



(12)发明专利申请

(10)申请公布号 CN 109563544 A

(43)申请公布日 2019.04.02

(21)申请号 201680071855.0

(74)专利代理机构 北京市金杜律师事务所
11256

(22)申请日 2016.10.11

代理人 王茂华 李春辉

(30)优先权数据

62/239,202 2015.10.08 US

(51)Int.Cl.

C12Q 1/6886(2018.01)

(85)PCT国际申请进入国家阶段日

2018.06.07

(86)PCT国际申请的申请数据

PCT/US2016/056451 2016.10.11

(87)PCT国际申请的公布数据

W02017/062989 EN 2017.04.13

(71)申请人 会聚基因学有限公司

地址 美国加利福尼亚州

(72)发明人 T·莱文 C·金 K·菲利普斯

M·伊万斯-霍姆 李越峒

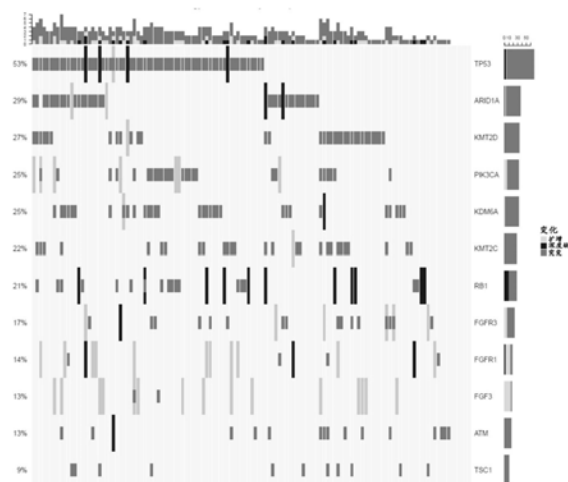
权利要求书4页 说明书34页 附图7页

(54)发明名称

用于膀胱癌的尿监测的诊断测定

(57)摘要

一种改进的诊断测定和与诊断测定有关的方法,该方法涉及突变聚焦的疾病诊断和监管生物标记小组,其中,基于它们涵盖患者群体的基因组多样性的能力选择潜在的基因组区域,最大化在每个患者内监测的独特标记的数目被最大化,同时平衡这些因素与经验测序性能、具有跨多样性患者的区域的事件的地理群集、以及与测量相应基因组区域相关联的大小和成本。方法还包括质量控制步骤以减少噪声并且最大化相关标记的存在。



1. 一种用于诊断受试者中的膀胱癌和/或监测受试者中的膀胱癌的复发的方法,包括:
 - a) 收集测试样本,所述测试样本选自来自受试者的含有核酸的一个或多个尿样和/或肿瘤样本;
 - b) 分离所述核酸,所述核酸来自所述受试者的所述一个或多个尿样和/或肿瘤样本中收集;
 - c) 分析所述核酸,以获得核酸序列数据;以及
 - d) 确定在所述核酸中的与膀胱癌相关联的至少一个基因中的至少一个突变和/或表观基因学改变的存在或不存在,其中,在所述至少一个基因中的所述至少一个突变和/或表观基因学改变选自表1中所列出的所述突变。
2. 根据权利要求1所述的方法,其中,分离的所述核酸在没有扩增步骤的情况下被直接分析。
3. 根据权利要求1所述的方法,其中,分离的所述核酸是RNA,并且其中,所述RNA在某些实施例中被逆转录成互补DNA,并且其中,逆转录能够单独执行或与扩增步骤组合执行。
4. 根据权利要求1所述的方法,其中,分离的所述核酸在没有扩增步骤的情况下被直接分析。
5. 根据权利要求1所述的方法,其中,在分析步骤c)之前,扩增所述核酸,以富集已知的膀胱癌基因。
6. 根据权利要求1所述的方法,其中,所述测试样本包括一个或多个尿样。
7. 根据权利要求1所述的方法,其中,所述测试样本包括在不同时间点所收集的一个或多个尿样。
8. 根据权利要求1所述的方法,还包括具有计算机可执行指令的计算机可读介质,所述计算机可执行指令用于确定受试者的膀胱癌的诊断和/或用于监测受试者的膀胱癌的复发。
9. 根据权利要求1所述的方法,其中,所述分析步骤c)还包括计算机可读介质,所述计算机可读介质包括一个或多个分析算法。
10. 根据权利要求1所述的方法,还包括:从源自所述受试者的基因型数据集来确定在与膀胱癌相关联的基因中一个或多个基因中的一个或多个突变的所述存在或不存在。
11. 根据权利要求1所述的方法,其中,在所述样本或更多样本中的所述核酸与一个或多个试剂接触,所述一个或多个试剂适合于检测在与膀胱癌相关联的一个或多个基因中的一个或多个突变和/或表观基因学改变的所述存在或不存在。
12. 根据权利要求1所述的方法,还包括:将在第一尿样核酸中所检测到的所述一个或多个突变和/或表观基因学改变的所述存在或不存在与在第二尿样核酸中所检测到的所述一个或多个突变和/或表观基因学改变的所述存在或不存进行比较,并且基于所述比较来监测所述受试者中膀胱癌的所述复发。
13. 根据权利要求1所述的方法,其中,使用杂交测定来检测所述至少一个基因中的所述至少一个突变,其中,基于一个或多个等位基因特异性寡核苷酸探针与来自所述受试者的DNA样本中的一个或多个核酸分子的杂交,确定基因突变的所述存在或不存在。
14. 根据权利要求1所述的方法,其中,通过测序技术来检测基因中的所述突变。
15. 根据权利要求1所述的方法,还包括:在不同时间点从所述受试者获得含有所述核

酸的第一尿样和第二尿样。

16. 一种用于监测受试者中的膀胱癌进展的方法,包括:

a) 收集测试样本,所述测试样本选自来自受试者的含有核酸的一个或多个尿样和/或肿瘤样本;

b) 分离所述核酸,所述核酸来自所述受试者的所述一个或多个尿样和/或肿瘤样本中收集;

c) 分析所述核酸,以获得核酸序列数据;以及

d) 确定在所述核酸中的与膀胱癌相关联的至少一个基因中的至少一个突变和/或表观基因学改变的存在或不存在,其中,所述至少一个基因中的所述至少一个突变和/或表观基因学改变选自表1中所列出的所述突变。

17. 根据权利要求16所述的方法,其中,分离的所述核酸在没有扩增步骤的情况下被直接分析。

18. 根据权利要求16所述的方法,其中,分离的所述核酸是RNA,其中,所述RNA在某些实施例中被逆转录成互补DNA,其中,逆转录能够单独执行或与扩增步骤组合执行。

19. 根据权利要求16所述的方法,其中,分离的所述核酸在没有扩增步骤的情况下被直接分析。

20. 根据权利要求16所述的方法,其中,在分析步骤c)之前,扩增所述核酸,以富集已知的膀胱癌基因。

21. 根据权利要求16所述的方法,还包括计算机可读介质,所述计算机可读介质具有用于确定患膀胱癌的受试者的癌症进展的计算机可执行指令。

22. 根据权利要求16所述的方法,其中,所述分析步骤c)还包括计算机可读介质,所述计算机可读介质包括一个或多个分析算法。

23. 根据权利要求16所述的方法,还包括:从源自所述受试者的基因型数据集中确定在与膀胱癌相关联的基因中一个或多个基因中的一个或多个突变的所述存在或不存在。

24. 根据权利要求16所述的方法,其中,所述样本或更多样本中的所述核酸与一个或多个试剂接触,所述一个或多个试剂适合于检测与膀胱癌相关联的一个或多个基因中的一个或多个突变和/或表观基因学改变的所述存在或不存在。

25. 根据权利要求16所述的方法,还包括:将在第一尿样核酸中检测到的所述一个或多个突变和/或表观基因学改变的所述存在或不存在与在第二尿样核酸中检测到的所述一个或多个突变和/或表观基因学改变的所述存在或不存在进行比较,并且基于所述比较来监测所述受试者中的癌症进展。

26. 根据权利要求16所述的方法,其中,使用杂交测定来检测所述至少一个基因中的所述至少一个突变,其中,基于一个或多个等位基因特异性寡核苷酸探针与来自所述受试者的DNA样本中的一个或多个核酸分子的杂交,确定基因突变的所述存在或不存在。

27. 根据权利要求16的方法,其中,通过测序技术来检测基因中的所述突变。

28. 根据权利要求16所述的方法,还包括:在不同时间点从所述受试者获得含有所述核酸的第一尿样和第二尿样。

29. 一种用于确定受试者对膀胱癌的易感性的方法,包括:

a) 收集测试样本,所述测试样本选自来自受试者的含有核酸的一个或多个尿样和/或

肿瘤样本；

b) 分离所述核酸,所述核酸从来自所述受试者的所述一个或多个尿样和/或肿瘤样本中收集;

c) 分析所述核酸,以获得核酸序列数据;以及

d) 确定在所述核酸中的与膀胱癌相关联的至少一个基因中的至少一个突变和/或表观基因学改变的存在或不存在,其中,在所述至少一个基因中的所述至少一个突变和/或表观基因学改变选自表1中所列出的所述突变。

30. 根据权利要求29所述的方法,其中分离的所述核酸在没有扩增步骤的情况下被直接分析。

31. 根据权利要求29所述的方法,其中,分离的所述核酸是RNA,其中,所述RNA在某些实施例中被逆转录成互补DNA,其中,逆转录能够单独执行或与扩增步骤组合执行。

32. 根据权利要求29所述的方法,其中,分离的所述核酸包括DNA和/或RNA,分离的所述核酸在没有扩增步骤的情况下被直接分析。

33. 根据权利要求29所述的方法,其中,在分析步骤c)之前,扩增所述核酸,以富集已知的膀胱癌基因。

34. 根据权利要求29所述的方法,其中,所述至少一个突变的所述存在和/或相对丰度指示所述受试者中对膀胱癌的易感性增加。

35. 根据权利要求29所述的方法,其中,所述分析步骤c)还包括计算机可读介质,所述计算机可读介质包括一个或多个分析算法。

36. 根据权利要求29所述的方法,还包括计算机可读介质,所述计算机可读介质具有用于确定受试者对膀胱癌的易感性的计算机可执行指令,其中,所述计算机可读存储器包括数据,所述数据指示被诊断患有膀胱癌的多个个体中的至少一个基因的至少一个突变和/或表观基因学改变的频率。

37. 根据权利要求29所述的方法,还包括计算机可读介质,所述计算机可读介质具有用于确定受试者对膀胱癌的易感性的计算机可执行指令,其中,所述计算机可读存储器包括数据,所述数据指示被诊断患有膀胱癌的多个个体中的至少一个基因的至少一个突变和/或表观基因学改变的频率,并且其中,针对所述受试者的风险测量能够基于针对所述受试者的基因型数据集与风险的比较,所述风险与所述至少一个基因或所述至少一个基因型数据集的所述至少一个突变和/或表观基因学改变相关联。

38. 根据权利要求29所述的方法,还包括计算机可读介质,所述计算机可读介质具有用于确定受试者对膀胱癌的易感性的计算机可执行指令,其中,所述计算机可读存储器包括指示被诊断患有膀胱癌的多个个体中的至少一个基因或至少一个基因型数据集的至少一个突变和/或表观基因学改变的频率的数据,所述存储器还能够包括指示多个参考个体中的至少一个基因或至少一个基因型数据集的至少一个突变和/或表观基因学改变的频率的数据。

39. 根据权利要求29所述的方法,还包括:确定在源自个体或受试者的基因型数据集中的至少一个或多个基因中与膀胱癌相关联的至少一个或多个突变的所述存在或不存在,其中,对所述至少一个突变的所述存在和/或相对丰度的确定指示所述受试者中对膀胱癌的易感性增加。

40. 根据权利要求29所述的方法,还包括:确定在源自所述受试者的基因型数据集的与膀胱癌相关联的基因中的一个或多个基因中的一个或多个突变的所述存在或不存在,其中,所述基因型数据集能够包括关于一个或多个突变或表观基因学标记的等位基因信息。

41. 根据权利要求29所述的方法,其中,对所述易感性的确定包括将所述核酸序列数据与数据库进行比较,所述数据库包含在所述至少一个突变和/或表观基因学标记与对膀胱癌的易感性之间的相关性数据。

42. 根据权利要求29所述的方法,其中,使用杂交测定来检测所述至少一个基因中的所述至少一个突变,其中,基于一个或多个等位基因特异性寡核苷酸探针与来自所述受试者的DNA样本中的一个或多个核酸分子的杂交,确定基因突变的所述存在或不存在。

43. 根据权利要求29的方法,其中,通过测序技术检测基因中的所述突变。

44. 一种用于评估受试者中对膀胱癌的易感性的试剂盒,包括:

a) 一种或多种试剂,对于选择性地检测所述受试者的基因组中与膀胱癌相关联的至少一个基因的至少一个突变和/或表观基因学改变所必需,其中,所述至少一个突变和/或表观基因学改变的存在指示对膀胱癌的易感性增加;以及

b) 数据,包括在所述至少一个突变和/或表观基因学改变与对膀胱癌的易感性之间的相关性数据。

45. 根据权利要求44所述的试剂盒,其中,所述一种或多种试剂包括至少一种连续寡核苷酸,所述至少一种连续寡核苷酸与包括所述至少一个突变的个体的所述基因组的片段杂交。

46. 根据权利要求44所述的试剂盒,其中,至少一对寡核苷酸与从所述受试者获得的基因组区段的相对链杂交,其中,每个寡核苷酸引物对被设计成选择性地扩增包括一个突变和/或表观基因学改变的个体的所述基因组的片段。

47. 根据权利要求44所述的试剂盒,其中,所述突变和/或表观基因学改变能够选自由表1中限定的所述突变和/或表观基因学改变组成的组。

用于膀胱癌的尿监测的诊断测定

技术领域

[0001] 本文的公开内容涉及癌症的识别,并且更具体地涉及通过基因生物标记和改进的基因测序和分析方法来检测、预后、诊断和处理个体或个体组的膀胱癌。

背景技术

[0002] 预计膀胱癌将成为北美第六最常见的实体癌,2014年美国估计有超过74,000例新病例(美国癌症协会,“Cancer Facts&Figures 2014.”2014)。诊断通常是在触发看医生的无痛性血尿症状(即尿中的血液)之后进行的。膀胱癌的常见风险因素包括吸烟、种族(高加索人发病率较高,亚洲人发病率较低)、职业性暴露和性别(膀胱癌是男性中第四位最常见的癌症,但女性中位居第十一位)。所有膀胱癌中大约三分之二将呈现为浅表性疾病,侵入性疾病存在于其余三分之一中。

[0003] 尽管在浅表性膀胱癌的手术和医疗管理方面有所改进,但在初始处理后大约70%-80%的膀胱癌复发,并且10%-20%的早期疾病将进展至膀胱壁的侵入(H.W.Herr, J.R.Faulkner, H.B.Grossman, R.B.Natale, R.deVere White, M.F.Sarosdy, and E.D.Crawford, “Surgical Factors Influence Bladder Cancer Outcomes: A Cooperative Group Report,” J.Clin.Oncol., vol.22, no.14, pp.2781-2789, Jul.2004); (“Bladder Cancer Treatment (PDQ®),” National Cancer Institute. www.cancer.gov/cancertopics/pdq/treatment/bladder/HealthProfessional/page1.[访问:2014年12月02日]);以及(American Urological Association, “Guideline for the Management of Nonmuscle Invasive Bladder Cancer: (Stages Ta, T1 and Tis: Update (2007).” 2007))。因此,具有通过经尿道切除术处理的浅表疾病的患者经历严格的筛查方案,至少5年中每3-6个月定期进行膀胱镜检查以评估疾病的复发情况,并且此后每年一次(American Urological Association, “Guideline for the Management of Nonmuscle Invasive Bladder Cancer: (Stages Ta, T1 and Tis: Update (2007).” 2007);以及(National Comprehensive Cancer Network, “NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): Bladder Cancer,” Version 2.2014))。通过膀胱镜检查的筛查对患者来说是高度侵入性的,需要通过尿道将显微镜插入膀胱中,并且因此与多达60%患者中的筛查不顺应性相关联(D.Schrag, L.J.Hsieh, F.Rabbani, P.B.Bach, H.Herr, and C.B.Begg, “Adherence to Surveillance Among Patients With Superficial Bladder Cancer,” J.Natl.Cancer Inst., vol.95, no.8, pp.588-597, Apr.2003)。此外,由于这些程序必须由泌尿科医师实行,并且随访频繁且是终生的,膀胱癌管理中所涉及的费用显著,结果是任何实体癌中每名患者的平均终生监管费用最高的癌症(M.F.Botteman, C.L.Pashos, A.Redaeli, B.Laskin, and R.Hauser, “The health economics of bladder cancer: a comprehensive review of the published literature,” PharmacoEconomics, vol.21, no.18, pp.1315-1330, 2003)。缺乏用于检测膀胱癌复发的足够敏感和特异性的基于尿的测定是一种重大的未满足的医疗需求。

发明内容

[0004] 当前实施例通过以下方式来解决以上讨论的需求：发展用于膀胱癌诊断的改进的测定和相关样本保存和处理方法，并且包括用于膀胱癌的敏感和特异性检测的尿核酸测序诊断。基于来自患有膀胱癌的人类患者的尿和肿瘤样本的数据，改进的测定呈现出高价值、具有成本效益的临床诊断测定。

[0005] 大规模的癌症基因组举措已经显著提高了我们对与膀胱癌相关联的基因组事件的理解。利用针对尿路上皮癌的癌症基因组图谱项目的近期完成，可获得膀胱癌中最常见突变的综合列表，包括已知基因，诸如TP53、PIK3CA、RBI和FGFR3 (The Cancer Genome Atlas Research Network, “Comprehensive molecular characterization of urothelial bladder carcinoma,” *Nature*, vol. 507, no. 7492, pp. 315–322, Mar. 2014)。许多其他出版物已经识别了膀胱癌中的附加突变，其中一些突变特别令人感兴趣，因为它们示出早期或低等级肿瘤中存在突变(表1)。例如，已经发现TERT启动子突变在早期侵入性膀胱癌中非常常见 ((C.D.Hurst, F.M.Platt, and M.A.Knowles, “Comprehensive Mutation Analysis of the TERT Promoter in Bladder Cancer and Detection of Mutations in Voided Urine,” *Eur.Urol.*) ; (P.J.Killela, Z.J.Reitman, Y.Jiao, C.Bettegowda, N.Agrawal, L.A.Diaz, A.H.Friedman, H.Friedman, G.L.Gallia, B.C.Giovanella, A.P.Grollman, T.-C.He, Y.He, R.H.Hruban, G.I.Jallo, N.Mandahl, A.K.Meeker, F.Mertens, G.J.Netto, B.A.Rasheed, G.J.Riggins, T.A.Rosenquist, M.Schiffman, I.-M.Shih, D.Theodorescu, M.S.Torbenson, V.E.Velculescu, T.-L.Wang, N.Wentzensen, L.D.Wood, M.Zhang, R.E.McLendon, D.D.Bigner, K.W.Kinzler, B.Vogelstein, N.Papadopoulos, and H.Yan, “TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal,” *Proc.Natl.Acad.Sci.*, vol. 110, no. 15, pp. 6021–6026, Apr. 2013) ; (X.Liu, G.Wu, Y.Shan, C.Hartmann, A.von Deimling, and M.Xing, “Highly prevalent TERT promoter mutations in bladder cancer and glioblastoma,” *Cell Cycle*, vol. 12, no. 10, pp. 1637–1638, May 2013) ; 以及 (I.Kinde, E.Munari, S.F.Faraj, R.H.Hruban, M.Schoenberg, T.Bivalacqua, M.Allaf, S.Springer, Y.Wang, L.A.Diaz, K.W.Kinzler, B.Vogelstein, N.Papadopoulos, and G.J.Netto, “TERT promoter mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine,” *Cancer Res.*, vol. 73, no. 24, pp. 7162–7167, Dec. 2013))。类似地，FGFR3中的突变早已知道在早期是常见的，非侵入性膀胱癌和STAG2突变最近已经被识别为具有类似的模式 ((C.Billerey, D.Chopin, M.H.Aubriot-Lorton, D.Ricol, S.Gil Diez de Medina, B.Van Rhijn, M.P.Bralet, M.A.Lefrere-Belda, J.B.Lahaye, C.C.Abbou, J.Bonaventure, E.S.Zafrani, T.van der Kwast, J.P.Thiery, and F.Radvanyi, “Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors,” *Am.J.Pathol.*, vol. 158, no. 6, pp. 1955–1959, Jun. 2001) ; (C.F.Taylor, F.M.Platt, C.D.Hurst, H.H.Thygesen, and M.A.Knowles, “Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes,” *Hum.Mol.Genet.*, vol. 23, no. 8,

pp.1964-1974, Apr. 2014); (D.A.Solomon, J.-S.Kim, J.Bondaruk, S.F.Shariat, Z.-F.Wang, A.G.Elkahlon, T.Ozawa, J.Gerard, D.Zhuang, S.Zhang, N.Navai, A.Siefker-Radtke, J.J.Phillips, B.D.Robinson, M.A.Rubin, B.Volkmer, R.Hautmann, R.Küfer, P.C.W.Hogendoorn, G.Netto, D.Theodorescu, C.D.James, B.Czerniak, M.Miettinen, and T.Waldman, "Frequent truncating mutations of STAG2 in bladder cancer," Nat.Genet., vol.45, no.12, pp.1428-1430, Dec.2013)). 发明人已经利用这些和类似研究的总和来设计测定, 该测定以高敏感度全面调查低等级和高等级膀胱癌两者中的全部突变谱。

[0006] 在过去几年中, 高敏感的下一代测序(NGS)技术已经作为检查癌症生物标记的强大方法而出现。虽然这些技术常规地准许对肿瘤进行广泛的测序, 以识别群体内5%或更高频率的丰富突变, 但标准方法和机器和测定噪声通常不准许重新识别低于1%-5%的等位基因频率的突变。另外, 大多数肿瘤测序方法依赖于来自患者的匹配正常组织的测序, 以筛选出它们基因组中的SNP(单核苷酸多态性)或非致病性变异。

[0007] 某些当前实施例利用基于DNA的标记的扩展小组改进了这种方法, 该扩展小组更充分地涵盖了膀胱癌的基因组多样性。这些新颖方法所给予的低变异系数和高敏感度准许技术敏感度与用于测量核酸突变的其他高敏感度临床平台相当。鉴于这些敏感度的改进, 发明人已经利用NGS来提供在患者的处理或复发监管过程中, 以个性化方式测量真正的肿瘤本征标记的无双能力。如本文所用的, NGS包括许多不同的现代测序技术, 包括: Illumina (Solexa) 测序, Roche 454测序, Ion torrent:Proton/PGM测序以及SOLiD测序。这些技术允许对DNA和RNA的快速测序(比起先前使用的Sanger测序技术)。

[0008] 发明人已经利用NGS的潜力, 以用于微创检测和癌症监测, 并且使得能够同时揭示驱动癌症的肿瘤抑制基因和启动子基因中的底层异常。这种有价值的洞察允许追踪处理和复发过程中的肿瘤演变, 以及哪些变化可能与进展风险、治疗反应以及复发时间相对应。

[0009] 在示例性实施例中, 分析算法在改进的测定中实施, 改进的测定允许在对患者的原发性肿瘤核酸进行初始评估之后, 纵向监测尿DNA。通过开发能够涵盖膀胱癌的基因组和临床多样性以及最终血尿的增强的靶向生物标记小组, 某些当前实施例的方法提供了高技术性能, 同时实现临床上可行的测定成本和处理时间。方法提供了以可能产生比现有技术高得多的敏感度和特异性的方式来监测膀胱患者的尿的机会, 并且提供了优于现有FDA批准的尿测定的优点。

[0010] 某些当前实施例提供了用于检测与膀胱癌相关联的一个或多个基因中的突变的方法。这些方法涉及从来自受试者的尿样中分离核酸、DNA或RNA, 并且分析核酸以获得核酸序列数据, 核酸序列数据适合于检测与膀胱癌相关联的基因中的一个或多个基因中的一个或多个突变的存在或不存在。任选地, 针对表观基因学标记, 诸如5-甲基胞嘧啶甲基化、CpG岛或核酸结构上的其他变异, 分析分离的核酸。任选地, 分离的核酸是无细胞核酸和从尿样中的细胞分离的核酸。

[0011] 本文中所使用的突变包括(但不限于): 基因或基因部分的缺失或重复, 基因或基因部分的易位或融合, 整个染色体或部分染色体的缺失和重复, 插入缺失或单一点突变。

[0012] 某些当前实施例提供了用于预后受试者中的膀胱癌的方法。实施例涉及确定从受试者的尿中获得的核酸样本中与膀胱癌相关联的至少一个基因中、或源自受试者的基因型

数据集中的至少一个突变或表观基因学改变的存在或不存在,其中,至少一个基因中的至少一个突变或表观基因学改变的存在和/或相对丰度指示膀胱癌预后。

[0013] 某些当前实施例提供了用于诊断受试者中的膀胱癌的方法。实施例涉及确定从受试者的尿中获得的核酸样本中的与膀胱癌相关联的至少一个基因中、或源自该受试者的基因型数据集中的至少一个突变的存在或不存在。在一个示例性实施例中,至少一个基因中的至少一个突变或至少一个表观基因学改变的存在和/或相对丰度指示膀胱癌。在一个示例性实施例中,受试者在他们的尿中呈现血液。在另一示例性实施例中,受试者是无症状的或以其它方式被认为是健康的个体。在另一实施例中,受试者是高风险个体或个体群体(诸如吸烟者、具有职业性致癌物暴露历史的个体、具有饮用受砷或其他可疑致癌物污染的井水或地下水的历史的个体、或生活在地理上癌症热点内的个体)。

[0014] 某些当前实施例确定受试者对膀胱癌的易感性,包括确定在源自个体或受试者的基因型数据集中的至少一个或多个基因中的与膀胱癌相关联的至少一个或多个突变的存在或不存在,其中,确定至少一个突变的存在和/或相对丰度指示受试者中对膀胱癌的易感性增加。

[0015] 在一个示例性实施例中,基因型数据集包括关于个体的等位基因状态的信息,即关于与膀胱癌相关联的突变的、由个体携带的两个等位基因的身份的信息。基因型数据集可以包含关于一个或多个突变或表观基因学标记的等位基因信息,一个或多个突变或表观基因学标记包括:两个或更多突变或表观基因学标记、三个或更多突变或表观基因学标记、五个或更多突变或表观基因学标记、一百个或更多突变或表观基因学标记等。在一些实施例中,基因型数据集包括来自个体的全基因组评估的基因型信息,全基因组评估可包括数十万个突变,或甚至一百万个或更多的突变。

[0016] 在某些实施例中,确定易感性包括将核酸序列数据与数据库进行比较,数据库包含在至少一个突变和/或表观基因学标记与膀胱癌易感性之间的相关性数据。在一些实施例中,数据库包括针对至少一个突变和/或表观基因学标记的膀胱癌易感性的至少一种风险测量。序列数据库可以例如作为查找表被提供,查找表包含指示针对任何一个或多个特定突变和/或表观基因学标记的膀胱癌易感性的数据。

[0017] 某些当前实施例提供了用于监测受试者中膀胱癌进展或膀胱癌复发的方法。该方法涉及在不同时间点从患有癌症的受试者中获得第一和第二尿样,从尿样中分离核酸,和/或分析核酸以获得核酸序列数据,核酸序列数据适合于检测与膀胱癌相关联的基因中的一个或多个基因中的一个或多个突变和/或表观基因学标记的存在或不存在。该方法还包括将第一样本中检测到的一个或多个突变和/或表观基因学标记的存在或不存在与第二样本中检测到的一个或多个突变和/或表观基因学标记的存在或不存在进行比较。通过允许实施算法来比较在不同时间从同一患者连续收集的样本之间的结果,该方法利用独特的优点,从而增强了测定的敏感度和特异性,同时还使针对每位患者的复发监测个性化,并且允许测定区分原发性肿瘤的生物学复发与散开的多病灶疾病的出现。

[0018] 在某些实施例中,至少一个基因中的至少一个突变和/或表观基因学改变选自表1中所列出的突变。

[0019] 某些实施例还提供了计算机实施的方面。在一个这样的方面,一个实施例提供了具有计算机可执行指令的计算机可读介质,该计算机可执行指令用于确定受试者对膀胱癌

的易感性,计算机可读介质包括:表示至少一个突变和/或表观基因学标记的数据;以及存储在计算机可读介质上的例程,并且例程适于由处理器执行,以基于受试者中的至少一个或多个基因的一个或多个突变和/或表观基因学改变来确定个体对膀胱癌的易感性。

[0020] 某些实施例还提供了用于确定受试者中膀胱癌的指示器的装置,该装置包括:处理器,具有计算机可执行指令的计算机可读存储器,该计算机可执行指令适于在处理器上执行,以分析至少一位受试者相关于膀胱癌的突变或基因信息,并且基于突变或基因信息生成输出。输出可包括作为受试者的膀胱癌指示器的至少一个突变和/或表观基因学改变的信息或风险测量。

[0021] 在一个示例性实施例中,计算机可读存储器包括指示诊断患有膀胱癌的多个个体中的至少一个基因的至少一个突变和/或表观基因学改变的频率的数据。存储器还可以包括指示多个参考个体中的至少一个基因的至少一个突变和/或表观基因学改变的频率的数据。风险测量可以基于针对受试者的至少一个突变和/或表观基因学改变和/或基因型数据集状态与针对诊断患有膀胱癌的多个个体的、指示至少一个突变的频率和/或基因型数据集信息的数据的比较。

[0022] 在一个备选实施例中,计算机可读存储器还包括指示与至少一个基因或至少一个基因型数据集的至少一个突变和/或表观基因学改变相关联的发展膀胱癌的风险的数据。受试者的风险测量可以基于受试者的基因型数据集与风险的比较,风险与至少一个基因或至少一个基因型数据集的至少一个突变和/或表观基因学改变相关联。

[0023] 在另一实施例中,计算机可读存储器还包括指示诊断患有膀胱癌的多个个体中的至少一个基因或至少一个基因型数据集的至少一个突变和/或表观基因学改变的频率的数据。存储器还可以包括指示多个参考个体中的至少一个基因或至少一个基因型数据集的至少一个突变和/或表观基因学改变的频率的数据。在此,发展膀胱癌的风险可以是基于诊断患有膀胱癌的个体中与参考个体中的基因型数据集或至少一个突变和/或表观基因学改变的频率的比较。在某个实施例中,至少一个突变或表观基因学改变选自表1中所阐述的那些突变或表观基因学改变。

[0024] 某些实施例还涉及试剂盒。在一个这样的方面,一个实施例涉及用于评估受试者对膀胱癌的易感性的试剂盒,试剂盒包括用于选择性地检测受试者的基因组中与膀胱癌相关联的至少一个基因的至少一个突变和/或表观基因学改变所必需的试剂,其中至少一个突变和/或表观基因学改变的存在指示对膀胱癌的易感性增加。在一个示例性实施例中,试剂盒还包括数据集合,其包括在至少一个突变与膀胱癌易感性之间的相关性数据。相关性数据可以是任何适合的形式,例如作为相对风险测量(RR)、比值比(OR)或技术人员已知的其他方便测量。在一个实施例中,数据的集合在计算机可读介质上。

[0025] 另一方面,一个实施例涉及用于评估受试者对膀胱癌的易感性的试剂盒,试剂盒包括用于选择性地检测受试者基因组中的至少一个基因的至少一个突变和/或表观基因学改变的试剂,其中,选择突变,并且其中,至少一个突变和/或表观基因学改变的存在指示对膀胱癌的易感性。在一个实施例中,至少一个突变和/或表观基因学改变选自表1中所阐述的那些突变和/或表观基因学改变。

[0026] 试剂盒试剂被用于某些实施例中。在一个实施例中,这样的试剂包括至少一个连续寡核苷酸,至少一个连续寡核苷酸与包括至少一个突变的个体的基因组片段杂交。在另

一实施例中,试剂盒包括至少一对寡核苷酸,该至少一对寡核苷酸与从受试者获得的基因组区段的相对链杂交,其中每个寡核苷酸引物对被设计成选择性地扩增包含一个突变和/或表观基因学改变的个体基因组片段。突变和/或表观基因学改变可以选自由表1中所限定的突变和/或表观基因学改变组成的组。在一个示例性实施例中,寡核苷酸与个体的基因组完全互补。在另一示例性实施例中,试剂盒还包含用于扩增片段的缓冲液和酶。在另一示例性实施例中,试剂还包括用于检测片段的标记。

[0027] 根据某些当前实施例的试剂盒也被用于实施例的其他方法中,包括:评估先前诊断患有膀胱癌的受试者中发展至少第二原发性肿瘤的风险的方法,评估受试者对膀胱癌治疗剂响应的可能性的方法,评估受试者的疾病病理阶段或等级进展可能性的方法,以及监测被诊断患有膀胱癌、并被给予疾病处理的受试者的处理进展的方法。

[0028] 根据以下详细描述,某些实施例的其他目的、特征和优点将变得显而易见。然而,应理解的是,当指示具体实施例时,详细描述和具体示例仅通过说明的方式被给出,因为根据该详细的描述,在本发明实施例的精神和范围内的各种变化和修改对于本领域技术人员来说将变得显而易见。

附图说明

[0029] 参考以下描述和所附权利要求以及附图,当前实施例的这些和其他特征、方面和优点将变得更好理解,其中:

[0030] 图1是显示了初步下一代测序小组的矩阵,该测序小组涵盖96%膀胱癌患者的生物多样性。

[0031] 图2描绘了应用错误抑制方法和高效率库转换之前和之后,RAD51基因中的相应噪声水平的图形表示。

[0032] 图3示出了线稀释曲线图,其中,这种线或稀释是使用标准市售方法稀释到另一尿DNA参考样本中的独特尿参考DNA样本。

[0033] 图4示出了线稀释曲线图,其中,这种线或稀释是在所公开的质量控制措施已实施后稀释到参考DNA样本中的独特DNA样本。

[0034] 图5图示了包括基因组库和原始患者测序数据的流算法,基因组库和原始患者测序数据用作到度量生成算法、突变调用算法和临床报告算法中的输入。

[0035] 图6图示了(A)由数据基础设施(B)支持的计算平台,该数据基础设施由专有(B i, ii, iv)和开源(B iii)基因组库两者组成。

[0036] 图7是一系列曲线图,曲线图示出了同一个体内一天的时间中的核酸完整性和变化。

[0037] 图8描绘了一系列患者分布,其描绘独特尿核酸分布。

[0038] 图9呈现了两个曲线图,这两个曲线图示出了两个患者匹配样本中的尿核酸和肿瘤中的等位基因频率之间的关系,其中在肿瘤存在于膀胱中的情况下收集尿。

[0039] 图10呈现了曲线图,该曲线图说明了非癌症和癌症患者的尿核酸中的过滤后突变丰度。

具体实施方式

[0040] 在下面的描述中,广泛地利用了许多术语。为了提供对说明书和权利要求的更清楚和一致的理解(包括向这些术语给出的范围),提供以下定义。

[0041] 当与术语“包括”、“包含”、“具有”或“含有”或它们的其他时态结合使用时,在权利要求和/或说明书中使用词语“一”或“一个”可以意味着“一个”,但也与“一个或多个”、“至少一个”以及“一个或多于一个”或“多个”的含义一致。

[0042] 贯穿本文的书面描述(其包括权利要求),术语“约”用于指示值,该值包括被用于确定该值的设备或方法的误差的标准偏差。

[0043] 在权利要求中使用术语“或”用于意指“和”或“或”(“和/或”),除非明确指出仅指代二择其一或者备选方案是否是相互排斥的,尽管本公开支持仅涉及二择其一以及“和/或”的定义。

[0044] 如在本说明书和权利要求中所使用的,词语“包括”(以及包括的任何形式)、“具有”(以及具有的任何形式)、“包含”(以及包含的任何形式)或“含有”(以及含有的任何形式)是包含性的或开放式的,并且不排除附加的、未记载的元素或方法步骤。

[0045] 还应明确理解的是,本文所记载的任何数值包括从低值到高值的所有值,包括这些值,并且所枚举的最低值与最高值之间的数值的所有可能组合都被认为是在本书面描述(其包括权利要求)中明确陈述。例如,如果范围被陈述为1%至50%,则旨在的是,在说明书和权利要求中明确枚举诸如2%至40%、10%至30%、或1%至3%等的值。

[0046] “接触”指代使至少两个不中的物种接触以使得它们能够反应的处理。然而,应领会的是,所得到的反应产物可以直接由所加入的试剂之间的反应产生,或由来自所加入的试剂中的一种或多种试剂的中间物产生,该中间物可以在反应混合物中产生。

[0047] “单核苷酸多态性”或“SNP”是当基因组中特定位置处的单核苷酸在物种成员之间、或在个体中的配对染色体之间不同时发生的DNA序列变异。大多数SNP多态性具有两个等位基因。在这种情况下,每个个体对于多态性的一个等位基因是纯合的(即,个体的两个染色体拷贝都在SNP位置具有相同的核苷酸),或者个体是杂合的(即,个体的两个姐妹染色体包含不同的核苷酸)。如本文报道的SNP命名是指由国家生物技术信息中心(NCBI)分配给每个独特SNP的官方参考SNP(rs) ID识别标签。

[0048] “核酸”、“寡核苷酸”以及“多核苷酸”是指单链或双链形式的脱氧核糖核酸(DNA)或核糖核酸(RNA)及其聚合物。除非特别限制,否则该术语涵盖包含天然核苷酸的已知类似物的核酸,该已知类似物具有与参考核酸类似的结合特性,并以类似于天然存在的核苷酸的方式代谢。术语核酸可与基因、cDNA和由基因编码的mRNA互换使用。

[0049] 如本文所描述的,术语“易感性”是指:个体倾向于发展某种状态(例如某种特征、表型或疾病,例如膀胱癌),或个体倾向于比一般个体更少地能够抵抗特定状态。该术语涵盖易感性增加和易感性降低两者。因此,如本文所描述的某些实施例的某些基因中的特定突变可能是膀胱癌易感性增加(即风险增加)的特征,其特征在于特定的突变、等位基因或单倍型的相对风险(RR)或比值比(OR)大于1。备选地,某些实施例的突变或其组合是膀胱癌易感性降低(即风险降低)的特征,其特征在于相对风险小于1。

[0050] “插入缺失(indel)”是一种常见的多态性形式,其包括通常只有几个核苷酸长的小插入或缺失。

[0051] “计算机可读介质”是一种信息存储介质,其可以使用市售的或定制化的接口由计算机访问。示例性的计算机可读介质包括存储器(例如RAM、ROM、闪存等)、光存储介质(例如CD-ROM)、磁存储介质(例如,计算机硬盘驱动器、软盘等)、穿孔卡或其他市售介质。信息可以在感兴趣的系统与介质之间,在计算机之间,或者在计算机与计算机可读介质之间传送,以用于存储或访问所存储的信息。这种传输可以是电的,或者通过其他可用的方法,诸如IR链路、无线连接等。

[0052] 词语“受试者”包括人、动物、禽类,例如马、驴、猪、老鼠、仓鼠、猴、鸡、绵羊、牛、山羊、水牛。

[0053] 对“瘤”或“癌症”的参考应理解为对包括瘤或癌细胞的损伤、肿瘤或其他包封或未包封的质量或其他形式的生长的参考。“癌细胞”应理解为对展现异常生长的细胞的参考。术语“生长”应在其最广泛的意义上理解,并且包括对增殖的参考。在这方面,异常细胞生长的一个示例是细胞不受控制的增殖。另一示例是细胞中的失败凋亡,从而延长其通常寿命。瘤细胞可以是良性细胞或恶性细胞。在某个实施例中,主题瘤是膀胱肿瘤。

[0054] 对“DNA区域”的参考应理解为对基因组DNA的特定分段的参考。这些DNA区域通过参考基因名称或一组染色体坐标来指定。基因名称和染色体坐标两者都是本领域技术人员所熟知和理解的。本文所呈现的染色体坐标对应于基因组的Hg19版本。一般地,可以通过参考它的名称来常规识别基因,经由该名称可以常规地获得它的序列和染色体位置两者,或者可以通过参考它的染色体坐标来常规识别基因,经由该染色体坐标也可以常规地获得基因名称和它的序列两者。

[0055] 关于基因/DNA,还应注意以下内容。对本文详述的基因/DNA区域中的每个基因/DNA区域的参考被理解为对所有形式的这些分子及其片段或变异体的参考。如本领域技术人员将领会的,已知一些基因展现出个体之间的等位基因变异或单核苷酸多态性。SNP包括简单序列重复(诸如二核苷酸和三核苷酸重复)和不同大小的插入和缺失。变异体包括来自相对于本文所描述DNA区域共享至少90%、95%、98%、99%序列一致性(即具有一个或多个缺失、添加、取代、反转序列等)的相同区域的核酸序列。因此,应理解某些当前实施例延伸到这样的变异体,就现在的诊断应用而言,尽管事实是实际核酸序列之间的微小基因变异可能存在于个体之间,但变异体实现了相同的结果。因此,当前实施例应被理解为延伸到由任何其他突变、多态性或等位基因变异所产出的所有形式的DNA。

[0056] 如本文所描述的癌症诊断指的是对癌症状态的性质进行确定或分类,例如癌症或肿瘤的突变或基因表型、与其进展相关联的癌症的临床阶段和/或癌症的转移性质。如本文所描述的,基于基因表型的癌症诊断可以帮助指导适当的治疗干预。

[0057] 如本文所描述的癌症预后包括确定癌症状况的可能进展和过程,并且确定患有癌症的受试者的恢复和存活机会,例如有利的预后指示癌症患者恢复和/或存活的可能性增加,而不利的预后指示癌症患者恢复和/或存活的可能性降低。受试者的预后可以通过适合处理(即,将增加患有癌症的受试者的恢复和存活的可能性的处理)的可用性来确定。某些当前实施例的这个方面还可以包括:基于确定的预后来选择适合的癌症治疗剂,并将所选择的治疗剂施用于受试者。

[0058] 预后还涵盖癌症的转移潜能。例如,基于存在或不存在基因表型的有利预后可以指示该癌症是具有低转移潜能的癌症类型,并且患者的长期恢复和/或存活的可能性增加。

备选地,基于存在或不存在基因表型的不利预后可以指示癌症是具有高转移潜能的癌症类型,并且患者的长期恢复和/或存活的可能性降低。预后部分地由病理等级和阶段来评估。其中,基于美国癌症联合委员会制定的标准,等级被定义为乳头状瘤或低等级或高等级。其中,阶段由肿瘤、结节、转移(TNM)分期系统来定义。例如,肿瘤阶段可以被定义为T、T0、Ta、Tis、T1、T2、T2a、T2b、T3、T3a、T3b、T4a、T4b。例如,结节阶段可以被定义为NX、N0、N1、N2、N3。例如,转移阶段可以被定义为M0、M1。在一个实施例中,基因组表型或一个或多个突变或表观基因学改变的组合可以与含有基因组表型和分期信息的数据库进行比较,并且其中该比较通过计算测量尿基因组表型与数据库中具有已知阶段、等级和患者结果的其他肿瘤的相似性,来粗略估计肿瘤阶段和等级。

[0059] 某些当前实施例的另一方面涉及识别存在的膀胱癌的类型。膀胱癌可以定义为移行细胞型或尿路上皮癌、鳞状细胞膀胱癌、膀胱腺癌、膀胱肉瘤、膀胱小细胞癌。在当前实施例的一个方面,可以将基因组表型或一个或多个突变或表观基因学改变的组合或一个或多个突变或表观基因学改变与包含基因组表型并定义癌细胞类型信息的数据库进行比较,并且其中该比较通过计算测量与数据库中具有已知细胞类型的其他肿瘤的尿基因组表型相似来,粗略估计肿瘤细胞类型。在当前实施例的另一方面,基因组表型或一个或多个表观基因学改变的组合可用于生成计算机模型,该计算机模型粗略估计肿瘤微环境、以及可调制癌细胞活性和生物性的非癌细胞的相对丰度。

[0060] 某些当前实施例的另一方面涉及监测受试者的癌症进展的方法,其涉及在不同时间点从患有癌症的受试者获得含有核酸的第一和第二尿样。样本中的核酸与一个或多个试剂接触,该试剂适用于检测与膀胱癌相关联的一个或多个基因中一个或多个突变和/或表观基因学改变的存在或不存在,并且检测与膀胱癌相关联的一个或多个基因中的一个或多个突变和/或表观基因学改变的存在或不存在。该方法进一步涉及将在第一尿样核酸中所检测到的一个或多个突变和/或表观基因学改变的存在或不存在与在第二尿样核酸中所检测到的一个或多个突变和/或表观基因学改变的存在或不存进行比较,并且基于该比较来监测受试者中的癌症进展。

[0061] 与膀胱癌有关的一个或多个基因的突变和/或表观基因学改变状态的改变(例如,检测到第二尿样中突变和/或表观基因学改变的存在,而未在第一尿样中检测到突变和/或表观基因学改变)指示癌症表型的改变已经随着疾病进展而发生。这种改变可能具有治疗意义,即它可以用信号指示需要改变受试者的处理过程。改变还可以指示癌症向转移表型的进展。因此,尿核酸突变和/或表观基因学状态的定期监测提供了用于检测原发性肿瘤进展、转移、并促进对癌症状况的最佳靶向或个性化处理的手段。

[0062] 如由医生确定为适合的,并且基于原发肿瘤的特征(肿瘤类型、阶段、位置等),获得第一尿核酸样本与第二或任何附加的随后尿核酸样本之间的时间可以是任何期望的时间段,例如数周、数月、数年。在该方面的一个实施例中,第一样本在处理前获得,并且第二样本在处理后的获得。备选地,两个样本都可以在一次或多次处理后获得;在晚于第一样本的某个时间点获得第二样本。备选地,可以在疾病存在之前获得一个或多个样本。

[0063] 几种基因中的突变和/或表观基因学改变已经证明与膀胱癌相关联。表1示出了从其选择用于测定与膀胱癌相关的突变和/或表观基因学改变的基因列表。突变可以包括插入、缺失、重复、扩增和易位。表观基因学特征可以包括胞嘧啶核苷酸的甲基化。发现与膀胱

癌相关联的其他基因也可以基于经验验证而被用于当前实施例中。使用单独合成的DNA或RNA杂交探针允许杂交捕获库的模块化和基于经验验证的迭代优化(去除/添加探针)。捕获探针的特异性可以在探针设计期间、但还在测序验证期间以计算方式解决。在CLIA实验室设置中,用于验证非决定性或意想不到的结果的示例性方法是用基于二级PCR扩增子的富集方法的补充杂交捕获,以提供不服从杂交捕获的区域的覆盖并确认新颖结果。大量并行扩增系统(诸如RainDance、AmpliSeq和Wafergen)为扩增子库制备提供了高效率 and 一致性。

[0064] 用于从尿中分离细胞、分离尿中的无细胞核酸、以及从尿中发现的细胞中分离核酸的任何已知方法以其相应整体并入本文。尿保存缓冲液可以包含以下种类的试剂:微生物静电剂(诸如EDTA)、异噻唑啉酮和/或其衍生物(诸如甲基异噻唑啉酮)、抗生素、pH缓冲试剂(诸如Tris盐)、DNase/RNase抑制剂(诸如EDTA和金精三羧酸),包括离液盐的核酸水合的改性剂(诸如硫氰酸胍)、醋酸铵、醋酸钠、十二烷基硫酸钠。在一方面,尿保存缓冲液结果指示在室温下保存DNA至少1周。根据本领域的知识和技能,可以使用其他缓冲液。在一个实施例中,优化缓冲液和试剂以避免盐的共沉淀,其抑制许多基于酶的反应(诸如PCR或连接),但同时最大化样本的高产率。

[0065] 癌症标记可以在尿内的细胞相关联核酸和无细胞核酸两者中被识别。如图8和图9中所示,尿中核酸的分布以及癌症核酸标记的相对丰度在这两个群体之间变化,并且可以取决于个体患者分布。归因于患者变异性,在一起检查这两种核酸群体时存在优势。如图8所示,这些核酸的大小分布也可以在无细胞核酸的情况下,通常在50-200bp大小范围内变化,而细胞相关联核酸通常大于1,000bp。一些患者显示出尿核酸大小的广泛范围和变异性,而另一些患者主要包含一个部分(超过另一部分)。DNA片段化分布的变异性不仅由收集和储存条件所引起,而且患者的多样化生理机能也是一个因素。除了收集的一天中的时间之外,某些个体似乎对一种尿分布类型具有天然偏置(超过另一种尿分布类型)。如图8所示,已经开发了三种类型的患者分布:“主要经肾”、“混合型”和主要泌尿学追踪。在以主要小型/经肾型分布为特征的个体中,为了获得具有改善的核酸分布的样本,还最好的是,当利用膀胱的尿孵育已经最大化(清晨)并且在其他情况下收集样本,立即将尿排入保存缓冲液中,该保存缓冲液抑制核酸酶活性以防止核酸降解。

[0066] 根据本发明的实施例,这些相应的患者分布使用核酸数据来确定,被归类,并且然后与健康患者和先前确定的以患有膀胱癌为特征的患者分布两者的对照组进行比较。

[0067] 图9描绘了点状图,其表示患者匹配样本中的肿瘤与尿核酸中的等位基因频率之间的关系,其中在肿瘤存在于膀胱中时收集尿。竖直轴是非参考等位基因频率,并且水平轴是靶向基因组区域内的基因组位置,点表示样本类型,并且在图说明中描述。患者A示出了一些患者在肿瘤(42%-71%范围)和尿(38%-60%)之间具有高的等位基因频率一致性,其中尿中的大部分核酸具有肿瘤起源。相反,患者B示出了另一种情况,其中尿中肿瘤来源的核酸丰度低得多,并且肿瘤(26%-51%)和尿(0.3%-2.2%)突变丰度不一致,而对于那些位置(灰色X),尿突变仍然保持丰度高于参考数据库范围。患者A和B两者都表明肿瘤和尿样本两者内不同等位基因频率群集的附加特征。在一个实施例中,等位基因频率群集的程度或类型可以用作诊断或预后疾病算法的一部分。因此,图8表明不同的患者可以在尿中且以显著不同的方式从根本上显示出膀胱癌指示,其中一些患者可以提供许多强信号,而另一些患者仅可以提供很少的检测到的标记。根据本发明的实施例,独特的患者分布数据相对于

这样的患者分布被捕获,并且被分析,并且然后结合随后的样本收集、制备以及正在进行的患者测试分析来使用。因此,随后的检测、诊断和预后可以照顾到个体患者分布,并且可以收集并分析展现类似分布的类似患者群体的数据,以用于随后跨广泛群体的处理结果分析。

[0068] 现在参考图3和图4,在具有和不具有质量控制度量的情况下,描绘了图示针对参考样本的迭代稀释水平绘制的SNP频率的曲线图。检测尿中正常核酸背景中稀释的癌症核酸的能力取决于库制备效率,库制备效率部分地由一系列样本质量控制来调制。为了确定样本质量控制对测序性能的影响,将具有已知变异体的参考样本稀释到背景参考样本中。如所指示的(水平轴),通过体积系列稀释进行DNA样本的稀释,生成测序库,计算测序的样本以及已知单核苷酸变异体的等位基因频率(垂直轴)。每条线上的点表示技术重复的平均表现,误差条是平均值的标准误差。在三个稀释系列中的两个中,单核苷酸变异体的测量的等位基因频率远低于理论预期的稀释信号。在开发增强样本质量控制方法之后,产生了新的稀释系列,其中使用质量控制结果将核酸输入标准化。这些增强的方法导致测序库制备效率与预期的稀释信号相匹配。

[0069] 当库制备效率差时(连接效率差归因于DNA大小、DNA超载、或存在端部修复、A拖尾和连接酶抑制剂、或存在被测量但不能连接的单链DNA),或当杂交捕获效率差时(归因于非人类核酸),预期如图3所示表现的样本,其中样本中的变异体被检测为远远低于基于我们认为我们投入的DNA量所预期的变异体。在图4中,在实施QC后,确保输入足量的较高分子量DNA(毛细管电泳和实时PCR),DNA具有人类起源(实时PCR,硝酸盐),DNA是双链的(毛细管电泳和荧光测定),DNA是功能性的并且具有可扩增的质量(实时PCR)。当所有这些都合在一起时,然后我们的测序性能可以看起来像图4那样,其中性能如理论上预期的那样。

[0070] 质量控制特征的性质和范围部分地取决于样本的性质。例如,尿中的硝酸盐指示可能存在高细菌水平。当从尿中提取的核酸中的细菌DNA丰富时,它具有破坏杂交捕获效率的能力。杂交捕获的这种破坏部分地归因于如下事实:大多数核酸量化技术不区分人类和非人类DNA(紫外吸光度、荧光测定和毛细管电泳全都未区分人类核酸与非人类核酸)。被设计为丰富人类基因的高效杂交捕获取决于准确的到反应中的前期DNA输入,其中该定义的输入是人类起源的。阳性硝酸盐结果可以作为实验室方案中的标志,并且指示需要附加的质量控制,其中使用PCR来量化人类DNA与非人类DNA的丰度,从而可以将足够的人类DNA加载到库制备反应中。在一些情况下,非人类DNA可能达到如下丰度水平:不管人类/非人类标准化,它开始超载或主动抑制库制备(端部修复、A拖尾、连接或杂交捕获反应)。在这种情况下,采取步骤来在库制备之前主动破坏或消耗非人类序列(这可以通过如下方式来执行:利用靶向细菌特异性序列基序的限制酶的处理、差异性核酸甲基化模式(例如甲基-CpG结合域,被描述在<http://dx.doi.org/10.1371/journal.pone.0076096>中)、利用非离子表面活性剂(诸如0.025%的皂角苷,如<http://jcm.asm.org/content/early/2016/01/07/JCM.03050-15>中所描述的)的处理)。

[0071] 因此,即使当样本被认为尿路感染阴性或通过尿化学对硝酸盐阴性时,对于非人类序列对库和富集效率的影响以及效率的其他细微降低两者的质量控制措施可以改善测定性能。在这方面,即使在“健康”和“正常”尿样中,来自正常微生物组的细菌和酵母水平也可能足以影响测序效率(见图3和图4)。此外,传统的现有技术生物标记和尿路感染(硝酸

盐,尿培养)的定义不如我们的基因组方法那样敏感,这可能导致早期尿路感染测试与非人类核酸水平之间的差异。由于这个原因,可以实施更敏感的基于PCR的方法来区分人类核酸和非人类核酸。在实施例中,所使用的PCR反应因此针对特定于人、细菌、酵母或病毒核酸序列的序列而设计。基于对膀胱癌中拷贝数变异数据的分析,来选择下面列出的基因,以选择拷贝数中性的基因组区域。备选地,分析ALU元件也可避免影响基因组等同物近似的基因组拷贝数。在一个实施例中,执行人类特异性PCR,其中针对ALU元件序列和/或选自以下列表中的基因中的一个或多个基因来设计反应引物:CTIF、MRO、STYX、TIMM9、PIGH、WRB、AIRE、MDFIC、PON3、ERMN和RND3。还选择位于基因组区域的PCR序列,该基因组区域不随正常或癌症相关联的基因组拷贝数变化而变化,以允许更直接地量化存在于核酸提取物中的基因组等同物的数量。

[0072] 附加的质量控制步骤涉及尿化学,尿化学包括:pH、血红蛋白、肌红蛋白、酮、尿胆素原和比重的水平。对这些标记进行测试,然后用于突变调用算法的标准化。这些分析物可以修改核酸的化学结构,从而在测序中引入错误。经验参考库(被表示在算法流程图中)的一个方面是使用来自具有这些异常的许多样本的测序数据,来针对这些分析物的不同分析范围而构建测序错误模式分布。然后可以使用这些错误模型来然后减少测序错误并纠正测序结果中潜在的假阳性信号。

[0073] 白细胞酯酶是尿中白血细胞(WBC)的标记。在具有高水平的白血细胞的尿样中,肿瘤标志可能被存在于这些细胞中的正常DNA稀释。本发明的实施例涉及两种方法来校正高的WBC:(1)在尿提取之前的主动消耗(方法的示例包括通过差速离心或暴露于溶质梯度的分离,通过盐溶液处理的差异溶解,使用细胞表面标记以通过抗体下拉或柱来消耗),和/或(2)算法阈值的调整以考虑针对非癌症DNA的水平上升。

[0074] 比重和肌酐值可用作肾功能和尿稀释的替代。在一些情况下,这些标记可以粗略估计全身(经肾)核酸相对于泌尿道核酸的水平。这些标记还可以告知大小分布如何关联于全身与泌尿道核酸。在实施例中,测试值,创建参考库,以及可以适当调整算法。比重和pH值可能与尿样中存在的双链DNA与单链DNA的水平相关。

[0075] 在一个实施例中,从尿中进行总核酸处理和提取的方法包括:

[0076] (i) 尿在溶解溶液(lysis solution)中的孵育的步骤。这样的溶液可以任选地包含去污剂、盐(例如5M NaCl)、离液盐(例如,硫氰酸胍、醋酸钠)、蛋白质消化酶(诸如蛋白酶K)、以及异丙醇或乙醇;

[0077] (ii) 添加核酸结合底物的步骤,核酸结合底物诸如为二氧化硅树脂浆(Norgen Urine DNA试剂盒)、或带负电荷的磁性核酸结合珠(诸如Invitrogen MagMax总核酸试剂盒)、或硅化柱(诸如Qiagen QIAprep Spin Miniprep试剂盒);

[0078] (iii) 用溶解溶液洗涤结合的DNA的步骤;

[0079] (iv) DNA在缓冲溶液中洗脱的步骤,例如,缓冲溶液包含Tris和EDTA;以及

[0080] (v) 将RNA转化和标签化/条形码成cDNA的任选步骤。

[0081] 该最终可选步骤可以通过本领域已知的任何方法完成。例如,使用ClonTech的Smarter(RNA模板的5'端部处的切换机制)的cDNA转化试剂盒。该技术允许在第一链合成期间在cDNA的两个端部处高效并入已知序列,而无需适配体连接。这些已知序列的存在对于下游应用至关重要,其中DNA和RNA衍生的cDNA(由SMARTER试剂盒生成)在相同库中制备并

在单次测序运行内一起测序。将DNA和RNA两者包含在单个库内允许从RNA/cDNA识别基因组易位,而可以从DNA或RNA识别突变和表观基因组学改变。并入已知序列的SMARTER允许DNA和RNA独特信号的下流信息解卷积。

[0082] 在一个实施例中,提取的核酸是DNA。在另一个实施例中,提取的核酸是RNA。在某些实施例中,RNA被逆转录为互补的DNA。这种逆转录可以单独进行或与扩增步骤组合进行,例如使用逆转录聚合酶链式反应(RT-PCR),其可以进一步被修改成量化的,例如美国专利No.5,639,606中描述的量化RT-PCR,其全文通过引用并入本文中。

[0083] 在一个实施例中,提取的核酸(包括DNA和/或RNA)不经扩增步骤而直接被分析。直接分析可以用不同的方法进行,包括但不限于nanosting技术。NanoString技术可以通过将颜色编码的荧光报告者连接到每个目标分子上,使得能够识别和量化生物样本中的单个目标分子。这种方法类似于通过扫描条形码来测量库存的概念。报告者可以使用数百甚至数千个不同的代码来制成,从而允许高度复用的分析。该技术被描述于Geiss等人的出版物“Direct Multiplexed Measurement of Gene Expression with Color-Coded Probe Pairs,”*Nat Biotechnol* 26(3):317-25(2008),其全文通过引用并入本文中。

[0084] 在另一实施例中,可能是有益的或另外期望的是,在分析核酸之前,扩增核酸以富集已知的膀胱癌基因。核酸扩增的方法是本领域中通常已知和常用的。如果需要,可以执行扩增以使其是量化的。量化扩增将允许量化确定各种核酸的相对量。通过PCR、乳液PCR、大量复用PCR、等位基因特异性PCR、分子倒置探针、跟随有环化的位点特异性探针的结合和片段化、或杂交捕获,膀胱癌基因的富集可以发生。特定实施例使用杂交捕获,其中适配体连接的DNA库被孵育有:1.与适配体序列互补的寡核苷酸(阻断寡核苷酸),2.针对DNA杂交优化的缓冲液(Illumina Nextera),以及3.一组与兴趣基因组区域互补的生物素化定制合成的寡核苷酸(Nextera Custom Capture或IDT XGen锁定探针)。

[0085] 核酸扩增方法包括但不限于聚合酶链式反应(PCR)(美国专利No.5,219,727,其全文通过引用并入本文中)及其变体,诸如原位聚合酶链式反应(美国专利No.5,538,871,其全文通过引用并入本文中)、量化聚合酶链式反应(美国专利No.5,219,727,其全文通过引用并入本文中)、巢式聚合酶链式反应(美国专利No.5,556,773)、自持续序列复制及其变体(Guatelli et al.“Isothermal, In vitro Amplification of Nucleic Acids by a Multienzyme Reaction Modeled after Retroviral Replication,”*Proc Natl Acad Sci USA* 87(5):1874-8(1990),其全文通过引用并入本文中)、转录扩增系统及其变体(Kwoh et al.“Transcription-based Amplification System and Detection of Amplified Human Immunodeficiency Virus type 1 with a Bead-Based Sandwich Hybridization Format,”*Proc Natl Acad Sci USA* 86(4):1173-7(1989),其全文通过引用并入本文中)、Qb复制酶及其变体(Miele et al.“Autocatalytic Replication of a Recombinant RNA.”*J Mol Biol* 171(3):281-95(1983),其全文通过引用并入本文中)、冷PCR(Li et al.“Replacing PCR with COLD-PCR Enriches Variant DNA Sequences and Redefines the Sensitivity of Genetic Testing.”*Nat Med* 14(5):579-84(2008),其全文通过引用并入本文中)、或任何其他核酸扩增方法,其后是使用本领域技术人员已知的技术检测扩增的分子。特别有用的是那些设计用于检测核酸分子的检测方案(如果这种分子存在的数目非常少)。

[0086] 检测来自受试者的肿瘤或源自尿的核酸样本中的膀胱癌基因中的一个或多个突变和/或表观基因学改变的存在或不存在的存在,可以使用本领域熟知的方法来执行。

[0087] 在一个实施例中,使用杂交测定来检测一个或多个识别的基因中的一个或多个突变。在杂交测定中,基于一个或多个等位基因特异性寡核苷酸探针与来自受试者的DNA样本中的一个或多个核酸分子的杂交,来确定基因突变的存在或不存在的存在。一个或多个寡核苷酸探针包含至少与含有感兴趣突变的基因区域互补的核苷酸序列。寡核苷酸探针被设计为与野生型、非突变型核苷酸序列和/或一个或多个基因的突变型核苷酸序列互补,以实现在使样本与寡核苷酸探针接触时,检测来自受试者的样本中的突变的存在或不存在的存在。本领域已知的各种杂交测定适用于当前实施例的方法。这些方法包括但不限于直接杂交测定,诸如northern印迹或Southern印迹(参见例如Ausabel et al.,Current Protocols in Molecular Biology,John Wiley&Sons,NY(1991))。

[0088] 备选地,可以使用基于阵列的方法进行直接杂交,其中将设计为与特定非突变或突变基因区域互补的一系列寡核苷酸探针附加到固体支持物(玻璃、硅、尼龙膜)上。使来自受试者的标记的DNA或cDNA样本与含有寡核苷酸探针的阵列接触,并且检测来自样本的核酸分子与阵列表面上的其互补寡核苷酸探针的杂交。直接杂交阵列平台的示例包括但不限于Affymetrix GeneChip或SNP阵列和Illumina's Bead Array。

[0089] 在另一个实施例中,将样本结合到固体支持物(通常是DNA或PCR扩增的DNA),并在溶液中用寡核苷酸标记(等位基因特异的或者短的,以便允许通过杂交进行测序)。

[0090] 检测特定突变可以通过本领域已知的用于检测特定位点处的序列的方法来完成。例如,基于荧光的技术(Chen,X.et al.,Genome Res.9(5):492-98(1999)),利用PCR,LCR,巢式PCR和其他用于核酸扩增的技术。可用的特定商业方法包括但不限于:TaqMan基因分型测定和SNPlex平台(Applied Biosystems),凝胶电泳(Applied Biosystems),质谱法(例如来自Sequenom的MassARRAY系统),微测序方法,实时PCR,Bio-Plex系统(BioRad),CEQ和SNPstream系统(Beckman),阵列杂交技术(例如Affymetrix GeneChip;Perlegen),BeadArray技术(例如Illumina GoldenGate和Infinium测定),阵列标签技术(例如Parallele)和基于核酸内切酶的荧光杂交技术(Invader;Third Wave)。一些可用的阵列平台(包括Affymetrix SNP Array 6.0和Illumina CNV370-Duo以及1M BeadChips)包括标签化某些CNV的SNP。这允许经由这些平台中包括的替代SNP来检测拷贝数变异(CNV)。因此,通过使用本领域技术人员可用的这些或其他方法,可以识别一个或多个突变和/或表观基因学改变。

[0091] 在某些实施例中,通过测序技术检测基因中的突变。获取关于个体的序列信息识别在序列的上下文中的特定核苷酸。对于SNP,关于单个独特序列位点的序列信息足以识别该特定SNP处的等位基因。对于包括多于一个核苷酸的标记,关于包含多态性位点的个体的核苷酸的序列信息识别针对特定位点的个体的等位基因。序列信息可以从来自受试者或个体的尿的核酸样本获得。

[0092] 用于获得核酸序列的各种方法是技术人员已知的,并且所有这些方法都可用于实践实施例。Sanger测序是用于生成核酸序列信息的众所周知的方法。近来已经开发了用于获得大量序列数据的方法,并且这些方法也被认为对于获得序列信息是有用的。这些包括:焦磷酸测序技术(Ronaghi,M.et al.Anal Biochem 267:65-71(1999);Ronaghi,et al.,

Biotechniques 25:876-878 (1998)), 例如, 454 焦磷酸测序 (Nyren, P., et al. Anal Biochem 208:171-175 (1993)), Illumina/Solexa 测序技术 (www.illumina.com; 也参见 Strausberg, R L, et al. Drug Disc Today 13:569-577 (2008)) 和支持寡核苷酸连接和检测平台 (SOLiD) 技术 (Applied Biosystems, www.appliedbiosystems.com); Strausberg, R L, et al. Drug Disc Today 13:569-577 (2008)。前述内容通过引用以其各自的整体并入。

[0093] 其他常见的基因分型方法包括但不限于: 限制性片段长度多态性分析; 基于扩增的测定, 诸如分子信标测定, 核酸阵列, 高分辨率熔解曲线分析 (Reed and Wittwer, "Sensitivity and Specificity of Single-Nucleotide Polymorphism Scanning by High Resolution Melting Analysis," Clinical Chem 50 (10):1748-54 (2004), 其全文通过引用并入本文中); 等位基因特异性 PCR (Gaudet et al., "Allele-Specific PCR in SNP Genotyping," Methods Mol Biol 578:415-24 (2009), 其全文通过引用并入本文中); 引物延伸测定, 诸如等位基因特异性引物延伸 (例如 Illumina™ Infinium™ 测定), 阵列化引物延伸 (参见 Krjutskov et al., "Development of a Single Tube 640-plex Genotyping Method for Detection of Nucleic Acid Variations on Microarrays," Nucleic Acids Res. 36 (12) e75 (2008), 其全文通过引用并入本文中), 均匀引物延伸测定, 通过质谱检测的引物延伸 (例如 Sequenom™ iPT EX SNP 基因分型测定) (参见 Zheng et al., "Cumulative Association of Five Genetic Variants with Prostate Cancer," N. Eng. J. Med. 358 (9):910-919 (2008), 其全文通过引用并入本文中), 在基因阵列上分选的复用引物延伸; 瓣状核酸内切酶测定 (例如 Invader™ 测定) (参见 Olivier M., "The Invader Assay for SNP Genotyping," Mutat. Res. 573 (1-2) 103-10 (2005), 其全文通过引用并入本文中); 5' 核酸酶测定, 诸如 TaqMan™ 测定 (参见 Gelfand 等人的美国专利 No. 5,210,015 和 Livak 等人的美国专利 No. 5,538,848, 它们的全文通过引用并入本文中); 以及寡核苷酸连接测定, 诸如用滚环扩增的连接, 均匀连接, OLA (参见 Landgren 等人的美国专利 No. 4,988,617, 其全文通过引用并入本文中), 复用连接反应, 然后是 PCR, 其中将邮编 (zipcodes) 并入连接反应探针中, 并且扩增的 PCR 产物通过电泳或通用的邮编 (zipcode) 阵列读出来确定 (参见 Barany 等人的美国专利 Nos. 7,429,453 和 7,312,039, 他们的全文通过引用并入本文中)。这些方法可以与检测机制结合使用, 检测机制诸如为例如发光或化学发光检测、荧光检测、时间分辨荧光检测、荧光共振能量转移、荧光偏振、质谱和电检测。一般地, 用于分析基因畸变的方法在许多出版物中报道, 不限于本文引用的那些, 并且对于本领域技术人员是可用的。适当的分析方法将取决于分析的具体目标, 患者的病情/病史, 以及待检测、监测或处理的特定的癌症、疾病或其他医学状况。

[0094] 备选地, 通过对来自患者样本的基因或在一个实施例中包括一个或多个识别的突变的特定基因区的直接测序, 可检测上文识别的一个或多个突变的存在或不存在。直接测序测定通常涉及使用本领域已知的任何合适方法从受试者分离 DNA 样本, 并且将待测序的兴趣区域克隆到合适的载体中, 以用于通过在宿主细胞 (例如细菌) 中生长进行扩增或通过 PCR 的直接扩增或其他扩增测定。扩增后, 可以使用任何合适的方法对 DNA 进行测序。由于某些测序方法涉及高吞吐量下一代测序 (NGS), 以识别基因变异。各种 NGS 测序化学品是可用的, 并且适用于施行实施例, 包括焦磷酸测序 (Roche™ 454), 通过可逆染料终止子的测序 (Illumina™ HiSeq, Genome Analyzer 和 MiSeq 系统), 通过寡核苷酸探针的序列连接的测序

(Life Technologies™ SOLiD) 和氢离子半导体测序 (Life Technologies™, Ion Torrent™)。备选地,可以使用本领域技术人员熟知的经典测序方法 (诸如Sanger链终止方法或Maxam-Gilbert测序) 来施行当前实施例的方法。

[0095] 某些当前实施例还提供了用于施行本文所阐述的公开内容有用的试剂盒。本试剂盒包括一个或多个包含上述测定组分的容器装置。试剂盒还包括其他容器装置,其包含施行实施例所必需或方便的溶液。容器装置可以由玻璃、塑料或箔制成,并且可以是小瓶、瓶子、囊、管、袋等。试剂盒还可以包含书面信息,诸如用于施行某些当前实施例的程序或分析信息 (诸如包含在第一容器装置中的试剂的量)。容器装置可以与书面信息一起位于另一容器装置 (例如,盒或袋) 中。

[0096] 包括以下示例以说明其某些实施例。本领域技术人员应该领会,在下面的示例中所公开的技术表示由发明人发现并认为在实施例的实践中良好地起作用的技术,并且因此可以被认为构成用于其实践的某些模式。然而,根据本公开内容,本领域技术人员应该领会,可以在所公开的具体实施例中做出许多改变,并且仍然获得相同或类似的结果,而不脱离所描述的精神和范围。

[0097] 本文所引用的所有文档作为整体通过对其的引入并入本文。

[0098] 在下面的示例中,使用以下材料和方法。

[0099] 示例1

[0100] DNA修复和测序适配体连接

[0101] 1. 通过利用一种或多种以下酶的处理来修复DNA链缺口或裂缝:Taq DNA连接酶,核酸内切酶IV,Bst DNA聚合酶,Fpg,尿嘧啶DNA糖基化酶 (UDG),T4 PDG (T4核酸内切酶V) 和核酸内切酶VIII,多核苷酸激酶,哺乳动物DNA聚合酶β和/或DNA连接酶I。

[0102] 2. 通过利用一种或多种以下酶处理DNA,来对DNA端部进行修复和A拖尾化 (A-tailing):T4 DNA聚合酶和Klenow片段。

[0103] 3. 测序适配体和核酸插入物的T4连接,其中适配体是Illumina TruSeq型适配体或等同物。在一个实施例中,适配体在适配体的双链茎部分中包含8-碱基对样本条形码,并且相同的条形码存在于p5端部和p7端部两者上。在这样的实施例中,使用匹配的双索引条形码,来避免池化样本之间的低频适配体污染或适配体交换/跳跃。适配体还可以在适配体的茎部或y部分中包含定义或随机序列的多样性库。并且其中这些定义或随机序列部分地被用于在库扩增之前对个体分子进行标签化。

[0104] 4. 或者备选地代替步骤2和步骤3:如Nature Protocols 8,737-748 (2013) 中所描述的,将核酸插入物连续地连接至单链适配体分子。简言之,用磷酸酶处理DNA,以从DNA链的5' 和3' 端部除去残余的磷酸基团。使用CircLigase II,将5' -磷酸化的适配体寡核苷酸和长的3' -生物素化间隔臂连接到DNA链的3' 端。将适配体连接的分子以及过量适配体分子固定在抗生蛋白链菌素珠上,并使用与适配体互补的引物来复制模板链。该反应使用Bst聚合酶2.0执行。在使用T4 DNA聚合酶去除3' 突出端后,通过利用T4 DNA连接酶进行平端连接,将第二适配体接合到新合成的链上。为了防止适配体之间的连接,只有一个适配体链可连接,而另一个适配体被3' -末端双脱氧修饰所阻断。洗去过量的适配体后,通过热变性将库分子从珠上释放出来。

[0105] 5. 设计适配体序列 (步骤3或4中使用的) 以包括位于适配体序列 (长度在6-10个核

核苷酸之间)内的特定数目的DNA碱基,适配体序列为简并或随机序列,或其中6-10个核苷酸序列是许多(50-200个独特)定义的序列之一。并且其中具有不同定义或简并序列的这些适配体存在于相同的混合物内,以便创建独特的适配体序列的多样性库。并且其中随后使用这些独特序列(与诸如DNA插入起始和终止位点之类的其他变量组合),以在PCR扩增适配体连接的插入物分子的多样性群体后,唯一地识别插入分子的克隆起源。

[0106] 示例2

[0107] (已知膀胱癌基因的富集)

[0108] 通过PCR、乳液PCR、大量复用PCR、等位基因特异性PCR、分子倒置探针、其后是环化的位点特异性探针的结合和片段化、或杂交捕获,可以发生膀胱癌基因的富集。

[0109] 一个实施例使用杂交捕获,其中适配体连接的DNA库孵育有:1.与适配体序列互补的寡核苷酸(阻断寡核苷酸),2.针对DNA杂交优化的缓冲液(Illumina Nextera),以及3.一组与兴趣基因组区域互补的生物素化定制合成的寡核苷酸(Nextera Custom Capture或IDT XGen锁定探针)。

[0110] 在各种温度下进行一系列孵育以促进寡核苷酸与其靶序列的杂交。

[0111] 利用抗生蛋白链菌素珠孵育杂交反应,以从溶液中富集结合的寡核苷酸。从珠子洗涤和洗脱结合的寡核苷酸。

[0112] 利用富集分级和定制寡核苷酸进行第二次重复的杂交捕获反应,以进一步富集兴趣目标。

[0113] 利用抗生蛋白链菌素珠捕获结合的寡核苷酸,洗涤并从珠上洗脱。

[0114] 将富集样本加载到测序仪上。

[0115] 示例3

[0116] (数据分析方法和结果的利用和解释)

[0117] 1.基于已知序列的DNA和cDNA序列的解卷积。

[0118] 2.将DNA和cDNA读数映射到参考基因组。

[0119] 3.使用简并或定义的适配体序列的独特对(两者都在分子的5'端和3'端上)以及DNA插入物的起始/终止位点,识别分子克隆家族。

[0120] 4.在克隆家族内,针对碱基对调用差异,比较测序读数。

[0121] 5.通过表决过程来过滤或纠正个体克隆内的差异,在该表决过程中,特定位置处的优势碱基调用获胜,并被定义为真正的碱基调用,并且不存在于来自相同克隆起源的大多数分子中的那些碱基调用变松,并且被该个体家族内的优势碱基调用取代。

[0122] 6.对针对特定基因识别的独特分子/克隆家族的数目进行计数,并将这些计数与同一样本内的一组参考基因进行比较,并且还将这些计数与该基因在多个样本中的经验分布计数进行比较。当基因的独特计数相对于参考基因和/或经验分布变化超过定义的阈值时,识别拷贝数损失或拷贝数增益。

[0123] 7.通过读取序列读数上的破坏位点,来分析针对特定基因的易位或融合的cDNA序列。

[0124] 8.比较DNA和cDNA之间的突变和拷贝数计数,以确认所调用的突变事件。

[0125] 9.利用拷贝数变化、和/或易位、和/或基于尿的突变的量化丰度,来确定先前针对膀胱癌处理的患者中的膀胱癌的存在或不存在。

[0126] 10. 利用拷贝数变化、和/或易位、和/或基于尿的突变的量化丰度,来确定诊断患有膀胱癌的患者疾病进展的预后或风险。

[0127] 11. 利用拷贝数变化、和/或易位、和/或基于尿的突变的量化丰度,来诊断在其尿中呈现有血液的患者的膀胱癌。

[0128] 12. 利用拷贝数变化、和/或易位、和/或基于尿的突变的量化丰度,来筛查无症状或在其他方面被认为是健康个体和/或高风险群体(诸如吸烟者、具有职业性致癌物暴露历史的个体、具有饮用来自受砷或其他可疑致癌物污染的井水或地下水的水的历史的个体、或生活在地理上癌症热点内的个体)中的膀胱癌或其他癌症风险。

[0129] 13. 利用拷贝数变化、和/或易位、和/或基于尿的突变的量化丰度,针对由外部刺激诱导的基因毒性应激进行短期个体筛查(暴露后数小时至数天至数周中的测试),诸如在哺乳动物中测试新的药物产品时评估潜在基因毒性,或者根据向环境或休闲致癌物(诸如烟雾或燃烧产物、酒精、烟草、UV辐射)的暴露,将个体的癌症风险分层。突变负担的变化可能是短暂的或持续的,并且这些基因组变化可以随着时间纵向追踪。

[0130] 示例4

[0131] DNA在尿中丰富,并且可以最佳提取用于测量膀胱癌基因组生物标记

[0132] 为了改善先前尝试以最小程度地检测尿中的膀胱癌,实施例聚焦于尿DNA作为分析物,这是因为下一代DNA测序的技术进步,下一代DNA测序允许在单个测序反应中对数十到数千个基因进行大量复用分析。DNA也具有如下优点:相对稳定,并且在肿瘤形成过程中经历独特的变化,该独特的变化对癌症具有高度特异性。为了评估利用尿DNA的可行性,使用多种提取方法,实施并优化从20ml-100ml尿的DNA提取。使用荧光双链DNA结合染料测定(生命技术公司的QuantIt)、毛细管电泳和实时PCR,测量总DNA产量。使用来自多个尿样的RNaseP基因的量化实时PCR扩增,来测量PCR扩增效率。随后的分析表明,当使用功能化磁珠方法提取DNA时,产量优异并且PCR扩增增强(更低的阈值循环(Ct))。在实施方式中,当在低体积、低浓度或退化的样本中使用,带正电荷的功能化磁珠提供有利的提取产量。

[0133] 示例5

[0134] 为了进一步验证作为有效的疾病生物标记的尿DNA的类型,分析了细胞团相关和尿无细胞DNA。其中将这两个群体及其各种大小的分级相互比较,以确定存在最丰富的疾病信号的位置,如对匹配肿瘤组织的先前分析所定义的。此外其中,将这些群体内疾病标记丰度的差异与尿化学、尿细胞学、核酸片段化模式和临床相关物进行比较,并且其中这些相关性被用于开发算法,算法针对未来患者预测哪个核酸群体将包含疾病特异性生物标记的最丰富水平。

[0135] 示例6

[0136] (包含膀胱癌的基因组多样性的生物标记小组的开发)

[0137] 核酸测序能力、速度、敏感度的显著发展和成本下降已经导致在临床分子病理学实验室中快速采用癌症DNA测序。先前FDA批准的用于监测膀胱癌的测定中的一个显著缺点是:所使用的生物标记没有特异性(检测血尿或炎症),或者它们不完全涵盖疾病的蛋白质组或基因组多样性。为了改进现有技术的膀胱癌测试,特别是其低敏感度,本发明的特定实施例针对多种DNA膀胱癌生物标记的小组,其更好地涵盖了膀胱癌的基因组多样性。为了评估使用NGS监测膀胱癌突变负担的功效,已经开发了基于复用扩增子的库富集试剂的小组,

其集中于膀胱癌中的12个反复突变或扩增的基因(图1)。在这个图1中,基因小组被表示为矩阵,每个行是小组中的独特基因,并且每个列是膀胱癌TCGA数据集中的独特患者。在主矩阵内,列表示独特患者,行表示基因。单元格用针对特定患者和基因的基因内相关联变异体的类型来编码,单元格编码在变化图例(右侧)中表示。

[0138] 嵌在矩阵右侧的图表示该群体中与特定基因相关联的突变变异体的丰度和类型。矩阵上方的顶部嵌入的条形图表示每个患者基础上的独特事件的数量和类型。根据这一分析,127名患者(94.8%)包含在我们的生物标记小组中的一个或多个异常,其中每名患者平均有2.2个SNV。开发该小组以创建信息最少的基于DNA的疾病特征,其涵盖疾病的基因组多样性,但也允许经济性高度深度测序、片段化DNA的富集、以及单次测序运行中的复用样本分析。我们的该小组的初步实施例使用690个PCR扩增子来扩增68kb的基因组材料,提供了靶基因的93%覆盖率,和通过胚细胞对准(blast alignment)的非常高(>99%)的预测靶向基因富集。

[0139] 示例7

[0140] (对膀胱癌负担的敏感和特异性检测)

[0141] 为了验证我们在本文中所公开的测定,我们分析了11种对照膀胱癌细胞系,其先前已经通过全外显子组测序来测序((J.Barretina,G.Caponigro,N.Stransky,K.Venkatesan,A.A.Margolin,S.Kim,C.J.Wilson,J.Lehár,G.V.Kryukov,D.Sonkin,A.Reddy,M.Liu,L.Murray,M.F.Berger,J.E.Monahan,P.Morais,J.Meltzer,A.Korejwa,J.Jané-Valbuena,F.A.Mapa,J.Thibault,E.Bric-Furlong,P.Raman,A.Shipway,I.H.Engels,J.Cheng,G.K.Yu,J.Yu,P.Aspesi,M.de Silva,K.Jagtap,M.D.Jones,L.Wang,C.Hatton,E.Palescandolo,S.Gupta,S.Mahan,C.Sougnéz,R.C.Onofrio,T.Liefeld,L.MacConaill,W.Winckler,M.Reich,N.Li,J.P.Mesirov,S.B.Gabriel,G.Getz,K.Ardlie,V.Chan,V.E.Myer,B.L.Weber,J.Porter,M.Warmuth,P.Finan,J.L.Harris,M.Meyerson,T.R.Golub,M.P.Morrissey,W.R.Sellers,R.Schlegel,and L.A.Garraway,“The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity,”*Nature*,vol.483,no.7391,pp.603-607,Mar.2012);以及(S.A.Forbes,D.Beare,P.Gunasekaran,K.Leung,N.Bindal,H.Boutselakis,M.Ding,S.Bamford,C.Cole,S.Ward,C.Y.Kok,M.Jia,T.De,J.W.Teague,M.R.Stratton,U.McDermott,and P.J.Campbell,“COSMIC:exploring the world’s knowledge of somatic mutations in human cancer,”*Nucleic Acids Res.*,p.gkul075,Oct.2014))。我们在我们的小组内针对它们的动态范围而选择了这些细胞系,一些细胞系不包含突变而其他细胞系包含多个突变。这个分析允许我们识别并掩盖归因于反复映射错误或归因于冗余(均聚物)序列上下文的反复假阳性调用。通过针对对准、碱基调用和突变调用质量得分、基因座特异性读取深度、和变异等位基因频率来建立多个测序质量阈值,优化了我们管线的敏感度。

[0142] 使用这种精炼的突变调用管线,对14位具有不同肿瘤阶段、等级和临床亚型的癌症患者进行分析(血液、肿瘤和手术前尿的分析)。支持小组扩展以包括经常在膀胱癌原位突变的附加基因组区域,以及膀胱癌的其他临床亚型。当性能随着可在患者中监测到的突变数量的增加而提高时,小组扩展被设想为进一步有益于测定敏感性

[0143] 为了评估这种类型的方法的特异性,本发明的实施例已经在7个非癌症对照上(血

液和尿)验证了该小组。该队列包括具有不同泌尿学状况的患者,该不同泌尿学状况包括:良性前列腺增生、尿潴留、肾结石、寻求生育咨询的个体、和健康对照。在这些之中,2名患者是吸烟者,吸烟史为10和60包年。未来的研究将扩展非癌症对照队列,以包括进一步分析吸烟者和患有慢性泌尿系炎症性疾病的个体,因为这些患者中的一些患者在没有临床上可检测的膀胱癌的情况下可能包含小组突变。

[0144] 示例8

[0145] (尿DNA的纵向分析可以预测未来的疾病复发)

[0146] 为了评估此方法预测纵向疾病复发的能力,本发明的另一实施例涉及分析已知复发和长期纵向随访的两名患者,包括在原发性与复发性肿瘤的经尿道切除术之间收集的尿样。

[0147] 使用基于PCR扩增子的库富集,确定了取决于测序深度和扩增子性能的范围为~1-5%等位基因分级的等位基因检测的下限。通过增加数据收集来迭代地改进分析管线,包括重新校准基本质量得分、应用阈值和修改突变调用算法,以滤除复发小组特定映射错误和分析噪声。

[0148] 示例9

[0149] (设计涵盖膀胱癌多样性的增强型基因组小组)

[0150] 采用基于杂交捕获的库富集方法、更深度的测序以及询问更多样化和涵盖更广的生物标记集合,具有在示例性实施例中将UriSeq复发测定的敏感度提高多达2个数量级的能力。我们选择排他地专注于突变(单核苷酸变体),而不是SNV和拷贝数改变。目前用于检测SNV的算法在较低的测序覆盖率下比用于检测拷贝数变化的算法更敏感,并提供敏感度与测序成本之间的良好折衷。为了扩展评估的生物标记小组,我们建立了一组排序标准,以优先考虑用于包含在增强型小组中的复发突变基因。这些标准包括:1.复发性突变的流行。2.已知致癌基因的优先级。3.基因的大小和其分析的边际成本(考虑到可池化到单一反应中的探针数量的限制)。4.通过将基因或外显子添加到小组而捕获的突变的相互排斥性和独特患者的数量。5.独特临床亚型中突变基因的差异流行(例如在CIS、低等级或高等级损伤中的富集)。

[0151] 基于这些标准,提供了针对增强型小组的实施例,其目标是用于包括在复发测定中的23个基因中的750个外显子。使用COSMIC数据库和2种其他公开可用的膀胱癌数据集,计算验证了该修订的基因小组的综合性质,这总结于表2中((The Cancer Genome Atlas Research Network,“Comprehensive molecular characterization of urothelial bladder carcinoma,”*Nature*,vol.507,no.7492,pp.315-322,Mar.2014);(S.A.Forbes,D.Beare,P.Gunasekaran,K.Leung,N.Bindal,H.Boutselakis,M.Ding,S.Bamford,C.Cole,S.Ward,C.Y.Kok,M.Jia,T.De,J.W.Teague,M.R.Stratton,U.McDermott,and P.J.Campbell,“COSMIC:exploring the world’s knowledge of somatic mutations in human cancer,”*Nucleic Acids Res.*,p.gkul075,Oct.2014);以及(P.H.Kim,E.K.Cha,J.P.Sfakianos,G.Iyer,E.C.Zabor,S.N.Scott,I.Ostrovnya,R.Ramirez,A.Sun,R.Shah,A.M.Yee,V.E.Reuter,D.F.Bajorin,J.E.Rosenberg,N.Schultz,M.F.Berger,H.A.Al-Ahmadie,D.B.Solit,and B.H.Bochner,“Genomic Predictors of Survival in Patients with High-grade Urothelial Carcinoma of the Bladder,”*Eur.Urol.*,Aug.2014))。

[0152] 这种设计增加了由测定所覆盖的患者的百分比,并增加了每名患者的平均SNV数目。

[0153]

研究	研究大小 (患者数)	涵盖的患者 (%)	每名患者的平均事件数
Kim PH, et al. 2014	109	98	3.5
TCGA, 2014	134	96	3.3

[0154] 基于这些先前研究的计算机验证可能低估了该生物标记小组将涵盖的患者的百分比。迄今为止,膀胱癌的大规模(外显子组)测序研究集中于晚期肌肉侵入性疾病。作为增加跨临床亚型的我们小组的综合性质的我们努力的一部分,我们包括TERT启动子、FGFR3和STAG2突变,所有这些突变在低等级疾病中显著更普遍。先前的外显子组测序研究不捕获TERT启动子突变,TERT启动子突变是70%-80%的膀胱癌患者中存在的高度普遍的生物标记((C.D.Hurst,F.M.Platt,and M.A.Knowles,“Comprehensive Mutation Analysis of the TERT Promoter in Bladder Cancer and Detection of Mutations in Voided Urine,”*Eur.Urol*);(P.J.Killela,Z.J.Reitman,Y.Jiao,C.Bettegowda,N.Agrawal,L.A.Diaz,A.H.Friedman,H.Friedman,G.L.Gallia,B.C.Giovanella,A.P.Grollman,T.-C.He,Y.He,R.H.Hruban,G.I.Jallo,N.Mandahl,A.K.Meeker,F.Mertens,G.J.Netto,B.A.Rasheed,G.J.Riggins,T.A.Rosenquist,M.Schiffman,I.-M.Shih,D.Theodorescu,M.S.Torbenson,V.E.Velculescu,T.-L.Wang,N.Wentzensen,L.D.Wood,M.Zhang,R.E.McLendon,D.D.Bigner,K.W.Kinzler,B.Vogelstein,N.Papadopoulos,and H.Yan,“TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal,”*Proc.Natl.Acad.Sci.*,vol.110,no.15,pp.6021-6026,Apr.2013);(X.Liu,G.Wu,Y.Shan,C.Hartmann,A.von Deimling,and M.Xing,“Highly prevalent TERT promoter mutations in bladder cancer and glioblastoma,”*Cell Cycle*,vol.12,no.10,pp.1637-1638,May 2013);以及(I.Kinde,E.Munari,S.F.Faraj,R.H.Hruban,M.Schoenberg,T.Bivalacqua,M.Allaf,S.Springer,Y.Wang,L.A.Diaz,K.W.Kinzler,B.Vogelstein,N.Papadopoulos,and G.J.Netto,“TERT promoter mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine,”*Cancer Res.*,vol.73,no.24,pp.7162-7167,Dec.2013))。除了小组设计的扩展和优化之外,在某些实施例中,我们从扩增子测序转变成杂交捕获库制备方法。杂交捕获试剂提供了跨我们目标的更均匀的覆盖范围、我们库中的增强基因组复杂性、计算上标记重复的更大能力、更少的PCR循环和减少的聚合酶引入的错误、以及降低的库制备成本,从而允许负担得起的更深度

测序,其中这些优点中的任何一个或多个优点有助于提高测定敏感度。

[0155] 示例10

[0156] (开发错误抑制方法以允许敏感和特定的基于尿的基因组监测)

[0157] 传统的NGS方法产生大量噪声,这限制了低于1-5%的等位基因变异体的检测。在图2中,我们表明了跨癌症基因Rad51内的核苷酸的标准噪声水平。即使使用不含PCR的方法,在低于0.6%的突变等位基因频率的水平下,在5,000x深度处被测序时,几乎所有核苷酸都表明非参考读数水平。使用我们的错误抑制方法和高效库库转换,我们能够在未检测标准噪声(右侧,底部)的情况下,从研究中的尖峰中检测真正的阳性事件。这些分析和库制备的增强适于开发随着时间的推移追踪导致基因组异常的低频疾病的诊断方法。

[0158] 某些基于计算机处理器的实施例

[0159] 在某些实施例中,上文所描述和/或所执行的步骤可以通过许多方式并且以许多方式来实施,包括但不限于被实施作为一个或多个系统或装置;一个或多个处理;物质的组成;一系列指令,该一系列指令驻留或非驻留到一个硬件设备、或者被耦合在一起和/或一起通信的多个硬件设备;一个或多个计算机程序产品,其有形地被实施在计算机可读存储介质上并且可在一个或多个处理器上操作;任何一个或多个处理器,被配置为执行由耦合到处理器的存储器所提供的指令;以及涉及由机器读取和/或执行指令的技术人员已知的任何技术。除非另有申明,否则被描述为被配置为执行任务的组件(诸如处理器或存储器)可以被实施为:临时配置为在给定时间执行任务的通用组件、或被制造为执行任务的特定组件。如本文所使用的,术语“处理器”是指(非限制性地)一个或多个设备、电路、处理核、或由一个机器或通信地耦合在一起的多个机器执行的其他指令,并且可以被配置为处理任何形式的驻留的或非驻留的数据。

[0160] 现在参考图5和图6,公开了流算法的图示,其包括(A)基因组库502和503以及原始患者测序数据501用作该算法的输入。在一些实施例中,基因组库包括专有的内部基因组库602以及从开源所收集的数据,该开源包括:桑格宇宙数据库604(<http://cancer.sanger.ac.uk/cosmic>)、dbSNP(www.ncbi.nlm.nih.gov/projects/SNP/)、参考基因组信息613(例如<http://hgdownload.cse.ucsc.edu/goldenPath/hgl9/chromosomes/>)、以及可用科学文献精选614。在一些实施例中,基因组库包括注释503。这些是从科学文献精选的开源库。这些包括AnnoVar (PMC2938201) 606、CBI0门户网站(www.cbioportal.org/) 607、OncoKB (www.oncokb.org) 619、癌症热点项目 (<http://cancerhotspots.org/>) 618、以及Mutation Assessor (<http://mutationassessor.org/r3/>) 617。在一些实施例中,原始患者测序数据输入包括匿名患者样本数据库503,匿名患者样本数据库503包含原始(未处理)测序数据609、对准测序数据610、患者的临床历史611和尿的尿化学分布612。(B) 度量生成器518提取50个度量。这些度量是通过如下方式生成的:过滤关于使用SAMTOOLS 506 (www.htslib.org) 计算出的质量测量Phred、MAPQ和Read Depth 516的读取信息。在一些实施例中,度量生成器包括噪声表征程序505。此外,为了确定患者样本数据中的基因组测量的质量,SAMTOOLS结合多路堆积命令517被用于量化phred得分(PHRED-定义测序器碱基调用置信度的统计)、映射质量(MAPQ-定义读取映射中的基因组对准器(aligner)的置信度的统计)、以及在感兴趣的患者基因内的每个基因座处的读取深度(READ DEPTH-定义位置被测量或计数的次数的度量)。然后,这些数据被传送给我们的质量控制(QC)过滤器515。过

过滤器执行一系列逻辑论证,以确保度量在预期范围内定义。如果遇到测量不良或映射不良的基因组信息,则丢弃该信息,并将算法移至新位置以确定噪声特性。如果符合QC标准,算法的下一步骤就开始了。如果读数具有质量信息,则确定读数的分子复杂度514(称为家族度量513),并且将读数的分子复杂度514与来自开源数据工具(包括dbSNP、癌症热点和OncoDB509)的注释508组合。然后,这些数据被留存到我们的数据库,以用于报告。量化分子复杂度提供了成功的库制备的测量和充分采样核酸库的测序方法能力。通过测序读数中不同测量的数量,来量化分子复杂度514。在库制备过程中,个体分子被复制,产生重复分子家族——称为家族。反过来,每个家族对应着独特的分子,独特的分子被用于识别基因组中当前位置处的碱基。家族数量提供了理解用于验证给定碱基和/或突变事件的存在的基础。家族度量513还用于抑制测序和PCR扩增引起的错误。注释包括先前表征的基因组事件的数据库,其描述正常人类变异、编目的癌症变异或算法,算法基于突变类型、基因内的位置、所产生的蛋白质结构变化、以及其他标准,来预测变异函数,以估算突变事件与和癌症或发展癌症的风险有关的基本细胞处理或病理性细胞处理的相关性。给定碱基位置处的注释的数量和类型部分地提供了定义给定突变的致癌性质的手段。该信息用于通知分子等级和临床报告。(C)由度量生成器确定的50个度量加注释用作突变调用器520的输入,突变调用器520过滤这些数据以对基因组变异体进行分类。基因组变异检测519是如下处理,该处理量化测量的患者DNA与健康人类基因组不一致的程度,并使用度量来帮助区分噪声(假阳性变异体)与真阳性基因组变异体。这种不一致性是通过将统计阈值放置在经验噪声量上来量化的——定义为难以测量或易于测量碱基的程度,存在的高质量数据的量——并且通过包括给定碱基位置处的测量群体的独特分子的数量,可以进一步支持或驳斥这种不一致性。特别地针对从尿提取和测序的核酸,对经验噪声进行建模。经验噪声模型527还可以整合尿内的各种患者临床特征和临床化学测量。创建错误分布,以涵盖从临床上注释的尿基因组样本产生的尿样的生理学和病理学多样性。分子复杂度模型521通过将分子复杂度度量组合成算法而生成。这些变异体(如果存在的话)根据它们的分子等级525进一步分类,并与患者的临床数据和癌症热点、OncoDB、TCGA中的基因组事件的先前表征两者进行比较。基因组变异体分类522包括如下方法,通过该方法,使用在度量生成器处理中生成的值来应用算法,其中这些值包括经验噪声模型527和分子复杂度模型521的组合(参见算法流程图图5)。其中,通过使用尿特异性参考样本和系列稀释和/或匹配的肿瘤-尿相关性研究,来确定这些度量的组合和阈值,以优化技术敏感性、特异性、真阳性、假阳性、真阴性523和假阴性比率,以及用于确定疾病状态524。而且其中,这些组合和阈值通过使用迭代测试和机器学习算法(诸如随机森林)进一步精炼。分子等级525是如下处理,其中变异体注释、对先前精选的文献的参考、基因组变异数据库(内部和公共的)、以及算法,算法基于患者体内各种基因组变异体的独特组合、以及它们与临床特征和/或传统的病理等级的关联。分子等级可以按照与传统病理等级类似的方式进行分类和报告,例如高等级或低等级。分子等级还可以提供预后信息,该预后信息与正常或处于升高的癌症发展风险下的有症状或无症状的个体中的复发、进展的风险或未来肿瘤发展的风险相关。如果正在针对癌症复发而监测个体,可以提供风险进展得分526,可以基于分子和临床特征的组合来分配风险进展得分。这些度量一起形成了临床报告的基础。流算法包括用于将数据保存到数据库的程序,由此数据库对象关系映射510提供了将程序变量存储到数据库中的表结构或文件结构的手

段。在一些实施例中，用于将程序变量存储到数据库中的表结构或文件结构的手段由提供该功能的开源工具来提供，该开源工具包括但不限于SQLAlchemy 512。所描述的算法的输出与临床数据组合成临床报告504。这些被留存到可共享的数据储存库，供公司研究人员和合作医生使用。

[0161] 在某些实施例中，本文中的任何一个实施例或组合多个实施例的方法由计算机可读介质指示，在计算机可读介质上存储有用于执行这样的方法的计算机可读指令。

[0162] 返回到附图，图6图示了(A) 计算平台，计算平台是包含在虚拟私有网络 (VPN) 601 中的基于云的计算基础设施。计算平台由数据基础设施 (B) 支持，该数据基础设施由专有 (B i, ii, iv) 501、502和504以及开源 (B iii) 503基因组库两者组成。这些库结合原始患者测序数据 (B i) 501用作计算算法 (D I, , ii) 518和520的输入，计算算法 (D I, , ii) 518和520提取用于表征基因突变的多种不同度量。数据处理在动态缩放的计算机集群 (C i, ii) 624上并行执行，该计算机集群将数据输出到中央数据储存库621。计算集群可以包括计算机网络622，其接收来自中央计算机的计算作业、以及输入数据和代码基623，以执行计算。集群中的每台计算机执行计算，并将数据返回到中央计算机620的共享存储设备。该数据储存库不仅用于传播发现 (B iv) 504，而且用于对突变编目录，以进一步精炼突变检测 (B ii, D iii) 502和503。代码基 (D) 623包括三个基本组分：来自患者测序数据和基因组库的度量提取 (D i) 518、基因突变量化 (D ii) 520、以及在存在计算的突变特征的情况下对这些方法的精炼 (D iii) 625。

[0163] 任何和所有前述执行方式中的指令的上述执行都是参考以下而被采用的：分析算法被实施在改进的测定中，改进的测定可以允许例如在初始评估患者的原发性肿瘤之后，或在对多个尿DNA核酸样本进行纵向分析之后，纵向监测尿DNA；开发增强的生物标记靶向小组，其例如能够涵盖膀胱癌（并且在某些实施例中，血尿）的基因组和临床多样性；在某些实施例中，提供高技术性能，同时实现临床上可行的测定费用和处理时间；以产生高敏感度和特异性的方式监测膀胱患者的尿；检测与膀胱癌相关的一个或多个基因中的突变；从来自受试者的尿样中分离核酸、DNA或RNA，并分析核酸以获得核酸序列数据，核酸序列数据适合于检测与膀胱癌相关的基因中的一个或多个基因中存在或不存在一个或多个突变；分离核酸，核酸是无细胞核酸和/或是从尿样中的细胞中分离的核酸；和/或根据包括但不限于通过在基因测序中使用基因生物标记和方法的实施例，关于个体或个体组，来执行本文中引用的方法中的一个或多个方法，以检测、预后、诊断和处理膀胱癌。

[0164] 在示例性实施例中，序列或其他数据被输入到处理器或其他计算机硬件组件。这里，处理器与测序设备耦合或以其它方式通信，测序设备读取和/或分析来自样本的核酸的序列。序列从处理工具或序列存储源被提供。一个或多个存储器设备缓存或存储序列。存储器还可以存储针对各种染色体或基因组的读数、标签、片段、相位信息和岛等，并可以存储用于分析和呈现序列或对准数据的指令。

[0165] 在某些实施例中，方法还包括收集关于多个核苷酸序列的数据。示例包括读数、标签和/或参考染色体序列。数据可以被发送到处理设备、硬件系统或其他计算系统。在一个示例性实施例中，处理器连接到实验室设备。这样的设备可以包括核苷酸扩增装置、样本收集装置、核苷酸测序装置和/或杂交装置。

[0166] 处理器然后可以收集已经由实验室设备采集的适用数据。在示例性实施例中，不

将其作为穷举列表,数据存储在机器或其他处理装置的驻留或非驻留存储装置中;数据是在数据传输之前、数据传输期间或与数据传输一起实时收集的;数据存储可在从处理器提取的计算机可读介质上;数据经由任何耦合或通信手段被传输到远程位置,包括但不限于经由计算机总线、经由局域网、经由广域网、通过内联网或因特网、经由有线、无线或卫星信号、以及通过任何已知的传输形式或媒介;数据在远程位置被处理和操作。

[0167] 现在参考图7,示出了在一天过程中从同一个体收集的尿核酸的毛细管电泳分布。核酸的相对大小由下和上标记参考峰(LM、UM)限定,其中LM表示0bp,并且UM表示5,000bp(水平轴,碱基对大小)。竖直轴反映核酸结合染料的荧光强度值,这允许量化特定大小下的核酸分子的摩尔浓度。

[0168] 如图7所示,清晨排泄最常见的特征是大量的高分子量核酸和低丰度的小分子。在这种状态下,这些核酸群体被分开,并且尿通常花费最长时间段与泌尿道和膀胱接触,从而丰富了尿中泌尿道生物标记的丰度。随着一天的进展,增加的中等大小的分子出现,这归因于尿中的增加的降解条件或改变的肾生理学修改了经肾DNA分级的大小分布。降解倾向性条件的进一步特征在于:高分子量峰向下移动(朝向下标记),以及较小的较低分子量峰的加宽和强度增加。在进行总DNA测量的情况下,较小(小于80bp)的分子可以稀释来自泌尿道核酸的信号,并且如果未相应地标准化,则会损害测序库制备效率。

[0169] 根据本发明的方法的一个实施例,从早晨排泄尿样收集尿以帮助丰富泌尿道信号,并且根据本文公开的一个或多个技术处理尿。接下来,分析核酸标记和/或其他标记。收集关于收集时间的数据以及其他患者数据,其他患者数据包括身份、年龄、体重、性别、用药、疾病、临床和其他个人数据,并且将这样的数据利用样本追踪并输入到数据库中。在实施例中,将产生的数据与在相同的收集时间和不同的收集时间下的同一患者数据进行比较。在其他实施例中,将数据与其他患者数据进行比较。根据毛细管电泳和/或实时PCR质量控制,对数据进行量化和标准化,并使用大小分布信息构建DNA库。

[0170] DNA片段化分布的变化可能由不同的生理和储存条件引起。除了收集的一天中的时间之外,某些个体似乎对一种尿分布类型具有天然偏置(超过另一种尿分布类型)。在主要为小的/经肾分布的个体中,更加重要的是,当利用膀胱的尿孵育已经最大化(清晨)时收集样本,并且在其他情况下立即将尿排入保存缓冲液中,保存缓冲液抑制核酸酶活性以防止核酸的降解。

[0171] 通过使用毛细管电泳和/或测序读取开始和停止位点分析的分析,跨各种年龄、性别、和疾病或健康状态的人的代表性样本,描述了尿中核酸大小的不均匀性。利用这一数据,可能评估尿核酸大小和片段化分布是否以及如何个体内、在时间过程(几小时、几天、几周或几个月)内、以及响应于生理扰动(诸如疾病、昼夜节律、饮食和水合作用)而发生变化。在该实施例的一个版本中,尿中的核酸大小和片段化模式被用作疾病分类算法的一个分量。

[0172] 此外,已经确定样本处理、保存和储存条件将影响在特定尿样内观察到的各种大小的分子或不均匀性。另外,核酸提取方法也会影响之前并未结合这样的尿分析表征的样本中的大小变化。如以下所讨论的,特定患者的样本也表现出不同的核酸大小分布。虽然各种大小分布对测序库制备效率和测序性能的最终影响尚未完全表征,但本发明的实施例涉及对这些变量中的每一个变量的表征,然后将收集的数据用于创建患者分布的数据库,并

最终改善膀胱癌的诊断和预后两者。在实施例中,收集的与核酸大小相关的数据与以上讨论的各种相关因素中的一个或多个因素相关联。在一个实施例中,来自具有预定分布的患者的样本根据其分布被标准化并通过测序被评估,以测量独特的片段化(测序开始/停止)位点,并且其中该序列上下文片段化模式作为一个方面被整合到疾病诊断算法中。

[0173] 已经确定尿核酸在其跨个体的大小分布上具有显著的不均匀性。此外,样本大小上的核酸的不均匀性在一天中在个体内是不统一的。另外,当尿在室温下放置数分钟至数小时至数天时,样本中的核酸降解可以显著减小核酸分子的大小。降解可能以各种方式发生,在一个示例中,较高分子量DNA降解,从而大小变小并且增加尿中小分子量DNA的丰度,这被称为低分子量池化。在另一示例中,高分子量DNA完全降解超出检测范围,不会在低分子量池内累积。此外,冷冻和解冻尿的处理对核酸大小和损伤具有显著影响。在一个实施例中,通过分析围绕读取开始和停止碱基的序列上下文,由于处理损伤引起的DNA降解可以区别于由于生物处理(例如细胞凋亡和坏死)而破碎的DNA。并且其中,这一信息用于创建样本质量比,以使测序数据标准化。

[0174] 根据本发明的一个方面,开发了一种数据库,其包括跨数千个独特的尿样以及跨数百个独特的生理学、病理学和处理条件的各种核酸大小分布和片段化序列上下文分析。然后将这一数据与采样数据相关联,并比较记录以提供与可变尿核酸大小的底层原因相关的输出。

[0175] 本发明的实施例涉及以下步骤:(1)对各种大小分布进行测序方法的迭代优化;(2)优化样本收集和储存技术以保持核酸大小的完整性;(3)实施质量控制以滤除质量差的样本;以及(4)将最终测序数据标准化回核酸大小分布的独特特征。合起来,这些步骤和针对测序方法的多次迭代已经产生高质量的基于尿的基因组学分析。

[0176] 用于检测泌尿道内疾病的诊断敏感度受样本中核酸大小分布的影响。基于这种理解,我们已经定义了以下参数/组合以增强测定性能。

[0177] 在一个实施例中,分析包括取决于样本分布的在120-5000bp范围内和/或在5,000-10,000bp范围内的核酸的靶向和富集,并且其中,通过(1)诸如由Covaris所公开的超声破碎法之类的机械破碎技术,(2)诸如由Kapa Hyper-plus执行的破碎之类的基于酶的破碎,(3)限制酶,或(4)各种限制酶的混合物,这些大小范围可以被片段化成大小为500-600bp的分子群体,并且其中,片段化的分子然后置于库制备反应中。

[0178] 在一个实施例中,对在第一或第二早晨排泄期间从个体收集的尿样进行分析,由此最大化尿与膀胱上皮接触所花费的时间。在使用片段化技术进行样本处理之后,执行基因组的分析,以确定样本中是否存在多个标记DNA或RNA标记区段。

[0179] 在一个实施例中,对在食用膳食或饮用流体之前收集的尿样进行分析,从而使肾的生理活动最小化,其中所述分析包括处理以确定多个标记DNA或RNA区段是否在样本中。

[0180] 在一个实施例中,针对核酸大小来执行样本标准化,以便(1)开发尿测序诊断,该尿测序诊断分析泌尿道中的信号,有利的是,富集并分析大小大于100-150bp的核酸,以及(2)开发尿测序诊断,该尿测序诊断分析来自体循环的核酸信号,有利的是,富集大小小于100个碱基对的核酸,并且具体地核酸的范围可以从20个至100个碱基对,这取决于肾功能/健康。常见的DNA测量(诸如紫外吸收和荧光测定)不提供大小信息,并且如果孤立地使用,常见的DNA测量可能会导致DNA过量或不足的加载到库制备反应中(见图3)。由于这个原因,

本发明的方法结合毛细管电泳和/或实时PCR反应来使用,其中反应的引物被设计为分开特定的核苷酸距离,以测量一种大小相对于另一种大小的丰度。可能有多(1、2、3)组的被设计用于不同大小的引物/扩增子(例如Kapa人类DNA量化试剂盒)。这些引物可以被设计用于大小为30-70个碱基对、大小为70-150个碱基对、大小为150-500个碱基对、或大小大于1,000个碱基对的扩增子。在确定来自尿的核酸样本的大小分布时,可将样本标准化,以确保将足够分位数的特定大小范围置于库制备反应中。在一个实施例中,这将是大小大于80bp的分子(参见图4)。基于总核酸加载,可以修改随后的库制备步骤,例如连接后的清洁中使用的羧化顺磁碱基的体积。备选地,在库制备之前,通过经过大小选择柱(诸如Pall Nanosep®设备)、用羧化顺磁珠(诸如AmpPureXP珠)的处理、凝胶毛细管电泳、凝胶电泳或阴离子交换(诸如Sage Sciences Pippin Prep或Pall Mustang膜),可以基于片段大小差异地分离核酸。

[0181] 图9描绘了点状图,其表示患者匹配样本中的肿瘤与尿核酸中的等位基因频率之间的关系,其中在肿瘤存在于膀胱中时收集尿。竖直轴是非参考等位基因频率,水平轴是靶向基因组区域内的基因组位置,点表示样本类型,并且在图说明中描述。患者A示出了一些患者在肿瘤(42%-71%范围)和尿(38%-60%)之间具有高的等位基因频率一致性,其中尿中的大部分核酸具有肿瘤起源。相反,患者B示出了另一种情况,其中尿中肿瘤来源的核酸丰度低得多,并且肿瘤(26%-51%)和尿(0.3%-2.2%)突变丰度不一致,而对于那些位置(灰色X),尿突变仍然保持丰度高于参考数据库范围。患者A和B两者都表明肿瘤和尿样两者内不同等位基因频率群集的附加特征。在一个实施例中,等位基因频率群集的程度或类型可以用作诊断或预后疾病算法的一部分。

[0182] 图10描绘了非癌症和癌症患者中过滤的突变丰度的条形图。从患者收集尿,其中已知癌症或没有癌症存在于其泌尿道内。水平轴表示独特的患者尿样,而竖直轴表示在算法过滤(在图5和6中描述)后,在对40个预选基因(取自表1)执行这一分析时,在样本中识别的突变数目。在实施本文所描述的质量控制和数据分析算法后,将这些样本内推定的突变的数量从133个与1,984个之间的事件调整至此处所示的事件(每个样本0-9个事件)。过滤后,尿相关联突变的丰度能够将疾病与非疾病状态分开,并具具有诊断、复发疾病检测和疾病表征的效用。

[0183] 结论

[0184] 应该注意的是,这里描述的顺序和标记的操作表示某些所呈现的方法的一个或多个示例性实施例。其他操作和方法可以由技术人员所设想,其在功能、逻辑或效果上等同于所示方法的一个或多个操作或其部分。虽然这里的方法的操作以特定的顺序示出和描述,但是可以改变每个方法的操作顺序,使得某些操作可以以相反的顺序执行,或者使得某些操作可以至少部分地与其他操作同时执行。在其他实施例中,不同操作的指令或子操作可以以间歇和/或交替的方式来实施。

[0185] 最后,尽管上面已经描述了本发明的各种实施例,但应该理解,它们仅仅是通过示例的方式被呈现,而不是限制性的。因此,本实施例的宽度和范围不应受到任何上述描述的限制。

[0186] 文献引用

[0187] [17] A.M. Newman, S.V. Bratman, J. To, J.F. Wynne, N.C.W. Eclöv, L.A. Modlin,

C.L.Liu, J.W.Neal, H.A.Wakelee, R.E.Merritt, J.B.Shrager, B.W.Loo Jr, A.A.Alizadeh, and M.Diehn, "An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage," *Nat.Med.*, vol. advance online publication, Apr. 2014.

[0188] [18] S.R.Kennedy, M.W.Schmitt, E.J.Fox, B.F.Kohn, J.J.Salk, E.H.Ahn, M.J.Prindle, K.J.Kuong, J.-C.Shen, R.-A.Risques, and L.A.Loeb, "Detecting ultralow-frequency mutations by Duplex Sequencing," *Nat.Protoc.*, vol. 9, no. 11, pp. 2586-2606, Nov. 2014.

[0189] [19] M.W.Schmitt, S.R.Kennedy, J.J.Salk, E.J.Fox, J.B.Hiatt, and L.A.Loeb, "Detection of ultra-rare mutations by next-generation sequencing," *Proc.Natl.Acad.Sci.U.S.A.*, vol. 109, no. 36, pp. 14508-14513, Sep. 2012.

[0190] [20] E.Crowley, F.Di Nicolantonio, F.Loupakis, and A.Bardelli, "Liquid biopsy: monitoring cancer-genetics in the blood," *Nat.Rev.Clin.Oncol.*, vol. 10, no. 8, pp. 472-484, Aug. 2013.

[0191] [21] M.Murtaza, S.-J.Dawson, D.W.Y.Tsui, D.Gale, T.Forsheew,

[0192] A.M.Piskorz, C.Parkinson, S.-F.Chin, Z.Kingsbury, A.S.C.Wong, F.Marass, S.Humphray, J.Hadfield, D.Bentley, T.M.Chin, J.D.Brenton, C.Caldas, and N.Rosenfeld, "Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA," *Nature*, vol. 497, no. 7447, pp. 108-112, May 2013.

[0193] [22] T.Forsheew, M.Murtaza, C.Parkinson, D.Gale, D.W.Y.Tsui, F.Kaper, S.-J.Dawson, A.M.Piskorz, M.Jimenez-Linan, D.Bentley, J.Hadfield, A.P.May, C.Caldas, J.D.Brenton, and N.Rosenfeld, "Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA," *Sci.Transl.Med.*, vol. 4, no. 136, pp. 136ra68-136ra68, May 2012.

[0194] [23] G.Sozzi, D.Conte, M.Leon, R.Ciricione, L.Roz, C.Ratcliffe, E.Roz, N.Cirenei, M.Bellomi, G.Pelosi, M.A.Pierotti, and U.Pastorino, "Quantification of free circulating DNA as a diagnostic marker in lung cancer," *J.Clin.Oncol.Off .J.Am.Soc.Clin.Oncol.*, vol. 21, no. 21, pp. 3902-3908, Nov. 2003.

[0195] [24] C.Fernandez, Shore, and A.Shuber, "Noninvasive multianalyte diagnostic assay for monitoring bladder cancer recurrence," *Res.Rep.Urol.*, p. 49, Oct. 2012.

[0196] [25] C.Fernandez, Millholland, Li, and A.Shuber, "Detection of low frequency FGFR3 mutations in the urine of bladder cancer patients using next-generation deep sequencing," *Res.Rep.Urol.*, p. 33, Jun. 2012.

[0197] [29] W.Ranasinghe and R.Pers, "The Changing Incidence of Carcinoma In-Situ of the Bladder Worldwide," in *Advances in the Scientific Evaluation of Bladder Cancer and Molecular Basis for Diagnosis and Treatment*, R.Persad, Ed. InTech, 2013.

[0198] [31] S.Myllykangas, J.D.Buenrostro, G.Natsoulis, J.M.Bell, and H.P.Ji,

“Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing,”*Nat.Biotechnol.*,vol.29,no.11,pp.1024-1027,Nov.2011.

[0199] [32]H.Lee,B.T.Lau,and H.P.Ji,“Targeted Sequencing Strategies in Cancer Research,”in *Next Generation Sequencing in Cancer Research*,W.Wu and H.Choudhry,Eds.Springer New York,2013,pp.137-163.

[0200] [33]“Press Announcements-FDA allows marketing of four ‘next generation’ gene sequencing devices.”[Online].Available:www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm375742.htm.[Accessed:02-Dec-2014].

[0201] [34]K.Bijwaard,J.S.Dickey,K.Kelm,and Z. **Težak**,“The first FDA marketing authorizations of next-generation sequencing technology and tests: challenges,solutions and impact for future assays,”*Expert Rev.Mol.Diagn.*, pp.1-8,Nov.2014.

[0202] [35]F.S.Collins and M.A.Hamburg,“First FDA Authorization for Next-Generation Sequencer,”*N.Engl.J.Med.*,vol.369,no.25,pp.2369-2371,Nov.2013.

[0203] [36]D.C.Koboldt,Q.Zhang,D.E.Larson,D.Shen,M.D.McLellan,L.Lin,C.A.Miller,E.R.Mardis,L.Ding,and R.K.Wilson,“VarScan 2:somatic mutation and copy number alteration discovery in cancer by exome sequencing,”*Genome Res.*, vol.22,no.3,pp.568-576,Mar.2012.

[0204] [37]A.Wilm,P.P.K.Aw,D.Bertrand,G.H.T.Yeo,S.H.Ong,C.H.Wong,C.C.Khor,R.Petric,M.L.Hibberd,and N.Nagarajan,“LoFreq:a sequence-quality aware,ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets,”*Nucleic Acids Res.*,vol.40,no.22, pp.11189-11201,Dec.2012.

[0205] [38]Z.We,W.Wang,P.Hu,G.J.Lyon,and H.Hakonarson,“SNVer:a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data,”*Nucleic Acids Res.*,vol.39,no.19,p.e132,Oct.2011.

[0206] [39]K.Cibulskis,M.S.Lawrence,S.L.Carter,A.Sivachenko,D.Jaffe,C.Sougnuez,S.Gabriel,M.Meyerson,E.S.Lander,and G.Getz,“Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples,”*Nat.Biotechnol.*,vol.31,no.3,pp.213-219,Mar.2013.

[0207] [41]J.Reading,R.R.Hall,and M.K.Parmar,“The application of a prognostic factor analysis for Ta.T1 bladder cancer in routine urological practice,”*Br.J.Urol.*,vol.75,no.5,pp.604-607,May 1995.

[0208] 表1

[0209]

基因符号	同义符号	完整基因名称	染色体	HG19 碱基对起始位点	HG19 碱基对终止位点
KDM6A		赖氨酸(K)-特异性脱甲基酶 6A	X	44732423	44971845
MLL2	KMT2B	赖氨酸(K)-特异性甲基转移酶 2D	12	49412758	49449107
TSC1		结节性硬化 1	9	135766735	135820020
NOTCH2		缺口 2	1	120454176	120612317
PTEN		磷酸酶和张力蛋白同系物	10	89623195	89728532
TP53		肿瘤蛋白 p53	17	7571720	7590868
NOTCH1		缺口 1	9	139388896	139440238
CDKN2A		细胞周期蛋白依赖性激酶抑制剂 2A	9	21967751	21994490
RB1		视网膜神经胶质瘤 1	13	48877883	49056026
ATM		ATM 丝氨酸/苏氨酸激酶	11	108093559	108239826
ERBB2		erb-b2 受体酪氨酸激酶 2	17	37844393	37884915
PIK3CA		磷脂酰肌醇-4,5-二磷酸 3-激酶,催化亚基 α	3	178866311	178952497
FGFR3		成纤维细胞生长因子受体 3	4	1795039	1810599
EGFR		表皮生长因子受体	7	55086725	55275031

[0210]

FGFR1		成纤维细胞生长因子受体 1	8	38268656	38326352
CREBBP		CREB 结合蛋白	16	3775056	3930121
LRP1B		低密度脂蛋白受体相关蛋白 1B	2	140988996	142889270
MYC		v-myc 禽髓细胞瘤病毒致癌基因同系物	8	128748315	128753680
ARID1A		富 AT 的交互域 1A (SWI 状)	1	27022522	27108601
MLL3	KMT2C	赖氨酸(K)-特异性甲基转移酶 2C	7	151832010	152133090
BIRC3		含 3 的杆状病毒 IAP 重复	11	102188181	102210135
WWOX		含氧化还原酶 WW 域	16	78133327	79246564
PALB2		BRCA2 的合作者和定位器	16	23614483	23652678
SOX4		SRY (性别确定区域 Y) - 框 4	6	21593972	21598849
YAP1		是相关联蛋白 1	11	101981192	102104154
CCND1		细胞周期蛋白 D1	11	69455873	69469242
BCL2L1		BCL2 状 1	20	30252261	30310656
MYCL1		v-myv 禽髓细胞瘤病毒致癌基因肺癌衍生同系物	1	40361096	40367687
MDM4		MDM4, p53 调节剂	1	204485507	204527248
FGF3		成纤维细胞生长因子 3	11	69624736	69634192
MDM2		MDM2 原癌基因, E3 泛素蛋白连接酶	12	69201971	69239320
CCNE1		细胞周期蛋白 E1	19	30302901	30315215
ZNF703		锌指蛋白 703	8	37553301	37556396
PRKCI		蛋白激酶 C, ι (iota)	3	169940220	170023770
NCOR1		核受体辅阻遏子 1	17	15933408	16118874
YWHAZ		酪氨酸 3-单氧酶/色氨酸 5-单氧酶活化蛋白 ζ	8	101930804	101965623
PPARG		过氧化物酶体增植物激活受体 γ	3	12329349	12475855
TBL1XR1		转导蛋白 (β) 状 1 X-连接受体 1	3	176738542	176915048
PDE4D		磷酸二酯酶 4D, cAMP 特异性	5	58264866	59783925

[0211]

IKZF2	IKAROS 家族锌指 2 (回照器 (helios))	2	213864411	21401633
SPAG1	精子相关抗原 1	8	101170263	101254132
E2F3	E2F 转录因子 3	6	20402137	20493945
NIT1	腓水解酶 1	1	161087862	161095235
BEND3	含 3 的 BEN 域	6	107386385	107435636
GDI2	GDP 离解抑制剂 2	10	5807186	5855512
PVRL4	骨髓灰质炎病毒受体相关 4	1	161040781	161059385
CCSER1	卷曲螺旋富含丝氨酸的蛋白 1	4	91048684	92523370
TERT Promoter	端粒酶逆转录酶启动子区域	5	1253287	1295162
SPTAN1	血影蛋白, α , 非红细胞 1	9	131314837	131395944
HRAS	哈维鼠肉瘤病毒致癌基因同系物	11	532242	535550
CTNNB1	连环蛋白 (钙粘蛋白相关蛋白), β 1, 88kDa	3	41240942	41281939
FBXW7	含 7, E3 泛素蛋白连接酶的 F 框和 WD 重复域	4	153242410	153456393
EP300	E1A 结合蛋白 p300	22	41488614	41576081
RHOA	ras 同系家族成员 A	3	49396579	49449526
CCND3	ras 同系家族成员 A	6	41902671	42016610
NOS1AP	细胞周期蛋白 D3	1	162039581	162339813
ELF3	一氧化氮合成酶 1 (神经元) 适配体蛋白	1	201979690	201986315
PTPRD	E74 状因子 3 (ets 域转录因子, 上皮特异性)	9	8314246	10612723
STAG2	蛋白酪氨酸磷酸酶, 受体类型, D	X	123094475	123236505
ERBB3	基质抗原 2	12	56473809	56497291
CDKN1A	erb-b2 受体酪氨酸激酶 3	6	36644237	36655116
NFE2L2	细胞周期蛋白依赖性激酶抑制剂	2	178095031	178129859

[0212]

		1A (p21, Cip1)			
AIRE		核因子, 红细胞 2 状 2	21	45705721	45718102
BTG2		自身免疫调节剂	1	203274664	203278729
TTC28		BTG 家族, 成员 2	22	28374002	29075853
IKZF3		三十四肽 (tetratricopeptide) 重复域 28	17	37913968	38020441
FHIT		IKAROS 家族锌指 3 (Aiolos)	3	59735036	61237133
SHANK2		脆性组氨酸三联体	11	70313961	70935808
ERCC2		SH3 和多个锚蛋白重复域 2	19	45854649	45873845
TPTE		切除修复交叉互补组 2	21	10906743	10990920
KLF5		跨膜磷酸酶与张力蛋白同源物	13	73633142	73651676
FOXA1		Kruppel 状因子 5 (肠)	14	38058757	38064325
PON3		叉头框 A1	7	94989184	95025687
RXRA		对氧磷酶 3	9	137218316	137332431
ZFP36L1		类视黄醇 X 受体, α	14	69254372	69262960
GPC5		ZFP36 环指蛋白状 1	13	92050935	93519487
PCSK5		磷脂酰肌醇蛋白聚糖 5	9	78505560	78977255
CTIF		前蛋白转化酶 subtilisin/kexin 5 型	18	46065427	46389586
FOXQ1		CBP80/20-依赖性翻译起始因子	6	1312675	1314993
TIMM9		叉头框 Q1	14	58875370	58874232
CX3CL1		内线粒体膜 9 转位酶同系物 (酵母)	16	57406414	57418956
TXNIP		趋化因子 (C-X3-C 基序) 配体 1	1	145438462	145442628
RHOB		硫氧还蛋白相互作用蛋白	2	20646835	20649201
PAIP1		ras 同系家族成员 B	5	43526370	43557521
PHACTR1		ras 同系家族成员 B	6	12717037	13288075
CDKAL1		聚 (A) 结合蛋白相互作用蛋白 1	6	20534688	21232634
TACC3		磷酸酶和肌动蛋白调节剂 1	4	1723217	1746905

[0213]

ASXL2		CDK5 调节亚基相关蛋白 1 状 1	2	25962253	26101312
HORMAD 1		转化, 含酸性卷曲螺旋蛋白 3	1	150670535	150693364
PHLDA3		附加的性梳状转录调节剂 2	1	201434607	201438299
MIPOL1		含 1 的 HORMA 域	14	37667118	38020464
ZFR2		普列克底物蛋白同源状域, 家族 A, 成员 3	19	3804022	3869027
PIGH		镜像多指 1	14	68056023	68067017
WRB		锌指 RNA 结合蛋白 2	21	40752213	40769815
MRO		磷脂酰肌醇聚糖锚形体生物合成, H 类	18	48321490	48351754
STYX		富含色氨酸的碱性蛋白	14	53196883	53241705
MDFIC		Maestro	7	114562209	114659970
ERMN		丝氨酸/苏氨酸/酪氨酸相互作用蛋白	2	158175125	158184146
RND3		含有 ermin 的 MyoD 家族抑制剂域, ERM 状蛋白 Rho 家族 GTP 酶 3	3	151324707	151344209

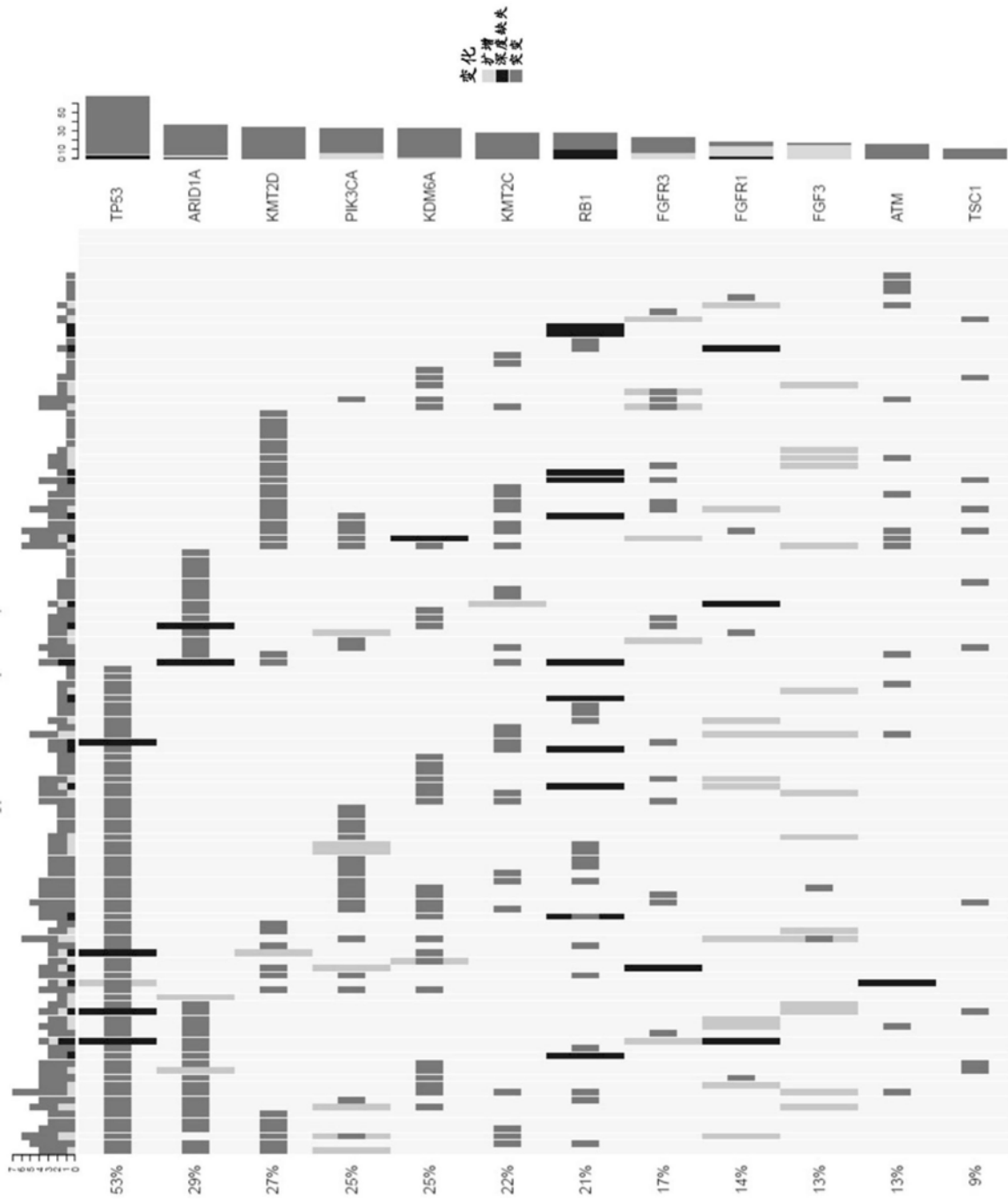


图1

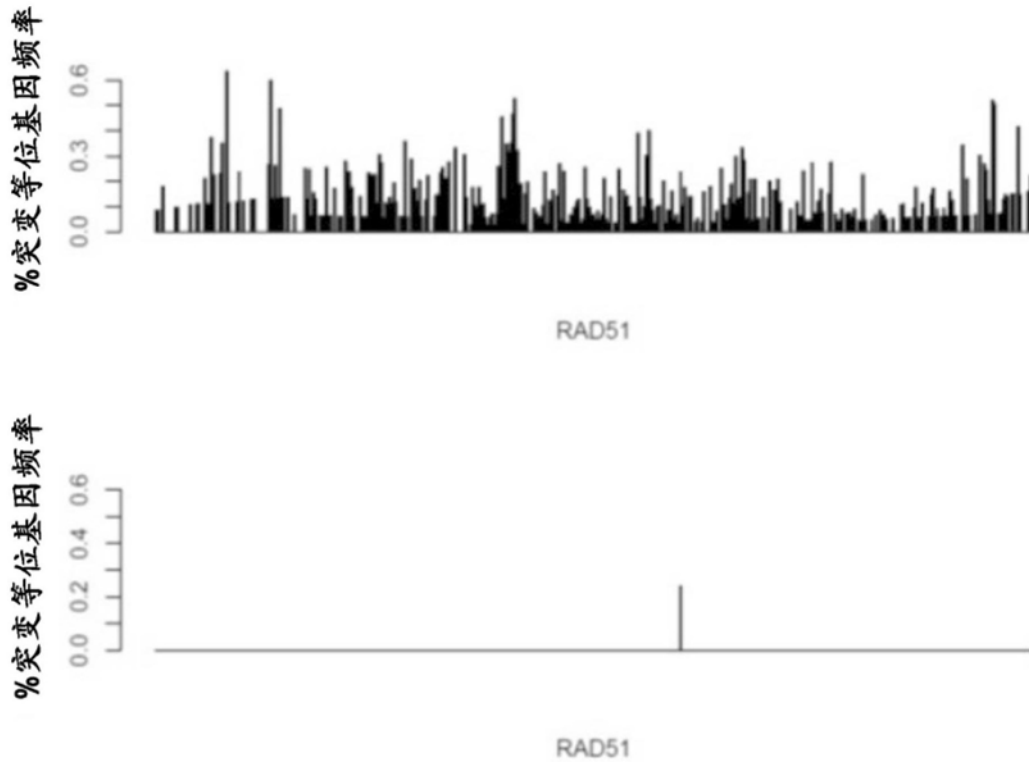


图2

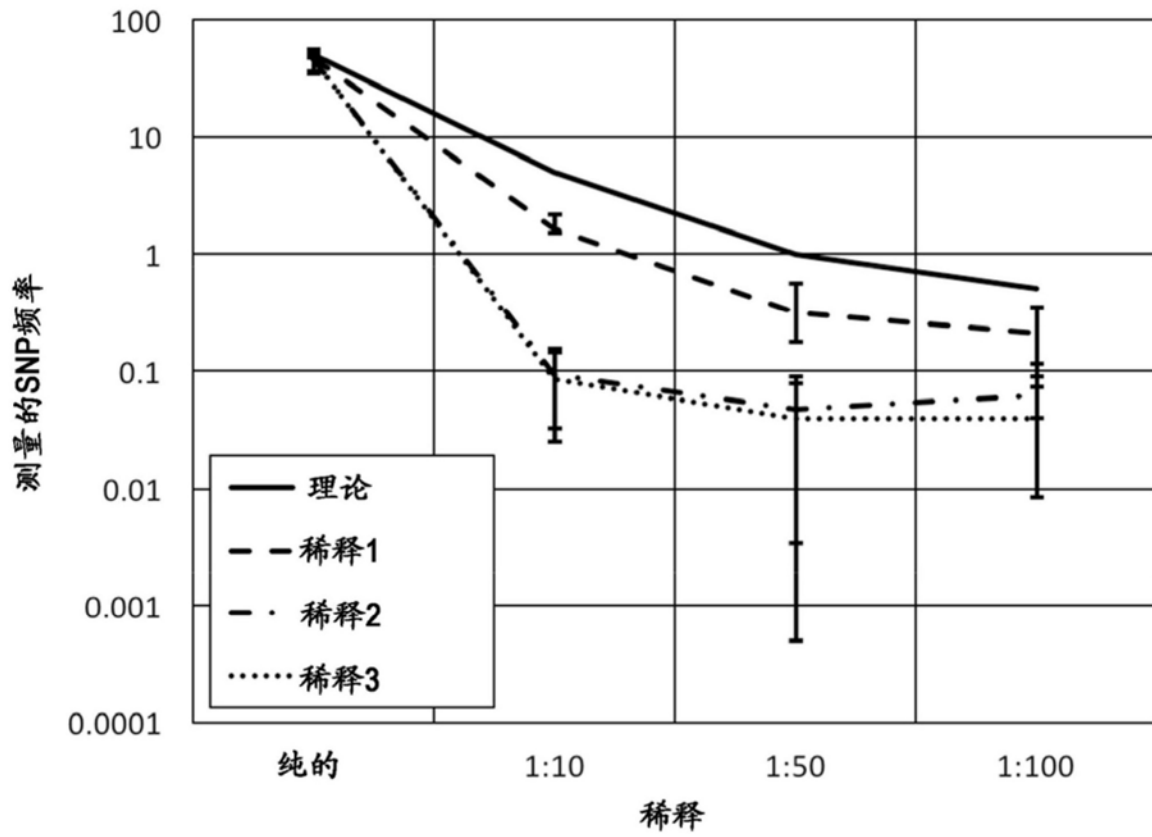


图3

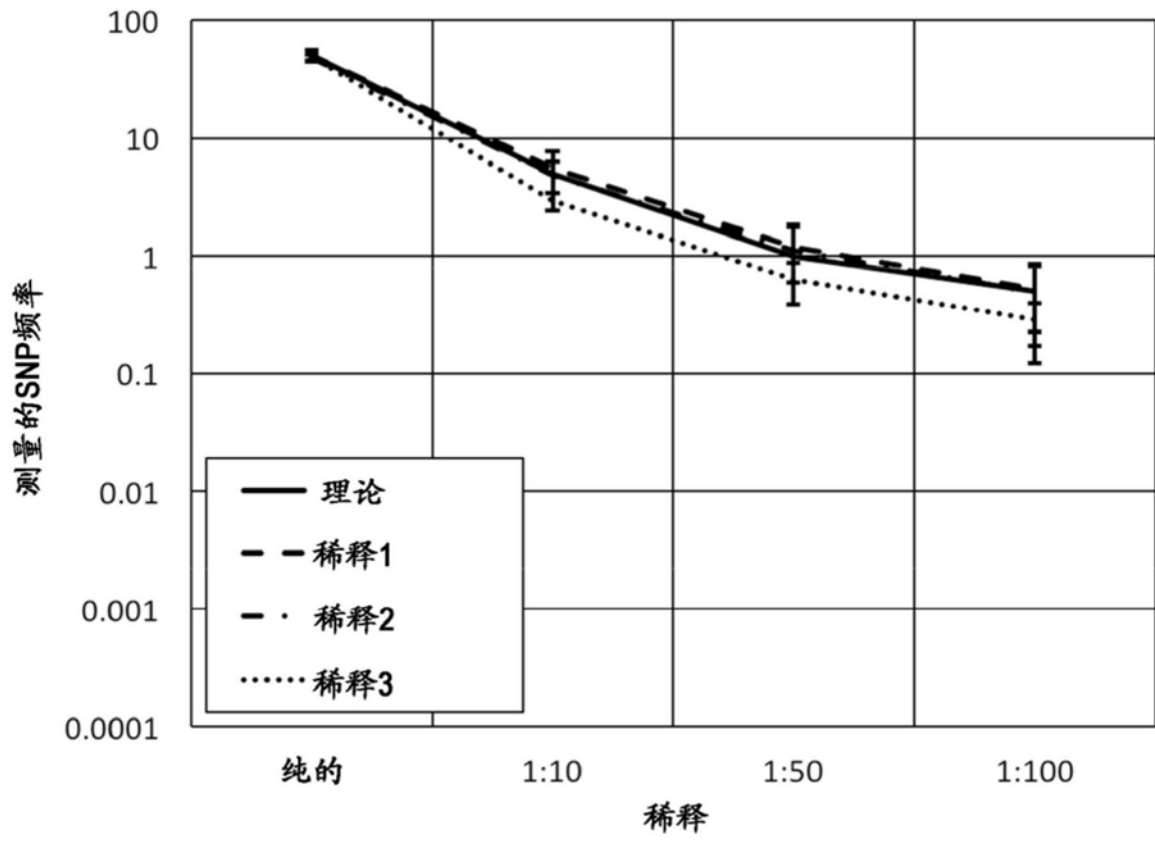


图4

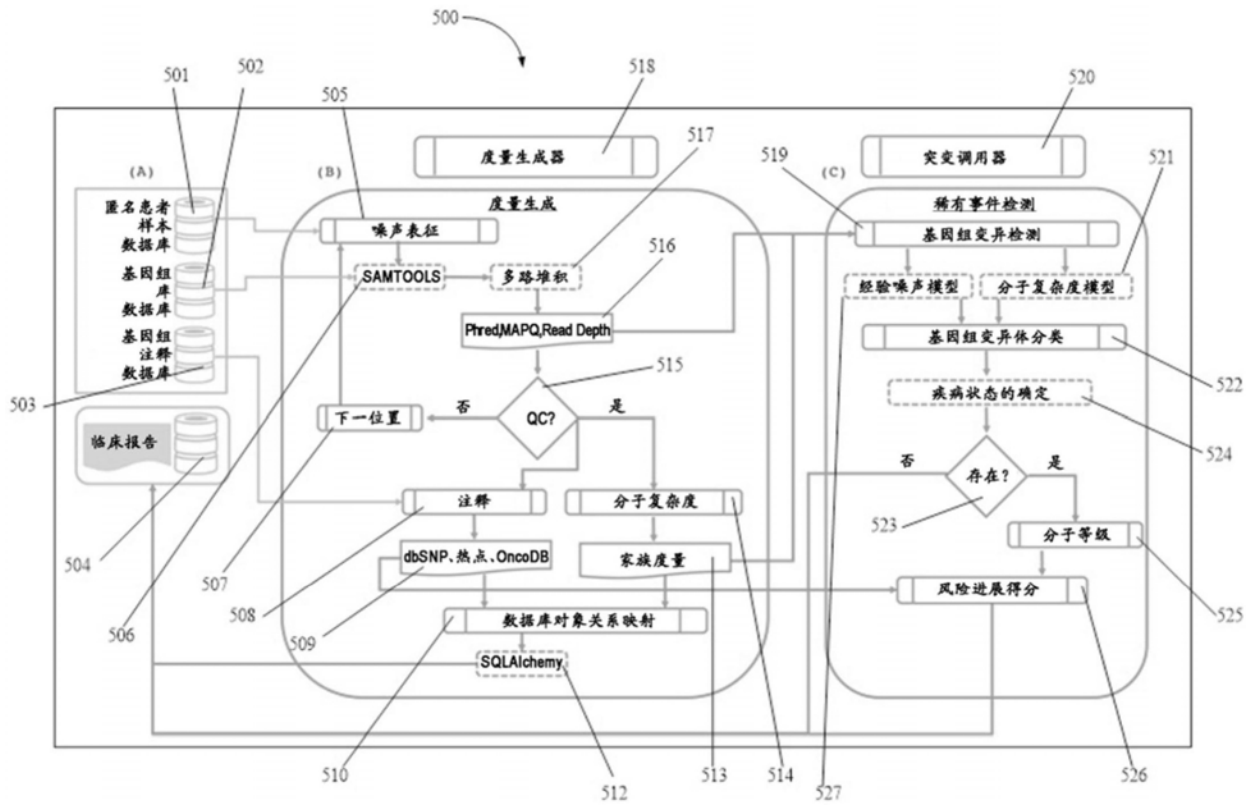


图5

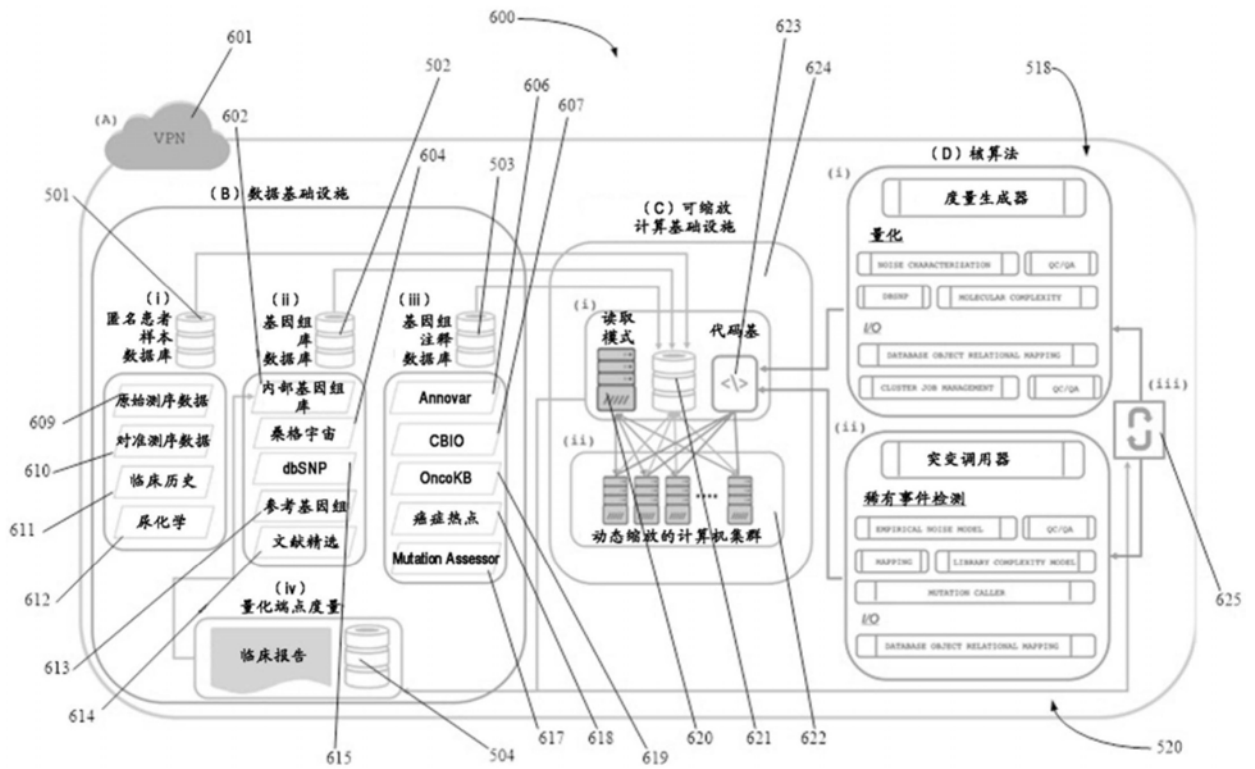


图6

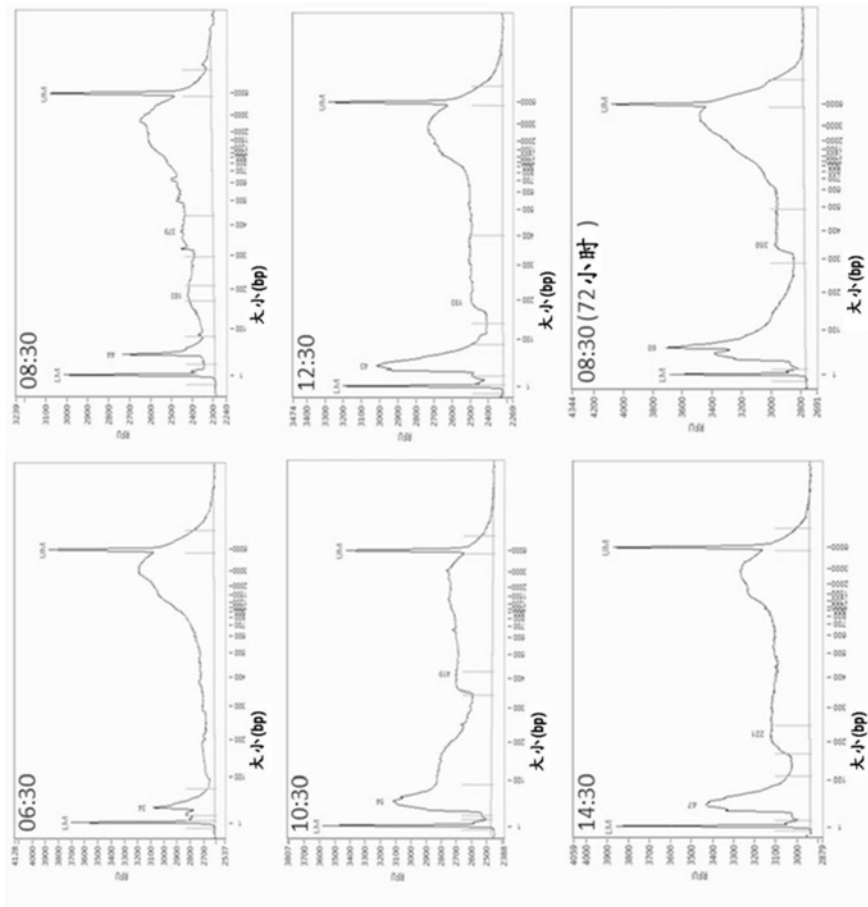


图7

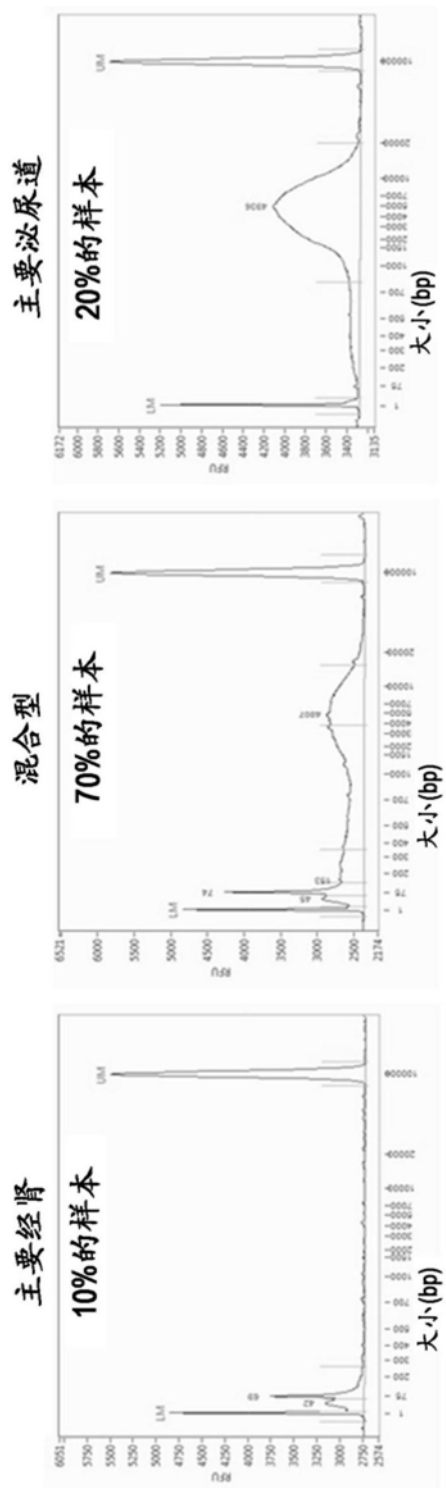


图8

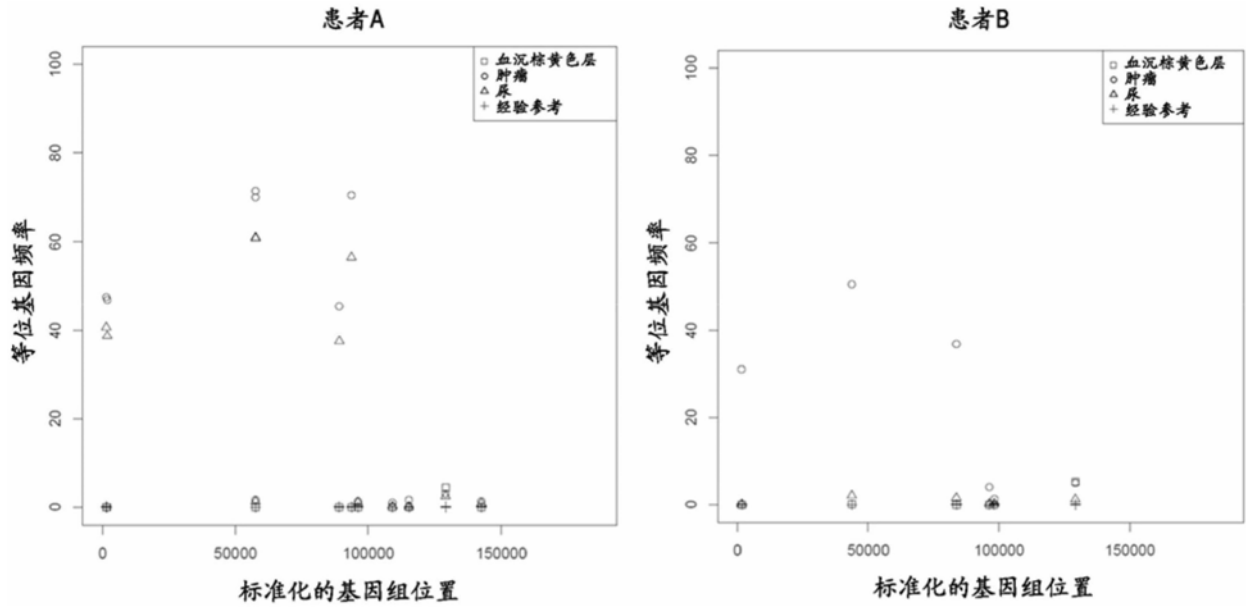


图9

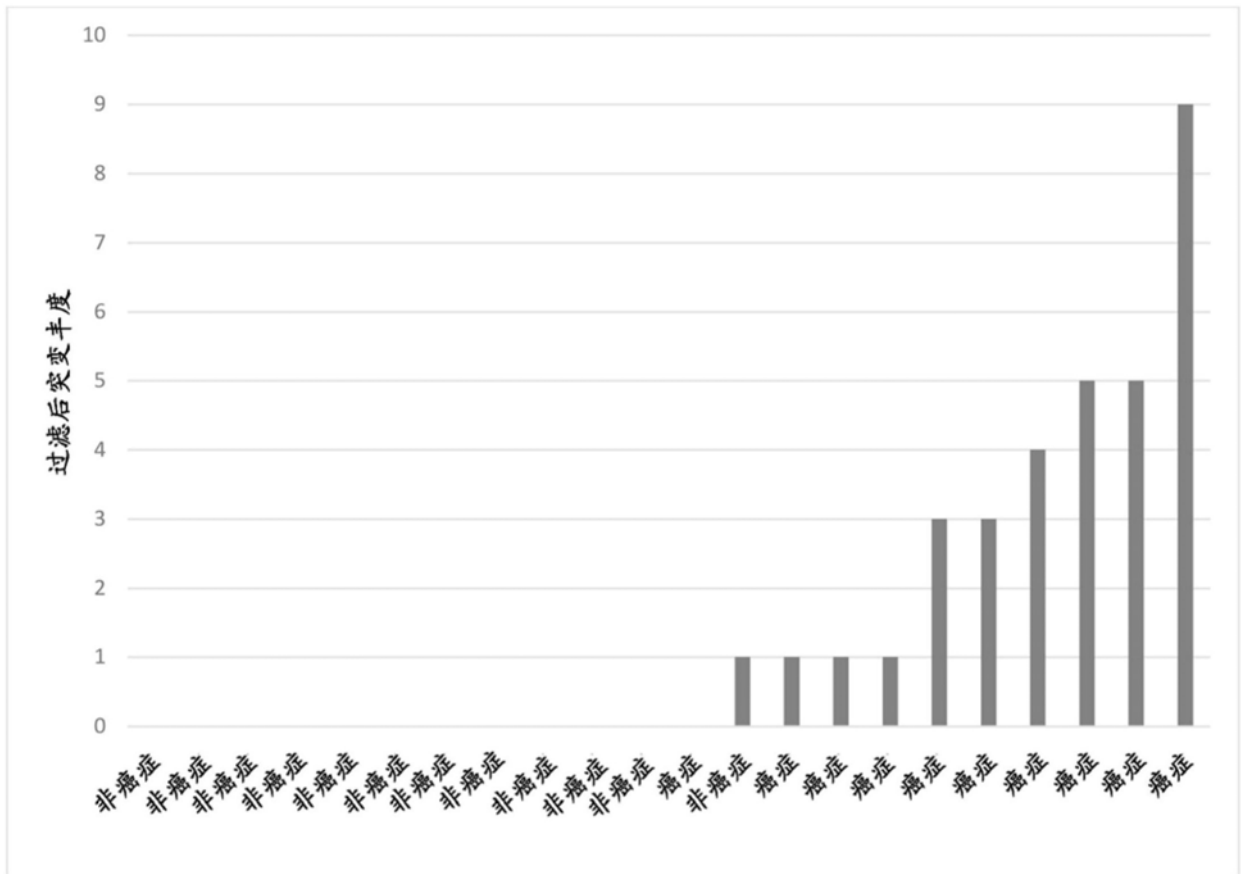


图10