



(12) 发明专利

(10) 授权公告号 CN 108368554 B

(45) 授权公告日 2022.09.09

(21) 申请号 201680069786.X

CN 103649335 A,2014.03.19

(22) 申请日 2016.09.29

WO 2010117817 A2,2010.10.14

(65) 同一申请的已公布的文献号
申请公布号 CN 108368554 A

George Wright 等.A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma.《Proceedings of the national academy of science of the united states of america》.2003,第100卷(第17期),第9991-9996页.

(43) 申请公布日 2018.08.03

(30) 优先权数据
62/234,491 2015.09.29 US

(85) PCT国际申请进入国家阶段日
2018.05.29

Alexandar Tzankov等.Multiparameter analysis of homogeneously R-CHOP-treated diffuse large B cell lymphomas identifies CD5 and FOXP1 as relevant prognostic biomarkers: report of the prospective SAKK 38/07 study.《Journal of hematology & oncology》.2015,第8卷(第70期),

(86) PCT国际申请的申请数据
PCT/US2016/054501 2016.09.29

(87) PCT国际申请的公布数据
WO2017/059108 EN 2017.04.06

(73) 专利权人 HTG分子诊断有限公司
地址 美国亚利桑那州

S Blenk等.Germinal center B cell-like (GCB) and activated B cell-like (ABC) type of diffuse large B cell lymphoma (DLBCL): analysis of molecular predictors, signatures, cell cycle state and patient survival.《Cancer Informatics》.2007,(第3期),第399-420页.

(72) 发明人 B·拉弗勒 Q·刘 J·W·吕克
P·C·罗氏

(74) 专利代理机构 北京华睿卓成知识产权代理
事务所(普通合伙) 11436
专利代理师 程淼

(51) Int.Cl.
G12Q 1/6886 (2018.01)

(56) 对比文件
CN 103649335 A,2014.03.19
CN 102224255 A,2011.10.19
WO 2015069790 A1,2015.05.14

审查员 娄菲

权利要求书4页 说明书68页
序列表4页 附图11页

(54) 发明名称
弥漫性大B细胞淋巴瘤(DLBCL)亚型分型的方法

DLBCL患者。还描述了可用于测量DLBCL标签基因的表达的阵列和试剂盒。

(57) 摘要
本文描述了用于分类DLBCL亚型的创新,以及将该分类结果用于诊断、预后以及治疗选择的创新。以此方式,该分类器可有效地分类DLBCL亚型并提供有意义的输出,有利于医疗实践和

CN 108368554 B

1. 由CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的表达值获得的概率评分在制备分类系统中的用途,所述分类系统用于分类弥漫性大B细胞淋巴瘤(DLBCL)的方法中,其中:

所述概率评分通过下述方法获得:

在从受试者获得的样品中测量CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的表达,由此获得CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的表达值,

为每个表达值加权,由此产生经加权的表达值,其中对每个表达值加权包括以下述系数乘以所述表达值:

对于CD47为-0.16,

对于CD86为0.83,

对于ENTPD1为-0.71,

对于FOXP1为-0.56,

对于FUT8为-0.30,

对于IL16为-0.51,

对于ITPKB为1.16,

对于LRMP为0.16,

对于MME为0.59,

对于NF2为-0.70,

对于PIM1为-0.13,

对于PTPRC为0.07,

对于REL为0.20,

对于STAT3为-0.17,

对于TNFRSF8为-0.01,

对于TYMS为-0.25,

对所述经加权的表达值求和,由此产生经加权的表达值总和,

使用所述经加权的表达值总和计算概率评分,其中采用下式计算概率评分:

$$Pr(Y = y|x) = \frac{e^{(\beta_0 + x^T \beta)}}{1 + e^{(\beta_0 + x^T \beta)}}$$

其中 β_0 为常数, β 为用于分类器基因的权重向量,以及 x^T 为针对给定样品的所述分类器基因的测量值向量;

所述分类弥漫性大B细胞淋巴瘤(DLBCL)的方法包括:

将所述概率评分与概率评分 t 截止值比较,所述概率评分 t 截止值为, < 0.43 用于针对ABC, > 0.57 用于针对GCB,以及 $0.43-0.57$ 用于针对未分类,

当所述概率评分为用于针对ABC的范围内的值时将所述DLBCL分类为活化B细胞样(ABC),

当所述概率评分为用于针对GCB的范围内的值时将所述DLBCL分类为生发中心B细胞样

(GCB), 或

当所述概率评分为用于针对未分类的范围内的值时将所述DLBCL分类为未分类;
其中测量表达为测量核酸表达。

2. 权利要求1的用途, 其中所述核酸为mRNA和/或miRNA。

3. 权利要求1至2中任一项的用途, 其中所述样品为经固定的样品。

4. 权利要求1至2中任一项的用途, 其中测量表达包括使用基于核酸酶的测定来分析所述样品。

5. 权利要求1至2中任一项的用途, 其中测量表达包括测序, 并且其中所述表达值为代表存在于所述样品中的分子数量的计数。

6. 权利要求1至2中任一项的用途, 其中测量表达包括对表达的定量。

7. 权利要求1至2中任一项的用途, 其中所述受试者为已知患有DLBCL或怀疑患有DLBCL。

8. 权利要求1至2中任一项的用途, 其中所述方法将少于6%的样品分类为未分类。

9. 权利要求1至2中任一项的用途, 其中所述表达值如下获得:

将所述样品与包括侧翼序列的核酸酶保护探针 (NPPF) 在足以让每个NPPF特异结合其靶核酸分子的条件接触;

将所述样品与包括与所述侧翼序列互补的序列的核酸分子 (CFS) 在足以让所述侧翼序列特异结合所述CFS的条件下接触;

将所述样品与特异于单链核酸分子的核酸酶在足以除去未结合核酸分子的条件接触, 由此产生经消化的样品, 所述经消化的样品包括杂交至所述靶核酸分子的NPPF和杂交至所述CFS的NPPF;

用扩增引物扩增所述经消化的样品中的NPPF, 由此产生NPPF扩增子;

对所述NPPF扩增子的至少一部分测序; 以及

对NPPF扩增子的数量计数; 由此确定所述样品中CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的表达值。

10. 权利要求1至2中任一项的用途, 其中所述表达值如下获得:

将所述样品与特异于CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS靶核酸分子的包括侧翼序列的核酸酶保护探针 (NPPF) 接触,

其中每个NPPF包括:

5'-末端和3'-末端,

与CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS靶核酸分子的区域互补的序列, 其允许所述NPPF和所述靶核酸分子之间特异结合,

其中所述侧翼序列位于与靶核酸分子互补的序列的5'、3'或两端, 其中5'-侧翼序列为所述与靶核酸分子互补的序列的5', 3'-侧翼序列为所述与靶核酸分子互补的序列的3',

其中所述侧翼序列包括未存在于所述样品内的核酸分子中发现的至少12个连续核苷酸,

如果所述NPPF包括5'-侧翼序列, 则将所述样品与包括与所述5'-侧翼序列互补的序列

的核酸分子(5CFS)、5'-末端磷酸酯在足以让所述5'-侧翼序列特异杂交所述5CFS的条件下接触;

如果所述NPPF包括3'-侧翼序列,则将所述样品与包括与所述3'-侧翼序列互补的序列的核酸分子(3CFS)在足以让所述3'-侧翼序列特异杂交所述3CFS的条件下接触;

其中所述3CFS和所述5CFS中的至少一个包括捕获部分;

产生杂交至所述靶核酸分子、杂交至所述3CFS、杂交至所述5CFS或者杂交至所述3CFS和所述5CFS二者的NPPF;

将所述样品与特异于单链核酸分子的核酸酶在足以除去未结合核酸分子的条件接触,由此产生经消化的样品,所述经消化的样品包括杂交至所述靶核酸分子、杂交至所述3CFS、杂交至所述5CFS或者杂交至所述3CFS和所述5CFS二者的NPPF;

捕获所述杂交至所述靶核酸分子、杂交至所述3CFS、杂交至所述5CFS或者杂交至所述3CFS和所述5CFS二者的NPPF;

将所述3CFS的5'-磷酸酯连接所述靶核酸分子的3'-末端,将所述5CFS的3'-末端连接所述靶核酸分子的5'-末端,由此产生经连接的靶核酸分子;

将所述NPPF与所述经连接的靶核酸分子分开,由此产生包括单链NPPF和单链经连接的靶核酸分子的混合物;以及

对所述单链经连接的靶核酸分子的至少一部分测序,由此确定所述样品中CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS靶核酸分子的序列。

11. 权利要求10的用途,其中所述NPPF包括至少一个dUTP,并且所述方法还包括在变性后和测序前将包括单链NPPF和单链经连接的靶核酸分子的所述混合物与尿嘧啶DNA脱糖基化酶(UDG)在足以降解所述单链NPPF的条件下接触。

12. 权利要求1至2中任一项的用途,其中所述分类弥漫性大B细胞淋巴瘤(DLBCL)的方法还包括,当所述DLBCL分类为GCB时,以治疗有效量的(1)环磷酰胺、多柔比星、长春新碱以及泼尼松或泼尼松龙(CHOP)化疗剂,(2)利妥昔单抗加CHOP(R-CHOP)化疗剂,或(3)依托泊苷加R-CHOP(R-EPOCH)化疗剂施用于所述受试者。

13. 权利要求1至2中任一项的用途,其中所述分类弥漫性大B细胞淋巴瘤(DLBCL)的方法还包括,当所述DLBCL分类为ABC时,以治疗有效量的苯达莫司汀、匹杉琼、吉西他滨/奥沙利铂、脂质体长春新碱、抗CD20 mAb、抗CD22 mAb、抗CD74 mAb、抗CD40 mAb、单链双特异性抗CD19和CD3 mAb构建体、I-131托西莫单抗、奥英妥珠单抗、90Y-替坦司依帕珠单抗、沙利度胺、来那度胺、硼替佐米、NPI-0052、依维莫司、替西罗莫司、伏林司他、奥利默森钠、PF-3512676、17-AAG、贝伐珠单抗、阿柏西普、CAL-101、丙戊、迪那西里、福斯坦替尼、达沙替尼、恩扎妥林、PCI-32765、SB1518和索拉非尼中的一种或更多种施用于所述受试者。

14. 一种或更多种计算机可读存储介质,其存储用于使计算系统在被编程时执行权利要求1至11中任一项中的获得概率评分的方法和分类弥漫性大B细胞淋巴瘤(DLBCL)的方法的计算机可执行指令。

15. 包含权利要求14所述的计算机可读存储介质的计算系统。

16. 用于权利要求1-11任一项中的获得概率评分的方法和分类弥漫性大B细胞淋巴瘤(DLBCL)的方法的试剂盒,包括:

包括至少16种不同NPPF和相应CFS的容器,其中所述至少16种不同NPPF特异于 CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和 TYMS;以及

以下的一种或更多种:

包括可特异结合所述至少16种不同NPPF的扩增子的微珠的容器;

包括特异于单链核酸的核酸酶的容器;

包括裂解缓冲液的容器;

包括洗涤缓冲液的容器;

包括用于PCR的试剂的容器;

包括乙醇的容器;

包括连接酶的容器;

包括连接缓冲液的容器;

包括变性油的容器;以及

包括蛋白酶K的容器。

17. 权利要求16的试剂盒,其中所述试剂盒包括:

所述包括至少16种不同NPPF和相应CFS的容器,其中所述至少16种不同NPPF特异于 CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和 TYMS;

包括特异于单链核酸的核酸酶的容器;

包括裂解缓冲液的容器;以及

包括核酸酶终止缓冲液的容器。

18. 权利要求16或17的试剂盒,还包括实现用于DLBCL亚型的分类器的计算系统。

19. 权利要求18的试剂盒,其中所述计算系统包括软件或计算机可读介质,其接收两个或更多个DLBCL标签基因的表达值,针对每个基因的相应阈值为所述多个值评分,以及在指示样品DLBCL亚型的框架中对样品分类。

弥漫性大B细胞淋巴瘤 (DLBCL) 亚型分型的方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2015年9月29日提交的美国临时申请No. 62/234,491的优先权,通过引用将其并入本文。

技术领域

[0003] 本公开涉及用于区分弥漫性大B细胞淋巴瘤 (DLBCL) 亚型的分类器(classifier)、阵列和试剂盒,以及将该分类结果用于诊断、预后和/或治疗选择。

背景技术

[0004] 淋巴瘤是最常见的血癌。两种主要形式的淋巴瘤是霍奇金淋巴瘤 (Hodgkin lymphoma) 和非霍奇金淋巴瘤。弥漫性大B细胞淋巴瘤(diffuse large B-cell lymphoma, DLBCL) 是最常见形式的非霍奇金淋巴瘤,在新诊断病例占多至30%。DLBCL可以基于基因表达标志物进一步表征为生发中心B细胞样 (GCB) 亚型或非GCB样(包括活化B细胞样(ABC)) 亚型。GCB和非GCB(例如ABC) 亚型分型可以用于(i) 预测淋巴瘤患者的预后,GCB样淋巴瘤比非GCB(或ABC) 亚型有更佳预后;和/或(ii) 确定对淋巴瘤患者的治疗,例如采用GCB或 ABC 靶向疗法,或者基于疾病侵袭性来确定(ABC通常比GCB病更有侵袭性)。

[0005] 对于区分GCB和非GCB(例如,ABC) DLBCL亚型和/或对于确定DLBCL患者的预后或对DLBCL患者的处理,没有单一的、可接受的诊断试验。在临床实践中,存在数种基于免疫组化(IHC) 的测试,包括Colomo等, (Blood 101(1):78-84,2003), Hans等, (Blood 103:275-282, 2004), Muris等 (J.Pathol.208(5):714-23,2006), Choi等 (Mod.Pathol.21:250A, 2008), 以及Tally (Meyer等, J.Clin.Oncol.29(2):200,2011) 测试。

[0006] 由于DLBCL具有各种各样的临床结果,因此认为该疾病的分子基础同样复杂,对临床上有用的生物标志物的发现正在进行中。此类工作的一个目标是“将临床异质性诊断类别细分为具有更佳均匀性临床行为的分子上独特的疾病”(Alizadeh等, Nature 403:503, 2000)。在这方面,至少由于仅可同时检测到(co-detected) 有限数量的蛋白标志物的原因,IHC是具有固有局限性的技术。类似地,基因组分析(例如,DNA) 受限至少是因为不清楚基因组事件是否表达为对该生物学系统有功能性影响。

[0007] mRNA转录组(mRNA transcriptome) 相对于蛋白和DNA分析具有优势,因为mRNA表达提供了对基因组活性的快照,并且适合于相同样品中很多生物标志物的多重检测。已有报道,数种mRNA在DLBCL样品中差异表达(例如,Alizadeh等, Nature, 403:503, 2000; Rosenwald等, NEJM 346(25):1937, 2002; Wright等, Proc.Natl.Acad.Sci.100:9991, 2003; 以及Roberts等, Lab.Invest.78:979, 2007)。但是,该领域中仍有重大差距仍有待解决。

发明概要

[0008] 对跨各种检测技术的生物标志物(其表达为DLBCL特征性的) 的鉴定为重要的第一步(例如,Roberts等, Lab.Invest.78:979, 2007)。然而,从实践和临床视角来看,仍然存在

的关键性挑战为将大量的神秘表达数据 (arcane expression data) 合成有意义的输出, 以有利于医疗实践和DLBCL患者。

[0009] 本公开描述了用于分类DLBCL亚型的创新, 以及将该分类结果用于诊断、预后以及治疗选择的创新。例如, 分类器(其可为计算机可实现的) 基于基因表达, 例如来自DLBCL受试者的样品中两个或更多个标签基因的表达水平, 来确定DLBCL的亚型。在一些实例中, 表达通过给定基因的测序读取数 (read) 占读取数总数的比例数来测量(其中读取数为可通过测序仪读取的(例如来自探针的) 特定序列的互补性拷贝的数量)。该分类器可有效地分类DLBCL亚型并提供有意义的输出, 有利于医疗实践和DLBCL患者。

[0010] 根据本创新的一个方面, 该方法测量样品中多个DLBCL标签基因表达水平(例如, 核酸或蛋白)。例如, 该方法可测量CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中至少两个, 或者CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS 中至少两个的表达(例如测量这些基因全部16个的表达)。可对每个基因获得的表达值加权, 并对加权的表达值求和。可确定采用该经加权的表达值总和的概率评分。该方法或分类器相对于用于该DLBCL分类器的相应阈值来评估或比较该概率评分。然后, 至少部分地基于该评分结果, 该方法将该DLBCL样品分类为活化B细胞样(ABC)、生发中心B细胞样(GCB) 或未分类。对DLBCL亚型的分类可作为方法的一部分、作为适于执行该方法的计算系统的一部分或作为有形的计算机可读介质(存储使计算系统执行该方法的计算机可执行指令)的一部分来实现。

[0011] 本文描述的创新还包括将该分类结果用于DLBCL患者的诊断、预后和/或治疗选择。本文还描述了可用于确定DLBCL标签基因的表达的阵列和试剂盒。

[0012] 根据以下参照附图进行的详细描述, 本公开的前述和其它特征将变得更加明显。

[0013] 附图简要说明

[0014] 图1为阐述用于分类DLBCL亚型的总体技术的流程图。

[0015] 图2为显示用于DLBCL分类器模型训练和验证的病例的流程图。

[0016] 图3为显示包括在DLBCL测定中用于训练和验证分类器实施方案的96个基因跨样品表达分布的图。

[0017] 图4为显示为未分类样品确定下限的手段的手段图。

[0018] 图5为显示DLBCL样品的验证集合的分类(即, ABC、GCB或未分类)的图, 这些 DLBCL样品事先已通过临床上接受的方法表征为ABC(黑点)或GCB(红点)。样品沿x轴标识。y轴显示基于基因表达谱的针对每个病例的预测概率。在本实施方案中, 用于针对GCB样品的估计概率的截止值为0.57(预测概率 >0.57 为GCB), 用于针对ABC分类的截止值为低于0.43(预测概率 <0.43), 以及未分类样品落入中间并包括该ABC和GCB截止值($0.43 \leq \text{预测概率} \leq 0.57$)。

[0019] 图6为显示从GCB FFPE样品制备的裂解物中用于CD274(PDL-1)和PDCD1(PD1)的Log2CPM的图。

[0020] 图7为显示从ABC FFPE样品制备的裂解物中用于CD274(PDL-1)和PDCD1(PD1)的Log2CPM的图。

[0021] 图8为显示针对每个重复和条件的分类图的图, 图9A-D为本发明的示例性实施方案的示意图。

[0022] 序列表

[0023] 本文列出的核酸序列采用用于核苷酸碱基的标准字母缩写来显示,如37C.F.R.1.822中所定义的。仅显示了每条核酸序列的一条链,但理解为互补链被包含在对所显示链的任何引用中。随本文提交的序列表通过引用并入(2016年9月1日生成,3.62KB)。在提供的序列中:

[0024] SEQ ID NO:1-16为示例性探针核酸序列。

[0025] 发明详细描述

[0026] 除非另有说明,否则技术术语按常规用法使用。对分子生物学中常见术语的定义可以在牛津大学出版社出版的Benjamin Lewin, Genes V, 1994 (ISBN 0-19-854287-9); Blackwell Science Ltd.出版的Kendrew等,(eds.), The Encyclopedia of Molecular Biology, 1994 (ISBN 0-632-02182-9); 以及VCH Publishers, Inc.出版的Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, 1995 (ISBN 1-56081-569-8)中找到。

[0027] 单数形式的“a”、“an”和“the”包括复数指代物,除非上下文另有明确说明。类似地,用词“or”旨在包括“and”,除非上下文另有明确说明。“包括(comprising)”意指“包含(including)”。因此,“包括A或B”意指包含A,或包含B,或包含A和B。还应该理解的是,给出的核酸或多肽的所有碱基大小或氨基酸大小以及所有分子量或分子量值是近似的,并提供用以进行描述说明。尽管与本文中所描述的那些相类似或者相等同的方法和材料可用于实施或试验本公开,但下文则描述了适合的方法和材料。通过引用将本文提及的所有出版物、专利申请、专利以及其它参考文献以其全文并入, GenBank®登记号也是如此(用于2015年9月29日提交的序列)。如有矛盾,以包括术语解释的本说明书为准。此外,所述材料、方法以及实施例仅是阐释性的,而无意进行限制。

[0028] 除非另有说明,本公开的方法和技术一般按照本领域公知的以及如本说明书通篇引用和讨论的各种通用和更专业的参考文献中所描述的各种常规方法实施。参见,例如, Sambrook 等, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press, 1989; Sambrook等, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press, 2001; Ausubel等, Current Protocols in Molecular Biology, Greene Publishing Associates, 1992(以及至2000的增补); Ausubel等, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, 4th ed., Wiley&Sons, 1999; Harlow和Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1990; 以及Harlow和Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1999; 通过引用将它们每个明确地以全文并入本文。

[0029] 概述

[0030] DLBCL为成人中最常见类型的非霍奇金淋巴瘤。有两种主要的DLBCL亚型:生发中心B细胞样(GCB)和非GCB(典型地,活化B细胞样(ABC))。GCB一般与有利的预后相关,而ABC或非GCB一般与较差的预后(例如,为更具侵袭性的疾病)相关。原发性纵膈B细胞淋巴瘤(PMBCL)被一些从业者认为是另一种亚型的DLBCL(Lossos, J. Clin. Oncol. 23:6351-6357, 2005)。PMBCL在胸腺中出现,并一般表现为纵膈中的肿块。DLBCL的非GCB亚型包括除GCB亚

型外的所有DLBCL亚型。类似地,DLBCL的非ABC亚型包括除 ABC亚型外的所有DLBCL亚型。

[0031] 本文描述了用于分类DLBCL亚型的创新,以及将该分类结果用于诊断、预后和/或治疗选择的创新。示例分类器基于来自DLBCL受试者的样品中两个或更多个标签基因的基因表达水平来确定DLBCL亚型。通过有效地确定DLBCL亚型,分类器提供有意义的输出,有利于医疗实践和DLBCL患者。

[0032] 本文描述的创新包括分类方法(其可全部或部分地用计算机实现),存储用于执行这些方法的计算机可执行指令的计算机可读介质,以及适于执行这些方法的计算系统。本文描述的创新还包括将这种分类的结果用于预后、治疗选择和/或其它目的,以及阵列和试剂盒,它们提供量化标签基因的表达水平的值作为对分类的输入。

[0033] 本文描述的实例的各种替代方案是可能的。例如,可通过改变所描述的方法行为的顺序,通过分割、重复或省略某些方法动作等来改变本文所描述方法中的任一个。所公开的技术的各种方面可以组合或独立使用。不同的实施方案使用所描述的创新中的一个或更多个。本文描述的一些创新解决背景技术中提到的一个或更多个问题。通常,给定的技术/工具不解决所有的这些问题。

[0034] DLBCL亚型分型的方法

[0035] 本文公开了基于来自患有DLBCL的受试者的样品中两个或更多基因(例如两个或更多 DLBCL标签基因)的基因表达来确定DLBCL(例如,GCB或非GCB(例如ABC)亚型)亚型的方法。通常,对每个评估的标签基因的表达数据赋值(例如基于或表示为占给定样品中总读取数的比例量)。对赋予给所评估的标签基因的所述值进行加权,并进行组合以提供评分(或概率评分)。然后基于截止值、阈值或其它标准由该评分确定DLBCL的亚型(例如,GCB 或非GCB的判定)。

[0036] 在一个实例中,本公开提供了对弥漫性大B细胞淋巴瘤(DLBCL)分类的方法。这些方法可包括在得自受试者的样品(例如经固定的样品,例如FFPE样品)中直接或间接地检测或测量多个DLBCL标签基因的表达(例如核酸表达或蛋白表达),该多个DLBCL标签基因例如为CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的至少两个、至少三个、至少四个或全部,或者为CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS 中的至少两个、至少三个、至少四个或全部。因此,获得每个DLBCL标签基因的表达值(其可为定量的),例如CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8中至少两个(或全部)或CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中至少两个(或全部)的表达值。可对每个DLBCL标签基因所获得的表达值加权(例如通过乘以系数值,例如表5中显示的那些),因而产生至少两个经加权的表达值。对所得的经加权的表达值相加或求和,由此产生相加的经加权的表达值。使用该相加的经加权的表达值(例如,在概率模型中)来计算概率评分,该概率评分与阈值或截止值比较。在一个实例中,采用下式来计算概率评分

$$[0037] \quad Pr(Y = y|x) = \frac{e^{(\beta_0 + x^T \beta)}}{1 + e^{(\beta_0 + x^T \beta)}} \quad (\text{等式 1})$$

[0038] 其中 β_0 为常数, β 为标签基因(例如表1中的一个标签基因)的权重, x^T 为该标签基因的值。在一个实例中,当该概率评分低于为ABC确立的评分时将该DLBCL样品分类为活化B细胞样(ABC),当该概率评分高于为GCB确立的截止值时分类为生发中心B细胞样(GCB);或

者当该概率评分为未分类评分范围内的值时,分类为未分类,该未分类评分范围典型地介于用于针对ABC的截止值评分和用于针对GCB的截止值评分之间并包括这两个截止值。

[0039] 在一些实例中,测量核酸分子如mRNA的表达。在一些实例中,测量蛋白分子的表达。在一些实例中表达测量为定量的或半定量的。在一些实例中,测量表达采用测序,并且其中每个DLBCL标签基因的表达值是代表样品中靶核酸(例如,mRNA)/蛋白的(相对和绝对)数量的计数。

[0040] 在一些实例中,所公开的方法和DLBCL分类器将少于6%,例如少于5%的样品分类为未分类,或者将4至6%分类为未分类。相比之下,其它的DLBCL分类器(例如测量这12个基因表达的那些:BCL6、CCND2、ENTPD1、FOXP1、FUT8、IRF4、ITPKB、LMO2、LRMP、MME、MYBL1以及SERPINA9)具有更高比例的未分类样品,例如至少10%、至少11%或至少12%,例如10至12%。因此,所公开的DLBCL分类器使得约4至8%的样品(采用之前的方法为未分类)现在分类为ABC或GCB。在一些实例中,所公开的方法以至少约90%例如至少92%的准确度对DLBCL亚型分型(例如,为GCB或非GCB,例如ABC)。

[0041] 在一些实例中,这些方法采用基于核酸酶的测定法测量表达。例如,可通过让样品(1)与至少两个包含侧翼序列的核酸酶保护探针(NPPF)在足以让每个NPPF特异结合至其靶核酸分子(例如,DLBCL标签基因)的条件下接触(例如,置于直接物理联系中),(2)与包括与侧翼序列互补的序列的核酸分子(CFS)在足以让侧翼序列特异结合至CFS的条件下接触,以及(3)与特异于单链核酸分子的核酸酶在足以去除未结合的核酸分子的条件下接触,来获得该表达值。用在本文时,“足以.....的条件”指允许所期望的活性的任何环境(例如,水性,包括或略去 Na^+ 、 Cl^- 、 Mg^{++} 、 Zn^{++} 或其它特定离子、离液剂或表面活性剂,或具有特定浓度的任何这些离子、离液剂或表面活性剂,或具有特定离子强度、pH、缓冲类型或浓度,或温度),例如,允许两核酸分子(例如探针和靶核酸)之间特异结合或杂交,或允许核酸酶去除(或消化)未结合的核酸分子。这导致得到经消化的样品,其包括与靶核酸分子和CFS杂交的NPPF。

[0042] 任选地,未结合的或单链分子被从样品中去除。用一个或多个适合的扩增引物扩增与靶和CFS杂交的NPPF,由此产生NPPF扩增子。对至少部分NPPF扩增子测序(例如,以确定其杂交至哪个靶核酸分子(例如,DLBCL标签基因))。可对NPPF扩增子的数量计数或赋值,由此确定样品中CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8的至少两个(或全部)或者CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的至少两个(或全部)中每个的表达值。因此,该测定中使用的NPPF使得能检测并且在一些实例中能定量靶核酸分子,例如mRNA。

[0043] 所公开的方法还可包括根据该DLBCL分类来施用治疗有效量的适当治疗剂。例如,当DLBCL分类为GCB时,可以以治疗有效量的(1)环磷酰胺(cyclophosphamide)、多柔比星(doxorubicin)、长春新碱(vincristine)以及泼尼松(prednisone)或泼尼松龙(prednisolone)(CHOP)化疗剂,(2)利妥昔单抗(ituximab)加CHOP(R-CHOP)化疗剂,或(3)依托泊苷(etoposide)加R-CHOP(R-EPOCH)化疗剂施用于受试者。在另一实例中,当DLBCL分类为ABC或非GCB时,可以以治疗有效量的苯达莫司汀(bendamustine)、匹杉琼(pixantrone)、吉西他滨/奥沙利铂(gemcitabine/oxaliplatin)、脂质体长春新碱(liposomal vincristine)、抗CD20 mAb、抗CD22 mAb、抗CD74 mAb、抗CD40 mAb、单链双特

异性抗CD19和CD3 mAb构建体、I-131托西莫单抗 (tositumomab) (抗CD20放射免疫疗法)、奥英妥珠单抗 (Inotuzumab ozogamicin) (CMC-544) (靶向CD22的细胞毒性免疫缀合物)、90Y-替坦司依帕珠单抗 (90Y-epratuzumab tetraacetan, 放射标记的抗CD22mAb)、本妥昔单抗 (Brentuximab vedotin) (SGN-35) (抗微管的单甲基阿里他汀 (auristatin) E抗CD30 mAb 缀合物)、沙利度胺 (thalidomide)、来那度胺 (lenalidomide)、硼替佐米 (Bortezomib) (蛋白酶体抑制剂)、NPI-0052 (蛋白酶体抑制剂)、依维莫司 (Everolimus) (mTOR抑制剂)、替西罗莫司 (Temsirolimus) (mTOR抑制剂)、伏林司他 (Vorinostat) (脱乙酰酶抑制剂)、奥利默森钠 (Oblimersen sodium) (Bcl-2反义寡核苷酸)、PF-3512676 (TLR9拮抗剂)、17-AAG (HSP90抑制剂)、贝伐珠单抗 (Bevacizumab, 抗VEGF mAb)、阿柏西普 (Aflibercept) (VEGF融合蛋白)、CAL-101 (PI3K抑制剂)、丙戊酸 (HDACI)、迪那西里 (Dinaciclib) (CDK1、2、5、9抑制剂)、福斯坦替尼 (Fostamatinib) (Syk抑制剂)、达沙替尼 (Dasatinib) (BCR-ABL、SRC、c-Kit、PDGF和蝶素 (ephrin) 受体激酶的RTK抑制剂)、恩扎妥林 (Enzastaurin) (蛋白激酶β抑制剂)、PCI-32765 (Bruton酪氨酸激酶抑制剂)、SB1518 (JAK2抑制剂)、索拉非尼 (Sorafenib) (RAF/MEK/ERK/c-kit/Flt3、VEGFRs、PDGFRs、RET的TKI抑制剂) 或本领域已知的其它治疗剂 (参见, 例如, Foon等, Adv. Hematology, Article ID 302570, 2012, doi: 10.1155/2012/302570) 施用于受试者。

[0044] 还提供了存储计算机可执行指令的一个或多个计算机可读存储介质, 该计算机可执行指令用于在被编程时使计算系统执行本文提供的DLBCL分类方法。还提供了适用于执行本文提供的DLBCL分类方法的计算系统。

[0045] 提供了可用于所公开的DLBCL分类方法的试剂盒。在一些实例中, 这些试剂盒具有包括至少两种不同NPPF和对应CFS的容器, 其中该至少两种不同NPPF特异于CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3或TNFRSF8中的两个或多个 (例如全部), 或特异于CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或多个 (例如全部)。在一些实例中, 这些试剂盒具有包括至少两种不同NPP的容器, 其中该至少两种不同NPP特异于 CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3或TNFRSF8中的两个或多个 (例如全部), 或特异于CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或多个 (例如全部)。该试剂盒还可包括以下的一种或更多种: 包括可特异结合该至少两种不同NPPF的扩增子的微珠 (bead) 的容器; 包括特异于单链核酸核酸的核酸酶的容器; 包括裂解缓冲剂的容器; 包括可中和 (neutralize) 该核酸酶的缓冲剂, 例如可将pH升高到6以上的缓冲剂的容器; 包括洗涤缓冲剂的容器; 包括用于PCR的试剂的容器; 包括乙醇的容器; 包括变性油 (denaturation oil) 的容器; 包括连接缓冲剂的容器 (例如, 包括连接酶的容器); 以及包括蛋白酶K的容器。在一些实例中, 该试剂盒还包括特异于NPP或NPPF的一部分的引物, 例如可在NPP或NPPF或CFS的3' - 或5' - 末端或两端添加实验标签和/或测序接头。在一些实例中, 该试剂盒还包括执行用于DLBCL亚型的分类器, 例如软件或计算机可读介质, 其接收两个或多个DLBCL标签基因的表达值, 针对每个基因的相应阈值为该多个值评分, 以及在指示样品DLBCL亚型的框架中对样品分类。

[0046] 图1显示了用于分类DLBCL样品的通用方法或技术 (100)。该方法 (100) 的至少一些操作可通过计算系统来进行, 例如通过专用诊断工具、台式计算机、便携式计算机、平板或

板式计算机、智能手机或其它移动计算设备。这样的计算系统可包括读取器(例如,网络连接、存储设备界面、直接接收用户输入的用户界面,或接收定量基因表达水平的值的其它界面),其接收定量样品中DLBCL标签基因的基因表达水平的值;分类器,其针对用于分类的阈值对基因表达值评分并至少部分地基于评分结果对样品分类;数据存储(例如,移动或非移动存储器,包括磁盘、磁带或盒式磁带、CD-ROM、DVD、或可用于非暂时性存储信息(例如测量的DLBCL标签基因的基因表达水平、用于运行分类器的软件、截止值)并在该计算系统中可读取的任何其它介质);以及输出模块(例如,声音输出设备、显示器、触摸屏、打印机、CD写入器,或其它从计算系统提供输出(例如关于所测试样品为ABC、GCB或未分类的决定)的设备)。该体系还可包括计算数据存储中存储的阈值或截止值的训练模块。该分类器可包括数据处理器(例如,将每个基因表达值与特定的权重值相乘并向评分求和器(score totaller)提供经加权的值)、评分求和器(将计算的经加权的值相加并将它们提供至概率模块),以及概率模块(例如,采用该经加权的值的总和确定特定分类的概率,并将该概率或DLBCL亚型提供至输出模块)。例如,该概率模块可将计算的概率值与对应的用于针对ABC、GCB以及未分类的阈值比较。分类器从存储了用于DLBCL分类器的阈值的数据存储获得概率阈值/截止值。作为分类结果,该分类器可提供总结性分类,例如非GCB(例如,ABC)、GCB或未分类。

[0047] 返回图1,首先,该系统接收(110)(例如,从读取器或测序仪,并因此可能从存储状态接收)分别定量样品的多个DLBCL标签基因的基因表达水平(例如,蛋白和/或mRNA)多个值(例如,相对光单位或所检测的靶的计数或相对或比值读取数)。在一些实例中,接收CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个、例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个的基因表达值。可采用本文描述的或本领域普通技术人员已知的任何手段来为试验样品测量DLBCL标签基因的基因表达水平的值。在一些实例中,原始数据可任选地被“条件化(conditioned)”,例如可以从数据减去背景值和/或数据可以被对数转换(例如, \log_2 转换)。该系统可任选地在评分前采用本文描述的任何手段将多个值标准化,条件是分类也采用与训练期间确定阈值所采用的相同手段(或类似手段)来标准化。

[0048] 对于每个基因,对其表达值加权(120)、例如通过乘以权重值,该权重值反映了该基因的表达对样品分类为ABC、GBC或未分类的贡献。因此,例如,如果所有分类器基因的权重都为1,则每个这样的基因将在该分类器输出中具有相等的权重。示例性权重值和它们对应的基因提供在表5中。系统对经加权的基因表达值求和(130)。即,对每个所分析基因的经加权的值求和,得到经加权的总和。然后将该经加权的总和用于(例如,在统计模型中)获得概率评分(140),例如采用实施例1中的等式1。执行该分类器的计算系统可用于实施这些任务中的一个或更多个任务。

[0049] 将该概率评分与用于所预测概率的对应阈值(样品将分类为ABC或GBC或非该二者(即,未分类))相比较。用于评分的阈值和规则取决于具体实施(implementation)。本文描述了示例阈值,下文描述了示例评分规则。作为另外的选择,该系统采用其它的阈值和/或其它评分规则。在一个实例中,该截止值低点为0.43,高点为0.57。

[0050] 至少部分地基于该评分,该细胞对样品分类(160)。用于分类的规则取决于具体实

施。示例分类规则在下文描述。作为另外的选择,该细胞采用其它的分离规则。在一个实例中,如果所预测的概率 <0.43 则该系统将样品分类为ABC,如果所预测的概率 >0.57 则该系统将样品分类为GCB,并且如果所预测的概率处于 $0.43-0.57$ 之间(含)则该系统将样品分类为未分类的。

[0051] I. 为获得基因表达数据做准备

[0052] 基因表达是将基因组(基因)中编码的信息转化(例如通过转录和翻译过程)成相应的基因产物(例如RNA和蛋白)的过程,这些基因产物相互作用产生一组特征(又称为,表型)。出于本公开的目的,可以通过现在或将来已知的任何技术测量基因表达。一般地,通过在收集自目标受试者(例如患有DLBCL的受试者)的样品中检测表达的基因产物(例如, RNA和/或蛋白)来测量基因表达。因此,可测量两个或更多个靶基因如两个或更多个DLBCL标签基因或标志物的表达。在一个实例中,通过检测(例如,通过测序)一个或更多个特异结合靶mRNA的探针并确定与该mRNA靶结合的探针数量(相对或绝对的)来测量基因表达。因此,所检测的探针数量用作对基因表达的衡量。DLBCL标志物的实例包括CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个,例如 CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个(例如全部)。DLBCL标志物的其它具体实例包括CD47、ENTPD1、FOXP1、FUT8、IL16、NF2、PIM1、STAT3、TNFRSF8和TYMS(它们与ABC分离相关)中的两个或更多个(例如2、3、4、5、6、7、8、9或10个),或者CD86、ITPKB、LRMP、MME、PTPRC和REL(它们与GCB分类相关)中的两个或更多个(例如2、3、4、5或6个)。可用所公开的方法分析的受试者包括人和非人哺乳动物(例如,兽医受试者,例如猫、狗和小鼠)。在一个实例中,受试者已知或怀疑患有肿瘤,例如淋巴瘤,例如DLBCL。

[0053] A. 受试者和样品

[0054] 用于本文公开的方法的合适样品包括来自受试者的任何生物学样品,该样品包含来自肿瘤(例如,淋巴结、脾、肝、骨髓、脑和/或骨髓中的肿瘤)的细胞(例如,B细胞),而对于该肿瘤期望获得关于基因或蛋白(例如表1中的那些基因或蛋白)表达的信息。样品包括从受试者获得的那些样品,例如从受试者获得的临床样品(包括来自健康的或表现健康的人类受试者或患有待诊断或检查的状况或疾病(例如,DLBCL,更具体地,ABC或GCB)的人类患者的样品)。在一个实例中,该受试者为已知或怀疑患有DLBCL。示例样品可以从正常细胞或组织(例如,作为对照样品)和/或从肿瘤细胞或组织获得。在具体实例中,生物学样品包括肿瘤样品,例如含有淋巴瘤细胞的样品。在一个实例中,样品包括RNA,例如mRNA。在一些实施方案中,样品来自以前通过不同于本文描述的组织学或临床方法(例如,IHC或原位杂交(ISH))诊断为DLBCL的受试者。

[0055] 用于确定所公开的DLBCL标签基因中的两个或更多个基因的表达的样品包括包含核酸(例如基因组DNA、cDNA、病毒DNA或RNA、rRNA、tRNA、mRNA、miRNA、寡核苷酸、核酸片段、修饰的核酸、合成的核酸,等等)或蛋白的任何生物学标本。在一些实例中,例如选择具有淋巴瘤如非Hodgkin淋巴瘤(例如,DLBCL)的受试者来确定该受试者的诊断结果或预后,或用于选择一个或更多个疗法。

[0056] 示例性样品包括但不限于细胞、细胞裂解物、血涂片、细胞离心制备物、细胞学涂片、体液(例如,血液、血清、唾液、痰液、尿液等)、组织活检(例如,肿瘤活检)、淋巴结组织或

活检、液体活检、细针抽吸物、手术样本、骨髓、羊膜穿刺样品、尸检材料和/或组织切片(例如,冷冻组织切片和/或石蜡包埋组织切片)。应理解,可采用从受试者获得样品(例如组织)的任何方法,对所用方法的选择将取决于各种因素,例如组织类型、受试者年龄或医师可用的操作。获得所述样品的标准技术是可得到的。参见,例如Tubbs和Stoler, *Cell and Tissue Based Molecular Pathology*, Philadelphia:Churchill Livingstone (2009)。例如,可通过手术切除全部或部分肿瘤、通过从该肿瘤收集细针抽吸物以及本领域已知的其它方法从肿瘤获得含有细胞材料的样品。

[0057] 所公开的方法是灵敏和特异的,使得能够在含有甚至有限数量的细胞的样品中检测靶核酸分子。包括少量细胞,例如少于250,000个细胞(例如少于100,000个、少于50,000个、少于10,000个、少于1,000个、少于500个、少于200个、少于100个或少于10个细胞)的样品包括,但不限于,FFPE样品、细针抽吸物(例如来自淋巴或脾的那些)、钻孔活检、针活检、分选细胞的小群体(例如,FACS)、少量的激光捕获或宏切细胞(macrodissected cell)、胞外体(exosome)和其它亚细胞颗粒,或体液(例如脊髓液)。例如,可在少至1000个细胞(例如包括1000或更多个细胞,例如1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、50,000或更多个细胞的样品)中检测靶RNA。在一些实例中,可在约1000至100,000个细胞,例如约1000至50,000、1000至15,000、1000至10,000、1000至5000、3000至50,000、6000至30,000,或10,000至50,000个细胞中检测靶RNA的表达。在一些实例中,可在约100至250,000个细胞,例如约100至100,000、100至50,000、100至10,000、100至5000、100至500、100至200,或100至150个细胞中检测靶RNA的表达。在其它实例中,可在约1至1000个细胞(例如约1至500个细胞、1至250个细胞、约1至100个细胞、约1至50个细胞、约1至25个细胞,或约1个细胞)中检测靶RNA的表达。

[0058] 在特定实例中,样品可直接使用(例如,新鲜的或冷冻的),或可在使用前例如通过固定(例如,福尔马林固定(例如,中性缓冲福尔马林、锌福尔马林和酸性福尔马林)、乙醇固定)和/或通过包埋在固体介质中来保存。包埋介质通常为惰性的,能够防潮和能够渗入组织(例如,蜡、石蜡、火棉、OCT™化合物、琼脂、塑料以及丙烯酸类)。一些有用的样品为福尔马林固定石蜡包埋(FFPE)的组织样品。在具体实例中,待分析的组织样品是固定的,或者更具体地,是固定的和蜡(石蜡)包埋的。术语脱蜡(deparaffinization或dewaxing)指从生物学样品部分或完全除去任何类型的包埋介质(例如在采用本公开的方法分析之前)。例如,可通过让石蜡包埋组织切片穿过有机溶剂如甲苯、二甲苯、柠檬烯或其它适合溶剂来脱蜡。在其它样品中,直接使用蜡包埋组织切片(例如,无脱蜡步骤)。

[0059] 可通过任何适合的方法来固定组织,包括灌注或通过浸入固定剂中。固定剂可分为交联剂(例如醛、例如,甲醛、多聚甲醛和戊二醛以及非醛交联剂)、氧化剂(例如,金属离子和配合物,例如四氧化锇和铬酸)、蛋白变性剂(例如,乙酸、甲醇和乙醇)、未知机制的固定剂(例如,氯化汞、丙酮和苦味酸)、组合试剂(例如,卡诺伊固定剂(Carnoy's fixative)、methacarn、布安液(Bouin's fluid)、B5固定剂、罗斯曼液(Rossman's fluid)以及让德尔液(Gendre's fluid)、微波以及其它固定剂(例如,体积排除固定(excluded volume fixation)和蒸汽固定(vapor fixation))。固定剂中还可以包括添加剂,例如缓冲剂、去垢剂、鞣酸、酚、金属盐(例如氯化锌、硫酸锌和锂盐)以及镧。

[0060] 在制备组织或细胞样品中常用的固定剂为甲醛,通常为福尔马林溶液的形式(缓

冲溶液中4%甲醛,也称为10%缓冲福尔马林)。在一个实例中,该固定剂为10%中性缓冲福尔马林。

[0061] 包括固定的(例如,FFPE)组织或细胞的一些样品可以通过将样品悬浮在适当溶液(例如,在下文实施例中所述的)中来制备,没有对该样品中存在的mRNA的任何纯化和/或逆转录。附着到固体表面(例如,显微镜载玻片)的样品的全部或部分(例如,预测量的面积或体积)可以直接刮进适当溶液中,伴有温和混合(例如使用移液管)以分散和/或破碎大的样品碎片,并产生包含细胞材料的悬浮液,这些细胞材料大部分至少在该悬浮液与适当试剂接触期间保持悬浮。在一些实施方案中,样品中的mRNA可以(但不必)为部分降解的,或者几乎不用担心样品中mRNA发生一些不完全降解。

[0062] 在一些实施方案中,样品为从肿瘤或组织获得的细胞和/或组织的裂解液。细胞裂解液含有细胞中包含的很多蛋白和核酸,包括例如表1显示的生物标志物。用于获得或制备细胞裂解液的方法为本领域所公知,例如可在Ausubel等(In Current Protocols in Molecular Biology, John Wiley&Sons, New York, 1998)中找到。在一些实例中,样品中的细胞为在水溶液中裂解或透化的(permeabilized)。该水溶液或裂解缓冲液可以包括去垢剂(例如十二烷基硫酸钠)和一种或更多种离液剂(例如甲酰胺、盐酸胍、异硫氰酸胍或脲)。该溶液还可以包含缓冲剂(例如SSC)。在一些实例中,该裂解缓冲液包括约8%至60%甲酰胺(v/v)、约0.01%至0.5%SDS,以及约0.5-6X SSC(例如,约3X SSC)。该缓冲液可以任选地包括0.001至约2.0mg/ml tRNA,或者核糖核酸酶;脱氧核糖核酸酶;蛋白酶K;降解蛋白、基质、脂类或某种类型寡核苷酸的酶(例如,胶原酶或脂肪酶),或它们的组合。该裂解缓冲液还可以包括pH指示剂,例如酚红(Phenol Red)。将细胞在水溶液中以足够时间(例如约1分钟至约60分钟,例如约5分钟至约20分钟,或约10分钟)和足够温度(例如约22°C至约115°C,例如,约37°C至约105°C,或约50°C至约95°C或约65°C至约100°C)进行温育,以裂解或透化细胞。在一些实例中,例如如果待检测的核酸为RNA时,在约50°C、65°C或95°C进行裂解。在其它实例中,例如如果待检测的核酸为DNA时,在约105°C进行裂解。在一些实例中,裂解条件可以是基因组DNA不能触及而RNA(例如mRNA)能触及探针的条件。在一些实例中,直接使用粗细胞裂解物,无需进一步纯化。

[0063] 在一些实例中,将组织或细胞样品应用至底物并分析以确定一种或更多种靶核酸的存在。可用于所公开方法的固体支持物仅需负载生物学样品,并任选地,但有利的是,允许方便地检测样品中的组分(例如,蛋白和/或核酸序列)。示例性支持物包括显微镜载玻片(例如,玻璃显微镜载玻片或塑料显微镜载玻片)、盖玻片(例如,玻璃盖玻片或塑料盖玻片)、组织培养皿、多孔板、膜(例如,硝酸纤维素或聚偏二氟乙烯(PVDF))或BIACORE™芯片。

[0064] 可从样品中分离或提取靶(例如RNA),由此将其从样品中的非靶生物学组分纯化出来。但是,该步骤为可选的。纯化指将靶与也存在于样品中的一种或更多种无关组分分开。例如,在用成对的靶特异引物进行基于PCR的mRNA检测之前,经常将总的或可溶性mRNA(包括靶mRNA)与样品中的细胞蛋白和其它核酸分开。与未纯化或非提取样品相比,从混合样本或样品分离、提取或纯化的组分典型地富集至少50%、至少60%、至少75%、至少90%,或者至少98%或甚至至少99%。

[0065] 在一些实例中,样品中的核酸分子在它们的分析之前例如采用PCR扩增(例如,实时RT-PCR)。在一个实例中,使用定量PCR方法。

[0066] B. 对照样品

[0067] 对照样品可与所公开的方法并行使用,并可包括用于将表1中显示的生物标志物表达与其比较的任何合适的对照样品。在一些实施方案中,对照样品为非肿瘤组织,例如数种非肿瘤组织样品。在一个实例中,非肿瘤组织为已知是良性的组织,例如组织学上正常的淋巴组织。在一些实例中,非肿瘤组织包括表观正常的淋巴或骨髓样品;即,其缺乏细胞异常或其它已知疾病(例如,淋巴瘤,例如DLBCL)指征。在一些实例中,该非肿瘤组织从相同受试者获得,例如临近或甚至远离淋巴恶性肿瘤(例如DLBCL)的非肿瘤组织。在其它实例中,该非肿瘤组织从健康对照受试者或几个健康对照受试者获得。例如,非肿瘤组织可从多个健康对照受试者(例如,不具有任何癌症,包括淋巴瘤(例如,DLBCL))获得,例如来自多个这样的受试者的含有正常细胞或组织的样品(例如来自淋巴结、脾、肝、骨髓、脑和/或脊髓)。在一些实施方案中,将一个或更多个(例如,数个)对照样品用于获得用于表1中显示的生物标志物表达水平的参考(例如,正常对照)值或范围值。在一些实施方案中,从对照样品获得的参考值可以为群体集中趋势(例如均值、中值或平均值),或值的参考范围例如围绕群体集中趋势+0.5、1.0、1.5、2.0标准差。

[0068] C. 样品分析选项

[0069] 一些方法实施方案采用经固定的样品(例如,FFPE组织样品)。固定技术可以随地点、国家、研究者等而变化(Dissecting the Molecular Anatomy of Tissue, ed. by Emmert-Buck, Gillespie and Chuaqui, New York: Springer-Verlag, 244 pages (2010))并且可以影响待检测的基因产物的完整性和/或可达性。在一些这样的方法(例如,涉及PCR)中,RNA回收(例如,采用可逆交联剂、基于乙醇的固定剂和/或RNA提取或纯化(总的或部分的))可以是有利的;然而,在其它的代表性方法(例如,涉及qNPA或qNPS)中RNA回收是任选的或RNA回收明确是不必要的。类似地,组织调整(tissue conditioning)可用于从经固定的组织回收蛋白基因产物,并由此辅助检测这些蛋白产物。

[0070] 生物学样品中肿瘤(例如,DLBCL)百分比可以变化;因此,在一些公开的实施方案中,样品中样品面积(或样品体积)或总细胞的至少5%、至少10%、至少25%、至少50%、至少75%、至少80%或至少90%为肿瘤(例如,DLBCL)。在其它实施方案中,样品可以富集肿瘤细胞,例如通过从样品宏切主要为或表观主要为肿瘤(例如,DLBCL)的区域或细胞。任选地,病理学家或其它经适当训练的专业人员可以观察样品(例如,H&E染色的组织切片)以确定是否有足够的肿瘤存在于样品中用于测试和/或标记待宏切的区域(例如,最密集肿瘤的区域)。在具体的实例中,肿瘤(例如,DLBCL)的宏切尽可能避免坏死和/或出血区域。以样品体积、或面积或总细胞计,可用在一些公开方法中的样品将具有少于25%、15%、10%、5%、2%、或1%的坏死。

[0071] 样品载荷可以影响可用于检测的基因产物(例如,表1中的生物标志物)的量和/或浓度。在具体实施方案中,至少1ng、10ng、100ng、1 μ g、10 μ g、100 μ g、500 μ g、1mg总RNA,至少1ng、10ng、100ng、1 μ g、10 μ g、100 μ g、500 μ g、1mg总DNA,或至少0.01ng、0.1ng、1ng、10ng、100ng、1 μ g、10 μ g、100 μ g、500 μ g、或1mg的总蛋白从样品(例如样品裂解物)分离和/或存在其中。一些实施方案采用组织样品(例如,FFPE淋巴结、脾、肝、骨髓、脑和/或脊髓组织),它们厚度至少为3、5、8或10 μ m(例如,约3至10 μ m)和/或面积至少为0.15、0.2、0.5、1、1.5、2、5或10cm²。在一些方法实施方案中,悬浮在缓冲液中的样品浓度为至少0.006cm²/ μ l(例如,

0.15cm²FFPE组织/25μL缓冲液(例如,裂解缓冲液))。

[0072] II. 基因和基因集合 (Gene Sets)

[0073] 本文公开的创新中包括可用于区分DLBCL亚型的基因(也称为生物标志物)和基因集合(也称为基因标签)。在具体实施方案中,公开了用于对ABC和GCB DLBCL亚型分型的基因和基因集合(见表1)。这些基因和基因集合的表达可用在DLBCL分类器和算法中,和/或为阵列或其它公开的组合物设计分析物特异的试剂(例如,核酸探针或抗体)。

[0074] 在一些实施方案中,确定生物样品(例如肿瘤活检)中的表达水平包括检测两种或更多种显示在表1中的基因产物(例如, RNA或蛋白),例如通过确定样品中这些核酸的相对或实际量,如本文所描述的。

[0075] A. DLBCL标签基因

[0076] 示例性DLBCL标签基因包括显示在表1中的那些。在一些实施方案中,该DLBCL基因标签包括CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和 TYMS中的两个或更多个(例如这些中的至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9或全部10个),例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个(例如这些中的至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少11、至少12、至少13、至少14、至少15或全部16个)。

[0077] 表1. 允许对GCB和非GCB亚型分类的示例DLBCL标志物

基因	GenBank®登记号
CD47	NM_001777.3, NM_198793.2, LN680437.1

[0078]

基因	GenBank®登记号
CD86	NM_175862.4, NG_029928.1, NM_176892.1, NM_001206924.1, NM_001206925.1
ENTPD1 (CD39)	NM_001776.5; NM_001098175.1; NM_001164178.1; NM_001164179.1; NM_001164181.1; NM_001164182.1; NM_001164183.1
FOXP1	NM_001012505.1; NM_032682.5; NM_001244808.1; NM_001244812.1; NM_001244814.1; NM_001244815.1, NM_001244813.1; NM_001244810.1; NM_001244816.1
FUT8	NM_178155.2, NM_004480.4, NM_178156.2, NR_038167.1, NR_038170.1
IL16	NM_004513.5, NM_172217.3, NM_001172128.1, XM_011521520.1, XR_931805.1
IITPKB	NM_002221.3; XM_005273120.1
LRMP	NM_006152.3, NM_001204127.1, NM_001204126.1, XM_006719076.2, XM_011520668.1, XM_011520669.1 XM_005253374.2, XM_011520670.1
MME (CD10)	NM_000902.3, NM_007287.2, NM_007288.2, BC143465.1, NM_007289.2
NF2	NM_000268.3, NM_016418.5, NM_181828.2, NM_181831.2
PIM1	NM_001243186.1, M24779.1, NM_002648.3
PTPRC	NM_002838.4, NM_080921.3, NR_052021.1, NM_001267798.1
REL	NM_002908.3, NM_001291746.1, DQ314888.1
STAT3	NM_139276.2, NM_003150.3, NM_213662.1, NG_007370.1
TNFRSF8	NM_001243.4, XM_011542442.1, XM_011542444.1, XM_011542443.1, AY498860.1
TYMS	NM_001071.2; NG_028255.1

[0079] [0080] 分化簇47 (CD47) (例如, OMIM 601028), 也称整联蛋白相关蛋白, 为跨膜蛋白, 其与膜整联蛋白相伴, 也结合配体血小板反应蛋白-1 (thrombospondin-1) 和信号调节蛋白 α 。已经鉴定了该基因的转录变体 (例如, 可变剪接转录物)。CD47的核酸和蛋白序列为公众可获得的。例如, GenBank®登记号NM_001777.3、NM_198793.2和LN680437.1公开了示例性人CD47核酸序列, GenBank®登记号NP_001768.1、NP_942088.1和CEJ95640.1公开了示例性人CD47蛋白序列。本领域技术人员应理解, 这些CD47序列的变体可用于DLBCL分类, 例如, 其表达在DLBCL中 (例如在ABC和/或GCB中) 有改变的那些, 例如与本文提供的任何CD47 GenBank®登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0081] 分化簇86 (CD86) (例如, OMIM 601020), 也称B7-2, 在抗原呈递细胞上表达, 并为T细胞活化和存活提供共刺激信号。已经鉴定了该基因的转录变体 (例如, 可变剪接转录物)。CD86的核酸和蛋白序列为公众可获得的。例如, GenBank®登记号NM_175862.4、NG_029928.1、NM_176892.1、NM_001206924.1和NM_001206925.1公开了示例性人CD86 核酸序列, GenBank®登记号NP_787058.4、NP_008820.3、NP_795711.1、NP_001193853.1 和NP_001193854.1公开了示例性人CD86蛋白序列。本领域技术人员应理解, 这些CD86 序列的变体可用于DLBCL分类, 例如, 其表达在DLBCL中 (例如在ABC和/或GCB中) 有改变的那些, 例如

与本文提供的任何CD86 GenBank® 登记号有至少90%、至少95%、至少 96%、至少97%、至少98%、至少99%序列一致性的序列。

[0082] 外核苷三磷酸二磷酸水解酶1 (ectonucleoside triphosphate diphosphohydrolase 1, ENTPD1) (例如, OMIM 601752), 也称CD39、ATPDase和NTPDase-1, 为水解ATP和 ADP至AMP, 引发腺苷生成的酶 (EC 3.6.1.5)。ENTPD1广泛表达, 包括在血管和免疫系统的细胞中表达。已经鉴定了该基因的转录变体 (例如, 可变剪接转录物)。ENTPD1的核酸和蛋白序列为公众可获得的。例如, GenBank® 登记号NM_001776.5; NM_001098175.1; NM_001164178.1; NM_001164179.1; NM_001164181.1; 和NM_001164182.1; NM_001164183.1 公开了示例性人ENTPD1核酸序列, GenBank® 登记号AAH47664.1、NP_001091645.1、NP_001157650.1、NP_001157653.1、NP_001767.3、NP_001157651.1、NP_001157654.1和NP_001157655.1公开了示例性人ENTPD1蛋白序列。本领域技术人员应理解, 这些ENTPD1序列的变体可用于DLBCL分类, 例如, 其表达在DLBCL中 (例如在 ABC和/或GCB中) 有改变的那些, 例如与本文提供的任何ENTPD1 GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0083] 叉头框蛋白P1 (forkhead box protein P1, FOXP1) (例如, OMIM 605515) 为含有DNA结合和蛋白-蛋白结合结构域的转录因子, 其起转录阻遏蛋白的作用。FOXP1调节各种发育重要方面, 包括肺、脑、胸腺和心脏的组织发育。已经鉴定了该基因的转录变体 (例如, 可变剪接转录物)。Foxp1的核酸和蛋白序列为公众可获得的。例如, GenBank® 登记号 NM_001012505.1; NM_032682.5; NM_001244808.1; NM_001244812.1; NM_001244814.1; NM_001244815.1; NM_001244813.1; NM_001244810.1和NM_001244816.1公开了示例性人 Foxp1核酸序列, GenBank® 登记号AAG47632.1、NP_001231743.1、NP_001012523.1、NP_001231737.1、NP_001231739.1、NP_001231741.1、NP_001231742.1和NP_001231744.1 公开了示例性人Foxp1蛋白序列。本领域技术人员应理解, 这些Foxp1序列的变体可用于 DLBCL分类, 例如, 其表达在DLBCL中 (例如在ABC和/或GCB中) 有改变的那些, 例如与本文提供的任何Foxp1 GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0084] 岩藻糖基转移酶8 (fucosyltransferase 8, α (1,6) 岩藻糖基转移酶) (FUT8) (例如, OMIM 602589), 也称GDP-L-Fuc:N-乙酰基- β -D-氨基葡萄糖苷 α 1,6-岩藻糖基转移酶; GDP-岩藻糖-糖蛋白岩藻糖基转移酶; 和 α -(1,6)-岩藻糖基转移酶, 为催化岩藻糖从GDP-岩藻糖转移至N连接类型的复杂糖肽的酶 (EC 2.4.1.68)。已经鉴定了该基因的转录变体 (例如, 可变剪接转录物) 和非编码RNA。FUT8的核酸和蛋白序列为公众可获得的。例如, GenBank® 登记号 NM_178155.2、NM_004480.4、NM_178156.2、NR_038167.1和NR_038170.1 公开了示例性人FUT8核酸序列, GenBank® 登记号AAI42959.1、NP_835368.1、NP_004471.4 和 NP_835369.1公开了示例性人FUT8蛋白序列。本领域技术人员应理解, 这些FUT8序列的变体可用于DLBCL分类, 例如, 其表达在DLBCL中 (例如在ABC和/或GCB中) 有改变的那些, 例如与本文提供的任何FUT8 GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0085] 白细胞介素16 (IL16) (例如, OMIM 603035) 为细胞因子, 对一些表达CD4的免疫细

胞起趋化因子的作用。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。IL16的核酸和蛋白序列为公众可获得的。例如, **GenBank®** 登记号NM_004513.5、NM_172217.3、NM_001172128.1、XM_011521520.1和XR_931805.1公开了示例性人IL16核酸序列, **GenBank®** 登记号AAI36661.1、AAD15990.1、NP_001165599.1、ACI00236.1和 XP_011519821.1公开了示例性人IL16蛋白序列。本领域技术人员应理解,这些IL16序列的变体可用于DLBCL分类,例如,其表达在DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何IL16 **GenBank®** 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0086] 肌醇1,4,5-三磷酸3-激酶B (ITPKB) (例如,OMIM 147522),也称IP3 3-激酶B、IP3KB 和增殖诱导蛋白37,将肌醇1,4,5-三磷酸磷酸化为Ins (1,3,4,5) P4。ITPKB的核酸和蛋白序列为公众可获得的。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。例如, **GenBank®** 登记No.NM_002221.3和XM_005273120.1公开了示例性人ITPKB核酸序列, **GenBank®** 登记号NP_002212.3和AAH15009.1公开了示例性人ITPKB蛋白序列。本领域技术人员应理解,这些ITPKB序列的变体可用于DLBCL分类,例如,其表达在DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何ITPKB **GenBank®** 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0087] 淋巴限制性膜蛋白(LRMP) (例如,OMIM 602003),也称JAW1,在淋巴细胞系和组织中表达。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。LRMP的核酸和蛋白序列为公众可获得的。例如, **GenBank®** 登记号NM_006152.3、NM_001204127.1、NM_001204126.1、XM_006719076.2、XM_011520668.1、XM_011520669.1、XM_005253374.2和XM_011520670.1公开了示例性人LRMP核酸序列, **GenBank®** 登记号 NP_006143.2、NP_001191056.1、XP_011518971.1、XP_011518972.1、Q12912.3和 NP_001191055.1公开了示例性人LRMP蛋白序列。本领域技术人员应理解,这些LRMP序列的变体可用于DLBCL分类,例如,其表达在DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何LRMP **GenBank®** 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0088] 膜金属内肽酶(membrane metallo-endopeptidase, MME) (例如,OMIM 120520),也称 CD10、心房肽酶(atriopeptidase)、常见急性淋巴细胞性白血病抗原(common acute lymphocytic leukemia antigen)、脑啡肽酶(enkephalinase)、脑啡肽酶(neprilysin)和中性内肽酶(neutral endopeptidase),为在疏水残基的氨基侧切割肽并让几种肽激素(包括胰高血糖素、脑啡肽、物质P、神经紧张素、催产素和缓激肽)失活的中性内肽酶(EC 3.4.24.11)。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。MME的核酸和蛋白序列为公众可获得的。例如, **GenBank®** 登记号NM_000902.3、NM_007287.2、NM_007288.2、BC143465.1和 NM_007289.2公开了示例性人MME核酸序列, **GenBank®** 登记号AAI43466.1、AAI01659.1、NP_000893.2、NP_009218.2、NP_009219.2和NP_009220.2公开了示例性人MME蛋白序列。本领域技术人员应理解,这些MME序列的变体可用于DLBCL分类,例如,其表达在DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何MME **GenBank®** 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0089] 神经纤维瘤蛋白2(NF2) (例如,OMIM 607379),也称merlin,为具有肿瘤遏制性质

的细胞骨架蛋白。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。NF2的核酸和蛋白序列为公众可获得的。例如, GenBank® 登记号NM_000268.3、NM_016418.5、NM_181828.2和NM_181831.2公开了示例性人NF2核酸序列, GenBank® 登记号 NP_000259.1、NP_057502.2、NP_861966.1和NP_861969.1公开了示例性人NF2蛋白序列。本领域技术人员应理解,这些NF2序列的变体可用于DLBCL分类,例如,其表达在 DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何NF2 GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0090] 原癌基因丝氨酸/苏氨酸蛋白激酶Pim-1 (PIM1) (例如,OMIM 164960) 为参与细胞因子信号转导的原癌基因。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。PIM1的核酸和蛋白序列为公众可获得的。例如, GenBank® 登记号NM_001243186.1、M24779.1和NM_002648.3公开了示例性人PIM1核酸序列, GenBank® 登记号NP_001230115.1、NP_002639.1和AAH20224.1公开了示例性人PIM1蛋白序列。本领域技术人员应理解,这些PIM1序列的变体可用于DLBCL分类,例如,其表达在DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何PIM1 GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0091] 蛋白酪氨酸磷酸酶,受体类型,C (PTPRC) (例如,OMIM 151460), 也称CD45,为在造血细胞中表达的酶(EC 3.1.3.48),其调节T和B细胞抗原受体信号转导。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。该基因含有34个外显子,初级转录物的三个外显子经可变剪接产生多至八种不同的成熟mRNA以及翻译后八种不同的蛋白产物。这三个外显子产生RA、RB和RC同种型。PTPRC的核酸和蛋白序列为公众可获得的。例如, GenBank® 登记号NM_002838.4、NM_080921.3、NR_052021.1和NM_001267798.1公开了示例性人PTPRC核酸序列, GenBank® 登记号NP_002829.3、NP_563578.2、P08575.2和 NP_001254727.1公开了示例性人PTPRC蛋白序列。本领域技术人员应理解,这些PTPRC 序列的变体可用于DLBCL分类,例如,其表达在DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何PTPRC GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0092] v-rel禽网状内皮增生症病毒癌基因同源物(REL) (例如,OMIM 164910),为含有Rel同源物结构域的转录因子,在B细胞存活和增殖中起作用。该REL基因在DLBCL中被扩增或突变。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。REL的核酸和蛋白序列为公众可获得的。例如, GenBank® Accession Nos. NM_002908.3、NM_001291746.1和DQ314888.1公开了示例性人REL核酸序列, GenBank® 登记号NP_002899.1、NP_001278675.1和AAI43886.1公开了示例性人REL蛋白序列。本领域技术人员应理解,这些REL序列的变体可用于DLBCL分类,例如,其表达在DLBCL中(例如在ABC和/或GCB中)有改变的那些,例如与本文提供的任何REL GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0093] 信号转导子和转录激活子(Signal transducer and activator of transcription, STAT3) (例如, OMIM 102582) 为转录激活因子,取决于肿瘤的遗传背景,其可具有致癌或肿瘤遏制作用。已经鉴定了该基因的转录变体(例如,可变剪接转录物)。

STAT3的核酸和蛋白序列为公众可获得的。例如，GenBank® 登记号NM_139276.2、NM_003150.3、NM_213662.1和 NG_007370.1公开了示例性人STAT3核酸序列，GenBank® 登记号NP_644805.1、NP_003141.2和NP_998827.1公开了示例性人STAT3蛋白序列。本领域技术人员应理解，这些STAT3序列的变体可用于DLBCL分类，例如，其表达在DLBCL中（例如在ABC和/或GCB中）有改变的那些，例如与本文提供的任何STAT3 GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0094] 肿瘤坏死因子受体超家族，成员8 (TNFRSF8) (例如，OMIM 153243)，也称为CD30，由活化的T和B细胞表达。已经鉴定了该基因的转录变体（例如，可变剪接转录物）。TNFRSF8的核酸和蛋白序列为公众可获得的。例如，GenBank® 登记号NM_001243.4、XM_011542442.1、XM_011542444.1、XM_011542443.1和AY498860.1公开了示例性人 TNFRSF8核酸序列，GenBank® 登记号NP_001234.3、XP_011540746.1和NP_001268359.2 公开了示例性人TNFRSF8蛋白序列。本领域技术人员应理解，这些TNFRSF8序列的变体可用于DLBCL分类，例如，其表达在DLBCL中（例如在ABC和/或GCB中）有改变的那些，例如与本文提供的任何TNFRSF8 GenBank® 登记号有至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0095] 胸苷酸合成酶 (TYMS) (例如，OMIM 188350) 为催化脱氧尿苷单磷酸 (dUMP) 转变为脱氧胸苷单磷酸的酶 (EC 2.1.1.45)。已经鉴定了该基因的转录变体（例如，可变剪接转录物）。TYMS的核酸和蛋白序列为公众可获得的。例如，GenBank® 登记号NM_001071.2和NG_028255.1公开了示例性人TYMS核酸序列，GenBank® 登记号NP_001062.1、EAX01716.1和EAX01717.1公开了示例性人TYMS蛋白序列。本领域技术人员应理解，这些TYMS序列的变体可用于DLBCL分类，例如，其表达在DLBCL中（例如在ABC和/或 GCB中）有改变的那些，例如与本文提供的任何TYMS GenBank® 登记号有至少90%、至少 95%、至少96%、至少97%、至少98%、至少99%序列一致性的序列。

[0096] 可用于DLBCL亚型分型的具体实施方案包括，但不限于，确定或测量如下基因的表达：

[0097] a. CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的一个或多个（例如，至少或固定的2、3、4、5、6、7、8、9或10个）；

[0098] b. CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的全部；

[0099] c. CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的一个或多个（例如，至少或固定的2、3、4、5、6、7、8、9、10、11、12、13、14、15或16个）；

[0100] d. CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的一个或多个（例如，至少或固定的2、3、4、5、6、7、8、9或10个）；

[0101] e. CD86、ITPKB、LRMP、MME、PTPRC和REL中的一个或多个（例如，至少或固定的2、3、4、5或6个）；

[0102] f. 任何基因集合，其包括如下两基因组合中的任何一个或任何两个（不重复的程度），或由其组成：{CD47, CD86} {CD47, IL16} {CD47, NF2} {CD47, PIM1} {CD47, PTPRC} {CD47,

REL} {CD47, STAT3} {CD47, TNFRSF8} {CD47, TYMS} {CD86, IL16} {CD86, NF2} {CD86, PIM1} {CD86, PTPRC} {CD86, REL} {CD86, STAT3} {CD86, TNFRSF8} {CD86, TYMS} {IL16, NF2} {IL16, PIM1} {IL16, PTPRC} {IL16, REL} {IL16, STAT3} {IL16, TNFRSF8} {IL16, TYMS} {NF2, PIM1} {NF2, PTPRC} {NF2, REL} {NF2, STAT3} {NF2, TNFRSF8} {NF2, TYMS} {PIM1, PTPRC} {PIM1, REL} {PIM1, STAT3} {PIM1, TNFRSF8} {PIM1, TYMS} {PTPRC, REL} {PTPRC, STAT3} {PTPRC, TNFRSF8} {PTPRC, TYMS} {REL, STAT3} {REL, TNFRSF8} {REL, TYMS} {STAT3, TNFRSF8} {STAT3, TYMS} {TNFRSF8, TYMS};

[0103] g. 任何基因集合, 其包括如下三基因组合中的任何一个或任何两个 (不重复的程度), 或由其组成: {CD47, CD86, IL16} {CD47, CD86, NF2} {CD47, CD86, PIM1} {CD47, CD86, PTPRC} {CD47, CD86, REL} {CD47, CD86, STAT3} {CD47, CD86, TNFRSF8} {CD47, CD86, TYMS} {CD47, IL16, NF2} {CD47, IL16, PIM1} {CD47, IL16, PTPRC} {CD47, IL16, REL} {CD47, IL16, STAT3} {CD47, IL16, TNFRSF8} {CD47, IL16, TYMS} {CD47, NF2, PIM1} {CD47, NF2, PTPRC} {CD47, NF2, REL} {CD47, NF2, STAT3} {CD47, NF2, TNFRSF8} {CD47, NF2, TYMS} {CD47, PIM1, PTPRC} {CD47, PIM1, REL} {CD47, PIM1, STAT3} {CD47, PIM1, TNFRSF8} {CD47, PIM1, TYMS} {CD47, PTPRC, REL} {CD47, PTPRC, STAT3} {CD47, PTPRC, TNFRSF8} {CD47, PTPRC, TYMS} {CD47, REL, STAT3} {CD47, REL, TNFRSF8} {CD47, REL, TYMS} {CD47, STAT3, TNFRSF8} {CD47, STAT3, TYMS} {CD47, TNFRSF8, TYMS} {CD86, IL16, NF2} {CD86, IL16, PIM1} {CD86, IL16, PTPRC} {CD86, IL16, REL} {CD86, IL16, STAT3} {CD86, IL16, TNFRSF8} {CD86, IL16, TYMS} {CD86, NF2, PIM1} {CD86, NF2, PTPRC} {CD86, NF2, REL} {CD86, NF2, STAT3} {CD86, NF2, TNFRSF8} {CD86, NF2, TYMS} {CD86, PIM1, PTPRC} {CD86, PIM1, REL} {CD86, PIM1, STAT3} {CD86, PIM1, TNFRSF8} {CD86, PIM1, TYMS} {CD86, PTPRC, REL} {CD86, PTPRC, STAT3} {CD86, PTPRC, TNFRSF8} {CD86, PTPRC, TYMS} {CD86, REL, STAT3} {CD86, REL, TNFRSF8} {CD86, REL, TYMS} {CD86, STAT3, TNFRSF8} {CD86, STAT3, TYMS} {CD86, TNFRSF8, TYMS} {IL16, NF2, PIM1} {IL16, NF2, PTPRC} {IL16, NF2, REL} {IL16, NF2, STAT3} {IL16, NF2, TNFRSF8} {IL16, NF2, TYMS} {IL16, PIM1, PTPRC} {IL16, PIM1, REL} {IL16, PIM1, STAT3} {IL16, PIM1, TNFRSF8} {IL16, PIM1, TYMS} {IL16, PTPRC, REL} {IL16, PTPRC, STAT3} {IL16, PTPRC, TNFRSF8} {IL16, PTPRC, TYMS} {IL16, REL, STAT3} {IL16, REL, TNFRSF8} {IL16, REL, TYMS} {IL16, STAT3, TNFRSF8} {IL16, STAT3, TYMS} {IL16, TNFRSF8, TYMS} {NF2, PIM1, PTPRC} {NF2, PIM1, REL} {NF2, PIM1, STAT3} {NF2, PIM1, TNFRSF8} {NF2, PIM1, TYMS} {NF2, PTPRC, REL} {NF2, PTPRC, STAT3} {NF2, PTPRC, TNFRSF8} {NF2, PTPRC, TYMS} {NF2, REL, STAT3} {NF2, REL, TNFRSF8} {NF2, REL, TYMS} {NF2, STAT3, TNFRSF8} {NF2, STAT3, TYMS} {NF2, TNFRSF8, TYMS} {PIM1, PTPRC, REL} {PIM1, PTPRC, STAT3} {PIM1, PTPRC, TNFRSF8} {PIM1, PTPRC, TYMS} {PIM1, REL, STAT3} {PIM1, REL, TNFRSF8} {PIM1, REL, TYMS} {PIM1, STAT3, TNFRSF8} {PIM1, STAT3, TYMS} {PIM1, TNFRSF8, TYMS} {PTPRC, REL, STAT3} {PTPRC, REL, TNFRSF8} {PTPRC, REL, TYMS} {PTPRC, STAT3, TNFRSF8} {PTPRC, STAT3, TYMS} {PTPRC, TNFRSF8, TYMS} {REL, STAT3, TNFRSF8} {REL, STAT3, TYMS} {REL, TNFRSF8, TYMS} {STAT3, TNFRSF8, TYMS};

[0104] h. 任何基因集合, 其包括如下四基因组合中的任何一个或任何两个 (不重复的程度), 或由其组成: {CD47, CD86, IL16, NF2} {CD47, CD86, IL16, PIM1} {CD47, CD86, IL16, PTPRC} {CD47, CD86, IL16, REL} {CD47, CD86, IL16, STAT3} {CD47, CD86, IL16, TNFRSF8}

{CD47,CD86,IL16,TYMS} {CD47,CD86,NF2,PIM1} {CD47,CD86,NF2,PTPRC} {CD47,CD86,NF2,REL} {CD47,CD86,NF2,STAT3} {CD47,CD86,NF2,TNFRSF8} {CD47,CD86,NF2,TYMS} {CD47,CD86,PIM1,PTPRC} {CD47,CD86,PIM1,REL} {CD47,CD86,PIM1,STAT3} {CD47,CD86,PIM1,TNFRSF8} {CD47,CD86,PIM1,TYMS} {CD47,CD86,PTPRC,REL} {CD47,CD86,PTPRC,STAT3} {CD47,CD86,PTPRC,TNFRSF8} {CD47,CD86,PTPRC,TYMS} {CD47,CD86,REL,STAT3} {CD47,CD86,REL,TNFRSF8} {CD47,CD86,REL,TYMS} {CD47,CD86,STAT3,TNFRSF8} {CD47,CD86,STAT3,TYMS} {CD47,CD86,TNFRSF8,TYMS} {CD47,IL16,NF2,PIM1} {CD47,IL16,NF2,PTPRC} {CD47,IL16,NF2,REL} {CD47,IL16,NF2,STAT3} {CD47,IL16,NF2,TNFRSF8} {CD47,IL16,NF2,TYMS} {CD47,IL16,PIM1,PTPRC} {CD47,IL16,PIM1,REL} {CD47,IL16,PIM1,STAT3} {CD47,IL16,PIM1,TNFRSF8} {CD47,IL16,PIM1,TYMS} {CD47,IL16,PTPRC,REL} {CD47,IL16,PTPRC,STAT3} {CD47,IL16,PTPRC,TNFRSF8} {CD47,IL16,PTPRC,TYMS} {CD47,IL16,REL,STAT3} {CD47,IL16,REL,TNFRSF8} {CD47,IL16,REL,TYMS} {CD47,IL16,STAT3,TNFRSF8} {CD47,IL16,STAT3,TYMS} {CD47,IL16,TNFRSF8,TYMS} {CD47,NF2,PIM1,PTPRC} {CD47,NF2,PIM1,REL} {CD47,NF2,PIM1,STAT3} {CD47,NF2,PIM1,TNFRSF8} {CD47,NF2,PIM1,TYMS} {CD47,NF2,PTPRC,REL} {CD47,NF2,PTPRC,STAT3} {CD47,NF2,PTPRC,TNFRSF8} {CD47,NF2,PTPRC,TYMS} {CD47,NF2,REL,STAT3} {CD47,NF2,REL,TNFRSF8} {CD47,NF2,REL,TYMS} {CD47,NF2,STAT3,TNFRSF8} {CD47,NF2,STAT3,TYMS} {CD47,NF2,TNFRSF8,TYMS} {CD47,PIM1,PTPRC,REL} {CD47,PIM1,PTPRC,STAT3} {CD47,PIM1,PTPRC,TNFRSF8} {CD47,PIM1,PTPRC,TYMS} {CD47,PIM1,REL,STAT3} {CD47,PIM1,REL,TNFRSF8} {CD47,PIM1,REL,TYMS} {CD47,PIM1,STAT3,TNFRSF8} {CD47,PIM1,STAT3,TYMS} {CD47,PIM1,TNFRSF8,TYMS} {CD47,PTPRC,REL,STAT3} {CD47,PTPRC,REL,TNFRSF8} {CD47,PTPRC,REL,TYMS} {CD47,PTPRC,STAT3,TNFRSF8} {CD47,PTPRC,STAT3,TYMS} {CD47,PTPRC,TNFRSF8,TYMS} {CD47,REL,STAT3,TNFRSF8} {CD47,REL,STAT3,TYMS} {CD47,REL,TNFRSF8,TYMS} {CD47,STAT3,TNFRSF8,TYMS} {CD86,IL16,NF2,PIM1} {CD86,IL16,NF2,PTPRC} {CD86,IL16,NF2,REL} {CD86,IL16,NF2,STAT3} {CD86,IL16,NF2,TNFRSF8} {CD86,IL16,NF2,TYMS} {CD86,IL16,PIM1,PTPRC} {CD86,IL16,PIM1,REL} {CD86,IL16,PIM1,STAT3} {CD86,IL16,PIM1,TNFRSF8} {CD86,IL16,PIM1,TYMS} {CD86,IL16,PTPRC,REL} {CD86,IL16,PTPRC,STAT3} {CD86,IL16,PTPRC,TNFRSF8} {CD86,IL16,PTPRC,TYMS} {CD86,IL16,REL,STAT3} {CD86,IL16,REL,TNFRSF8} {CD86,IL16,REL,TYMS} {CD86,IL16,STAT3,TNFRSF8} {CD86,IL16,STAT3,TYMS} {CD86,IL16,TNFRSF8,TYMS} {CD86,NF2,PIM1,PTPRC} {CD86,NF2,PIM1,REL} {CD86,NF2,PIM1,STAT3} {CD86,NF2,PIM1,TNFRSF8} {CD86,NF2,PIM1,TYMS} {CD86,NF2,PTPRC,REL} {CD86,NF2,PTPRC,STAT3} {CD86,NF2,PTPRC,TNFRSF8} {CD86,NF2,PTPRC,TYMS} {CD86,NF2,REL,STAT3} {CD86,NF2,REL,TNFRSF8} {CD86,NF2,REL,TYMS} {CD86,NF2,STAT3,TNFRSF8} {CD86,NF2,STAT3,TYMS} {CD86,NF2,TNFRSF8,TYMS} {CD86,PIM1,PTPRC,REL} {CD86,PIM1,PTPRC,STAT3} {CD86,PIM1,PTPRC,TNFRSF8} {CD86,PIM1,PTPRC,TYMS} {CD86,PIM1,REL,STAT3} {CD86,PIM1,REL,TNFRSF8} {CD86,PIM1,REL,TYMS} {CD86,PIM1,STAT3,TNFRSF8} {CD86,PIM1,STAT3,TYMS} {CD86,PIM1,TNFRSF8,TYMS} {CD86,PTPRC,REL,STAT3} {CD86,PTPRC,REL,TNFRSF8} {CD86,PTPRC,REL,TYMS} {CD86,PTPRC,STAT3,TNFRSF8} {CD86,PTPRC,STAT3,TYMS} {CD86,PTPRC,TNFRSF8,

TYMS} {CD86,REL,STAT3,TNFRSF8} {CD86,REL,STAT3,TYMS} {CD86,REL,TNFRSF8,TYMS} {CD86,STAT3,TNFRSF8,TYMS} {IL16,NF2,PIM1,PTPRC} {IL16,NF2,PIM1,REL} {IL16,NF2,PIM1,STAT3} {IL16,NF2,PIM1,TNFRSF8} {IL16,NF2,PIM1,TYMS} {IL16,NF2,PTPRC,REL} {IL16,NF2,PTPRC,STAT3} {IL16,NF2,PTPRC,TNFRSF8} {IL16,NF2,PTPRC,TYMS} {IL16,NF2,REL,STAT3} {IL16,NF2,REL,TNFRSF8} {IL16,NF2,REL,TYMS} {IL16,NF2,STAT3,TNFRSF8} {IL16,NF2,STAT3,TYMS} {IL16,NF2,TNFRSF8,TYMS} {IL16,PIM1,PTPRC,REL} {IL16,PIM1,PTPRC,STAT3} {IL16,PIM1,PTPRC,TNFRSF8} {IL16,PIM1,PTPRC,TYMS} {IL16,PIM1,REL,STAT3} {IL16,PIM1,REL,TNFRSF8} {IL16,PIM1,REL,TYMS} {IL16,PIM1,STAT3,TNFRSF8} {IL16,PIM1,STAT3,TYMS} {IL16,PIM1,TNFRSF8,TYMS} {IL16,PTPRC,REL,STAT3} {IL16,PTPRC,REL,TNFRSF8} {IL16,PTPRC,REL,TYMS} {IL16,PTPRC,STAT3,TNFRSF8} {IL16,PTPRC,STAT3,TYMS} {IL16,PTPRC,TNFRSF8,TYMS} {IL16,REL,STAT3,TNFRSF8} {IL16,REL,STAT3,TYMS} {IL16,REL,TNFRSF8,TYMS} {IL16,STAT3,TNFRSF8,TYMS} {NF2,PIM1,PTPRC,REL} {NF2,PIM1,PTPRC,STAT3} {NF2,PIM1,PTPRC,TNFRSF8} {NF2,PIM1,PTPRC,TYMS} {NF2,PIM1,REL,STAT3} {NF2,PIM1,REL,TNFRSF8} {NF2,PIM1,REL,TYMS} {NF2,PIM1,STAT3,TNFRSF8} {NF2,PIM1,STAT3,TYMS} {NF2,PIM1,TNFRSF8,TYMS} {NF2,PTPRC,REL,STAT3} {NF2,PTPRC,REL,TNFRSF8} {NF2,PTPRC,REL,TYMS} {NF2,PTPRC,STAT3,TNFRSF8} {NF2,PTPRC,STAT3,TYMS} {NF2,PTPRC,TNFRSF8,TYMS} {NF2,REL,STAT3,TNFRSF8} {NF2,REL,STAT3,TYMS} {NF2,REL,TNFRSF8,TYMS} {NF2,STAT3,TNFRSF8,TYMS} {PIM1,PTPRC,REL,STAT3} {PIM1,PTPRC,REL,TNFRSF8} {PIM1,PTPRC,REL,TYMS} {PIM1,PTPRC,STAT3,TNFRSF8} {PIM1,PTPRC,STAT3,TYMS} {PIM1,PTPRC,TNFRSF8,TYMS} {PIM1,REL,STAT3,TNFRSF8} {PIM1,REL,STAT3,TYMS} {PIM1,REL,TNFRSF8,TYMS} {PIM1,STAT3,TNFRSF8,TYMS} {PTPRC,REL,STAT3,TNFRSF8} {PTPRC,REL,STAT3,TYMS} {PTPRC,REL,TNFRSF8,TYMS} {PTPRC,STAT3,TNFRSF8,TYMS} {REL,STAT3,TNFRSF8,TYMS};

[0105] i.任何基因集合,其包括如下两基因组合中的任何一个或任何两个(不重复的程度),或由其组成:{CD47,CD86} {CD47,ENTPD1} {CD47,FOXP1} {CD47,FUT8} {CD47,IL16} {CD47,ITPKB} {CD47,LRMP} {CD47,MME} {CD47,NF2} {CD47,PIM1} {CD47,PTPRC} {CD47,REL} {CD47,STAT3} {CD47,TNFRSF8} {CD47,TYMS} {CD86,ENTPD1} {CD86,FOXP1} {CD86,FUT8} {CD86,IL16} {CD86,ITPKB} {CD86,LRMP} {CD86,MME} {CD86,NF2} {CD86,PIM1} {CD86,PTPRC} {CD86,REL} {CD86,STAT3} {CD86,TNFRSF8} {CD86,TYMS} {ENTPD1,FOXP1} {ENTPD1,FUT8} {ENTPD1,IL16} {ENTPD1,ITPKB} {ENTPD1,LRMP} {ENTPD1,MME} {ENTPD1,NF2} {ENTPD1,PIM1} {ENTPD1,PTPRC} {ENTPD1,REL} {ENTPD1,STAT3} {ENTPD1,TNFRSF8} {ENTPD1,TYMS} {FOXP1,FUT8} {FOXP1,IL16} {FOXP1,ITPKB} {FOXP1,LRMP} {FOXP1,MME} {FOXP1,NF2} {FOXP1,PIM1} {FOXP1,PTPRC} {FOXP1,REL} {FOXP1,STAT3} {FOXP1,TNFRSF8} {FOXP1,TYMS} {FUT8,IL16} {FUT8,ITPKB} {FUT8,LRMP} {FUT8,MME} {FUT8,NF2} {FUT8,PIM1} {FUT8,PTPRC} {FUT8,REL} {FUT8,STAT3} {FUT8,TNFRSF8} {FUT8,TYMS} {IL16,ITPKB} {IL16,LRMP} {IL16,MME} {IL16,NF2} {IL16,PIM1} {IL16,PTPRC} {IL16,REL} {IL16,STAT3} {IL16,TNFRSF8} {IL16,TYMS} {ITPKB,LRMP} {ITPKB,MME} {ITPKB,NF2} {ITPKB,PIM1} {ITPKB,PTPRC} {ITPKB,REL} {ITPKB,STAT3} {ITPKB,TNFRSF8} {ITPKB,TYMS} {LRMP,MME} {LRMP,NF2} {LRMP,PIM1} {LRMP,PTPRC} {LRMP,REL} {LRMP,STAT3} {LRMP,TNFRSF8} {LRMP,

TYMS} {MME,NF2} {MME,PIM1} {MME,PTPRC} {MME,REL} {MME,STAT3} {MME,TNFRSF8} {MME, TYMS} {NF2,PIM1} {NF2,PTPRC} {NF2,REL} {NF2,STAT3} {NF2,TNFRSF8} {NF2,TYMS} {PIM1, PTPRC} {PIM1,REL} {PIM1,STAT3} {PIM1,TNFRSF8} {PIM1,TYMS} {PTPRC,REL} {PTPRC, STAT3} {PTPRC,TNFRSF8} {PTPRC,TYMS} {REL,STAT3} {REL,TNFRSF8} {REL,TYMS} {STAT3, TNFRSF8} {STAT3,TYMS} {TNFRSF8,TYMS};以及

[0106] j.任何基因集合,其包括如下三基因组合中的任何一个或任何两个(不重复的程度),或由其组成:{CD47,CD86,ENTPD1} {CD47,CD86,FOXP1} {CD47,CD86,FUT8} {CD47, CD86,IL16} {CD47,CD86,ITPKB} {CD47,CD86,LRMP} {CD47,CD86,MME} {CD47,CD86,NF2} {CD47,CD86,PIM1} {CD47,CD86,PTPRC} {CD47,CD86,REL} {CD47,CD86,STAT3} {CD47,CD86, TNFRSF8} {CD47,CD86,TYMS} {CD47,ENTPD1,FOXP1} {CD47,ENTPD1,FUT8} {CD47,ENTPD1, IL16} {CD47,ENTPD1,ITPKB} {CD47,ENTPD1,LRMP} {CD47,ENTPD1,MME} {CD47,ENTPD1,NF2} {CD47,ENTPD1,PIM1} {CD47,ENTPD1,PTPRC} {CD47,ENTPD1,REL} {CD47,ENTPD1,STAT3} {CD47,ENTPD1,TNFRSF8} {CD47,ENTPD1,TYMS} {CD47,FOXP1,FUT8} {CD47,FOXP1,IL16} {CD47,FOXP1,ITPKB} {CD47,FOXP1,LRMP} {CD47,FOXP1,MME} {CD47,FOXP1,NF2} {CD47, FOXP1,PIM1} {CD47,FOXP1,PTPRC} {CD47,FOXP1,REL} {CD47,FOXP1,STAT3} {CD47,FOXP1, TNFRSF8} {CD47,FOXP1,TYMS} {CD47,FUT8,IL16} {CD47,FUT8,ITPKB} {CD47,FUT8,LRMP} {CD47,FUT8,MME} {CD47,FUT8,NF2} {CD47,FUT8,PIM1} {CD47,FUT8,PTPRC} {CD47,FUT8, REL} {CD47,FUT8,STAT3} {CD47,FUT8,TNFRSF8} {CD47,FUT8,TYMS} {CD47,IL16,ITPKB} {CD47,IL16,LRMP} {CD47,IL16,MME} {CD47,IL16,NF2} {CD47,IL16,PIM1} {CD47,IL16, PTPRC} {CD47,IL16,REL} {CD47,IL16,STAT3} {CD47,IL16,TNFRSF8} {CD47,IL16,TYMS} {CD47,ITPKB,LRMP} {CD47,ITPKB,MME} {CD47,ITPKB,NF2} {CD47,ITPKB,PIM1} {CD47, ITPKB,PTPRC} {CD47,ITPKB,REL} {CD47,ITPKB,STAT3} {CD47,ITPKB,TNFRSF8} {CD47, ITPKB,TYMS} {CD47,LRMP,MME} {CD47,LRMP,NF2} {CD47,LRMP,PIM1} {CD47,LRMP,PTPRC} {CD47,LRMP,REL} {CD47,LRMP,STAT3} {CD47,LRMP,TNFRSF8} {CD47,LRMP,TYMS} {CD47, MME,NF2} {CD47,MME,PIM1} {CD47,MME,PTPRC} {CD47,MME,REL} {CD47,MME,STAT3} {CD47, MME,TNFRSF8} {CD47,MME,TYMS} {CD47,NF2,PIM1} {CD47,NF2,PTPRC} {CD47,NF2,REL} {CD47,NF2,STAT3} {CD47,NF2,TNFRSF8} {CD47,NF2,TYMS} {CD47,PIM1,PTPRC} {CD47, PIM1,REL} {CD47,PIM1,STAT3} {CD47,PIM1,TNFRSF8} {CD47,PIM1,TYMS} {CD47,PTPRC, REL} {CD47,PTPRC,STAT3} {CD47,PTPRC,TNFRSF8} {CD47,PTPRC,TYMS} {CD47,REL,STAT3} {CD47,REL,TNFRSF8} {CD47,REL,TYMS} {CD47,STAT3,TNFRSF8} {CD47,STAT3,TYMS} {CD47, TNFRSF8,TYMS} {CD86,ENTPD1,FOXP1} {CD86,ENTPD1,FUT8} {CD86,ENTPD1,IL16} {CD86, ENTPD1,ITPKB} {CD86,ENTPD1,LRMP} {CD86,ENTPD1,MME} {CD86,ENTPD1,NF2} {CD86, ENTPD1,PIM1} {CD86,ENTPD1,PTPRC} {CD86,ENTPD1,REL} {CD86,ENTPD1,STAT3} {CD86, ENTPD1,TNFRSF8} {CD86,ENTPD1,TYMS} {CD86,FOXP1,FUT8} {CD86,FOXP1,IL16} {CD86, FOXP1,ITPKB} {CD86,FOXP1,LRMP} {CD86,FOXP1,MME} {CD86,FOXP1,NF2} {CD86,FOXP1, PIM1} {CD86,FOXP1,PTPRC} {CD86,FOXP1,REL} {CD86,FOXP1,STAT3} {CD86,FOXP1, TNFRSF8} {CD86,FOXP1,TYMS} {CD86,FUT8,IL16} {CD86,FUT8,ITPKB} {CD86,FUT8,LRMP} {CD86,FUT8,MME} {CD86,FUT8,NF2} {CD86,FUT8,PIM1} {CD86,FUT8,PTPRC} {CD86,FUT8, REL} {CD86,FUT8,STAT3} {CD86,FUT8,TNFRSF8} {CD86,FUT8,TYMS} {CD86,IL16,ITPKB}

{CD86, IL16, LRMP} {CD86, IL16, MME} {CD86, IL16, NF2} {CD86, IL16, PIM1} {CD86, IL16, PTPRC} {CD86, IL16, REL} {CD86, IL16, STAT3} {CD86, IL16, TNFRSF8} {CD86, IL16, TYMS} {CD86, ITPKB, LRMP} {CD86, ITPKB, MME} {CD86, ITPKB, NF2} {CD86, ITPKB, PIM1} {CD86, ITPKB, PTPRC} {CD86, ITPKB, REL} {CD86, ITPKB, STAT3} {CD86, ITPKB, TNFRSF8} {CD86, ITPKB, TYMS} {CD86, LRMP, MME} {CD86, LRMP, NF2} {CD86, LRMP, PIM1} {CD86, LRMP, PTPRC} {CD86, LRMP, REL} {CD86, LRMP, STAT3} {CD86, LRMP, TNFRSF8} {CD86, LRMP, TYMS} {CD86, MME, NF2} {CD86, MME, PIM1} {CD86, MME, PTPRC} {CD86, MME, REL} {CD86, MME, STAT3} {CD86, MME, TNFRSF8} {CD86, MME, TYMS} {CD86, NF2, PIM1} {CD86, NF2, PTPRC} {CD86, NF2, REL} {CD86, NF2, STAT3} {CD86, NF2, TNFRSF8} {CD86, NF2, TYMS} {CD86, PIM1, PTPRC} {CD86, PIM1, REL} {CD86, PIM1, STAT3} {CD86, PIM1, TNFRSF8} {CD86, PIM1, TYMS} {CD86, PTPRC, REL} {CD86, PTPRC, STAT3} {CD86, PTPRC, TNFRSF8} {CD86, PTPRC, TYMS} {CD86, REL, STAT3} {CD86, REL, TNFRSF8} {CD86, REL, TYMS} {CD86, STAT3, TNFRSF8} {CD86, STAT3, TYMS} {CD86, TNFRSF8, TYMS} {ENTPD1, FOXP1, FUT8} {ENTPD1, FOXP1, IL16} {ENTPD1, FOXP1, ITPKB} {ENTPD1, FOXP1, LRMP} {ENTPD1, FOXP1, MME} {ENTPD1, FOXP1, NF2} {ENTPD1, FOXP1, PIM1} {ENTPD1, FOXP1, PTPRC} {ENTPD1, FOXP1, REL} {ENTPD1, FOXP1, STAT3} {ENTPD1, FOXP1, TNFRSF8} {ENTPD1, FOXP1, TYMS} {ENTPD1, FUT8, IL16} {ENTPD1, FUT8, ITPKB} {ENTPD1, FUT8, LRMP} {ENTPD1, FUT8, MME} {ENTPD1, FUT8, NF2} {ENTPD1, FUT8, PIM1} {ENTPD1, FUT8, PTPRC} {ENTPD1, FUT8, REL} {ENTPD1, FUT8, STAT3} {ENTPD1, FUT8, TNFRSF8} {ENTPD1, FUT8, TYMS} {ENTPD1, IL16, ITPKB} {ENTPD1, IL16, LRMP} {ENTPD1, IL16, MME} {ENTPD1, IL16, NF2} {ENTPD1, IL16, PIM1} {ENTPD1, IL16, PTPRC} {ENTPD1, IL16, REL} {ENTPD1, IL16, STAT3} {ENTPD1, IL16, TNFRSF8} {ENTPD1, IL16, TYMS} {ENTPD1, ITPKB, LRMP} {ENTPD1, ITPKB, MME} {ENTPD1, ITPKB, NF2} {ENTPD1, ITPKB, PIM1} {ENTPD1, ITPKB, PTPRC} {ENTPD1, ITPKB, REL} {ENTPD1, ITPKB, STAT3} {ENTPD1, ITPKB, TNFRSF8} {ENTPD1, ITPKB, TYMS} {ENTPD1, LRMP, MME} {ENTPD1, LRMP, NF2} {ENTPD1, LRMP, PIM1} {ENTPD1, LRMP, PTPRC} {ENTPD1, LRMP, REL} {ENTPD1, LRMP, STAT3} {ENTPD1, LRMP, TNFRSF8} {ENTPD1, LRMP, TYMS} {ENTPD1, MME, NF2} {ENTPD1, MME, PIM1} {ENTPD1, MME, PTPRC} {ENTPD1, MME, REL} {ENTPD1, MME, STAT3} {ENTPD1, MME, TNFRSF8} {ENTPD1, MME, TYMS} {ENTPD1, NF2, PIM1} {ENTPD1, NF2, PTPRC} {ENTPD1, NF2, REL} {ENTPD1, NF2, STAT3} {ENTPD1, NF2, TNFRSF8} {ENTPD1, NF2, TYMS} {ENTPD1, PIM1, PTPRC} {ENTPD1, PIM1, REL} {ENTPD1, PIM1, STAT3} {ENTPD1, PIM1, TNFRSF8} {ENTPD1, PIM1, TYMS} {ENTPD1, PTPRC, REL} {ENTPD1, PTPRC, STAT3} {ENTPD1, PTPRC, TNFRSF8} {ENTPD1, PTPRC, TYMS} {ENTPD1, REL, STAT3} {ENTPD1, REL, TNFRSF8} {ENTPD1, REL, TYMS} {ENTPD1, STAT3, TNFRSF8} {ENTPD1, STAT3, TYMS} {ENTPD1, TNFRSF8, TYMS} {FOXP1, FUT8, IL16} {FOXP1, FUT8, ITPKB} {FOXP1, FUT8, LRMP} {FOXP1, FUT8, MME} {FOXP1, FUT8, NF2} {FOXP1, FUT8, PIM1} {FOXP1, FUT8, PTPRC} {FOXP1, FUT8, REL} {FOXP1, FUT8, STAT3} {FOXP1, FUT8, TNFRSF8} {FOXP1, FUT8, TYMS} {FOXP1, IL16, ITPKB} {FOXP1, IL16, LRMP} {FOXP1, IL16, MME} {FOXP1, IL16, NF2} {FOXP1, IL16, PIM1} {FOXP1, IL16, PTPRC} {FOXP1, IL16, REL} {FOXP1, IL16, STAT3} {FOXP1, IL16, TNFRSF8} {FOXP1, IL16, TYMS} {FOXP1, ITPKB, LRMP} {FOXP1, ITPKB, MME} {FOXP1, ITPKB, NF2} {FOXP1, ITPKB, PIM1} {FOXP1, ITPKB, PTPRC} {FOXP1, ITPKB, REL} {FOXP1, ITPKB, STAT3} {FOXP1, ITPKB, TNFRSF8}

{FOXP1, ITPKB, TYMS} {FOXP1, LRMP, MME} {FOXP1, LRMP, NF2} {FOXP1, LRMP, PIM1} {FOXP1, LRMP, PTPRC} {FOXP1, LRMP, REL} {FOXP1, LRMP, STAT3} {FOXP1, LRMP, TNFRSF8} {FOXP1, LRMP, TYMS} {FOXP1, MME, NF2} {FOXP1, MME, PIM1} {FOXP1, MME, PTPRC} {FOXP1, MME, REL} {FOXP1, MME, STAT3} {FOXP1, MME, TNFRSF8} {FOXP1, MME, TYMS} {FOXP1, NF2, PIM1} {FOXP1, NF2, PTPRC} {FOXP1, NF2, REL} {FOXP1, NF2, STAT3} {FOXP1, NF2, TNFRSF8} {FOXP1, NF2, TYMS} {FOXP1, PIM1, PTPRC} {FOXP1, PIM1, REL} {FOXP1, PIM1, STAT3} {FOXP1, PIM1, TNFRSF8} {FOXP1, PIM1, TYMS} {FOXP1, PTPRC, REL} {FOXP1, PTPRC, STAT3} {FOXP1, PTPRC, TNFRSF8} {FOXP1, PTPRC, TYMS} {FOXP1, REL, STAT3} {FOXP1, REL, TNFRSF8} {FOXP1, REL, TYMS} {FOXP1, STAT3, TNFRSF8} {FOXP1, STAT3, TYMS} {FOXP1, TNFRSF8, TYMS} {FUT8, IL16, ITPKB} {FUT8, IL16, LRMP} {FUT8, IL16, MME} {FUT8, IL16, NF2} {FUT8, IL16, PIM1} {FUT8, IL16, PTPRC} {FUT8, IL16, REL} {FUT8, IL16, STAT3} {FUT8, IL16, TNFRSF8} {FUT8, IL16, TYMS} {FUT8, ITPKB, LRMP} {FUT8, ITPKB, MME} {FUT8, ITPKB, NF2} {FUT8, ITPKB, PIM1} {FUT8, ITPKB, PTPRC} {FUT8, ITPKB, REL} {FUT8, ITPKB, STAT3} {FUT8, ITPKB, TNFRSF8} {FUT8, ITPKB, TYMS} {FUT8, LRMP, MME} {FUT8, LRMP, NF2} {FUT8, LRMP, PIM1} {FUT8, LRMP, PTPRC} {FUT8, LRMP, REL} {FUT8, LRMP, STAT3} {FUT8, LRMP, TNFRSF8} {FUT8, LRMP, TYMS} {FUT8, MME, NF2} {FUT8, MME, PIM1} {FUT8, MME, PTPRC} {FUT8, MME, REL} {FUT8, MME, STAT3} {FUT8, MME, TNFRSF8} {FUT8, MME, TYMS} {FUT8, NF2, PIM1} {FUT8, NF2, PTPRC} {FUT8, NF2, REL} {FUT8, NF2, STAT3} {FUT8, NF2, TNFRSF8} {FUT8, NF2, TYMS} {FUT8, PIM1, PTPRC} {FUT8, PIM1, REL} {FUT8, PIM1, STAT3} {FUT8, PIM1, TNFRSF8} {FUT8, PIM1, TYMS} {FUT8, PTPRC, REL} {FUT8, PTPRC, STAT3} {FUT8, PTPRC, TNFRSF8} {FUT8, PTPRC, TYMS} {FUT8, REL, STAT3} {FUT8, REL, TNFRSF8} {FUT8, REL, TYMS} {FUT8, STAT3, TNFRSF8} {FUT8, STAT3, TYMS} {FUT8, TNFRSF8, TYMS} {IL16, ITPKB, LRMP} {IL16, ITPKB, MME} {IL16, ITPKB, NF2} {IL16, ITPKB, PIM1} {IL16, ITPKB, PTPRC} {IL16, ITPKB, REL} {IL16, ITPKB, STAT3} {IL16, ITPKB, TNFRSF8} {IL16, ITPKB, TYMS} {IL16, LRMP, MME} {IL16, LRMP, NF2} {IL16, LRMP, PIM1} {IL16, LRMP, PTPRC} {IL16, LRMP, REL} {IL16, LRMP, STAT3} {IL16, LRMP, TNFRSF8} {IL16, LRMP, TYMS} {IL16, MME, NF2} {IL16, MME, PIM1} {IL16, MME, PTPRC} {IL16, MME, REL} {IL16, MME, STAT3} {IL16, MME, TNFRSF8} {IL16, MME, TYMS} {IL16, NF2, PIM1} {IL16, NF2, PTPRC} {IL16, NF2, REL} {IL16, NF2, STAT3} {IL16, NF2, TNFRSF8} {IL16, NF2, TYMS} {IL16, PIM1, PTPRC} {IL16, PIM1, REL} {IL16, PIM1, STAT3} {IL16, PIM1, TNFRSF8} {IL16, PIM1, TYMS} {IL16, PTPRC, REL} {IL16, PTPRC, STAT3} {IL16, PTPRC, TNFRSF8} {IL16, PTPRC, TYMS} {IL16, REL, STAT3} {IL16, REL, TNFRSF8} {IL16, REL, TYMS} {IL16, STAT3, TNFRSF8} {IL16, STAT3, TYMS} {IL16, TNFRSF8, TYMS} {ITPKB, LRMP, MME} {ITPKB, LRMP, NF2} {ITPKB, LRMP, PIM1} {ITPKB, LRMP, PTPRC} {ITPKB, LRMP, REL} {ITPKB, LRMP, STAT3} {ITPKB, LRMP, TNFRSF8} {ITPKB, LRMP, TYMS} {ITPKB, MME, NF2} {ITPKB, MME, PIM1} {ITPKB, MME, PTPRC} {ITPKB, MME, REL} {ITPKB, MME, STAT3} {ITPKB, MME, TNFRSF8} {ITPKB, MME, TYMS} {ITPKB, NF2, PIM1} {ITPKB, NF2, PTPRC} {ITPKB, NF2, REL} {ITPKB, NF2, STAT3} {ITPKB, NF2, TNFRSF8} {ITPKB, NF2, TYMS} {ITPKB, PIM1, PTPRC} {ITPKB, PIM1, REL} {ITPKB, PIM1, STAT3} {ITPKB, PIM1, TNFRSF8} {ITPKB, PIM1, TYMS} {ITPKB, PTPRC, REL} {ITPKB, PTPRC, STAT3} {ITPKB, PTPRC, TNFRSF8} {ITPKB, PTPRC, TYMS} {ITPKB, REL, STAT3} {ITPKB, REL, TNFRSF8}

{ITPKB,REL,TYMS} {ITPKB,STAT3,TNFRSF8} {ITPKB,STAT3,TYMS} {ITPKB,TNFRSF8,TYMS} {LRMP,MME,NF2} {LRMP,MME,PIM1} {LRMP,MME,PTPRC} {LRMP,MME,REL} {LRMP,MME,STAT3} {LRMP,MME,TNFRSF8} {LRMP,MME,TYMS} {LRMP,NF2,PIM1} {LRMP,NF2,PTPRC} {LRMP,NF2,REL} {LRMP,NF2,STAT3} {LRMP,NF2,TNFRSF8} {LRMP,NF2,TYMS} {LRMP,PIM1,PTPRC} {LRMP,PIM1,REL} {LRMP,PIM1,STAT3} {LRMP,PIM1,TNFRSF8} {LRMP,PIM1,TYMS} {LRMP,PTPRC,REL} {LRMP,PTPRC,STAT3} {LRMP,PTPRC,TNFRSF8} {LRMP,PTPRC,TYMS} {LRMP,REL,STAT3} {LRMP,REL,TNFRSF8} {LRMP,REL,TYMS} {LRMP,STAT3,TNFRSF8} {LRMP,STAT3,TYMS} {LRMP,TNFRSF8,TYMS} {MME,NF2,PIM1} {MME,NF2,PTPRC} {MME,NF2,REL} {MME,NF2,STAT3} {MME,NF2,TNFRSF8} {MME,NF2,TYMS} {MME,PIM1,PTPRC} {MME,PIM1,REL} {MME,PIM1,STAT3} {MME,PIM1,TNFRSF8} {MME,PIM1,TYMS} {MME,PTPRC,REL} {MME,PTPRC,STAT3} {MME,PTPRC,TNFRSF8} {MME,PTPRC,TYMS} {MME,REL,STAT3} {MME,REL,TNFRSF8} {MME,REL,TYMS} {MME,STAT3,TNFRSF8} {MME,STAT3,TYMS} {MME,TNFRSF8,TYMS} {NF2,PIM1,PTPRC} {NF2,PIM1,REL} {NF2,PIM1,STAT3} {NF2,PIM1,TNFRSF8} {NF2,PIM1,TYMS} {NF2,PTPRC,REL} {NF2,PTPRC,STAT3} {NF2,PTPRC,TNFRSF8} {NF2,PTPRC,TYMS} {NF2,REL,STAT3} {NF2,REL,TNFRSF8} {NF2,REL,TYMS} {NF2,STAT3,TNFRSF8} {NF2,STAT3,TYMS} {NF2,TNFRSF8,TYMS} {PIM1,PTPRC,REL} {PIM1,PTPRC,STAT3} {PIM1,PTPRC,TNFRSF8} {PIM1,PTPRC,TYMS} {PIM1,REL,STAT3} {PIM1,REL,TNFRSF8} {PIM1,REL,TYMS} {PIM1,STAT3,TNFRSF8} {PIM1,STAT3,TYMS} {PIM1,TNFRSF8,TYMS} {PTPRC,REL,STAT3} {PTPRC,REL,TNFRSF8} {PTPRC,REL,TYMS} {PTPRC,STAT3,TNFRSF8} {PTPRC,STAT3,TYMS} {PTPRC,TNFRSF8,TYMS} {REL,STAT3,TNFRSF8} {REL,STAT3,TYMS} {REL,TNFRSF8,TYMS} {STAT3,TNFRSF8,TYMS}。

[0107] III. 获得基因表达信息

[0108] 基因表达是将继承的生物体基因组内信息(例如,基因;DNA区域)转变为特定的功能性产物(有时称为基因产物)的过程。基因产物包括RNA(核糖核酸)和蛋白。特定生物学样品中的基因表达可通过检测这种样品中基因产物(RNA(例如mRNA、tRNA、rRNA和/或 miRNA)或蛋白)的表达来测量。在一些实例中,基因表达通过对结合靶核酸(例如,mRNA)的一种或更多种探针测序来测量。

[0109] 在目标样品中测量基因表达,有各种技术是(或可以成为)可获得的。但是,本公开并不限于获得、测量或检测基因表达的具体方法。很多这样的技术涉及在这些样品中检测表达的基因产物(例如,核酸(例如RNA)和/或蛋白)。不依赖于其产生的基因产物而直接检测基因或染色体DNA的活性(例如,转录率)也是可能的(或成为可能的),这些技术也可用在本文公开的方法中。

[0110] 基因表达可用在确定DLBCL亚型中。例如,将两个或更多个基因(例如两个或更多个 DLBCL标签基因)的基因表达用在确定DLBCL肿瘤为GCB或非GCB(例如,ABC)亚型,还是未分类中。在一些实施方案中,该DLBCL标签包括(1) CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个,或者(2) CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个(例如全部)。DLBCL标志物的其它具体实例包括CD47、ENTPD1、FOXP1、FUT8、IL16、NF2、PIM1、STAT3、TNFRSF8和 TYMS(它们与ABC分类相关)中的两个或更多个(例如2、3、4、5、6、7、8、9或10个),或CD86、ITPKB、LRMP、MME、PTPRC和REL(它们与GCB分类相关)中的两个或更多个(例如2、3、

4、5或6个)。

[0111] A. 检测核酸基因产物

[0112] 顾名思义,核酸基因产物就是将核酸作为的基因表达产物。示例性核酸包括DNA或RNA,例如cDNA、编码蛋白的RNA(例如,mRNA)或非编码RNA(例如,长的非编码(lnc) RNA)。RNA或DNA的互补链之间的碱基配对(即,核酸杂交)形成了用于检测核酸基因产物的主要代表性技术的全部或部分基础。其它的代表性检测技术涉及核酸测序,其可以涉及或可以不涉及杂交步骤和/或生物信息学步骤(例如,将核酸序列信息与其对应的基因相关联)。这些以及其它的检测核酸的方法为本领域所知,尽管本文描述代表性技术,但本公开不限于核酸检测的特定方法。

[0113] 1. 可选的核酸分离

[0114] 在一些实例中,在将样品中的这些核酸与互补核酸探针或引物接触和/或以其他方式检测样品中的这些核酸之前,从测试样品中分离或提取核酸。可根据多种方法中的任何一种从样品中分离核酸(例如RNA(例如,mRNA或lncRNA)或DNA)。分离和纯化核酸的代表性方法在Chapter 3 of Laboratory Techniques in Biochemistry and Molecular Biology:Hybridization With Nucleic Acid Probes,Part I.Theory and Nucleic Acid Preparation,P.Tijssen,ed.Elsevier, N.Y.(1993)和Chapter 3 of Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes,Part I.Theory and Nucleic Acid Preparation,P.Tijssen, ed.Elsevier,N.Y.(1993)中有详细描述。类似地,用于RNA(例如,mRNA或lncRNA)提取的代表性方法为本领域所公知,并且在分子生物学的标准教科书中有公开,包括Ausubel等,Current Protocols of Molecular Biology,John Wiley and Sons(1997)。

[0115] 在从样品分离或提取核酸(例如,RNA(例如mRNA或lncRNA)或DNA)后,可以进行数种可选的其它步骤的任一种来准备这样的核酸用于检测,包括测量所分离的核酸的浓度、修复(或回收)降解或损坏的RNA、RNA逆转录和/或RNA或DNA扩增。

[0116] 在其它实例中,将样品(例如,FFPE组织样品)悬浮在缓冲液(例如,裂解缓冲液)中,并且悬浮样品中存在的核酸(例如,RNA或DNA)不从该悬浮样品中分离或提取(例如,完全或部分纯化),而是在这种悬浮液中与一种或更多种互补核酸探针(例如,NPP或NPPF)接触;因此,无需从样品分离或提取核酸(例如,RNA)。当悬浮样品中存在的核酸(例如 RNA或DNA)为交联的或固定至细胞结构而不容易分离或提取时,这种实施方案是尤其有利的。较短的(例如,少于100碱基对,例如75-25碱基对或50-25碱基对)检测探针可用在一些非提取方法实施方案中。本文描述了用于无需预先提取核酸而在样品中检测该核酸(例如,RNA)的具体方法(例如qNPA和qNPS)。

[0117] 2. 核酸杂交

[0118] 在一些实例中,在本文提供的方法中确定两个或更多个DLBCL生物标志物(例如表1 标志物中的两个或更多个)的表达水平可包括让样品与多个核酸探针(例如NPP或NPPF+CFS)或成对的扩增引物在一定条件下接触,其中该多个中每个探针或成对的引物特异于并互补于表1中的生物标志物,所述一定条件允许该多个核酸探针或成对引物与表1中它的/它们的互补生物标志物杂交。在一个实例中,该方法还可包括在该样品与该多个核酸探针(例如NPP或NPPF_s+)接触后,将该样品与消化单链核酸分子的核酸酶接触。在其它实例中,

将表2中至少两个生物标志物中的每个与由特异于每个这样的生物标志物的多个(例如,至少2、或至少5或至少10个探针,例如2、3、4、5、6、7、8、9或10个)探针(例如, NPP或NPPF)组成的“平铺探针”接触,这种设计例如可用于增强从这样的基因产物获得的信号或者检测相同基因产物的多个变体。例如,该平铺探针组中的每个探针可以结合靶核酸分子的不同区域(或轻微重叠)。

[0119] 在一些实例中,通过核酸杂交检测核酸。核酸杂交涉及在探针和其互补靶通过互补碱基配对可形成稳定的杂交双链体条件下提供探针和靶核酸(例如,表1那些中的至少两个)。在一些实例中,接着除去不形成杂交双链体的核酸(例如,洗去、通过核酸酶消化或物理去除),留下准备例如通过检测(直接或间接)连接的可检测标签来检测的杂交核酸。在具体实例中,不形成杂交双链体的核酸,例如不与其各自靶杂交的任何过量探针和与探针互补的靶序列区域,可通过添加核酸酶来消化掉,仅留下互补探针的靶核酸的杂交双链体。

[0120] 通常认识到,通过增加温度和/或降低含有核酸的缓冲液的盐浓度来使核酸变性的情况下,在低严格条件(例如,低温和/或高盐)下,即使退火序列不完全互补也会形成杂交双链体(例如,DNA:DNA、RNA:RNA、或RNA:DNA)。因此,在较低严格下,杂交特异性降低。相反,在较高严格时(例如,较高温度或较低盐)成功杂交需要较少错配。本领域技术人员应理解,可设计杂交条件来提供不同程度的严格性。可增加杂交强度而不降低杂交严格性,因此可通过在探针中包括非天然碱基、例如通过包括锁核酸或肽核酸而在高严格性缓冲液中维持杂交特异性。

[0121] 由这些方法检测的核酸表达和/或核酸存在的变化例如可包括这些核酸的功能活性水平(量)、它们的表达或翻译成蛋白的增加或降低,或者它们的定位或稳定性。例如相对于标准化生物标志物,增加或降低可为特定核酸如与显示在表1中的生物标志物对应的核酸的表达至少1倍、至少2倍、至少5倍(例如约1倍、2倍、3倍、4倍、5倍)的变化(增加或降低)和/或存在的变化。

[0122] 在一个实例中,采用多重技术(multiplexed methodology)测量基因表达。在这些方法中,可在单一样品中进行多个测量(例如,基因表达测量)。已经开发了允许在单一样品中监测大量基因的各种技术(例如,传统微阵列、多重PCR、基因表达系列分析(SAGE;例如,美国专利号5,866,330)、多重连接依赖性探针扩增(MLPA)、高通量测序、基于标记微珠的技术(例如,美国专利5,736,330和6,449,562)、数字分子条形码技术(例如,美国专利号7,473,767))。

[0123] 阵列为一组用于多重检测基因表达的工具。阵列为元素(例如,分析物捕获试剂(例如,靶特异的寡核苷酸探针、适体(aptamer)或抗体))的系统性排列,其中一组值(例如,基因表达值)可与识别关键字(identification key)相关联。采用分开的可识别表面(例如,流动通道或微珠),或通过它们的组合,可以在单一表面(例如,通过空间映射(spatial mapping)或通过差异标记(differential tagging))上系统地鉴定阵列元素。

[0124] 其它有用的实施方案涉及高通量技术,通过它可以同时质询(query)多个样品。还考虑高通量、多重实施方案(在多个样品中同时测量多个基因的表达)。可用于检测所公开的生物标志物的方法和测定系统的实例为公开在国际专利公开号WO 2003/002750和WO 2008/121927、WO 1999/032663、WO 2000/079008、WO/2000/037684和WO 2000/037683和美国专利号6,232,066、6,458,533、6,238,869和7,659,063中公开的高通量测定技术,通过

引用将它们并入本文,只要它们描述高通量测定技术。

[0125] 在一些阵列实施方案中,将设计来(直接或间接地)捕获显示在表1中的基因的两种或更多种产物的感兴趣核酸序列(例如寡核苷酸)覆盖或排列在微芯片基底上。例如,该阵列可包括与显示在表2中的基因中至少两个(例如显示在表1中的至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10个或全部16个基因)具有足够互补性的寡核苷酸。在其它实例中,该阵列可包括互补于核酸酶保护探针(NPP)的一部分的寡核苷酸,而该核酸酶保护探针具有一部分互补于CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中至少两个,例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF中两个或更多个(例如显示在表1中基因的至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10个或全部16个)的产物。

[0126] 接着让排列的序列与来自测试样品(例如,从期望表征为ABC或GCB的受试者获得的DLBCL样品)的核酸如cDNA或RNA(例如,mRNA、miRNA和/或lncRNA)杂交。在一个实例中,对来自测试样品的核酸(其可以为分离的)进行标记,使得它们与阵列上特异互补的寡核苷酸的杂交可被确定。作为选择,不标记测试样品核酸,而是采用夹心测定法,例如采用额外的互补于靶的经标记寡核苷酸来检测阵列上的寡核苷酸与靶核酸之间的杂交。

[0127] 在一个实施方案中,通过检测连接至样品核酸或直接或间接杂交至靶核酸的核酸探针的一个或更多个标记物来检测杂交的核酸。这种标记物可通过数种方法的任一种来并入。在一个实例中,该标记物在样品核酸制备的扩增步骤期间并入。因此,例如,采用标记的引物或标记的核苷酸的聚合酶链式反应(PCR)将提供被标记的扩增产物。在一个实施方案中,采用经标记核苷酸(例如荧光素标记的UTP和/或CTP)的转录扩增将标记物并入转录的核酸。

[0128] 适用于所公开方法(例如,用于核酸或蛋白检测或测序)任一种的可检测标记物包括任何直接或间接偶联至另一分子(例如核酸分子或核苷酸)以协助检测该分子的化合物或组分,例如可通过光谱学、光化学、生物化学、免疫化学、电学、光学或化学手段检测的。标记物的非限定性实例包括发荧光和产荧光的部分、显色部分、半抗原、亲和标签、酶以及放射性同位素。有用的标记物包括用于与标记的链霉亲和素缀合物一起染色的生物素、磁珠(例如DYNABEADS™)、荧光染料(例如,荧光素、德克萨斯红、罗丹明、绿色荧光蛋白等)、放射标记物(例如,³H、¹²⁵I、³⁵S、¹⁴C、或³²P)、酶(例如,辣根过氧化物酶、碱性磷酸酶和其它常用在ELISA中的酶)以及比色标记物例如胶体金或彩色玻璃或塑料(例如,聚苯乙烯、聚丙烯、乳胶等)珠。教导这些标记物使用的专利包括美国专利号3,817,837;美国专利号3,850,752;美国专利号3,939,350;美国专利号3,996,345;美国专利号4,277,437;美国专利号4,275,149;以及美国专利号4,366,241。

[0129] 检测这些标记物的手段也是公知的。因此,例如,可以采用照相胶片或闪烁计数器检测放射性标记物,可以采用光检测器检测发射的光来检测荧光标记物。酶标记物典型地通过向酶提供底物并检测酶对底物作用产生的反应产物来检测,而比色标记物通过简单地观察有色标记物来检测。

[0130] 可以在杂交之前或之后将标记物添加至靶(样品)核酸。所谓的“直接标记物”为可检测标记物,它们在杂交前直接连接至或并入靶(样品)核酸中。相反,所谓的“间接标记物”在杂交后连接至杂交双链体。该间接标记物经常连接至在杂交前已连接至靶核酸的结合部

分。因此,例如,可以在杂交前将靶核酸生物素化。杂交后,亲和素缀合的荧光团将结合带有生物素的杂交双链体,提供容易检测的标记物(参见Laboratory Techniques in Biochemistry and Molecular Biology, Vol.24:Hybridization With Nucleic Acid Probes, P.Tijssen, ed. Elsevier, N.Y., 1993)。

[0131] 原位杂交(in situ hybridization, ISH), 例如显色原位杂交(CISH)和银原位杂交(SISH), 为检测和比较目标基因(例如表1中基因的至少两个)表达的示例性方法。ISH为一种类型的杂交, 其采用互补核酸在组织的一部分或组织切片中定位一种或更多种特异核酸序列(原位), 或者如果该组织太小, 则在整个组织中定位(整体ISH(whole mount ISH))。RNA ISH可用于测定组织中的表达模式, 例如表1中生物标志物的表达。DNA ISH(例如CISH和SISH)可用于在基因组水平上检测核酸。

[0132] 处理样品细胞或组织以增强它们的渗透性, 使得探针, 例如特异于表1中的生物标志物的探针, 能够进入细胞。将探针加入经处理的细胞, 让其在相关温度下杂交, 并且洗去过量探针。可以用诸如放射性、荧光或抗原性标签的可检测标记物来标记互补性探针, 以便可例如采用放射自显影、荧光显微镜或免疫测定来确定组织中该探针的位置和数量。

[0133] 原位PCR为在ISH前的基于PCR的靶核酸序列扩增。为了检测RNA, 在原位PCR之前引入细胞内逆转录步骤, 以从RNA模板生成互补DNA。这使得能够检测低拷贝的RNA 序列。在原位PCR前, 将细胞或组织样品固化和透化, 以便保持形态并允许PCR试剂接近待扩增的细胞内序列。之后在维持在悬浮液中的完整细胞中, 或者直接在玻璃载玻片上的细胞离心制备物或组织切片中, 进行靶序列的PCR扩增。在前一种方式中, 采用常规热循环仪对悬浮在PCR反应混合物中的经固定细胞热循环。PCR后, 将细胞离心到玻璃载玻片上, 通过ISH或免疫组织化学观察细胞内PCR产物。通过在盖玻片下以PCR混合物覆盖样品并随后密封以防止反应混合物蒸发来进行玻璃载玻片上原位PCR。通过将玻璃盖玻片直接置于常规或特别设计的热循环仪的加热模块上或者通过使用热循环炉来实现热循环。

[0134] 通常通过两种不同的技术之一实现对细胞内PCR产物的检测: 通过以PCR产物特异性探针进行的ISH进行的间接原位PCR, 或者通过直接检测经标记的核苷酸(例如地高辛-11-dUTP、荧光素-dUTP、3H-CTP或生物素-16-dUTP)进行的直接原位PCR(无需ISH), 这些经标记的核苷酸在热循环期间已掺入PCR产物中。

[0135] 3. 核酸扩增

[0136] 在一些实例中, 将靶核酸分子(例如核酸基因产物(例如, mRNA或lncRNA))进行扩增, 作为对它们的检测的手段(或者作为它们检测中的步骤)。在一些实例中, 在扩增期间, 例如通过使用实时RT-PCR, 确定核酸表达水平。由此, 采用扩增方法和适当引物可在样品中检测(例如定量地)表1中DLBCL生物标志物的一种或更多种。

[0137] 可使用的体外扩增方法的实例包括, 但不限于, 定量实时PCR、实时定量RT-PCR、链置换扩增; 无转录恒温扩增; 修复链式反应扩增; 连接酶链式反应扩增; 缺口填补连接酶链式反应扩增; 偶联的连接酶检测和PCR以及NASBA™RNA无转录扩增。在一个实例中, 使用基于连接的扩增方法。

[0138] 4. RNA测序

[0139] RNA测序可用于获得多重以及在一些实施方案中获得高通量基因表达信息, 例如关于表1中标志物表达的信息。RNA测序的方法是已知的(例如, 参见Chu和Corey, Nuc. Acid

Therapeutics, 22:271 (2012))。全转录组测序和靶向RNA测序技术每个都是可获得的并可用在所公开的方法中。用于基于测序的基因表达分析的代表性方法包括基因表达系列分析(SAGE)、通过大规模平行标签测序(MPSS)进行的基因表达分析、全转录组鸟枪测序(也称, WTSS或RNA-Seq), 或者核酸酶保护测序(也称, qNPS或NPSeq; 参见PCT公开号WO 2012/151111; 下文更详细讨论)。

[0140] 5. 定量核酸酶保护测定(qNPA)

[0141] 在一些实例中, 样品中其表达待测量的核酸分子采用定量核酸酶保护测定在样品中检测, 例如在国际专利申请WO 99/032663; WO 00/037683; WO 00/037684; WO 00/079008; WO 03/002750; WO 08/121927; 以及WO 2012/151111和美国专利号6,238,869; 6,458,533; 7,659,063和8,741,564中描述的, 通过引用将它们全文并入本文。还参见, Martel等, Assay and Drug Development Technologies, 2002, 1(1-1):61-71; Martel等, Progress in Biomedical Optics and Imaging, 2002, 3:35-43; Martel等, Gene Cloning and Expression Technologies, Q. Lu and M. Weiner, Eds., Eaton Publishing, Natick (2002); Seligmann Pharmacogenomics, 2003, 3:36-43; Martel等, "Array Formats" in "Microarray Technologies and Applications," U.R. Muller and D. Nicolau, Eds., Springer-Verlag, Heidelberg (2005); Sawada等, Toxicology in Vitro, 20:1506-1513, 2006; Bakir等, Bioorg. & Med. Chem Lett, 17:3473-3479, 2007; Kris等, Plant Physiol. 144:1256-1266, 2007; Roberts等, Laboratory Investigation, 87:979-997, 2007; Rimsza等, Blood, 2008 Oct 15, 112(8):3425-3433; Pechhold等, Nature Biotechnology, 27:1038-1042, 2009。通过引用将它们全部都完全并入本文。

[0142] 采用qNPA方法, 让核酸酶保护探针(NPP)与靶序列(例如表1中的序列)杂交, 接着将样品与消化单链核酸分子的核酸酶一起温育。于是, 如果检测到NPP, (例如, 其未被核酸酶消化), 那么探针的靶例如显示在表1中的靶核酸则存在于样品中, 并且可对其存在进行定量。可为各个靶设计NPP, 并将它们作为用于识别的混合物(cocktail)添加到测定中(例如, 在阵列上(称为qNPA))或通过测序(称为qNPS)。因此可在相同测定和/或阵列内测量多个基因靶(例如通过采用多个NPP)。

[0143] NPP为核酸分子(例如DNA或RNA), 与靶核酸(例如, mRNA)有足够的互补性, 使得其能够特异杂交至靶核酸。NPP防止互补的靶核酸分子被核酸酶(例如特异于单链核酸的核酸酶)切割。在一些实例中, NPP为至少35个核苷酸(例如40至80或50至150个核苷酸), 并与靶核酸分子有至少90%、至少95%或100%互补性。NPP可包括非天然碱基。在一些实例中, 所公开的方法用于采用多个NPP(例如NPPF, 见下文)对样品中的几种不同靶核酸分子(例如列在表1中的那些)进行检测(或测序), 其中每个NPP(或NPPF)特异结合特定的靶核酸分子。

[0144] 在具体实例中, 该NPP还在5'-末端和/或3'-末端包括一个或多个侧翼序列, 称为NPPF(见U.S. 8,741,564)。该NPPF包括与靶核酸分子的全部或部分互补的序列, 因而允许靶核酸分子和该NPPF之间特异结合或杂交。例如, 与靶核酸分子区域互补的NPPF区域以高特异性结合或杂交靶核酸分子的该区域(并且在一些实例中还可结合双功能接头区域)。与靶核酸分子区域互补的NPPF部分可为至少6个核苷酸长, 例如至少10个、至少25个或至少60个, 例如6至60个核苷酸长。该NPPF还在NPPF的5'-末端和/或3'-末端包括一个或多个侧翼序列。因此, 该一个或多个侧翼序列位于互补于靶核酸分子的序列的5'、3'或二者。如

果该NPPF在5'-末端和3'-末端都包括侧翼序列,在一些实例中,该每个NPPF的该序列不同且不互补。该侧翼序列包括几个邻接核苷酸,它们具有存在于样品中的核酸分子中所没有的序列(例如至少6个核苷酸、至少12个核苷酸或至少25个核苷酸,例如12至50个核苷酸的序列),提供了通用杂交和/或扩增序列。这种通用杂交和/或扩增序列,当具有与扩增引物的至少一部分互补的序列时,允许进行多重操作,因为该相同的扩增引物可用于扩增特异于不同靶核酸分子的NPPF。还提供了用于所有NPPF的通用杂交序列,其可用于将可检测标记物添加至NPPF或捕获和浓缩NPPF。例如,如果该相同侧翼序列存在于特异于不同靶核酸分子的NPPF上,则该相同引物可用于扩增具有该相同侧翼序列的任何NPPF,即使该NPPF靶向不同的核酸分子。例如,侧翼序列可用于将NPPF例如捕获在表面上。侧翼序列可含有可变量序列,例如特异于每个特异NPPF的序列,可用于将NPPF捕获在表面上或用于其它目的,例如鉴定NPPF。在一些实例中,该NPPF为至少35个核苷酸,例如40至80或50至150个核苷酸。在一些实例中,该NPPF包括两个侧翼序列:一个在5'-末端,另一个在3'-末端。在一些实例中,在5'-末端的侧翼序列不同于在3'-末端的序列。另外,如果NPPF包括两个侧翼序列,理想的是该两个侧翼序列具有类似的解链温度(melting temperature, T_m),例如 $\pm 5^\circ\text{C}$ 的 T_m 。

[0145] 在一些实施方案中,该NPP或NPPF特异杂交和/或互补于靶序列的全部或部分。因此,该NPP或NPPF可包括与靶核酸分子相互补的至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少55、至少60、至少65、至少70、至少75个或更多个连续核苷酸。在一些实例中,该NPP或NPPF为与靶核酸分子相互补的不超过500个核苷酸,例如不超过400个、不超过300个、不超过250个、不超过200个、不超过100个、不超过50个或甚至不超过25个的连续核苷酸(例如为与靶核酸分子相互补的10至500个核苷酸、10至400个核苷酸、10至250个核苷酸、10至200个核苷酸、10至100个核苷酸、25至75个核苷酸、10至60个核苷酸、40至80核苷酸、100至200个核苷酸或10至50个的连续核苷酸)。在具体实例中,该NPP和NPPF为至少10个核苷酸长度,例如与靶核酸序列或其部分相互补的至少10个连续氨基酸,例如本文公开的靶序列(例如,表1中的那些)。可用于实施本公开的方法的NPP和NPPF的具体长度包括具有与靶核酸分子或其部分相互补的至少10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75个或更多个连续核苷酸的探针。在一些实例中,NPP和NPPF包括与靶序列的反向互补序列有至少95%序列一致性(例如至少95%、96%、97%、98%、99%或甚至100%序列一致性)的核酸序列(或NPP由其组成)。在一些实例中,与靶核酸分子的反向互补序列相比,NPP和NPPF可包括1、2、3或更多个错配。

[0146] 在一些实例中,NPP和NPPF可包括一个或更多个(例如1、2、3或更多个)合成碱基或可变碱基(例如次黄嘌呤核苷)。在其它实例中,NPP可包括一个或更多个修饰核苷酸或核苷酸类似物,例如一个或更多个锁定核酸(参见,例如美国专利号6,794,499)或一个或更多个肽核酸。可以在NPP或NPPF的一个或更多个位点(例如1、2、3、4、5或更多个位点)使用修饰核苷酸、非天然核苷酸、合成的或可变换核苷酸(alternative nucleotide)。NPP和NPPF还可以在一个或更多个位点(例如1、2、3、4、5或更多个位点)为简并的(degenerate),例如,在NPP或NPPF中特定位点包括核苷酸的混合(例如2、3或4个核苷酸)的NPP或NPPF。

[0147] 在一些实施方案中,来自受试者的样品(例如包括核酸例如RNA的样品)可与多个

NPP 或NPPF接触,其中NPP和NPPF的至少一些特异结合至少两种不同的靶核酸分子(例如,表1中显示的那些)。在一些实例中,NPP或NPPF特异结合CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的至少两个(例如这些中的1、2、3、4、5、6、7、8、9个或全部10个),例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个(例如这些中的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15个或全部16个)。在一些实施方案中,来自受试者的样品(例如包括核酸例如RNA的样品)还与一个或更多个特异于一个或更多个管家(housekeeper)基因/标准化基因的NPP或NPPF(例如,一个或更多个特异于ANT的NPP或NPPF)接触。

[0148] 在一些实例中,该多个NPP或NPPF包括不止一个(例如2、3、4、5、6、7、8、9、10、11、12、13、14、15或16个或更多个)不同的NPP或NPPF,每个特异于单一靶mRNA或靶mRNA的独特区域(例如,参见表2中NPPF序列的靶特异性部分)。该多个NPP或NPPF与样品在该NPP或NPPF足以特异杂交它们各自的靶mRNA的条件下一同温育。如果使用NPPF,该方法还包括让样品与包含与侧翼序列互补的序列的核酸分子(CFS)在足以让NPPF的侧翼序列特异结合CFS的条件下接触。让样品与特异于单链核酸的核酸酶(例如,S1核酸酶)接触,并对每个NPP或NPPF的存在进行检测(或测序)。在一些实例中,该NPP或NPPF在检测(或测序)前被扩增。当检测到它们各自的NPP或NPPF时,mRNA被鉴定为存在于样品中和/或被定量。

[0149] a. 样品的处理

[0150] 在一些实例中,首选将来自受试者的样品(例如,细胞或组织)在水溶液中裂解或透化(permeabilized)(例如采用裂解缓冲液)。让细胞在足够温度下(例如约22°C至约115°C,例如约37°C至约105°C或约90°C至约110°C)在该水溶液中温育足够长的时间(例如约1分钟至约60分钟,例如约5分钟至约20分钟或约10分钟)以裂解或透化细胞。在一些实例中,裂解在约95°C进行。在一些实例中,该裂解步骤包括让样品在约95°C温育约5至15分钟,以便让样品中的RNA变性,而基因组DNA不变性。在其它实例中,该裂解步骤包括将样品在约105°C温育约5至15分钟以让样品中的核酸变性。在一个实例中,在裂解缓冲液中包括有蛋白酶K。

[0151] 在一些实例中,可将一个或更多个与一个或更多个靶互补的NPP或NPPF(例如包括或具有显示在SEQ ID NO:1-16中的序列的那些)以在缓冲液中约10pM至约10nM(例如约30pM至5nM、约100pM至约1nM)的浓度范围添加至样品中,该缓冲液例如为含有NaCl、KCl、 H_2PO_4 、EDTA、0.05% Triton X-100,或其组合(例如,6X SSPE-T:0.9M NaCl、60 mM NaH_2PO_4 、6mM EDTA和0.05% Triton X-100)的缓冲液或裂解缓冲液。在一个实例中,每个NPP或NPPF以至少10pM、例如至少20pM、至少30pM、至少50pM、至少100pM、至少150pM、至少200pM、至少500pM、至少1nM或至少10nM的终浓度添加至样品中。在一个实例中,每个NPP或NPPF以约30pM的终浓度添加至样品中。在另一实例中,每个NPP或NPPF以约167pM的终浓度添加至样品中。在其它实例中,每个NPP或NPPF以约1nM的终浓度添加至样品中。在一个实例中,如果使用NPPF,则每个CFS以约至少6倍探针量,例如至少10倍或至少20倍探针量(例如6至20倍探针量)的终浓度添加至样品中。在一个实例中,每个CFS以至少1nM、至少5nM、至少10nM、至少50nM、至少100nM、至少200nM,例如1至100、5至100或5至50nM加入。例如,如果有六个探针,每个166pM,则每个CFS以5至50nM加入。在这些实例中,如果NPP或NPPF杂交至互补序列,例如靶序列(与其形成双链体),则NPP或NPPF不被诸如S1的核酸酶消化。

[0152] b. 示例性杂交条件

[0153] 本领域技术人员可鉴定足以用于NPP或NPPF (+CFS) 与其存在于测试样品中的靶进行特异杂交的条件。例如,本领域技术人员可通过试验确定将使得核酸(例如融合探针)在所选严格性条件下杂交另一核酸(例如,表2至8任一个中的靶核酸)、同时让与其它物质或分子非特异杂交最小化的特性(例如长度、碱基组成以及互补程度)。典型地,该NPP的核酸序列将与相应的靶序列有足够互补性,使得其能够在所选的严格杂交条件下杂交,例如在约37°C或更高温度(例如约37°C、42°C、50°C、55°C、60°C、65°C、70°C、75°C或更高)杂交。杂交反应参数中可变的有盐浓度、缓冲剂、pH、温度、温育时间、诸如甲酰胺的变性剂的类型等。

[0154] 让样品中的核酸分子变性(例如在约95°C至约105°C约5至15分钟)并在约4°C至约70°C的温度范围(例如,约37°C至约65°C、约42°C至约60°C或约50°C至约60°C)杂交多个NPP或NPPF (+CFS) 约10分钟至约72小时(例如约10分钟至约24小时、例如至少约1小时至约20小时或约6小时至约16小时)。在一些实例中,该多个NPP或NPPF (+CFS) 与样品在至少约37°C、至少约40°C、至少约45°C、至少约50°C、至少约55°C、至少约60°C、至少约65°C或至少约70°C的温度一起温育。在一个实例中,该多个NPP或NPPF (+CFS) 与样品在约37°C一起温育。在另一实例中,该多个NPP或NPPF (+CFS) 与样品在约42°C一起温育。在其它实例中,该多个NPP或NPPF (+CFS) 与样品在约50°C一起温育。这些杂交温度为示例性的,本领域技术人员可基于诸如NPP或NPPF (+CFS) 的长度和核苷酸组成的因素来选择适合的杂交温度。

[0155] 在一些实施方案中,该方法不包括核酸纯化(例如,在将样品与NPP或NPPF (+CFS) 接触前不进行核酸纯化和/或在样品与NPP或NPPF (+CFS) 接触后不进行核酸纯化)。在一些实例中,该方法不包括核酸扩增(例如,在将样品与NPP或NPPF (+CFS) 接触前不进行核酸扩增和/或在样品与NPP或NPPF (+CFS) 接触后不进行核酸扩增)。在一些实例中,除了细胞裂解外,不需要样品预处理。在一些实例中,细胞裂解和将样品与多个NPP或NPPF (+CFS) 接触依顺序进行。在其它实例中,细胞裂解和将样品与多个NPP或NPPF (+CFS) 接触同时进行,在一些非限定性实例中无需任何居间步骤。

[0156] c. 核酸酶处理

[0157] 在一个或更多个NPP或NPPF (+CFS) 和样品中的核酸杂交后,样品接受了核酸酶保护程序。杂交至靶核酸的NPP或NPPF不会被核酸酶水解,随后可被检测。

[0158] 以一种或更多种核酸酶处理将破坏所有的ss核酸分子(单链核酸分子)(包括样品中未与NPP或NPPF杂交(因而未被保护)的RNA和DNA、未与靶核酸杂交的NPP或NPPF和未与NPPF杂交的CFS(当使用时)),但不会破坏ds核酸分子(双链核酸分子),例如已与CFS和样品中存在的靶核酸分子杂交的NPPF和已与样品中存在的靶核酸分子杂交的NPP。例如,如果该样品包括细胞提取物或裂解物,则不需要的核酸,例如非靶基因组DNA、tRNA、rRNA、mRNA、miRNA以及不与互补的NPP或NPPF序列杂交的靶核酸分子部分(例如突出序列),可在该步骤中被充分破坏,如果mRNA或DNA核酸靶构成核酸靶序列的主要部分。这留下了化学计量的靶核酸/NPP双链体或靶核酸/CFS/NPPF双链体。如果靶分子与组织交联(产生自固定化),则NPP或NPPF杂交至交联的靶分子,而无需反交联(reverse cross-linking)或将以其它方式将靶核酸分子从其交联的组织释放。

[0159] 可采用各种核酸酶中的任一种,包括胰腺RNA酶、绿豆核酸酶、S1核酸酶、RNA酶A、

核糖核酸酶T1、外切核酸酶III、外切核酸酶VII、RNA酶CLB、RNA酶PhyM、RNA酶 U2等,这取决于杂交的复合物和样品中不需要的核酸的性质。在具体实例中,该核酸酶特异于单链核酸,例如S1核酸酶。在一些本文公开的方法实施方案中采用特异于单链核酸的核酸酶的好处是从随后的反应步骤中去除这样的单链(“粘性”)分子,在那里它们可能导致不期望的背景或交叉反应性。S1核酸酶可例如从Promega, Madison, WI (cat.no.M5761); Life Technologies/Invitrogen, Carlsbad, CA (cat.no.18001-016); Fermentas, Glen Burnie, MD (cat.no.EN0321) 以及其它地方购得。这些酶的反应条件在本领域是公知的,并且可以凭经验优化。

[0160] 在一些实例中,将稀释在缓冲液(例如含有醋酸钠、NaCl、KCl、ZnSO₄、KATHON,或其组合的缓冲液)中的S1核酸酶添加至杂交的探针/样品混合物中,并在约37°C至约60°C(例如约50°C)温育10至120分钟(例如,10至30分钟、30至60分钟、60至90分钟或120分钟),以便消化来自样品的未杂交核酸和未杂交NPP或NPPF。

[0161] 可任选地处理样品以除去未杂交的材料和/或失活或除去残余的酶(例如,通过加热、酚提取、沉淀、柱过滤、添加蛋白酶K、添加核酸酶抑制剂、核酸酶活性所需的螯合二价阳离子,或它们的组合)。在一些实例中,可任选地处理样品以让靶核酸和CFS与其互补的NPPF解离或让靶核酸与其互补的NPP解离(例如,采用碱水解和加热)。在一些实例中,在杂交和核酸酶处理后,可降解杂交至NPP或NPPF的靶核酸分子,例如通过让与NPP或NPPF的双链体在碱中解离,接着通过核酸酶或通过化学/物理处理破坏该核酸,例如在升高温度下的碱水解,留下与已杂交至靶核酸的量成正比的NPP或NPPF。作为选择,可对样品进行处理以便留下靶核酸的(单链)杂交部分或杂交的靶核酸和探针形成的双链体,进行进一步分析。

[0162] 在一些实例中,在与核酸酶温育后,添加碱(例如NaOH或KOH)以将pH升高至约9至12,并加热样品(例如至约95°C 10分钟)。这解离了二聚体,留下处于单链状态的NPP或NPPF,并且对于RNA的情形,水解RNA靶分子。该步骤还可中和或失活核酸酶,例如通过将pH升高约6以上。

[0163] 在一些实例中,处理样品以将pH调节至约7至约8,例如通过添加酸(例如HCl)。在一些实例中,pH在Tris中升高至约7至约8。升高pH可防止DNA脱嘌呤,还防止许多ss 特异性核酸酶(例如S1)完全起作用。

[0164] 在一些实例中,例如通过凝聚纯化或其它分离方法纯化或分离样品,以除去不需要的核酸或其它分子。

[0165] d. 任选的扩增

[0166] 在一些实例中,NPP或NPPF在它们检测前扩增。例如,可例如采用常规方法例如PCR或其它形式的酶学扩增或基于连接的扩增方法来扩增得到的NPP、NPPF,或得到的已与NPP或NPPF分开的靶核酸分子。

[0167] 可使用的体外扩增方法包括,但不限于,定量实时PCR、链置换扩增;无转录恒温扩增;修复链式反应扩增;连接酶链式反应扩增;缺口填补连接酶链式反应扩增;偶联的连接酶检测和PCR;以及NASBATMRNA无转录扩增。在一个实例中,使用基于连接的扩增方法,其中引物为NPPF特异性的并一起对接(butt-up together),使得它们可连接在一起、熔化(melt-off)以及随后新鲜引物连接在一起,进行一序列循环。连接可为酶学或非酶学的。如果该NPPF侧翼序列用于引物的杂交,则扩增可为通用的。

[0168] 扩增过程中, 试验标签和/或测序接头可例如作为引物和延伸构建体的一部分例如在 NPP或NPPF的3'-或5'-末端或两末端掺入。在一些实例中, 这样的标签或接头在扩增前添加。扩增也可用于向产生的NPP或NPPF扩增子(例如, 如果该NPP或NPPF最初未标记或如果需要额外标记)或其它允许检测或淬灭的分子中引入可检测标记物。例如, 该扩增引物可包括在扩增期间掺入NPP或NPPF的可检测标记物、半抗原(hapten)、淬灭物。这样的标记物、半抗原、或淬灭物可在NPP或NPPF扩增子的任一末端(或两末端)引入, 或在之间任何地方引入。

[0169] 在一些实例中, 在检测或测序前清洁(clean up)所获得的NPP或NPPF扩增子。例如可采用本领域公知的方法清洁扩增反应混合物(例如, 凝胶纯化、生物素/亲和素捕获和释放、毛细管电泳)。在一个实例中, 该NPP或NPPF扩增子被生物素化(或包括其它半抗原)并被捕获在亲和素或抗半抗原包被的微珠或表面上、洗涤, 然后释放, 用于检测或测序。所扩增的产物也可在扩增的最后步骤后仍为双链时通过采用水解单链寡核苷酸的核酸酶(例如外切核酸酶I)的方法清洁, 该核酸酶继而可在继续下一步骤(例如杂交至表面)前灭活。

[0170] e. 采用标记物检测NPP或NPPF

[0171] 可检测NPP或NPPF(或剩余的靶、靶:NPP复合物或靶:NPPF:CFS复合物)、NPP扩增子或NPPF扩增子的存在。任何适合的方法都可用于检测NPP或NPPF探针或扩增子(或剩余的靶、靶:NPP复合物或靶:NPPF:CFS复合物)。在一些实例中, NPP或NPPF(或其扩增子)包括可检测标记物, 对该NPP或NPPF的存在的检测包括检测该可检测标记物。在一些实例中, NPP或NPPF(或其扩增子)被相同的可检测标记物标记。在其它实例中, NPP或NPPF(或其扩增子)被不同的可检测标记物标记(例如针对每个靶的不同标记物)。在其它实例中, 间接检测NPP或NPPF(或其扩增子), 例如通过与经标记的核酸杂交。

[0172] 在特定的、非限定性实例中, NPP或NPPF(或其扩增子)包括生物素标记物。在该实例中, 该NPP或NPPF(或其扩增子)可通过将它们(例如在含有NPP或NPPF(或其扩增子)的支持物, 例如阵列或微珠上)与亲和素-HRP或链霉亲和素-HRP或与诸如碱性磷酸酶的另一适合酶的缀合物一起温育, 接着让该支持物与显色、化学发光或产生荧光的底物接触来检测。在一个非限定性实例中, 该底物为TMA-3(Lumigen, Southfield, MI)。另外的化学发光底物可购得, 例如LumiGlo®(KPL, Gaithersburg, MD)、SuperSignal®(Pierce, Rockford, IL)以及ECL™(Amersham/GE Healthcare, Piscataway, NJ)。底物产生的信号例如采用微阵列成像仪(例如OMIX、OMIX HD、CAPELLA或SUPERCAPELLA成像仪, HTG Molecular Diagnostics, Tucson, Arizona)扫描仪检测, 或例如在测流装置中视觉检测。可使用基于铈的发光, 以及电致发光或光散射, 或与电有关的(例如, 电导率或电阻)。

[0173] 在另一实例中, 该NPP或NPPF(或其扩增子)包括荧光标记物, 例如Cy-3或Cy-5。可采用标准微阵列成像仪(例如Typhoon™成像仪(GE Life Sciences, Piscataway, NJ)、GenePix®微阵列扫描仪(Molecular Devices, Sunnyvale, CA)、GeneChip®扫描仪(Affymetrix, Santa Clara, CA)、流式细胞术或荧光显微镜方法来检测该NPP或NPPF(或其扩增子)。本领域技术人员可选择适合的检测方法和试剂用于这些或其它的可检测标记物。

[0174] f. 采用捕获分子检测NPP或NPPF

[0175] 在一些实施方案中, 在杂交和核酸酶处理(以及任选的扩增)后, 让样品与包括多个空间上离散的区域(每个包括捕获分子)的表面接触, 或者与每个包括捕获分子的多个表

面接触。例如,该表面可为一群微珠,其中微珠亚群的每个包括至少一种捕获分子。例如,第一亚群可包括至少一种捕获分子,而第二亚群可包括不同于第一亚群的至少一种捕获分子,以此类推。在一些实例中,该捕获分子包括至少一个与双功能接头(也成为“程序接头(programming linker)”)缔合的锚。作为选择,该捕获分子包括核酸捕获探针,其具有与NPP或NPPF(或其扩增子)的至少一部分互补的序列,例如互补于NPPF或其扩增子的侧翼序列的全部或部分。

[0176] 在捕获分子包括至少一个与双功能接头缔合的锚的实例中,该锚和双功能接头通过杂交、退火、共价键或其它结合缔合。该双功能接头包括特异结合(例如,互补于)锚的第一部分和特异结合(例如,互补于)该多个NPP或NPPF(或其扩增子)之一的第二部分。

[0177] 在一些实施方案中,所公开的方法包括表面上(例如阵列上)的锚,其与双功能接头缔合,该双功能接头用于在核酸酶或扩增步骤后捕获NPP或NPPF(或其扩增子)。在一些实施方案中,锚为约8至150个核苷酸长的寡核苷酸(例如,约8至100、15至100、20至80、25至75或25至50,例如约15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140或150个核苷酸)。在一个非限定性实例中,该锚为约25个核苷酸长。在一些实例中,该锚包括特异结合双功能接头的第一部分的第一部分,和用作表面和锚的第一部分之间的间隔物的第二部分。在一些实例中,该锚的第二部分为约6至60个碳原子或核苷酸长(例如约6、12、24、30、36、42、48、54或60个碳原子或核苷酸)。在其它实例中,该锚的第二部分为约5至100个碳原子或核苷酸长(例如约10至50、15至40、20至30或约25个碳原子或核苷酸)。

[0178] 用于本公开方法的锚的碱基组成使得锚和双功能接头配对的热力学稳定性高。在一些实例中,锚的百分比碱基组成为约30-40%G、30-40%C、10-20%A和10-20%T。在一些实例中,锚中的最近邻频率(nearest neighbor frequency)让G-G或C-C最近邻最小化,以降低通过G四连体形成(G-quartet formation)介导的副反应。在其它实例中,可将非天然碱基或肽核酸掺入锚或双功能接头中以修饰其性质。

[0179] 设计和合成用在本公开方法中的锚的方法例如在PCT公开号WO 98/24098中有描述,通过引用并入本文。在一些实例中,需要一组基本上彼此不类似的锚。用于获得一组不类似锚的示例性算法如下:

[0180] 1) 定义组大小。在一些实施方案中,16、24、36、48、49、64、81、96和100构成可用的大小。

[0181] 2) 定义该组锚的总体序列结构。将上文所述的长度和碱基组成用于定义这样的参数。一般地,将G碱基和C碱基的数目保持相等,A碱基和T碱基的数目也是如此。该相等优化了最终锚组的配置多样性。因此,这样的组通过等式 $G_n C_n A_m T_m$ 描述。

[0182] 3) 对于由m和n定义的一组结构,采用随机数生成器产生一组随机序列异构体。

[0183] 4) 选择随机序列组的一个成员作为组的元素#1。

[0184] 5) 定义组成员之间可允许的最大相似性。就局部的成对碱基比较来定义相似性。

例如,当两个相同长度n的寡聚体链进行比对以使5'-末端和3'-末端对齐时,缺乏错配指这样的情况,即在所有的位置1-n,两条链中的碱基都是相同的。完全错配指这样的情况,即在所有的位置1-n,两条链中的碱基都是不同的。例如,有用的最大相似性可以为一组16个16聚体捕获探针内10个或更多个错配。

[0185] 6) 选择随机序列组的第二成员,并确定其与元素#1的相似性。如果元素#2与元素#1 具有小于最大允许相似性,则其保持在该组中。如果元素#2具有大于最大允许相似性,则将其丢弃并选择新序列来比较。重复该过程直至确定第二元素。

[0186] 7) 以有序方式,选择随机序列组的另外成员,它们相对于所有之前选择的元素满足不相似约束。

[0187] 在其它实例中,其中捕获分子包括至少一个核酸捕获探针,其具有与NPP或NPPF (或其扩增子)的至少一部分互补,例如与NPPF扩增子的侧翼序列的全部或部分互补的序列。例如,该核酸捕获探针可包括与NPP或NPPF (或其扩增子)互补的区域,并且可以包括不与其互补的区域(例如允许探针附着到表面的区域)。该核酸捕获探针可直接附着至表面。例如,该核酸捕获探针可包括用于共价附着至表面的氨。在一些实例中,核酸捕获探针为至少8个核苷酸长的寡核苷酸,例如至少10、至少15、至少20、至少30、至少50或至少100个核苷酸长(例如,约8至100、15至100、20至80、25至75、或25至50个,例如约15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140或150个核苷酸)。本领域技术人员应理解,与NPP或NPPF (或其扩增子)区域互补的核酸捕获探针区域不必100%互补,只要在核酸捕获探针和适合的NPP或NPPF (或其扩增子)之间可发生杂交即可。在一些实例中,与NPP或NPPF (或其扩增子)区域互补的核酸捕获探针区域为至少8个核苷酸长,例如至少8、至少10、至少15、至少20、至少30、至少50、至少100个核苷酸长(例如,约8至100、15至100、20至80、25至75、或25至50个,例如约15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140或150个核苷酸长)。

[0188] 在一些实例中,将含有NPP或NPPF (或其扩增子)的样品在与阵列表面接触前变性(例如通过加热至95°C 5分钟,并在冰上快速冷却样品)。在一些实例中,将含有NPP或NPPF (或其扩增子)的样品在与表面接触前进行调整(例如调整盐或甲酰胺浓度)。将含有NPP或NPPF (或其扩增子)的样品与表面(例如,阵列或微珠)一起温育足够长的时间,用于使NPP或NPPF (或其扩增子)特异结合(例如杂交)捕获分子。在一些实例中,将样品与表面在约37°C至约65°C(例如,约45°C至约60°C或约50°C至约60°C,例如50°C)温育至少1小时(1至8小时、1至36小时、12至24小时或16至24小时,或过夜),以使NPP或NPPF (或其扩增子)杂交至捕获分子。例如,如果采用微流体或宏流体(macrofluidic)装置、测流装置,或者通过减少扩散并使用主动流动或混合,捕获时间可缩短。

[0189] 可用作本公开方法中的一些表面(或基体)可容易地从供应商获得。在一些实施方案中,该表面为96孔、384孔或1536孔微量滴定板,例如Corning Costar销售的经修饰的板。在其它实施方案中,基底包括一种或多种微珠(例如可通过大小或颜色,例如通过流式细胞术区分的微珠群体)。作为选择,可通过如下方式形成包括孔(well)、孔又包含凹陷(indentation)或“小窝(dimples)”的表面:微机械加工诸如铝或钢的物质来制备磨具,然后将塑料或类似材料微注射进该磨具中以形成一结构。或者,可组装由玻璃、塑料、陶瓷等构成的结构。分隔物(separator)可例如为一块材料,例如硅树脂,布满了隔开的洞(hole),这样当三块接合时每个洞将形成测试孔的壁。细分物(subdivider)可例如为薄片材料,例如硅树脂,成型为筛网(screen)或细网(fine meshwork)形式。分隔不同反应的表面上的分割物(divider)还可以是有涂层的表面(不附着溶液)、纳米结构,或者简单地是独立液滴,或毛细或微流体通道或位置。在一些实例中,基底(base)为成块的平坦材料(例如玻璃或塑

料),具有例如用于生物化学测定的典型微孔板的下部的形状。该基底的上表面可为平坦的,或者可形成将与细分物形状对齐的凹陷以在每个样品孔内提供完全的细分或孔。这三块可通过标准程序拼接,例如用在硅片组装中的程序。

[0190] 按照本公开,可使用各种各样的用于排列锚的阵列形式。一种适合的形式包括二维模式的离散孔(例如64×64阵列中的4096个方块)。本领域技术人员应理解,其它的阵列形式,包括但不限于槽(矩形)和环形阵列,也是同样适用的(参见美国专利5,981,185)。在一些实例中,该阵列为多孔板。

[0191] 在一个实施方案中,具有与NPP或NPPF(或其扩增子)的至少一部分互补的序列的预先形成的核酸锚(例如,寡核苷酸锚)或核酸捕获探针可通过各种常规技术的任一种而位于测试区域的表面上或表面内,包括光刻或丝网化学附着,通过喷墨技术布置,毛细管、筛网或流体通道芯片,使用电极阵列的电化学图案形成,与针或刺接触,或变性,随后烘烤或UV 辐射至过滤器上(参见,例如,Rava等,(1996).U.S.Patent No.5,545,531;Fodor等,(1996).U.S.Patent No.5,510,270;Zanzucchi等,(1997).U.S.Patent No.5,643,738;Brennan(1995).U.S.Patent No.5,474,796;PCT WO 92/10092;PCT WO 90/15070)。寡核苷酸锚或探针可被置于测试区域的表面上,或者例如对于聚丙烯酰胺凝胶垫的情况,可以以这样的方式被包埋在表面内:使得锚或探针的一部分从凝胶结构伸进凝胶内的水性部分和凝胶表面,并可用于与接头、NPP、NPPF(或其扩增子)相互作用。对于可渗透表面和部分可渗透表面来说确实是这样,例如这样的表面,第一部分(例如与含有双功能接头、NPP、或NPPF的溶液接触的表面的区域)为可渗透的,而第二部分(例如表面内一定距离处)不是可渗透的。在一个实施方案中,预先形成的寡核苷酸锚或探针在5'-末端以自由氨基基团衍生化;以通常经验上确定的浓度(例如,约1 μ M)溶解在诸如50mM磷酸盐缓冲液,pH 8.5和1mM EDTA的缓冲液中;并通过Pixus纳米喷射分配器(Cartesian Technologies)以约10.4纳升的小滴分布在测试孔内的特定位置,该测试孔的上表面为新鲜、干燥的DNA结合板(Corning Costar)表面。根据寡核苷酸附着和蒸发的相对速度,可能需要在制备过程中控制孔内的湿度。在另一实施方案中,可用常规方法直接在测试区表面上合成寡核苷酸锚或探针,例如,延伸寡核苷酸链的光活化脱保护(例如,与使用定点“屏蔽”联合)或通过采用纳米喷射分配器图案化分配去活化化合物(deactivating compound)的纳升液滴。例如,可完成所有有待接收单核苷酸的延伸寡核苷酸的脱保护,接着将该核苷酸在整个表面添加。在另一实施方案中,采用常规技术将寡核苷酸锚或探针通过寡核苷酸的3'-末端附着至表面。

[0192] g. 采用备选方法检测NPP或NPPF

[0193] 在一些实施方案中,在杂交、核酸酶处理以及任选的扩增后,采用诸如高通量平台的替代方法检测NPP或NPPF(或其扩增子)。在一些实例中,NPP或NPPF(或其扩增子)采用凝胶电泳、色谱、质谱、测序、常规微阵列分析检测,在扩增或杂种捕获期间检测。在一些实施方案中,该NPP或NPPF(或其扩增子)不包括可检测标记物,并采用间接检测方法。这些方法为本领域技术人员所知,并包括但不限于本文中描述的那些。

[0194] 在一个实例中,NPP或NPPF(或其扩增子)采用基于微珠的测定法(例如微珠阵列)检测。基于微珠的测定法的一个实例采用X-MAP®微珠(Luminex,Austin,TX),例如QBEAD测定。在一些实例中,该NPP或NPPF(或其扩增子)通过杂交至与微珠缔合的寡核苷酸而被捕获在X-MAP®微珠或其它微珠上(例如在约50°C约1小时)。可例如通过流式细胞术(例如采

用Luminex 200、Flexmap 3D或其它适合仪器)检测包括在该NPP或NPPF(或其扩增子)中的可检测标记物。

[0195] 在另一实例中,采用标准微阵列检测NPP或NPPF(或其扩增子)。这种阵列的一个实例为Nimblegen微阵列(Nimblegen, Madison, WI)。在一些实例中,该NPP或NPPF(或其扩增子)杂交至包括特异结合该NPP或NPPF(或其扩增子)的寡核苷酸的阵列。可检测包括在该NPP或NPPF(或其扩增子)中的可检测标记物。

[0196] 在一些实例中,用“条形码”测定法检测NPP或NPPF(或其扩增子)。这种测定的一个实例为nCounter® Analysis System(Nanostring Technologies, Seattle, WA)。在一些实例中,该NPP或NPPF(或其扩增子)杂交至包括一种或更多种彩色编码标签(“条形码”)的探针。对彩色编码标签的检测提供对该NPP或NPPF(或其扩增子)的鉴定。参见,例如,WO 07/0761282;WO 07/076129;WO 07/139。

[0197] h. 对扩增子测序(qNPS)

[0198] 在一些实例中,例如通过对全长NPP或NPPF(或其扩增子)或其部分(例如足以允许鉴定靶核酸分子的量)测序,来对得到的NPP或NPPF(或其扩增子)测序。本公开不限于具体的测序方法。在一些实例中,在单一反应中对不同的NPP或NPPF(或其扩增子)测序。在一个实例中,可对NPP或NPPF(或其扩增子)的实验标签(其可设计为对应具体靶序列)测序。因此,如果NPPF扩增子的3'-末端在末端的2至25个核苷酸(例如末端的2至5或2至7个核苷酸,例如末端的2、3、4、5、6、7、8、9或10个核苷酸)具有这样的序列:其代表针对所测量的各个靶的独特序列,那么这就是需要测序以鉴定靶的所有NPPF扩增子,并且通过对所测序的这些实验标签计数,就可确定样品中各靶的量。

[0199] 在一个实例中,使用链终止方法对所得到的NPP或NPPF(或其扩增子),例如由DNA组成的,进行测序。所得到的片段例如通过在平板聚丙烯酰胺凝胶或在填充了粘性聚合物的狭小玻璃管(毛细管)中电泳来进行大小分离。可选的方法为染料终止子测序(dye terminator sequencing)。在另一实例中,使用焦磷酸测序,例如由Biotage(用于低通量测序)和454Life Sciences(用于高通量测序)商业化的方法。在另一实例中,采用桥式PCR(例如, Illumina®)(例如, HiSeq)或 Ion Torrent®、454®、Helicos®、PacBio®、Solid®(Applied Biosystems)、或任意数量的其它商业测序系统来对NPP或NPPF(或其扩增子)测序。将测序衔接头(例如存在于NPP或NPPF(或其扩增子)上的poly-A或poly-T尾,例如使用PCR引入)用于捕获。

[0200] 6. 直接定量核酸酶保护测序(qNPS)

[0201] 在一些实例中,样品中其表达待测量的核酸分子采用上文描述的定量核酸酶保护测定法的替代方法在样品中检测,例如如在2016年2月11日提交的美国临时申请No. 62/294,143(通过引用并入本文)中所描述的(图9A)。该方法也使用NPPF和CFS,但是不同于对NPPF替代物进行检测或测序,该方法能够对靶核酸分子直接测序。

[0202] 该方法包括让样品与至少一个NPPF在该NPPF足以特异结合或杂交该靶核酸分子的条件接触。如果NPPF在5'-末端和3'-末端都包括测序序列,则在一些实例中每个NPPF的序列都是不同的并且彼此不互补。在一个实例中,NPPF(例如在一个或两个侧翼序列和/或互补于靶核酸分子的全部或部分的区域的序列中)包括至少一个dUTP,例如至少两个、至少三个、至少四个或至少五个dUTP。在一个实例中,NPPF中的全部“T”都由“U”替换。这些碱

基的存在使得能够在随后步骤中以尿嘧啶DNA脱糖基化酶 (UDG) 降解或破坏单链 NPPF (例如, 在NPPF从结合的靶变性后, 而在测序前, 例如在结合的靶扩增之前或扩增期间)。

[0203] 侧翼序列互补于CFS。在一些实例中, 至少一个侧翼序列包括至少一个dUTP。在一些实例中, NPPF包括两个侧翼序列, 每个具有至少一个dUTP。在一些实例中, NPPF包括两个侧翼序列, 但仅一个具有至少一个dUTP。在一些实例中, NPPF包括具有至少一个dUTP 的单个侧翼序列。在一些实例中, 该dUTP的位置靠近与靶核酸分子区域互补的序列, 例如与互补于靶核酸分子区域的序列相隔1、2、3、4、5个碱基 (例如在1、2、3、4、5个碱基内)。在一些实例中, 该dUTP的位置距离与靶核酸分子区域互补的序列至少两个碱基 (例如至少3个、至少4个或至少5个碱基)。

[0204] 该方法还包括让样品与对该侧翼序列有互补性的至少一个核酸分子 (CFS) 在足以让该 CFS特异结合或杂交该NPPF的侧翼序列的条件下接触。例如, 如果该NPPF具有5'-侧翼序列, 则让该样品与对该5'-侧翼序列有序列互补性的核酸分子 (5CFS) 和5'-末端磷酸酯在足以让该5'-侧翼序列特异结合该5CFS的条件下接触。类似地, 如果该NPPF具有3'-侧翼序列, 则让该样品与对该3'-侧翼序列有序列互补性的核酸分子 (3CFS) 在足以让该3'-侧翼序列特异结合该3CFS的条件下接触。在一个实例中, 该3CFS和5CFS的至少一个包括允许捕获、分开、回收或分离靶序列的捕获部分。本领域技术人员应理解, 可将多个CFS用于保护侧翼序列, 而不是采用单个CFS保护侧翼序列 (例如, 多个5CFS可用于保护5'-侧翼序列)。在一些实例中, 靶核酸分子为DNA, 5CFS和3CFS为DNA, 或者该5CFS为DNA且该 3CFS为RNA。在一些实例中, 靶核酸分子为RNA, 5CFS为DNA且3CFS为RNA, 或者该5CFS为RNA且该3CFS为RNA。在一些实例中, 靶核酸分子为RNA或DNA, 5CFS和/ 或3CFS为RNA-DNA杂种寡核苷酸, 例如, 其中该5CFS的5'碱基或5'的数个碱基和/或该 3CFS的3'碱基或3'的数个碱基为RNA, 而该5CFS和该3CFS的剩余部分为DNA。

[0205] 这导致产生结合至靶核酸分子 (或其部分) 以及CFS的NPPF分子, 由此产生包括NPPF 的参与与靶和CFS上的互补碱基杂交的碱基的双链分子。CFS杂交并由此防止其对应的侧翼序列被随后步骤中的核酸酶消化。在一些实例中, 每个CFS恰好为其对应的侧翼序列的长度。在一些实例中, CFS完全与其对应的侧翼序列完全互补。但是, 本领域技术人员应理解, 保护5'-末端侧翼序列的5CFS的3'-末端或保护3'-末端侧翼序列的3CFS的5'-末端可具有差异, 例如在这些位置每一处的一个核苷酸。

[0206] 在让靶核酸分子和CFS结合NPPF后, 该方法还包括让样品与特异于单链 (ss) 核酸分子或核酸分子的ss区的核酸酶 (例如S1核酸酶) 在足以除去未杂交至互补碱基的核酸碱基的条件下接触。因而, 例如, 未结合靶核酸分子或CFS的NPPF以及未结合的靶核酸分子、样品中其它ss核酸分子和未结合的CFS被降解。这产生了经消化的样品, 其包括杂交至5CFS、3CFS或二者以及靶核酸的一部分的以双链加合物存在的完整NPPF。在一些实例中, 例如如果该NPPF由DNA构成, 则核酸酶可包括外切核酸酶、内切核酸酶或其组合。

[0207] 随后, 该CFS可连接至靶核酸, 例如5CFS的5'-磷酸酯连接至靶核酸分子的3'-末端, 并且3CFS的3'-末端连接至靶核酸分子的5'-末端, 由此产生连接的靶核酸分子 (本文称为连接的靶)。

[0208] 将该双链的NPPF连接靶核酸分子分开成ss核酸分子 (例如通过加热或升高样品的pH), 由此产生ss NPPF和ss连接的靶核酸分子的混合物。任选地回收该ss连接的靶核酸分

子,例如通过将其捕获,例如通过杂交或使用捕获部分(例如,在5CFS或3CFS上)。在一些实例中,将ss NPPF和ss连接的靶核酸分子的混合物与尿嘧啶DNA脱糖基化酶(UDG)在足以降解或破坏具有至少一个dUTP的ss NPPF的条件下一同温育。

[0209] 该方法包括一个或更多步骤,其例如通过使用NPPF、5CFS或3CFS使得能够捕获或回收待测序的靶。例如,如图9B至9D中所显示,这种捕获可发生在杂交期间、在核酸酶消化后,或在杂交后,但在连接之前。图9A和9C显示了在核酸酶消化后完成捕获的实例。此外,可例如在连接后包括额外的捕获步骤。在一个实例中,将NPPF、5CFS或3CFS附着至固体基体(例如多孔板),以允许在分子杂交至锚定的NPPF、5CFS或3CFS时捕获它们。在另一实例中,捕获部分(例如附着靶、NPPF、5CFS或3CFS的颗粒或标记物)可用于将附着或杂交至含有该捕获部分的分子的分子与样品的其它组分物理分开。取决于捕获方法,这些方法可包括离心以收集固体捕获部分、磁珠捕获、结合或杂交至固体支持物、过滤等。此外,这些捕获步骤可包括洗涤所捕获的核酸分子(由此让未捕获的分子,例如未杂交的NPPF和非杂交的CFS,被分开或除去)。这些方法可包括从固体支持物(例如多孔板的孔或单反应管或容器)释放捕获的核酸分子,或者捕获的核酸分子可保持粘附至固体支持物或保存在固体支持物内。在一些实例中,该步骤可包括离心,例如以便将靶核酸分子捕获至容器或管的底部,并洗涤以除去不希望的或不需要的试剂。在一些实例中,该步骤包括靶核酸分子的磁珠捕获,并洗涤以除去不希望的或不需要的试剂。在一些实例中,该步骤包括过滤来捕获靶核酸分子,并洗涤以除去不希望的或不需要的试剂。该步骤可包括洗涤靶核酸分子,在一些实例中,其除去不再需要的试剂(例如,核酸酶)和/或允许靶核酸分子悬浮在另一种溶液中,这在具体的方法实施方案中可能有助于一个或更多后续步骤。

[0210] 在一些实例中,该方法的多个步骤(例如显示在图9A中步骤的两个、三个或四个)在相同容器或器皿中进行。例如,所公开的方法使得期望的组分能够保留在相同容器中用于多个步骤,而不希望的或不需要的组分被去除(例如,采用重复捕获和洗涤步骤)。例如,可将待分析样品在容器中裂解。将适当的NPPF和CFS添加至该容器,将核酸酶添加至该容器,并且在该容器内捕获ds核酸分子(例如,用磁珠拉至容器底部)。洗涤所捕获的ds核酸分子以除去不希望的试剂,并将需要的缓冲剂或试剂添加至该容器(例如,连接酶缓冲剂)。然后可将ss连接的靶核酸分子捕获在容器中,并进行洗涤。

[0211] 任选地例如采用PCR扩增来扩增连接的靶核酸分子。这样的方法可用于向连接的靶添加实验标签和/或序列衔接物,和/或增加该连接的靶的拷贝数。对该ss连接的靶核酸分子(或其扩增子)的至少一部分测序,由此确定样品中该至少一个靶核酸分子的序列。在一些实例中,该靶核酸分子为RNA,并该方法包括在测序前将它们逆转录为DNA。

[0212] 连接的靶可采用一个或更多扩增引物来扩增,由此产生连接的靶扩增子。至少一个扩增引物包括与连接至靶的CFS互补的区域。在一个实例中,该靶在其5'-末端连接至5CFS,在其3'-末端连接至3CFS,并且使用两个扩增引物,其中一个扩增引物具有与该5CFS的区域互补的区域,另一个扩增引物具有与该3CFS的区域互补的区域。在一些实例中,该靶分别在其5'-末端和3'-末端连接至5CFS或3CFS,并且使用一个扩增引物(例如使用快速扩增或cDNA末端),其中该扩增引物具有与该5CFS或3CFS的区域互补的区域。扩增引物的一个或两个可包括允许在扩增期间将实验标签和/或测序衔接物附着至连接的靶扩增子的序列,并且可标记一个或两个引物以允许标记该NPPF扩增子。在一些实例中,添加实验标签和

测序衔接物,例如在连接的靶扩增子的相对末端。例如,这些引物的使用可产生从连接的靶扩增子的5'-末端或3'-末端或者从5'-末端和3'-末端延伸的实验标签和/或测序衔接物,以增加可能的复用(multiplexing)程度。该实验标签可包括允许鉴定样品、受试者或靶核酸序列的独特核酸序列。该测序衔接物可包括允许将获得的连接的靶扩增子捕获在测序平台上的核酸序列。在一些实例中,在测序前从混合物除去引物。

[0213] 例如采用本文描述的方法对该连接的靶或连接的靶扩增子(或其部分)测序。任何方法都可用于对该连接的靶或连接的靶扩增子测序,并且本公开不限于具体的测序方法。在一些实例中,所用的测序方法为链终止测序、染料终止测序、焦磷酸测序、纳米孔测序或大规模平行测序(也称为下一代测序(或NGS)),这通过ThermoFisher Ion Torrent™ Personal Genome Machine (PGM™)、Illumina-branded NGS测序仪(例如,MiSeq™,HiSeq™)(或以其它方式源自Solexa™测序的)以及可从Roche Life Sciences获得的454测序来举例说明。在一些实例中,使用单分子测序。在一些实例中,该方法还包括将获得的靶序列与序列数据库比较,例如以确定是否存在靶突变。在一些实例中,该方法包括确定每个获得的靶序列的数量(例如计数)。

[0214] 图9A为显示用于所公开方法的采用NPPF 202对核酸分子测序的实施方案概况的示意图。如步骤1中所显示,让样品(例如已知或怀疑含有靶核酸200,并已经用样品破坏或裂解缓冲液(例如,裂解或以其它方式处理使得核酸可达)处理的样品)与多个具有一个或更多个侧翼序列的核酸酶保护探针(NPPF) 202(这里显示分别带有5'-和3'-侧翼序列204和206)接触或一起温育,该多个NPPF中包括至少一个NPPF特异结合第一靶核酸200(例如靶DNA或RNA)。该反应还可包括特异结合第二靶核酸序列的其它NPPF,以此类推。例如,该方法可使用一个或更多个设计的特异于每个独特靶核酸分子的不同NPPF,例如特异于CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中每一个的NPPF。在一些实例中,该多个NPPF可包括不止一个(例如2、3、4、5、10、20、50或更多个)特异于单一靶核酸分子的NPPF(其称为NPPF的平铺集(tiled set))。该反应还包括与侧翼序列互补的核酸分子(CFS) 208、210。因此,如果该NPPF具有5'-侧翼序列204,则反应将包括与该5'-侧翼序列互补的序列(5CFS) 208,并且如果该NPPF具有3'-侧翼序列206,则反应将包括与该3'-侧翼序列互补的序列(3CFS) 210。在一些实例中,该5CFS 208具有5'-末端磷酸酯,使得能够在随后步骤中连接至靶200。在一些实例中,该CFS的至少一个包括捕获部分212,其允许在期望的时间回收靶。尽管该捕获部分212显示在5CFS 208中,但是应理解,在其存在的实施方案中,其可选地处于3CFS 210、该NPPF 202或靶200上。本领域技术人员应理解,该CFS的序列将随存在的侧翼序列而变化。另外,可将不止一个CFS用来保证侧翼序列被保护(例如,可使用结合单一侧翼序列的不同区域的至少两个CFS)。该CFS可包括天然或非天然碱基。

[0215] 图2A中的捕获部分212是任选的,可位于3CFS 210或靶200上。将样品、NPPF和CFS在足以让NPPF特异结合(例如杂交)它们各自的靶核酸分子以及让CFS结合(例如杂交)NPPF侧翼序列上它们的互补序列的条件下一起温育。在一些实例中,添加CFS 208、210超过NPPF 202,例如CFS比NPPF多至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少20倍、至少40倍、至少50倍或至少100倍。在一些实例中,添加NPPF 202超过样品中的总核酸分子,例如NPPF比样品中的总核酸分子多至少50倍(摩尔过量),例如NPPF比样品中的总核酸分子多

至少75倍、至少100倍、至少200倍、至少500倍或至少1000倍。为了实验方便,可包括相似浓度的每种NPPF以形成混合物,这样对于所测量的最丰富核酸靶来说,将有至少多50倍的NPPF针对该核酸靶,例如至少过量100倍。所用的全部NPPF的实际过量和总量仅由破坏未与靶核酸靶杂交的全部NPPF的核酸酶(例如,S1核酸酶)的能力所限制。在一些实例中,该反应是加热的,例如在50℃温育过夜(例如16小时)。

[0216] 如图9A中步骤2显示的,在允许结合/杂交反应发生之后,将样品与特异于单链(ss)核酸分子的试剂(例如,核酸酶)在足以除去(或消化)诸如未结合核酸分子(例如,未结合的NPPF、未结合的CFS和未结合的靶核酸分子或这些分子保持单链的部分,例如未结合NPPF的靶核酸分子的部分)的ss核酸分子的条件接触。这导致产生ds NPPF/靶杂交复合物(或双链体)214。如图9A中显示,将样品与特异于ss核酸分子的核酸酶一起温育导致存在的任何ss核酸分子的降解,留下完整的双链核酸分子,包括已结合的NPPF以及CFS和靶核酸分子。例如,反应可用S1核酸酶在50℃温育1.5小时(尽管水解可在其它温度下发生和可进行其它时长,并且一定程度上,所需的时间和温度将是核酸酶的量、需要水解的核酸的量以及被保护的双链区域的 T_m 的函数)。

[0217] 如图9A步骤3中所显示,NPPF/靶复合物214(有时在连接前被捕获或回收,参见图9B至9D,未杂交材料被去除)暴露于连接条件,其使得靶序列200(其可为DNA或RNA)能够连接至CFS(例如连接至5CFS 208、3CFS 210或二者),由此产生连接的靶216。可使用能够共价连接靶序列和CFS的任何连接酶,例如DNA或RNA连接酶,例如T4DNA连接酶、T4RNA连接酶或Taq连接酶(因此,这些酶可存在于本文提供的试剂盒中,例如单独地或者处于连接缓冲液中)。所用的连接酶可以随存在于CFS上最接近靶序列的核苷酸(例如核糖核苷酸或脱氧核糖核苷酸)而变化。在一些实例中,连接酶稀释在缓冲液(例如含有必要辅助因子如NAD⁺或ATP、缓冲剂的缓冲液)中,将阳离子添加至含有NPPF/靶双链体的混合物并在约4℃至60℃(取决于连接酶)(例如至少4℃、至少20℃、至少25℃、至少30℃、至少37℃、至少50℃或至少60℃,例如4至37℃、4至25℃或20至25℃)温育至少15分钟、至少30分钟、至少1小时、至少2小时、至少4小时、至少6小时、至少12小时或至少16小时以将CFS连接至靶核酸分子(其中该CFS和靶核酸分子已杂交至NPPF)。在一个实例中,用于T4DNA连接酶的缓冲液包括50mM Tris-HCl、pH 7.5、10mM MgCl₂、10 mM二硫苏糖醇、1mM ATP。5CFS 208可包括5'-末端磷酸酯,其可连接至靶200的3'-末端。3CFS 210可包括3'-末端(例如,3'-末端核苷酸),其可连接至靶的5'-末端磷酸酯。如果仅一个侧翼序列存在于NPPF上,则仅一个CFS将结合在NPPF/靶复合物中,并且靶将仅连接至一个CFS。该CFS可为DNA或RNA(或两核苷酸类型的混合物)。如果该CFS或靶为DNA,则用于连接至DNA的5'-末端磷酸酯供体不为核糖核苷酸。在一个实例中,如果靶为RNA,则5CFS 208为DNA或RNA并且3CFS 210为RNA。在一个实例中,如果靶为DNA,则5CFS 208为DNA(或者至少将连接至靶的3'-末端的5CFS 208的5'-末端核苷酸为dNTP)或RNA并且3CFS 210为RNA。

[0218] 如图9A的步骤4所显示的,在CFS连接至靶后,NPPF 202与连接的靶216分开。即,NPPF/靶复合物214(例如,双链核酸分子)分开为两单链核酸分子NPPF 202和连接的靶216。该步骤还除去未连接的核酸分子(例如,未连接的CFS)。例如,可在允许NPPF 202从连接的靶216解离的条件下对该反应加热或调整pH(例如,导致该反应具有碱性pH),产生单链NPPF 202和单链连接的靶216的混合群体。在一些实例中,采用CFS上的捕获部分212将该连

接的靶与该混合物分离。本文描述的捕获方法可用于捕获连接的靶216。例如,如果捕获部分212为微珠,可采用离心回收连接的靶216。如果捕获部分212为金属微珠,可采用磁体回收连接的靶216。如果捕获部分212包括羧基基团,可采用适当的标记胺的固体基质捕获连接的靶216。在捕获连接的靶216后,可除去(例如,通过洗涤)反应的剩余物(例如, NPPF 202)。在一些实例中,采用过滤捕获该连接的靶216。

[0219] 如图9A的步骤5所显示的,接着可对分离的连接的靶216(或其扩增子)测序。在一些实例中,多个连接的靶平行测序,例如同时或同期测序。该方法因此可用于对多个连接的靶序列测序。

[0220] 任选地,连接的靶216在测序前进行扩增(例如,采用PCR)、洗涤或进行扩增和洗涤。在一些实例中,例如如果至少一个侧翼序列或捕获序列包括dUTP,则用尿嘧啶DNA脱糖基化酶(UDG)降解NPPF 202。这种降解可在扩增前或扩增期间发生。在一些实例中,例如如果靶为RNA,则可将逆转录步骤作为扩增的一部分包括进来。因此,随后可对得到的连接的靶扩增子226测序。PCR引物或探针可包括一个或更多个实验标签222、224和/或测序衔接物218、220(例如,允许连接的靶被特定的测序平台测序,并因此这些衔接物互补于测序芯片上的捕获序列)。PCR引物/探针的至少一部分特异于5CFS 208和/或3CFS 210。在一些实例中,引物的浓度超过连接的靶216,例如过量至少10,000倍、至少50,000倍、至少100,000倍、至少150,000倍、至少200,000倍、至少400,000倍、至少500,000倍、至少600,000倍、至少800,000倍或至少1,000,000倍。在一些实例中,反应中引物208的浓度为至少200nM(例如至少400nM、至少500nM、至少600nM、至少750nM或至少1000nM)。

[0221] 如图9B至9D中显示的,在连接前的时点(例如,图9A的步骤3),将靶(图9A的200)和附着其上/与其杂交的其它分子回收或捕获(步骤320),由此使得能够除去样品中的剩余试剂(例如,未杂交材料)并添加新试剂。此外,对靶的捕获使得能够随后对与该靶杂交的或以其它方式附着至该靶的核酸分子进行洗涤(例如,失活或除去残留的酶)。

[0222] 在一个实例中,靶在杂交时回收或捕获(图9B)。如图9B中显示的,在任选的变性300后,让含有靶的样品与NPPF和CFS在允许杂交310的条件下一起温育,由此产生杂交的复合物。在该实例中,将靶、连同将与其杂交的NPPF以及可杂交该NPPF的CFS在固体支持物存在下杂交,而该固体支持物含有能够捕获产生的杂交复合物320的材料。例如,存在于3CSF、5CSF和/或NPPF上的捕获部分(例如,图9A中的212)可用作该固体支持物。例如,通过该捕获部分为微珠或板(plate)(例如,采用羧基-胺键合,例如CFS上的氨基基团和微珠上的羧基),则该固体支持物为该微珠或板。这种情况下,该捕获部分/固体支持物使得杂交复合物结合至该固体支持物(例如,因为该NPPF将杂交至附着至固体支持物的CFS,并且靶和其它CFS将杂交至该NPPF)。该方法包括杂交后的洗涤步骤。一个或更多个随后的步骤,例如核酸酶消化330、洗涤340、连接350、UDG消化370(任选地,仅如果该NPPF包括dUTP)和/或PCR扩增380,可在固体支持物存在下进行,其使得试剂能够去除或添加至被捕获的靶。

[0223] 在一个实例中,在核酸酶处理后(例如,在图9A中步骤2后)但连接前(例如图9A中步骤3之前)回收或捕获靶(图9C)。如图9C中显示的,在任选的变性400、杂交410和核酸酶消化430(其产生NPPF/靶复合物(图9A中的214))后,在步骤420将该NPPF/靶复合物从样品混合物回收。在一些实例中,通过使用CFS上的捕获部分回收该NPPF/靶复合物,无需将该NPPF/靶双链体分开成ss核酸分子。例如,存在于3CSF、5CSF和/或NPPF上的捕获部分(例如,

图9A中的212)可用作该固体支持物。例如,通过该捕获部分为微珠或板(plate)(例如,采用羧基-胺键合,例如CFS上的氨基基团和微珠上的羧基),则该固体支持物为该微珠或板。这种情况下,捕获部分/固体支持物为NPPF/靶复合体的一部分,其允许对其捕获或回收。该方法可包括回收后的洗涤步骤440。一个或更多个随后的步骤,例如连接450、洗涤460、UDG消化470(任选地,仅如果该NPPF包括dUTP)和/或PCR扩增480,可在固体支持物存在下进行,其使得试剂能够去除或添加至被捕获的靶。

[0224] 在一个实例中,在杂交后(例如,在图9A中步骤1后)但核酸酶消化前(例如如图9A中步骤2之前)回收或捕获靶(图9D)。如图9D中显示的,在任选的变性500和杂交510(其产生杂交复合体)后,在步骤520将该得到的杂交复合体从样品混合物回收。在一些实例中,该杂交复合体通过使用CFS上的捕获部分回收。例如,存在于3CSF、5CSF和/或NPPF上的捕获部分(例如,图9A中的212)可用作该固体支持物。例如,通过该捕获部分为微珠或板(plate)(例如,采用羧基-胺键合,例如CFS上的氨基基团和微珠上的羧基),则该固体支持物为该微珠或板。这种情况下,捕获部分/固体支持物为杂交复合体的一部分,其允许对其捕获或回收。该方法可包括回收后的洗涤步骤。一个或更多个随后的步骤,例如核酸酶消化530、连接550、洗涤540、560、UDG消化570(任选地,仅如果该NPPF包括dUTP)和/或PCR扩增580,可在固体支持物存在下进行,其使得试剂能够去除或添加至被捕获的靶。

[0225] B. 用于检测基因表达的蛋白

[0226] 在所公开方法的一些实施方案中,确定样品中基因表达水平包括检测样品中一种或更多种蛋白(例如通过确定这些蛋白的相对或实际量)。检测蛋白的常规技术在本领域内是已知的,本公开并不限于蛋白检测的具体方法。

[0227] 可借由被用在免疫测定中的特异于靶蛋白(例如表1中的那些)的蛋白特异性结合剂(例如抗体或适体)识别的新表位来检测蛋白基因产物(例如,表1中的那些)和确定样品中的蛋白表达水平,所述免疫测定例如为ELISA测定、免疫印迹测定、流式细胞术测定、免疫组织化学测定、酶联免疫测定、放射免疫测定、Western印迹测定、免疫荧光测定、化学发光测定以及其它肽检测策略。这些方法一般采用单克隆抗体或多克隆抗体。

[0228] 因此,在一些实施方案中,采用靶蛋白特异性结合剂,例如可被可检测地标记的抗体或其片段,检测存在于生物学样品中的靶蛋白表达(例如表1中的那些)的水平以及表达的蛋白的量。在一些实施方案中,该特异性结合剂为抗体,例如多克隆或单克隆抗体,其特异结合靶蛋白(例如表1中的那些)。因此某些实施方案中,确定生物学样品中蛋白的水平或量包括让来自受试者的样品与蛋白特异结合剂(例如特异结合表1中显示的蛋白的抗体)接触,检测该结合剂是否被样品结合,并由此测量样品中存在的蛋白的量。在一个实施方案中,该特异性结合剂为单克隆或多克隆抗体,其特异结合靶蛋白(例如表1中的那些)。本领域技术人员应理解,针对靶蛋白(例如表1中的那些)的抗体有商业来源。

[0229] 靶蛋白(例如表1中的那些)的存在可用多种特异性结合剂检测,例如一种、两种、三种或更多种特异性结合剂。因此,这些方法可利用不止一种抗体。在一些实施方案中,抗体中的一种附着至固体支持物,例如多孔板(例如,微量滴定板)、微珠、膜等。实践中,微量滴定板可方便地用作该固体支持物。但是,抗体反应也可在液相中进行。

[0230] 在一些实例中,该方法可包括让样品与特异结合第一抗体的第二抗体接触,而该第一抗体特异结合靶蛋白(例如表1中的那些)。在一些实例中,该第二抗体例如被荧光团

(例如 FITC、PE、荧光蛋白等)、酶(例如HRP)、放射标记物或纳米颗粒(例如金颗粒或半导体纳米晶体,例如量子点(QDOT®))可检测地标记。在该方法中,结合至抗体的酶将与适当的底物例如显色底物以如此方式反应,使得能够产生可例如通过分光光度计、荧光计或通过视觉手段检测的化学部分。可用于可检测地标记抗体的酶包括,但不限于,苹果酸脱氢酶、葡萄球菌核酸酶、 δ -5-甾体异构酶、酵母醇脱氢酶、 α -甘油磷酸、脱氢酶、磷酸丙糖异构酶、辣根过氧化物酶、碱性磷酸酶、天冬酰胺酶、葡萄糖氧化酶、 β -半乳糖苷酶、核糖核酸酶、脲酶、过氧化氢酶、葡萄糖-6-磷酸脱氢酶、葡糖淀粉酶和乙酰胆碱酯酶。检测可通过使用该酶的显色底物的比色法来完成。

[0231] 检测也可通过与类似制备的标准品相比,视觉比较底物的酶促反应程度来完成。还可以用荧光化合物标记抗体,示例性荧光标记化合物包括异硫氰酸荧光素、罗丹明、藻红蛋白、藻蓝蛋白、别藻蓝蛋白、邻苯二甲醛、Cy3、Cy5、Cy7、四甲基若丹明异硫氰酸酯、藻红蛋白、别藻蓝蛋白、德克萨斯红和荧光胺。还可使用荧光发射金属例如 ^{152}Eu 或其它镧系元素对抗体进行可检测标记。可与抗体缀合的其它金属化合物包括,但不限于铁蛋白、胶体金,例如胶体超顺磁珠。这些金属可使用诸如二亚乙基三胺五乙酸(DTPA)或乙二胺四乙酸(EDTA)的金属螯合基团连接到抗体上。该抗体还可通过将其偶联至化学发光化合物来可检测地标记。化学发光标记化合物的实例为鲁米诺(luminol)、异鲁米诺(isoluminol)、theromatic吡啶酯(theromatic acridinium ester)、咪唑、吡啶盐和草酸酯。类似地,生物发光化合物可用于标记抗体。在一个实例中,用诸如荧光素、荧光素酶或水母素的生物发光化合物标记抗体。可缀合至抗体的半抗原包括,但不限于,生物素、地高辛、噁唑酮以及硝基酚。可缀合或掺入抗体的放射性化合物包括但不限于 $^{99\text{m}}\text{Tc}$ 、 ^{125}I 以及包括任何放射性同位素(radionucleotides)的氨基酸,包括但不限于 ^{14}C 、 ^3H 和 ^{35}S 。

[0232] 一般地,蛋白(例如表1中的那些)的免疫测定通常包括将生物学样品在抗体存在下温育,并通过本领域公知的数种技术的任一种检测结合的抗体。在一个实例中,可让该生物学样品(例如含有黑素细胞的样品)与诸如硝酸纤维素或多孔板的固相支持物或载体或者其它能够固定细胞、细胞颗粒或可溶性蛋白的其它固体支持物接触并固定于其上。该支持物随后可以用适合的缓冲液洗涤,接着用特异结合该靶蛋白(例如表1中的那些)的抗体处理。然后可用缓冲液再次洗涤该固相支持物以除去未结合的抗体。如果该抗体为直接标记的,则可通过常规手段检测固体支持物上结合的标记物的量。如果该抗体为未标记的,则可使用检测特异结合靶蛋白(例如表1中的那些)的抗体的经标记的第二抗体。

[0233] 作为选择,抗体可固定在固体支持物上,接着与分离自生物学样品(例如组织活检)的蛋白在使得该抗体和蛋白能够彼此特异结合的条件下接触。然后可检测所得的抗体:蛋白复合物,例如通过添加特异于该蛋白的另一抗体(因此形成抗体:蛋白:蛋白三明治)。如果添加的第二抗体为被标记的,则可检测该复合物,或者,可使用特异于该添加的第二抗体的经标记的二级抗体(secondary antigen)。

[0234] 固相支持物或载体包括能够结合样品、抗体或抗体的材料。示例性的载体包括玻璃、聚苯乙烯、聚丙烯、聚乙烯、葡聚糖、尼龙、淀粉酶、天然和改性纤维素、聚丙烯酰胺、辉长岩和磁铁矿。载体的性质可在一定程度上可溶或不可溶。载体材料可以实际上具有任何可能的结构构型,只要偶联的分子能够结合其靶(例如抗体或蛋白)。因此,支持物构型可以为球形的,例如在微珠中,或者圆柱形的,例如在试管的内表面或杆的外表面。作为选择,该表

面可以为平坦的,例如片材或测试条。

[0235] 在一个实施方案中,酶联免疫吸附测定(ELISA)可用于检测靶蛋白。ELISA可用于检测样品中蛋白的存在,例如通过使用特异结合靶蛋白例如表1中的那些)的抗体。在一些实例中,抗体可例如直接缀合或通过二级抗体连接至酶,并添加该酶可转化成可检测信号的底物。因此,对于荧光ELISA的情况,当适当波长的光照射到样品上时,任何抗原:抗体复合物都将发荧光,使得样品中的抗原量可从该荧光强度推断出。蛋白(例如从含有黑素细胞的样品提取或分离的蛋白)通常非特异性地(例如通过吸附至表面)或特异性地(例如在“三明治”ELISA中通过特异于该相同抗原的另一抗体捕获)固定在固体支持物(例如聚苯乙烯微量滴定板)上。在每个步骤之间通常用温和去垢剂溶液(例如含或不含NP40或TWEEN的磷酸缓冲盐水)洗涤板,以除去任何未特异结合的蛋白或抗体。在最后洗涤步骤之后,通过添加酶学底物产生可见信号来让板显色,这指示样品中蛋白的量。

[0236] 还可使用各种其它免疫测定的任一种完成检测。例如,通过放射性标记抗体或抗体片段,有可能通过使用放射免疫测定(RIA)检测指纹基因野生型或突变肽。在另一实例中,可以使用灵敏和特异的串联免疫放射测定(tandem immunoradiometric assay)(参见Shen和Tai, J. Biol. Chem., 261:25, 11585-11591, 1986)。放射性同位素可通过例如使用伽马计数器或闪烁计数器的手段或通过放射自显影来检测。

[0237] 在一个实例中,将光谱测定法用于检测或定量靶蛋白(例如表1中的那些)的表达水平。示例性光谱测定法包括质谱法、核磁共振光谱法以及它们的组合。在一个实例中,将质谱法用于检测生物学样品中存在的靶蛋白(例如表1中的那些)。

[0238] 也可通过与免疫亲和测定偶联的质谱测定来检测靶蛋白(例如表1中的那些),使用基质辅助的激光解吸/电离飞行时间(MALDI-TOF)质量映射和液相色谱/四极杆飞行时间电雾化电离串联质谱(LC/Q-TOF-ESI-MS/MS)通过二维聚丙烯酰胺凝胶电泳(2D-PAGE)分离的蛋白质的序列标签。

[0239] 也可将定量质谱法例如SELDI用于分析样品中的蛋白表达。在一个实例中,将表面增强的激光解吸电离飞行时间(SELDI-TOF)质谱用于检测蛋白表达,例如使用ProteinChip(Ciphergen Biosystems, Palo Alto, CA)。这些方法在本领域内是公知的。SELDI为用于解吸的固相方法,其中分析物呈现在增强分析物捕获或解吸的表面上的能量流中。

[0240] 简言之,一种版本的SELDI使用带有化学物的色谱表面,该化学物选择性捕获目的分析物(例如表2至8任一个中的那些)。色谱表面可由疏水性、亲水性、离子交换、固定化金属或其它化学物构成。例如,该表面化学物可包括基于氧依赖性、碳依赖性、硫依赖性和/或氮依赖性的共价或非共价分析物固定手段的结合功能性。该活化的表面用于共价固定特异的“诱饵”分子,例如常常用于生物分子相互作用研究例如蛋白-蛋白和蛋白-DNA相互作用的抗体、受体、或寡核苷酸。

[0241] 该表面化学物使得结合的分析物能够保留,而未结合的材料被洗去。随后,结合至表面的分析物(例如表1中的那些)可被解吸并通过几种手段的任一种,例如通过采用质谱法进行分析。当分析物在解吸过程中离子化,例如在激光解吸/离子化质谱仪中,检测器可为离子检测器。质谱仪通常包括用于确定解吸离子的飞行时间的装置。该信息被转化为质量。然而,不需要确定解吸离子的质量来解析和检测它们:电离的分析物在不同时间撞击检测器的事实提供了对它们的检测和解析。作为选择,分析物可被可检测地标记(例如用荧光

团或放射性同位素)。这些情况下,检测器可为荧光或放射性检测器。可将多个检测装置串联实施,以充分质询分析物组分和与阵列中每个位置处保留的分子相关的功能。

[0242] 因此,在特定实例中,该色谱表面包括特异结合靶蛋白(例如表1中的那些)的抗体。在其它实例中,该色谱表面主要由,或者由特异结合靶蛋白(例如表1中的那些)的抗体组成。在一些实例中,该色谱表面包括结合其它分子例如标准化蛋白(例如表1中的那些)的抗体。

[0243] 在另一个实例中,采用细菌Fc结合支持物将抗体固定在该表面上。该色谱表面与样品(例如痣样品)一起温育。存在于样品中的抗原可识别该色谱表面上的抗体。洗去未结合的蛋白质和质谱干扰化合物,并通过SELDI-TOF分析和检测保留在该色谱表面上的蛋白。然后采用差异蛋白表达作图(differential protein expression mapping)比较来自样品的MS谱,由此通过各种统计学技术和生物信息学软件系统比较特定分子量蛋白的相对表达水平。

[0244] 作为选择,靶蛋白的量可使用荧光法确定。例如,量子点(例如, Qdots® 标记物)可用在越来越多的应用清单中,包括免疫组织化学、流式细胞术以及板基测定(plate-based assays),因此可以与本公开结合使用。量子点纳米晶体具有独特的光学性质,包括用于灵敏度和定量的极亮信号;以及用于成像和分析的高度光稳定性。需要单个激发源,不断扩大范围的缀合物(例如,抗体缀合物)使得它们可用于广泛的应用中。来自量子点的发射是狭窄并对称的,这意味着与其它颜色的重叠最小化,导致对临近检测通道的渗入最小化和衰减的串扰,尽管可同时使用更多的颜色。例如,可用量子点缀合的二级抗体或链霉亲和素缀合的量子点联合生物素标记的初级或二级抗体来进行IHC。

[0245] C. 任选的测定对照测量

[0246] 任选地,用于检测基因表达产物(例如,核酸(例如mRNA、lncRNA)或蛋白)的测定包括用于评估测定性能的阳性和阴性过程对照元件。

[0247] 阳性对照可以是任何已知的元件,优选与靶(例如, RNA靶,然后RNA(或cDNA)阳性对照)性质类似的元件,其可包括在测定(或样品)中,与靶平行检测,并且不干扰(例如,交叉反应)这种靶检测。在一个实例中,该阳性对照为体外转录物(IVT),作为分开的样品平行运行,或以已知量“掺(spiked)”在每个样品中的。IVT特异性结合剂(例如,寡核苷酸探针,例如NPP或NPPF)和(如果适用的话)IVT特异性检测剂也包括在每个测定中,以确保这种体外转录物的阳性结果。

[0248] 在一些情况下,异常信号可能由未例如通过对标准化物(normalizer)进行分析来加以另外控制的预料之外的过程相关事件所导致;因此,在一些实施方案中,有用的是包括独立于样品的过程对照元件,以指示在任何样本上成功或失败的测定,而不考虑样本稳定性、完整性或输入水平。检测核酸基因表达产物的方法实施方案可以在每个测定中包括已知浓度的RNA样品(例如,体外转录RNA或IVT)。这样的对照元件(例如,IVT)可在每个测定中测量,并用作测定过程质量对照。

[0249] 一些公开的涉及RNA基因表达产物的方法实施方案可以,但不必,包括含有通用人参考RNA(Universal Human Reference RNA)的平行处理样品。如果这种通用RNA样品包括被通过适当测定进行的检测所靶向的RNA的全部或部分,则可期望针对这些包括的RNA的阳性信号,其可以充当一种(或另一种)测定过程质量对照。

[0250] 阴性过程对照元件可包括分析物特异性结合剂(例如,寡核苷酸或抗体),其被设计或选择来检测预期不在测试样品中表达的基因产物。例如,不识别人转录组或蛋白组中任何基因表达产物的分析物特异性结合剂可以包括在多重测定中(例如分别特异于植物或昆虫或线虫 RNA或蛋白的寡核苷酸探针或抗体,而人基因表达产物为期望的靶)。这种阴性对照元件不应在适用的测定中产生信号。这种阴性过程对照元件的高于背景信号的任何信号都是测定失败的指示。在一个实例中,该阴性对照为ANT。

[0251] IV. 基因表达数据

[0252] 基因表达数据“包含解决与疾病的预防和治疗、生物进化机制或药物发现相关的基本问题的关键因素”(Lu和Han, Information Systems, 28:243-268 (2003))。在一些实例中,从这样的数据中提取信息就像从检测到的一种或更多种基因产物的存在、不存在或定性量(例如,高、中、低)中做定性决定一样简单。在其它实例中,原始基因表达数据可以被预处理(例如,减去背景、log转换和/或校正)、标准化和/或应用在分类算法中。但是,这种数据预处理是任选的。

[0253] 很多生物学变量(例如,基因表达数据)不符合参数统计检验的假设,例如,这样的变量不是正态分布,方差不齐,或二者(Durbin等, Bioinformatics, 18:S105 (2002))。一些情况下,变换数据将使得其更好地吻合统计假设。在一些方法实施方案中,有用的数据变换可包括(i) 对数变换,其由对每个观察结果取对数组成,例如以10为底的对数、以2为底的对数、以e为底的对数(也称为自然对数);该对数选择不造成差异,因为这样的对数区别在于常数因子;或者方差稳定变换,例如如Durbin所描述的(上文)。在具体实例中,在这种方法中检测的每个生物标志物(例如,至少两个表1中的生物标志物)的原始表达值经对数变换(例如,log₂或log₁₀)。当计算计数/读取数(read)表达时,变换可包括每个样品的总读取数,该值被缩放至每百万读取数:基因水平读取数相对于样品总读取数按比例求值并进行log₂变换,得到以2为底的每百万个计数的对数, $\log_2\left(\frac{r_{gi}+0.5}{R_i+1} \times 1,000,000\right)$, 其中 r_{gi} 为针对每个探针(g)和样品(i)的序列读取数的数量,(缩放以避免零计数),并针对每个样品的映射读取数数量(文库计数)进行调整(R_i ,以常数1放大,确保按比例的读取数与文库大小的比值大于零)。

[0254] V. 方法实施

[0255] 本文描述的方法,例如那些涉及分类器的方法,可以以多种方式实施。以下描述几种代表性的非限定性实施方案。在一些方法实施方案中,将基因表达数据输入(例如,手工或自动地)计算机或其它设备、机器或装置以应用本文描述的各种算法,这在收集和处理大量的基因表达数据点时是尤其有利的。因此,在一些实例中,采用所公开的方法获得的基因表达数据或概率值保存在计算机存储器中。其它的实施方案涉及使用通信基础设施,例如互联网。各种形式的硬件、软件、固件、处理器,或它们的组合可用于实施具体的分类器和方法实施方案。软件可作为具体体现在程序存储设备上的应用程序来实现,或者软件的不同部分在用户的计算环境中(例如,作为小应用程序)和在审阅者的计算环境中实现,其中审阅者可能位于相关联的远程站点(例如,在服务提供商的设施处)。

[0256] 例如,在用户进行数据输入期间或之后,在用户侧计算环境中可进行部分数据处理。例如,可对用户侧计算环境进行编程以提供定义的检验码(defined test codes)来指

示可能性“评分”，其中该评分以检验码的形式作为对审阅者的计算环境的经处理或经部分处理的响应进行传送，用于随后执行一种或更多种算法以在审阅者的计算环境中提供结果和/或生成报告。该评分可为数字评分（代表数值）或代表数值或数值范围的非数字评分（例如，代表结果90- 95%可能性的“A”）。

[0257] 用于执行本文描述的算法的应用程序可以上传至包括任何适当架构的机器并由其执行。通常，该机器包括计算机平台，该计算机平台具有诸如一个或多个中央处理单元（CPU）、随机存储器（RAM）和输入/输出（I/O）界面的硬件。该计算机平台还包括操作系统和微指令代码。本文描述的各种处理和功能可以通过操作系统执行的微指令代码的一部分或应用程序的一部分（或其组合）。此外，可以将各种其它的外围设备（例如额外的数据存储设备和打印设备）连接至该计算机平台。

[0258] 作为计算机系统，该系统通常包括处理器单元。该处理器单元运行以接收信息，其可包括试验数据（例如，应答基因的水平、参考基因产物的水平；应答基因的标准化水平；并且还可以包括其它数据，例如患者数据。这种接收的信息可至少暂时地保存在数据库中，并且如上文所描述对数据进行分析以生成报告。

[0259] 还可将输入和输出数据的部分或全部以电子方式发送；可以以电子方式或通过电话（例如，通过传真，使用设备如传真回复）传送某些输出数据（例如，报告）。示范性输出接收设备可包括显示元件、打印机、传真设备等。电子形式的传输和/或显示可包括电子邮件、交互式电视等。在一个实施方案中，将输入数据的全部或部分和/或输出数据的全部或部分（例如，通常至少最终报告）保持在web服务器上用于使用一般浏览器进行访问，优选加密访问。可以对该数据进行访问或根据需要发送至健康专业人员。该输入和输出数据，包括最终报告的全部或部分，可用来填充可以存在于卫生机构的保密数据库中的患者医疗记录。在一些实例中，该方法包括生成报告。在一些实例中，该报告包括表明样品分类的图标，例如“ABC”用于针对ABC DLBCL或“GCB”用于针对GCB DLBCL。

[0260] 用于本文描述的方法的系统通常包括至少一个计算机处理器（例如，在该方法以其整体在单一站点进行时）或至少两个联网的计算机处理器（例如，在数据准备由使用者（在本文也称为“客户”）输入并传输至远程站点第二计算机处理器进行分析时，其中该第一和第二计算机处理器通过网络（例如，经由内联网或互联网）连接）。该系统还可包括用于输入的用户组件；以及用于检查数据、生成报告和人工干预的检查者组件（reviewer component）。该系统的其它组件可包括服务器组件；和用于存储数据的数据库（例如，如在报告元件（例如，解释性报告元件）数据库中，或在可包括由用户输入数据和数据输出的关系数据库（RDB）中）。该计算机处理器可为一般存在于个人台式计算机（例如，IBM、Dell、Macintosh）、便携式计算机、大型计算机、微型计算机、平板电脑或其它计算设备中的处理器。

[0261] 联网的客户端/服务器架构可按照需要选择，并且可例如为经典的两层或三层客户端服务器模型。关系数据库管理系统（RDMS），作为应用服务器组件的一部分或作为单独的组件（RDB机器），提供到该数据库的接口。

[0262] 在一个实例中，将该架构提供为以数据库为中心的客户端/服务器架构，其中该客户端应用一般请求来自应用服务器的服务，该应用服务器对数据库（或数据库服务器）发出请求来按照需要以各种报告元件填充报告，特别是解释性报告元件、尤其是解释文本和警

告。该服务器(例如,作为一个应用服务器机器的一部分或单独的RDB/关系数据库机器)响应客户端请求。

[0263] 该输入客户端组件可为完整的、独立个人计算机,提供了全范围的能力和特征来运行应用程序。该客户端组件通常在任何希望的操作系统下操作,并包括通信元件(例如,调制解调器或其它用于连接到网络的硬件)、一个或更多个输入设备(例如,用于传输信息和命令的键盘、鼠标、小键盘、或其他设备)、存储元件(例如,硬盘驱动器或其它计算机可读、计算机可写存储介质)以及显示元件(例如,将信息传送给用户的监视器、电视、LCD、LED或其它显示设备)。用户通过输入设备将输入命令输入到计算机处理器。通常,用户界面是写入的用于web浏览器应用程序的图形用户界面(GUI)。

[0264] 该服务器组件可为个人计算机、小型计算机或大型计算机并提供数据管理、客户间的信息共享、网络管理以及安全性。所使用的应用程序和任何数据库可在相同或不同的服务器上。

[0265] 可考虑其它用于客户端和服务器的计算机布局,包括在单个机器如主机、一组机器或其它适合的装置上进行处理。通常,客户端和服务器机器一起运行来完成对本公开的处理。

[0266] 当使用时,数据库一般连接到数据库服务器组件,并且可为任何容纳数据的设备。例如,该数据库可为任何用于计算机的磁存储设备或者光学存储设备(例如,CD-ROM、内部硬盘驱动器、磁带驱动器)。该数据库可位于服务器组件远端(通过网络、调制解调器等访问)或位于服务器组件本地。

[0267] 当用于本系统和方法中时,所述数据库可以是根据数据项之间的关系对其进行组织和访问的关系数据库。该关系数据库通常包括多个表(实体)。该表格的行代表记录项(关于独立项的信息集合),列代表字段(记录项的具体属性)。在其最简单的概念中,所述关系数据库是通过至少一个共同字段彼此“联系”的数据条目集合。

[0268] 可以将配备有计算机和打印机的其它工作站用在服务点以输入数据,并在一些实施方案中,生成适当的报告(如果需要的话)。该计算机具有启动该应用程序的快捷方式(例如,在桌面上),以便于根据需要起始数据输入、传输、分析、报告接收等。

[0269] 1. 计算机可读存储介质

[0270] 本发明还关注计算机可读存储介质(例如,CD-ROM、存储键、闪存卡、磁盘等),其上存储有程序,当在计算环境中执行时,该程序提供对算法的执行,以完成如本文所描述的全部或部分应答可能性评估结果。当该计算机可读介质包含用于进行本文所描述的方法的完整程序时,该程序包括用于收集、分析和生成输出结果的程序指令,并且通常包括计算机可读代码装置,用于如本文所描述与用户互动、与分析信息结合来处理数据,以及为该用户生成唯一的印刷或电子介质。

[0271] 当该存储介质提供用于实施本文描述方法的一部分的程序(例如,该方法的用户侧方面(例如,数据输入、报告接收能力等))时,该程序提供将由用户输入的数据(例如,通过因特网、通过内联网等)传输到远程站点处的计算环境。

[0272] 可在远程位点处进行该数据处理或完成该数据处理以生成报告。在检查报告以及完成任何所需的人工干预以提供完整的报告后,接着可将完整报告作为电子文档或打印文档(例如,传真或邮寄纸质报告)传送回用户。含有本文描述的程序的该存储介质可与记录

在适当基质上的说明书(例如,用于程序安装、使用等)或可以获得这些说明书的网址一起包装。该计算机可读存储介质还可与用于进行应答可能性评估的一种或多种试剂(例如,引物、探针、阵列或其它这样的试剂盒组分)联合提供。

[0273] 2. 输出

[0274] 在一些实施方案中,一旦确定了特定样品(患者)的评分,可将该评分的指示显示和/或传送给临床医生或其它护理人员。例如,可将该测试的结果以提供关于该测试结果的信息的可感知输出的形式提供给用户(例如临床医生或其它医护人员、实验室人员、或患者)。在一些实例中,该输出是纸面输出(例如,书写或打印输出)、屏幕上的显示、图形输出(例如,图、表或其它图示)或音频输出。因此,该输出可包括生成的报告。

[0275] 例如,该输出可为文本(任选地,带有相应的)评分。例如,文本输出可以为“ABC”等,或“GCB”等,或“未定”等。这样的文本输出可例如用于提供对DLBCL亚型GCB或 DLBCL亚型ABC的诊断结果,或者可简单地用于辅助临床医生区分亚型ABC和GCB。

[0276] 在其它实例中,该输出为数值(例如,数量输出),例如预测的概率值。在另外的实例中,该输出为图形表现形式,例如指示预测的概率的图。在具体实例中,该输出(例如图形输出)显示或提供表征所测试样品为ABC或GCB的截止值或水平。在其它实例中,该输出为图标,例如,如果该样品分类为ABC则为“ABC”;如果该样品分类为GCB,则为“GCB”;如果该样品分类为未分类的(例如,与ABC或GCB都不一致),则为“U”或“?”。在一些实例中,该输出例如通过提供经由物理、听觉或电子手段的输出而传送至用户(例如,通过邮件、电话、传真发送、电子邮件或传送至电子医疗记录)。

[0277] 在一些实例中,该输出伴有用于解释数据的指导,例如,表明该DLBCL为ABC或 GCB的数值或其它界限。该指导不必明确指出DLBCL为ABC还是GCB,尽管其可以包括这样的诊断结果。输出中的标识可例如包括正常或异常范围或截止值,然后该输出的接收者可用其对结果进行解释,例如,以达成诊断或治疗计划。在其它实例中,该输出可提供建议的治疗方案。在一些实例中,该测试可以包括对其它临床信息(例如确定样品中一种或更多种额外DLBCL生物标记物的量)的确定。

[0278] VI. 基因集合和分类器输出的临床用途

[0279] 所公开的基因集合或分类器可以导致DLBCL样品被表征(例如,诊断)为GCB亚型或非GCB亚型(例如ABC),或未确定。这些(以及其它可能的)结果的每个对于经培训临床专业人员是有用的。一些代表性临床用途在下文有更详细描述。诊断或预后可以通过任何适合的手段提供给受试者(例如,患者或医生),包括但不限于通过物理的、听觉的或电子手段(例如,通过邮件、电话、传真发送、电子邮件、或传送至电子医疗记录)。

[0280] A. 诊断指示

[0281] 诊断结果告知受试者(例如,患者)他/她患有或可能患有何种疾病或状况。如遍及本公开更具体描述的,任何公开的对DLBCL亚型进行分类的方法的任何结果都可作为诊断结果例如提供给受试者或健康专业人员。因此,诊断结果包括确定DLBCL为GCB亚型或非GCB亚型(例如ABC),或者未确定。

[0282] B. 预后指示

[0283] 预后为其样品接收特定测试结果(例如,DLBCL的DCB亚型或非GCB亚型(例如ABC))的受试者可能的健康后果。不良预后指该受试者的长期展望不佳,例如,1、2、3或5年存活率

为50%或更低(例如,40%、30%、25%、20%、15%、10%、5%、2%或1%或更低)。另一方面,预后良好指该受试者的长期展望是适中到良好,例如,1、2、3或5年存活率大于30%、40%、50%、60%、70%、75%、80%或90%。

[0284] 非GCB亚型DLBCL,例如ABC亚型DLBCL,已显示出具有比GCB亚型DLBCL更不良的预后。相应地,通过所公开方法的任一种对非GCB亚型DLBCL、例如ABC亚型 DLBCL的发现结果可用于预测受试者(从其获取测试样品)相对不良的预后。相反,通过所公开方法的任一种对GCB亚型DLBCL的发现结果可用于预测该相应受试者相对良好的预后。

[0285] C. 治疗(预测)指示

[0286] 所公开的方法还可包括选择(或不选择)受试者来处理GCB亚型DLBCL或非GCB亚型DLBCL,例如ABC亚型DLBCL,如果他们对应的样品被如此亚型分型。示例性处理方法提供在下文段落中。因此,在一个实例中,如果样品被确定为GCB亚型DLBCL,则从其获得该样品的受试者用(1)环磷酰胺、多柔比星、长春新碱以及泼尼松或泼尼松龙(CHOP)化疗剂、(2)利妥昔单抗加CHOP(R-CHOP)化疗剂、或(3)依托泊苷加R-CHOP(R-EPOCH)处理。在另一实例中,如果该样品被确定为ABC亚型DLBCL,则从其获得该样品的受试者较不可能应答于CHOP或R-CHOP治疗,并且该受试者可以选择为用一种或更多种备选疗法处理(例如本文公开的那些)。

[0287] 在一些实施方案中,取决于患者的诊断结果,所公开的方法还包括如下的一种或更多种: a) 如果受试者的确定的诊断结果为GCB亚型DLBCL,则对该受试者指定(prescribe)处理方案(例如以一种或更多种化疗剂处理;例如CHOP、R-CHOP或R-EPOCH); b) 如果受试者的确定的诊断结果为ABC亚型DLBCL,则对该受试者指定处理方案;或c) 如果受试者的确定的诊断结果为ABC亚型DLBCL,则不对该受试者指定处理方案(例如不指定CHOP、R-CHOP或R-EPOCH)。

[0288] VII. 处理DLBCL受试者的方法

[0289] 在一些实施方案中,在对DLBCL亚型分类后,例如如果该受试者评分为GCB DLBCL或非GCB(例如,ABC)DLBCL,则该受试者可选为用一种或更多种疗法处理。另外,该一种或更多种疗法可随后施用至该选择的受试者。

[0290] 在一些实例中,以环磷酰胺、多柔比星、长春新碱和泼尼松或泼尼松龙(CHOP)化疗剂或利妥昔单抗加CHOP(R-CHOP)化疗剂施用给该受试者。在一个实例中,还以依托泊苷至R-CHOP施用给该受试者,得到称为R-EPOCH的药物组合。

[0291] 在其它实例中,以一种或多种单克隆抗体向该受试者给药,例如抗CD20(例如,利妥昔单抗、奥法木单抗(ofatumumab)、奥瑞珠单抗(ocrelizumab)、GA101、托西莫单抗(tositumomab)、或替伊莫单抗(ibritumomab tiuxetan)),抗CD22(例如,依帕珠单抗(epratuzumab))、抗CD19(例如,SAR3419或布那图单抗(blinatumumab))、抗CD30(例如布瑞图单抗(brentuximab vedotin))、抗CD40(例如,达西图单抗(dacetuzumab)或卢卡图单抗(lucatumumab))、或抗CD70(例如,SGN-75)。在一些实例中,单克隆抗体缀合至毒素、放射性标记物或药物。在其它实例中,以Bcl-2抑制剂(例如,ABT-737、奥巴克拉(obatoclastax)或奥默森钠),mTOR抑制剂(例如,雷帕霉素(rapamycin)、依维莫司(everolimus)、西罗莫司(temsirolimus)、或deforolimus(ridaforolimus)),Syk抑制剂(例如,福斯坦替尼(fostamatinib)(他曼替尼(tamatinib)),蛋白酶体抑制剂(例如,硼替佐米

(bortezomib)), 蛋白激酶C抑制剂(例如, 泽兰素(enzastaurin)或sotrastaurin), 免疫调节剂(例如, 来那度胺(lenalidomide)和pomalidomide), aurora A激酶抑制剂(例如 alisertib), 或组蛋白脱乙酰酶抑制剂(例如, 伏立诺他(vorinostat)、罗米地辛(romidepsin)、帕比司他(panobinostat)、belinostat、mocetinostat、abexinostat或entinostat)的一种或更多种向该受试者给药。

[0292] 在一些实例中, 以CHOP、R-CHOP、或R-EPOCH作为一线治疗向受试者给药, 而另外的治疗, 例如上文描述的那些(并在一些实例中干细胞移植)在复发时或者该DLBCL对CHOP、R-CHOP或R-EPOCH处理不应答时给药。在其它实例中, 用R-CHOP或R-EPOCH以及上文列出的一种或更多种另外的治疗向DLBCL受试者给药。也可用利妥昔单抗和上文列出的一种或更多种另外的治疗向该受试者给药。

[0293] 可以用于处理GCB或非GCB DLBCL受试者的另外的治疗包括放疗和/或一种或更多种化疗剂。在一些实例中, 化疗剂包括, 但不限于, 烷化剂, 例如氮芥(例如, 苯丁酸氮芥(chlorambucil)、氮芥(chlormethine)、环磷酰胺(cyclophosphamide)、异环磷酰胺(ifosfamide)和美法仑(melphalan), 亚硝脲类(例如, 卡莫司汀(carmustine)、福莫司汀(fotemustine)、罗莫司汀(lomustine)和链佐星(streptozocin), 铂化合物(例如, 卡铂、顺铂、奥沙利铂(oxaliplatin)和BBR3464), 白消安(usulfan), 达卡巴嗪(dacarbazine), mechlorethamine, 甲基苄肼, 替莫唑胺(temozolomide), 噻替派(thiotepa)和乌拉莫司汀(uramustine); 抗代谢药, 例如叶酸(例如, 甲氨喋呤、培美曲塞(pemetrexed)和雷替曲塞(alitretrexed), 嘌呤(例如, 克拉屈滨(cladribine)、氯法拉滨(clofarabine)、氟达拉滨(fludarabine)、巯嘌呤和tioguanine), 嘧啶(例如卡培他滨(capecitabine)), 阿糖胞苷(cytarabine), 氟尿嘧啶, 和吉西他滨(gemcitabine); 植物生物碱类, 例如鬼臼(podophyllum)(例如, 依托泊苷(etoposide)和替尼泊苷(teniposide)), 紫杉烷(例如多西紫杉醇(docetaxel)和紫杉醇(paclitaxel)), 长春花(vinca)(例如, 长春碱(vinblastine)、长春新碱(vincristine)、长春地辛(vindesine)和长春瑞滨(inorelbine)); 细胞毒性/抗肿瘤抗生素, 例如蒽环类化合物家族成员(例如, 柔红霉素(daunorubicin)、多柔比星(doxorubicin)、表柔比星(epirubicin)、伊达比星(idarubicin)、米托蒽醌(mitoxantrone)和戊柔比星(valrubicin), 博来霉素, 羟基脲和丝裂霉素; 拓扑异构酶抑制剂, 例如拓扑替康(topotecan)和伊立替康(irinotecan); 单克隆抗体, 例如阿仑单抗(alemtuzumab), 贝伐单抗(bevacizumab), 西妥昔单抗(cetuximab), 吉姆单抗(gemtuzumab), 利妥昔单抗(rituximab), 帕尼单抗(panitumumab)和曲妥珠单抗(trastuzumab); 光敏剂, 例如氨基乙酰丙酸, 氨基乙酰丙酸甲酯, 卟菲尔钠和维替泊芬; 以及其它药物, 例如阿利维A酸(alitretinoin), 六甲蜜胺(altretamine), 安吡啶(amsacrine), 阿那格雷(anagrelide), 三氧化二砷, 天冬酰胺酶, 贝沙罗汀(bexarotene), 硼替佐米(bortezomib), 塞来考昔(celecoxib), 地尼白介素(denileukin diftitox), 埃罗替尼(erlotinib), 雌莫司汀(estrामustine), 吉非替尼(gefitinib), 羟基脲, 伊马替尼(imatinib), 喷司他丁(pentostatin), masoprocol, 米托坦(mitotane), 培加帕酶(pegaspargase)和维甲酸(tretinoin)。

[0294] 可以用包括上文描述的那些的一种或更多种的治疗剂组合处理受试者。在具体实例中, 可以用蛋白激酶C抑制剂联合利妥昔单抗、吉西他滨和奥沙利铂(R-GEMOX)向受试者

给药。在另一具体实例中,可以用蛋白酶体抑制剂(例如硼替佐米)联合R-CHOP或联合依托泊苷、长春新碱和多柔比星加环磷酰胺和泼尼松(EPOCH)向受试者给药。在具体实例中,用R-ACVBP(利妥昔单抗加多柔比星、环磷酰胺、长春新碱、博来霉素和泼尼松)向受试者给药。

[0295] 本领域普通技术人员可为受试者选择合适的疗法、剂量和施用时间表。所需的剂量将随受试者而变化,取决于受试者的物种、年龄、体重和一般状况、所用的治疗剂,以及施用模式。

[0296] 检测表达的阵列

[0297] 本文公开了可用于检测或测量基因表达(例如表1中两种或更多种生物标志物的表达)的阵列,例如用于在DLBCL受试者中确定DLBCL亚型。在一些实例中,可将所公开的阵列用于检测一种或更多种对照基因(例如,标准化生物标志物)的表达。在具体实例中,该阵列表面包括板、微珠或流动池。

[0298] 在一些实施方案中,阵列可包括固体表面,该固体表面包括特意离散的区域或可寻址位置,每个区域具有至少一个固定化的能够直接或间接杂交表1中的生物标志物的寡核苷酸(于是该阵列可具有多个区域,每个例如特异于表1中的生物标志物)。例如,该阵列可包括特意离散的区域,每个区域具有能够直接或间接与表1中的至少2、至少3、至少5、至少10个或全部16个生物标志物杂交的至少一个或至少两个固定化的捕获探针(例如能够直接或间接与CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的每一个杂交的捕获探针)。在一些实例中,寡核苷酸通过它们在阵列上的位置而可识别。在其它实例中,寡核苷酸通过可检测标记物而可识别(直接或间接与该寡核苷酸缔合)。在另一实例中,阵列可包括特异离散的区域,每个区域具有至少一个或至少两个固定化的寡核苷酸。

[0299] 在一些实例中,该阵列包括两个或更多个能够直接或间接与DLBCL标签基因或对照基因的NPP特异杂交的捕获探针。例如,该阵列可包括这样的寡核苷酸,其包括如下核苷酸或由它们组成:与表1中的DLBCL标签基因中的至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15个或全部16个互补的寡核苷酸(例如包括如下寡核苷酸或由它们组成的寡核苷酸:与CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中每一个互补的寡核苷酸,例如显示在SEQ ID NO:1-16中的那些)。在一些实例中,该阵列(例如包括能够直接或间接与CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中每一个特异杂交的捕获探针的阵列)还包括与至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8或至少9个对照基因互补的寡核苷酸或由这样的寡核苷酸组成。

[0300] 在具体实例中,该阵列包括这样的寡核苷酸(例如,程序化接头(programming linkers)),其包括靶核酸序列,例如包括表1中公开的靶核酸的全部或部分区域的核酸(或与这些序列互补的寡核苷酸)。在一些实例中,程序化接头为双功能寡核苷酸,其包括与同基底缔合的寡核苷酸(例如锚)互补的部分和与NPP的至少一部分互补的部分。

[0301] 在一些实例中,该阵列可包括具有空间上离散的区域(例如,多孔表面的孔、微珠、或流动池内的通道)的表面,每个区域包括稳定(例如,共价地)附着至该表面的锚和寡核苷酸

酸(例如,程序化接头),其中该程序化接头为异质双功能接头(hetero-bifunctional linker),其具有与该锚互补的第一部分和互补于NPP的第二部分,其中该NPP互补于靶核酸(例如如下的两个或更多个:CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS,例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个,例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的全部)。在一些实施方案中,该阵列包括双功能接头或主要由它们组成,其中双功能接头第一部分与锚互补,第二部分与NPP互补。在一些实例中,该阵列还包括双功能接头,其中双功能接头第一部分与锚互补,第二部分与互补于对照标志物的NPP互补。在另外的实例中,该阵列还包括双功能接头,其中双功能接头第一部分与锚互补,第二部分与互补于阴性对照(例如ANT)的NPP互补。这样的阵列附着有杂交至不与NPP互补的双功能接头的至少一段的锚。这样的阵列还可包括(1)杂交至程序化接头的第一部分的锚探针,(2)杂交至程序化接头的第二部分的NPP,(3)具有杂交NPP的第一部分和杂交检测探针的第二部分的双功能检测接头,(4)检测探针;(5)标记物(例如亲和素 HRP),或它们的组合。

[0302] 试剂盒

[0303] 本文还公开了试剂盒,它们可用于检测或测量DLBCL标签基因的水平(例如表1中两个或更多个生物标志物的表达),例如用于表征样品为ABC、GCB或未分类DLBCL,如本文所讨论的。在一些实施方案中,这些公开的试剂盒还可用于检测一种或更多个对照标志物(例如,ANT)的表达。在具体实例中,该试剂盒包括本文提供的一个或更多个阵列,例如包括能够直接或间接特异杂交CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中每一个的捕获探针的阵列。

[0304] 在一些实例中,这些试剂盒包括用于检测核酸表达的探针(例如,NPP或NPPF),例如检测CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个,例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个,或例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的全部。例如,这些试剂盒可包括至少两个不同的NPPF和相应的CFS,其中该至少两个不同的NPPF特异于CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3或TNFRSF8中的两个或更多个(例如至少3、至少4、至少5、至少6、至少7、至少8、至少9个或全部10个),或特异于CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个(例如至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15个或全部16个)。例如,这些试剂盒可包括特异于CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中每一个的NPP(或NPPF),例如包括显示在SEQ ID NO:1-16中的序列或由其组成的NPP。

[0305] 在一些实例中,这些试剂盒还包括一个或更多个对照生物标志物。在一些实例中,这些试剂盒还包括检测一个或更多个对照生物标志物(例如,ANT)的探针和/或引物。

[0306] 包括在试剂盒中的探针包括一个或更多个互补于或特异杂交至本文公开的靶序列的NPP或NPPF,例如包括显示在SEQ ID NO:1-16中的序列的探针。在一些实例中,这些试剂盒可包括一个或更多个为构建用于检测本文公开的生物标志物的阵列所需的特酸探针。

[0307] 在一些实例中,这些试剂盒包括特异结合表1中列出的一种或更多种生物标志物的抗体,以及任选地特异结合一种或更多种对照生物标志物的抗体。例如,这些试剂盒可包括用于检测蛋白表达的抗体(例如,单克隆抗体、多克隆抗体和/或抗体片段),例如检测CD47、CD86、IL16、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个,例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个(例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的全部),并且在一些实例中,检测一种或更多种对照生物标志物。

[0308] 在一些实例中,该试剂盒另外包括如下的一个或更多个:包括缓冲液(例如裂解缓冲液、杂交缓冲液、洗涤缓冲液和/或测序缓冲液)的容器;包括特异于单链核酸的核酸酶(例如S1核酸酶)的容器;包括核酸程序化接头的容器;包括乙醇的容器;包括变性油的容器;包括蛋白酶K的容器;包括可特异结合至少两个不同NPPF的扩增子的微珠的容器;以及包括用于PCR的试剂(例如PCR缓冲液、聚合酶、dNTP)的容器。在一些实例中,试剂盒还包括对照样品,例如特定量的CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS核酸或蛋白。

[0309] 在一个实例中,试剂盒包括互补于或特异杂交至CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS的NPP或NPPF,例如包括SEQ ID NO:1-16中显示的序列或由它们组成的NPP。如果试剂盒包括NPPF,则该试剂盒还可包括对应的CFS。在一些实例中,这样的试剂盒还包括裂解缓冲液、核酸酶溶液(例如包括S1的核酸酶溶液)、终止缓冲液(例如,可中和或失活该核酸酶的终止缓冲液,例如可将pH升高到6以上、例如约9至12的终止缓冲液)。在一些实例中,这样的试剂盒还包括蛋白酶K;变性油;以及任选地微阵列板和移液器吸头。在一些实例中,这样的试剂盒还包括连接缓冲液(例如,包括连接酶的连接缓冲液)。在一些实例中,该试剂盒还包括特异于NPP或NPPF或CFS的一部分的引物,例如可在NPP或NPPF的3'-末端或5'-末端或两末端添加实验标签和/或测序衔接物的引物。

[0310] 在一个实例中,该试剂盒包括执行用于DLBCL亚型的分类器的计算系统,例如本文描述的软件或计算机可读介质,其接收测量或检测DLBCL标签基因(例如CD47、CD86、ENTPD1、FOXP1、FUT8、IL16、ITPKB、LRMP、MME、NF2、PIM1、PTPRC、REL、STAT3、TNFRSF8和TYMS中的两个或更多个)的mRNA表达的多个值,采用等式1处理该多个值,计算预测的概率,以及在表明样品的DLBCL亚型的框架内对样品分类(例如,ABC、GCB、或未分类)。

[0311] 在一个实例中,该试剂盒包括图或表,其显示DLBCL ABC和/或GCB亚型以及任选地未分类样品中期望的值或值的范围(例如预测性概率值)。

[0312] 这些试剂盒还可以包括另外的组分,例如说明性材料和另外的试剂,例如检测试剂,例如基于酶的检测系统(例如,检测试剂包括辣根过氧化物酶或碱性磷酸酶和适当的底物)。这些试剂盒还可以包括另外的组分(例如微量滴定板、移液器吸头等)以方便为其而设计该试剂盒的特定应用。这些试剂盒还可以包括次级抗体(例如特异结合特异结合表1中的蛋白的初级抗体的抗体,或特异结合特异结合对照蛋白的初级抗体的抗体),或标记抗体的手段。在一个实例中,该试剂盒还包括对照核酸和/或对照蛋白。这些试剂盒和适当的内容物本领域普通技术人员所公知。说明性材料可以为电子形式(例如计算机磁盘或光盘)书写

的,或者可以为可视的(例如视频文件)。

[0313] 通过以下非限定性实施例进一步阐述本公开。

实施例

[0314] 实施例1

[0315] DLBCL标签基因列表的鉴定

[0316] 本实施例描述用于构建将DLBCL亚型分类为两类ABC和GCB的分类器的统计方法。另外,将包括不确定区域的第三亚型(称为未分类)加入到该分类中。采用R程序语言版本3.2.2和Bioconductor软件包版本3.1(64位版本)进行所有的数据分析。

[0317] 当采用单个数据集合用于开发和测试估计性能时,采用与如Hastie,Tibshirani,和 Freidman(The Elements of Statistical Learning:Data Mining,Inference and Prediction Springer Science and Business Media,Inc.,2009)中概述相的类似的策略开发通用统计设计程序。在开发该亚型分类器中使用两个通用步骤,模型开发和模型评价。模型开发包括模型/变量选择、参数调整,以及对定义亚型的最终模型进行测试的内部评估。所述评价包括对预测误差的评估。这些步骤二者都利用预期获得的DLBCL FFPE卷(curl)单位点方便样本(single site convenience sample)进行,这些FFPE卷具有由专家临床医生评估的相关DLBCL状态。

[0318] 质量控制

[0319] 在测序完成后将三个质量控制(QC)指标(metrics)应用于每个样本。针对按顺序评估的每个互斥QC指标鉴定失败模式。QC指标的第一个为映射读取数与探针序列的比对率(alignment rate)(基于Bowtie)。期望值为总体比对 $\geq 85\%$,如果不满足这一点则样本具有QC1失败模式。应用于满足QC1质量指标的所有案例的第二QC指标为该样本具有差异表达。通过评估表达第一四分位处(first quartile of expression,Q1)和表达第三四分位(third quartile of expression,Q3)处探针间的表达差异来建立表达;如果平均表达之间的比值小于2.0($Q1/Q3 < 2.0$)则该样本将不符合第二QC指标(QC2)(称为S1失败或数字墙(a wall of numbers))。应用于通过QC1和QC2的样本/案例的第三QC指标(QC3)涉及阴性对照:ANT 探针(ANT1、ANT2、ANT3)。用于阴性对照的统计过程控制(SPC)指标在基线研究中建立,该基线研究对来自SUDHL6细胞系裂解物的90孔单一技术重复测序。该SPC表征研究建立用于阴性对照的上公差,然后将该公差应用于所有新样本以证实性能在预期性能内。通过评估平均ANT信号变异性来建立预期性能。该预期为,与均值水平相比,平均ANT不应波动超过预定的截止值。该截止值基于 \log_2 (CPM)为7.75;样本平均ANT大于该值的任何样本都将不符合第三QC指标(QC3)。具有这些失败模式任一种的所有样本都可能损害分类器判定(compromised classifier calls)。

[0320] 样本

[0321] 总共257个病例具有足够代表性的用于测序的组织,通过了上文描述的测序质量控制程序,并具有通过基于表达谱(GEP)确认的DLBCL状态。用在该分析中的样本通过了用于平均ANT值(阴性对照)的质量控制指标,并具有 $\geq 85\%$ 的探针与序列比对。RNAseq是对映射至特定序列的读取数数量计数的方法。比对率是量化有多少碱基对被正确映射到从PCR产生的RNA拷贝的方式,预期样品内有至少85%的可获得读取数将被映射。小于该量意味着

样品中存在未映射产物。图2显示了从总的可用病例中进行病例选择的数据流程图。

[0322] 从FFPE组织切片制备DLBCL组织裂解物。测量该组织切片,接着使用刀片将其刮入已标记的eppendorf管,并避免载玻片上任何过量的石蜡。将样品面积用于计算所添加的裂解缓冲液的量:总面积(cm^2) \times $28\mu\text{l}/0.25\text{cm}^2$ 组织或总面积(cm^2) \times $25\mu\text{l}/0.25\text{cm}^2$ 组织。将该样品悬浮在裂解缓冲液(预加热的(50°C)包括甲酰胺和SDS的SSC)中。接着在该组织悬浮液上覆加五百(500) μl 含有表面活性剂(例如Brij-97)的矿物油(“非水层”),将该裂解反应在 95°C 温育10至20分钟。在简单冷却该反应混合物后,以裂解缓冲液体积1/20的比例添加蛋白酶K,并在 50°C 继续温育60至120分钟。该裂解反应可立即使用或冷冻并保存在 -70°C 至 -80°C 。在随后使用前将冷冻的裂解反应于 50°C 解冻10至15分钟。对于已知样品输入的样品,该样品输入为 $0.02\text{cm}^2/28\mu\text{l}$ 。

[0323] qNPS测定

[0324] 将二十八(28) μl 的每个裂解反应混合物置于96孔板的孔中,覆加 $70\mu\text{l}$ 非水层。向每个孔添加 $5\mu\text{l}$ 含有具有5'-和3'-侧翼序列的核酸酶保护探针(NPPF)和用于每个侧翼序列的适当互补性侧翼序列(CFS)的混合物。该NPPF的序列包括特异结合靶的捕获序列以及3'-和5'-侧翼序列。表2显示了用于该DLBCL分类器的16个基因的NPPF的捕获序列。因此,靶序列为所显示序列的互补序列。在混合物中存在与待检测的多个mRNA靶中每一个互补的(1) nM(过量的)NPPF。每个96孔板包括30(每个一式三份)个不同的DLBCL裂解物和6个对照样品(两个无样品对照和4个细胞系裂解物)。以不知情的方式将DLBCL样品随机化到平板上(即,盲样注释)。

[0325] 表2.用于检测靶序列的NPPF的捕获部分

靶	NPPF 捕获序列	SEQ ID NO:
CD47	AAACTGGTTACCTAGAGGCCAGAGGCCCTAGGACCTGA AAGGCATCATT	1
CD86	GTCTCCTCTTGGCATAACGGAGCAGAGCTGGAGTTACAG GGAGGCTATTCC	2
ENTPD1	CCATGAGGAAGACATAGGTGGAGTGGGAGAGAGGTTGTG GACAATGGTTGC	3
FOXP1	CTCTTCCCGTATTGCGCTGGCTAAGTTGCCAGAGTGGG ATTTCCCATGG	4
FUT8	CAGTTTGGCTGAAATGGTTTGAAGTGGTTCGAAC TGAGTTTGGTC	5
IL16	GTCCATCAGGCAGTGCCTTGATGATGTTCCAGGCTTCAA ACCGTGTGAGG	6
ITPKB	GCGCCTCAAACATGCCACTTTCTGGTTCACCTGCACGT TCTGCAACTCG	7
LRMP	GCTATTTCTACAGCGGTTAAAGTCCTGATATGTCAAGTT GCCTCTTTCTG	8
MME	GGCCTTGCAGAAAGCATTCTGGACTCCTGTAGGTTCCG GCTGAGGCTGC	9
NF2	GAGAGAAATCTCTACAGGGTCGTAGTTCAAGGCAATTG CACATAAGAGGG	10
PIM1	GCTCACCTTCTCAGCAGGACCACTTCCATGGGCACTCG AGTGCCATTAG	11
PTPRC	GGAACAATTCCTCCTCTGTTACCCTAAGAACAACCCAC TTGCTAGCTGC	12
REL	CAGAGGTGCCATTGAGGCATGATGTGACAATCCACTTG AGATGGGCCAG	13
STAT3	CAGCTTCAGGATGCTCCTGGCTCTCTGGCCGACAATACT TTCCGAATGCC	14
TNFRSF8	CTGGGACCAATGCTGTTCTCGGCAGTGCCCATCCTGCCA TCTGTTTGCTG	15
TYMS	CTTCAGGCCCGTGATGTGCGCAATCATGTACGTGAGCAG GGCGTAGCTGG	16

[0327] 将该96孔测试板在85℃加热10分钟让核酸变性,然后让其在52℃温育16小时,以允许NPPF杂交至它们各自的mRNA靶,以及允许CFS杂交至它们各自的测序序列靶。

[0328] 杂交步骤后,将含过量S1核酸酶的20μl醋酸钠缓冲液(2.5U/μl)添加至每个孔的水相中。该S1反应在52℃进行90分钟,以消化未结合的mRNA和未结合的NPPF以及未结合的CFS。

[0329] 在S1消化步骤期间,对应96孔测试板中的反应,通过向每个孔添加5μl含1M Tris-HCL的溶液(pH 9.0)制备96孔“终止”板。将96孔测试板中每个孔40μl反应转移至第二96孔终止板中的相应孔,并覆加70μl非水层。将终止板在100℃温育20分钟,接着室温下冷却30分钟。

[0330] 在该捕获反应后,可将板保存在-20℃或直接用于PCR标记(PCR tagging)。如果板为冷冻保存,则可在用于PCR之前将它们室温轻轻摇动解冻30至40分钟。

[0331] 按下文进行PCR标记,用于测序文库制备。将包括6μl水、15μl NEB OneTaq® HotStart X2Master Mix GC缓冲液、3μl正向引物、3μl反向引物以及3μl样品的PCR试剂添加至干净的管中。作为阴性对照,一些样品仅接收水。PCR进行20个循环。

[0332] 为了在PCR后除去过量的引物,将15μl的每个PCR样品(PCR阴性对照除外)合并微量离心管中。让350μl的合并PCR反应与135μl 5M NaCl、137μl 40%PEG溶液以及35 μl

AMPure微珠混合。室温下将该管涡旋震动并温育5分钟。将该管置于磁体架上5分钟,以允许收集微珠。弃去上清液。向微珠添加30 μ L 10mM Tris-HCl pH 8.0和75 μ L AMPure XP。室温下将该管涡旋震动并温育5分钟。将该管置于磁体架上5分钟,以允许收集微珠。弃去上清液。用80%乙醇洗涤微珠两次,并让微珠在37 $^{\circ}$ C干燥5分钟。将微珠悬浮在40 10mM Tris-HCl pH 8.0中。将30 μ L上清液添加至新管,其含有该净化产物(the cleaned-up)。通过在2%琼脂糖凝胶上运行10 μ L的该文库来确认引物的去除。

[0333] 采用KapaQuant试剂盒(Kapa Biosystems,Catalog#10176100),按照制造商指导,将净化的扩增子1:10,000稀释。

[0334] 采用MiSeq测序仪(Illumina)对该净化的扩增子测序。按照制造商关于文库变性以及测序仪设置和上样的说明书,将30pM的文库上样至MiSeq 150bp V3试剂盒。

[0335] 模型开发

[0336] 作为建立分类器算法的第一步,将257个经质量控制的DLBCL样本(它们的基因表达(GEP)判定(calls)是可供使用的)随机分为两个不重叠的数据集合子集,其包括172个随机样本用于分类器开发;剩下的86个用作验证集合。对这两个随机样本的相似度进行评估,以确保该随机选择没有引入偏倚(bias)。表3描述了通过GEP判定的这两个子集的分布。

[0337] 最初,采用qNPS在所有的DLBCL样品中确定96个基因的表达。图3显示了在去除了横跨这两个集合的阴性、阳性和管家基因后可能的分类器基因的随机性质;偏倚将以 \log_2 (CPM)计数之间的差异而不是重叠显示出来,如图3中所示。在进行统计分析之前,将管家和对照探针(表4)去除。对这些探针的去除导致有83个基因考虑用于分类器。

[0338] 表3.开发和验证集合之间的GEP判定的分布

[0339]		ABC	GCB
	开发	86	86
	验证	36	49

[0340] 表4.通过Parser命名的管家和对照探针

[0341]	管家探针	对照探针
	ACTB	POS_CTRL_POS1
	EEF1G	NEG_CTRL_ANT1
	EIF4A1	NEG_CTRL_ANT2
	GAPDH	NEG_CTRL_ANT3
[0342]	RPL19	
	RPL4	
	RPL6	
	RPS29	
	TBP	

[0343] 采用带有logit联系函数的通用线性模型,也称为逻辑回归(logistic regression),来进行分类器开发。逻辑回归为响应(结果)数据具有二个分类(也称为二元)时使用的模型;在本情形中,二个分类为ABC和GCB分类/亚型。用于评估给定分类的概率的统计模型书写为:

[0344]
$$Pr(Y = y|x) = \frac{e^{(\beta_0 + x^T \beta)}}{1 + e^{(\beta_0 + x^T \beta)}} \quad (\text{等式 1})$$

[0345] 其中 β_0 表示截距, β 为与探针向量 x^T 相关的参数估计。将开发样本用于拟合(1)描述的逻辑模型以获得参数估计 $\hat{\beta}_0$ 和 $\hat{\beta}$,它们称为“权重”。这些权重与每个基因的表达水平计数值联合使用,以确定在任何新的(独立于开发数据集合)数据集合中分类为ABC或GCB的预测概率 $Pr(Y = y | x)$ 。最终的模型可具有与用来计算该概率的探针一样多的 $\hat{\beta}$ 值。该估计的概率介于0和1之间,并且将概率的最佳截止值用于对未来的患者分类。

[0346] 该模型的开发部分包括两个步骤,训练和测试。测试阶段是指用于鉴别最佳探针/靶并在靶/探针选择和系数收缩/缩放后获取参数(权重)的内部度量指标。这与一些建议应使用三个随机数据集合:训练、测试、验证的统计文献(Tibshirani, J. Royal Statistical Society (Series B) 58:267-288, 2006)稍有不同。出于本分类器的目的,在相同的开发集合中进行训练和测试。在测试阶段进行用于针对ABC、GCB和未分类患者分类概率的最佳截止值的确定。这些步骤都采用病例的开发集合进行。

[0347] 对用于概率计算以及优化权重的83个靶的选择采用称为“lasso”的算法进行。统计上,这些方法是通过作为该优化的一部分应用的“惩罚(penalty)”类型来分类的,这种惩罚缩小了权重以适应基于小样本大小的可能偏倚。描述这种方式的方法已有公开(例如,参见 Tibshirani, J. Royal Statistical Society (Series B) 58:267-288, 2006)。开发本分类器所使用的 lasso 算法在R语言“glmnet”软件包中执行,并在Friedman等(J. Statistical Software, 33(1):1-22, 2010)中有描述。采用10折交叉验证作为重采样来确定最佳变量。这是如下的方法,其中将最初的开发数据集合分割成大小为10的相同大小的子样本,并产生了100个这样的子样本。将这些子样本用于发现能最优分类成两个亚型(即,ABC、GCB)的模型(例如,拟合的回归参数)和确定将用作权重的值(例如,在可能发生的任何缩放后的系数)。

[0348] 采用这种方法,将83个靶的范围缩小成表5中的16个靶。表5还列出了用于最佳系数集合的“权重” $\hat{\beta}$ (即,针对每个靶)。具有小于1的权重(即,负值)的那些靶与ABC分类相关,而具有大于1的权重的那些靶与GCB分类相关。

[0349] 表5. 用于DLBCL分类器的系数

[0350]

基因	系数(权重)
CD47	-0.16
CD86	0.83
ENTPD1	-0.71
FOXP1	-0.56
FUT8	-0.30
IL16	-0.51
ITPKB	1.16
LRMP	0.16
MME	0.59

NF2	-0.70
PIM1	-0.13
PTPRC	0.07
REL	0.20
STAT3	-0.17
TNFRSF8	-0.01
TYMS	-0.25

[0351] 预测概率截止值

[0352] 一旦获得最佳靶和权重,就获得了预测概率的截止值。当恰好存在两个分类时,通过在预测概率的整个范围上最大化敏感度(正确指定样品为ABC的概率)和特异度(正确指定样品为GCB的概率)而获得定义类别的概率。该DLBCL分类器还包括“未分类”样本子集。这些样本鉴定为这样的样本集合:它们可能既不是ABC也不是GCB,并因此潜在地落在独立亚型下(例如,对其不能做出ABC或DCB判定)。确定敏感度+特异度(其组合称为准确度)开始从最大点下降的点,在该相同点误分类率增加。这被称为下限;上限是与此值相关的对称高点。图4以图表描述了这种判定;数值上,对于低点,截止值为0.43,对于高点为0.57。获得的ABC/GCB判定为:具有估计概率 <0.43 的那些病例为ABC,具有估计概率 >0.57 的那些为GCB,介于0.43-0.57(含0.43和0.57)的每个病例为未分类。

[0353] 实施例2

[0354] 将DLBCL分类器应用于独立样本

[0355] 本实施例证实实施例1开发的DLBCL分类器允许在福尔马林固定、石蜡包埋(FFPE)样本中鉴定DLBCL的GCB和ABC亚型。

[0356] 该分类器的最终测试显示了低倾向的“过拟合”(例如,过度优化),并将得到的权重应用于验证集合以证实性能。因此,分类器的开发在该数据集合上完成。

[0357] 将未用于开发分类器的剩余86个DLBCL样本用作验证集合。如上文描述进行表5中16个基因的qNPS分析。让所得的数据经分类器处理(例如,每个基因采用qNPS获得的值以等式1分析,并确定预测概率)。如本文其它地方提到的,CD86、ITPKAB、LRMP、MME、PTPRC和REL为趋向与DLBCL的GCB亚型相关的基因,而CD47、ENTPD1、FOXP1、FUT8、IL16、NF2、PIM1、STAT3、TNFRSF8和TYMS为趋向与DLBCL的ABC亚型相关的基因。将16个基因中的每一个从qNPS获得的输出值乘以表5中显示的权重值,确定这16个基因的总和,并确定预测概率,其中所得的预测概率值确定样品为ABC、GCB或未分类。

[0358] 分类器性能的指标包括误分类率,定义为该分类器能够获得与GEP判定相同的ABC/GCB判定的病例比例。其它的指标包括阳性预测值和阴性预测值(分别为PPV和NPV;即正确分类为ABC或GCB的病例)。注意,未分类样本不包括在这些指标的评估中。表6显示了分类器应用至验证集合的验证总结,图5显示了对每个样品的分类。

[0359] 表6. 分类器在验证集合上的性能指标

[0360]

	病例比例
正确分类为GCB (PPV)	0.95 (42/44)
正确分类为ABC (NPV)	0.84 (32/38)
误分类	0.10 (8/82)

总体正确判定	0.96 (74/82)
--------	--------------

[0361] 实施例3

[0362] 体外DLBCL诊断性测定

[0363] 本实施例描述可用于从FFPE组织确定弥漫性大B细胞淋巴瘤 (DLBCL) 肿瘤的细胞起源 (cell of origin, COO) 亚型的方法, 其使用HTG EdgeSeq系统, 检测在Illumina MiSeq新一代测序仪上进行。通过分类算法评价综合的数据 (profiled data), 并确定肿瘤为活化的B细胞样 (ABC)、生发B细胞样 (GCB) 或非分类亚型。

[0364] 该HTG EdgeSeq系统将定量核酸酶保护测定 (qNPA) 化学与新一代测序 (NGS) 平台结合, 使得能够在单一测定中进行96个靶向基因的半定量分析。该qNPA化学不需要从大多数样品中提取核酸, 因此显著降低的样品输入要求。其也消除了与核酸提取、大小选择、cDNA 合成以及衔接物连接相关的偏倚。该自动化程序降低了操作错误, 因此有助于确保从诸如 FFPE组织的宝贵样品制备的测序文库能够可重复地生成。HTG EdgeSeq系统的使用让实验人员能够在大约36小时内从原始样品开始到准备好文库, 且上机时间小于3小时。

[0365] 该qNPA技术 (参见, 例如美国专利号8,741,564, 通过引用并入本文) 使得能够以无需提取、无RNA扩增方式进行mRNA定量。HTG EdgeSeq化学方法的两个主要元素为DNA 或RNA杂交和S1核酸酶消化。侧翼有通用翼序列 (wing sequence) 的功能性DNA核酸酶保护探针杂交靶RNA, 该靶RNA可为可溶性的和在生物基质中交联的。通用DNA wingmen 探针杂交该翼序列以防止S1核酸酶消化。添加S1核酸酶以消化过量的、未杂交的DNA探针和未杂交的RNA; 仅剩下的完好、功能性DNA保护探针是与靶RNA杂交的那些。这产生了基本上1:1比例的DNA检测探针和样品中最初靶向的RNA。热变性将保护探针从 DNA:RNA双链体释放。该释放的DNA保护探针为计数做好准备。

[0366] 在热循环仪中用测序衔接物和标签标记DNA保护探针。浓缩、合并该经标记的DNA保护探针, 并为在Illumina MiSeq平台上采用标准NGS方案测序做好准备。处理并报告来自NGS仪器的基因表达数据, 例如通过HTG EdgeSeq主机系统软件。

[0367] 与靶mRNA中互补序列杂交的特异设计的寡核苷酸 (探针) 用作测量mRNA水平的手段。这些寡核苷酸设计成这样的序列, 它们经检查以确保缺乏潜在二级结构, 并且针对其它报告的人类序列进行BLAST检索以确保与测定中的其它基因缺乏明显同源性或互补性。每个寡核苷酸为为100聚体, 包含5'-末端处的25聚体“翼”, 3'-末端处的25聚体“翼”, 以及它们之间的与靶mRNA互补的50聚体序列。

[0368] 该HTG EdgeSeq DLBCL细胞起源测定包括四个主要步骤。(1) 样品制备 (手动步骤): 添加HTG裂解缓冲液以裂解和/或透化FFPE组织样品, 使可用的RNA随后结合相应的靶特异性核酸酶保护探针 (NPP)。将裂解的样品转移至标准96孔微量滴定板, 也称为样品板。(2) 通过HTG EdgeSeq化学的靶捕获 (自动化步骤): 通过HTG EdgeSeq化学完成靶捕获。简言之, 将核酸酶保护探针 (NPP) 过量添加至样品板内的裂解样品中, 并增加靶RNA。然后添加S1核酸酶以消化未杂交的mRNA和过量的NPP, 因此产生了化学计量量的靶 mRNA/NPP双链体。在S1消化完成后, 将经加工的样品转移至具有v型底的称为终止板的新96孔微量滴定板, 并通过添加终止溶液和随后热变性S1酶来终止S1消化。(3) 用于添加衔接物和条形码的PCR扩增 (手动步骤): 将来自终止板的每个经加工样品用作模板来用特异设计的引物 (称为标签) 建立PCR反应。这些标签共有与探针的5'-末端和3'-末端“翼”序列互补的共同序列和

Illumina MiSeq测序平台上簇生成(cluster generation)所需的共同衔接物。此外,每个标签含有唯一的条形码,用于样品鉴定和多重操作。可使用12个正向和8个反向引物标签,由此可在Illumina测序平台上同时分析多至96个样品。HTG EdgeSeq DLBCL 细胞起源测定被配置为在Illumina MiSeq仪器上同时分析8、24或96个样品。在PCR扩增完成后,进行清理程序(clean-up procedure)以从PCR产物(称为文库)除去未并入的标签。然后用各个PCR反应进行PCR清理。(4)文库定量和测序(半自动化步骤):采用MiSeq v3 试剂盒在Illumina MiSeq平台上进行测序。可平衡和调节样品文库的浓度,以确保合适的簇生成。关于靶基因的mRNA表达的测序数据被输入HTG EdgeSeq parser软件,于此DLBCL C00亚型可用表格形式报告给实验人员。

[0369] **样品加工:**对于FFPE样品,所测试的切片应为4至6微米厚,并安置在玻璃显微镜载玻片上。待测试的切片区域应确定为DLBCL。对于每个样品,强烈建议用清晰标记了感兴趣区域的一张H&E染色载玻片或扫描的图像,以指导从载玻片上对感兴趣区域进行宏切(macro-dissection)。通常,需要未染色切片。如果切片中含有的肿瘤部分极少,则可以刮取多个切片并合并到单只管中,用于测试。

[0370] 当测量肿瘤大小时,从该组织切片排除任何正常的、其它非肿瘤或坏死组织。在所述排除为不可行的情况下,在大小计算中不包括非肿瘤组织。当组织切片提供与最大建议的样品输入相比更大的表面积时,刮取整个相关的肿瘤部分,并采用裂解缓冲液稀释至适当的工作浓度。如果测试极小的组织,则可以将多个切片合并到单只管中以实现期望的组织输入量,并作为单个样品加工。当保存在室温时,变性油可以形成白色沉淀物。在50°C加热该变性油10分钟,并简单混合底部以溶解该沉淀物。不立即使用的裂解样品可以在微量离心管中冷冻在-80°C长至一(1)周。样品不应冷冻在96孔样品板中。

[0371] FFPE组织加工可如下进行。用H&E染色、IHC或其它适当手段鉴定待测试的正确区域;采用不可擦除记号笔在盖玻片背面对此区域做标记。测量并计算该标记区域涵盖的总表面积(mm^2)。排除低细胞性、非靶组织、坏死或以其它方式损害测试结果的组织区域。用于HTG EdgeSeq DLBCL细胞起源测定的可接受输入量为1.56至12.5 mm^2 组织/测试。在极小样品的情况下(例如,核芯针活检),可以将多个组织切片合并到相同管中,以得到所需的样品量。在50°C解冻裂解缓冲液和变性油30分钟。蛋白酶K应保存在-20°C,直至准备使用。从未染色切片开始,采用标记的载玻片作指导,用刀片从未染色载玻片刮下适当面积的组织。将该组织置于适当标记的无RNase的微量离心管中(可以使用移液管帮助从刀片移除并置于该微量离心管中)。短暂离心(~20秒),以将FFPE组织拉至管底。对于每个组织靶向面积,向该管添加35 μL HTG裂解缓冲液。例如:

[0372] 表7. 组织和裂解缓冲液的示例量

[0373]

组织量	裂解缓冲液体积(μl)
1.56-12.5 mm^2	35
12.6-25 mm^2	70
25-50 mm^2	140
50-100 mm^2	280

[0374] 极大切片可以采用上文列出的多倍裂解缓冲液体积来加工。向每个管添加500 μL 变性油,并盖上该管。不进行涡旋震动。将样品加热至95°C 15分钟,让样品在室温冷却10分

钟。以裂解缓冲液体积的1/20的比例向每只管的下层添加蛋白酶K。例如：

[0375] 表8. 裂解缓冲液和蛋白酶K的示例量

裂解缓冲液体积 (μl)	蛋白酶K
35	1.75
70	3.5
140	7
280	14

[0377] 通过移液管吸放 (pipetting) 该溶液六次来混合蛋白酶K。将裂解样品在50℃温育3小时, 伴有在定轨摇床上混合, 或者每30分钟的移液管吸放。可对样品进行加工或者冷冻保存在-70至-80℃多至4周。

[0378] 在HTG EdgeSeq处理器上的初始化运行: 从包装取出HTG EdgeSeq测定试剂包 (Assay Reagent Pack), 并在室温解冻80分钟。不要摇动或倒置试剂盘, 这些试剂将在使用前在仪器上混合。使用10%漂白剂, 然后用70%异丙醇清洁HTG EdgeSeq测定试剂包。如果裂解的样品为冷冻的, 则在50℃解冻45分钟, 并在使用前用移液管吸取混合每个样品的水相。短暂离心样品以让所有液体返回管底。吸移35μL的每个样品至样品板的适当孔中。采用HTG EdgeSeq系统条形码扫描仪, 从HTG EdgeSeq Plate Pack扫描该样品板。于是启动HTG EdgeSeq处理器, 并运行大概20小时。处理后, 移除终止板并分析。

[0379] 文库加工: 可例如采用NEB OneTaq®HotStart 2X Master Mix GC缓冲液、分子级别水以及测序标签来制备PCR预混液 (master mixes)。将该混合物, 连同反向引物, 添加至终止板内一部分适当的样品中, 并如下进行PCR: 95℃4分钟, 接着20个循环的: 95℃15秒, 56℃45秒, 以及68℃45秒, 然后68℃10分钟, 并保持在4℃。所得的反应可立即处理或保存在-20℃。可如下合并和清洁该文库。将15μL的每个PCR产物转移至新的预先经标记的96孔PCR板。向每个孔添加37.5μL的AMPure XP微珠 (提前升温至室温并通过涡旋震动混合充分)。充分混合并室温下 (RT) 温浴5分钟。将样品放置在磁体架上, 保持3分钟用于微珠收集。应形成可见的微珠沉淀。去除上清液, 不要扰动微珠沉淀。为了防止微珠污染, 如果微珠沉淀松散, 则可以留下5-10μL的残余液体。添加200μL新制备的80%乙醇, 静置30秒至1分钟。从每个孔去除乙醇。重复该乙醇洗涤, 总共洗涤两次。从磁体架移除板, 让其在室温干燥5分钟。向每个孔添加40μL的10mM Tris-HCl, pH 8.0, 并进行移液管吸放混合以重悬微珠。室温静置5分钟。将板放置在磁体架上2分钟, 以收集微珠。转移30 μL的每个文库上清液至新孔。

[0380] 可例如采用下文的KAPA定量系统, 在测序前对生成的各个样品文库定量, 并平衡各个文库浓度。进行1:100系列稀释 (例如, 一式三份)。向该1:100稀释的每个孔添加297μL 10mM Tris/0.05% Tween。向深孔板的对应孔添加3μL的经清洁文库。混合样品。进行第二次1:100稀释, 得到最终稀释因子1:10,000。向该1:10,000稀释的每个孔添加297μL 10 mM Tris/0.05% Tween。向该“1:10,000稀释”板的对应孔添加3μL来自“1:100稀释”板的3μL样品。充分混合。如果需要1:100,000稀释, 则进行另外的1:10稀释。室温解冻 KAPA文库定量试剂盒30分钟。对于首次使用, 向2X KAPA SYBR FAST qPCR Master Mix (5ml) 瓶添加1mL的10X Primer Premix, 并混合。按制造商说明书, 计算用于HTG EdgeSeq样品和标准品的2X Mastermix、50X ROX High Dye (如果需要) 和分子生物学级别的H₂O的体积。为所使用的qPCR系统生成适当的顺序条目 (order entry) 和规程。向该 qPCR加样并按制造商说明书循

环样品。在循环结束后,将结果作为Microsoft Excel®文件输出。保证用于标准曲线的QC指标与制造商的说明一致。

[0381] 可例如使用HTG EdgeSeq IVD文库计算器生成合并的文库。

[0382] 在Illumina MiSeq系统上的测序:可例如采用Illumina MiSeq仪器按制造商说明书如下进行合并文库的测序。对于每个Illumina MiSeq仪器,最佳的最终合并文库浓度可以变化。可分析得到的测序数据。

[0383] 在产生样品分类之前,HTG EdgeSeq DLBCL细胞起源测定分类系统进行数据质量评价。未通过这些要求的样品报告为“QC失败”,带有所附的编号。如果没有获得足够的测序读取数用于分类,则确保在文库合并和定量阶段期间样品被正确定量且添加了足够的文库;再次运行来自任何剩余样品裂解物的样品,再次裂解来自FFPE组织新切片的样品。对组织再次评价,以确保所用的区域为肿瘤,或它们的组合。如果数据概况没有反映从DLBCL获得的表达典型模式,则这表明了潜在的样品制备或处理器问题。这种情况下,再次运行来自任何剩余样品裂解物的样品,和/或再次裂解来自FFPE组织的新切片以及再次评价组织以确保所用的区域为肿瘤。

[0384] 进行HTG EdgeSeq DLBCL细胞起源测定:让随HTG EdgeSeq DLBCL细胞起源测定提供的冷冻组分子室温解冻80分钟,不要混合或搅动试剂。在吸取样品至样品板中之前,通过在50°C加热30分钟解冻任何冷冻的样品裂解物。HTG EdgeSeq处理器内消耗品的放置可以随待加工的样品数量而变化。

[0385] 测定结果:测定报告产生三个与DLBCL病例的(1)“ABC”或活化B细胞亚型(2)“GCB”或生发B细胞亚型、(3)未分类(其中基因表达模式不指示ABC或GCB)一致的潜在结果,或者(4)“QC失败”,其表明该样品在分类前未产生通过样品水平质量控制指标的结果,并且应重新运行样品。

[0386] 与那些GCB亚型的相比,显示ABC亚型的肿瘤通常具有更差的预后和对标准治疗(standard-of-care)DLBCL疗法的应答。DLBCL肿瘤的亚型分型可以指导临床医师在复发患者中使用其它治疗形式。(Vose, J Clin Oncol 2011 29:4065-4066;Thieblemont等, J Clin Oncol 2011 29:4079-87)。世界卫生组织DLBCL分类已做出了最新改动,以便认识到这些差异并促进对ABC亚型更有效处理的研究(Swerdlow等, Blood 2016;127:2375-90)。包括在这些选项中的有用于众多新疗法的研究试验。

[0387] 性能特征

[0388] 样品输入:在每孔12.5和0.39mm²之间的两倍稀释度下测试来自FFPE样品(一个ABC 和一个GCB)的重复样品裂解物。在100%的6个测试稀释点,甚至低于建议的1.56mm²组织情况下,分类保持一致。还在两个低表达探针CD274 (PD-L1) 和PDCD1 (PD-1) 中评估了这种一致性。图6和7描绘了从低到高样品输入量的ABC和GCB样品中的样品表达。

[0389] 测定可重复性:采用事先通过Affymetrix™基因表达谱(GEP)表征的80个样品(40ABC,40GCB)评估日常(day to day)重复性。有意排除之前表征为“未分类”的样品。每个样品孔中存在等量的大约5mm²FFPE组织,在单一HTG EdgeSeq处理器和MiSeq测序仪上在三个独立的日子进行测序。在产生的240个总数据集合中,236(98.33%)个通过了预分类QC。一个ABC和一个GCB样品在第1天导致QC失败,两个ABC样品在第2天失败,没有样品在第3天失败。表9描述了与GEP相比较的来自HTG EdgeSeq DLBCL细胞起源测定的样品亚型分型最终

结果,所有天数的总体一致率为100%。

[0390] 表9.与GEP比较的HTG EdgeSeq DLBCL细胞起源测定分类

GEP	第1天		第2天		第3天	
	ABC	GCB	ABC	GCB	ABC	GCB
[0391] ABC	39	0	38	0	40	0
GCB	0	39	0	40	0	40

[0392] 与IHC的测定一致性:HTG EdgeSeq DLBCL细胞起源测定与IHC方法学的一致性通过使用两个独立队列的回顾性研究进行。第一队列由事先通过Visco-Young法 (Lenz等, N Engl J Med 2008;359 (22) :2313-23) 表征为ABC或GCB的132个样本组成。第二队列的24个样本采用Choi法 (Rosenwald等, N Engl J Med 2002;346 (25) :1937-47) 表征。排除未分类样品,为每个队列计算一致率。该相对于GEP优化的HTG EdgeSeq DLBCL细胞起源测定表现与 Choi IHC算法更一致,有91%的总体一致。HTG EdgeSeq DLBCL细胞起源测定与Visco-Young IHC之间的总体一致性为83% (表11)。总体一致率都不包括“未分类”样品,这是由于IHC算法不包括这种亚型分类类别。

[0393] 表10.HTG EdgeSeq DLBCL COO测定分类与Choi IHC一致

		HTG EdgeSeq DLBCL COO		
		ABC	GCB	UNC
Choi	ABC	11	1	1
	GCB	1	10	0

[0394]

与 Choi IHC 总体一致	91%
未分类比例	4%

[0395] 表11:HTG EdgeSeq DLBCL细胞起源测定分类与Visco-Young IHC一致 HTGEdgeSeq DLBCL COO测定

Visco-Young IHC	ABC	GCB	UNC
ABC	60	12	6
GCB	9	40	5

[0396]

与 Visco-Young IHC 的总体一致	83%
未分类比例	6%

[0397] 诊断可重复性:将14个事先被表征的临床样品以两个样品输入量裂解。所有样品的三个重复在三个HTG EdgeSeq处理器上运行,并在单日内测序。汇集所有条件产生了总共252个GCB/ABC分类判定。

[0398] 总共97%的样品 (244/252) 通过了QC指标,用于随后的分类。当未分类样品被排除时,获得了百分之九十九 (99%) 一致 (总共1个错分类);当未分类结果包括在内时,获得了97%一致。总共有8个未分类结果,其中6个获得自单一样品;这些一致率总结在表12中。对每个重复的分类的图形表示显示在图8中。

[0399] 表12:处理器和样品输入的一致

		%通过 QC	%一致
[0400]	处理器		
	仪器 1(N=83)	98.8%	94.0%
	仪器 2(N=81)	96.4%	97.5%
	仪器 3(N=80)	95.2%	97.5%
样品输入	1.5 mm (N=118)	93.7%	94.9%
	5 mm (N=126)	100.0%	97.6%

[0401] 有鉴于其中可以应用本公开的原理的许多可能的实施方案,应当认识到,所阐述的实施方案仅是示例,并不应被视为限制本发明的范围。更确切地,本发明的范围通过所附的权利要求进行限定。因此,我们主张这些权利要求范围和精神内的所有发明均为我们的发明。

- [0001] 序列表
- [0002] <110> HTG分子诊断有限公司
- [0003] <120> 弥漫性B细胞淋巴瘤 (DLBCL) 亚型分型的方法
- [0004] <130> 8736-95651-02
- [0005] <150> 62/234491
- [0006] <151> 2015-09-29
- [0007] <160> 16
- [0008] <170> SIPOSequenceListing 1.0
- [0009] <210> 1
- [0010] <211> 50
- [0011] <212> DNA
- [0012] <213> Artificial Sequence
- [0013] <220>
- [0014] <221> misc_feature
- [0015] <223> 用于检测CD47的NPPF的捕获部分
- [0016] <400> 1
- [0017] aaactggtta cctagaggcc agaggcccta ggacctgaaa ggcattcattc 50
- [0018] <210> 2
- [0019] <211> 50
- [0020] <212> DNA
- [0021] <213> Artificial Sequence
- [0022] <220>
- [0023] <221> misc_feature
- [0024] <223> 用于检测CD86的NPPF的捕获部分
- [0025] <400> 2
- [0026] gtctcctctt ggcatacggg gcagagctgg agttacaggg aggctattcc 50
- [0027] <210> 3
- [0028] <211> 50
- [0029] <212> DNA
- [0030] <213> Artificial Sequence
- [0031] <220>
- [0032] <221> misc_feature
- [0033] <223> 用于检测ENTPD1的NPPF的捕获部分
- [0034] <400> 3
- [0035] ccatgaggaa gacataggtg gagggtggaga gaggtgtgga caatggttgc 50
- [0036] <210> 4
- [0037] <211> 50
- [0038] <212> DNA

- [0039] <213> Artificial Sequence
[0040] <220>
[0041] <221> misc_feature
[0042] <223> 用于检测FOXP1的NPPF的捕获部分
[0043] <400> 4
[0044] ctcttcccgt attgcgctgg ctaagttgcc cagagtggga tttcccatgg 50
[0045] <210> 5
[0046] <211> 50
[0047] <212> DNA
[0048] <213> Artificial Sequence
[0049] <220>
[0050] <221> misc_feature
[0051] <223> 用于检测FUT8的的NPPF的捕获部分
[0052] <400> 5
[0053] cagtttggt gaaatggttt gaactgagtt tggtcgaact gagtttggtc 50
[0054] <210> 6
[0055] <211> 50
[0056] <212> DNA
[0057] <213> Artificial Sequence
[0058] <220>
[0059] <221> misc_feature
[0060] <223> 用于检测IL16的NPPF的捕获部分
[0061] <400> 6
[0062] gtccatcagg cagtgccttg atgatgttcc aggcttcaaa ccggttgagg 50
[0063] <210> 7
[0064] <211> 50
[0065] <212> DNA
[0066] <213> Artificial Sequence
[0067] <220>
[0068] <221> misc_feature
[0069] <223> 用于检测ITPKB的NPPF的捕获部分
[0070] <400> 7
[0071] ggcctcaaa catgcccact ttctggttca cctgcacgtt ctgcaactcg 50
[0072] <210> 8
[0073] <211> 50
[0074] <212> DNA
[0075] <213> Artificial Sequence
[0076] <220>
[0077] <221> misc_feature

[0078] <223> 用于检测LRMP的NPPF的捕获部分
[0079] <400> 8
[0080] gctatttcta cagcggtaa agtcctgata tgtcaagttg cctctttctg 50
[0081] <210> 9
[0082] <211> 50
[0083] <212> DNA
[0084] <213> Artificial Sequence
[0085] <220>
[0086] <221> misc_feature
[0087] <223> 用于检测MME的NPPF的捕获部分
[0088] <400> 9
[0089] ggccttgceg aaagcatttc tggactcctt gtaggttcgg ctgaggctgc 50
[0090] <210> 10
[0091] <211> 50
[0092] <212> DNA
[0093] <213> Artificial Sequence
[0094] <220>
[0095] <221> misc_feature
[0096] <223> 用于检测NF2的NPPF的捕获部分
[0097] <400> 10
[0098] gagagaaatc tctacagggt cgtagttcaa ggcaattgca cataagaggg 50
[0099] <210> 11
[0100] <211> 50
[0101] <212> DNA
[0102] <213> Artificial Sequence
[0103] <220>
[0104] <221> misc_feature
[0105] <223> 用于检测PIM1的NPPF的捕获部分
[0106] <400> 11
[0107] gctcaccttc ttcagcagga ccacttccat gggcactcga gtgccattag 50
[0108] <210> 12
[0109] <211> 50
[0110] <212> DNA
[0111] <213> Artificial Sequence
[0112] <220>
[0113] <221> misc_feature
[0114] <223> 用于检测PTPRC的NPPF的捕获部分
[0115] <400> 12
[0116] ggaacaattt cctcctctgt taccctaaga acaaaccact tgctagctgc 50

- [0117] <210> 13
[0118] <211> 50
[0119] <212> DNA
[0120] <213> Artificial Sequence
[0121] <220>
[0122] <221> misc_feature
[0123] <223> 用于检测REL的NPPF的捕获部分
[0124] <400> 13
[0125] cagaggtgcc attgaggcat gatgtgacaa tccacttgag atgggcccag 50
[0126] <210> 14
[0127] <211> 50
[0128] <212> DNA
[0129] <213> Artificial Sequence
[0130] <220>
[0131] <221> misc_feature
[0132] <223> 用于检测STAT3的NPPF的捕获部分
[0133] <400> 14
[0134] cagcttcagg atgctcctgg ctctctggcc gacaatactt tccgaatgcc 50
[0135] <210> 15
[0136] <211> 50
[0137] <212> DNA
[0138] <213> Artificial Sequence
[0139] <220>
[0140] <221> misc_feature
[0141] <223> 用于检测TNFRSF8的NPPF的捕获部分
[0142] <400> 15
[0143] ctgggaccaa tgctgttctc ggcagtgcc atcctgcat ctgtttgctg 50
[0144] <210> 16
[0145] <211> 50
[0146] <212> DNA
[0147] <213> Artificial Sequence
[0148] <220>
[0149] <221> misc_feature
[0150] <223> 用于检测TYMS的NPPF的捕获部分
[0151] <400> 16
[0152] cttcaggccc gtgatgtgcg caatcatgta cgtgagcagg gcgtagctgg 50

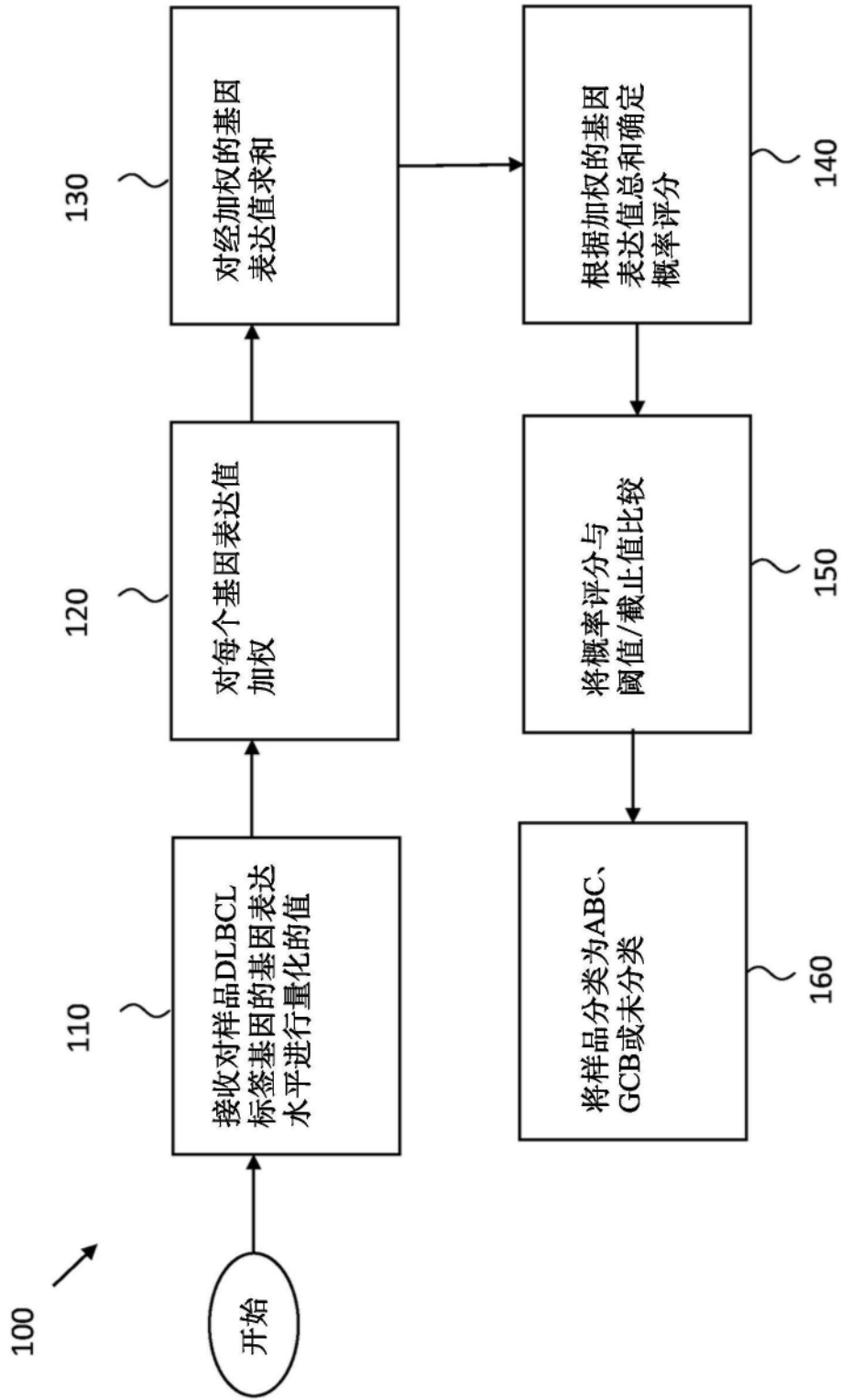


图1

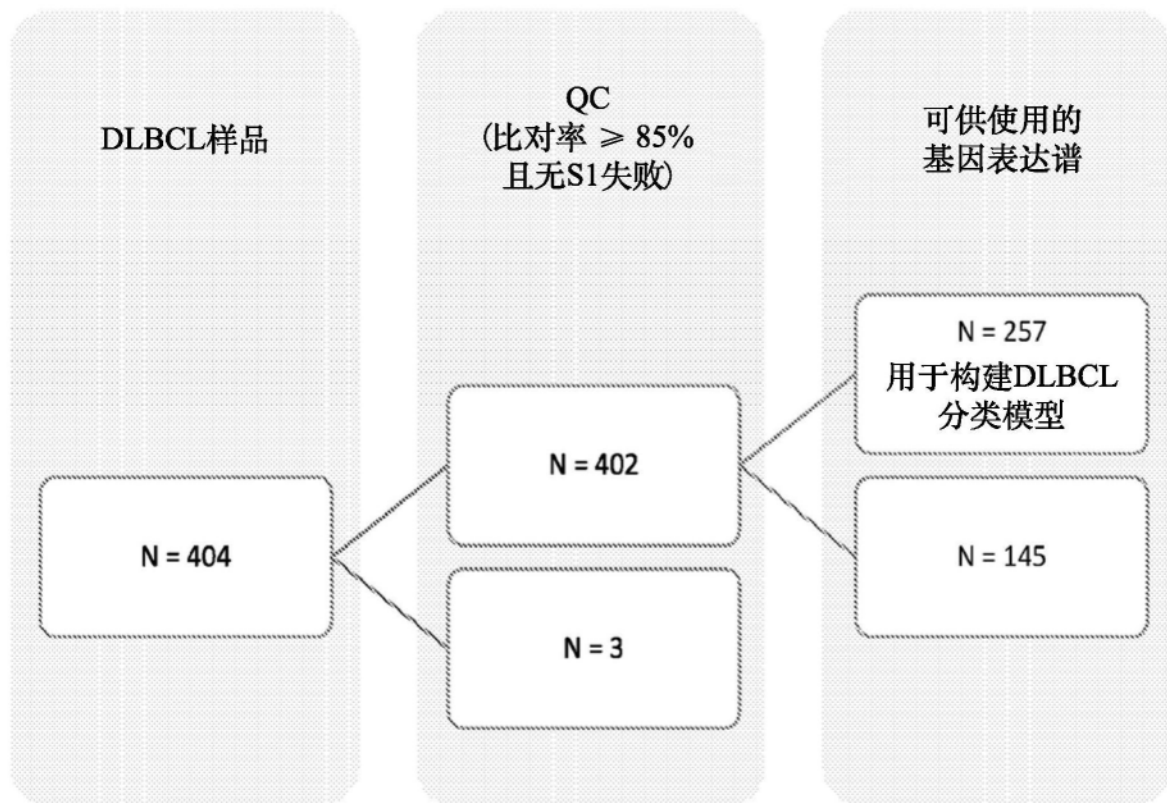


图2

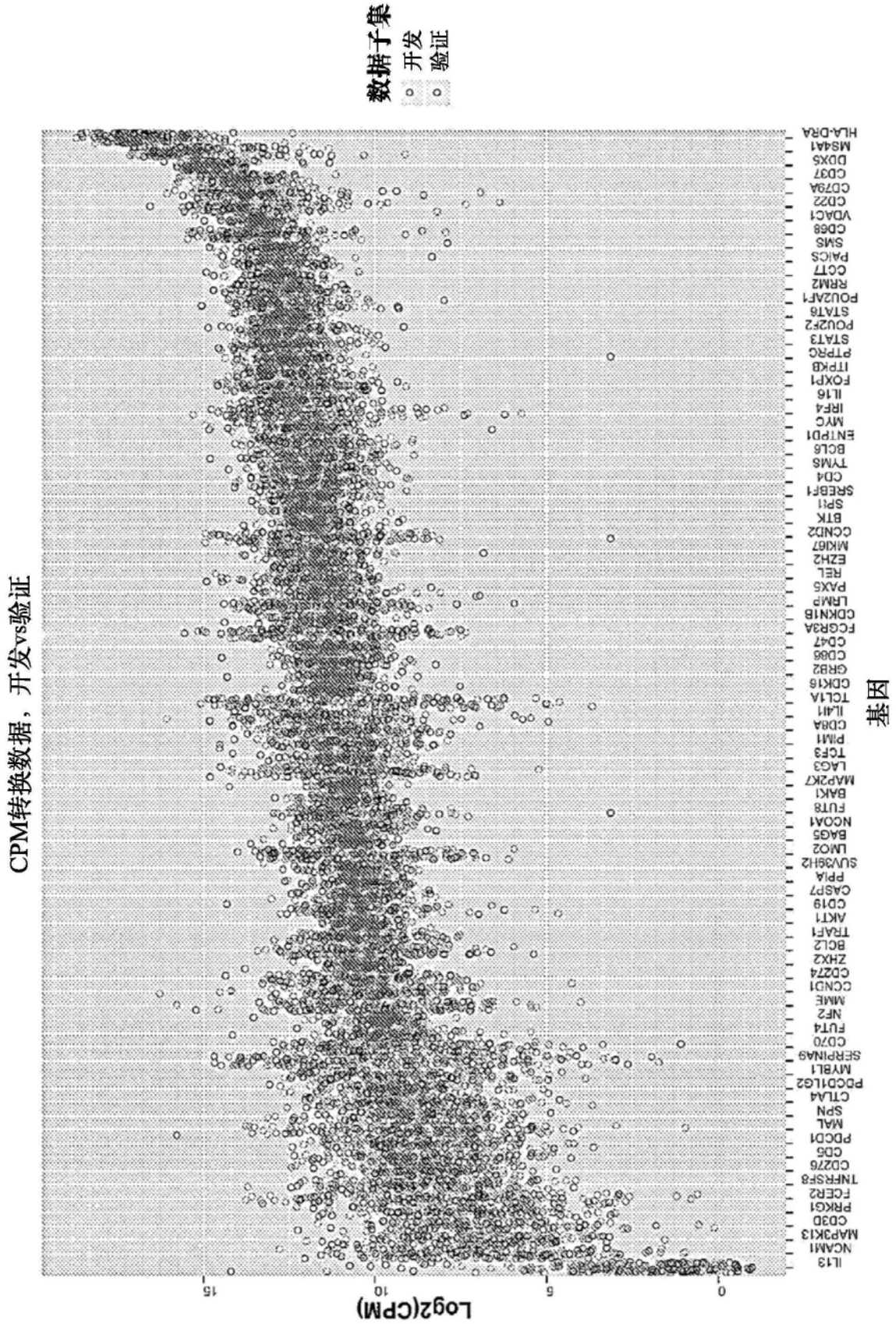


图3

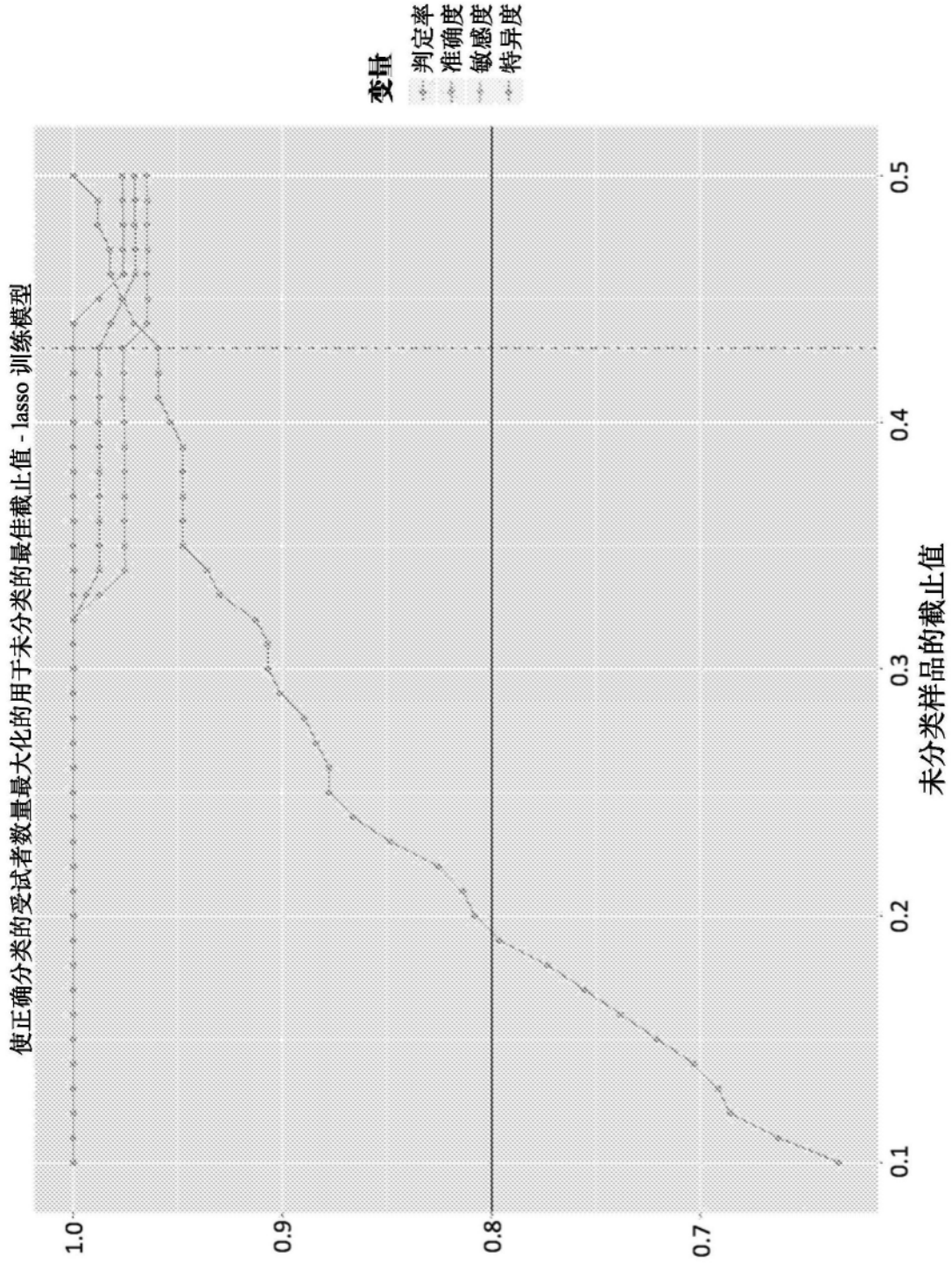


图4

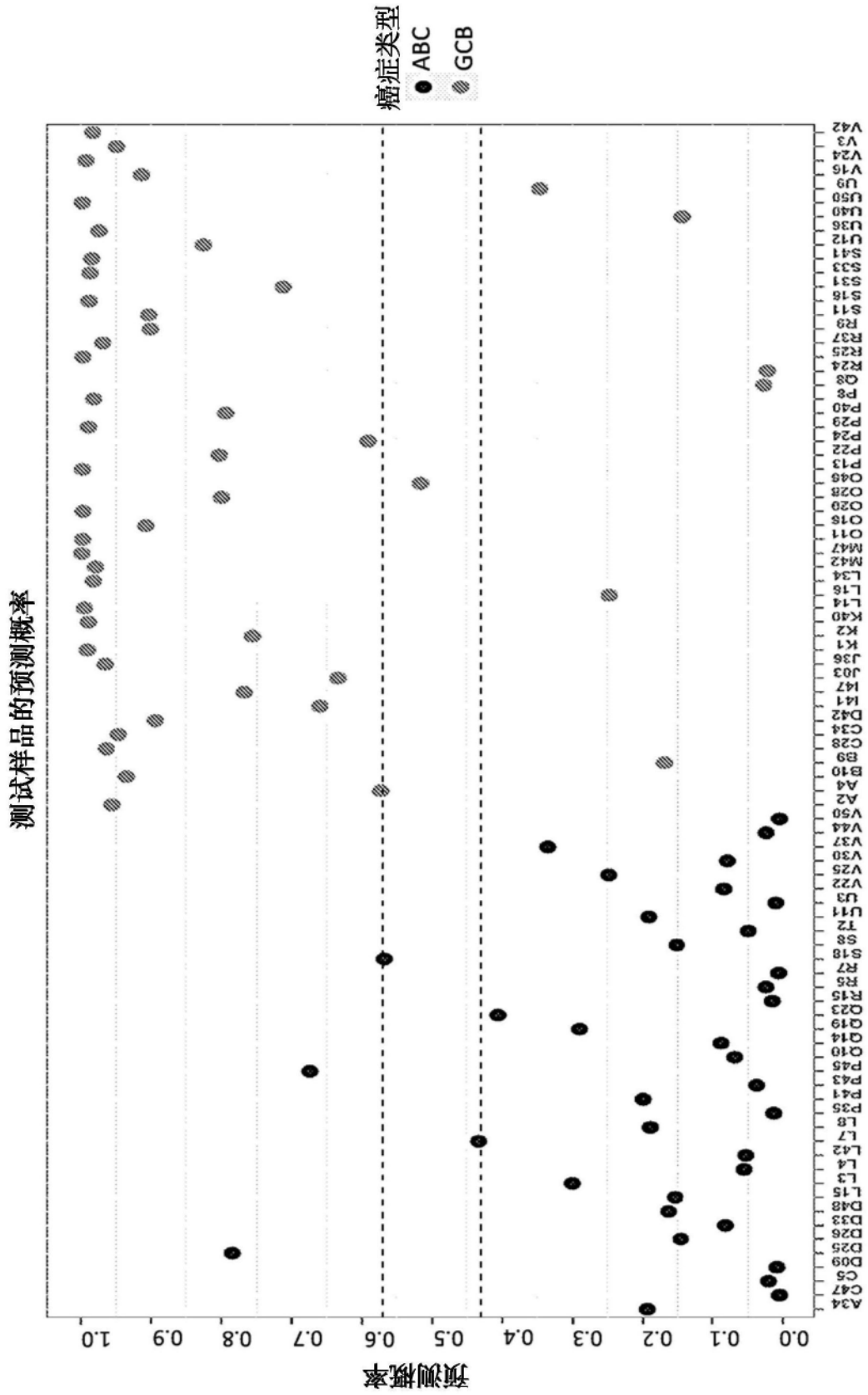


图5

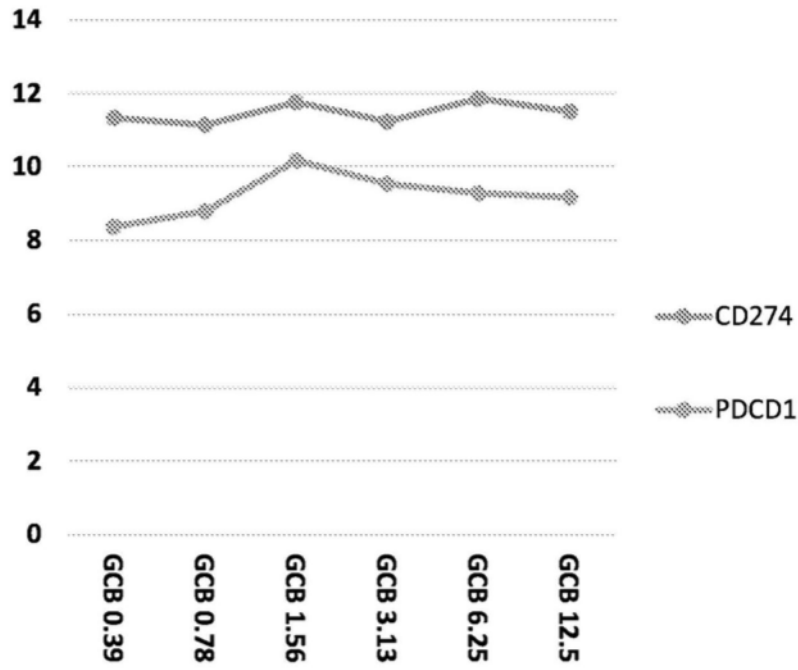


图6

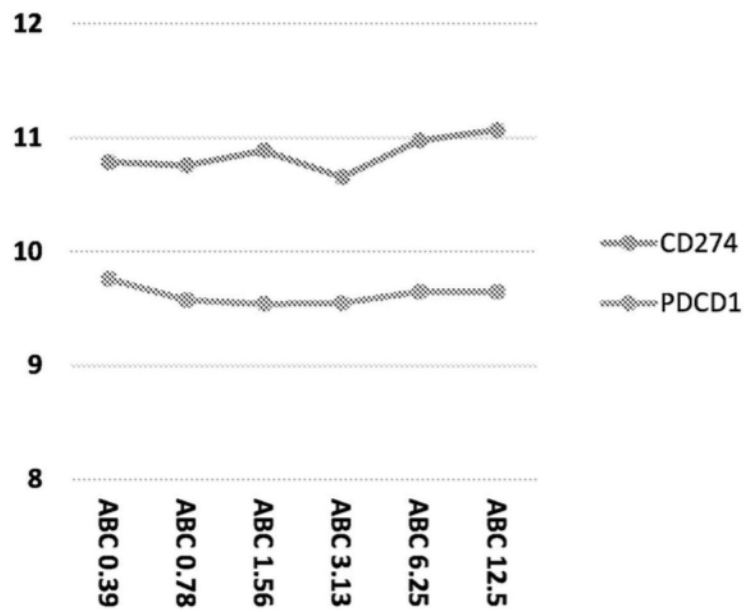


图7

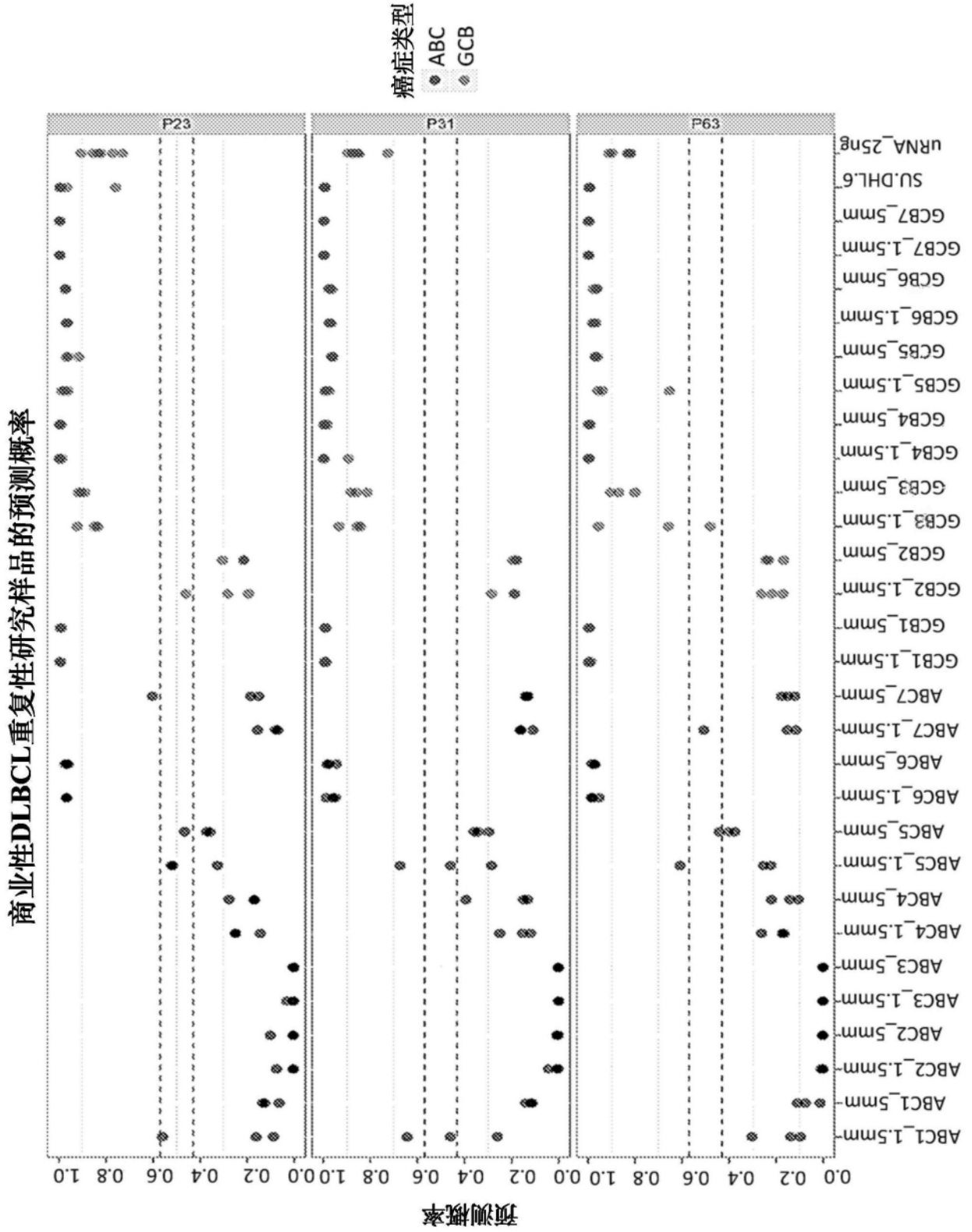


图8

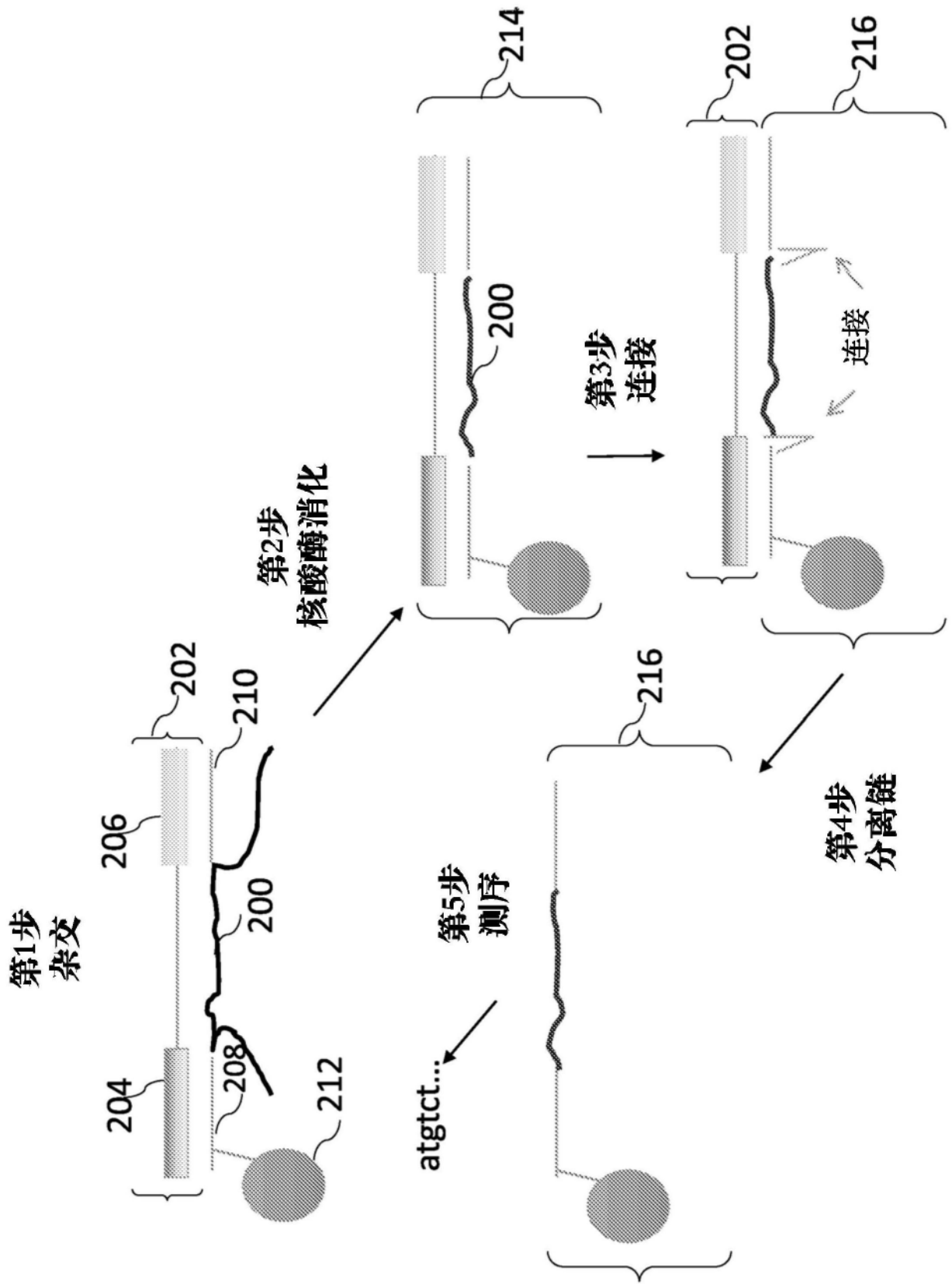


图9A

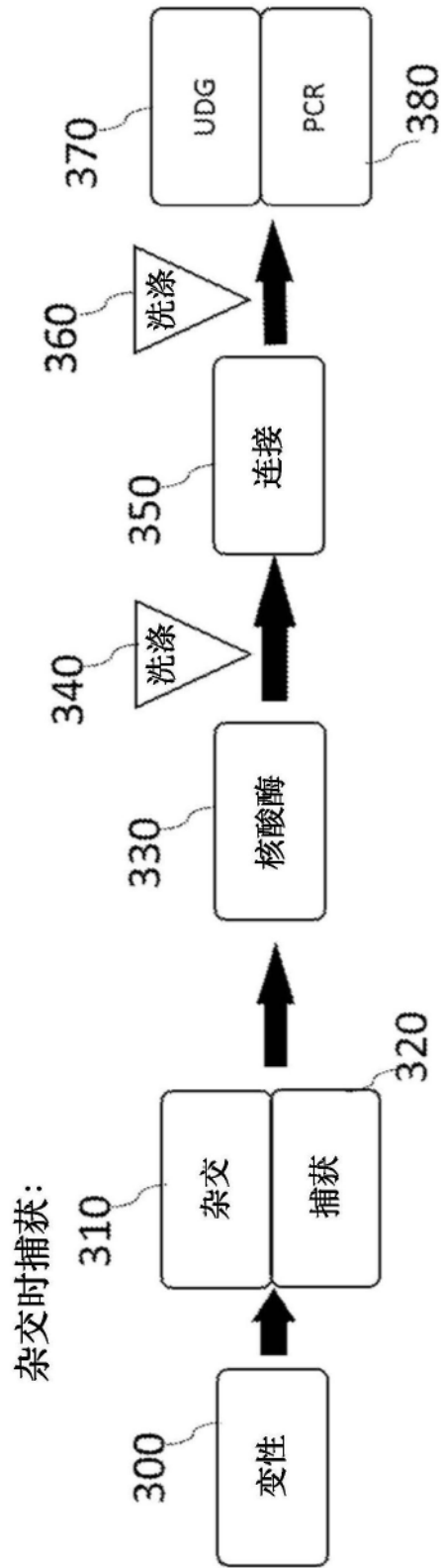


图9B

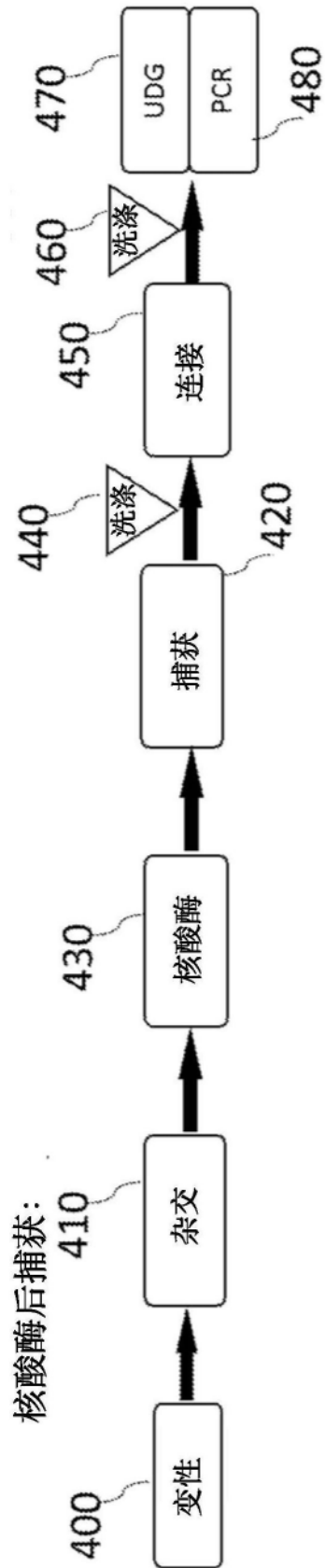


图9C

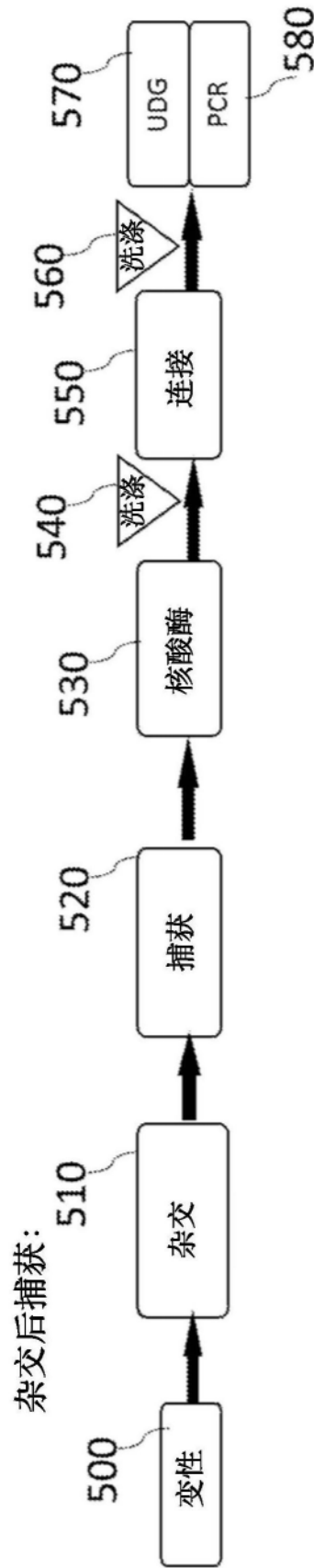


图9D