减少受试者体内存在的癌细胞的数目。

[0061] 在一些方面,癌症治疗可以是化疗剂。化疗剂的非限制性实例包括:安吡啶、阿扎胞苷、硫唑嘌呤、贝伐单抗(或其抗原结合片段)、博来霉素、白消安、卡铂、卡培他滨、苯丁酸氮芥、顺铂、环磷酰胺、阿糖胞苷、达卡巴嗪、柔红霉素、多西他赛、去氧氟尿苷、阿霉素、表柔比星、盐酸埃罗替尼、依托泊苷、氟达拉滨、氟尿苷、氟达拉滨、氟尿嘧啶、吉西他滨、羟基脲、伊达比星、异环磷酰胺、伊立替康、洛莫司汀、氮芥、美法仑、巯嘌呤、甲氨蝶呤、丝裂霉素、米托蒽醌、奥沙利铂、紫杉醇、培美曲塞、丙卡巴肼、全反式维甲酸、链佐星、他氟平、替莫唑胺、替尼泊苷、硫鸟嘌呤、拓扑替康、乌拉莫司汀、戊柔比星、长春碱、长春新碱、长春地辛、长春瑞滨,及其组合。抗癌疗法的附加实例是本领域已知的;参见例如来自美国临床肿瘤学会(ASCO)、欧洲医学肿瘤学会(ESMO)或国家综合癌症网络(NCCN)的疗法指南。

[0062] 当监测患有或疑似患有如本文所述的癌症的受试者时,监测可以在癌症治疗的历程之前、期间和/或之后进行。本文提供的监测方法可以用于确定一种或多种癌症治疗的功效和/或选择进行增强监测的受试者。在一些方面,监测可以包括分析从受试者获得的样本中的cfDNA片段大小密度的曲线的形状。例如,可以在向患有或疑似患有癌症的受试者施用一种或多种癌症治疗之前获得cfDNA片段大小密度的曲线的形状,可以向受试者施用一种或多种癌症治疗,并且可以在癌症治疗的历程期间获得cfDNA片段大小密度的一条或多条曲线。在一些方面,cfDNA片段大小密度的曲线的形状可以在癌症治疗(例如,本文所述的癌症治疗中的任一种)的历程期间改变。例如,指示受试者患有癌症的cfDNA片段大小密度的曲线的形状可以改变为指示受试者未患有癌症的cfDNA片段大小密度的曲线的形状。

[0063] 在一些方面,监测可以包括能够监测一种或多种癌症治疗(例如,一种或多种癌症治疗的功效)的常规技术。在一些方面,与未被选择进行增强监测的受试者相比,可以以增加的频率向被选择进行增强监测的受试者施用诊断测试(例如,本文公开的诊断测试中的任一种)。例如,可以以每天两次、每天一次、每周两次、每周一次、每月两次、每月一次、每季度一次、每半年一次、每年一次或其中的任何频率向被选择进行增强监测的受试者施用诊断测试。

[0064] 在各个方面,DNA存在于取自受试者的生物样本中并在本发明的方法中使用。生物样本实际上可以是包括DNA的任何类型的生物样本。生物样本通常是流体,如全血或其具有循环cfDNA的一部分。在实施例中,样本包括来自肿瘤或液体活检的DNA,该肿瘤或液体活检为如但不限于羊水、房水、玻璃体液、血液、全血、分级分离的血液、血浆、血清、母乳、脑脊液(CSF)、耳垢(耳垢)、乳糜、chime、内淋巴、外淋巴、粪便、呼吸、胃酸、胃液、淋巴、粘液(包括鼻引流液和痰)、心包液、腹膜液、胸膜液、脓液、发炎性分泌物、唾液、呼出气冷凝物、皮脂、精液、痰液、汗液、滑液、眼泪、呕吐物、前列腺液、乳头抽吸液、泪液、汗水、颊拭子、细胞裂解物、胃肠液、活检组织和尿液或其它生物流体。在一个方面,样本包括来自循环肿瘤细胞的DNA。

[0065] 如上所公开,生物样本可以是血液样本。该血液样本可以使用本领域已知的方法获得,如手指穿刺或静脉切开术。合适地,血液样本为大约0.1至20ml,或替代地大约1至15ml,其中血液的体积为大约10ml。也可以使用较少的量,以及血液中的循环游离DNA。微量采样和通过针吸活组织检查、导管、排泄或产生含有DNA的体液的采样也是潜在的生物样本来源。

[0066] 本公开的方法和系统利用核酸序列信息,并且因此可以包括任何方法或测序装置以用于执行核酸测序,包括核酸扩增、聚合酶链反应(PCR)、纳米孔测序、454测序、插入标签测序。在一些方面,本公开的方法或系统利用如由依诺米那公司(Illumina, Inc)提供的系统(包括但不限于HiSeq™ X10、HiSeq™ 1000、HiSeq™ 2000、HiSeq™ 2500、Genome Analyzers™、MiSeq™、NextSeq、NovaSeq 6000系统)、应用生物系统生命技术公司(Applied Biosystems Life Technologies)提供的系统(SOLiD™ System、Ion PGM™ Sequencer、ion Proton™ Sequencer)或Genapsys或BGI MGI和其它系统。核酸分析还可以通过由牛津纳米孔技术公司(Oxford Nanopore Technologies)提供的系统(GridiON™、MiniON™)或由太平洋生物科学公司(Pacific Biosciences)提供的系统(Pacbio™ RS II或Sequel I或II)执行。

[0067] 本发明包括用于执行所公开的方法的步骤的系统,并且部分地根据功能组件和各种处理步骤来描述。此类功能组件和处理步骤可以通过被配置成执行指定功能并实现各种结果的任何数目的组件、操作和技术来实现。例如,本发明可以采用各种生物样本、生物标记物、元素、材料、计算机、数据源、存储系统和介质、信息收集技术和过程、数据处理标准、统计分析、回归分析等,它们可以执行各种功能。

[0068] 因此,本发明还提供了用于检测、分析和/或评估癌症的系统。在各个方面,该系统包括:(a) 测序仪,该测序仪被配置成生成样本的低覆盖度全基因组测序数据集;和(b) 具有执行本发明的方法的功能的计算机系统和/或处理器。

[0069] 在各个方面,该计算机系统具有非暂时性计算机可读介质,该非暂时性计算机可读介质具有指令以执行以下各项中的一者或多者:(i) 处理低覆盖度全基因组测序数据集,以产生样本的片段大小密度的曲线;(ii) 将样本的片段大小密度的曲线拟合到至少两组不同的已建立的统计参数,以产生至少两个建议曲线拟合;(iii) 显示该至少两个建议曲线拟合,使得用户能够选择所述至少两个建议曲线拟合中的至少一个来用于进一步处理;以及(iv) 显示对应于(iii)的所选建议曲线拟合的建议曲线拟合线以及参考曲线拟合线,使得能够在所选建议曲线拟合线与参考曲线拟合线之间进行比较。

[0070] 在一些方面,计算机系统具有非暂时性计算机可读介质,该非暂时性计算机可读介质具有指令,以针对基因组的一个或多个染色体臂确定长度在约260bp与440bp之间的片段的数目,并计算每个臂的二核小体片段的比例。在一些方面,计算机系统具有非暂时性计算机可读介质,该非暂时性计算机可读介质具有指令,以针对基因组的一个或多个染色体臂确定长度在约260bp与440bp之间的片段的数目,计算每个臂的二核小体片段的比例,并计算每个染色体臂中的二核小体片段的数目。在一些方面,计算机系统具有非暂时性计算机可读介质,该非暂时性计算机可读介质具有指令,以针对基因组的一个或多个染色体臂确定长度在约260bp与440bp之间的片段的数目,计算每个臂的二核小体片段的比例,计算每个染色体臂中的二核小体片段的数目,并生成片段大小密度的曲线。

[0071] 在一些方面,计算机系统还包括一个或多个附加模块。例如,该系统可以包括以下各项中的一者或多者:提取单元,该提取单元可操作以选择合适的分量用于曲线拟合分析;曲线拟合单元,该曲线拟合单元可操作以通过使用用户定义的等式来执行正态分布拟合的有限混合或执行多项式回归拟合;曲线拟合优度分析单元,该曲线拟合优度分析单元可操作以提供由曲线拟合单元生成的拟合质量的指示符;曲线拟合参数表征单元,该曲线拟合参数表征单元可操作以将曲线拟合参数与参考值进行比较来分类;表征数据库,该表征数据库用于存储拟合系数及其表征,或其任何组合。

[0072] 在一些方面,计算机系统还包括视觉显示装置。该视觉显示装置可以可操作以显示曲线拟合线、参考曲线拟合线和/或两者的比较。

[0073] 根据本发明的各个方面的用于检测和分析的方法可以以任何合适的方式实施,例如使用在计算机系统上操作的计算机程序。如本文所讨论的,根据本发明的各个方面的示范性系统可以结合计算机系统来实施,该计算机系统为例如包含处理器和随机存取存储器的常规计算机系统,如远程可访问的应用服务器、网络服务器、个人计算机或工作站。该计算机系统还合适地包括附加存储装置或信息存储系统,如大容量存储系统和用户界面,例如常规监视器、键盘和跟踪装置。然而,计算机系统可以包括任何合适的计算机系统和相关联的设备,并且可以以任何合适的方式来配置。在一个实施例中,计算机系统包含独立系统。在另一实施例中,计算机系统是包括服务器和数据库的计算机的网络的一部分。

[0074] 接收、处理和分析信息所需的软件可以在单个装置中实施或在多个装置中实施。软件可以经由网络访问,使得信息的存储和处理相对于用户远程进行。根据本发明的各个方面的系统及其各种元件提供便于检测和/或分析的功能和操作,如数据收集、处理、分析、报告和/或诊断。例如,在本方面,计算机系统执行计算机程序,其可以接收、存储、搜索、分析和报告与人类基因组或其区域相关的信息。计算机程序可以包含执行各种功能或操作的多个模块,如用于处理原始数据并生成补充数据的处理模块以及用于分析原始数据和补充数据以生成疾病状态模型和/或诊断信息的定量评估的分析模块。

[0075] 由系统执行的程序可以包含任何合适的过程以便于分析和/或癌症诊断。在一个实施例中,系统被配置成建立疾病状态模型和/或确定患者的疾病状态。确定或鉴定疾病状态可以包括生成关于患者相对于疾病的状况的任何有用信息,如执行诊断、提供有助于诊断的信息、评估疾病的阶段或进程、鉴定可以指示对疾病的易感性的状况、鉴定是否会推荐进一步的测试、预测和/或评估一个或多个治疗程序的功效,或以其它方式评估患者的疾病状态、疾病的可能性或其它健康方面。

[0076] 提供以下实例以进一步绘示本发明的优点和特征,但其不旨在限制本发明的范围。虽然该实例是可以使用的实例中的典型实例,但是也可以替代地使用本领域技术人员已知的其它程序、方法或技术。

[0077] 实例1

[0078] 检测癌症

[0079] 在该实例中,本公开的方法用于检测癌症。以下提供了对用于癌症检测的方法和过程的深入讨论。使用从供体(由207个患有癌症的个体和214个无癌症个体组成)收集的421个血浆样本的全基因组数据,证明了片段大小密度的形状是如何与核小体重新定位相关的方式计算的。附加地,示出了如何将结果单独和与其它已开发的方法结合用于预测

癌症状态。

[0080] 先前在本文中描述了用于总结片段大小密度的形状的方法，并且更详细地描述如下。

[0081] 在第一种方法(多项式回归方法)中，使用对给定长度的片段的计数的多项式回归拟合的系数，分别通过样本拟合。大小小于105bp和大于170bp的片段由于片段的低计数和不容易被多项式回归捕捉的不显著模式而被排除。详细地，大小为N的片段的计数通过样本标准化为具有平均值0和方差1。该回归的输入是片段大小的一至十二次正交多项式。因此，对于每个回归模型，输入是大小为 66×12 的矩阵，并且输出是缩放计数。从每个多项式回归模型中提取的是系数。多项式回归模型明确地捕捉片段大小密度的形状，并且隐含地对核小体滑动和DNA环的贡献建模。

[0082] 在第二种方法(贝叶斯有限混合模型)中，将截断正态分布的有限混合拟合到片段大小的计数。小于105bp和大于220bp的片段大小由于该范围之外的DNA片段的低计数而被排除。在分布的上升侧可以看到对应于各模式的七个分量。三个分量对应于在下降侧看到的分量。一个分量表征总体基料并且在98.33%的样本中占混合的50%以上。最后的分量捕捉在片段大小密度的下降侧看到的偏斜。每个分量具有220bp的截断，以防止用较大的平均值低估分量的方差。这些分量中的每一者的特征在于平均值、方差和对总体混合的贡献。将不可交换的、适度信息量的先验置于混合平均值和方差上，并将弱信息量的Dirichlet先验置于混合比例上。使用禁止回转采样器(Hoffman, Matthew D., 和Andrew Gelman. 2014. “禁止回转采样器:在哈密尔顿蒙特卡罗中自适应设置路径长度(The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.)”《机器学习研究杂志(J.Mach.Learn.Res.)》15(1):1593-1623.)用2,000个样本(1,000个预热样本)对每个样本拟合该模型。通过检查所有参数的多变量潜在比例缩减因子(Gelman, Andrew, 和Donald B. Rubin. 1992. “使用多个序列的迭代模拟的推断(Inference from Iterative Simulation Using Multiple Sequences.)”《统计科学(Statistical Science)》7(4):457-72. <https://doi.org/10.1214/ss/1177011136>.) 小于或等于1.1来评估收敛。还可以使用变分推断来拟合模型。

[0083] 结果

[0084] 如图2所示，虽然多项式模型捕捉了片段大小密度的形状的一些细微差别，但混合模型提供了对片段大小密度的形状的几乎准确的拟合。重要的是，图3示出了混合模型的可解释性，并且使得与底层DNA-核小体相互作用的连接更加明确。

[0085] 据信，片段大小密度的形状可以反映DNA-核小体相互作用动力学。Lequieu等人(Lequieu, Joshua, David C. Schwartz, 和Juan J. de Pablo. 2017. “序列依赖性核小体滑动的计算机模拟证据(In Silico Evidence for Sequence-Dependent Nucleosome Sliding.)”《美国国家科学院院刊(Proceedings of the National Academy of Sciences)》114(44):E9197-E9205. <https://doi.org/10.1073/pnas.1705685114>.) 描述了核小体在基因组内重新定位的各种方法。特别地，他们开发了分子模型来描述DNA环，这些DNA环被“引入核小体的一侧，并且然后以蠕虫样方式沿着组蛋白核心移动”。Lequieu等人发现这些DNA环的位置是“对DNA序列不敏感的”，并且基于他们的分子模型，在图6中按在组蛋白上的位置证明了DNA环的密度图。该密度图反映了我们在图2中所示的密度的许多形

状。值得注意的是,两种密度表现出类似的小模式图式。这些环的引入可以产生内切核酸酶酶切割接头组蛋白内部的核小体DNA的机会。例如,在核小体DNA末端处的DNA环可以允许产生稍短于167bp的片段。环沿着组蛋白“寸动”越远,所得切割片段越短。附加地,通过DNA环重新定位的核小体可以在接头组蛋白处被切割,并且可以表现为长于167bp的cfDNA片段。

[0086] 基于Lequieu等人的描述,混合模型中的许多分量可以由DNA环位置和稳定性驱动。通过这种理解,片段大小的周期性和计数将不会捕捉DNA环的这些特征。在给定样本中,DNA环在组蛋白上的特定位置处可能更频繁。这将在混合比例中反映。附加地,环的稳定性可以由分量的方差反映。

[0087] 为了评价混合模型参数预测供体的癌症状态的能力,以与Cristiano等人描述的方法相同的方法将系数用作机器学习模型中的特征(Cristiano,Stephen,Alessandro Leal,Jillian Phallen,Jacob Fiksel,Vilmos Adleff,Daniel C.Bruhm,Sarah Østrup Jensen等人2019.“癌症患者的全基因组游离DNA片段化(Genome-Wide Cell-Free DNA Fragmentation in Patients with Cancer.)”《自然(Nature)》570(7761):385-89.<https://doi.org/10.1038/s41586-019-1272-6>.)。简言之,使用无超参数调节的梯度提升机(Friedman,Jerome H.2000.“贪婪函数近似:梯度提升机(Greedy Function Approximation:A Gradient Boosting Machine.)”《统计学年鉴(Annals of Statistics)》29:1189-1232.)的10倍、10次重复交叉验证。使用了以下特征。

[0088] 1)混合模型系数:用12个平均值、12个方差和12个混合比例充分描述混合模型。对于建模,我们排除第12个比例,因为它是其它11个比例的线性组合。35个特征未转化。

[0089] 2)短/长比率:类似于Cristiano等人,我们计算了整个基因组中504个相互排斥的5MB仓中GC含量校正的短(100-150bp)片段和长(151-220bp)片段的数目,并以样本为中心将短片段的计数除以长片段的计数。

[0090] 3)短/总体覆盖度:在比率中,我们使用短(100-150bp)片段和总体(100-220bp)片段的覆盖度作为特征。这些覆盖度按样本和类型(短,总体)标准化为具有平均值0和标准偏差1。

[0091] 这三个特征集分别用作GBM的输入以及混合模型系数与(2)和(3)的组合用作GBM的输入。结果作为ROC曲线下面积(AUC)和95%和98%特异性下的灵敏度报告于表1中。

[0092] 表1:癌症检测的交叉验证结果

特征集	AUC	灵敏度(95%特异性)	灵敏度(98%特异性)
短/长比率	0.8902	0.5898	0.5146
短/总体覆盖度	0.9307	0.7670	0.6650
混合	0.9466	0.7330	0.6990
混合, 比率	0.9547	0.7621	0.7039
混合, 覆盖度	0.9737	0.8301	0.7573

[0094] 如表1所绘示,混合模型系数的AUC指示癌症状态的强预测因子。更能指示对早期癌症检测的实用性的是高特异性下的灵敏度。虽然混合模型系数具有比覆盖度参数稍低的灵敏度,但两者组合时示出了改进的AUC和灵敏度。组合特征的结果证明了片段大小密度的这种新总结如何能够补充其它基因组特征以确定癌症状态。

[0095] 在图4中,绘制了这5个模型的ROC曲线。这些曲线证明了混合系数作为特征以及混

合和覆盖度特征的组合在甚至更高的特异性下的高灵敏度。参考线指示95%和98%的特异性。

[0096] 实例2

[0097] 使用二核小体片段大小密度检测癌症

[0098] 在该实例中,本公开的方法用于检测癌症。如本文所讨论的,已经表明ctDNA片段平均比来自非肿瘤细胞的其它cfDNA更短。先前的工作已经探索了将片段分成不同大小的组(例如,短和长,或相互排斥的大小组)并使用这些仓的计数来量化ctDNA和/或将各个样本分类为存在/不存在肿瘤。

[0099] 如本文所证明的,片段大小密度可以被建模为分布的混合,其参数可以预测癌症状态。这些方法均集中于紧密对应于单个核小体的大小的片段大小(大小小于260bp)。这些通常包括用147bp的DNA包裹的组蛋白八聚体,连同H1组蛋白和接头DNA(20bp),从而得到观察到的167bp大小的主峰。

[0100] 该实例中描述的研究集中于对来自低覆盖度全基因组测序(1-2x覆盖率)的小于260bp的cfDNA片段的分析,可能包括两个核小体,它们导致中值为334bp的峰。

[0101] 方法

[0102] 对于每个样本,片段的数目由每个染色体臂(不包括近端着丝粒染色体臂)的260bp与440bp之间的碱基对确定。按宽度计算每个臂的二核小体片段的比例。在给定臂中所有181个片段大小的这些比例总和为1。该组181个比例被定义为二核小体片段大小密度。每个样本具有39组二核小体片段大小密度,这是通过非近端着丝粒染色体臂计算的。

[0103] 在一组48个购买的专门无癌症样本中,对每个臂计算这些片段大小密度,并在所有样本中合计以表示39个参考片段大小密度。

[0104] 对于数据集中的每个样本,通过计算经验样本/臂与参考/臂之间的Kullback-Leibler散度来确定给定样本/臂与参考的类似性。附加地,计算每个染色体臂中的二核小体片段的数目。

[0105] 为了评价这些二核小体参数预测供体的癌症状态的能力,本发明人使用与Cristiano等人描述的方法相同的方法将参数用作机器学习模型中的特征(Cristiano等人,2019.“癌症患者的全基因组游离DNA片段化”《自然》570(7761):385-389)。简言之,使用无超参数调节的梯度提升机(Friedman,Jerome H.2000.“贪婪函数近似:梯度提升机”《统计学年鉴》29:1189-1232)的10倍、10次重复交叉验证。使用了以下特征集。

[0106] 1) 二核小体:本发明人计算了每个样本/臂的尺寸比例与参考之间的Kullback-Leibler散度。还确定了每个臂的二核小体读段(260bp-440bp)的数目。

[0107] 2) 短/总体覆盖度:类似于Cristiano等人(2019),本发明人计算了整个基因组中504个相互排斥的5MB仓中GC含量校正的短(100-150bp)片段和总体(100-220bp)片段的数目。这些覆盖度按样本和类型(短,总体)标准化为具有平均值0和标准偏差1。

[0108] 这两个特征集分别用作GBM的输入以及两个特征的组合用作GBM的输入。结果作为ROC曲线下面积(AUC)和95%和98%特异性下的灵敏度报告于下表2中。

[0109] 结果

[0110] 在图5中,按癌症和无癌症个体绘制了通过DNA读段的量归一化的260bp与440bp之间的平均片段大小密度。

[0111] 如图5所示,很明显,平均而言,无癌症个体在二核小体范围内具有较长的cfDNA片段(平均大小为334.75bp),而癌症患者具有较短的cfDNA片段(平均大小为329.6bp)。一些研究集中于单核小体cfDNA,并且已经表明,平均而言,在患有癌症的个体中小于260bp的cfDNA片段比未患癌症的个体短。然而,这些工作未评价癌症患者或健康个体中二核小体cfDNA(大小小于260bp)的大小分布。

[0112] 本发明人预期334bp处的峰表示二核小体,即具有相关联H1和接头DNA的两个邻近核小体,它们各自包含167bp的DNA。假定在核小体之间通常存在附加接头DNA,则需要重新定位一个或两个核小体以支持该假设。一项研究使用原子力显微镜(ATM)证明核小体可以通过RSC(重塑染色质结构)复合物以此类方式重新定位,如图6所示。在所鉴定的五种状态中,最稳定的状态中的一个(#4)表示在裸DNA末端直接邻近的核小体,并且切割可以发生在该构型的5'和3'末端处。重塑与DNA长度无关。

[0113] 基于使用ATM观察到的核小体构型,在较低频率下(260bp与334bp之间)的数据中观察到的较小的ctDNA片段可以表示在核小体的一侧上的切割和在不同位置(直至邻近核小体)处的居间接头DNA的进一步内切核酸酶消化(图7;状态2或5)。较大的片段(334bp与440bp之间)可以由作用于两个邻近单核小体之外的接头DNA的内切核酸酶产生(图7;状态1、3或4)。

[0114] 二核小体形成可以基于它们与RSC的关联而链接到启动子区域,因为该复合物富含高度表达的基因。使用酵母模型的数据也证明,除了一个核小体之外的所有核小体在启动子激活后均被去除,这符合以下模型:核小体通过RSC滑动,DNA解链,并且组蛋白八聚体沿着路径排出,在过程结束时仅留下单个核小体结合到RSC。在RSC水平枯竭的酵母模型中使用ChIP-Seq的附加数据示出了转录起始位点(TSS)上游和下游的组蛋白增加。总之,这些数据涉及在NDR(核小体枯竭区域)中去除核小体时的RSC介导的重新定位以及转录的调节。

[0115] 在图8中,每个样本的染色体臂1p中的二核小体片段大小密度以蓝色绘制,每个样本类型的中值片段大小密度以黑色绘制,并且参考以橙色绘制。从图8可以明显看出,与1至3期的样本相比,来自4期癌症的样本与参考样本更不类似。特别地,很明显,相对于1至3期癌症,在4期中二核小体片段大小密度变得更偏斜,并且总体片段变得更短,并且相对于无癌症,在1至3期癌症中也是如此。

[0116] 表2绘示了机器学习模型使用从游离DNA提取的该新基因组特征来区分癌症个体和无癌症个体的能力。使用二核小体系数的机器学习模型的AUC指示癌症状态的强预测因子。更能指示对早期癌症检测的实用性的是高特异性下的灵敏度。组合这两个特征示出了与最佳特征一样好或比最佳特征更好的AUC和灵敏度。组合特征的结果证明了片段大小密度的这种新总结如何能够补充其它基因组特征以确定癌症状态。

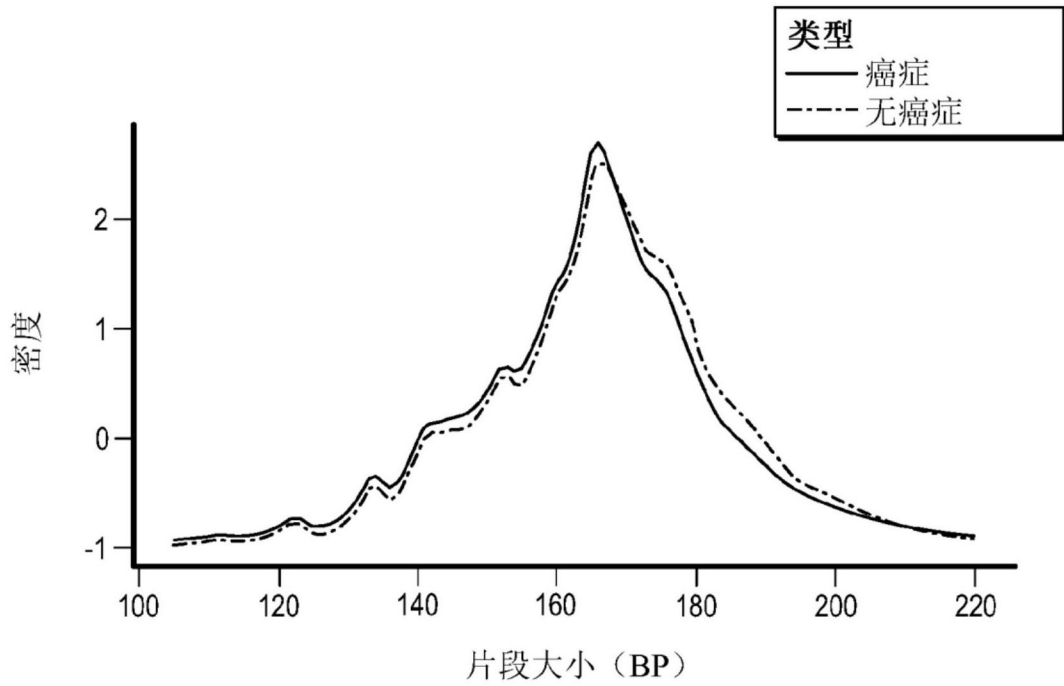
[0117] 表2:使用每个臂的与二核小体参考的散度进行的癌症检测的交叉验证结果

[0118] 特征集	AUC	灵敏度(95%特异性)	灵敏度(98%特异性)
短/总体覆盖度	0.7963	0.5342	0.5205
[0119] 二核小体	0.8493	0.6164	0.4931
二核小体,覆盖度	0.8847	0.6164	0.5890

[0120] 图9示出了这三个模型的ROC曲线。这些曲线证明了二核小体系数作为特征以及二核小体和覆盖度特征的组合在甚至更高的特异性下的高灵敏度。参考线指示90%、95%和

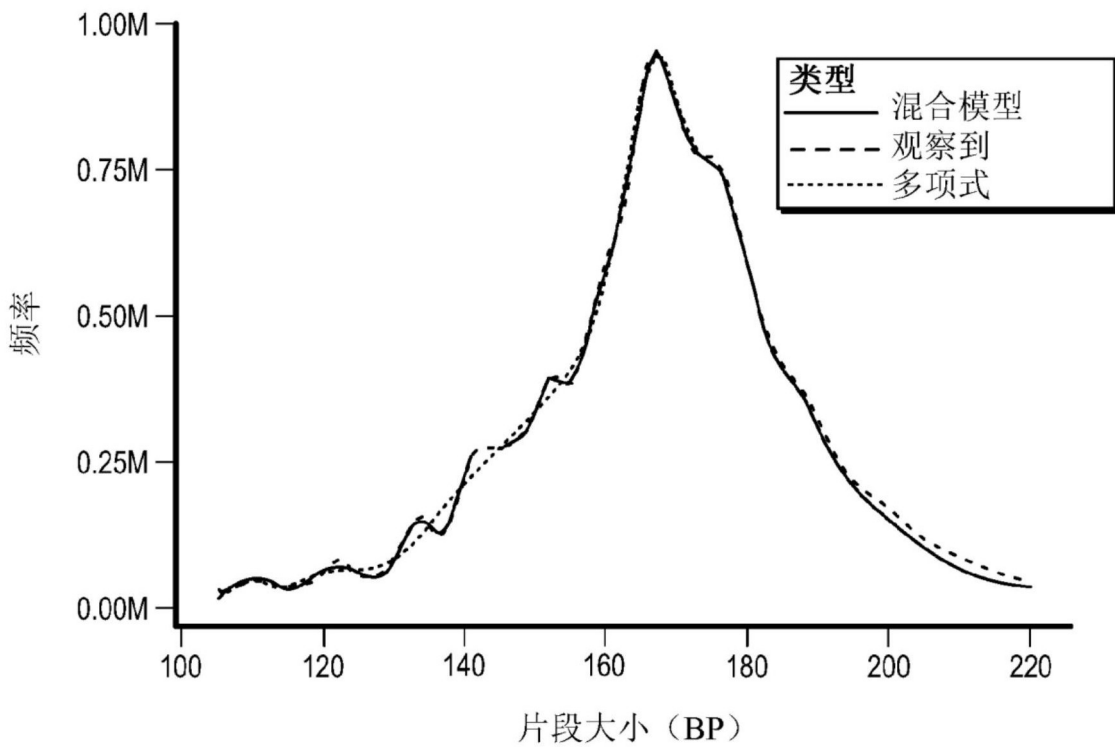
98%的特异性。

[0121] 尽管已经参考以上实例描述了本发明,但应当理解,修改和变化均涵盖在本发明的精神和范围之内。因此,本发明仅由以下权利要求限制。



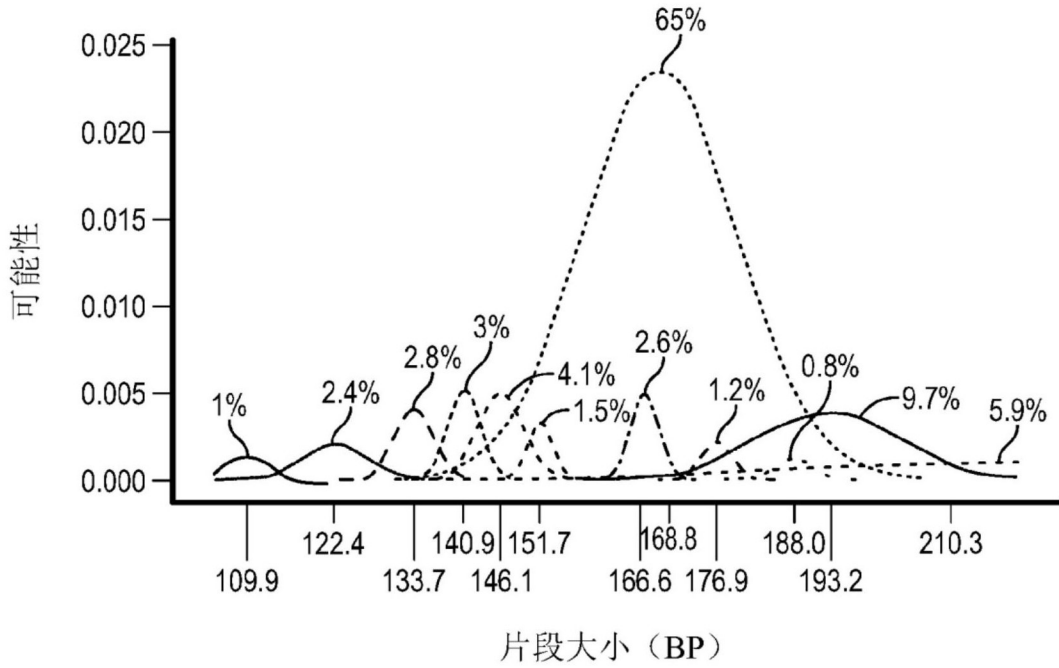
平均片段大小密度的比较。

图1



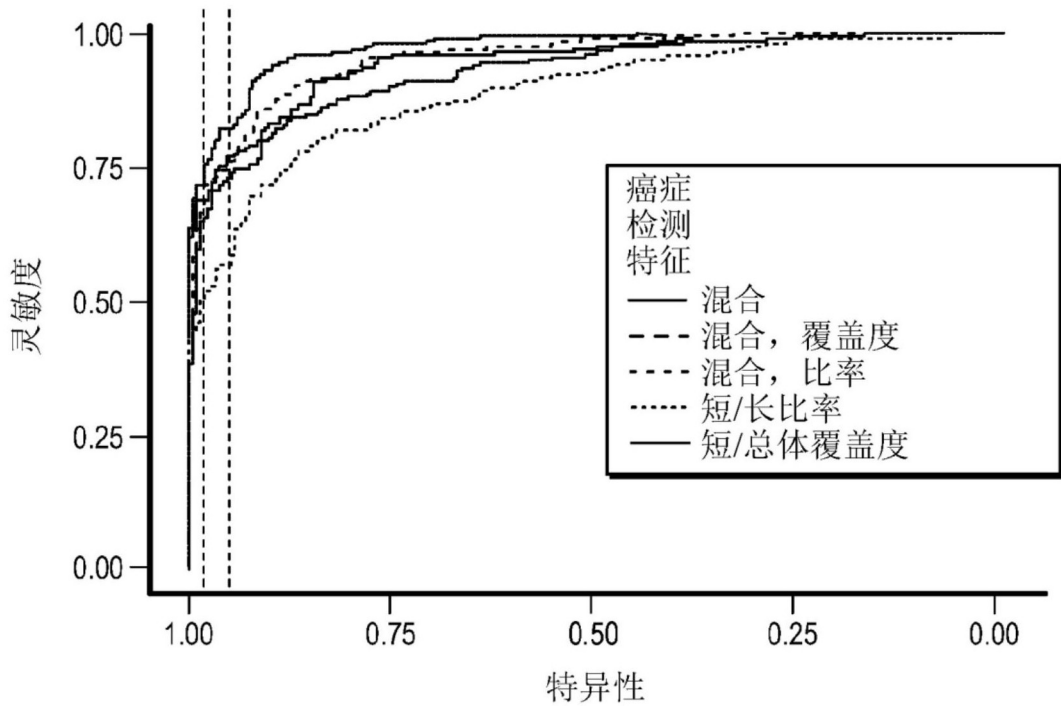
使用两种模型表示游离序列的大小分布

图2



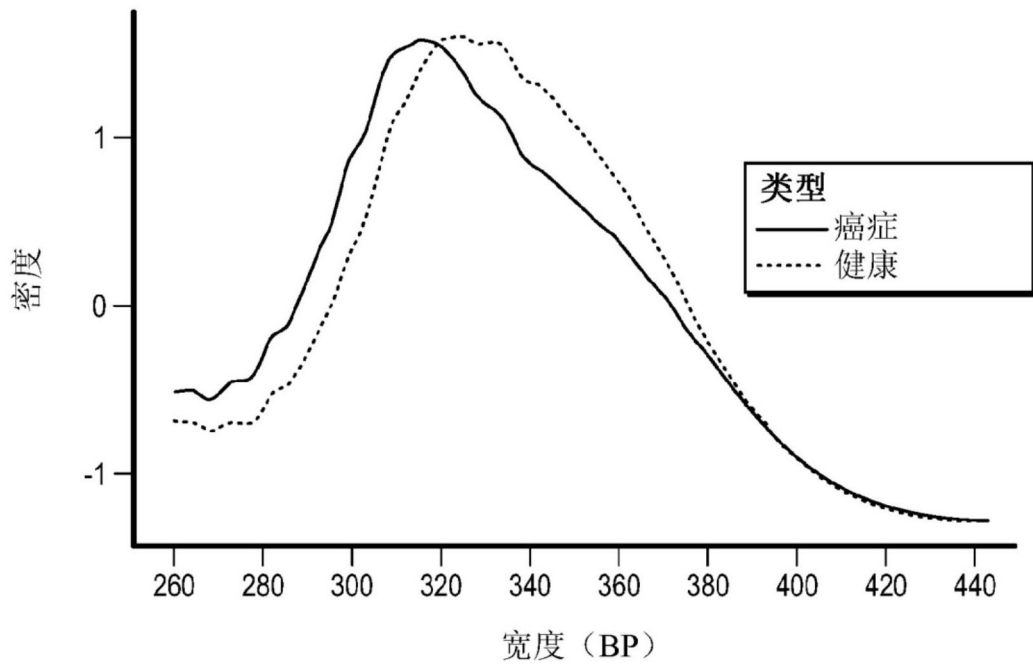
来自一个样本的混合模型分量

图3



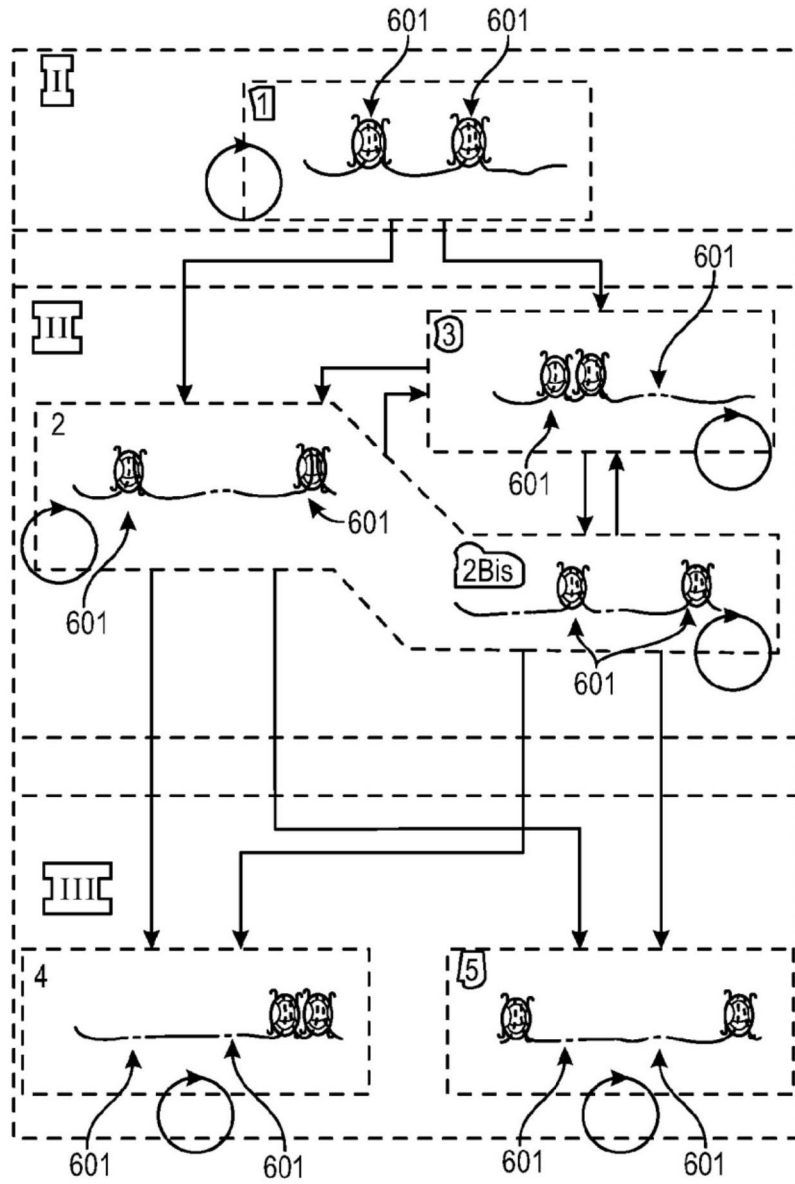
平均片段大小密度的比较

图4



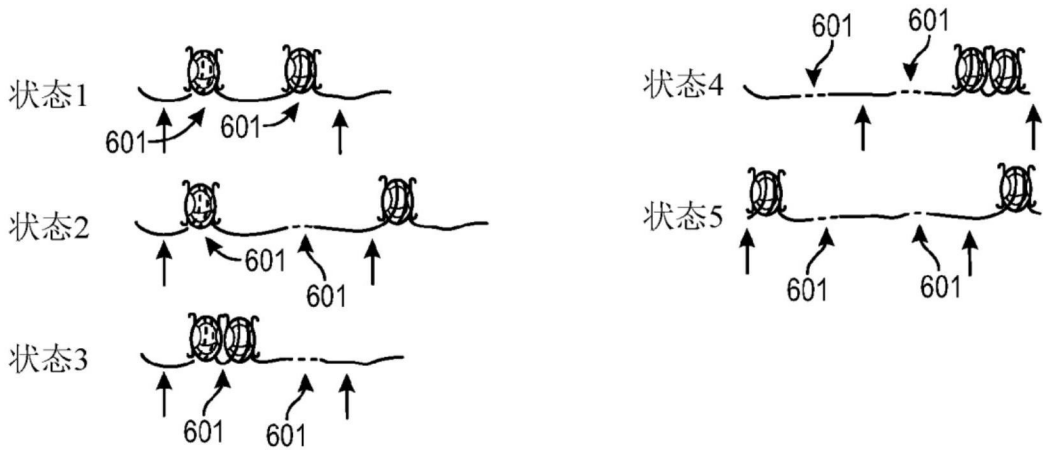
健康个体和癌症患者中的大小大于260 bp的cfDNA中的平均片段大小密度的比较

图5



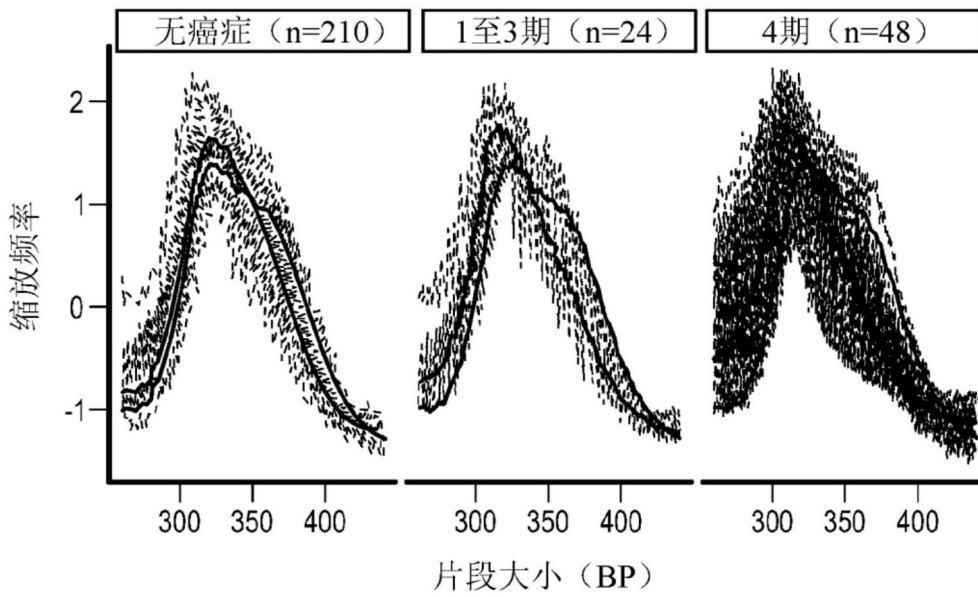
示出由二核小体的RSC滑动产生的各种状态的运动方案。

图6



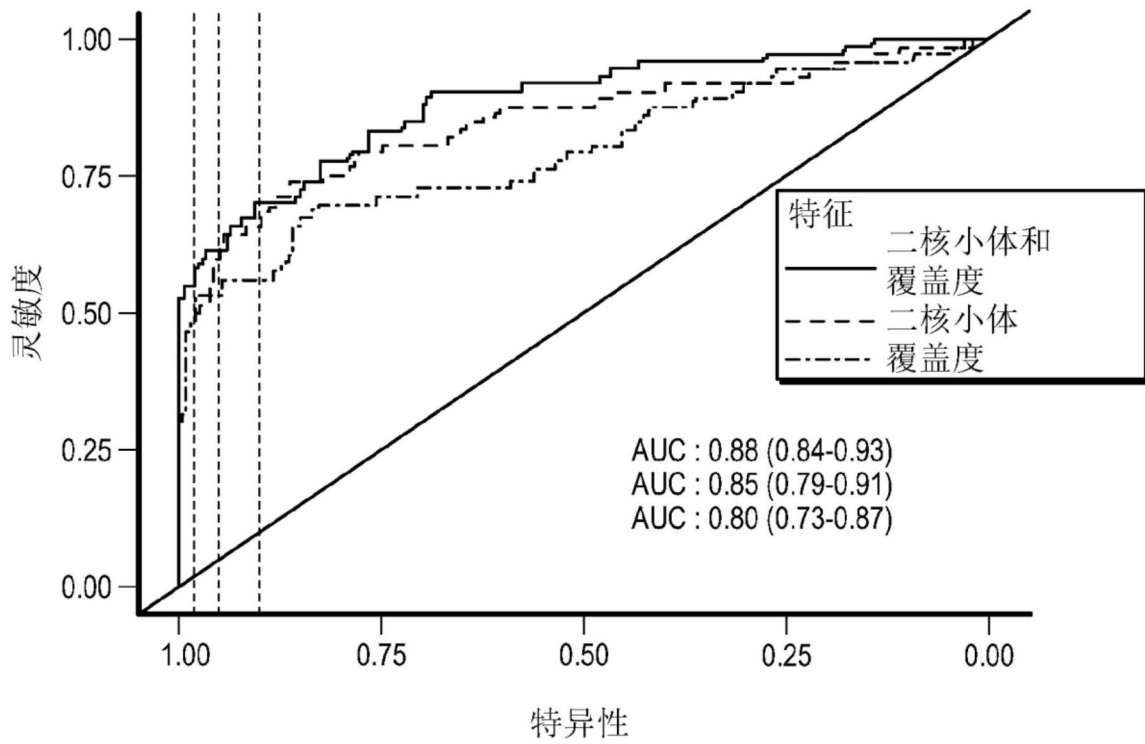
导致变化片段大小的潜在内切核酸酶切割位点的实例。每种状态仅示出了两个位点，但切割可以发生在接头DNA中的多个不同位置处。

图7



晚期癌症变得与非癌症参考更不类似的图示，以上在染色体臂1p中证明。

图8



使用不同特征的模型的ROC曲线。

图9