



(12) 发明专利

(10) 授权公告号 CN 112955570 B

(45) 授权公告日 2025. 01. 21

(21) 申请号 201980069237.6

(22) 申请日 2019.09.13

(65) 同一申请的已公布的文献号  
申请公布号 CN 112955570 A

(43) 申请公布日 2021.06.11

(30) 优先权数据  
62/731,718 2018.09.14 US

(85) PCT国际申请进入国家阶段日  
2021.04.20

(86) PCT国际申请的申请数据  
PCT/US2019/051138 2019.09.13

(87) PCT国际申请的公布数据  
W02020/056347 EN 2020.03.19

(73) 专利权人 莱森特生物公司  
地址 美国马萨诸塞州

(72) 发明人 亚历山大·德·永·罗伯逊  
妮可·哈辛达·兰伯特  
哈鲁克·泰兹詹

拉姆·耶拉曼奇里 尼尔·皮泰曼  
罗希斯·卡纳潘·斯里瓦斯

(74) 专利代理机构 北京东方亿思知识产权代理  
有限责任公司 11258  
专利代理师 肖善强

(51) Int.Cl.  
C12Q 1/6869 (2018.01)  
C12Q 1/6827 (2018.01)  
C12Q 1/6886 (2018.01)  
G16B 20/00 (2019.01)  
G16B 30/10 (2019.01)

(56) 对比文件  
US 2014235456 A1, 2014.08.21  
US 2017213008 A1, 2017.07.27  
Arthur Gilly et al..Very low-depth  
whole-genome sequencing in complex trait  
association studies.《Sequence analysis》  
.2018,第35卷(第15期),2555-2561.

审查员 王雨洁

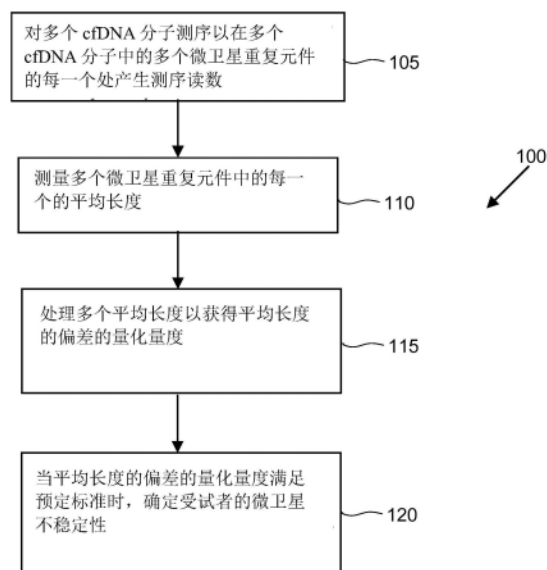
权利要求书6页 说明书21页 附图5页

(54) 发明名称

评估微卫星不稳定性的方法和系统

(57) 摘要

本文公开的发明通常涉及评估受试者中的微卫星不稳定性的方法。一方面,本公开提供一种评估受试者的微卫星不稳定性的计算机实现的方法,所述方法包括:从受试者的血液样品获得多个微卫星重复元件的量化量度;处理所述多个量化量度以获得所述多个量化量度的偏差的统计量度;并且当所述多个量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的存在,或当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的不存在。



1. 一种用于评估受试者的微卫星不稳定性的计算机实现的方法,所述方法包括:

计算机处理器接收从受试者的血液样品中获得的多个来源于肿瘤的无细胞DNA (cfDNA) 分子的多个微卫星重复元件的量化量度,其中所述多个微卫星重复元件的量化量度包括所述多个微卫星重复元件中的每一个的平均长度,长度落入预定大小范围内的多个微卫星重复元件的数量或分数,或所述多个微卫星重复元件中的每一个的平均插入或缺失(indel)长度;

所述计算机处理器处理所述多个量化量度以获得所述多个量化量度的偏差的统计量度,其中所述偏差的统计量度是平均z分数;和

当所述平均z分数的绝对值大于预定数量时,所述计算机处理器检测所述受试者的微卫星不稳定性(MSI)的存在,或当所述平均z分数的绝对值不大于预定数量时,所述计算机处理器检测所述受试者的微卫星不稳定性(MSI)的不存在。

2. 根据权利要求1所述的方法,其中所述受试者被诊断出患有癌症。

3. 根据权利要求1所述的方法,其中所述多个量化量度从在所述多个cfDNA分子中的所述多个微卫星重复元件的每一个处的一组测序读数测量。

4. 根据权利要求3所述的方法,所述方法还包括对所述多个cfDNA分子测序以产生所述一组测序读数。

5. 根据权利要求3或4所述的方法,其中所述测序包括全基因组测序(WGS)。

6. 根据权利要求5所述的方法,其中所述测序在不大于10X的深度下进行。

7. 根据权利要求5所述的方法,其中所述测序在不大于8X的深度下进行。

8. 根据权利要求5所述的方法,其中所述测序在不大于6X的深度下进行。

9. 根据权利要求1所述的方法,其中测量所述多个量化量度包括在所述多个微卫星重复元件中的每一个处进行对所述多个cfDNA分子的结合测量。

10. 根据权利要求1所述的方法,所述方法还包括针对所述多个微卫星重复元件的至少一个子集富集所述多个cfDNA分子。

11. 根据权利要求10所述的方法,其中所述富集包括扩增所述多个cfDNA分子。

12. 根据权利要求11所述的方法,其中所述扩增包括选择性扩增。

13. 根据权利要求11所述的方法,其中所述扩增包括通用扩增。

14. 根据权利要求10所述的方法,其中所述富集包括选择性分离所述多个cfDNA分子的至少一部分用于靶向富集。

15. 根据权利要求14所述的方法,其中所述至少一部分包含单核苷酸。

16. 根据权利要求14所述的方法,其中所述至少一部分包含二核苷酸。

17. 根据权利要求1所述的方法,其中所述偏差的统计量度是相对于参考血液样品的平均z分数。

18. 根据权利要求17所述的方法,其中所述参考血液样品从具有微卫星不稳定性的受试者获得。

19. 根据权利要求17所述的方法,其中所述参考血液样品从不具有微卫星不稳定性的受试者获得。

20. 根据权利要求1所述的方法,其中所述预定数量是3。

21. 根据权利要求1所述的方法,其中所述多个微卫星重复元件包括单核苷酸或二核苷

酸。

22. 根据权利要求21所述的方法,其中所述多个微卫星重复元件包括单核苷酸和二核苷酸。

23. 根据权利要求1所述的方法,其中所述多个微卫星重复元件包括至少一百万个不同的微卫星重复元件。

24. 根据权利要求1所述的方法,其中所述多个微卫星重复元件包括至少500万个不同的微卫星重复元件。

25. 根据权利要求1所述的方法,其中所述多个微卫星重复元件包括至少1000万个不同的微卫星重复元件。

26. 根据权利要求1所述的方法,其中所述多个微卫星重复元件包括至少2000万个不同的微卫星重复元件。

27. 根据权利要求1所述的方法,其中以至少90%的灵敏度检测所述受试者的微卫星不稳定性的存在。

28. 根据权利要求1所述的方法,其中以至少95%的灵敏度检测所述受试者的微卫星不稳定性的存在。

29. 根据权利要求1所述的方法,其中以至少99%的灵敏度检测所述受试者的微卫星不稳定性的存在。

30. 根据权利要求1所述的方法,其中以至少90%的特异性检测所述受试者的微卫星不稳定性的不存在。

31. 根据权利要求1所述的方法,其中以至少95%的特异性检测所述受试者的微卫星不稳定性的不存在。

32. 根据权利要求1所述的方法,其中以至少99%的特异性检测所述受试者的微卫星不稳定性的不存在。

33. 根据权利要求1所述的方法,其中以至少90%的阳性预测值 (PPV) 检测所述受试者的微卫星不稳定性的存在。

34. 根据权利要求1所述的方法,其中以至少0.90的曲线下面积 (AUC) 检测所述受试者的微卫星不稳定性的存在或不存在。

35. 根据权利要求1所述的方法,还包括:当所述平均z分数的绝对值不大于预定数量时,检测所述受试者的微卫星稳定性 (MSS) 的存在。

36. 一种系统,所述系统包括控制器,所述控制器包括或能够访问包括机器可执行指令的非暂时性计算机可读介质,所述机器可执行指令在由一个或多个计算机处理器执行时进行用于评估受试者的微卫星不稳定性的方法,所述方法包括:

接收从受试者的血液样品中获得的多个来源于肿瘤的无细胞DNA (cfDNA) 分子的多个微卫星重复元件的量化量度,其中所述多个微卫星重复元件的量化量度包括所述多个微卫星重复元件中的每一个的平均长度,长度落入预定大小范围内的多个微卫星重复元件的数量或分数,或所述多个微卫星重复元件中的每一个的平均插入或缺失 (indel) 长度;

处理所述多个量化量度以获得所述多个量化量度的偏差的统计量度,其中所述偏差的统计量度是平均z分数;和

当所述平均z分数的绝对值大于预定数量时,检测所述受试者的微卫星不稳定性 (MSI)

的存在,或当所述平均z分数的绝对值不大于预定数量时,检测所述受试者的微卫星不稳定性(MSI)的不存在。

37.根据权利要求36所述的系统,其中所述受试者被诊断出患有癌症。

38.根据权利要求36所述的系统,其中所述多个量化量度从在所述多个cfDNA分子中的所述多个微卫星重复元件的每一个处的一组测序读数测量。

39.根据权利要求38所述的系统,其中所述方法还包括对所述多个cfDNA分子测序以产生所述一组测序读数。

40.根据权利要求38或39所述的系统,其中所述测序包括全基因组测序(WGS)。

41.根据权利要求40所述的系统,其中所述测序在不大于10X的深度下进行。

42.根据权利要求40所述的系统,其中所述测序在不大于8X的深度下进行。

43.根据权利要求40所述的系统,其中所述测序在不大于6X的深度下进行。

44.根据权利要求36所述的系统,其中测量所述多个量化量度包括在所述多个微卫星重复元件中的每一个处进行对所述多个cfDNA分子的结合测量。

45.根据权利要求36所述的系统,其中所述方法还包括基于所检测的所述受试者的微卫星不稳定性,识别针对所述受试者的治疗或待施用到所述受试者的治疗有效量的治疗。

46.根据权利要求45所述的系统,其中所述治疗选自化学疗法、放射疗法和免疫疗法。

47.根据权利要求46所述的系统,其中所述治疗包括免疫疗法。

48.根据权利要求47所述的系统,其中所述免疫疗法包括派姆单抗。

49.根据权利要求36所述的系统,其中所述方法还包括针对所述多个微卫星重复元件的至少一个子集引导所述多个cfDNA分子的富集。

50.根据权利要求49所述的系统,其中所述富集包括扩增所述多个cfDNA分子。

51.根据权利要求50所述的系统,其中所述扩增包括选择性扩增。

52.根据权利要求50所述的系统,其中所述扩增包括通用扩增。

53.根据权利要求49所述的系统,其中所述富集包括选择性分离所述多个cfDNA分子的至少一部分用于靶向富集。

54.根据权利要求53所述的系统,其中所述至少一部分包含单核苷酸。

55.根据权利要求53所述的系统,其中所述至少一部分包含二核苷酸。

56.根据权利要求36所述的系统,其中所述偏差的统计量度是相对于参考血液样品的平均z分数。

57.根据权利要求56所述的系统,其中所述参考血液样品从具有微卫星不稳定性的受试者获得。

58.根据权利要求56所述的系统,其中所述参考血液样品从不具有微卫星不稳定性的受试者获得。

59.根据权利要求36所述的系统,其中所述预定数量是3。

60.根据权利要求36所述的系统,其中所述多个微卫星重复元件包括单核苷酸或二核苷酸。

61.根据权利要求60所述的系统,其中所述多个微卫星重复元件包括单核苷酸和二核苷酸。

62.根据权利要求36所述的系统,其中所述多个微卫星重复元件包括至少一百万个不

同的微卫星重复元件。

63. 根据权利要求36所述的系统,其中所述多个微卫星重复元件包括至少500万个不同的微卫星重复元件。

64. 根据权利要求36所述的系统,其中所述多个微卫星重复元件包括至少1000万个不同的微卫星重复元件。

65. 根据权利要求36所述的系统,其中所述多个微卫星重复元件包括至少2000万个不同的微卫星重复元件。

66. 根据权利要求36所述的系统,其中以至少90%的灵敏度检测所述受试者的微卫星不稳定性的存在。

67. 根据权利要求36所述的系统,其中以至少95%的灵敏度检测所述受试者的微卫星不稳定性的存在。

68. 根据权利要求36所述的系统,其中以至少99%的灵敏度检测所述受试者的微卫星不稳定性的存在。

69. 根据权利要求36所述的系统,其中以至少90%的特异性检测所述受试者的微卫星不稳定性的不存在。

70. 根据权利要求36所述的系统,其中以至少95%的特异性检测所述受试者的微卫星不稳定性的不存在。

71. 根据权利要求36所述的系统,其中以至少99%的特异性检测所述受试者的微卫星不稳定性的不存在。

72. 根据权利要求36所述的系统,其中以至少90%的阳性预测值 (PPV) 检测所述受试者的微卫星不稳定性的存在。

73. 根据权利要求36所述的系统,其中以至少0.90的曲线下面积 (AUC) 检测所述受试者的微卫星不稳定性的存在或不存在。

74. 根据权利要求36所述的系统,其中所述方法还包括:当所述平均z分数的绝对值不大于预定数量时,检测所述受试者的微卫星稳定性 (MSS) 的存在,或当所述平均z分数的绝对值大于预定数量时,检测所述受试者的微卫星稳定性 (MSS) 的不存在。

75. 一种包括机器可执行代码的非暂时性计算机可读介质,所述机器可执行代码在由一个或多个计算机处理器执行后进行一种用于评估受试者的微卫星不稳定性的方法,所述方法包括:

接收从受试者的血液样品中获得的多个来源于肿瘤的无细胞DNA (cfDNA) 分子的多个微卫星重复元件的量化量度,其中所述多个微卫星重复元件的量化量度包括所述多个微卫星重复元件中的每一个的平均长度,长度落入预定大小范围内的多个微卫星重复元件的数量或分数,或所述多个微卫星重复元件中的每一个的平均插入或缺失 (indel) 长度;

处理所述多个量化量度以获得所述多个量化量度的偏差的统计量度,其中所述偏差的统计量度是平均z分数;和

当所述平均z分数的绝对值大于预定数量时,检测所述受试者的微卫星不稳定性 (MSI) 的存在,或当所述平均z分数的绝对值不大于预定数量时,检测所述受试者的微卫星不稳定性 (MSI) 的不存在。

76. 根据权利要求75所述的非暂时性计算机可读介质,其中所述受试者被诊断出患有

癌症。

77. 根据权利要求75所述的非暂时性计算机可读介质,其中所述多个量化量度从在所述多个cfDNA分子中的所述多个微卫星重复元件的每一个处的一组测序读数测量。

78. 根据权利要求77所述的非暂时性计算机可读介质,其中所述方法还包括对所述多个cfDNA分子测序以产生所述一组测序读数。

79. 根据权利要求77或78所述的非暂时性计算机可读介质,其中所述测序包括全基因组测序(WGS)。

80. 根据权利要求79所述的非暂时性计算机可读介质,其中所述测序在不大于10X的深度下进行。

81. 根据权利要求79所述的非暂时性计算机可读介质,其中所述测序在不大于8X的深度下进行。

82. 根据权利要求79所述的非暂时性计算机可读介质,其中所述测序在不大于6X的深度下进行。

83. 根据权利要求75所述的非暂时性计算机可读介质,其中测量所述多个量化量度包括在所述多个微卫星重复元件中的每一个处进行对所述多个cfDNA分子的结合测量。

84. 根据权利要求75所述的非暂时性计算机可读介质,其中所述方法还包括基于所检测的所述受试者的微卫星不稳定性的存在或不存在,识别针对所述受试者的治疗或待施用到所述受试者的治疗有效量的治疗。

85. 根据权利要求84所述的非暂时性计算机可读介质,其中所述治疗选自化学疗法、放射疗法和免疫疗法。

86. 根据权利要求85所述的非暂时性计算机可读介质,其中所述治疗包括免疫疗法。

87. 根据权利要求86所述的非暂时性计算机可读介质,其中所述免疫疗法包括派姆单抗。

88. 根据权利要求75所述的非暂时性计算机可读介质,其中所述方法还包括针对所述多个微卫星重复元件的至少一个子集引导所述多个cfDNA分子的富集。

89. 根据权利要求88所述的非暂时性计算机可读介质,其中所述富集包括扩增所述多个cfDNA分子。

90. 根据权利要求89所述的非暂时性计算机可读介质,其中所述扩增包括选择性扩增。

91. 根据权利要求89所述的非暂时性计算机可读介质,其中所述扩增包括通用扩增。

92. 根据权利要求88所述的非暂时性计算机可读介质,其中所述富集包括选择性分离所述多个cfDNA分子的至少一部分用于靶向富集。

93. 根据权利要求92所述的非暂时性计算机可读介质,其中所述至少一部分包含单核苷酸。

94. 根据权利要求92所述的非暂时性计算机可读介质,其中所述至少一部分包含二核苷酸。

95. 根据权利要求75所述的非暂时性计算机可读介质,其中所述偏差的统计量度是相对于参考血液样品的平均z分数。

96. 根据权利要求95所述的非暂时性计算机可读介质,其中所述参考血液样品从具有微卫星不稳定性的受试者获得。

97. 根据权利要求95所述的非暂时性计算机可读介质,其中所述参考血液样品从不具有微卫星不稳定性的受试者获得。

98. 根据权利要求75所述的非暂时性计算机可读介质,其中所述预定数量是3。

99. 根据权利要求75所述的非暂时性计算机可读介质,其中所述多个微卫星重复元件包括单核苷酸或二核苷酸。

100. 根据权利要求99所述的非暂时性计算机可读介质,其中所述多个微卫星重复元件包括单核苷酸和二核苷酸。

101. 根据权利要求75所述的非暂时性计算机可读介质,其中所述多个微卫星重复元件包含至少一百万个不同的微卫星重复元件。

102. 根据权利要求75所述的非暂时性计算机可读介质,其中所述多个微卫星重复元件包括至少500万个不同的微卫星重复元件。

103. 根据权利要求75所述的非暂时性计算机可读介质,其中所述多个微卫星重复元件包括至少1000万个不同的微卫星重复元件。

104. 根据权利要求75所述的非暂时性计算机可读介质,其中所述多个微卫星重复元件包括至少2000万个不同的微卫星重复元件。

105. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少90%的灵敏度检测所述受试者的微卫星不稳定性的存在。

106. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少95%的灵敏度检测所述受试者的微卫星不稳定性的存在。

107. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少99%的灵敏度检测所述受试者的微卫星不稳定性的存在。

108. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少90%的特异性检测所述受试者的微卫星不稳定性的不存在。

109. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少95%的特异性检测所述受试者的微卫星不稳定性的不存在。

110. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少99%的特异性检测所述受试者的微卫星不稳定性的不存在。

111. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少90%的阳性预测值(PPV)检测所述受试者的微卫星不稳定性的存在。

112. 根据权利要求75所述的非暂时性计算机可读介质,其中以至少0.90的曲线下面积(AUC)检测所述受试者的微卫星不稳定性的存在或不存在。

113. 根据权利要求75所述的非暂时性计算机可读介质,其中所述方法还包括:当所述平均z分数的绝对值不大于预定数量时,检测所述受试者的微卫星稳定性(MSS)的存在,或当所述平均z分数的绝对值大于预定数量时,检测所述受试者的微卫星稳定性(MSS)的不存在。

## 评估微卫星不稳定性的方法和系统

[0001] 相关申请的交叉引用

[0002] 本申请要求美国临时专利申请号62/731,718的权益,该申请于2018年9月14日提交,其全部内容通过引用并入本文。

### 背景技术

[0003] 微卫星不稳定性(Microsatellite Instability,MSI)通常可以指由于受试者DNA错配修复(MMR)受损而导致的遗传易感性突变。在患有MSI的受试者中,MMR功能异常的细胞可能会在DNA复制期间积聚错误,产生突变的微卫星片段或重复的DNA序列。MSI可能在许多类型的癌症,例如结肠癌、胃癌、子宫内膜癌、卵巢癌、肝胆道癌、泌尿道癌、脑癌和皮肤癌中发挥重要作用。例如,MSI是检测遗传性非息肉性结直肠癌(HNPCC)或Lynch综合症(一种具有结肠癌和其他类型癌症的高风险的常染色体显性遗传疾病)的良好标志物。另外,微卫星状态可以指示受试者进行癌症治疗的预后。例如,在结肠癌患者中的MSI研究表明,与低MSI(MSI-L)或微卫星稳定(Microsatellite Stable,MSS)肿瘤的患者相比,MSI高的患者(MSI-H)的预后更好。

### 发明内容

[0004] 本文提供用于通过分析受试者的血液样品来评估受试者(例如患有癌症的患者)的微卫星不稳定性(MSI)的方法、系统和介质。微卫星不稳定性(MSI)可以通过在与包含单核苷酸和二核苷酸的微卫星对应的多个遗传基因座中分析受试者样品的肿瘤DNA(例如,来自无细胞DNA),和基于肿瘤DNA的分析测量来自受试者的血液样品的多个微卫星重复元件中的每一个的平均长度来评估和/或监测。例如,可以通过识别受试者中MSI的存在或不存在来评估受试者的MSI。可以基于例如相对于参考基因组或患者特定参考长度的微卫星重复元件的测量的平均插入或缺失(indel)长度、包含超过一定大小的插入或缺失(indel)的一组微卫星重复元件(例如两个重复单元的缺失)的分数、或每个微卫星基因座的测序数据中微卫星长度的平均数从一组选定的重复元件中产生MSI状态。受试者的MSI状态可以指示受试者的诊断、预后或治疗选择。

[0005] 在一些实施方式中,MSI状态可以在一段时间(例如,在两个或更多个不同的时间点)上变化(例如增加或减少)。在一些实施方式中,该持续时间可以对应于例如受试者的癌症的治疗过程或外科手术切除或肿瘤的其他疗法之后的监测时期(例如,以检测受试者中肿瘤的复发)。在一些实施方式中,MSI状态的产生可以包括针对对应于微卫星的多个遗传基因座中的每一个产生cfDNA测序读数的量化量度(quantitative measure)。多个遗传基因座可包含微卫星,例如人类参考基因组中的整组微卫星重复序列(或其子集),为了最小化微卫星稳定(MSS)数据中的噪声而优化的一组微卫星重复序列(或其子集),都是相同类别的一组微卫星重复序列(例如,重复单元的长度为1的所有重复序列,或其子集),在一定大小(例如长度)范围内的一组微卫星重复单元,其中测序数据表明缺少混杂的种系插入或缺失(indel)的一组微卫星重复序列(或其子集),为了最大化给定一组训练数据的算法性

能而优化的一组微卫星重复单元(或其子集)或其组合的并集或交集。在一些情况下,cfDNA的量化量度(例如,测序读数)可以包括与多个遗传基因座中的每一个对准的测序读数的计数。或者,获得cfDNA的量化量度可以包括在多个微卫星重复元件的每一个上进行多个cfDNA分子的结合测量。在一些实施方式中,MSI状态的产生可以包括产生针对cfDNA(例如,测序读数)的量化量度的比较(例如,差异或比率)。通过评估对应于微卫星的不同组的基因座之间的测序读数计数的比较,本文提供的方法可以允许产生MSI状态,这对于通过无创实验室测试(例如,基于血液的测试)对受试者的诊断、预后或治疗选择是有用的。

[0006] 一方面,本公开提供一种评估受试者的微卫星不稳定性的计算机实现的方法,所述方法包括:从受试者的血液样品获得多个微卫星重复元件的量化量度;处理所述多个量化量度以获得所述多个量化量度的偏差的统计量度(statistical measure);和当所述多个量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的存在,或当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的不存在。

[0007] 在一些实施方式中,多个微卫星重复元件的量化量度选自多个微卫星重复元件中的每一个(或其子集)的平均长度,长度落入预定大小范围内的多个微卫星重复元件(或其子集)的数量、频率或分数,以及多个微卫星重复元件(或其子元素)的平均插入或缺失(indel)长度。在一些实施方式中,受试者被诊断出患有癌症。在一些实施方式中,受试者对癌症无症状。在一些实施方式中,受试者具有一种或多种癌症危险因素(例如,年龄,性别,人种,种族,家族病史,吸烟或饮酒史,遗传变异的存在或其他临床健康特征)。在一些实施方式中,从多个无细胞DNA(cfDNA)分子测量多个量化量度。在一些实施方式中,从在多个cfDNA分子中的多个微卫星重复元件的每一个处的一组测序读数测量多个量化量度。在一些实施方式中,该方法还包括对多个cfDNA分子测序以产生该组测序读数。在一些实施方式中,测序包括全基因组测序(WGS)。在一些实施方式中,测序在不大于约50X,不大于约48X,不大于约46X,不大于约44X,不大于约42X,不大于约40X,不大于约38X,不大于约36X,不大于约34X,不大于约32X,不大于约30X,不大于约28X,不大于约24X,不大于约22X,不大于约20X,不大于约18X,不大于约16X,不大于约14X,或不大于约12X的深度下进行。在一些实施方式中,测序在不大于约10X的深度下进行。在一些实施方式中,测序在不大于约8X的深度下进行。在一些实施方式中,测序在不大于约6X的深度下进行。在一些实施方式中,测序在不大于约5X,不大于约4X,不大于约3X,不大于约2X或不大于约1X的深度下进行。在一些实施方式中,测量多个量化量度包括在多个微卫星重复元件(或其子集)的每一个处对多个cfDNA分子进行结合测量。

[0008] 在一些实施方式中,该方法还包括,基于检测的受试者的微卫星不稳定性的存在或不存在,识别对该受试者的治疗和/或向该受试者施用治疗有效量的治疗。在一些实施方式中,治疗选自化学疗法、放射疗法和免疫疗法。在一些实施方式中,治疗包括免疫疗法。在一些实施方式中,免疫疗法包含派姆单抗(pembrolizumab)。在一些实施方式中,该方法还包括针对多个微卫星重复元件的至少一个子集富集多个cfDNA分子。在一些实施方式中,富集包括扩增多个cfDNA分子。在一些实施方式中,扩增包括选择性扩增(例如,靶向PCR或靶向富集,然后进行通用或靶向PCR)。在一些实施方式中,扩增包括通用扩增(例如通用PCR)。在一些实施方式中,富集包括选择性分离多个cfDNA分子的至少一部分(例如,靶向富集)。

在一些实施方式中,该至少一部分包含单核苷酸。在一些实施方式中,该至少一部分包含二核苷酸。

[0009] 在一些实施方式中,偏差的统计量度是平均z分数。在一些实施方式中,偏差的统计量度是相对于参考血液样品的平均z分数。在一些实施方式中,参考血液样品从具有微卫星不稳定性的受试者(例如,MSI阳性受试者)获得。在一些实施方式中,参考血液样品从不具有微卫星不稳定性的受试者(例如,MSI阴性或MSS受试者)获得。在一些实施方式中,预定标准是平均z分数的绝对值大于预定数量。在一些实施方式中,预定数量是约1。在一些实施方式中,预定数量是约2。在一些实施方式中,预定数量是约3。在一些实施方式中,多个微卫星重复元件包含单核苷酸或二核苷酸。在一些实施方式中,多个微卫星重复元件包含单核苷酸和二核苷酸。

[0010] 在一些实施方式中,多个微卫星重复元件包括至少约一百万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约500万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约1000万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约2000万个不同的微卫星重复元件。

[0011] 在一些实施方式中,以至少约70%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约80%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约90%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约95%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约99%的灵敏度检测受试者的微卫星不稳定性的存在。

[0012] 在一些实施方式中,以至少约70%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约80%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约90%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约95%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约99%的特异性检测受试者的微卫星不稳定性的不存在。

[0013] 在一些实施方式中,以至少约70%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约80%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约90%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约95%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约99%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。

[0014] 在一些实施方式中,以至少约70%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约80%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约90%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约95%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%

的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约 99% 的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。

[0015] 在一些实施方式中,用至少约 0.70 的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约 0.80 的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约 0.90 的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约 0.95 的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约 0.96, 至少约 0.97 或至少约 0.98 的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约 0.99 的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。

[0016] 在一些实施方式中,该方法还包括:当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星稳定性 (MSS) 的存在,或当所述多个量化量度的偏差的统计量度满足预定标准时,检测微卫星稳定性 (MSS) 的不存在。

[0017] 在一些实施方式中,以至少约 70% 的灵敏度检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 80% 的灵敏度检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 90% 的灵敏度检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 95% 的灵敏度检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 96%, 至少约 97% 或至少约 98% 的灵敏度检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 99% 的灵敏度检测受试者的微卫星稳定性的存在。

[0018] 在一些实施方式中,以至少约 70% 的特异性检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 80% 的特异性检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 90% 的特异性检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 95% 的特异性检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 96%, 至少约 97% 或至少约 98% 的特异性检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 99% 的特异性检测受试者的微卫星稳定性的不存在。

[0019] 在一些实施方式中,以至少约 70% 的阳性预测值 (PPV) 检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 80% 的阳性预测值 (PPV) 检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 90% 的阳性预测值 (PPV) 检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 95% 的阳性预测值 (PPV) 检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 96%, 至少约 97% 或至少约 98% 的阳性预测值 (PPV) 检测受试者的微卫星稳定性的存在。在一些实施方式中,以至少约 99% 的阳性预测值 (PPV) 检测受试者的微卫星稳定性的存在。

[0020] 在一些实施方式中,以至少约 70% 的阴性预测值 (NPV) 检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 80% 的阴性预测值 (NPV) 检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 90% 的阴性预测值 (NPV) 检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 95% 的阴性预测值 (NPV) 检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 96%, 至少约 97% 或至少约 98% 的阴性预测值 (NPV) 检测受试者的微卫星稳定性的不存在。在一些实施方式中,以至少约 99% 的阴性预测值 (NPV) 检测受试者的微卫星稳定性的不存在。

[0021] 在一些实施方式中,用至少约0.70的曲线下面积(AUC)检测受试者的微卫星稳定性的存在或不存在。在一些实施方式中,用至少约0.80的曲线下面积(AUC)检测受试者的微卫星稳定性的存在或不存在。在一些实施方式中,用至少约0.90的曲线下面积(AUC)检测受试者的微卫星稳定性的存在或不存在。在一些实施方式中,用至少约0.95的曲线下面积(AUC)检测受试者的微卫星稳定性的存在或不存在。在一些实施方式中,以至少约0.96,至少约0.97或至少约0.98的曲线下面积(AUC)检测受试者的微卫星稳定性的存在或不存在。在一些实施方式中,用至少约0.99的曲线下面积(AUC)检测受试者的微卫星稳定性的存在或不存在。

[0022] 在另一方面,本公开提供一种系统,所述系统包括控制器,所述控制器包括或能够访问包括机器可执行指令的非暂时性计算机可读介质,所述机器可执行指令在由一个或多个计算机处理器执行时执行用于评估受试者的微卫星不稳定性的方法,所述方法包括:从受试者的血液样品获得多个微卫星重复元件的量化量度;处理所述多个量化量度以获得所述多个量化量度的偏差的统计量度;并且当所述多个量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的存在;或当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的不存在。

[0023] 在一些实施方式中,多个微卫星重复元件的量化量度选自多个微卫星重复元件中的每一个(或其子集)的平均长度,长度落入预定大小范围内的多个微卫星重复元件(或其子集)的数量、频率或分数,以及多个微卫星重复元件(或其子元素)中的每一个的平均插入或缺失(indel)长度。在一些实施方式中,受试者被诊断出患有癌症。在一些实施方式中,受试者对癌症无症状。在一些实施方式中,受试者具有一种或多种癌症危险因素(例如,年龄,性别,人种,种族,家族病史,吸烟或饮酒史,遗传变异的存在或其他临床健康特征)。在一些实施方式中,从多个无细胞DNA(cfDNA)分子测量多个量化量度。在一些实施方式中,从在多个cfDNA分子中的多个微卫星重复元件的每一个处的一组测序读数测量多个量化量度。在一些实施方式中,该系统的方法还包括对多个cfDNA分子测序以产生测序读数的组。在一些实施方式中,测序包括全基因组测序(WGS)。在一些实施方式中,测序在不大于约50X,不大于约48X,不大于约46X,不大于约44X,不大于约42X,不大于约40X,不大于约38X,不大于约36X,不大于约34X,不大于约32X,不大于约30X,不大于约28X,不大于约24X,不大于约22X,不大于约20X,不大于约18X,不大于约16X,不大于约14X,或不大于约12X的深度下进行。在一些实施方式中,测序在不大于约10X的深度下进行。在一些实施方式中,测序在不大于约8X的深度下进行。在一些实施方式中,测序在不大于约6X的深度下进行。在一些实施方式中,测序在不大于约5X,不大于约4X,不大于约3X,不大于约2X或不大于约1X的深度下进行。在一些实施方式中,测量多个量化量度包括在多个微卫星重复元件(或其子集)的每一个处对多个cfDNA分子进行结合测量。

[0024] 在一些实施方式中,该系统的方法还包括,基于检测的受试者的微卫星不稳定性的存在或不存在,识别对该受试者的治疗或待施用到该受试者的治疗有效量的治疗。在一些实施方式中,治疗选自化学疗法、放射疗法和免疫疗法。在一些实施方式中,治疗包括免疫疗法。在一些实施方式中,免疫疗法包括派姆单抗。在一些实施方式中,该系统的方法还包括针对多个微卫星重复元件的至少一个子集引导多个cfDNA分子的富集。在一些实施方式中,富集包括扩增多个cfDNA分子。在一些实施方式中,扩增包括选择性扩增(例如,靶向

PCR或靶向富集,然后进行通用或靶向PCR)。在一些实施方式中,扩增包括通用扩增(例如通用PCR)。在一些实施方式中,富集包括选择性分离多个cfDNA分子的至少一部分(例如,靶向富集)。在一些实施方式中,该至少一部分包含单核苷酸。在一些实施方式中,该至少一部分包含二核苷酸。

[0025] 在一些实施方式中,偏差的统计量度是平均z分数。在一些实施方式中,偏差的统计量度是相对于参考血液样品的平均z分数。在一些实施方式中,参考血液样品从具有微卫星不稳定性的受试者(例如,MSI阳性受试者)获得。在一些实施方式中,参考血液样品从不具有微卫星不稳定性的受试者(例如,MSI阴性或MSS受试者)获得。在一些实施方式中,预定标准是平均z分数的绝对值大于预定数量。在一些实施方式中,预定数量是约1。在一些实施方式中,预定数量是约2。在一些实施方式中,预定数量是约3。在一些实施方式中,多个微卫星重复元件包含单核苷酸或二核苷酸。在一些实施方式中,多个微卫星重复元件包含单核苷酸和二核苷酸。

[0026] 在一些实施方式中,多个微卫星重复元件包括至少约一百万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约500万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约1000万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约2000万个不同的微卫星重复元件。

[0027] 在一些实施方式中,以至少约70%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约80%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约90%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约95%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约99%的灵敏度检测受试者的微卫星不稳定性的存在。

[0028] 在一些实施方式中,以至少约70%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约80%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约90%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约95%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约99%的特异性检测受试者的微卫星不稳定性的不存在。

[0029] 在一些实施方式中,以至少约70%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约80%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约90%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约95%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约99%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。

[0030] 在一些实施方式中,以至少约70%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约80%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约90%的阴性预测值(NPV)检测受试者的微

卫星不稳定性的不存在。在一些实施方式中,以至少约95%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约99%的阴性预测值(NPV)检测受试者的微卫星不稳定性的不存在。

[0031] 在一些实施方式中,用至少约0.70的曲线下面积(AUC)检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.80的曲线下面积(AUC)检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.90的曲线下面积(AUC)检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.95的曲线下面积(AUC)检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.96,至少约0.97或至少约0.98的曲线下面积(AUC)检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.99的曲线下面积(AUC)检测受试者的微卫星不稳定性的存在或不存在。

[0032] 在一些实施方式中,该系统的方法还包括:当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星稳定性(MSS)的存在;或当所述多个量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星稳定性(MSS)的不存在。

[0033] 在另一方面,本公开提供一种包括机器可执行代码的非暂时性计算机可读介质,所述机器可执行代码在由一个或多个计算机处理器执行时进行一种用于评估受试者的微卫星不稳定性的方法,所述方法包括:获得来自受试者的血液样品的多个微卫星重复元件的量化量度;处理所述多个量化量度以获得所述多个量化量度的偏差的统计量度;并且当所述多个量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的存在;或当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的不存在。

[0034] 在一些实施方式中,多个微卫星重复元件的量化量度选自多个微卫星重复元件中的每一个(或其子集)的平均长度,长度落入预定大小范围内的多个微卫星重复元件(或其子集)的数量、频率或分数,以及多个微卫星重复元件(或其子元素)中的每一个的平均插入或缺失(indel)长度。在一些实施方式中,受试者被诊断出患有癌症。在一些实施方式中,受试者对癌症无症状。在一些实施方式中,受试者具有一种或多种癌症危险因素(例如,年龄,性别,人种,种族,家族病史,吸烟或饮酒史,遗传变异的存在或其他临床健康特征)。在一些实施方式中,从多个无细胞DNA(cfDNA)分子测量多个量化量度。在一些实施方式中,从在多个cfDNA分子中的多个微卫星重复元件的每一个处的一组测序读数测量多个量化量度。在一些实施方式中,非暂时性计算机可读介质的方法还包括对多个cfDNA分子测序以产生测序读数的组。在一些实施方式中,测序包括全基因组测序(WGS)。在一些实施方式中,测序在不大于约50X,不大于约48X,不大于约46X,不大于约44X,不大于约42X,不大于约40X,不大于约38X,不大于约36X,不大于约34X,不大于约32X,不大于约30X,不大于约28X,不大于约24X,不大于约22X,不大于约20X,不大于约18X,不大于约16X,不大于约14X,或不大于约12X的深度下进行。在一些实施方式中,测序在不大于约10X的深度下进行。在一些实施方式中,测序在不大于约8X的深度下进行。在一些实施方式中,测序在不大于约6X的深度下进行。在一些实施方式中,测序在不大于约5X,不大于约4X,不大于约3X,不大于约2X或不大于约1X的深度下进行。在一些实施方式中,测量多个量化量度包括在多个微卫星重复元件(或其子

集)的每一个上对多个cfDNA分子进行结合测量。

[0035] 在一些实施方式中,非暂时性计算机可读介质的方法还包括,基于检测的受试者的微卫星不稳定性的存在或不存在,识别针对受试者的治疗或待施用到受试者的治疗有效量的治疗。在一些实施方式中,治疗选自化学疗法、放射疗法和免疫疗法。在一些实施方式中,治疗包括免疫疗法。在一些实施方式中,免疫疗法包含派姆单抗。在一些实施方式中,非暂时性计算机可读介质的方法还包括针对多个微卫星重复元件的至少一个子集引导多个cfDNA分子的富集。在一些实施方式中,富集包括扩增多个cfDNA分子。在一些实施方式中,扩增包括选择性扩增(例如,靶向PCR或靶向富集,然后进行通用或靶向PCR)。在一些实施方式中,扩增包括通用扩增(例如通用PCR)。在一些实施方式中,富集包括选择性分离多个cfDNA分子的至少一部分(例如,靶向富集)。在一些实施方式中,该至少一部分包含单核苷酸。在一些实施方式中,该至少一部分包含二核苷酸。

[0036] 在一些实施方式中,偏差的统计量度是平均z分数。在一些实施方式中,偏差的统计量度是相对于参考血液样品的平均z分数。在一些实施方式中,参考血液样品从具有微卫星不稳定性的受试者(例如,MSI阳性受试者)获得。在一些实施方式中,参考血液样品从不具有微卫星不稳定性的受试者(例如,MSI阴性或MSS受试者)获得。在一些实施方式中,预定标准是平均z分数的绝对值大于预定数量。在一些实施方式中,预定数量是约1。在一些实施方式中,预定数量是约2。在一些实施方式中,预定数量是约3。在一些实施方式中,多个微卫星重复元件包含单核苷酸或二核苷酸。在一些实施方式中,多个微卫星重复元件包含单核苷酸和二核苷酸。

[0037] 在一些实施方式中,多个微卫星重复元件包括至少约一百万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约500万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约1000万个不同的微卫星重复元件。在一些实施方式中,多个微卫星重复元件包括至少约2000万个不同的微卫星重复元件。

[0038] 在一些实施方式中,以至少约70%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约80%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约90%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约95%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的灵敏度检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约99%的灵敏度检测受试者的微卫星不稳定性的存在。

[0039] 在一些实施方式中,以至少约70%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约80%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约90%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约95%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的特异性检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约99%的特异性检测受试者的微卫星不稳定性的不存在。

[0040] 在一些实施方式中,以至少约70%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约80%的阳性预测值(PPV)检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约90%的阳性预测值(PPV)检测受试者的微卫星

不稳定性的存在。在一些实施方式中,以至少约95%的阳性预测值 (PPV) 检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的阳性预测值 (PPV) 检测受试者的微卫星不稳定性的存在。在一些实施方式中,以至少约99%的阳性预测值 (PPV) 检测受试者的微卫星不稳定性的存在。

[0041] 在一些实施方式中,以至少约70%的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约80%的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约90%的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约95%的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约96%,至少约97%或至少约98%的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。在一些实施方式中,以至少约99%的阴性预测值 (NPV) 检测受试者的微卫星不稳定性的不存在。

[0042] 在一些实施方式中,用至少约0.70的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.80的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.90的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.95的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.96,至少约0.97或至少约0.98的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。在一些实施方式中,用至少约0.99的曲线下面积 (AUC) 检测受试者的微卫星不稳定性的存在或不存在。

[0043] 在一些实施方式中,非暂时性计算机可读介质的方法还包括:当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星稳定性 (MSS) 的存在,或当所述多个量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星稳定性 (MSS) 的不存在。

[0044] 本公开的另一方面提供一种包括机器可执行代码的非暂时性计算机可读介质,所述机器可执行代码在由一个或多个计算机处理器执行时实现本文上面或本文其他地方的任何方法。

[0045] 本公开的另一方面提供一种系统,所述系统包括一个或多个计算机处理器和与其耦合的计算机存储器。该计算机存储器包括机器可执行代码,所述机器可执行代码在由一个或多个计算机处理器执行时实现本文上面或本文其他地方的任何方法。

[0046] 通过下面的详细描述,本公开的其他方面和优点对于本领域技术人员将变得显而易见,其中,仅示出和描述本公开的说明性实施方式。将会认识到,本公开内容能够具有其他和不同的实施方式,并且其若干细节能够在各种明显的方面进行修改,而所有这些都脱离本公开内容。因此,附图和描述本质上应被认为是说明性的,而不是限制性的。

[0047] 通过引用并入

[0048] 本说明书中提及的所有出版物、专利和专利申请都以相同的程度通过引用并入本文,就好像每个单独的出版物、专利或专利申请被明确地并单独地指出通过引用并入一样。在通过引用并入的出版物和专利或专利申请与本说明书中包含的公开内容相抵触时,本说明书旨在取代和/或优先于任何此类矛盾的材料。

## 附图说明

[0049] 专利或申请文件包含至少一幅彩色附图。在请求并支付必要的费用后,专利局将提供带有彩色附图的本专利或专利申请公开的副本。

[0050] 本发明的新颖特征在所附的权利要求书中具体阐述。通过参考阐述其中利用本发明的原理的说明性实施方式的以下详细描述和附图(本文也称为“图(Figure)”和“图(FIG.)”),可以更好地理解本发明的特征和优点,其中:

[0051] 图1示出根据一些实施方式的评估受试者中的微卫星不稳定性的示例方法。

[0052] 图2示出四个不同患者群组的累积密度函数(CDF, y轴)与微卫星插入或缺失(indel)长度(x轴)的图:肿瘤TCGA-A6-A566-01A-11D-A28G,微卫星稳定(MSS)(左上);肿瘤TCGA-A6-A566-01A-11D-A28G,微卫星不稳定性高(MSI-H)(右上);肿瘤TCGA-D7-55,微卫星稳定(MSS)(左下);和肿瘤TCGA-D7-55,微卫星不稳定性高(MSI-H)(右下)。

[0053] 图3示出表示从微卫星稳定(MSS)患者(左,蓝色)和微卫星不稳定性高(MSI-H)患者(右,红色)测定的一组微卫星的平均插入或缺失(indel)长度的箱形图。

[0054] 图4示出表示从微卫星稳定(MSS)患者(左,蓝色)和微卫星不稳定性高(MSI-H)患者(右,红色)测定的一组微卫星的平均插入或缺失(indel)长度的箱形图。

[0055] 图5示出被编程或以其他方式配置为实现本文提供的方法的计算机系统。

## 具体实施方式

[0056] 尽管已经在本文中示出和描述本发明的各种实施方式,对于本领域技术人员显而易见的是,这些实施方式仅以示例的方式提供。在不脱离本发明的情况下,本领域技术人员可以想到许多变化、改变和替换。应当理解,可以采用本文所述的本发明的实施方式的各种替代方案。

[0057] 如说明书和权利要求书中所使用的,单数形式“一个”、“一种”(或不含冠词的名词)和“该”包括复数对象,除非上下文另外明确指出。例如,术语“核酸”包括多个核酸,包括其混合物。

[0058] 如本文所用,术语“核酸”或“多核苷酸”通常是指包含一个或多个核酸亚基或核苷酸的分子。核酸可以包括一个或多个选自腺苷(A)、胞嘧啶(C)、鸟嘌呤(G)、胸腺嘧啶(T)和尿嘧啶(U)或其变体的核苷酸。核苷酸通常包括核苷和至少1、2、3、4、5、6、7、8、9、10或更多个磷酸根( $PO_3$ )。核苷酸可以单独地或组合地包括核碱基、五碳糖(核糖或脱氧核糖)以及一个或多个磷酸根。

[0059] 核糖核苷酸是其中糖为核糖的核苷酸。脱氧核糖核苷酸是其中糖是脱氧核糖的核苷酸。核苷酸可以是核苷单磷酸或核苷多磷酸。核苷酸可以是易于掺入的形式,例如脱氧核糖核苷多磷酸,例如脱氧核糖核苷三磷酸(dNTP),其可以选自脱氧腺苷三磷酸(dATP)、脱氧胞苷三磷酸(dCTP)、脱氧鸟苷三磷酸(dGTP)、尿苷三磷酸(dUTP)和脱氧胸苷三磷酸(dTTP) dNTP,它们包括可检测的标签,例如发光标签或标记(例如,荧光团)。核苷酸可以包括可以掺入正在生长的核酸链中的任何亚基。此类亚基可以是A、C、G、T或U,或对一个或多个互补A、C、G、T或U特有的任何其他亚基,或与嘌呤(例如A或G,或其变体)或嘧啶(例如C、T或U,或其变体)互补的任何其他亚基。在一些实例中,核酸是脱氧核糖核酸(DNA)、核糖核酸(RNA)或其衍生物或变体。核酸可以是单链或双链的。核酸分子可以是线性的、弯曲的或圆形的或

其任何组合。

[0060] 如本文所用的术语“核酸分子”、“核酸序列”、“核酸片段”、“寡核苷酸”和“多核苷酸”通常是指可以具有各种长度的多核苷酸,例如脱氧核糖核苷酸或核糖核苷酸(RNA)或其类似物。核酸分子可以具有至少约5个碱基,10个碱基,20个碱基,30个碱基,40个碱基,50个碱基,60个碱基,70个碱基,80个碱基,90,100个碱基,110个碱基,120个碱基,130个碱基,140个碱基,150个碱基,160个碱基,170个碱基,180个碱基,190个碱基,200个碱基,300个碱基,400个碱基,500个碱基,1千个碱基(kb),2kb,3kb,4kb,5kb,10kb,或50kb的长度,或它可以具有在上述两个值之间的任何数量的碱基。寡核苷酸通常由四个核苷酸碱基的特定序列组成:腺嘌呤(A);胞嘧啶(C);鸟嘌呤(G);和胸腺嘧啶(T)(当多核苷酸是RNA时,对于胸腺嘧啶(T)是尿嘧啶(U))。因此,术语“核酸分子”、“核酸序列”、“核酸片段”、“寡核苷酸”和“多核苷酸”至少部分意图是多核苷酸分子的字母表示。或者,该术语可以应用于多核苷酸分子本身。该字母表示可以输入到具有中央处理单元的计算机中的数据库中和/或用于诸如功能基因组学和同源性搜索的生物信息学应用。寡核苷酸可以包括一种或多种非标准核苷酸、核苷酸类似物和/或修饰的核苷酸。

[0061] 如本文所用,术语“样品”通常是指生物学样品。生物学样品的实例包括核酸分子、氨基酸、多肽、蛋白质、碳水化合物、脂肪或病毒。在一个实例中,生物学样品是包括一个或多个核酸分子的核酸样品。核酸分子可以是无细胞的或无细胞的核酸分子,例如无细胞DNA(cfDNA)或无细胞RNA(cfRNA)。核酸分子可源自多种来源,包括人类、哺乳动物、非人类哺乳动物、猿、猴、黑猩猩、爬虫类、两栖或禽类来源。此外,可以从含有无细胞序列的各种动物体液中提取样品,这些动物体液包括但不限于血液、血清、血浆、玻璃体、痰液、尿液、眼泪、汗液、唾液、精液、粘膜排泄物、粘液、脊髓液、羊水、淋巴液等。无细胞的多核苷酸(例如,cfDNA)可以是胎儿来源的(经由从怀孕的受试者获取的液体),或可以源自受试者本身的组织。

[0062] 如本文所用,术语“受试者”通常是指具有正在被加工或分析的生物学样品的个体。受试者可以是动物或植物。受试者可以是哺乳动物,例如人、狗、猫、马、猪或啮齿动物。受试者可以是患者,例如患有或被怀疑患有疾病,例如一种或多种癌症(例如脑癌,乳腺癌,子宫颈癌,结肠直肠癌,子宫内膜癌,食道癌,胃癌,肝胆道癌症,白血病,肝癌,肺癌,淋巴瘤,卵巢癌,胰腺癌,皮肤癌,泌尿道癌),一种或多种传染病,一种或多种遗传性疾病或一种或多种肿瘤或其任意组合。对于患有或怀疑患有的一种或多种肿瘤的受试者,肿瘤可以是一种或多种类型。

[0063] 如本文所用,术语“全血”通常是指尚未被分离成子组分(例如,通过离心)的血液样品。血液样品的全血可能包含cfDNA和/或种系DNA。可以从血液样品中提取全血DNA(可能包含cfDNA和/或种系DNA)。全血DNA测序读数(可能包含cfDNA测序读数和/或种系DNA测序读数)可从全血DNA中提取。

[0064] 微卫星不稳定性(MSI)通常可以指由于受试者DNA错配修复(MMR)受损而导致的遗传易感性突变。在这类受试者中,MMR功能异常的细胞可能会在DNA复制期间积聚错误,导致突变的微卫星片段或重复的DNA序列。MSI可能在许多类型的癌症中发挥重要作用,例如结肠癌、胃癌、子宫内膜癌、卵巢癌、肝胆道癌、泌尿道癌、脑癌和皮肤癌。例如,MSI是检测遗传性非息肉性结直肠癌(HNPCC)或Lynch综合症(一种具有结肠癌和其他类型癌症的高风险的

常染色体显性遗传病症)的良好标志物。另外,微卫星状态可以指示受试者进行癌症治疗的预后。例如,在结肠癌患者中进行的MSI研究表明,与低MSI (MSI-L) 或微卫星稳定 (MSS) 肿瘤的患者相比,MSI-高的患者 (MSI-H) 的预后更好。

[0065] 可以根据美国国家癌症研究所 (NCI) 建立的方法确定MSI状态,该方法可以使用五个微卫星标记来指示MSI存在:两个单核苷酸 (BAT25和BAT26) 和三个二核苷酸重复序列 (D2S123, D5S346和D17S250)。可将MSI大于不稳定MSI生物标志物的约30%的那些称为MSI-H肿瘤,而将MSI小于不稳定MSI生物标志物的约30%的那些称为MSI-L肿瘤。

[0066] MSI-L肿瘤可分类为其他病因性肿瘤。研究可能表明,MSI-H患者对单独的手术反应是最佳的,而不是化学疗法和手术组合。对MSI-H状态的准确识别可能阻止向患者进行和施用可能无效的治疗例如化疗。

[0067] 另外,可以至少部分地基于对患者中的MSI的识别来向患者进行和施用癌症治疗。例如,美国食品药品监督管理局 (FDA) 加快Keytruda™ (派姆单抗) 的批准,用于具有特征在于高度微卫星不稳定性或错配修复缺陷的不可切除或转移性实体瘤的成人和儿科患者,这些患者已经在替代药物方面取得进展。MSI状态的准确识别可以允许做出准确的临床决策,例如向患者进行和施用靶向治疗例如Keytruda™ (派姆单抗)。

[0068] 确定患者中的MSI状态的方法可以包括组织分析。例如,可以在一组遗传基因座 (例如,一组标准的五个NCI推荐基因座) 中的每个基因座上进行聚合酶链反应 (PCR) 和成对的正常和肿瘤组织样品的片段分析以确定微卫星不稳定性 (MSI)。组织分析可能会将报告的阳性测试结果产生为MSI高 (表明至少两个标志物不稳定) 或将报告的阴性测试结果产生为MSI低 (表明一个标志物不稳定)。MSI状态确定的此类方法可能需要获取肿瘤组织用于分析。在某些情况下,获取肿瘤组织可能带来挑战。获取组织可能很耗时且昂贵,需要与病理医生协调。如果不是不可能获得,则获得活检组织可能很困难,可能昂贵且涉及痛苦的过程,并且由于潜在的癌症基因组进化而可能产生低至中等的临床相关性。在某些情况下,可能需要等到初步癌症诊断后的几年才能确定患者是否符合Keytruda™ (派姆单抗) 的资格。因此,用于确定MSI状态的液体活检测试可提供比肿瘤活检更早、侵入性更小且成本更低的替代方法的优点。

[0069] 评估来自受试者的DNA序列数据中的微卫星不稳定性

[0070] 当从受试者获取的样品的显著部分 (例如,大于约50%、约60%、约70%、约80%或约90%) 来自或源自肿瘤细胞时,对微卫星不稳定性 (MSI) 状态的评估可能相对简单。然而,在从血液样品中衍生的受试者血浆中制备无细胞DNA (cfDNA) 中,从cfDNA中检测肿瘤DNA以及由此评估微卫星不稳定性 (MSI) 状态可能是不灵敏且噪声多的过程。由于来自非肿瘤DNA (例如来自非肿瘤来源的种系细胞的种系DNA) 的压倒性信号,从这种不敏感和/或噪声多的信号中检测肿瘤DNA和评估微卫星不稳定性 (MSI) 状态可能具有挑战性。本公开提供用于从无细胞DNA (cfDNA) 序列数据 (例如,cfDNA测序读数) 或源自受试者样品的cfDNA分子的结合测量评估微卫星不稳定性 (MSI) 状态的方法、系统和介质。一旦从来自受试者的样品分析获得cfDNA序列数据,可以使用一个或多个生物信息学过程来评估受试者的微卫星不稳定性 (MSI) 状态。

[0071] 一方面,本公开提供一种用于评估受试者的微卫星不稳定性 (MSI) 的计算机实现的方法,所述方法包括:从受试者的血液样品获得多个微卫星重复元件的量化量度;处理所述多

个量化量度以获得所述多个量化量度的偏差的统计量度;并且当所述多个量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的存在,或当所述多个量化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的不存在。

[0072] 图1示出根据一些实施方式的评估受试者中的微卫星不稳定性的示例方法。在一些实施方式中,从多个无细胞DNA(cfDNA)分子测量量化量度(例如,多个平均长度)(如在105中)。在一些实施方式中,测量多个平均长度包括对多个cfDNA分子测序以在多个cfDNA分子中的多个微卫星重复元件的每一个处产生测序读数(如在110中)。

[0073] 例如,可以使用任何合适的测序方法从cfDNA产生测序读数。测序方法可以是第一代测序方法,例如Maxam-Gilbert或Sanger测序,或高通量测序(例如,下一代测序或NGS)方法。高通量测序方法可以同时(或基本上同时)对至少约10,000,约100,000,约100万,约1000万,约1亿,约10亿或多于约10亿的多核苷酸分子测序。测序方法可以包括但不限于:焦磷酸测序,合成测序,单分子测序,纳米孔测序,半导体测序,连接测序,杂交测序,数字基因表达(Helicos),大规模平行测序,例如Helicos,克隆单分子阵列(Solexa/Illumina),使用PacBio、SOLiD、Ion Torrent或Nanopore平台测序。

[0074] 在一些实施方式中,测序包括全基因组测序(WGS)。测序可以在足以评估具有期望性能(例如准确性,敏感性,特异性,阳性预测值(PPV),阴性预测值(NPV)或接收器操作者特征(ROC)的曲线下面积(AUC))的受试者中的微卫星不稳定性深度下进行。在一些实施方式中,以“低通”方式测序,例如深度不大于约12X,不大于约11X,不大于约10X,不大于约9X,不大于约8X,不大于约7X,不大于约6X,不大于约5X,不大于约4X,不大于约3X,或不大于约2X。

[0075] 在一些实施方式中,评估受试者中的微卫星不稳定性可以包括将cfDNA测序读数与参考基因组比对。参考基因组可以包含基因组(例如,人类基因组)的至少一部分。参考基因组可以包含整个基因组(例如,整个人类基因组)。参考基因组可以包括数据库,该数据库包含对应于基因组的编码和/或非编码基因组区域的多个基因组区域。该数据库可以包括多个基因组区域,其对应于基因组的癌症相关(或肿瘤相关)的编码和/或非编码基因组区域,例如癌症驱动基因突变(例如,单核苷酸变体(SNV),拷贝数变异(CNV),插入或缺失(indels),融合基因和微卫星重复元件(例如单核苷酸和/或二核苷酸))。例如,可以使用Burrows-Wheeler算法或任何其他合适的对准算法来执行对准。

[0076] 在一些实施方式中,评估受试者中的微卫星不稳定性可以包括产生针对多个遗传基因座中的每一个的cfDNA测序读数的量化量度。可能会产生cfDNA测序读数的量化量度,例如与给定基因座(例如微卫星重复元件)对准的DNA测序读数的计数。具有与给定的微卫星重复元件对准的部分或全部测序读数的cfDNA测序读数可以计入该微卫星重复元件的量化量度。

[0077] 在一些实施方式中,多个微卫星重复元件选自人类参考基因组中的整组微卫星重复序列(或其子集),为了最小化MSS数据中的噪声而优化的一组微卫星重复序列(或其子集),都是相同类别的一组微卫星重复序列(例如,重复单元的长度为1的所有重复序列),在一定大小(例如长度)范围内的一组微卫星重复单元,其中测序数据表明缺少混杂的种系indel的一组微卫星重复序列,为了最大化给定一组训练数据的算法性能而优化的一组微

卫星重复单元(或其子集)或其组合的并集或交集。特定和非特定微卫星重复元件的模式可以指示微卫星不稳定性(MSI)状态或微卫星稳定性(MSS)状态。这些微卫星重复元件的模式随时间的变化可能表示微卫星不稳定性(MSI)状态或微卫星稳定性(MSS)状态的变化。

[0078] 在一些实施方式中,测量多个平均长度包括在多个微卫星重复元件的每一个上执行多个cfDNA分子的结合测量。在一些实施方式中,进行结合测量包括使用对多个cfDNA分子中的多个微卫星重复元件的至少一部分具有选择性的探针来分析多个cfDNA分子。在一些实施方式中,探针是与多个微卫星重复元件的核酸序列具有序列互补性的核酸分子。在一些实施方式中,这些核酸分子是引物或富集序列。在一些实施方式中,测定包括使用阵列杂交或聚合酶链反应(PCR)或核酸测序。

[0079] 在一些实施方式中,该方法还包括富集多个微卫星重复元件的至少一部分的多个cfDNA分子。在一些实施方式中,富集包括扩增多个cfDNA分子。例如,可以通过选择性扩增(例如,通过使用包含与多个微卫星重复元件的核酸序列具有序列互补性的核酸分子的一组引物或探针)来扩增多个cfDNA分子。替代地或组合地,可以通过通用扩增(例如,通过使用通用引物)来扩增多个cfDNA分子。在一些实施方式中,富集包括选择性分离多个cfDNA分子的至少一部分(例如,单核苷酸和/或二核苷酸)。

[0080] 在一些实施方式中,评估受试者中的微卫星不稳定性的方法包括处理多个平均长度以获得平均长度的偏差的量化量度(例如,统计量度)(如在115中)。在一些实施方式中,偏差的统计量度是相对于一种或多种参考血液样品的平均z分数。参考血液样品可以从具有微卫星不稳定性的受试者和/或从不具有微卫星不稳定性的受试者获得。参考血液样品可以从患有癌症的受试者或从不具有癌症(例如乳腺癌,子宫颈癌,结肠直肠癌,子宫内膜癌,食道癌,胃癌,肝胆道癌,白血病,肝癌,肺癌,淋巴瘤,卵巢癌,胰腺癌,皮肤癌,尿路癌)的受试者获得。

[0081] 在一些实施方式中,评估受试者中的微卫星不稳定性的方法还包括:当平均长度的偏差的统计量度满足预定标准时,确定受试者的微卫星不稳定性(MSI)(如在120中)。偏差的统计量度可以是平均z分数或相对于参考样品或参考值的平均z分数。在一些实施方式中,预定标准是平均z分数的绝对值大于预定数量。预定数量可以是约0.1,约0.2,约0.3,约0.4,约0.5,约0.6,约0.7,约0.8,约0.9,约1,约1.5,约2,约2.5,约3,约3.5,约4,约4.5,约5,或大于约5。

[0082] 在一些实施方式中,多个微卫星重复元件包含单核苷酸和/或二核苷酸。多个微卫星重复元件可以包括至少约10个不同的微卫星重复元件,至少约50个不同的微卫星重复元件,至少约100个不同的微卫星重复元件,至少约500个不同的微卫星重复元件,至少约1000个不同的微卫星重复元件,至少约5000个不同的微卫星重复元件,至少约1万个不同的微卫星重复元件,至少约5万个不同的微卫星重复元件,至少约10万个不同的微卫星重复元件,至少约50万个不同的微卫星重复元件,至少约100万个不同的微卫星重复元件,至少约200万个不同的微卫星重复元件,至少约300万个不同的微卫星重复元件,至少约400万个不同的微卫星重复元件,至少约500万个不同的微卫星重复元件,至少约1000万个不同的微卫星重复元件,至少约1500万个不同的微卫星重复元件,至少约2000万个不同的微卫星重复元件,至少约2500万个不同的微卫星重复元件,至少约3000万个不同的微卫星重复元件,或多于3000万个不同的微卫星重复元件。

[0083] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的灵敏度检测所述受试者的微卫星不稳定性 (MSI) 的存在。

[0084] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的特异性检测所述受试者的微卫星不稳定性 (MSI) 的不存在。

[0085] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的阳性预测值 (PPV) 检测所述受试者的微卫星不稳定性 (MSI) 的存在。

[0086] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的阴性预测值 (NPV) 检测受试者的微卫星不稳定性 (MSI) 不存在。

[0087] 在一些实施方式中,以至少约0.50,至少约0.55,至少约0.60,至少约0.65,至少约0.70,至少约0.75,至少约0.80,至少约0.85,至少约0.90,至少约0.95,至少约0.96,至少约0.97,至少约0.98,或至少约0.99的接收器操作者特征 (ROC) 的曲线下面积 (AUC) 检测受试者的微卫星不稳定性 (MSI)。

[0088] 在一些实施方式中,评估受试者中的微卫星不稳定性的方法还包括:当平均长度的偏差的统计量度不满足预定标准时,确定受试者的微卫星稳定性 (MSS) 的存在,或当平均长度的偏差的统计量度满足预定标准时,确定受试者的微卫星稳定性 (MSS) 的不存在。

[0089] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的灵敏度检测所述受试者的微卫星稳定性 (MSS) 的存在。

[0090] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的特异性检测所述受试者的微卫星稳定性 (MSS) 的不存在。

[0091] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的阳性预测值 (PPV) 检测所述受试者的微卫星稳定性 (MSS) 的存在。

[0092] 在一些实施方式中,以至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约55%,至少约60%,至少约65%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约96%,至少约97%,至少约98%,或至少约99%的阴性预测值 (NPV) 检测受试者的微卫星稳定性 (MSS) 不存在。

[0093] 在一些实施方式中,以至少约0.50,至少约0.55,至少约0.60,至少约0.65,至少约0.70,至少约0.75,至少约0.80,至少约0.85,至少约0.90,至少约0.95,至少约0.96,至少约0.97,至少约0.98,或至少约0.99的接收器操作者特征(ROC)的曲线下面积(AUC)检测所述受试者的微卫星稳定性(MSS)的不存在。

[0094] 在一些实施方式中,受试者已被诊断出患有癌症。例如,癌症可以是一种或多种类型,包括:脑癌、乳腺癌、子宫颈癌、结肠直肠癌、子宫内膜癌、食道癌、胃癌、肝胆道癌、白血病、肝癌、肺癌、淋巴瘤、卵巢癌、胰腺癌、皮肤癌或泌尿道癌。

[0095] 在一些实施方式中,该方法还包括基于确定的受试者微卫星不稳定性的存在或不存在,施用治疗有效量的治疗和/或识别治疗以治疗受试者的微卫星不稳定性。在一些实施方式中,治疗包括化学疗法、放射疗法或免疫疗法。例如,治疗可以包括免疫疗法,例如Keytruda™(派姆单抗)。

[0096] 可以评估受试者的微卫星不稳定性(MSI)或微卫星稳定性(MSS)以确定受试者中的癌症诊断、癌症预后或肿瘤进展或消退的迹象。此外,可基于微卫星不稳定性(MSI)或微卫星稳定性(MSS)评估或监测(例如,两个或更多个时间点之间的微卫星不稳定性(MSI)或微卫星稳定性(MSS)状态差异)分配一项或多项临床结果。这样的临床结果可以包括诊断患有包括一种或多种类型的肿瘤的癌症的受试者,诊断患有包括一种或多种类型和阶段的肿瘤的癌症的受试者,预测患有癌症的受试者(例如,指示治疗的临床过程(例如手术,化学疗法、放射疗法,免疫疗法或其他治疗),指示另一种临床作用过程(例如,无治疗,在规定的的时间间隔基础上继续监测,停止当前治疗,改用另一种治疗),或指示受试者的预期生存时间。

[0097] 在一些实施方式中,评估受试者的微卫星不稳定性(MSI)的方法还包括确定微卫星不稳定性(MSI)或微卫星稳定性(MSS)是否大于预定阈值。可以通过对来自一个或多个对照受试者的一个或多个样品(例如,已知患有某种肿瘤类型的患者,已知患有某个阶段的某种肿瘤类型的患者或未表现出任何癌症的健康受试者)进行一个微卫星不稳定性(MSI)或微卫星稳定性(MSS)评估并基于对照样品的微卫星不稳定性(MSI)或微卫星稳定性(MSS)评估识别合适的预定阈值来产生预定阈值。

[0098] 可以基于期望的灵敏度、特异性、阳性预测值(PPV)、阴性预测值(NPV)或评估受试者的微卫星不稳定性(MSI)或微卫星稳定性(MSS)状态的准确性来调节预定阈值。例如,如果期望评估受试者的微卫星不稳定性(MSI)或微卫星稳定性(MSS)状态的高灵敏度,则可以将预定阈值调节为较低。或者,如果期望评估受试者的微卫星不稳定性(MSI)或微卫星稳定性(MSS)状态的高特异性,则可以将预定阈值调节为更高。可以调节预定阈值,以最大化从对照受试者获得的对照样品的接收器操作者特征(ROC)的曲线下面积(AUC)。可以调节预定阈值,以便在评估包含一种或多种类型的肿瘤的癌症的微卫星不稳定性(MSI)或微卫星稳定性(MSS)中的假阳性(FP)和假阴性(FN)之间实现期望的平衡。

[0099] 在一些实施方式中,评估微卫星不稳定性(MSI)或微卫星稳定性(MSS)的方法还包括在第二稍后的时间点重复该评估。可以选择第二时间点以相对于第一时间点进行微卫星不稳定性(MSI)或微卫星稳定性(MSS)评估的适当比较。第二时间点的实例可以对应于外科切除后的时间,在治疗施用期间或在治疗施用以治疗受试者中的癌症后监测治疗的有效性的时间,或在治疗后在受试者中无法检测癌症后监测受试者中的残留疾病或癌症复发的时间。

[0100] 在一些实施方式中,评估微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 的方法还包括确定第一微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态与第二微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态之间的差异,其差异指示受试者的肿瘤的进展或消退。替代地或组合地,该方法还可以包括由计算机处理器产生第一微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态和第二微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态随第一时间点和第二时间点变化的图,该图指示受试者的肿瘤的进展或消退。例如,计算机处理器可以产生y轴上的两个或更多个微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态相对于与x轴上的所述两个或更多个微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态对应的数据的收集时间对应的时间的图。

[0101] 确定的差异或说明第一微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态与第二微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态之间的差异的图可指示受试者的肿瘤的进展或消退。如果第二微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态大于第一微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态,则该差异可以指示例如肿瘤进展,对受试者中的肿瘤的治疗无效,肿瘤对正在进行治疗有抗性,肿瘤向受试者中的其他部位转移,或受试者中的残留的疾病或癌症复发。如果第二微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态小于第一微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态,则该差异可以指示例如肿瘤消退,手术切除受试者中的肿瘤的有效性,对受试者中的肿瘤的治疗功效或受试者中没有残留疾病或癌症复发。

[0102] 在评估和/或监控微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态之后,可以基于微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态评估或监视 (例如,两个或更多个时间点之间的微卫星不稳定性 (MSI) 或微卫星稳定性 (MSS) 状态差异) 分配一个或多个临床结果。这样的临床结果可以包括诊断患有包括一种或多种类型的肿瘤的癌症的受试者,诊断患有包括一种或多种类型和阶段的肿瘤的癌症的受试者,预测患有癌症的受试者 (例如,指示治疗的临床过程 (例如手术、化学疗法、放射疗法、免疫疗法或其他治疗),指示另一种临床作用过程 (例如,无治疗,在规定的的时间间隔基础上继续监测,停止当前治疗,改用另一种治疗),或指示受试者的预期生存时间。

[0103] 实施例

[0104] 实施例1:通过全基因组测序从患者正常-肿瘤配对样品中测定MSI

[0105] 从来自癌症患者的受试者获得的约500组正常-肿瘤配对组织样品中收集全基因组测序数据。针对短重复单元 (例如,单核苷酸和二核苷酸) 富集对应于所评估的微卫星的一组130万个遗传基因座。在MSI-H肿瘤中,单核苷酸重复序列可能丰富且突变频率更高。对于每个微卫星,测量每个正常-肿瘤配对组织样品的平均长度,并计算平均长度的差。由于MSI-H正常-肿瘤配对在微卫星中具有更多的缺失,而微卫星稳定 (MSS) 肿瘤则没有,对正常-肿瘤配对的每个微卫星的测量的平均长度进行分析以确定受试者的MSI状态。

[0106] 图2示出累积密度函数 (CDF, y轴) 与患者的4个不同队列中的每个队列的微卫星插入或缺失 (indel) 长度 (x轴): 肿瘤TCGA-A6-A566-01A-11D-A28G, 微卫星稳定 (MSS) (左上); 肿瘤TCGA-A6-A566-01A-11D-A28G, 微卫星不稳定性高 (MSI-H) (右上); 肿瘤TCGA-D7-55, 微卫星稳定 (MSS) (左下); 和肿瘤TCGA-D7-55, 微卫星不稳定性高 (MSI-H) (右下)。如图2所示, 对于MSS状态的患者的两个队列, 所测量的累积密度函数 (CDF) 表明, 在所测定的肿瘤组织

和正常组织样品中,所测量的的大多数微卫星的indel长度约为零。该结果表明MSS正常-肿瘤配对具有基本上相同的微卫星长度。相反,对于MSI-H状态的患者的两个队列,所测量的累积密度函数(CDF)表明,所测量的的大部分微卫星具有在测定的肿瘤组织样品中约为零的负indel长度(约-6至约0的范围)。该结果表明,MSI-H正常-肿瘤配对具有不同微卫星长度的微卫星的统计学显著部分。

[0107] 图3示出表示从微卫星稳定(MSS)患者(左,蓝色)和微卫星不稳定性高(MSI-H)患者(右,红色)测定的一组微卫星的平均插入或缺失(indel)长度的箱形图。如图3所示,对于MSS状态的患者,测量的平均indel长度具有以约零的中值为中心的分布,具有小的标准偏差。相反,对于MSI-H状态的患者,测量的平均indel长度具有以约0.5的中值为中心的分布,具有显著更大的标准偏差。特别地,几乎所有的平均indel长度的绝对值均显著大于零。如果样品的平均indel长度的z分数小于约-3(例如,具有大于约3的预定阈值的绝对值),则将其视为MSI-H。以约98.9%的高灵敏度和93.1%的高特异性,基于通过组织的全基因组测序(WGS)获得的下一代测序(NGS)数据来确定患者的MSI状态。

#### [0108] 实施例2:通过全基因组测序从患者血液样品中进行MSI测定

[0109] 从来自癌症患者的受试者获得的约几组血液样品中收集全基因组测序数据。从患者收集血液样品以分析无细胞DNA(cfDNA),以测定循环肿瘤DNA(ctDNA)的微卫星不稳定性状态。针对短重复单元(例如单核苷酸和二核苷酸)富集对应于所评估的微卫星的一组130万个遗传基因座。在MSI-H肿瘤中,单核苷酸重复序列可能丰富且突变频率更高。对于每个微卫星,测量每个血液样品的平均长度。由于MSI-H正常-肿瘤配对在微卫星中具有更多的缺失,而微卫星稳定(MSS)肿瘤则没有,可以对血液样品中每个微卫星的测量平均长度进行分析,以确定受试者的MSI状态。

[0110] 通过计算机模拟从肿瘤组织获得的1%测序读数掺入到患者匹配的正常背景读数(例如,从受试者的正常-肿瘤配对样品的正常组织获得的测序读数)来模拟通过对从患者获得的血液样品进行下一代测序(NGS)获得的全基因组测序数据。即使在低肿瘤分数下(例如,诸如倾向于在血液中观察到的那些),也观察到微卫星长度的差异,从而使得能够在受试者中区分MSI-H和MSS状态。

[0111] 图4示出表示从微卫星稳定(MSS)患者(左,蓝色)和微卫星不稳定性高(MSI-H)患者(右,红色)测定的一组微卫星的平均插入或缺失(indel)长度的箱形图。如图4所示,对于MSS状态的患者,所测量的平均indel长度具有以约零的中值为中心的分布,具有小的标准偏差。相反,对于MSI-H状态的患者,测量的平均indel长度具有以约0.01的中值为中心的分布,具有显著更大的标准偏差。特别地,几乎所有的平均indel长度的绝对值均显著大于零。如果样品的平均indel长度的z分数的绝对值大于预定阈值,则将其视为MSI-H。以95.7%的高灵敏度、99.1%的高特异性和1.7的分类差距,基于计算机模拟的测序数据确定患者的MSI状态,该测序数据从具有低的1%肿瘤分数的血液样品测量。

#### [0112] 计算机系统

[0113] 本公开提供被编程为实现本公开的方法的计算机系统。图5示出计算机系统501,该计算机系统501被编程或以其他方式配置为例如从受试者的血液样品获得微卫星重复元件的量化量度,处理该量化量度以获得该量化量度的偏差的统计量度,以及当量化量度的偏差的统计量度满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的存在,或当量

化量度的偏差的统计量度不满足预定标准时,检测所述受试者的微卫星不稳定性(MSI)的不存在。计算机系统501可以调节本公开的分析、计算和产生的各个方面,例如,从受试者的血液样品获得微卫星重复元件的量化量度,处理该量化量度以获得量化量度的偏差的统计量度,并且当量化量度的偏差的统计量度满足预定标准时,检测受试者的微卫星不稳定性(MSI)的存在,或当所述多个量化量度的偏差的统计量度不满足预定标准时,检测受试者的微卫星不稳定性(MSI)的不存在。计算机系统501可以是用户的电子设备或相对于电子设备位于远程的计算机系统。该电子设备可以是移动电子设备。

[0114] 计算机系统501包括中央处理单元(CPU,在本文中也称为“处理器”和“计算机处理器”)505,其可以是单核或多核处理器,或用于并行处理的多个处理器。计算机系统501还包括存储器或存储器位置510(例如,随机存取存储器,只读存储器,闪存),电子存储单元515(例如,硬盘),用于与一个或多个其他系统通信的通信接口520(例如,网络适配器),和外围设备525,例如高速缓存,其他存储器,数据存储和/或电子显示适配器。存储器510,存储单元515,接口520和外围设备525通过诸如主板的通信总线(实线)与CPU 505通信。存储单元515可以是用于存储数据的数据存储单元(或数据存储库)。计算机系统501可以借助于通信接口520可操作地耦合到计算机网络(“网络”)530。网络530可以是Internet、因特网和/或外部网,或与Internet通信的内部网和/或外部网。在某些情况下,网络530是电信和/或数据网络。网络530可以包括一个或多个计算机服务器,其可以启用分布式计算,例如云计算。例如,一个或多个计算机服务器可以使网络530上的云计算(“云”)能够执行本公开的分析、计算和产生的各个方面,例如从受试者的血液样品获得微卫星重复元件的量化量度,处理量化量度以获得量化量度的偏差的统计量度,并且当量化量度的偏差的统计量度满足预定标准时,确定受试者的微卫星不稳定性。这样的云计算可以由诸如亚马逊网络服务(AWS)、微软Azure、谷歌云平台和IBM云的云计算平台提供。网络530在某些情况下借助于计算机系统501可以实现对等网络,该对等网络可以使耦合到计算机系统501的设备能够充当客户端或服务器。

[0115] CPU 505可以执行可以在程序或软件中实现的一系列机器可读指令。指令可以存储在诸如存储器510的存储器位置中。指令可以针对CPU 505,其可以随后对CPU 505进行编程或以其他方式配置CPU 505以实现本公开的方法。由CPU 505执行的操作的示例可以包括获取、解码、执行和写回。

[0116] CPU 505可以是诸如集成电路的电路的一部分。系统501的一个或多个其他部件可以被包括在电路中。在某些情况下,该电路是专用集成电路(ASIC)。

[0117] 存储单元515可以存储文件,例如驱动器、库和保存的程序。存储单元515可以存储用户数据,例如,用户偏好和用户程序。在某些情况下,计算机系统501可以包括位于计算机系统501外部的一个或多个其他数据存储单元,例如位于通过内部网或Internet与计算机系统501通信的远程服务器上。

[0118] 计算机系统501可以通过网络530与一个或多个远程计算机系统通信。例如,计算机系统501可以与用户(例如,医师,护士,看护者,患者或受试者)的远程计算机系统通信。远程计算机系统的示例包括个人计算机(例如,便携式PC),平板电脑或平板计算机(例如, Apple® iPad, Samsung® Galaxy Tab),电话,智能电话(例如, Apple® iPhone,支

持Android的设备, **Blackberry®**) 或个人数字助理。用户可以通过网络530访问计算机系统501。

[0119] 可以通过机器(例如, 计算机处理器)可执行代码来实现本文所描述的方法, 该机器可执行代码存储在计算机系统501的电子存储位置上, 例如存储在存储器510或电子存储单元515上。机器可执行或机器可读代码可以以软件的形式提供。在使用期间, 代码可以由处理器505执行。在某些情况下, 可以从存储单元515中检索代码并将其存储在存储器510中以供处理器505随时访问。在某些情况下, 可以排除电子存储单元515, 并且将机器可执行指令存储在存储器510中。

[0120] 可以对代码进行预编译并配置为与具有适用于执行代码的处理器一起使用, 或可以在运行期间编译。可以以可以选择的编程语言来提供代码, 以使代码能够以预编译或编译时的方式执行。

[0121] 本文提供的系统和方法(例如计算机系统501)的各个方面可以体现在编程中。可以将技术的各个方面视为机器(或处理器)可执行代码和/或关联数据的形式“产品”或“制品”, 其通常以机器可读介质的类型承载或体现。机器可执行代码可以存储在电子存储单元, 例如存储器(例如, 只读存储器, 随机存取存储器, 闪存)或硬盘上。“存储”类型的介质可以包括计算机、处理器等的任何或所有有形存储器, 或其相关模块, 例如可以对软件编程随时提供非暂时性存储的各种半导体存储器、磁带驱动器、磁盘驱动器等。软件的全部或部分有时可以通过Internet或其他各种电信网络通信。例如, 这样的通信可以使得能够将软件从一个计算机或处理器加载到另一计算机或处理器中, 例如从管理服务器或主机计算机加载到应用服务器的计算机平台中。因此, 可以承载软件元件的另一种类型的介质包括例如通过有线和光学座机网络并通过各种空中链路跨本地设备之间的物理接口使用的光波、电波和电磁波。携带这种波的物理元件例如有线或无线链路、光链路等也可以被视为承载软件的介质。如本文所使用, 除非限于非暂时性的有形“存储”介质, 否则术语例如计算机或机器“可读介质”是指参与向处理器提供指令以供执行的任何介质。

[0122] 因此, 机器可读介质例如计算机可执行代码可以采取许多形式, 包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质包括例如光盘或磁盘, 例如任何计算机中的任何存储设备等, 例如可用于实现附图中所示的数据库等。易失性存储介质包括动态存储器, 例如这种计算机平台的主存储器。有形的传输介质包括同轴电缆; 铜线和光纤, 包括构成计算机系统内总线的电线。载波传输介质可以采用电信号或电磁信号或声波或光波的形式, 例如在射频(RF)和红外(IR)数据通信期间产生的那些。因此, 计算机可读介质的常见形式包括: 软盘, 软性盘, 硬盘, 磁带, 任何其他磁介质, CD-ROM, DVD或DVD-ROM, 任何其他光学介质, 打孔卡纸磁带, 带孔图案的任何其他物理存储介质, RAM, ROM, PROM和EPROM, FLASH-EPROM, 任何其他存储芯片或盒带, 用于传输数据或指令的载波, 用于传输此类载波的电缆或链路或计算机可以从中读取编程代码和/或数据的任何其他介质。这些形式的计算机可读介质中的许多可能涉及将一个或多个指令的一个或多个序列传送到处理器用于执行。

[0123] 计算机系统501可以包括电子显示器535或与电子显示器535通信, 该电子显示器535包括用户界面(UI)540, 用于提供例如来自受试者的血液样品的微卫星重复元件的测量的平均长度, 平均长度的偏差的统计量度, 以及检测的受试者的微卫星不稳定性(MSI)或微

卫星稳定性 (MSS) 的存在或不存在。UI 的示例包括但不限于图形用户界面 (GUI) 和基于 Web 的用户界面。

[0124] 本公开的方法、系统和介质可以通过一种或多种算法来实现。可以由中央处理单元 505 在执行时通过软件来实现算法。该算法例如可以从受试者的血液样品获得微卫星重复元件的量化量度, 处理该量化量度以获得量化量度的偏差的统计量度, 并且当量化量度的偏差的统计量度满足预定标准时, 检测所述受试者的微卫星不稳定性 (MSI) 的存在, 或当所述多个量化量度的偏差的统计量度不满足预定标准时, 检测所述受试者的微卫星不稳定性 (MSI) 的不存在。

[0125] 尽管已经在本文中示出和描述本发明的优选实施方式, 但是对于本领域技术人员而言显而易见的是, 仅通过示例的方式提供这样的实施方式。并非意图通过说明书中提供的特定实施例来限制本发明。尽管已经参考前述说明书描述本发明, 本文中的实施方式的描述和图示并不意味着以限制性的意义来解释。在不脱离本发明的情况下, 本领域技术人员现在将想到许多变化、改变和替代。此外, 应当理解, 本发明的所有方面不限于本文所阐述的具体描述、构造或相对比例, 其取决于各种条件和变量。应当理解, 本文所述的本发明的实施方式的各种替代方案可以用于实施本发明。因此, 可以预期的是, 本发明也将涵盖任何这样的替代、修改、变化或等同形式。意图是所附权利要求限定本发明的范围, 并且由此涵盖这些权利要求范围内的方法和结构及其等同物。

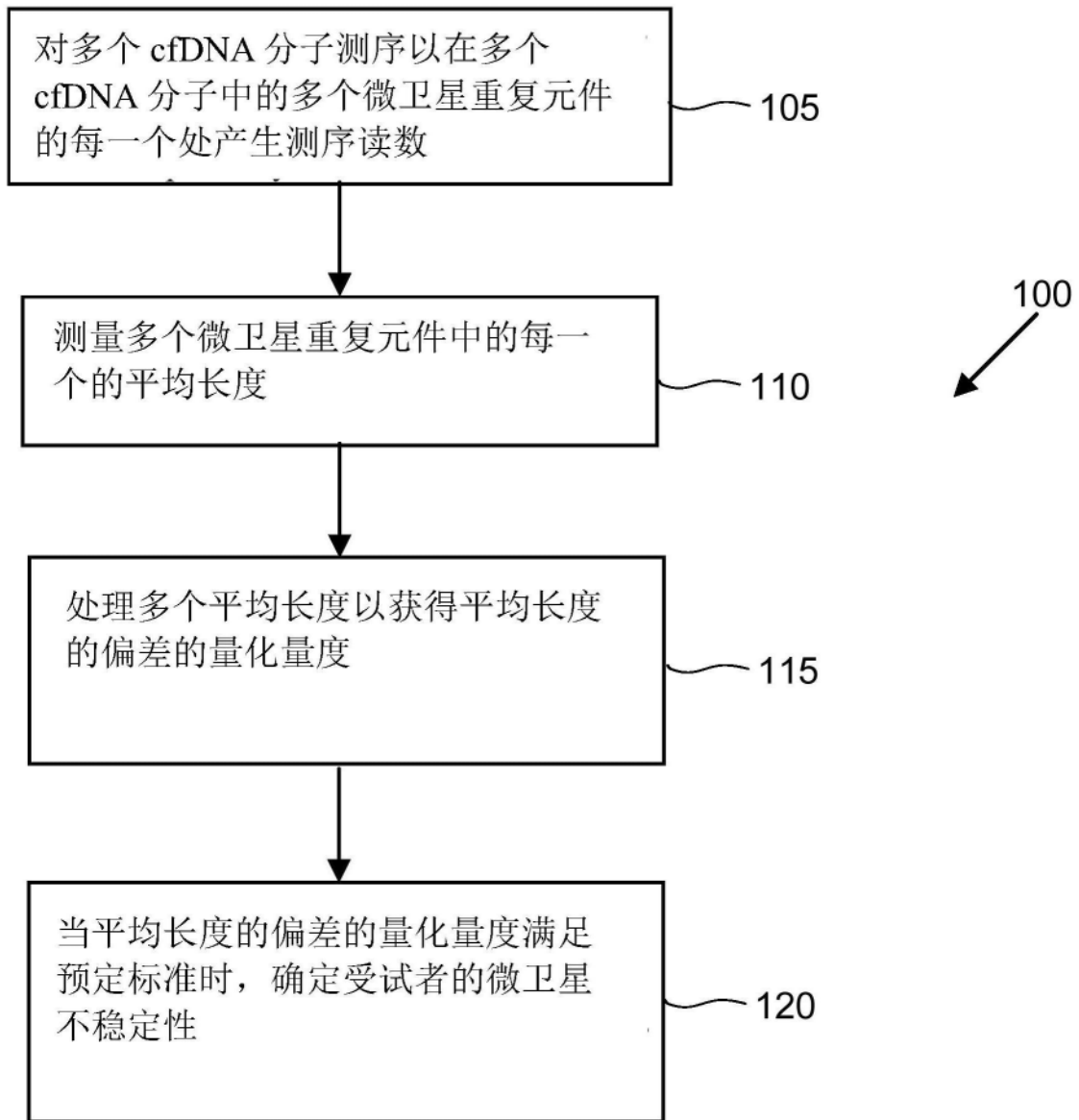
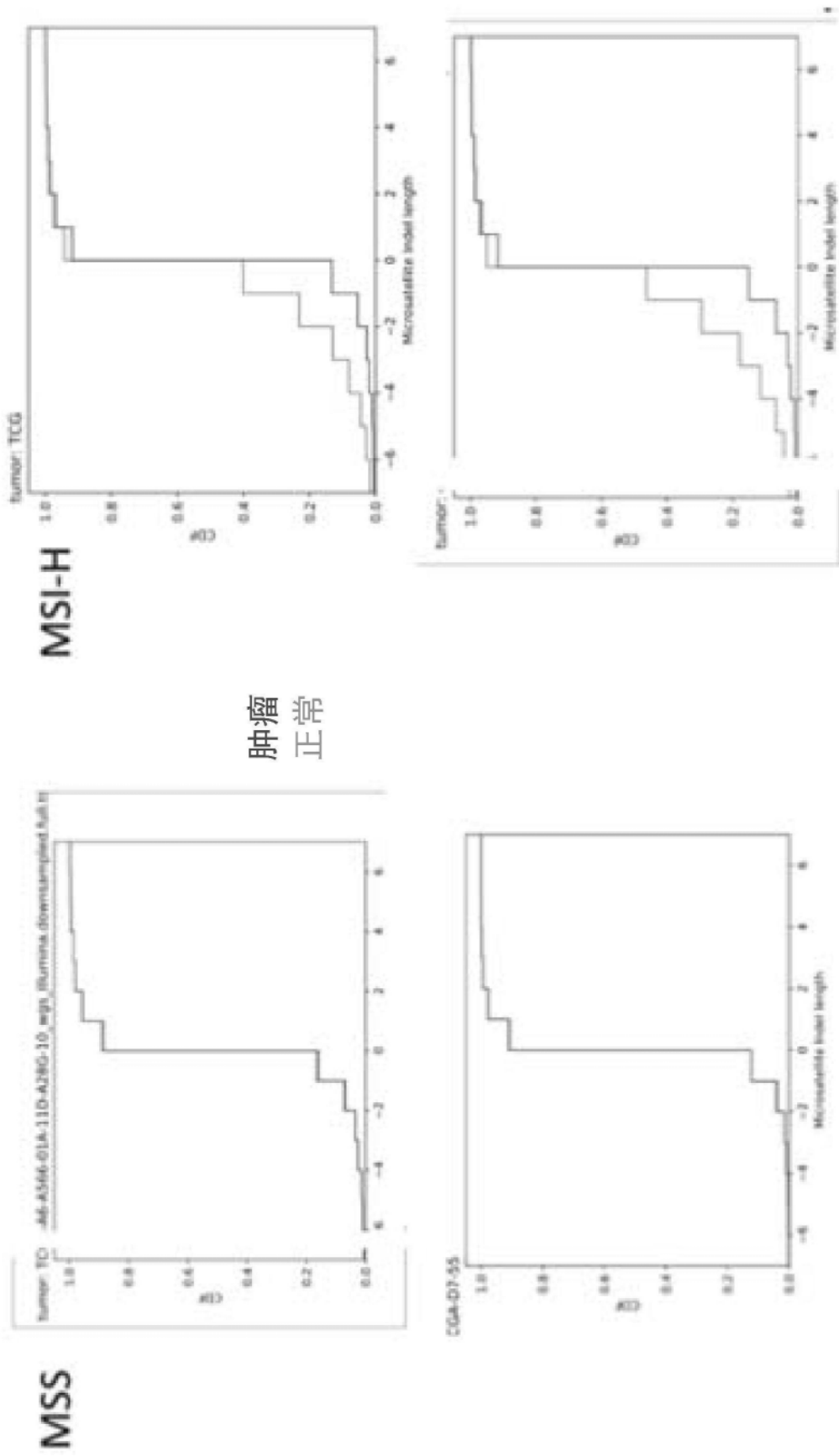


图1



肿瘤  
正常

图2

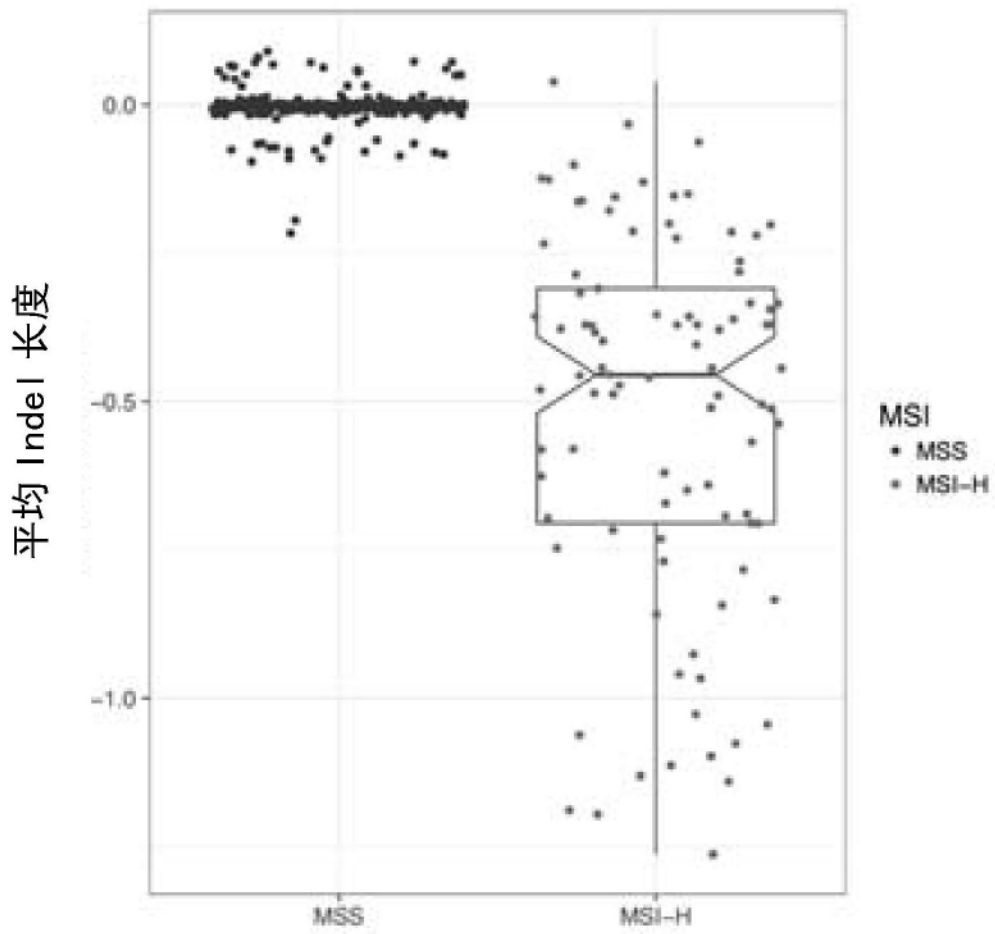


图3

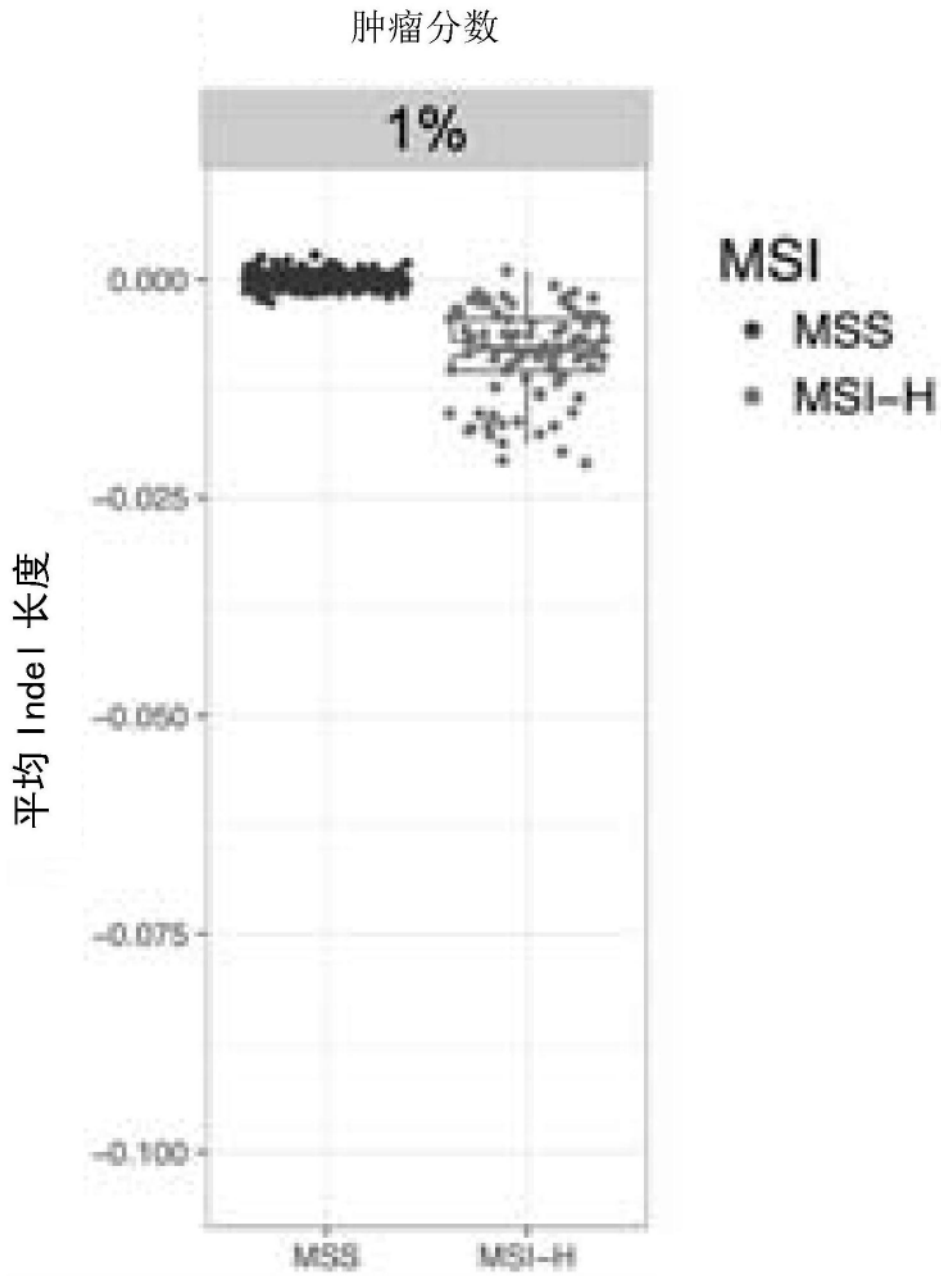


图4

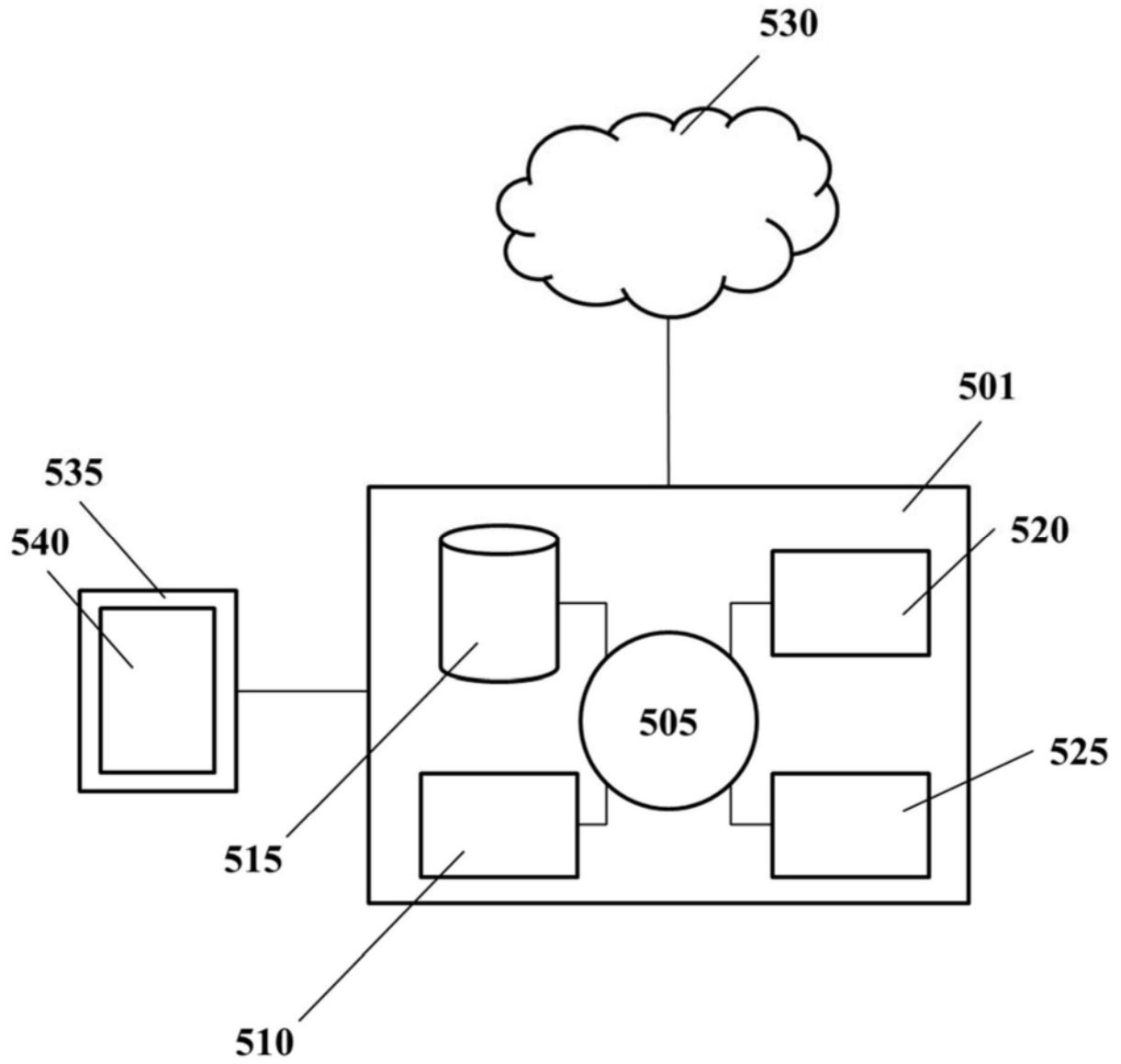


图5