



(12)发明专利

(10)授权公告号 CN 104781668 B

(45)授权公告日 2018.07.24

(21)申请号 201380056857.9

尼尔·伍德伯里

(22)申请日 2013.08.29

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262

(65)同一申请的已公布的文献号  
申请公布号 CN 104781668 A

代理人 刘小立 郑霞

(43)申请公布日 2015.07.15

(51)Int.Cl.

(30)优先权数据

G01N 33/53(2006.01)

61/694,598 2012.08.29 US

G01N 33/00(2006.01)

(85)PCT国际申请进入国家阶段日  
2015.04.29

G01N 33/567(2006.01)

A61K 38/00(2006.01)

G07K 1/00(2006.01)

(86)PCT国际申请的申请数据  
PCT/US2013/057373 2013.08.29

(56)对比文件

CN 1438324 A,2003.08.27,全文.

(87)PCT国际申请的公布数据  
W02014/036312 EN 2014.03.06

WO 2007/068240 A2,2007.06.21,全文.

US 2008/0188618 A1,2008.08.07,全文.

(73)专利权人 亚利桑那州评议委员会,亚利桑那州法人团体,代理和代表亚利桑那州立大学

WO 01/56691 A2,2001.08.09,全文.

US 2012/0190574 A1,2012.07.26,全文.

CN 102099372 A,2011.06.15,全文.

地址 美国亚利桑那州

WO 2012/055069 A1,2012.05.03,全文.

审查员 贾静

(72)发明人 斯蒂芬·阿尔伯特·约翰斯顿  
菲利普·斯塔福德

权利要求书3页 说明书46页

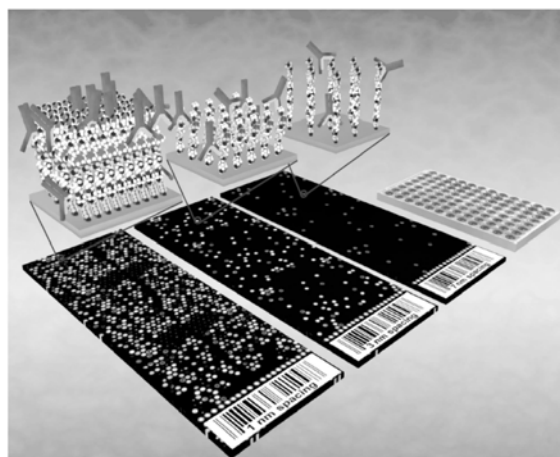
序列表13页 附图25页

(54)发明名称

免疫特征分析:通向早期诊断和健康监测的途径

(57)摘要

健康是一种复杂的状态,其代表了几乎所有人类活动和互动的不断变化的结果。本发明提供了用于健康监测、诊断、治疗和预防性保健的有效方法和阵列。本发明监测来自受试者的广泛的标识分子,如循环抗体,并且本发明评估这些分子与肽阵列的结合的模式。用本发明的方法表征此类分子与肽阵列的结合模式提供了对受试者健康状态的稳健测量。



1. 一种检测少量样品与阵列中的多个肽的结合模式的试剂盒,该试剂盒包括:  
装置,其中所述装置为肽阵列,其中所述肽阵列包含能够脱靶结合复杂生物样品中的至少一种抗体的不同的肽,其中所述阵列包含多个固定至固体支持体上不同位置的、原位合成的具有可变长度的聚合物,其中聚合物的原位合成包括以下步骤:
  - a. 将第一单体添加至所述固体支持体上预定部分的位置;
  - b. 将第二单体添加至所述固体支持体上预定部分的位置,其中用于所述第二单体的预定部分的位置包括含有第一单体的位置和不含单体的位置;
  - c. 将第三单体添加至所述固体支持体上预定部分的位置,其中用于所述第三单体的预定部分的位置包括含有第一和第二单体的位置、仅含有第二单体的位置以及不含单体的位置;以及
  - d. 用限定的一组单体重复步骤a-c,直到所述聚合物达到所期望的平均长度并且所述预定部分的总和总计至少为100%。
2. 根据权利要求1所述的试剂盒,其中所述肽阵列上的不同的肽为8至35个残基的长度。
3. 根据权利要求1所述的试剂盒,其中所述肽阵列上的不同的肽为15至25个残基的长度。
4. 根据权利要求1所述的试剂盒,其中所述肽阵列上的不同的肽具有范围为2-4nm的平均间距。
5. 根据权利要求1所述的试剂盒,其中所述肽阵列上的不同的肽具有范围为3-6nm的平均间距。
6. 根据权利要求1所述的试剂盒,其中所述不同的肽以约 $10^3\text{M}^{-1}$ 的缔合常数与分子结合。
7. 根据权利要求1所述的试剂盒,其中所述不同的肽以 $10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。
8. 根据权利要求1所述的试剂盒,其中所述不同的肽以 $2 \times 10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。
9. 根据权利要求1所述的试剂盒,其中所述不同的肽以 $10^4$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。
10. 根据权利要求1所述的试剂盒,其中所述不同的肽包含肽模拟物。
11. 根据权利要求1所述的试剂盒,其中所述不同的肽具有随机氨基酸序列。
12. 根据权利要求1所述的试剂盒,其中所述不同的肽包含非天然氨基酸。
13. 一种阵列,其包含多个固定至固体支持体上不同位置的、原位合成的具有可变长度的聚合物,其中聚合物的原位合成包括以下步骤:
  - a. 将第一单体添加至所述固体支持体上预定部分的位置;
  - b. 将第二单体添加至所述固体支持体上预定部分的位置,其中用于所述第二单体的预定部分的位置包括含有第一单体的位置和不含单体的位置;
  - c. 将第三单体添加至所述固体支持体上预定部分的位置,其中用于所述第三单体的预定部分的位置包括含有第一和第二单体的位置、仅含有第二单体的位置以及不含单体的位置;以及

d. 用限定的一组单体重复步骤a-c,直到所述聚合物达到所期望的平均长度并且所述预定部分的总和总计至少为100%。

14. 根据权利要求13所述的阵列,其中所述阵列是伪随机阵列。

15. 根据权利要求13所述的阵列,其中所述阵列是随机阵列。

16. 根据权利要求13所述的阵列,其中所述单体选自氨基酸、核酸和肽核酸。

17. 根据权利要求13所述的阵列,其中所述限定的单体组中的单体出现一次或多次。

18. 根据权利要求13所述的阵列,其中所述聚合物具有至少10个残基的平均长度。

19. 根据权利要求13所述的阵列,其中所述聚合物具有至少12个残基的平均长度。

20. 根据权利要求13所述的阵列,其中所述聚合物具有不少于5个残基的平均长度。

21. 根据权利要求13所述的阵列,其中至少5%的聚合物具有至少12个残基的长度。

22. 根据权利要求13所述的阵列,其中所述聚合物与样品中的成分结合。

23. 根据权利要求13所述的阵列,其中所述预定部分的总和总计为100%。

24. 根据权利要求13所述的阵列,其中所述预定部分的总和大于100%。

25. 根据权利要求13所述的阵列,其中所述聚合物的数目大于3,000。

26. 根据权利要求13所述的阵列,其中所述聚合物的数目大于10,000。

27. 根据权利要求13所述的阵列,其中所述聚合物的数目大于100,000。

28. 根据权利要求13所述的阵列,其中所述聚合物的数目大于330,000。

29. 一种制造阵列的方法,该阵列包含多个固定至固体支持体上不同位置的、原位合成的具有可变长度的聚合物,该方法包括以下步骤:

a. 提供基底作为在其上合成聚合物的固体支持体;

b. 将第一单体添加至所述固体支持体上预定部分的位置;

c. 将第二单体添加至所述固体支持体上预定部分的位置,其中用于所述第二单体的预定部分的位置包括含有第一单体的位置和不含单体的位置;

d. 将第三单体添加至所述固体支持体上预定部分的位置,其中用于所述第三单体的预定部分的位置包括含有第一和第二单体的位置、仅含有第二单体的位置以及不含单体的位置;以及

e. 用限定的一组单体重复步骤b-d,直到所述聚合物达到所期望的平均长度并且所述预定部分的总和总计至少为100%。

30. 根据权利要求29所述的方法,其中所述阵列是伪随机阵列。

31. 根据权利要求29所述的方法,其中所述阵列是随机阵列。

32. 根据权利要求29所述的方法,其中所述单体选自氨基酸、核酸和肽核酸。

33. 根据权利要求29所述的方法,其中所述限定的单体组中的单体出现一次或多次。

34. 根据权利要求29所述的方法,其中所述聚合物具有至少10个残基的平均长度。

35. 根据权利要求29所述的方法,其中所述聚合物具有至少12个残基的平均长度。

36. 根据权利要求29所述的方法,其中所述聚合物具有不少于5个残基的平均长度。

37. 根据权利要求29所述的方法,其中至少5%的聚合物具有至少12个残基的长度。

38. 根据权利要求29所述的方法,其中所述聚合物与样品中的成分结合。

39. 根据权利要求29所述的方法,其中所述预定部分的总和总计为100%。
40. 根据权利要求29所述的方法,其中所述预定部分的总和大于100%。
41. 根据权利要求29所述的方法,其中所述聚合物的数目大于3,000。
42. 根据权利要求29所述的方法,其中所述聚合物的数目大于10,000。
43. 根据权利要求29所述的方法,其中所述聚合物的数目大于100,000。
44. 根据权利要求29所述的方法,其中所述聚合物的数目大于330,000。

## 免疫特征分析:通向早期诊断和健康监测的途径

### [0001] 交叉引用

[0002] 本申请要求2012年8月29日提交的、名称为“免疫特征分析:通向早期诊断和健康监测的途径”的美国临时申请号61/694,598的权益,该临时申请通过引用整体并入本文。

[0003] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同特别且单独地指出每个单独的出版物、专利或专利申请通过引用而并入。

### 背景技术

[0004] 监测个体的健康是一个巨大的挑战。状况的早期检测可能对疾病的结果具有显著影响,然而,对于大多数状况,并不存在可在主要症状出现前检测出疾病的单一测试。许多研究小组曾尝试开发可以诊断特定状况的测试;然而此类测试仅限于特定疾病或诊断。此外,对目前可用的诊断测试而言,在一段时间内监测健康耗费太多的成本和时间。

### 发明内容

[0005] 本文公开了用于监测受试者的健康的方法、阵列和试剂盒。在本文公开的实施方案中,本发明提供了一种快速、稳健且可再现的健康监测方法,从而允许在一段时间内监测个体的健康。在一些实施方案中,该方法包括:a)使复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述生物样品中的至少一种抗体的不同的肽;b)测定所述抗体与所述肽阵列中的多个不同的肽的脱靶结合以形成免疫特征(immunosignature);以及c)将所述免疫特征与健康状态相关联。

[0006] 在一些实施方案中,本发明提供了一种提供治疗的方法,该方法包括:a)接收来自受试者的复杂生物样品;b)使所述复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述生物样品中的至少一种抗体的不同的肽;c)测定所述抗体与多个不同的肽的脱靶结合以形成免疫特征;d)将所述免疫特征与状况相关联;以及e)提供针对所述状况的治疗。

[0007] 在一些实施方案中,本发明提供了一种预防状况的方法,该方法包括:a)提供来自受试者的复杂生物样品;b)使所述复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述复杂生物样品中的至少一种抗体的不同的肽;c)测定所述复杂生物样品与多个不同的肽的脱靶结合以形成免疫特征;d)将所述免疫特征与状况相关联;以及e)接受针对所述状况的处理。

[0008] 在一些实施方案中,本发明提供了一种诊断方法,该方法包括:a)接收来自受试者的复杂生物样品;b)使所述复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述生物样品中的至少一种抗体的不同的肽;c)测定所述抗体与所述肽阵列中的多个不同的肽的脱靶结合以形成免疫特征;以及d)基于所述免疫特征诊断状况。

[0009] 在一些实施方案中,本发明提供了一种用于接收、记录(log)和稀释来自受试者的生物样品的系统。在一些实施方案中,用于接收、记录和稀释来自受试者的生物系统的系统是全自动化的。

[0010] 在一些实施方案中,一种免疫特征分析(immunosignaturing)系统包含自动化装置,该自动化装置由下列组件组成:1)用于接收、记录和稀释来自受试者的生物样品的自动化系统;2)用于自动化的免疫特征分析测试的隔室,所述免疫特征分析测试包括:a)将稀释的样品施加至肽阵列,b)孵育特定时间,c)洗涤并除去未结合的样品,d)施加第二抗体溶液特定时间,e)除去第二抗体,以及f)干燥并扫描所述阵列,以测定各个斑点的荧光;以及3)用检测器检测所述荧光。

[0011] 本文提供了用于产生可与本文所述的免疫特征测试结合使用的新型阵列的方法和装置。在一些实施方案中,制造所述阵列以减少在阵列上生成杂聚物所必需的图案化步骤的数目。在其他实施方案中,本文公开的方法和装置采用新型图案化算法来同时添加多个单体。在一些实施方案中,本文公开的算法可显著减少合成大阵列所需的图案化步骤的数目,从而导致更低的成本和更短的制造时间。

[0012] 在一些实施方案中,本文公开的方法和装置提供了一种阵列,该阵列包含多个固定至固体支持体上不同位置的、原位合成的具有可变长度的聚合物,其中聚合物的原位合成包括以下步骤:将第一单体添加至所述固体支持体上预定部分的位置;将第二单体添加至所述固体支持体上预定部分的位置,其中用于第二单体的预定部分的位置包括含有第一单体的位置和不含单体的位置;将第三单体添加至所述固体支持体上预定部分的位置,其中用于第二单体的预定部分的位置包括含有第一和第二单体的位置、含有第二单体的位置以及不含单体的位置;以及用限定的一组单体重复步骤a-c,直到所述聚合物达到所期望的平均长度并且所述部分的总和总计至少为100%。

[0013] 在其他实施方案中,本文公开的方法和装置还提供了一种制造阵列的方法,该阵列包含多个固定至固体支持体上不同位置的、原位合成的具有可变长度的聚合物,该方法包括以下步骤:提供基底作为在其上合成聚合物的固体支持体;将第一单体添加至所述固体支持体上预定部分的位置;将第二单体添加至所述固体支持体上预定部分的位置,其中用于第二单体的预定部分的位置包括含有第一单体的位置和不含单体的位置;将第三单体添加至所述固体支持体上预定部分的位置,其中用于第二单体的预定部分的位置包括含有第一和第二单体的位置、含有第二单体的位置以及不含单体的位置;以及用限定的一组单体重复步骤b-d,直到所述聚合物达到所期望的平均长度并且所述部分的总和总计至少为100%。

[0014] 在另外的其他实施方案中,本文公开的方法和装置提供了一种使用本文所述的阵列来监测受试者的健康状态的方法,该方法包括以下步骤:从所述受试者采集生物样品;将所述生物样品与所述阵列杂交;确定所述样品中与所述阵列杂交的成分;评估杂交程度;以及确定所述受试者的健康状态。所公开的阵列可用于产生如本文所述的免疫特征,而且还可用于使用微芯片阵列确定复杂生物样品中靶标的结合活性的其他诊断和治疗测试。

[0015] 在一些实施方案中,本发明提供了一种试剂盒。试剂盒可包含用于从受试者抽取少量血液的手指刺破装置和用于收集血液样品的接收表面。在一些实施方案中,该试剂盒包含关于其使用的书面说明。

## 附图说明

[0016] 图1为测试#1中的组间和组内相对差异的视觉表示。对120个肽和120个患者样品

中的每一个的值作图,蓝色指示低结合而红色指示高结合。使用欧几里德距离 (Euclidean distance) 作为相似性度量的分级聚类用来对肽 (Y轴) 和患者 (X轴) 进行聚类 (cluster)。最左边的层级就是基于这种聚类。

[0017] 图2示出了来自测试#1的样品的热图。A图示出了使用120个选定特征 (feature) 的训练数据集的热图。B图示出了使用相同的120个肽进行聚类的未混合的测试数据。

[0018] 图3示出了来自测试#2的样品的热图。1516个样品 (X轴) 与255个预测肽 (Y轴) 中的每一个的值一起示出。每种疾病与患者总数 (在括号内注明) 一起列出。

[0019] 图4为测试#1的受试者工作特征 (ROC) 曲线的图形表示。针对来自测试#1的测试数据的各个疾病组群 (cohort), 计算灵敏度和特异性。对于每个分类算法, 针对每种疾病绘制单独的ROC曲线并计算曲线下面积 (AUC) 值。SVM的AUC以灰色显示。A图为针对乳腺癌的SVM图的特异性/灵敏度AUC的图形表示。B图为针对脑癌的SVM图的特异性/灵敏度AUC的图形表示。C图为针对食管癌的SVM图的特异性/灵敏度AUC的图形表示。D图为针对多发性骨髓瘤的SVM图的特异性/灵敏度AUC的图形表示。E图为针对健康对照的SVM图的特异性/灵敏度AUC的图形表示。F图为针对胰腺癌的SVM图的特异性/灵敏度AUC的图形表示。

[0020] 图5为测试#1的受试者工作特征 (ROC) 曲线的图形表示。PCA的曲线下面积 (AUC) 以灰色显示。A图为针对乳腺癌的PCA图的特异性/灵敏度AUC的图形表示。B图为针对脑癌的PCA图的特异性/灵敏度AUC的图形表示。C图为针对食管癌的PCA图的特异性/灵敏度AUC的图形表示。D图为针对多发性骨髓瘤的PCA图的特异性/灵敏度AUC的图形表示。E图为针对健康对照的PCA图的特异性/灵敏度AUC的图形表示。F图为针对胰腺癌的PCA图的特异性/灵敏度AUC的图形表示。

[0021] 图6为测试#1的受试者工作特征 (ROC) 曲线的图形表示。NB的曲线下面积 (AUC) 以灰色显示。A图为针对乳腺癌的NB图的特异性/灵敏度AUC的图形表示。B图为针对脑癌的NB图的特异性/灵敏度AUC的图形表示。C图为针对食管癌的NB图的特异性/灵敏度AUC的图形表示。D图为针对多发性骨髓瘤的NB图的特异性/灵敏度AUC的图形表示。E图为针对健康对照的NB图的特异性/灵敏度AUC的图形表示。F图为针对胰腺癌的NB图的特异性/灵敏度AUC的图形表示。

[0022] 图7为测试#1的受试者工作特征 (ROC) 曲线的图形表示。LDA的曲线下面积 (AUC) 以灰色显示。A图为针对乳腺癌的LDA图的特异性/灵敏度AUC的图形表示。B图为针对脑癌的LDA图的特异性/灵敏度AUC的图形表示。C图为针对食管癌的LDA图的特异性/灵敏度AUC的图形表示。D图为针对多发性骨髓瘤的LDA图的特异性/灵敏度AUC的图形表示。E图为针对健康对照的LDA图的特异性/灵敏度AUC的图形表示。F图为针对胰腺癌的LDA图的特异性/灵敏度AUC的图形表示。

[0023] 图8为测试#1的受试者工作特征 (ROC) 曲线的图形表示。k-NN的曲线下面积 (AUC) 以灰色显示。A图为针对乳腺癌的k-NN图的特异性/灵敏度AUC的图形表示。B图为针对脑癌的k-NN图的特异性/灵敏度AUC的图形表示。C图为针对食管癌的k-NN图的特异性/灵敏度AUC的图形表示。D图为针对多发性骨髓瘤的k-NN图的特异性/灵敏度AUC的图形表示。E图为针对健康对照的k-NN图的特异性/灵敏度AUC的图形表示。F图为针对胰腺癌的k-NN图的特异性/灵敏度AUC的图形表示。

[0024] 图9为四个分类器 (classifier) 的图形表示。A图为PCA的图形表示, 对前两个主分

量作图。B图为LDA的图形表示，X和Y轴描绘了前两个线性判别量 (discriminant)。C图为NB的图形表示，对预测变量作图。D图为k-NN的图形表示，对组间距离 (groupwise distance) 作图。

[0025] 图10为来自测试#2的255个分类器肽中的3个分类器肽的线型图。该强度谱在X轴上显示了个体，疾病用空格隔开，并且在Y轴上显示了各个肽的 $\log_{10}$ 强度。A图示出了对疾病6和9为高而对所有其他疾病为低的肽的线型图。B图示出了对疾病11为高的肽的线型图。C图示出了对疾病1和部分疾病9为高的肽。

[0026] 图11为示出了可与本发明的示例实施方案结合使用的计算机系统的第一示例架构的框图。

[0027] 图12为示出了可与本发明的示例实施方案结合使用的计算机网络的示意图。

[0028] 图13为示出了可与本发明的示例实施方案结合使用的计算机系统的第二示例架构的框图。

[0029] 图14示出了具有不同肽密度的本发明的示例性阵列。

[0030] 图15为示出了多名受试者在接受流感疫苗后一段时间内的免疫特征谱的热图。

[0031] 图16为示出了免疫特征分析结合模式的热图。A图示出了在1天的过程中来自同一受试者的不同生物样品的免疫特征。B图示出了A图的一部分的特写图。

[0032] 图17为示出了在数月内监测的1名受试者的免疫特征分析结合模式的热图。

[0033] 图18为示出了3名受试者在21天的时程中的免疫特征分析结合模式的热图。A图示出了当检测到IgM免疫球蛋白的结合时，具有约10,000个肽的肽微阵列的聚类。B图示出了当检测到IgM免疫球蛋白的结合时，具有50个人肽 (personal peptides) 的肽微阵列的聚类。C图示出了当检测到IgG免疫球蛋白的结合时，具有约10,000个肽的肽微阵列的聚类。D图示出了当检测到IgG免疫球蛋白的结合时，具有50个人肽的肽微阵列的聚类。

[0034] 图19为示出了采用免疫特征分析结合模式分析对两名受试者进行的30天健康监测分析的热图。

[0035] 图20为示出了在30天时程中的第17天接受流感疫苗的受试者的免疫特征分析结合模式的热图。在不同的时间范围 (time-frame) 上示出了该受试者对22个所选肽序列的免疫特征分析结合谱。

[0036] 图21为示出了在图20中表征的受试者在2013年3月5日被诊断为支气管炎的热图。

[0037] 图22为示出了在图20中表征的受试者在2011年12月11日被症状后诊断为流感的热图。

[0038] 图23为示出了接受肝炎疫苗治疗并在此后3个月接受第一次加强治疗的受试者的免疫特征分析结合模式的热图。

[0039] 图24示出了多种传染病的分类的汇总。A图为示出了登革热、西尼罗病毒 (WNV)、梅毒、乙型肝炎病毒 (HBV)、正常血液、溪谷热和丙型肝炎病毒的聚类免疫特征分析结合谱的热图。B图为PCA分类的图形表示。

[0040] 图25为本发明的免疫特征分析系统的组件的示意图。

[0041] 图26:A图示出了噬菌体展示文库。B图示出了肽微阵列。

[0042] 图27示出了作为图案化步骤数的函数的合成肽的平均长度。X轴为图案化循环的数目，而Y轴为平均肽长度。通过使用任意数目的图案化循环，本文公开的图案化算法将图

案化步骤减少了几乎2倍。

[0043] 图28示出了使用标准的逐层法(layer by layer approach)与新型图案化算法相比,通过以单体形式施加所有20种氨基酸而产生的结果。

[0044] 图29示出了免疫特征分析实施方案的结果,在该实施方案中用少于60个步骤即获得平均12个残基的长度。

[0045] 图30示出了使用16种不同的氨基酸由优化算法的70个步骤产生的分布。

[0046] 图31示出了在肽阵列产生后选择的肽的长度分布。

[0047] 图32为示出了3个、4个或5个氨基酸长的可能序列的分布的图。

[0048] 图33示出了对于所选择的肽文库,作为肽中的位置的函数的氨基酸组成。

[0049] 发明详述

[0050] “健康”是一种复杂的状态,其代表了几乎所有人类活动和交互的不断变化的结果。这使得难以定量地定义健康状态。必须系统性地测量数以千计的生化和物理属性。健康监测中的一大挑战是受试者对各种刺激的响应的复杂性。大多数生物每天暴露于多种不同的刺激,然而一些生物具有能够响应于此类刺激并防止引发或形成疾病的生物结构和过程的系统。为了适当地行使功能,此类系统必须检测很多种刺激,如病毒或寄生虫的存在,并在体内引发针对这些物质、异常细胞和/或组织的响应。

[0051] 健康监测中的一个必然的挑战是受试者对复杂刺激的响应的复杂性。例如,由自身体内的病变细胞产生的生理响应可以不同于对感染的生理响应。然而,检测、处理、识别和作用于例如感染或癌症的早期体征的能力可对受试者的健康具有显著影响。如果在肿瘤细胞有时间繁殖、抑制免疫系统、形成转移性集落并造成组织损伤前诊断出癌症,则人们可有望更有利地对治疗作出响应。

[0052] 类似地,如果在感染后不久检测出受试者体内存在病原体,则可在宿主炎症阻止侵入物进入之前以及在病原体负荷变得在免疫上具有压倒性优势之前施用抗微生物剂。如果早期检测到诸如狼疮的自身免疫疾病,而自身抗体水平低,则用以减弱免疫爆发的治疗可能有效得多。幸运的是,免疫系统持续监测受试者的健康状态。然而,用于健康监测和早期检测的稳健、可靠且有效的方法仍然是一个尚未得到满足的需求。

[0053] 免疫特征分析是微阵列和噬菌体技术的结合体,其展示了体液免疫应答的复杂性并将其转化成机器可读的、量化的形式。免疫特征分析在早期准确地检测健康状态的甚至微小的扰动。对于抗体组库(repertoires)的这些综合性测量提供了用于快速、廉价及早期诊断任何患病状态的手段;最终,对免疫特征的持续监测可提供在症状出现之前检测危险疾病状态的手段。

[0054] 因此,本文公开的发明提供了用于健康监测、诊断、治疗和预防性保健的灵敏、稳健、有效且可靠的方法。本文公开的实施方案通过提供将受试者的体液状态与状况相关联的大规模平台而解决了多种不同的状况和健康状态缺乏关联及替代标志物的问题。

[0055] 生理系统的任何组件,不论是外来的还是自身的,均可充当状况或健康状态的阳性或阴性标志物。免疫系统是有机体内被设计用于检测很多种标志物(包括外来的和自身的物质)的生物结构和过程的生理系统。免疫系统可产生多种抗体,这些抗体可存在于个体的外周血样品中,并且可内源性地扩增至高浓度。抗体可以是丰富的,可具有高靶标亲和力,并可显示极为多样化的表位和结构柔性。

[0056] 免疫系统的组件,例如抗体,可以是非常稳健的,并且可充当受试者健康状态的合适的标志物。血液、血浆和/或血清中的抗体在经历加热、干燥和/或暴露于宽范围的pH值时可保持其完整性。血液、血浆和/或血清中的抗体在以干、冷冻或干燥形式经历长期贮存时可保持其完整性。例如,当抗体保持在干滤纸上并邮寄时,抗体可保持部分和/或完全的完整性。此类性质可使得大多数血液、血浆和/或血清样品成为供监测、诊断、预防和治疗状况的方法使用的生物标志物的潜在来源。

[0057] 本发明提供了用于将诸如血液、干血、血清、血浆、唾液样品、检查拭子(check swab)、活检物、组织、皮肤、毛发、脑脊液样品、粪便或尿液样品等生物样品与受试者的健康状态相关联的阵列和方法。在一些实施方案中,生物样品是与非天然肽序列的肽阵列接触的血液样品。在一些实施方案中,受试者可以,例如,使用“手指针刺(fingerstick)”或“手指刺破(fingerprick)”来抽取少量血液并将其添加到表面如滤纸或其他可吸收源,或者添加到小瓶或容器中并任选地干燥。可将例如从受试者血液的血滴中获得并放置在滤纸上的生物样品直接邮寄到本发明方法的提供者而无需处理该样品。可对受试者提供的生物样品进行浓缩或稀释。

[0058] 本发明的肽阵列可被构造为以高灵敏度检测少量生物样品与阵列中的多个肽的结合模式。在一些实施方案中,本发明提供了检测、处理、分析生物样品与多个肽的结合模式并将该结合模式与状况相关联的方法。在一些实施方案中,本发明产生了与受试者的健康状态相关联的“免疫特征”。

[0059] 免疫特征分析检测抗体应答并将其分割成可经数学方法解读的一组明晰的信号。来自采用本发明的阵列和方法获得的免疫特征的一组明晰的信号可提供用于诊断各种状况(包括癌症、炎症、感染和其他生理状况)的稳健及综合的方法。免疫特征分析不同于用于诊断各种状况的传统的单个蛋白质或遗传生物标志物,并且是后者的备选方案。来自采用本发明的阵列和方法获得的免疫特征的一组明晰的信号可用作预防性保健、健康监测、诊断的有效方法并可用作治疗方法。

#### [0060] 抗体生物标志物的多重检测

[0061] 被设计用于检测宿主产生的抗体(而非其他较不丰富的生物标志物)的诊断方法更有可能总够灵敏以检测罕见事件。由于在循环血液中已经存在这些标志物的非常多样化的来源,因此无需形成高亲和力、高特异性抗体的充足供应。在设计用于检测抗体的多重阵列中,将多组(panels)蛋白质或肽附接至固体支持体上,随后暴露于血液。

[0062] 蛋白质阵列作为能够在单一实验中同时检测大量参数的高容量方法而出现。蛋白质靶标提供了构象表位的来源以供抗体结合,尽管并不总是暴露线性表位。Invitrogen生产了一种更全面的蛋白质微阵列,其包含在单一载玻片上排列的约9000种不同的杆状病毒产生的人蛋白质。这些蛋白质阵列的大规模化潜能受到每个载玻片的高成本、可扩大性的缺乏以及重组蛋白质生产、纯化和稳定性的不一致性的限制。使用体外合成的蛋白质已改善了蛋白质生产的通量和成功率,但在排列的量、稳定性、翻译后修饰的不一致性方面以及对膜(表面)、多聚体和大蛋白质的偏性方面仍然存在问题。这两种方法均局限于检测自身抗体,除非特异性地合成已知的突变体或由病原体衍生的候选蛋白质。对病变细胞的生化分级分离使得能够探查针对修饰的和突变的抗原的抗体,但这是一个实质上更复杂的过程(Hanash,S. (2003) Disease proteomics. Nature 422,226-232)。

[0063] 与蛋白质相比之下,肽能够化学合成,从而可大量获得高度可再现且纯的产品,且该产品具有长贮存期。生物学相关修饰或检测分子的衔接是简单的,并且非天然的设计也是可能的 (Reddy, M.M. 等人, Identification of Candidate IgG Biomarkers for Alzheimer's Disease via Combinatorial Library Screening. *Cell* 144, 132-142)。

[0064] 肽在溶液中以相似的方式呈现,即使与固体支持体结合时也是如此;因此,不论批次间的生产差异如何,均可针对高度一致的结构来筛选抗体相互作用。肽微阵列可被使用的时间远远长于蛋白质微阵列 (Panicker, R.C. 等人, (2004) Recent advances in peptide-based microarray technologies. *Comb Chem High Throughput Screen* 7, 547-556), 并且已经用于多种应用。酶 (Fu, J. 等人, (2010) Exploring peptide space for enzyme modulators. *J Am Chem Soc* 132, 6419-6424; 以及 Fu, J. 等人, (2011) Peptide-modified surfaces for enzyme immobilization. *PLoS One* 6, e18692)、蛋白质 (Diehnelt, C.W. 等人, Discovery of high-affinity protein binding ligands-backwards. *PLoS One* 5, e10728; Greving, M.P. 等人, High-throughput screening in two dimensions: binding intensity and off-rate on a peptide microarray. *Anal Biochem* 402, 93-95; Greving, M.P. 等人, Thermodynamic additivity of sequence variations: an algorithm for creating high affinity peptides without large libraries or structural information. *PLoS One* 5, e15432; Gupta, N. 等人, Engineering a synthetic ligand for tumor necrosis factor- $\alpha$ . *Bioconjug Chem* 22, 1473-1478)、DNA 和小分子 (Boltz, K.W. 等人, (2009) Peptide microarrays for carbohydrate recognition. *Analyst* 134, 650-652; Foong, Y.M. 等人, (2012) Current advances in peptide and small molecule microarray technologies. *Curr Opin Chem Biol* 16, 234-242; Morales Betanzos, C. 等人, (2009) Bacterial glycoprofiling by using random sequence peptide microarrays. *Chembiochem* 10, 877-888)、全细胞 (Falsey, J.R. 等人, (2001) Peptide and small molecule microarray for high throughput cell adhesion and functional assays. *Bioconjug Chem* 12, 346-353) 和抗体 (Cerecedo, I. 等人, (2008) Mapping of the IgE and IgG4 sequential epitopes of milk allergens with a peptide microarray-based immunoassay. *J Allergy Clin Immunol* 122, 589-594; Cretich, M. 等人, (2009) Epitope mapping of human chromogranin A by peptide microarrays. *Methods Mol Biol* 570, 221-232; Lin, J. 等人, (2009) Development of a novel peptide microarray for large-scale epitope mapping of food allergens. *J Allergy Clin Immunol* 124, 315-322, 322e311-313; Lorenz, P. 等人, (2009) Probing the epitope signatures of IgG antibodies in human serum from patients with autoimmune disease. *Methods Mol Biol* 524, 247-258; Perez-Gordo, M. 等人, (2012) Epitope mapping of Atlantic salmon major allergen by peptide microarray immunoassay. *Int. Arch Allergy Immunol* 157, 31-40; 以及 Shreffler, W.G. 等人, (2005) IgE and IgG4 epitope mapping by microarray immunoassay reveals the diversity of immune response to the peanut allergen, Ara h 2. *J Allergy Clin Immunol* 116, 893-899) 仅是可对其与肽的结合进行分析的生物分子的子集。一个经典的实例是表位作图: 可将跨越抗原的肽平铺以有效地破译单克隆抗

体的表位。通常,高特异性抗体将识别并结合其表位序列,而与抗原衍生的其他肽没有或具有很少的可测量的结合。利用这种方法,可以区分并表征针对相同抗原的不同的单克隆抗体。

[0065] 交叉反应性的相关性

[0066] 免疫系统已经进化为产生并扩增忽略自身蛋白质并以显著的强度结合非自身靶标的抗体。免疫系统已经进化为产生并扩增忽略自身蛋白质并以显著的强度结合非自身靶标的抗体。这些不一致的压力在分子水平上变得明显。典型的抗体识别约15个氨基酸的表位,其中约5个氨基酸主导结合能。这5个残基中任何一个的变化将极大地影响结合强度。

[0067] 其他表位位置的序列变化将改变结合区的空间构象并适度地影响总强度。因此,如果欲将结合强度最大化,则必须调整条件以允许高亲和力和低亲和力残基相互作用。这意味着严格性降低,因此允许表位序列的变体结合抗体。抗体具有包含许多重叠互补位的50个氨基酸的可变区(抗体的表位识别部分),从而进一步有助于交叉反应性的潜能(Mohan,S.等人,(2009) Association energetics of cross-reactive and specific antibodies. *Biochemistry* 48,1390-1398;Thorpe,I.F.和Brooks,C.L.,第3版(2007) *Molecular evolution of affinity and flexibility in the immune system*. *Proceedings of the National Academy of Sciences of the United States of America* 104,8821-8826;以及Zhou,Z.H.等人,(2007) Properties and function of polyreactive antibodies and polyreactive antigen-binding B cells. *J Autoimmun.* 29,219-228. Epub 2007 Sep 2020)。

[0068] 这些互补位中的每一个均由约15个氨基酸组成,使得互补位和表位为大小相似的序列段,其限定了形状和电荷的互补区。互补位可结合多于一个表位,并且单一表位可与多于一个互补位结合,每一对都表现出独特的结合性质。由于单一抗体携带多个互补位,因此抗体能以不同的强度结合不同的而潜在多样化的一组表位。这种交叉反应性以及特异性与亲和性的复杂相互影响是复杂的免疫系统的标志,该免疫系统协调对抗即时威胁的直接攻击以及对抗对将来可能的变体暴露的间接攻击。

[0069] 在体外,对特定线性表位具有特异性的抗体已显示出既结合序列相关肽,又结合序列非相关肽(Folgori,A.等人,(1994) A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *Embo J* 13,2236-2243)。通常表现出构象相关性的这些序列非相关肽被称为模拟表位,并且最初描述于早期噬菌体展示研究中(Folgori,A.等人,(1994) A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *Embo J* 13,2236-2243;Christian,R.B.等人,(1992) Simplified methods for construction,assessment and rapid screening of peptide libraries in bacteriophage. *Journal of Molecular Biology* 227,711-718;Liu,R.等人,(2003) Combinatorial peptide library methods for immunobiology research. *Experimental Hematology* 31,11-30;Wang,Y.等人,(1995) Detection of Mammary Tumor Virus ENV Gene-like Sequences in Human Breast Cancer. *Cancer Research* 55,5173-5179)。

[0070] 基于噬菌体的系统提供了构建和筛选比使用大多数其他系统所可能达到的具有

大得多的配体多样性的文库的机会。例如,针对特定单克隆抗体淘选展示肽的大随机序列文库。反复数轮选择常常导致鉴定出同源表位,但也会鉴定出几种非相关肽序列。随机肽多样性比生物序列多样性大几个数量级的事实意味着该肽将不对应于任何生物肽。所有的结合反应均依赖于非同源交叉反应性,这是抗体的固有性质。这意味着可以鉴定针对任何类别抗体的配体:自身抗原、修饰的抗原、突变的表位或非肽模拟表位。尽管在随机序列空间中进行筛选具有这些优势,但噬菌体展示技术受到使用噬菌体和细菌培养物进行反复数轮淘选(一种二元选择过程)和缺乏可扩大性的限制(Derda,R.等人,(2011) Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* 16,1776-1803; Szardenings,M. (2003) Phage display of random peptide libraries: applications, limits, and potential. *J Recept Signal Transduct Res* 23,307-349;53,54)。迄今为止,随机噬菌体文库尚未产生抗体生物标志物。

#### [0071] 免疫特征分析

[0072] 免疫特征分析是上述技术的综合。首先,并非在噬菌体上以生物学方式展示肽,而将合成的较长的肽以可寻址的有序的阵列连接至载玻片上是一种更为系统化的方法。虽然噬菌体文库可以有超过 $10^{10}$ 个单独的克隆,但微阵列已经从每个载玻片几千个斑点增加到数百万个斑点。成本、可靠性、精确度和分析速度使微阵列具有显著的优势。微阵列由于其低成本和可扩大性已证明其对于基因组学和蛋白质组学是无价的,并且商业化的阵列室和扫描仪已经存在多年。

[0073] 其次,使用抗体作为疾病的生物标志物利用了稳定且易于获得的分子以及免疫系统的多样性、监视和生物扩增的适宜性质。哺乳动物免疫系统的复杂性令人吃惊(Janeway, C.和Travers,J. (1997) *Immunobiology: The Immune System in Health and Disease*. Current Biology Limited),因此信息量也是如此。随着免疫学家探索免疫组,日渐达成以下共识:能够有 $>10^{10}$ 个不同的分子种类的抗体组库(Nobrega,A.等人,(1998) Functional diversity and clonal frequencies of reactivity in the available antibody repertoire. *European Journal of Immunology* 28,1204-1215)是个体健康状态的过去、现在乃至前驱扰动(prodromic perturbation)的动态数据库。

[0074] 第三,随机序列肽的使用能够使抗体组库的多样性被待筛选的无偏、综合配体文库所匹配。随机序列肽可用于噬菌体展示文库,但它们携带偏性,并且不是无序的、控制不良的形式。由于随机肽序列对生物空间没有限制且没有刻意的同源性,因此微阵列包含极少的但覆盖范围非常广的序列空间。可在相同的微阵列上筛选与任何疾病或有机体相对应的正常的、突变的、翻译后修饰的和模拟的表位。本领域中最近的出版物已使用了10,000个独特的随机序列20-mer肽来表征大量的疾病状态1、10(Brown,J.R.等人,(2011) Statistical methods for analyzing immunosignatures. *BMC Bioinformatics* 12,349; Hughes,A.K.等人,(2012) Immunosignaturing can detect products from molecular markers in brain cancer. *PLoS One* 7,e40201; Kroening,K.等人,(2012) Autoreactive antibodies raised by self derived de novo peptides can identify unrelated antigens on protein microarrays. Are autoantibodies really autoantibodies? *Exp Mol Pathol* 92,304-311; Kukreja,M.等人,(2012) Comparative study of classification algorithms for immunosignaturing data. *BMC Bioinformatics* 13,

139;Kukreja,M.等人,(2012) Immunosignaturing Microarrays Distinguish Antibody Profiles of Related Pancreatic Diseases. *Journal of Proteomics and Bioinformatics*;以及Legutki,J.B.等人,(2010) A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28,4529-4537)。

[0075] 由噬菌体展示获得的结果相对于由免疫特征分析微阵列获得的结果存在几个显著的差异。免疫特征分析探查阵列上所有的肽,并产生每个肽的结合值。噬菌体展示产生在限制性选择后留下的序列,并且通常仅鉴定共有序列。处理免疫特征分析微阵列需要数小时,而非数周。图14显示了这些技术之间的差别。在技术上,‘免疫特征’是指肽的统计学上显著的模式,每一个都具有可稳健地将一种疾病状态与其他疾病状态区分开的特定结合值。

[0076] 技术的这种整合可代表朝向普遍适用的早期诊断平台的目标的进展。有待解决的关键问题为:i) 免疫系统是否对传染病和慢性疾病均引起一致的疾病特异性体液应答,ii) 抗体是否足够早地响应于疾病的病因学从而是临床上有用的,以及iii) 该测试是否足够灵敏、提供足够多的信息、足够廉价以及可扩大,从而筛选大量的患者样品以用于可信的确定。如果这几项能够得到满足,则可以发现任何免疫相关疾病的免疫特征。而后可使用这些限定的反应性模式来早期及全面地诊断疾病。如果可以使这些检测广泛适用于群体,则免疫特征分析可以构成在个体以及流行病学水平上具有重要意义的长期健康监测系统的基础。我们提出在这方面很有前景的平台的特点。

[0077] 免疫特征分析是上述技术的综合。首先,并非在噬菌体上以生物学方式展示肽,而将合成的较长的肽以可寻址的有序阵列连接至载玻片上是一种更为系统化的方法。虽然噬菌体文库可以有超过 $10^{10}$ 个单独的克隆,但微阵列已经从每个载玻片几千个斑点增加到数百万个斑点。成本、可靠性、精确度和分析速度使微阵列具有显著的优势。微阵列由于其低成本和可扩大性已证明其对于基因组学和蛋白质组学是无价的,并且商业化的阵列室和扫描仪已经存在多年。其次,使用抗体作为疾病的生物标志物利用了稳定且易于获得的分子以及免疫系统的多样性、监视和生物扩增的适宜性质。

[0078] 哺乳动物免疫系统的复杂性令人吃惊,因此信息量也是如此。随着免疫学家探索免疫组,日渐达成以下共识:能够具有 $>10^{10}$ 个不同的分子种类的抗体组库是个体健康状态的过去、现在乃至前驱扰动的动态数据库。第三,随机序列肽的使用能够使抗体组库的多样性被待筛选的无偏、综合的配体文库所匹配。随机序列肽可用于噬菌体展示文库,但它们携带偏性,并且不是无序的、控制不良的形式。由于随机肽序列对生物空间没有限制且没有刻意的同源性,因此微阵列包含极少的但覆盖范围非常广的序列空间。

[0079] 可在相同的微阵列上筛选与任何疾病或有机体相对应的正常的、突变的、翻译后修饰的和模拟的表位。本领域的出版物已使用了10,000个独特的随机序列20-mer肽来表征大量的疾病状态。由噬菌体展示获得的结果与由免疫特征分析微阵列获得的结果存在几个显著的差异。免疫特征分析探查阵列上所有的肽,并产生每个肽的结合值。噬菌体展示产生在限制性选择后留下的序列,并且通常仅鉴定共有序列。

[0080] 处理免疫特征分析微阵列可能需要数小时而非数周。“免疫特征”是指肽的统计学上显著的模式,每一个均具有可稳健地将一种疾病状态与其他疾病状态区分开的特定结合

值。因此,本文公开的实施方案的一个方面是用复杂生物样品探查免疫特征阵列的相对较短的处理时间,其中探查和处理时间可能花费最多10分钟、最多20分钟、最多30分钟、最多45分钟、最多60分钟、最多90分钟、最多2小时、最多3小时、最多4小时或最多5小时。或者,探查和处理时间可花费不超过10分钟、不超过20分钟、不超过30分钟、不超过45分钟、不超过60分钟、不超过90分钟、不超过2小时、不超过3小时、不超过4小时或不超过5小时。

[0081] 技术的这种整合可代表朝向普遍适用的早期诊断平台的目标的进展。有待解决的关键问题为:i) 免疫系统是否对传染病和慢性疾病均引起一致的疾病特异性体液应答,ii) 抗体是否足够早地响应于疾病病因学从而是临床上有用的,以及iii) 该测试是否足够灵敏、提供足够多的信息、足够廉价以及可扩大,从而筛选大量的患者样品以用于可信的确定。如果这几项能够得到满足,则可以发现任何免疫相关疾病的免疫特征。

[0082] 而后可以使用这些限定的反应性模式来早期及全面地诊断疾病。如果可以使这些检测广泛适用于群体,则免疫特征分析可以构成在个体以及流行病学水平上具有重要意义的长期健康监测系统的基础。我们提出在这方面很有前景的平台的特点。

#### [0083] 免疫特征分析的独特特点

[0084] 除了抗体的互补位对配体的亲和力外,结合强度还能受血清中的抗体种类的浓度影响。与噬菌体展示不同,免疫特征分析可以定量地测量这些参数的结果,并且可在非常大的动态范围内进行这种测量(Legutki, J.B. 等人, (2010) A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28, 4529-4537; Stafford, P. 和 Johnston, S. (2011) Microarray technology displays the complexities of the humoral immune response. *Expert Rev Mol Diagn* 11, 5-8; Halperin, R.F. 等人, (2011) Exploring Antibody Recognition of Sequence Space through Random-Sequence Peptide Microarrays. *Molecular & Cellular Proteomics* 10)。

[0085] 科学家利用这种能力来检查高亲和力单克隆抗体与免疫特征分析微阵列的结合。他们发现单一单克隆抗体识别数百个随机序列,并且可以测量和比较这些独特结合反应的不同强度(Halperin, R.F. 等人, (2011) Exploring Antibody Recognition of Sequence Space through Random-Sequence Peptide Microarrays. *Molecular & Cellular Proteomics* 10)。令人好奇的是,这些脱靶模拟表位相互作用中有许多比同源表位具有更强的结合。虽然这些相互作用的相应溶液相结合较低,但构建免疫特征分析微阵列的方式增强了这些相互作用。抗体与免疫特征分析微阵列脱靶结合的这种免疫学现象对该技术极为重要。

[0086] 另一重要的观察结果是对低亲和力相互作用的检测而言,免疫特征分析比噬菌体展示或基于ELISA的测定具有更高的灵敏度(Stafford, P. 和 Johnston, S. (2011) Microarray technology displays the complexities of the humoral immune response. *Expert Rev Mol Diagn* 11)。高灵敏度是载玻片表面上肽的高密度的结果,并且曾被称为“免疫特征分析效应”。这已通过将肽打印在官能化玻璃表面上并检测肽在其上的不同空间布置而建立。如果打印阵列使得肽间隔约9至约12nm,则同源表位比脱靶随机肽(除非常强的模拟表位以外)更有利地竞争抗体。

[0087] 我们通常使肽平均间隔1-2nm,但在肽间隔3-4nm时观察到脱靶结合。如果肽间隔

约1至约1.5nm,则观察到脱靶结合的增加。紧密填充的肽似乎通过亲合力和快速重结合来捕获抗体。这一概念已经显示出有极大的可再现性,并示于图26中(Stafford,P.等人,(2012)Physical characterization of the“immunosignaturing effect”.Mol Cell Proteomics 11,M111011593;Chase,B.A.等人,(2012)Evaluation of biological sample preparation for immunosignature-based diagnostics.Clin Vaccine Immunol19,352-358;Hughes,A.K.等人,(2012)Immunosignaturing can detect products from molecular markers in brain cancer.PLoS One 7,e40201;Restrepo,L.等人,(2011)Application of immunosignatures to the assessment of Alzheimer’s disease.Annals of Neurology 70,286-295)。虽然肽的序列是完全随机的,但它们对抗体的脱靶捕获明显不是这样;相反,血清与阵列结合的模式非常一致。早期对这种技术的担心是任何血清样品中抗体种类的高度多样性可能导致重叠的结合竞争,从而导致平坦的、无法提供信息的强度域。数据不支持这一观点。事实上,即使是稀释至血清中的纯化的单克隆抗体仍会保持其独特的反应性模式,没有或具有极少的结合损失(Uhlen,M.和Hober,S.(2009)Generation and validation of affinity reagents on a proteome-wide level.J Mol Recognit 22,57-64)。

[0088] 用于解释常规核酸微阵列的经典统计模型(Draghici,S.(2012)Statistics and Data Analysis for Microarrays Using R and Bioconductor.Chapman&Hall/CRC)对于解决由免疫特征阵列所呈现的新复杂性不具有灵活性。并非是描述RNA或DNA与微阵列上的互补探针结合的一对一结合模型,免疫特征分析肽可与多于一种抗体结合,并且许多不同的抗体可与相同的肽结合。三个不同的报告特别针对免疫特征分析比较了图像分析方法(Yang,Y.等人,(2011)Segmentation and intensity estimation for microarray images with saturated pixels.BMC Bioinformatics 12,462)、因素分析和混合模型(Brown,J.R.等人,(2011)Statistical methods for analyzing immunosignatures.BMC Bioinformatics 12,349)以及分类(Kukreja,M.等人,(2012)Comparative study of classification algorithms for immunosignaturing data.BMC Bioinformatics 13,139)。免疫特征分析微阵列存在许多能够区分疾病的基本性质。

[0089] 首先,来自健康志愿者的对照血清表现出相当宽的基线结合反应性分布。这要求使用该技术的大规模研究必须从大量未患病个体中采集血清以适应群体变化性。其次,来自患有给定疾病的人的血清的特征(signatures)非常一致,这不同于非疾病血清的特征。这一观察结果暗示免疫系统不断探查局部环境并与局部环境进行反应,从而导致特征的巨大差异。然而,一旦针对于抗原,抗体趋于形成具有很小的个体变异性的窄且明确限定的特征。

[0090] 即使如此,该技术仍能够识别疾病的亚型(Hughes,A.K.等人,(2012)Immunosignaturing can detect products from molecular markers in brain cancer.PLoS One 7,e40201),同时仍提供对照与患病者之间的区别。对多种肽之间的共同关系和协方差的分析提供了极大的辨别能力,该能力在单一表位水平上是不可能的。在免疫学上,抗体:肽结合模式并非由非特异性危险信号或天然抗体的活性而创建:它们由可识别的刺激而创建。抗体吸附实验证明,来自流感感染的肽主要结合病毒特异性抗体,而阿尔茨海默病的特征结合许多抗A $\beta$ 抗体(Legutki,J.B.等人,(2010)A general method for

characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28,4529-4537; Restrepo, L. 等人, (2011) Application of immunosignatures to the assessment of Alzheimer's disease. *Annals of Neurology* 70,286-295)。

[0091] 通过免疫特征分析进行的疾病确定与使用目前的诊断测试获得的结果具有良好的相关性 (Hughes, A.K. 等人, (2012) Immunosignaturing can detect products from molecular markers in brain cancer. *PLoS One* 7, e40201; Kukreja, M. 等人, (2012) Immunosignaturing Microarrays Distinguish Antibody Profiles of Related Pancreatic Diseases. *Journal of Proteomics and Bioinformatics*; Legutki, J.B. 等人, (2010) A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28,4529-4537)。免疫特征携带采用传统诊断法无法获得的历史健康信息; 即, 即时应答和记忆应答均可检测 (Legutki, J.B. 等人, (2010) A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28,4529-4537)。迄今为止, 该方法已经应用于超过33种不同的疾病和后遗症, 包括小鼠、大鼠、狗、猪和人宿主中的病毒、细菌、真菌和寄生虫感染, 癌症, 糖尿病, 自身免疫疾病, 移植患者和许多慢性疾病。可建立与病理学相关的肽结合模式的高度可再现模式。

[0092] 这些结合谱将从患者和健康志愿者获得的盲血清样品正确地分类, 并且在灵敏度和准确度上优于经典的免疫学检验。在大规模研究中, 免疫特征能够以非常高的准确度诊断溪谷热 (Valley Fever) 患者, 包括正确诊断最初在标准ELISA中检测为阴性的患者。对患者的免疫特征的分析能够在癌症之间和癌症内进行区分 (Brown, J.R. 等人, (2011) Statistical methods for analyzing immunosignatures. *BMC Bioinformatics* 12,349; Hughes, A.K. 等人, (2012) Immunosignaturing can detect products from molecular markers in brain cancer. *PLoS One* 7, e40201; Kukreja, M. 等人, (2012) Immunosignaturing Microarrays Distinguish Antibody Profiles of Related Pancreatic Diseases. *Journal of Proteomics and Bioinformatics*; 以及 Yang, Y. 等人, (2011) Segmentation and intensity estimation for microarray images with saturated pixels. *BMC Bioinformatics* 12,462), 甚至达到准确地诊断将会响应于以及将不会响应于药物治疗的癌症类型的程度 (Hughes, A.K. 等人, Immunosignaturing can detect products from molecular markers in brain cancer. *PLoS One* 7, e40201)。

[0093] 免疫特征分析技术的最独特的特点之一可能是对特定肽: 抗体反应性 (先前不可测量的一类相互作用) 的降低的测量。即, 尽管来自患病个体的血清产生比正常血清更高的信号, 但也存在始终表现出相对于健康人降低的结合的肽。 (Kukreja, M. 等人, (2012) Immunosignaturing Microarrays Distinguish Antibody Profiles of Related Pancreatic Diseases. *Journal of Proteomics and Bioinformatics*; 以及 Legutki, J.B. 等人, (2010) A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28,4529-4537)。这些“下降”肽在免疫应答中的作用是有趣的。尽管在其最简单的层面上, 这些“下降”肽增加了疾病分类, 但可能存在一些以其他方式无法观察到的潜在免疫学现象。

[0094] 分子与阵列的结合

[0095] 根据美国国家癌症研究所(National Cancer Institute),存在约150类癌症以及(取决于人们如何定义它们)数百种不同的亚型。抗体通常针对于肿瘤细胞表达的抗原而产生,并且随后在B细胞成熟期间扩增。抗体也在对疫苗或感染应答期间产生。在受试者日常暴露于各种病原性以及非病原性刺激期间,也可产生抗体。

[0096] 受试者体内抗体扩增的过程可产生与状况相关的受试者特异性标志物的充足供应。抗体扩增可提供与受试者的特定健康状态和/或状况相关的充足数目的抗体。样品中充足数目的抗体的存在可降低健康监测方法中对人工生物标志物扩增的需求。样品中充足数目的抗体的存在可允许将少量样品成功应用于例如健康监测方法中。

[0097] 本发明的方法和阵列允许采用少量来自受试者的生物样品进行健康监测、诊断、治疗和预防。在一些实施方案中,生物样品可以无需进一步处理且以少量应用于本发明的方法中。在一些实施方案中,生物样品包括血液、血清、唾液、汗液、细胞、组织或任何体液。在一些实施方案中,由本发明的阵列和方法进行的分析需要约0.5nl、约1nl、约2nl、约3nl、约4nl、约5nl、约6nl、约7nl、约8nl、约9nl、约10nl、约11nl、约12nl、约13nl、约14nl、约15nl、约16nl、约17nl、约18nl、约19nl、约20nl、约21nl、约22nl、约23nl、约24nl、约25nl、约26nl、约27nl、约28nl、约29nl、约30nl、约31nl、约32nl、约33nl、约34nl、约35nl、约36nl、约37nl、约38nl、约39nl、约40nl、约41nl、约42nl、约43nl、约44nl、约45nl、约46nl、约47nl、约48nl、约49nl或约50nl、约51nl、约52nl、约53nl、约54nl、约55nl、约56nl、约57nl、约58nl、约59nl、约60nl、约61nl、约62nl、约63nl、约64nl、约65nl、约66nl、约67nl、约68nl、约69nl、约70nl、约71nl、约72nl、约73nl、约74nl、约75nl、约76nl、约77nl、约78nl、约79nl、约80nl、约81nl、约82nl、约83nl、约84nl、约85nl、约86nl、约87nl、约88nl、约89nl、约90nl、约91nl、约92nl、约93nl、约94nl、约95nl、约96nl、约97nl、约98nl、约99nl、约0.1、约0.2 $\mu$ l、约0.3 $\mu$ l、约0.4 $\mu$ l、约0.5 $\mu$ l、约0.6 $\mu$ l、约0.7 $\mu$ l、约0.8 $\mu$ l、约0.9 $\mu$ l、约1 $\mu$ l、约2 $\mu$ l、约3 $\mu$ l、约4 $\mu$ l、约5 $\mu$ l、约6 $\mu$ l、约7 $\mu$ l、约8 $\mu$ l、约9 $\mu$ l、约10 $\mu$ l、约11 $\mu$ l、约12 $\mu$ l、约13 $\mu$ l、约14 $\mu$ l、约15 $\mu$ l、约16 $\mu$ l、约17 $\mu$ l、约18 $\mu$ l、约19 $\mu$ l、约20 $\mu$ l、约21 $\mu$ l、约22 $\mu$ l、约23 $\mu$ l、约24 $\mu$ l、约25 $\mu$ l、约26 $\mu$ l、约27 $\mu$ l、约28 $\mu$ l、约29 $\mu$ l、约30 $\mu$ l、约31 $\mu$ l、约32 $\mu$ l、约33 $\mu$ l、约34 $\mu$ l、约35 $\mu$ l、约36 $\mu$ l、约37 $\mu$ l、约38 $\mu$ l、约39 $\mu$ l、约40 $\mu$ l、约41 $\mu$ l、约42 $\mu$ l、约43 $\mu$ l、约44 $\mu$ l、约45 $\mu$ l、约46 $\mu$ l、约47 $\mu$ l、约48 $\mu$ l、约49 $\mu$ l或约50 $\mu$ l的生物样品。

[0098] 来自受试者的生物样品可以例如从受试者采集并直接与本发明的阵列接触。在一些实施方案中,生物样品在与本发明的阵列接触之前,不需要制备或处理步骤。在一些实施方案中,来自受试者的干血样在与本发明的阵列接触之前,在稀释步骤中重建。稀释可为免疫特征分析提供来自受试者的生物样品的抗体的最佳浓度。

[0099] 本发明的方法和阵列允许采用少量来自受试者的生物样品进行健康监测、诊断、治疗和预防。在一些实施方案中,本发明的方法需要不多于约0.5nl至约50nl、不多于约1nl至约100nl、不多于约1nl至约150nl、不多于约1nl至约200nl、不多于约1nl至约250nl、不多于约1nl至约300nl、不多于约1nl至约350nl、不多于约1nl至约400nl、不多于约1至约450nl、不多于约5nl至约500nl、不多于约5nl至约550nl、不多于约5nl至约600nl、不多于约5nl至约650nl、不多于约5nl至约700nl、不多于约5nl至约750nl、不多于约5nl至约800nl、不多于约5nl至约850nl、不多于约5nl至约900nl、不多于约5nl至约950nl、不多于约5nl至

约1 $\mu$ l、不多于约0.5 $\mu$ l至约1 $\mu$ l、不多于约0.5 $\mu$ l至约5 $\mu$ l、不多于约1 $\mu$ l至约10 $\mu$ l、不多于约1 $\mu$ l至约20 $\mu$ l、不多于约1 $\mu$ l至约30 $\mu$ l、不多于约1 $\mu$ l至约40 $\mu$ l或不多于约1 $\mu$ l至约50 $\mu$ l。

[0100] 在一些实施方案中,本发明的方法需要至少0.5nl至约50nl、至少约1nl至约100nl、至少约1nl至约150nl、至少约1nl至约200nl、至少约1nl至约250nl、至少约1nl至约300nl、至少约1nl至约350nl、至少约1nl至约400nl、至少约1nl至约450nl、至少约5nl至约500nl、至少约5nl至约550nl、至少约5nl至约600nl、至少约5nl至约650nl、至少约5nl至约700nl、至少约5nl至约750nl、至少约5nl至约800nl、至少约5nl至约850nl、至少约5nl至约900nl、至少约5nl至约950nl、至少约5nl至约1 $\mu$ l、至少约0.5 $\mu$ l至约1 $\mu$ l、至少约0.5 $\mu$ l至约5 $\mu$ l、至少约1 $\mu$ l至约10 $\mu$ l、至少约1 $\mu$ l至约20 $\mu$ l、至少约1 $\mu$ l至约30 $\mu$ l、至少约1 $\mu$ l至约40 $\mu$ l、至少约1 $\mu$ l至约50 $\mu$ l或至少50 $\mu$ l。

[0101] 受试者可以提供多种生物样品,包括固体生物样品,例如,来自活检或组织。在一些实施方案中,本发明的阵列和方法需要约1mg、约5mg、约10mg、约15mg、约20mg、约25mg、约30mg、约35mg、约40mg、约45mg、约50mg、约55mg、约60mg、约65mg、约7mg、约75mg、约80mg、约85mg、约90mg、约95mg或约100mg的生物样品。

[0102] 在一些实施方案中,本发明的方法和阵列需要不多于约1mg至约5mg、不多于约1mg至约10mg、不多于约1mg至约20mg、不多于约1mg至约30mg、不多于约1mg至约40mg、不多于约1mg至约50mg、不多于约50mg至约60mg、不多于约50mg至约70mg、不多于约50mg至约80mg、不多于约50mg至约90mg、不多于约50mg至约100mg的生物样品。

[0103] 在一些实施方案中,本发明的方法和阵列需要至少约1mg至约5mg、至少约1mg至约10mg、至少约1mg至约20mg、至少约1mg至约30mg、至少约1mg至约40mg、至少约1mg至约50mg、至少约50mg至约60mg、至少约50mg至约70mg、至少约50mg至约80mg、至少约50mg至约90mg、至少约50mg至约100mg的生物样品。

[0104] 本发明的方法和阵列提供了采用少量来自受试者的生物样品针对状况进行健康监测、诊断、治疗和预防的灵敏方法。在一些实施方案中,来自受试者的生物样品的浓度过高,因此在与本发明的阵列接触之前需要稀释。在生物样品与本发明的阵列接触之前,可对该样品进行多次稀释。稀释可以是系列稀释,其可以以对数方式产生几何级数的浓度。例如,10倍系列稀释可以为1M、0.01M、0.001M及其几何级数。稀释可以为,例如,一倍稀释、两倍稀释、三倍稀释、四倍稀释、五倍稀释、六倍稀释、七倍稀释、八倍稀释、九倍稀释、十倍稀释、十六倍稀释、二十五倍稀释、三十二倍稀释、六十四倍稀释和/或一百二十五倍稀释。

[0105] 生物样品可来源于受试者体内的多种来源,并且生物样品可在多种不同情况下从受试者采集。例如,可以在日常医疗咨询期间采集生物样品,如在年度体检期间抽血。生物样品可以在非日常咨询过程中采集,例如,生物样品可在活检过程中采集。受试者还可从自身采集生物样品,并且受试者可以以直接面向消费者的方式提供将要通过本发明的方法和系统来分析的生物样品。在一些实施方案中,可将生物样品邮寄至本发明的方法和阵列的提供者。在一些实施方案中,将干生物样品,如在滤纸上的来自受试者的干血样邮寄至本发明的方法和阵列的提供者。

[0106] 分子与本发明阵列的结合产生可与状况相关的结合模式。分子与阵列中的肽的结合亲和力可在数学上与状况相关。抗体与本发明的多个不同肽的脱靶结合模式可在数学上与状况相关。分子与本发明的多个不同肽的结合亲合力可在数学上与状况相关。脱靶结合

和亲合力可包含生物样品中的分子与肽阵列中的多个不同肽的相互作用。分子与肽阵列中的多个不同肽的结合亲合力可决定该分子与肽阵列的缔合常数。在一些实施方案中,样品中的抗体的浓度有助于与肽阵列的结合亲合力,例如,通过捕获阵列中临界数目的抗体,并允许抗体与阵列的快速重结合。

[0107] 生物分子与阵列的结合亲合力可由多种键相互作用的组合来决定。抗体与肽阵列中的多个肽的交叉反应性可有助于结合亲合力。在一些实施方案中,抗体可识别约3个氨基酸、约4个氨基酸、约5个氨基酸、约6个氨基酸、约7个氨基酸、约8个氨基酸、约9个氨基酸、约10个氨基酸、约11个氨基酸、约12个氨基酸、约13个氨基酸、约14个氨基酸、约15个氨基酸、约16个氨基酸或约17个氨基酸的表位。在一些实施方案中,约5个氨基酸的序列主导抗体与肽的结合能。

[0108] 分子与本发明阵列的脱靶结合和/或亲合力可以,例如,有效地将跨越飞摩尔 (fM) 至微摩尔 ( $\mu\text{M}$ ) 解离常数的结合亲和力压缩至仅使用3log的动态范围即可定量测量的范围。分子可以以 $10^3\text{M}^{-1}$ 或更高的缔合常数与阵列中的多个肽结合。分子可以以范围为 $10^3$ 至 $10^6\text{M}^{-1}$ 、 $2 \times 10^3\text{M}^{-1}$ 至 $10^6\text{M}^{-1}$ 的缔合常数和/或范围为 $10^4\text{M}^{-1}$ 至 $10^6\text{M}^{-1}$ 的缔合常数与阵列中的多种肽结合。分子可以以约1fM、约2fM、约3fM、约4fM、约5fM、约6fM、约7fM、约8fM、约9fM、约10fM、约20fM、约30fM、约40fM、约50fM、约60fM、约70fM、约80fM、约90fM、约100fM、约200fM、约300fM、约400fM、约500fM、约600fM、约700fM、约800fM、约900fM、约1皮摩尔 (pM)、约2pM、约3pM、约4pM、约5pM、约6pM、约7pM、约8pM、约9pM、约10pM、约20pM、约30pM、约40pM、约50pM、约60pM、约7pM、约80pM、约90pM、约100pM、约200pM、约300pM、约400pM、约500pM、约600pM、约700pM、约800pM、约900pM、约1纳摩尔 (nM)、约2nM、约3nM、约4nM、约5nM、约6nM、约7nM、约8nM、约9nM、约10nM、约20nM、约30nM、约40nM、约50nm、约60nM、约70nM、约80nM、约90nM、约100nM、约200nM、约300nM、约400nM、约500nM、约600nM、约700nM、约800nM、约900nM、约1 $\mu\text{M}$ 、约2 $\mu\text{M}$ 、约3 $\mu\text{M}$ 、约4 $\mu\text{M}$ 、约5 $\mu\text{M}$ 、约6 $\mu\text{M}$ 、约7 $\mu\text{M}$ 、约8 $\mu\text{M}$ 、约9 $\mu\text{M}$ 、约10 $\mu\text{M}$ 、约20 $\mu\text{M}$ 、约30 $\mu\text{M}$ 、约40 $\mu\text{M}$ 、约50 $\mu\text{M}$ 、约60 $\mu\text{M}$ 、约70 $\mu\text{M}$ 、约80 $\mu\text{M}$ 、约90 $\mu\text{M}$ 或约100 $\mu\text{M}$ 的解离常数与阵列中的多种肽结合。

[0109] 分子可以以至少1fM、至少2fM、至少3fM、至少4fM、至少5fM、至少6fM、至少7fM、至少8fM、至少9fM、至少10fM、至少20fM、至少30fM、至少40fM、至少50fM、至少60fM、至少70fM、至少80fM、至少90fM、至少100fM、至少200fM、至少300fM、至少400fM、至少500fM、至少600fM、至少700fM、至少800fM、至少900fM、至少1皮摩尔 (pM)、至少2pM、至少3pM、至少4pM、至少5pM、至少6pM、至少7pM、至少8pM、至少9pM、至少10pM、至少20pM、至少30pM、至少40pM、至少50pM、至少60pM、至少7pM、至少80pM、至少90pM、至少100pM、至少200pM、至少300pM、至少400pM、至少500pM、至少600pM、至少700pM、至少800pM、至少900pM、至少1纳摩尔 (nM)、至少2nM、至少3nM、至少4nM、至少5nM、至少6nM、至少7nM、至少8nM、至少9nM、至少10nM、至少20nM、至少30nM、至少40nM、至少50nm、至少60nM、至少70nM、至少80nM、至少90nM、至少100nM、至少200nM、至少300nM、至少400nM、至少500nM、至少600nM、至少700nM、至少800nM、至少900nM、至少1 $\mu\text{M}$ 、至少2 $\mu\text{M}$ 、至少3 $\mu\text{M}$ 、至少4 $\mu\text{M}$ 、至少5 $\mu\text{M}$ 、至少6 $\mu\text{M}$ 、至少7 $\mu\text{M}$ 、至少8 $\mu\text{M}$ 、至少9 $\mu\text{M}$ 、至少10 $\mu\text{M}$ 、至少20 $\mu\text{M}$ 、至少30 $\mu\text{M}$ 、至少40 $\mu\text{M}$ 、至少50 $\mu\text{M}$ 、至少60 $\mu\text{M}$ 、至少70 $\mu\text{M}$ 、至少80 $\mu\text{M}$ 、至少90 $\mu\text{M}$ 或约100 $\mu\text{M}$ 的解离常数与阵列中的多个肽结合。

[0110] 来自生物样品的抗体与肽阵列结合的动态范围可被描述为检测到的结合信号的最大值与最小值之间的比值。结合信号可以是,例如,利用第二抗体检测到的荧光信号。

传统测定受限于结合的预定的且狭窄的动态范围。本发明的方法和阵列可检测宽动态范围的抗体与本发明阵列中的肽的结合。在一些实施方案中,可在对数尺度上检测宽动态范围的抗体结合。在一些实施方案中,本发明的方法和阵列允许使用最高2log的动态范围、最高3log的动态范围、最高4log的动态范围或最高5log的动态范围来检测多种抗体与阵列的结合模式。

[0111] 阵列中的分子组成可决定分子与阵列的结合亲合力。在健康状况的预防、治疗、诊断或监测中使用的阵列中可存在多种不同的分子。生物分子的非限制性实例包括氨基酸、肽、肽模拟物、蛋白质、重组蛋白质抗体(单克隆或多克隆的)、抗体片段、抗原、表位、碳水化合物、脂质、脂肪酸、酶、天然产物、核酸(包括DNA、RNA、核苷、核苷酸、其结构类似物或组合)、营养物、受体和维生素。在一些实施方案中,阵列中的分子为模拟表位,即模拟表位的结构并能够结合由表位引发的抗体的分子。在一些实施方案中,阵列中的分子为互补位或互补位模拟物,包含与抗原表位结合的抗体(或T细胞受体)的可变区中的位点。在一些实施方案中,本发明的阵列是包含随机肽序列的肽阵列。

[0112] 肽阵列中的氨基酸内距离为肽微阵列中的每个肽之间的距离。氨基酸内距离可有助于脱靶结合和/或分子与阵列的结合亲合力。氨基酸内差异可以是约0.5nm、约1nm、约1nm、约1.1nm、约1.2nm、约1.3nm、约1.4nm、约1.5nm、约1.6nm、约1.7nm、约1.8nm、约1.9nm、约2nm、约2.1nm、约2.2nm、约2.3nm、约2.4nm、约2.5nm、约2.6nm、约2.7nm、约2.8nm、约2.9nm、约3nm、约3.1nm、约3.2nm、约3.3nm、约3.4nm、约3.5nm、约3.6nm、约3.7nm、约3.8nm、约3.9nm、约4nm、约4.1nm、约4.2nm、约4.3nm、约4.4nm、约4.5nm、约4.6nm、约4.7nm、约4.8nm、约4.9nm、约5nm、约5.1nm、约5.2nm、约5.3nm、约5.4nm、约5.5nm、约5.6nm、约5.7nm、约5.8nm、约5.9nm和/或约6nm。在一些实施方案中,氨基酸内距离小于6纳米(nm)。

[0113] 氨基酸内差异可以是至少0.5nm、至少1nm、至少1nm、至少1.1nm、至少1.2nm、至少1.3nm、至少1.4nm、至少1.5nm、至少1.6nm、至少1.7nm、至少1.8nm、至少1.9nm、至少2nm、至少2.1nm、至少2.2nm、至少2.3nm、至少2.4nm、至少2.5nm、至少2.6nm、至少2.7nm、至少2.8nm、至少2.9nm、至少3nm、至少3.1nm、至少3.2nm、至少3.3nm、至少3.4nm、至少3.5nm、至少3.6nm、至少3.7nm、至少3.8nm、至少3.9nm、至少4nm、至少4.1nm、至少4.2nm、至少4.3nm、至少4.4nm、至少4.5nm、至少4.6nm、至少4.7nm、至少4.8nm、至少4.9nm、至少5nm、至少5.1nm、至少5.2nm、至少5.3nm、至少5.4nm、至少5.5nm、至少5.6nm、至少5.7nm、至少5.8nm或至少5.9nm。

[0114] 氨基酸内差异可以是不大于0.5nm、不大于1nm、不大于1nm、不大于1.1nm、不大于1.2nm、不大于1.3nm、不大于1.4nm、不大于1.5nm、不大于1.6nm、不大于1.7nm、不大于1.8nm、不大于1.9nm、不大于2nm、不大于2.1nm、不大于2.2nm、不大于2.3nm、不大于2.4nm、不超过2.5nm、不大于2.6nm、不大于2.7nm、不大于2.8nm、不大于2.9nm、不大于3nm、不大于3.1nm、不大于3.2nm、不大于3.3nm、不大于3.4nm、不大于3.5nm、不大于3.6nm、不大于3.7nm、不大于3.8nm、不大于3.9nm、不大于4nm、不大于4.1nm、不大于4.2nm、不大于4.3nm、不大于4.4nm、不大于4.5nm、不大于4.6nm、不大于4.7nm、不大于4.8nm、不大于4.9nm、不大于5nm、不大于5.1nm、不大于5.2nm、不大于5.3nm、不大于5.4nm、不大于5.5nm、不大于5.6nm、不大于5.7nm、不大于5.8nm、不大于5.9nm和/或不大于6nm。在一些实施方案中,氨基酸内距离不大于6纳米(nm)。

[0115] 氨基酸内差异的范围可以是0.5nm至1nm、0.5nm至2nm、0.5nm至3nm、0.5nm至3nm、0.5nm至4nm、0.5nm至5nm、0.5nm至6nm、1nm至2nm、1nm至3nm、1nm至4nm、1nm至5nm、1nm至6nm、2nm至3nm、2nm至4nm、2nm至5nm、2nm至6nm、3nm至4nm、3nm至5nm、3nm至6nm、4nm至5nm、4nm至6nm和/或5nm至6nm。

[0116] 肽阵列可以在表面上包含多种不同的肽模式。肽阵列可包含,例如,不同的多个肽和/或分子的单份、一式两份、一式三份、一式四份、一式五份、一式六份、一式七份、一式八份、一式九份和/或一式十份重复。在一些实施方案中,将多个不同的肽重复地点印在肽阵列表面上。肽阵列可以,例如,包含均匀地分布在阵列上的多个肽。肽阵列可以,例如,包含非均匀地分布在阵列上的多个肽。

[0117] 可将肽“点印”于肽阵列中。肽斑点(peptide spot)可具有各种几何形状,例如,肽斑点可以是圆形、正方形、矩形和/或三角形。肽斑点可具有多种直径。肽斑点直径的非限制性实例为约3 $\mu$ m至约8 $\mu$ m、约3至约10mm、约5至约10mm、约10 $\mu$ m至约20 $\mu$ m、约30 $\mu$ m、约40 $\mu$ m、约50 $\mu$ m、约60 $\mu$ m、约70 $\mu$ m、约80 $\mu$ m、约90 $\mu$ m、约100 $\mu$ m、约110 $\mu$ m、约120 $\mu$ m、约130 $\mu$ m、约140 $\mu$ m、约150 $\mu$ m、约160 $\mu$ m、约170 $\mu$ m、约180 $\mu$ m、约190 $\mu$ m、约200 $\mu$ m、约210 $\mu$ m、约220 $\mu$ m、约230 $\mu$ m、约240 $\mu$ m和/或约250 $\mu$ m。

[0118] 肽阵列可包含许多不同的肽。在一些实施方案中,肽阵列包含约10个肽;约50个肽;约100个肽;约200个肽;约300个肽;约400个肽;约500个肽;约750个肽;约1000个肽;约1250个肽;约1500个肽;约1750个肽;约2,000个肽;约2,250个肽;约2,500个肽;约2,750个肽;约3,000个肽;约3,250个肽;约3,500个肽;约3,750个肽;约4,000个肽;约4,250个肽;约4,500个肽;约4,750个肽;约5,000个肽;约5,250个肽;约5,500个肽;约5,750个肽;约6,000个肽;约6,250个肽;约6,500个肽;约7,500个肽;约7,725个肽;8,000个肽;约8,250个肽;约8,500个肽;约8,750个肽;约9,000个肽;约9,250个肽;约10,000个肽;约10,250个肽;约10,500个肽;约10,750个肽;约11,000个肽;约11,250个肽;约11,500个肽;约11,750个肽;约12,000个肽;约12,250个肽;约12,500个肽;约12,750个肽;约13,000个肽;约13,250个肽;约13,500个肽;约13,750个肽;约14,000个肽;约14,250个肽;约14,500个肽;约14,750个肽;约15,000个肽;约15,250个肽;约15,500个肽;约15,750个肽;约16,000个肽;约16,250个肽;约16,500个肽;约16,750个肽;约17,000个肽;约17,250个肽;约17,500个肽;约17,750个肽;约18,000个肽;约18,250个肽;约18,500个肽;约18,750个肽;约19,000个肽;约19,250个肽;约19,500个肽;约19,750个肽;约20,000个肽;约20,250个肽;约20,500个肽;约20,750个肽;约21,000个肽;约21,250个肽;约21,500个肽;约21,750个肽;约22,000个肽;约22,250个肽;约22,500个肽;约22,750个肽;约23,000个肽;约23,250个肽;约23,500个肽;约23,750个肽;约24,000个肽;约24,250个肽;约24,500个肽;约24,750个肽;约25,000个肽;约25,250个肽;约25,500个肽;约25,750个肽;和/或约30,000个肽。

[0119] 在一些实施方案中,在健康监测方法、治疗方法、诊断方法和预防状况的方法中使用的肽阵列包含多于30,000个肽。在一些实施方案中,在健康监测方法中使用的肽阵列包含约330,000个肽。在一些实施方案中,该阵列包含约30,000个肽;约35,000个肽;约40,000个肽;约45,000个肽;约50,000个肽;约55,000个肽;约60,000个肽;约65,000个肽;约70,000个肽;约75,000个肽;约80,000个肽;约85,000个肽;约90,000个肽;约95,000个肽;约100,000个肽;约105,000个肽;约110,000个肽;约115,000个肽;约120,000个肽;约125,000

个肽;约130,000个肽;约135,000个肽;约140,000个肽;约145,000个肽;约150,000个肽;约155,000个肽;约160,000个肽;约165,000个肽;约170,000个肽;约175,000个肽;约180,000个肽;约185,000个肽;约190,000个肽;约195,000个肽;约200,000个肽;约210,000个肽;约215,000个肽;约220,000个肽;约225,000个肽;约230,000个肽;约240,000个肽;约245,000个肽;约250,000个肽;约255,000个肽;约260,000个肽;约265,000个肽;约270,000个肽;约275,000个肽;约280,000个肽;约285,000个肽;约290,000个肽;约295,000个肽;约300,000个肽;约305,000个肽;约310,000个肽;约315,000个肽;约320,000个肽;约325,000个肽;约330,000个肽;约335,000个肽;约340,000个肽;约345,000个肽;和/或约350,000个肽。在一些实施方案中,在健康监测方法中使用的肽阵列包含多于330,000个肽。

[0120] 肽阵列可包含许多不同的肽。在一些实施方案中,肽阵列包含至少2,000个肽;至少3,000个肽;至少4,000个肽;至少5,000个肽;至少6,000个肽;至少7,000个肽;至少8,000个肽;至少9,000个肽;至少10,000个肽;至少11,000个肽;至少12,000个肽;至少13,000个肽;至少14,000个肽;至少15,000个肽;至少16,000个肽;至少17,000个肽;至少18,000个肽;至少19,000个肽;至少20,000个肽;至少21,000个肽;至少22,000个肽;至少23,000个肽;至少24,000个肽;至少25,000个肽;至少30,000个肽;至少40,000个肽;至少50,000个肽;至少60,000个肽;至少70,000个肽;至少80,000个肽;至少90,000个肽;至少100,000个肽;至少110,000个肽;至少120,000个肽;至少130,000个肽;至少140,000个肽;至少150,000个肽;至少160,000个肽;至少约170,000个肽;至少180,000个肽;至少190,000个肽;至少200,000个肽;至少210,000个肽;至少220,000个肽;至少230,000个肽;至少240,000个肽;至少250,000个肽;至少260,000个肽;至少270,000个肽;至少280,000个肽;至少290,000个肽;至少300,000个肽;至少310,000个肽;至少320,000个肽;至少330,000个肽;至少340,000个肽;至少350,000个肽。在一些实施方案中,在健康监测方法中使用的肽阵列包含至少330,000个肽。

[0121] 肽可通过连接体分子物理链接至肽阵列上。肽的N末端或C末端可附接至连接体分子。连接体分子可以为,例如,存在于阵列表面上的多种官能化分子,如酰亚胺官能团、胺官能团、羟基官能团、羧基官能团、醛官能团和/或巯基官能团。连接体分子可以为,例如,聚合物。在一些实施方案中,连接体为马来酰亚胺。在一些实施方案中,连接体为甘氨酸-丝氨酸-半胱氨酸(GSC)或甘氨酸-甘氨酸-半胱氨酸(GGC)连接体。在一些实施方案中,连接体由不同长度或组成的多肽组成。在一些情况下,连接体为不同长度的聚乙二醇。在另外一些情况下,连接体为羟甲基苯甲酸、4-羟基-2-甲氧基苯甲醛、4-氨基磺酰苯甲酸或适用于将肽附接至固体基底的其他连接体。

[0122] 肽阵列的表面可包含多种不同的材料。肽阵列的表面可以是,例如,玻璃。可构成肽阵列表面的材料的非限制性实例包括玻璃、官能化玻璃、硅、锗、砷化镓、磷化镓、二氧化硅、氧化钠、氮化硅(silicon nitride)、硝化纤维、尼龙、聚四氟乙烯、聚偏二氟乙烯(polyvinylidene difluoride)、聚苯乙烯、聚碳酸酯、甲基丙烯酸酯或其组合。

[0123] 肽阵列的表面可以是平的、凹的或凸的。肽阵列的表面可以是均质的,并且阵列的表面可以是异质的。在一些实施方案中,肽阵列的表面是平的。

[0124] 肽阵列的表面可涂覆有涂层。涂层可以例如改善本发明的阵列的粘附能力。涂层可以例如降低生物样品与本发明的阵列的背景粘附。在一些实施方案中,本发明的肽阵列

包含具有氨基硅烷涂层的载玻片。

[0125] 肽阵列可具有多种尺寸。肽阵列可以是肽微阵列。

[0126] 制造阵列

[0127] 本文还公开了促进用于制造复杂生物阵列如上述肽阵列的图案化技术的方法。现有的方法已经表明了使用光刻术 (lithography) 或其他图案化技术来制备在表面上具有限定位置的杂聚物文库的可行性。该方法已广泛地应用于DNA和肽阵列。最简单的方法是制备成层的杂聚物文库。考虑由M个单体的序列组成的、长度为N的杂聚物。通常,每层存在M个图案化步骤,每个单体一个步骤。对于序列长度N,将存在N层图案化。因此,图案化步骤的总数为 $N \times M$ 。

[0128] 图案化的另一方面是它是二元 (binary) 事件。换言之,在每个图案化步骤中表面的任何区域是“暴露的”或者“未暴露的”,其中暴露是暴露于在图案化中使用的任何辐射、化学品、效应物或力。以这种方法对阵列进行图案化涉及将 $M^N$ 种可能性的序列空间投射至 $2^R$ 种可能性的二元空间上,其中R是图案化步骤的总数。原则上,R的最小值是通过设置两个表达式相等给出的,且求解该方程得到:

$$[0129] \quad R = N \frac{\ln M}{\ln 2}$$

[0130] 如果人们希望能够由在 $2^R$ 空间中的一系列图案化步骤代表在 $M^N$ 空间中的任何杂聚物,则表示图案化步骤的理论最小数目。人们可将逐层算法中的图案化步骤的数目 ( $M \times N$ ) 与上述R给出的最小数目进行比较,如以下表1中所示。

[0131]

N \ M	10		11		12		13		14		15	
	M×N	R	M×N	R	M×N	R	M×N	R	M×N	R	M×N	R
8	80	27	88	28	96	29	104	30	112	30	120	31
9	90	30	99	31	108	32	117	33	126	34	135	35
10	100	33	110	35	120	36	130	37	140	38	150	39
11	110	37	121	38	132	39	143	41	154	42	165	43
12	120	40	132	42	144	43	156	44	168	46	180	47
13	130	43	143	45	156	47	169	48	182	49	195	51

[0132]

<b>14</b>	140	47	154	48	168	50	182	52	196	53	210	55
<b>15</b>	150	50	165	52	180	54	195	56	210	57	225	59
<b>16</b>	160	53	176	55	192	57	208	59	224	61	240	63
<b>17</b>	170	56	187	59	204	61	221	63	238	65	255	66
<b>18</b>	180	60	198	62	216	65	234	67	252	69	270	70
<b>19</b>	190	63	209	66	228	68	247	70	266	72	285	74
<b>20</b>	200	66	220	69	240	72	260	74	280	76	300	78

表 1, 第 2 部分

N \ M	16		17		18		19		20	
	M×N	R	M×N	R	M×N	R	M×N	R	M×N	R
<b>8</b>	128	32	136	33	144	33	152	34	160	35
<b>9</b>	144	36	153	37	162	38	171	38	180	39
<b>10</b>	160	40	170	41	180	42	190	42	200	43
<b>11</b>	176	44	187	45	198	46	209	47	220	48
<b>12</b>	192	48	204	49	216	50	228	51	240	52
<b>13</b>	208	52	221	53	234	54	247	55	260	56
<b>14</b>	224	56	238	57	252	58	266	59	280	61
<b>15</b>	240	60	255	61	270	63	285	64	300	65
<b>16</b>	256	64	272	65	288	67	304	68	320	69
<b>17</b>	272	68	289	69	306	71	323	72	340	73
<b>18</b>	288	72	306	74	324	75	342	76	360	78
<b>19</b>	304	76	323	78	342	79	361	81	380	82
<b>20</b>	320	80	340	82	360	83	380	85	400	86

[0133] 与二元表示法的理论最小值相比,逐层法中包括的步骤的数目是非常大的。问题在于:在正常的光刻术加工中,单独地添加每种氨基酸,因此没有办法使用这种方法直接加印将会分选出不同氨基酸的二元代码。然而,也不必一次仅向一层上添加氨基酸,从而导致所需循环数的显著变化。

[0134] 本发明的新的图案化过程以下述方式进行描述。在一个示例实施方案中,使用通过采用10种不同种类的单体形成的杂聚物的阵列,并且用于形成杂聚物的单体的百分比是相等的,即每种单体为10%。第一图案化步骤添加单体A;即10%的杂聚物将在第一层中具有A。第二步考虑单体B。在该实施方案中,分配至第一层的单体中的10%将具有B;但另外,当前可用的第二层中的10%(即,在第一层中接收A的10%)也将准备接收B单体。因此,B实际上将偶合至总胺位点的11%。在添加单体C的第三步中,有10%的杂聚物接收C作为第一层,然后是第二层中的位点的10%,以及当前对已经添加有A和B两者的第三层开放的位点中的10%。该过程持续进行并最终稳定于放置在表面上的每种单体代表接近可用胺的

20%的水平上,即使每层仅有10%添加有任何特定单体。与逐层合成相比,这导致对于特定数目的步骤,所制备的聚合物的平均长度增加了几乎两倍。

[0135] 该过程可以以算法的形式来描述。简而言之,该过程是递归地将系列中的下一种单体添加至序列指示的每一可用层中。当在单体中以特定顺序反复地循环时,该算法具有最大效果。通常,该算法以下述方式工作:

[0136] • 选择用于制备杂聚物的一组单体。

[0137] • 向每种单体分配每层(每个残基)覆盖的添加位点(例如,在肽合成中的胺)的部分

[0138] ○在一个实施方案中,选择该部分为 $1/(\text{单体数})$ 。

[0139] ○在另一个实施方案中,单体具有总和为100%的不同的部分。

[0140] ○在另一个实施方案中,当产生伪随机肽序列时,与单体相关的部分可等于一个大于100%的值。

[0141] • 产生一组期望的杂聚物序列,其包括使用伪随机或随机序列。

[0142] • 使用图案化的化学方法:一次将一种单体添加至所有位置,所述位置将根据所期望的序列适当地延伸该肽,而不论杂聚物中的哪个残基位置可用于添加。

[0143] ○在一个实施方案中,所述步骤包括以预定的顺序在单体中循环。这对于最小数目的图案化步骤给出了最长的肽。

[0144] 也可将每个循环中的顺序或添加完全地改变或随机化,但随机顺序的图案化将增加获得特定的平均长度所需的图案化步骤的数目。

[0145] 鉴于上述部分分配,即使任何特定的层仅具有单体的部分,使用该算法在图案化步骤中添加的实际部分也要显著更高。在图案化步骤中添加的单体的量可如下估算。以 $f_j$ 表示加入特定单体的层的部分。计算加入至所有层中的单体的所有部分的总和得到:

$$[0146] \left( f_i \sum_{j=i-Z}^{i-1} f_j \right) + f_i$$

[0147] 其中,下标 $i$ 表示当前的图案化步骤数; $Z$ 是自上一次添加当前单体起已经添加的不同单体的数目;因此,该总和超过与自上一次添加当前单体起已经添加的所有单体相关的每层的部分。

[0148] 一个示例实施方案在此示出。在这种情况下,使用16种氨基酸来构建10,000个预定义的肽序列。图27示出了作为图案化步骤数目的函数的合成肽的平均长度。Y轴为平均肽长度而X轴为图案化循环的数目。注意:对于几乎任何数目的图案化循环,该优化的模型使制造效率提高了几乎两倍。

[0149] 利用本发明产生伪随机肽的其他方法是使用这种方法经计算产生非常多的肽序列,但随后在阵列的产生中只包括最长的肽序列。这种方法导致偏向于具有与添加氨基酸的顺序(虽然通常不是连续的)相似的顺序的序列。所得序列仍然覆盖大量的空间,并且随机性程度取决于实施者选择什么分布部分。参考图30,下面描述了使用16种不同的氨基酸由优化算法的70个步骤产生的分布的实施方案。这些序列中的前5%平均为约12个残基的长度并且可被选择用于在阵列中实际合成,而其他序列被弃去。如果将图案化步骤的数目降低到60,则可通过选择前0.5%的最长的肽来获得大约相同的平均肽长度。同样地,使用

的图案化步骤的数目越少,肽文库所具有的序列偏性越大,但在一定的偏性可被容忍的程度上,图案化步骤的数目可极大地减少。

[0150] 在一些实施方案中,使用本发明可通过在稍多于100个图案化步骤中使用16种不同的氨基酸来产生平均长度为12个残基的限定序列的肽阵列。然而,当实施方案试图制备具有特定一组序列的特定一组杂聚物时,直到基本上经过 $M \times N$ 个图案化步骤为止,我们将无法直至各个序列的末端。在一部分序列比所预定的短一个或两个单体终止的实施方案中,我们可以通过比 $M \times N$ 少得多的步骤制备序列。图28示出了相对于优化的算法,对于标准的逐层法使用所有20种氨基酸得到的结果。

[0151] 另一个实施方案考虑产生重叠的肽序列,这些肽序列在它们之间覆盖整个蛋白质组,如人类蛋白质组。人们可使用这样的阵列进行表位发现或鉴定蛋白质或小分子的结合位点。例如,可以使用具有3-5个氨基酸重叠的、12-15个残基长的肽的阵列来鉴定线性表位。产生这样的阵列将需要在表面上有几百万个肽。这可通过制备平均长度为13个残基的阵列来实现,使用优化的算法这将需要约140个步骤,而与之相比,使用逐层法将需要260个步骤。

[0152] 本文公开的阵列可与如上所述的免疫特征分析结合使用。阵列上的肽的可变长度是可接受的,并且当与免疫特征分析结合使用时,有时也是所希望的。使用16种不同的氨基酸并具有平均12个残基的肽已显示非常适用于免疫特征分析,并且此类肽的随机阵列可以通过稍多于100个图案化步骤来制备,如图27所示。相比之下,使用逐层合成将需要192个步骤。

[0153] 免疫特征分析也可采用非完全随机的肽有效地实现。有两种方式使用该算法通过比纯随机肽更少的步骤来产生伪随机肽。考虑使用16种类型的单体(例如氨基酸)的阵列的实例。我们可以简单地运行氨基酸的循环,如同仅存在8种氨基酸而不是16种,但随后在所使用的8种氨基酸的组之间交替进行。这种添加单体的方式意味着该系列中的初始几个氨基酸将偏向于第一组的8种氨基酸。最终,偏性将减弱,尽管并未完全消失。图29示出了该实施方案的结果;我们可以在不到60个步骤后获得平均12个残基的长度。

#### [0154] 检测

[0155] 可以以多种形式检测样品成分与阵列之间的结合相互作用。在一些形式中,对样品的成分进行标记。标记物可以是放射性同位素或染料等。可通过在获得样品前对患者施用标记物或通过将标记物连接至样品或其选择性成分上来提供标记物。

[0156] 结合相互作用还可使用第二检测试剂如抗体来检测。例如,样品中的抗体与阵列的结合可使用对抗体的同种型(例如,IgG(包括任何亚型,如IgG1、IgG2、IgG3和IgG4)、IgA、IgM)具有特异性的第二抗体来检测。第二抗体通常被标记,并且可与被分析样品中特定同种型的所有抗体结合。可使用具有不同的同种型特异性的不同的第二抗体。虽然在被相同样品中不同同种型的抗体所结合的化合物中通常存在大量重叠,但在谱中也存在差异。

[0157] 结合相互作用还可使用无标记方法如表面等离子体共振(SPR)和质谱分析法来检测。SPR可提供对解离常数和解离率的量度。例如,A-100Biocore/GE仪器适合于这种类型的分析。FLEX芯片(FLEXchip)可用于分析同一支持体上的多达400个结合反应。

[0158] 任选地,样品成分与阵列之间的结合相互作用可以以竞争形式来检测。在存在与不存在结合的竞争性抑制剂的情况下阵列与样品的结合谱的差异在表征样品时可能是有

用的。该竞争性抑制剂可以是，例如，与疾病状况相关的已知蛋白质，如病原体或病原体的抗体。在这样的竞争剂的存在下阵列成员与样品的结合的降低提供了存在该病原体的指示。可通过改变盐、离子强度、有机溶剂含量和文库成员与靶标接触时的温度来调节严格性。

[0159] 基于抗体的检测方法，如酶联免疫吸附测定 (ELISA) 方法，可用于检测与本发明阵列的结合模式。例如，检测免疫球蛋白的特定同种型 (例如，IgM同种型) 的第二抗体可用于检测来自受试者的复杂生物样品的多种IgM抗体与阵列的结合模式。例如，第二抗体可以缀合至可检测标记物，如荧光部分或放射性标记物。

[0160] 本发明提供了用于检测多种不同抗体与本发明阵列的脱靶结合的阵列和方法。复杂生物样品中的多种抗体能够与肽微阵列中的多个肽脱靶结合。在一些实施方案中，检测至少一种抗体与肽阵列中的多个肽的脱靶结合可形成免疫特征。抗体的众多类别或同种型可提供与阵列的脱靶结合模式。抗体或免疫球蛋白可以为IgA、IgD、IgE、IgG和/或IgM抗体。

[0161] 来自复杂生物样品的至少一种IgM抗体与肽阵列的结合模式可形成免疫特征。IgM抗体可形成聚合物，在该聚合物中多个免疫球蛋白通过二硫键共价连接在一起。IgM聚合物可为五聚体。IgM同种型的抗体的聚合性质可增强样品与阵列的脱靶结合。抗体的聚合性质可提高样品与阵列的结合亲合力。例如，聚合IgM同种型抗体的抗体与肽微阵列的结合模式可形成独特的五聚体驱动的免疫特征。

[0162] IgA抗体可以是IgA1或IgA2抗体。IgA同种型的抗体可形成二聚体。IgG抗体可以是IgG1、IgG2、IgG3或IgG4抗体。IgG同种型的抗体可作为单体存在。IgD和/或IgE抗体可形成单体。在一些实施方案中，本发明可检测来自受试者的复杂生物样品的至少一种IgM抗体与肽阵列的脱靶结合。

[0163] 在受试者的整个生命期间监测受试者

[0164] 本发明的方法、装置、试剂盒、阵列和系统可用于在受试者的整个生命期间监测受试者。受试者的生命期间可指受试者自出生起已经经历的时间。用本发明的方法、阵列、试剂盒和系统对受试者健康的监测可以并入受试者的病历或受试者电子病历 (EMR) 中。

[0165] 电子病历 (EMR) 可涉及由受试者的医生、临床医生、保险公司、医院和/或受试者在其中作为患者的其他设施获得并保存的记录。在一些实施方案中，医生可包括医师、牙医、验光师、治疗师、按摩师以及向受试者提供医疗保健服务的任何人。电子病历 (EMR) 可以包括，例如，CAT扫描、MRI、超声波、血糖水平、诊断、变态反应、实验室检验结果、EKG、药物治疗、日常图表、药物施用、身体评估、入院护理记录、护理计划、转诊 (referral)、现在和既往的症状、病史、生活方式、体检结果、检测、程序、治疗、药物治疗、出院、历史记录、日记、问题、发现的情况、免疫接种、入院记录、服务记录、病程记录、术前记录、手术记录、术后记录、程序记录、分娩记录、产后记录和出院记录。

[0166] 治疗和状况

[0167] 本发明的阵列和方法可用于，例如，诊断、监测、表征和指导治疗受试者的多种不同的状况。受试者可以是人、豚鼠、狗、猫、马、小鼠、兔和各种其他动物。受试者可为任何年龄，例如，受试者可为婴儿、学步孩童、儿童、青春期前儿童、青少年、成年或老年个体。

[0168] 受试者的状况可对应于疾病或健康状况。在一些实施方案中，受试者的状况为健康状况，并且本发明的方法监测该健康状况。在一些实施方案中，受试者的状况为疾病状

况,并且本发明的方法用于诊断/监测该状况的状态和/或进展。本发明的方法还可用于预防状况。在一些实施方案中,本发明的方法与预防性处理结合使用。

[0169] 重要的是,本文公开的阵列装置和方法同时检测和监测多种疾病和/或状况。例如,本文公开的阵列装置和方法能够在同一阵列上同时检测炎性状况、癌症疾病和病原性感染。因此,检测广谱的疾病和状况仅需要一个阵列,即一个免疫特征测定。因此,在受试者的整个生命期间对受试者的监测将利用所进行的每次免疫特征来提供受试者的健康状态随时间的快照。这提供了检测受试者健康状态的总体和具体改变的有力手段,并且与高灵敏度的免疫特征测定一起提供了一种能够非常早期地检测个体健康状态的任何变化的系统。

[0170] 因此,本文公开的方法、系统和阵列装置能够在疾病和/或状况的早期检测、诊断、监测、预防和/或治疗该疾病和/或状况。例如,本文公开的方法、系统和阵列装置能够在传统的基于生物标志物的测定之前数天或数周检测、诊断和监测疾病和/或状况。此外,仅需要一个阵列,即一个免疫特征测定,即可检测、诊断和监测疾病和状况(包括炎性状况、癌症和病原性感染)的副谱(side spectra)。

[0171] 本发明的阵列和方法也可用于,例如,诊断、监测、预防和/或治疗癌症。可用本发明的阵列和方法诊断、监测、预防和/或治疗的癌症的非限制性实例可包括:急性淋巴母细胞性白血病、急性髓样白血病、肾上腺皮质癌、AIDS相关癌症、AIDS相关淋巴瘤、肛门癌、阑尾癌、星形细胞瘤、基底细胞癌、胆管癌、膀胱癌、骨癌、脑肿瘤(如小脑星形细胞瘤、大脑星形细胞瘤/恶性胶质瘤)、室管膜瘤、髓母细胞瘤、幕上原始神经外胚层肿瘤、视觉通路和下丘脑神经胶质瘤、乳腺癌、支气管腺瘤、伯基特淋巴瘤(Burkitt lymphoma)、未知原发来源的癌症、中枢神经系统淋巴瘤、小脑星形细胞瘤、宫颈癌、儿童期癌症、慢性淋巴细胞性白血病、慢性髓细胞性白血病、慢性骨髓增生性病症、结肠癌、皮肤T细胞淋巴瘤、促结缔组织增生性小圆细胞肿瘤、子宫内膜癌、室管膜瘤、食管癌、尤因氏肉瘤(Ewing's sarcoma)、生殖细胞肿瘤、胆囊癌、胃癌、胃肠道类癌瘤、胃肠道间质瘤、神经胶质瘤、毛细胞白血病、头颈癌、心脏癌、肝细胞(肝)癌、霍奇金淋巴瘤、下咽癌、眼内黑素瘤、胰岛细胞癌、卡波西肉瘤、肾癌、喉癌、唇及口腔癌、脂肪肉瘤、肝癌、肺癌(例如非小细胞和小细胞肺癌)、淋巴瘤、白血病、巨球蛋白血症、骨恶性纤维组织细胞瘤/骨肉瘤、髓母细胞瘤、黑素瘤、间皮瘤、隐性原发性转移性鳞状颈癌、口癌、多发性内分泌肿瘤综合征、骨髓增生异常综合征、髓样白血病、鼻腔和副鼻窦癌、鼻咽癌、神经母细胞瘤、非霍奇金淋巴瘤、非小细胞肺癌、口腔癌、口咽癌、骨肉瘤/骨恶性纤维组织细胞瘤、卵巢癌、卵巢上皮瘤、卵巢生殖细胞肿瘤、胰腺癌、胰岛细胞胰腺癌、副鼻窦和鼻腔癌、甲状旁腺癌、阴茎癌、咽癌、嗜铬细胞瘤、松果体星形细胞瘤(pineal astrocytoma)、松果体生殖细胞瘤(pineal germinoma)、垂体腺瘤、胸膜肺母细胞瘤、浆细胞瘤、原发性中枢神经系统淋巴瘤、前列腺癌、直肠癌、肾细胞癌、肾盂和输尿管移行细胞癌、视网膜母细胞瘤、横纹肌肉瘤、唾液腺癌、肉瘤、皮肤癌、梅克尔细胞皮肤癌(skin carcinoma merkel cell)、小肠癌、软组织肉瘤、鳞状细胞癌、胃癌、T细胞淋巴瘤、喉癌、胸腺瘤、胸腺癌、甲状腺癌、滋养细胞肿瘤(妊娠期)、未知原发部位的癌症、尿道癌、子宫肉瘤、阴道癌、外阴癌、瓦尔登斯特伦巨球蛋白血症(Waldenström macroglobulinemia)和肾母细胞瘤(Wilms tumor)。

[0172] 在一些实施方案中,本发明的方法可用于诊断、监测、预防和/或治疗与免疫系统

相关的状况。与免疫系统相关的病症的非限制性实例可包括：自身免疫病症、炎性疾病、HIV、类风湿性关节炎、I型糖尿病、系统性红斑狼疮、硬皮病、多发性硬化、重症联合免疫缺陷(SCID)、迪乔治综合征(DiGeorge syndrome)、共济失调性毛细血管扩张症(ataxia-telangiectasia)、季节性变态反应、常年性变态反应、食物变态反应、过敏反应(anaphylaxis)、肥大细胞增多症、变应性鼻炎、特应性皮炎、帕金森病、阿尔茨海默病、脾功能亢进、白细胞粘附缺陷症、X连锁淋巴增生性疾病、X连锁无丙种球蛋白血症、选择性免疫球蛋白A缺乏症、高IgM综合征、自身免疫性淋巴增生性综合征、Wiskott-Aldrich综合征、慢性肉芽肿病、常见变异型免疫缺陷病(CVID)、高免疫球蛋白E综合征和桥本甲状腺炎(Hashimoto's thyroiditis)。

[0173] 本发明可提供一种预防状况的方法，该方法包括：a) 提供来自受试者的复杂生物样品；b) 使所述复杂生物样品与肽阵列接触，其中所述肽阵列包含能够脱靶结合复杂生物样品中的至少一种抗体的不同肽；c) 测量所述复杂生物样品与多个不同肽的脱靶结合以形成免疫特征；d) 将所述免疫特征与状况相关联；以及e) 接受针对所述状况的处理。

[0174] 在一些实施方案中，本发明的方法可与预防性处理结合使用。例如，疫苗可以是预防性处理。起到预防性处理作用的疫苗的非限制性实例包括脊髓灰质炎疫苗、天花疫苗、麻疹疫苗、腮腺炎疫苗、人乳头瘤病毒(HPV)疫苗和流感疫苗。在一些实施方案中，本发明的方法用于监测，例如，受试者对预防性疫苗的应答。

[0175] 在一些实施方案中，本发明提供了一种提供治疗的方法，该方法包括：a) 接收来自受试者的复杂生物样品；b) 使复杂生物样品与肽阵列接触，其中所述肽阵列包含能够脱靶结合生物样品中的至少一种抗体的不同肽；c) 测量抗体与多个不同肽的脱靶结合以形成免疫特征；d) 将所述免疫特征与状况相关联；以及e) 提供针对所述状况的治疗。

[0176] 在一些实施方案中，本发明可提供一种诊断方法，该方法包括：a) 接收来自受试者的复杂生物样品；b) 使所述复杂生物样品与肽阵列接触，其中所述肽阵列包含能够脱靶结合生物样品中的至少一种抗体的不同肽；c) 测量抗体与肽阵列中的一组不同肽的脱靶结合以形成免疫特征；以及d) 基于所述免疫特征诊断状况。

[0177] 在一些实施方案中，本发明的方法可用作诊断、监测和治疗状况的方法。治疗状况的方法可能需要开出以治疗受试者的状况或疾病为目标的治疗剂的处方。在一些实施方案中，可以以下述剂量范围开出治疗剂：约1mg至约2000mg；约5mg至约1000mg、约10mg至约500mg、约50mg至约250mg、约100mg至约200mg、约1mg至约50mg、约50mg至约100mg、约100mg至约150mg、约150mg至约200mg、约200mg至约250mg、约250mg至约300mg、约300mg至约350mg、约350mg至约400mg、约400mg至约450mg、约450mg至约500mg、约500mg至约550mg、约550mg至约600mg、约600mg至约650mg、约650mg至约700mg、约700mg至约750mg、约750mg至约800mg、约800mg至约850mg、约850mg至约900mg、约900mg至约950mg或约950mg至约1000mg。

[0178] 在一些实施方案中，开出至少1mg、至少5mg、至少15mg、至少15mg、至少20mg、至少25mg、至少30mg、至少35mg、至少40mg、至少45mg、至少50mg、至少55mg、至少60mg、至少65mg、至少70mg、至少80mg、至少85mg、至少90mg、至少100mg、至少150mg、至少200mg、至少250mg、至少300mg、至少350mg、至少400mg、至少450mg、至少500mg、至少550mg、至少600mg、至少650mg、至少700mg、至少750mg、至少800mg、至少850mg、至少900mg、至少950mg或至少1000mg的治疗剂。

[0179] 本发明的阵列和方法可由使用者使用。许多使用者可使用本发明的方法来监测、诊断、治疗或预防状况的发作。使用者可以是，例如，希望监测其自身健康的人。使用者可以是，例如，卫生保健提供者。卫生保健提供者可以是，例如，医师。在一些实施方案中，使用者是照顾受试者的卫生保健提供者。可以是本发明的使用者的医师和卫生保健提供者的非限制性实例可包括麻醉师、减肥手术专业人员、血库输血医学专业人员、心脏电生理学医生、心脏外科医生、心脏病科医生、注册护理助理、临床心脏电生理学专业人员、临床神经生理学专业人员、临床护理专业人员、结直肠外科医生、重症护理医学专业人员、重症护理外科手术专业人员、牙科保健员、牙科医生、皮肤科医师、急诊医疗技师、急诊医学医师、胃肠外科医生、血液病医生、临终关怀和姑息医学专业人员、顺势疗法专业人员、传染病专业人员、内科医生、颌面外科医生、医务助理、体检医生 (medical examiner)、医学遗传学家、医疗肿瘤学家、助产师、新生儿-围产期专业人员、肾病医生、神经科医生、神经外科医生、核医学专业人员、护士、开业护士、产科医师、肿瘤科医生、口腔外科医生、口腔正畸医师、整形外科专业人员、疼痛控制专业人员、病理学医师、儿科医师、灌注师、牙周病医师、整形外科医生、足病医生、直肠科医生、假肢专业人员、精神科医生、肺脏科医生、放射科医生 (radiologist)、外科医生、胸科专业人员、移植专业人员、血管专业人员、血管外科医生和兽医。利用本发明的阵列和方法确定的诊断可并入受试者的病历中。

#### [0180] 试剂盒

[0181] 本发明的装置可包装成试剂盒。在一些实施方案中，试剂盒包括关于该装置的使用的书面说明。该书面材料可以是，例如，标签。该书面材料可以建议状况和施用方法。该说明为受试者和指导医师提供达到施用疗法的最佳临床结果的最佳指导。

### 实施例

#### [0182] 实施例1：作为癌症诊断方法的免疫特征分析

[0183] 以下实施例显示了用本发明的示例性阵列诊断癌症的方法。该实施例描述了两个测试：测试#1和测试#2，这两个测试在从多个不同位点、从多名受试者采集的生物样品上测试了本发明的方法。

#### [0184] 肽阵列

[0185] 将10,000个包含不同序列的非天然序列肽的两个不同文库打印在两个不同的肽阵列上。肽阵列#1包含10,420个肽并在测试#1上进行了实验测试。肽阵列#2包含10,286个肽并在测试#2上进行了实验测试。

[0186] 打印文库1，使得在一个载玻片上可进行两个完整的测试，但对每个测试而言每个序列仅有一个肽可用。将文库1载玻片划分成两个物理上隔开的区室，用柔性垫圈 (Agilent, Santa Clara, CA) 隔开各个区室。文库2打印有重复的肽，但每个载玻片仅可进行一个测试。

[0187] 用于测试#1的肽由Sigma Genosys (St. Louis, MO) 合成，而用于测试#2的肽由Alta Biosciences (Birmingham, UK) 合成，且在胺末端 (测试#1) 或羧基末端 (测试#2) 上具有共同的GSC连接体，其后为17个完全随机化的氨基酸。

[0188] 由Applied arrays (Tempe, AZ) 使用非接触式压电打印将阵列打印到涂覆有氨基硅烷的载玻片 (Schott, Jena, Germany) 上。使阵列与封闭缓冲液 (BB=10nM磷酸盐缓冲盐水

(pH 7.3)和0.5%BSA[Sigma Aldrich],0.5%吐温)预孵育1小时,之后添加血清在样品缓冲液(SB=BB减去0.5%吐温)中的1:500稀释液,在25℃下保持1小时。在25℃下,将载玻片在SB中暴露于5nM的AlexaFluor 647-缀合的抗人第二抗体(Rockland Antibodies, Gilbertsville,PA)1小时并且在SB中洗涤3次,继而在180mM水中洗涤5次,随后在1800g下离心5'至干。在Agilent'C'扫描仪中采用高激光功率和70%PMT以10μm分辨率在647nm下对阵列进行扫描。将TIFF图像与将肽名称与强度联系起来的相应gal文件进行比对。

#### [0189] 研究设计和生物样品

[0190] 设计有对照的实验以测试用于癌症诊断的免疫特征系统。测试#1采用经典的训练/盲测范式(train/blinded test paradigm)检查从每种疾病的2-3个不同组群中采集的少数样品。测试#2为经多年处理的大量疾病样品的交叉验证,该疾病样品由来自大量采集地点的不平衡且多样化的组群构成。

[0191] 测试#1的研究设计和生物样品:使用五种不同癌症中的每一种的20个非盲训练样品加上20个其他健康对照的三份技术重复物来构建盲测-训练测试。使用相同的随机选择方法但只选择仍为盲(blinded)的样品来产生同等大小的测试组群。将采集地点、采集日期、年龄和性别随机化。样品是来自在-20℃储存不同时间长度的各为2-10ml的静脉抽取物的血清或血浆。样品在表2中进一步描述。样品从多个不同地点采集,这些地点缩写如下:ASU:Arizona State University collection, Tempe, AZ; BNI: Barrow Neurological Institute, St. Joseph's Hospital and Medical Center, Phoenix, AZ; CC: Cleveland Clinic, Cleveland, OH; FHCRC: Fred Hutchison Cancer Research Center, Seattle, WA; MSKCC: Memorial Sloan-Kettering Cancer Center, New York, NY; MMRF: Multiple Myeloma Research Foundation, Norwalk, CT; MS: Mt. Sinai Hospital, New York, NY; PCRT: Pancreatic Cancer Research Team, Phoenix, AZ; UTSW: University of Texas Southwestern Medical Center, Dallas, TX; UCI: University of California Irvine, Irvine, CA; UPitt: University of Pittsburgh Dept. of Immunology, Pittsburgh, PA; UW: University of Washington Medical Center, Seattle, WA。在表2中,合作者通常在不同地点进行采集。缩写对应于合作者所来自的地方。

疾病, 健康状态	训练	测试	采集地点
健康对照	20	20	ASU, PCRT, FHCRC, UTSW
多形性胶质母细胞瘤	20	20	BNI
胰腺癌	20	20	CC, PCRT, UW
肺癌	20	20	FHCRC
多发性骨髓瘤	20	20	MMRC
乳腺癌	20	20	ASU, FHCRC, UTSW

[0192] [0193] 在测试#1中对来自患有晚期胰腺癌(PC)、首次治疗的多形性胶质母细胞瘤(therapy-naïve Glioblastoma multiformae) (星形细胞瘤的侵袭性形式) (GBM)、食管腺癌(EC)、多发性骨髓瘤(MM)和IV期乳腺癌(BC)的患者的二十个随机选择的血清样品以及20个混合的“非疾病”对照(表2)进行了测试。表2描述了在诊断时记录的并用于分类的原发疾病状态。为分类的目的,忽略任何报告的副发病变。

[0194] 测试#2的研究设计和生物样品:在测试#2中,在2007年9月与2011年1月间,对来自10个不同合作者的2118个样品进行免疫特征分析。就每种疾病的患者数、年龄、性别、种族、报告的副发病变和构成“非疾病”组群的对照数目而言,在本测试中所分析的血清库是固有地不平衡的。未对其技术重复物的皮尔森相关系数(Pearson's correlation coefficient) $<0.85$ 的独立阵列进行分析。使用ComBat来分析其余阵列的阵列批次偏性。1516个样品被认为可用于该测试。

[0195] 表3是对在测试#2中使用的1516个样品的描述。对于在第1列中列出的各个疾病状态,在第2列疾病组群中列出了可用样品数。100倍重采样方法选择各个疾病的约1/4的样品用于训练。第3列训练大小中列出了训练组群的平均值和标准差。第4列合作者中列出了提供样品的合作者的机构名称。

[0196]

疾病, 状态	疾病组群	训练大小	合作者
健康对照	249	62 ± 4	UCI
继发性乳腺癌	61	15 ± 1	BNI
II、III 期乳腺癌	141	35 ± 3	ASU, FHCRC, UTSW, UCI
IV 期乳腺癌	42	11 ± 1	UTSW
星形细胞瘤	166	42 ± 3	Barrow Neurological Institute
多形性胶质母细胞瘤	27	7 ± 1	ASU, BNI, CC, FHCRC, MSKCC, PCRT, UTSW, UCI, UPitt, UW
肺癌	107	25 ± 2	FHCRC
多发性骨髓瘤	112	28 ± 2	MMRC
少突神经胶质瘤	48	12 ± 1	BNI
混合的少突神经胶质瘤 / 星形细胞瘤 (Oligo/Astro)	97	25 ± 2	BNI
卵巢癌	86	22 ± 2	MS, MSKCC
胰腺炎	82	20 ± 1	CC, UW
胰腺癌	136	34 ± 3	CC, UW
尤因氏肉瘤	20	5 ± 0	ASU
溪谷热	142	36 ± 3	UA

[0197] 测试#1

[0198] 测试#1显示采用本发明的方法对多种癌症的同时、高准确度分类。测试#1描述了采用来源于多个采集地点的相等数目的训练和测试样品的有对照的实验(表2)。对来自患有晚期胰腺癌(PC)、首次治疗的多形性胶质母细胞瘤(GMB)、食管腺癌(EC)、多发性骨髓瘤(MM)和IV期乳腺癌(BC)的患者的二十个血清样品以及二十个混合的“非疾病”对照进行了测试,这些样品在不同地点采集。

[0199] 训练集中全部120个样品的两种技术重复物间的平均皮尔森相关系数为 $0.92 \pm 0.05$ 。乳腺癌显示了最低平均重复物相关性(0.87),而食管癌为最高(0.96)。为了估测每种疾病与健康对照间的差异的大小和一致性,我们在 $N=20$ 的癌症组与 $N=20$ 的对照组的每一个之间逐个地进行了T-检验。表4中列出了 $p < 9.6 \times 10^{-5}$ (允许一个FP)的肽的数目以及最小p

值的绝对值。

[0200] 表4汇总了测试#1肽的T-检验统计分析的结果。使用T-检验来比较每种疾病的20个训练样品与20个对照。第1列列出了疾病组群。第2列列出了 $p$ 值 $<9.6 \times 10^{-5}$  (对应于1FP/10,480个肽)的肽的数目。第3列为该比较的最小 $p$ 值。第4列为前100个最显著的肽中与来自至少一种其他疾病的肽相重叠的肽的数目。乳腺癌与任何其他疾病均没有重叠,而GMB与来自其他3种疾病的肽重叠。

疾病	$p < 9.6 \times 10^{-5}$ 的肽的数目	用于比较的最小 $p$ 值	共同的肽/100
健康的	NA	NA	NA
[0201] 乳腺癌	608	$1.54 \times 10^{-14}$	0
食管癌	3103	$4.8 \times 10^{-25}$	14
GBM	3596	$9.05 \times 10^{-30}$	26
骨髓瘤	4478	$3.52 \times 10^{-34}$	19
胰腺癌	1126	$3.67 \times 10^{-11}$	12

[0202] 当仅使用来自 $\text{FWER} = 5\%$ 的T-检验的肽时,使用支持向量机(SVM)作为分类器可能实现分成疾病与健康者的完美的二元分类。然而,这并没有解决多重疾病分类性能的问题。表4的最右列示出了当对每种疾病的100个最显著的T-检验肽进行比较时,与至少一种其他疾病相重叠的肽的数目。一些疾病具有比其他疾病更大的肽重叠。

[0203] 为了改善对多种疾病进行分类的能力,将过滤器应用于具有重叠特异性的肽。首先,将ANOVA/ $\text{FWER} = 0.05\%$ 应用于训练集,以鉴定与这六类中的每一类的总平均值显著不同的4,620个肽。其次,使用GeneSpring 7.3.1中的模式匹配来除去在多于一种疾病中具有高信号的肽。因此,每种疾病选择24个肽,总计120个肽作为最终的特征集。胰腺癌和乳腺癌具有相对较低的整体信号,食管癌和脑癌具有高得多的信号,但该选择方法阻止分类器被具有较强信号和许多显著肽的疾病所压倒。当使用支持向量机(SVM)时,训练集的弃一法交叉验证产生了两个误判(miscall)。然后,使用这120个肽对测试数据集进行分类,得到示于表5中的得分。

疾病 (SVM)	乳腺癌(BC)	脑癌(BC)	食管癌(EC)	多发性骨髓瘤	非疾病	胰腺癌
[0204] 乳腺癌	20	0	0	0	0	2
脑癌	0	19	1	0	0	0
食管癌	0	0	19	0	0	0
多发性骨髓瘤	0	1	0	20	0	0
非疾病	0	0	0	0	20	2
胰腺癌	0	0	0	0	0	16
灵敏度	1	0.95	0.95	1	1	0.80
特异性	0.98	0.99	1	0.99	0.98	1
PPV	0.91	0.95	1	0.95	0.91	1
NPV	1	0.99	0.99	1	1	0.96
患病率	0.17	0.17	0.17	0.17	0.17	0.17
检出率	0.17	0.16	0.16	0.17	0.17	0.13
检出患病率	0.18	0.17	0.15	0.18	0.18	0.13

## [0205] 阵列数据分析

[0206] 对于测试#1,对三种技术重复物取平均,未对生物重复物取平均。重新处理未能达到皮尔森相关系数 $>0.85$ 的任何技术重复物。对数据进行中值归一化和 $\log_{10}$ 转化以用于线型图的可视化。使用ANOVA进行用于分类的肽的初始选择,并使用设置为5%的族错误率(Family Wise Error Rate,FWER)来针对多重检验校正T-检验。使用GeneSpring中的“表达谱(Expression Profile)”以欧几里德距离/平均连接(Average Linkage)作为相似性度量来进行肽的进一步过滤。对于这种过滤器,将每个疾病组(“疾病”)与所有其他疾病组(统称为“非疾病”)相比较。选择在“疾病”中具有一致的高信号且在“非疾病”中具有一致的低信号的肽,确保在“疾病”与“非疾病”信号之间有 $>3$ 倍的平均差异。对于多疾病分类,每种疾病同等数目的肽(特征)阻止高平均信号使特征选择具有偏性;然而,未进行进一步的数据预处理,以确保分类性能依赖于近乎原始的值。

[0207] 使用支持向量机(SVM)作为分类器在2.6.2版R中进行分类。使用支持向量机(SVM)的测试#1的误分类得分示于表6中。正确和错误的判定在灰色区域中列出,性能统计数据在白色区域中列出。平均准确度为0.95,95%CI=0.8943,0.9981, $\kappa=0.94$ 。正确判定在特征向量中,给定类别的任何误判在另一类别中产生假阳性。

[0208]

疾病 (SVM)	乳腺癌(BC)	脑癌 (BC)	食管癌(EC)	多发性骨髓瘤	非疾病	胰腺癌
乳腺癌	20	0	0	0	0	2
脑癌	0	19	1	0	0	0
食管癌	0	0	19	0	0	0
多发性骨髓瘤	0	1	0	20	0	0
非疾病	0	0	0	0	20	2
胰腺癌	0	0	0	0	0	16
灵敏度	1	0.95	0.95	1	1	0.80
特异性	0.98	0.99	1	0.99	0.98	1
PPV	0.91	0.95	1	0.95	0.91	1
NPV	1	0.99	0.99	1	1	0.96
患病率	0.17	0.17	0.17	0.17	0.17	0.17
检出率	0.17	0.16	0.16	0.17	0.17	0.13
检出患病率	0.18	0.17	0.15	0.18	0.18	0.13

[0209] 图1示出了如通过本文所述的聚类和分类方法所确定的组间和组内相对差异的视觉表示。在图1中示出的定量差异在表6中描述。左上:来自PCA的前两个主分量(未用于分类,仅用于显示)绘制在X轴和Y轴上。将来自测试数据集的20个样品用以下疾病标记:BC=乳腺癌;EC=食管癌;N=健康供者;PC=胰腺癌;MM=多发性骨髓瘤;BrC=GMB脑癌。右上:来自LDA的前两个线性判别量绘制在X轴和Y轴上,疾病缩写如上所述。左下:在选择后留下的两个支持向量绘制在X轴和Y轴上。右下:两个单纯贝叶斯(Naïve Bayes)预测变量绘制在X轴和Y轴上。

[0210] 图2示出了通过使用欧几里德距离与平均连接来评估节点的分裂式分级聚类获得的120个肽(Y轴)和120名患者(X轴)的热图。

[0211] 这种层级结构在左侧的彩色系统树图(dendrogram)中明确描绘。肽的k-均值聚类(其中 $k=5$ 个类别,编号为1至V)的结果示于每张热图的右侧。非癌症对照未用于选择“非疾

病”肽,因此存在五组肽和六组患者。A图示出了使用120个选定特征的训练数据集的热图。B图示出了使用相同的120个肽进行聚类的未混合的测试数据。注意,肽类别编号遵循k-均值着色,但肽被重新聚类。

[0212] 测试#2

[0213] 测试#2测试免疫特征是否能对十四种不同的疾病(包括乳腺癌的三种亚型)进行分类。使用1536个样品产生一组255个判别肽。在交叉验证测试中,免疫特征为98%的准确率。表3描述了在测试#2中分析的1516-样品组群中使用的样品。如表4所述,对于每个疾病组与对照组,选择100个T-检验肽。

[0214] 对于测试#2,使用了用于提供分类性能的无偏估计的重采样方法。以下程序重复100次;结果为100次不同训练/测试重复的平均值。首先,取出每种疾病的样品中的25%±7%且不替换,并用作特征选择的训练。特征选择每次挑选总计恰好255个肽。组群大小的7%变化模拟疾病患病率和/或样品可用性的自然变化。通过使用从训练中选出的255个特征对剩余的约75%的样品进行分类来进行交叉验证。对所有统计学评价计算95%置信区间。测试#2使用如在测试#1中应用的支持向量机(SVM)。

[0215] 表7显示了由100次重采样和重分析得到的具有95%置信区间的LDA、NB和SVM分类结果。如果所预测的疾病作为预测出现在任何其他疾病类别中,则该预测被评分为假阳性,而如果错过了正确的类别,则该预测被评分为假阴性。鉴于测试#1的高准确度,甚至具有高固有患者变异性的小组群也允许使用线性超平面进行准确的免疫特征分析,该线性超平面优化从任何训练点到该平面的距离。

[0216]

疾病 / (LDA)	准确度	灵敏度	特异性	PPV	NPV
继发性 BC	97.8 ± 0.14	69.1 ± 2.82	99.21 ± 0.1	81.05 ± 3.46	98.48 ± 0.11

[0217]

星形细胞瘤 (Astro)	96.93 ± 0.17	90.1 ± 1.3	97.82 ± 0.17	83.79 ± 3.46	98.73 ± 0.18
BC	99.51 ± 0.05	99.71 ± 0.2	99.49 ± 0.08	95.45 ± 0.68	99.97 ± 0.18
BCIVa	99.62 ± 0.06	89.85 ± 1.49	100 ± 0	100 ± 0	99.6 ± 0.06
GBM	99.18 ± 0.1	94.33 ± 2	99.25 ± 0.09	62.1 ± 4.24	99.92 ± 0.03
肺癌	99.02 ± 0.12	92.37 ± 0.58	99.59 ± 0.09	94.79 ± 1.27	99.35 ± 0.05
MM	98.72 ± 0.11	100 ± 0	98.62 ± 0.12	85.13 ± 1.13	100 ± 0
ND	96.62 ± 0.17	85.45 ± 0.77	99.31 ± 0.1	96.66 ± 0.47	96.6 ± 0.23
少突神经胶质 瘤(Oligo)	99.65 ± 0.17	92.57 ± 1.95	99.86 ± 0.03	95.21 ± 1.19	99.78 ± 0.06
少突神经胶质 瘤/星形细胞瘤 (OligoAstro)	98.94 ± 0.15	98.45 ± 0.82	98.95 ± 0.12	86.41 ± 1.78	99.91 ± 0.04
卵巢癌	99.92 ± 0.03	100 ± 0	99.91 ± 0.03	98.67 ± 0.47	100 ± 0
胰腺炎	99.67 ± 0.05	95.42 ± 1	99.91 ± 0.03	98.5 ± 0.54	99.74 ± 0.05
PC	97.69 ± 0.11	86.61 ± 1.39	98.79 ± 0.08	87.22 ± 1.19	98.67 ± 0.12
肉瘤	98.81 ± 0.11	54.15 ± 5.48	99.67 ± 0.07	71.55 ± 5.65	99.12 ± 0.12
VF	99.67 ± 0.08	100 ± 0	99.64 ± 0.09	96.87 ± 0.74	100 ± 0
总计	98.77 ± 0.04	89.87 ± 1.32	99.33 ± 0.08	88.89 ± 1.59	99.33 ± 0.07
<b>疾病 / (NB)</b>	<b>准确度</b>	<b>灵敏度</b>	<b>特异性</b>	<b>PPV</b>	<b>NPV</b>
继发性 BC	96 ± 0.16	56.07 ± 1.46	99.46 ± 0.07	90.37 ± 11.68	96.31 ± 0.15
星形细胞瘤 (Astro)	91.92 ± 0.23	91.96 ± 1.25	91.91 ± 0.25	31.39 ± 10.61	99.66 ± 0.06
BC	98.78 ± 0.07	97.75 ± 0.46	98.91 ± 0.12	90.55 ± 9.81	99.73 ± 0.06
BCIVa	99.4 ± 0.09	84.48 ± 2.05	100 ± 0	100 ± 0	99.38 ± 0.09
GBM	96.08 ± 0.1	43.19 ± 2.17	99.72 ± 0.05	88.81 ± 16.61	97.04 ± 0.19
肺癌	99.08 ± 0.1	92.4 ± 0.89	99.74 ± 0.06	97.32 ± 6.18	99.25 ± 0.08
MM	96.45 ± 0.15	81.51 ± 2.07	97.76 ± 0.14	75.72 ± 11.16	98.38 ± 0.2
ND	95.84 ± 0.17	93.18 ± 0.62	96.41 ± 0.18	83.88 ± 7.21	98.54 ± 0.14
星形细胞瘤 (Astro)	98.54 ± 0.14	74.38 ± 2.24	99.94 ± 0.03	98.56 ± 5.95	98.85 ± 0.09
少突神经胶质 瘤/星形细胞瘤 (OligoAstro)	97.75 ± 0.15	86.11 ± 0.86	98.72 ± 0.13	84.75 ± 13.01	99.87 ± 0.04
卵巢癌	99.79 ± 0.05	98.48 ± 0.43	99.9 ± 0.03	98.81 ± 3.75	99.45 ± 0.11
胰腺炎	99.3 ± 0.11	92.27 ± 1.49	99.8 ± 0.05	97.4 ± 5.82	97.13 ± 0.17
PC	95.91 ± 0.2	78.67 ± 0.96	98.26 ± 0.09	85.62 ± 7.73	96.69 ± 0.21
肉瘤	96.73 ± 0.2	25.21 ± 1.44	100 ± 0	100 ± 0	99.73 ± 0.07
VF	97.96 ± 0.22	97.48 ± 0.6	97.99 ± 0.22	84.63 ± 12.45	98.57 ± 0.12
总计	97.35 ± 0.15	79.52 ± 1.27	98.57 ± 0.1	87.19 ± 8.13	98.57 ± 0.12
<b>疾病 / (SVM)</b>	<b>准确度</b>	<b>灵敏度</b>	<b>特异性</b>	<b>PPV</b>	<b>NPV</b>
继发性 BC	98.89 ± 0.03	91.04 ± 0.59	99.19 ± 0.04	81.16 ± 8.55	99.65 ± 0.03
星形细胞瘤 (Astro)	97.12 ± 0.06	84.11 ± 0.31	98.93 ± 0.03	91.96 ± 2.18	97.82 ± 0.06
BC	99.78 ± 0.02	99.39 ± 0.13	99.82 ± 0.02	98.4 ± 1.34	99.93 ± 0.01
BCIVa	99.89 ± 0.02	96.26 ± 0.75	100 ± 0	100 ± 0	99.88 ± 0.02
GBM	99.08 ± 0.03	100 ± 0	99.07 ± 0.03	46.42 ± 21.1	100 ± 0

[0218]

肺癌	99.73 ± 0.02	96.82 ± 0.18	99.97 ± 0.01	99.65 ± 1.12	99.73 ± 0.02
MM	99.58 ± 0.01	99.89 ± 0.08	99.55 ± 0.01	94.7 ± 1.19	99.99 ± 0.01
ND	98.13 ± 0.07	91.33 ± 0.35	99.7 ± 0.02	98.6 ± 0.81	98.03 ± 0.09
少突神经胶质瘤(Oligo)	99.82 ± 0.01	94.76 ± 0.3	99.96 ± 0.01	98.67 ± 3.38	99.85 ± 0.01
少突神经胶质瘤/星形细胞瘤(OligoAstro)	99.29 ± 0.03	100 ± 0	99.24 ± 0.03	89.66 ± 4.09	100 ± 0
卵巢癌	99.92 ± 0.01	98.7 ± 0.1	100 ± 0	100 ± 0	99.92 ± 0.01
胰腺炎	99.73 ± 0.02	96.27 ± 0.27	99.94 ± 0.01	99.07 ± 1.58	99.77 ± 0.02
PC	98.62 ± 0.03	90.98 ± 0.21	99.45 ± 0.02	94.74 ± 2.16	99.02 ± 0.02
肉瘤	99.19 ± 0.04	100 ± 0	99.18 ± 0.03	38.81 ± 31.06	100 ± 0

[0219] 低置信区间表明,线性和概率性分类器对大量的不平衡类别或大量的肽特征均没有特别的偏向。如在测试#1中,以SVM观察各个类别相对于正常样品的分离。T-检验肽的重叠产生至少一个、平均15个与至少一种其他疾病相重叠的肽。没有一组T-检验肽不包含至少一个与至少一种其他疾病相重叠的肽。

[0220] 表8。表8中的值包含针对各个分类器(PCA、NB和LDA以及k-NN)进行的实际判定。判定按预测(列标题)与真实疾病(行)列出,使得第1列第1行包含由PCA分类器正确鉴定的判定数。第1列第2行包含分类器将乳腺癌(BC)鉴定为脑癌(BrC)的次数。该表中包括多种分类器,以确保没有分类算法产生严重不符的判定。

疾病 / (PCA)	BC	BrC	EC	MM	ND	PC
乳腺癌	15	0	1	1	8	2
脑癌	0	7	0	5	1	3
食管癌	0	0	14	2	5	0
多发性骨髓瘤	2	11	3	11	0	5
非疾病	3	0	1	1	5	3
胰腺癌	0	2	1	0	1	7
灵敏度	0.75	0.35	0.70	0.55	0.25	0.35
特异性	0.88	0.91	0.93	0.79	0.92	0.96
PPV	0.56	0.44	0.67	0.34	0.38	0.64
NPV	0.95	0.88	0.94	0.90	0.86	0.88
患病率	0.17	0.17	0.17	0.17	0.17	0.17
检出率	0.13	0.06	0.12	0.09	0.04	0.06
检出患病率	0.23	0.13	0.18	0.27	0.11	0.09

[0221]

[0222]

疾病/ (NB)	BC	BrC	EC	MM	ND	PC
乳腺癌	13	0	0	0	0	0
脑癌	0	19	0	4	0	0
食管癌	0	0	20	0	9	0
多发性骨髓瘤	0	1	0	16	0	0
非疾病	0	0	0	0	10	1
胰腺癌	7	0	0	0	1	19
灵敏度	0.65	0.95	1	0.80	0.50	0.95
特异性	1	0.96	0.91	0.99	0.99	0.92
PPV	1	0.83	0.69	0.94	0.91	0.70
NPV	0.93	0.99	1	0.96	0.91	0.99
患病率	0.17	0.17	0.17	0.17	0.17	0.17
检出率	0.11	0.16	0.17	0.13	0.08	0.16
检出患病率	0.11	0.19	0.24	0.14	0.09	0.23
疾病/ (LDA)	BC	BrC	EC	MM	ND	PC
乳腺癌	20	0	0	0	1	3
脑癌	0	16	0	1	0	0
食管癌	0	0	20	0	0	0
多发性骨髓瘤	0	4	0	19	0	0
非疾病	0	0	0	0	19	2
胰腺癌	0	0	0	0	0	15
灵敏度	1	0.80	1	0.95	0.95	0.75
特异性	0.96	0.99	1	0.96	0.98	1
PPV	0.83	0.94	1	0.83	0.91	1
NPV	1	0.96	1	0.99	0.99	0.95
患病率	0.17	0.17	0.17	0.17	0.17	0.17
检出率	0.17	0.13	0.17	0.16	0.16	0.13
检出患病率	0.20	0.14	0.17	0.19	0.18	0.13
疾病 / (k-NN)	BC	BrC	EC	MM	ND	PC
乳腺癌	20	0	0	0	0	4
脑癌	0	17	0	0	0	0
食管癌	0	0	20	0	0	0
多发性骨髓瘤	0	3	0	20	0	0
非疾病	0	0	0	0	20	3
胰腺癌	0	0	0	0	0	13
灵敏度	1	0.85	1	1	1	0.65
特异性	0.96	1	1	0.97	0.97	1
PPV	0.83	1	1	0.87	0.87	1
NPV	1	0.97	1	1	1	0.93
患病率	0.17	0.17	0.17	0.17	0.17	0.11
检出率	0.17	0.14	0.17	0.17	0.17	0.11
检出患病率	0.20	0.14	0.17	0.19	0.19	0.11

[0223] 图3为描绘了255个分类器肽在1516个患者样品中的热图,并在括号内列出了组群大小。颜色区分低(蓝色)强度与高(红色)强度,而且在肽(Y轴)和患者(X轴)均分级聚类后

仍保留的模式有助于将疾病组群之内和之间的相对差异可视化。已知患有副发病变的患者未被排除,且对照样品显示出高度多样化的信号。

[0224] 图4示出了选自255个分类器肽的选定肽的表现。一些肽对特定的癌症具有高度选择性,并完全有助于分类的准确度。许多肽具有不完美的疾病内一致性。另外一些肽对多于一种疾病为高。对于每个分类算法,针对每种疾病绘制单独的受试者工作特征(ROC)曲线并且计算曲线下面积(AUC)值。针对SVM的AUC以灰色显示。A图为乳腺癌的ROC曲线的图形表示。B图为脑癌的ROC曲线的图形表示。C图为食管癌的ROC曲线的图形表示。D图为多发性骨髓瘤的ROC曲线的图形表示。E图为健康对照的ROC曲线的图形表示。F图为胰腺癌的ROC曲线的图形表示。

[0225] 图5为测试#1的受试者工作特征(ROC)曲线的图形表示。PCA的曲线下面积(AUC)以灰色显示。

[0226] 图6为测试#1的受试者工作特征(ROC)曲线的图形表示。NB的曲线下面积(AUC)以灰色显示。

[0227] 图7为测试#1的受试者工作特征(ROC)曲线的图形表示。LDA的曲线下面积(AUC)以灰色显示。

[0228] 图8为测试#1的受试者工作特征(ROC)曲线的图形表示。k-NN的曲线下面积(AUC)以灰色显示。

[0229] 图9汇总了四种分类器,PCA、LDA、NB和k-NN,它们可产生相关分类性能的图形解读,如在图1中针对SVM所示的。A图为PCA的图形表示,对前两个主分量作图。B图为LDA的图形表示,X轴和Y轴描绘了前两个线性判别量。C图为NB的图形表示,对预测变量作图。D图为k-NN的图形表示,对组间距离作图。

[0230] 图10为来自测试#2的255个分类器肽中的3个分类器肽的线型图。该强度谱在X轴上显示了个体,疾病用空格隔开,并且在Y轴上显示了各个肽的 $\log_{10}$ 强度。显示了特异性的三个实例。A图示出了对疾病6和9为高而对所有其他疾病为低的肽的线型图。这增强了针对其他9种疾病的特异性,但在疾病6和9之间产生了可能的误读。B图示出了对疾病11为高且比任何其他疾病平均高9倍的肽。虽然疾病3、5和6具有大的变化,但疾病11高度一致并增强了对疾病11的特异性。C图示出了对疾病1和部分疾病9为高的肽。在组群内不同但为疾病特异性的肽并没有负面地影响针对该疾病的特异性,但可能影响灵敏度。鉴于在疾病1内的相对高信号,这种肽在仅区别疾病9时仅取得中度成功,但在针对疾病2-8和10-11进行判别时是非常成功的。

[0231] 作为健康监测方法、诊断方法、治疗方法和预防性保健方法的免疫特征分析

[0232] 在疾病的诊断、健康监测、治疗和预防中所面临的挑战是样品组群的变异性、不同的血液采集方法和对样品进行的冻融循环。测试#1和测试#2表明本发明能够克服这些挑战,并且测试#1和测试#2显示了采用本发明方法在宽范围的受试者中具有状况分类的高特异性。测试#1和测试#2还表明免疫特征分析可用于高容量样品处理,从而允许在发现阶段使用更多疾病和对照样品。免疫特征分析的这一特征可以克服过度拟合——标准的生物标志物发现所共有的问题。

[0233] 测试#1和测试#2表明了免疫特征分析是一种能够对标准训练、盲测测试中不同类型的癌症进行高准确度分类的方法。阵列中肽的数目的变化、阵列中肽的接近度的优化以

及阵列中分子的类型的变化可使免疫特征分析成为一种用于多种不同健康状态的健康监测、诊断、治疗和预防的有力方法。

#### [0234] 实施例2:与免疫特征系统一起使用的计算机架构

[0235] 由本发明的阵列检测到的数据可通过具有不同计算机架构的多台计算机进行分析。图11为示出了可与本发明的示例实施方案结合使用的计算机系统1100的第一示例架构的框图。如图11所示,示例计算机系统可包括用于处理指令的处理器1102。处理器的非限制性实例包括: Intel Core i7™处理器、Intel Core i5™处理器、Intel Core i3™处理器、Intel Xeon™处理器、AMD Opteron™处理器、Samsung 32-bit RISC ARM 1176JZ (F) -S v1.0™处理器、ARM Cortex-A8 Samsung S5PC100™处理器、ARM Cortex-A8 Apple A4™处理器、Marvell PXA930™处理器或功能等同的处理器。多线程执行可用于并行处理。在一些实施方案中,可使用多个处理器或多核处理器,无论是在单一计算机系统中,在集群中,还是分布在经由网络包含多台计算机、蜂窝电话和/或个人数据助理设备的系统中。

#### [0236] 数据采集、处理和存储

[0237] 如图11所示,高速缓冲存储器1101可连接到或并入处理器1102,以便为处理器1102最近使用或频繁使用的指令或数据提供高速存储。处理器1102通过处理器总线1105连接到北桥1106。北桥1106通过存储器总线1104连接到随机存取存储器 (RAM) 1103,并且管理处理器1102对RAM 1103的访问。北桥1106还通过芯片组总线1107连接到南桥1108。南桥1108又连接至外围总线1109。例如,外围总线可以是PCI、PCI-X、PCI Express或其他外围总线。北桥和南桥通常被称作处理器芯片组并管理处理器、RAM和外围总线1109上的外围组件之间的数据传送。在一些架构中,北桥的功能可并入处理器,而不是使用单独的北桥芯片。

[0238] 在一些实施方案中,系统1100可包括连接至外围总线1109的加速器卡1112。加速器可包括现场可编程门阵列 (FPGA) 或用于加速某些处理的其他硬件。

#### [0239] 软件接口

[0240] 软件和数据被存储在外部存储1113中,并且可加载到RAM 1103和/或高速缓冲存储器1101中以供处理器使用。系统1100包括用于管理系统资源的操作系统;操作系统的非限制性实例包括:Linux、Windows™、MACOS™、BlackBerry OS™、iOSTM和其他功能等同的操作系统,以及在操作系统之上运行的应用软件。

[0241] 在本实施例中,系统1100还包括连接到外围总线的网络接口卡 (NIC) 1110和1111,以用于提供通向外部存储如网络附加存储 (NAS) 和其他可用于分布式并行处理的计算机系统的网络接口。

#### [0242] 计算机系统

[0243] 图12为示出了具有多个计算机系统1202a和1202b、多个蜂窝电话和个人数据助理1202c以及网络附加存储 (NAS) 1201a和1201b的网络1200的示意图。在一些实施方案中,系统1202a、1202b和1202c可管理数据存储并优化对存储在网络附加存储 (NAS) 1201a和1202b中的数据的数据存取。数学模型可用于该数据,并使用在计算机系统1202a和1202b以及蜂窝电话和个人数据助理系统1202c中的分布式并行处理进行评估。计算机系统1202a和1202b以及蜂窝电话和个人数据助理系统1202c还可提供并行处理,以用于在网络附加存储 (NAS) 1201a和1202b中存储的数据的适应性数据重构。图12仅示出了一个实例,而很多种其他计算机架构和系统可与本发明的各个实施方案一起使用。例如,刀片式服务器可用于提

供并行处理。处理器刀片可通过底板连接以提供并行处理。存储还可连接至底板或通过单独的网络接口作为网络附加存储 (NAS) 连接。

[0244] 在一些实施方案中,处理器可保持单独的存储空间并通过网络接口、底板或其他连接器传输数据以便由其他处理器进行并行处理。在一些实施方案中,一些或所有处理器可使用共享虚拟地址存储空间。

#### [0245] 虚拟系统

[0246] 图13为使用共享虚拟地址存储空间的多处理器计算机系统的框图。该系统包括多个处理器1301a-f,所述处理器可访问共享存储器子系统1302。该系统将多个可编程的硬件存储器算法处理器 (MAP) 1303a-f并入存储器子系统1302中。各个MAP 1303a-f可包括存储器1304a-f以及一个或多个现场可编程门阵列 (FPGA) 1305a-f。MAP提供可配置的功能单元,可向FPGA 1305a-f提供特定算法或算法的一部分,以与相应的处理器密切配合地进行处理。在本实施例中,为了这些目的,每个MAP可被所有处理器全局访问。在一种配置中,每个MAP可使用直接存储器存取 (DMA) 来访问相关存储器1304a-f,从而允许其独立于相应的微处理器1301a-f并与之异步地执行任务。在这种配置中,MAP可将结果直接提供给另一个MAP,以用于算法的流水操作 (pipelining) 及并行执行。

[0247] 上述计算机架构和系统仅为示例,且很多种其他计算机、蜂窝电话以及个人数据助理架构和系统可与示例实施方案结合使用,包括使用通用处理器、协处理器、FPGA和其他可编程逻辑设备、片上系统 (system on chip) (SOC)、专用集成电路 (ASIC) 以及其他处理和逻辑元件的任意组合的系统。任何种类的数据存储介质均可与示例实施方案结合使用,包括随机存取存储器、硬盘驱动器、闪存存储器、磁带驱动器、磁盘阵列、网络附加存储 (NAS) 及其他本地或分布式数据存储设备和系统。

[0248] 在示例实施方案中,计算机系统可使用在任一上述或其他计算机架构和系统上执行的软件模块来实现。在其他实施方案中,系统的功能可部分地或完全地在固件、可编程逻辑设备 (例如图13中示出的现场可编程门阵列 (FPGA))、片上系统 (SOC)、专用集成电路 (ASIC) 或其他处理和逻辑元件中实现。例如,集处理器和优化器 (Set Processor and Optimizer) 可通过使用硬件加速器卡,例如图11中示出的加速器卡1112,采用硬件加速来实现。

[0249] 图14示出了具有不同肽密度的本发明的示例性阵列。上述的任何计算机架构均可用于检测、处理及分析免疫特征。

#### [0250] 实施例3:健康监测方法、诊断方法、治疗方法和预防状况的方法

[0251] 可在受试者生命期间的多个时间点监测受试者的健康,包括实施治疗前和实施治疗后。以下的实施例阐述了本发明的方法和示例性阵列在监测6名受试者的健康中的应用。在本文所述的实施例中,用示例性的肽阵列来测试用于诊断、治疗、监测和预防6名受试者中的一名或多名的状况的方法。用约10,000个肽的特定微阵列来进行本实施例中所述的实验。本发明的任何微阵列可与本发明的方法结合使用。

#### [0252] 健康监测

[0253] 在采用一定剂量的流感疫苗对受试者进行处理“之前”和“之后”对多名受试者的健康进行追踪。图15为示出了6名受试者在接受流感疫苗后的一段时间内的免疫特征谱的热图。在图15中,“前”指接种疫苗前1-2周,而“后”可指接种疫苗后21天的时间段内的6个不

同时间点中的一个。在图15中,对6名受试者的免疫特征分析结合模式阐述如下:1)6名去除标识信息的(de-identified)受试者用数字112、113、33、43、73和84表示。2)免疫特征分析结合模式被聚类为:a)受试者112,“红色标签”,接种前、第1天、第5天、第7天、第14天、第21天;b)受试者113,“绿色标签”,接种前、第1天、第5天、第7天、第14天、第21天;c)受试者33,“蓝色标签”,接种前、第1天、第5天、第7天、第14天、第21天;d)受试者43,“橙色标签”,接种前、第1天、第5天、第7天、第14天、第21天;e)受试者73,“浅粉色标签”,接种前、第21天;以及f)受试者84,“黄色标签”,接种前、第1天、第5天、第7天、第14天、第21天。

#### [0254] 生物样品的类型

[0255] 从本实施例所述的六名受试者之一的体内的不同来源采集生物样品。每小时监测其中一名受试者的健康,持续1天。图16的A图示出了在当天的过程中来自同一受试者的不同生物样品的免疫特征分析结合模式的热图。从三处,即两个不同的唾液来源和静脉血采集生物样品。两个唾液采集部位是:a)腮腺,聚类在“黄色标签”上;以及b)下颌样品,聚类在“蓝色标签”上。来自血液的生物样品来源于受试者的静脉血。A图示出了不同的生物样品在11个不同的时间点的聚类的热图。B图为对A图中示出的热图区域的更高分辨率的分析。B图示出了不同生物样品在10,000个肽的阵列中的聚类的差异。

[0256] 可以使用生物样品的其他来源并可用本发明的阵列和方法对其进行检测。

#### [0257] 预防性保健

[0258] 在数月内阶段性追踪其中一名受试者的健康。在这段时间内,受试者在2010年11月25日前报告感觉不适。图17示出了在数月内监测的受试者的免疫特征分析结合模式的热图。A图示出了受试者在2010年11月7日前后的免疫特征分析结合模式中的峰。A图中的免疫特征分析结合模式指示了在受试者报告症状之前的峰,接着是随后的下降。B图显示了在所有10,000个肽中的一致性,且疾病特征隐藏在抗体的正常变化中。这表明:本发明的方法可鉴定在症状出现前与状况相关的免疫特征分析结合模式。

[0259] 在症状出现前与状况相关的结合模式可用于预防状况,包括状况的发作或进展。例如,医师可以在症状出现前开出药方来治疗所鉴定的状况。

#### [0260] 对与阵列结合的独特模式进行检测和聚类

[0261] 有超过一种方法可应用于检测生物样品与阵列结合的模式。我们在这里显示了检测IgM和IgG抗体与本发明阵列的结合模式的应用。

[0262] 用本发明的阵列和方法监测3名受试者的健康。在肽阵列中分析了来自这3名受试者的IgM抗体和IgG抗体的结合模式的检测和聚类。图18示出了3名受试者在21天的时程中在第0天、第1天、第2天、第5天、第7天和第21天的免疫特征分析结合模式的热图。A图示出了当检测到IgM免疫球蛋白的结合时具有约10,000个肽的肽阵列的聚类。B图示出了当检测到IgM免疫球蛋白的结合时具有50个个人肽的肽阵列的聚类。C图示出了当检测到IgG免疫球蛋白的结合时具有约10,000个肽的肽阵列的聚类。D图示出了当检测到IgG免疫球蛋白的结合时具有50个个人肽的肽阵列的聚类。当IgM免疫球蛋白与肽阵列的结合模式采用层级距离进行检测并聚类时,具有10,000个肽的组的阵列没有将个体受试者组织成与他们的血液抽取日期相对应的正确的组(B图)。当IgG免疫球蛋白与阵列的结合模式采用层级距离进行检测并聚类时,受试者的身份和抽血日期被正确地聚类。对于B图,显示了来自双因素ANOVA分析的前50个肽。对于C图,显示了来自双因素ANOVA分析的前50个肽。每个类别与受

试者对应。

[0263] 健康监测

[0264] 在数月内阶段性追踪其中一名受试者的健康。图19为示出了使用免疫特征分析结合模式分析对两名受试者的30天时程分析的热图。该时程包括两名受试者的免疫特征分析结合谱的逐年聚类。

[0265] 其中一名受试者——受试者84,在所述时程的第17天接受一定剂量的流感疫苗。图20为示出了受试者84与22个特定肽序列的免疫特征分析结合谱的热图。图20包括受试者84的免疫特征结合谱的逐年聚类。这22个肽的序列为:SEQ ID NO.1:CSGSYNMDKYFTYSWYREER;SEQ ID NO.2:CSGWDSFRHYERITDRHQGD;SEQ ID NO.3:CSGRYFMHMEPTINHYYEGM;SEQ ID NO.4:CSCVMMPDYRIHVHWSNWTG;SEQ ID NO.5:CSGLRHYNVYDFRSNDRHWA;SEQ ID NO.6:CSGVMAHTGHSGRMGPPDFQ;SEQ ID NO.7:CSGNDHSQHDFAPVESYIMM;SEQ ID NO.8:CSGILFFTRETVDVHYPANEG;SEQ ID NO.9:CSGVPWRSHANQREYAJAN;SEQ ID NO.10:CSGNGVHEFSAMLIMDMIIF;SEQ ID NO.11:CSGIGDHMPLNEPNPLRDLK;SEQ ID NO.12:CSGTHIATNPLNVQYVMVQS;SEQ ID NO.13:CSGTRKEHYLEHVAKHMEVW;SEQ ID NO.14:CSGPTDITELMMRPKYSRIN;SEQ ID NO.15:CSGDQQGTWGRVDMWSNRMH;SEQ ID NO.16:CSGIMKRIHAQTMWYSPITD;SEQ ID NO.17:CSGSFFYVVKQVNNKNYQTI;SEQ ID NO.18:CSGLYAKQVAAQRPIKYWDH;SEQ ID NO.19:CSGMMWYHGYPHVHANDAHW;SEQ ID NO.20:CSGRYHPNYGDAKKHBMSRF;SEQ ID NO.21:CSGHWKGDLSRGRHYHHQEF;和SEQ ID NO.22:CSGEDTRRGHAWKFSEISPH。

[0266] 图21为示出了受试者84在诊断出支气管炎后约20天内的血液样品的免疫特征结合谱的热图。图21显示了生物样品与本发明的14个所选肽的结合模式。这14个肽的序列为:SEQ ID NO.23:CSGWRKILKKRIWTDPTNY;SEQ ID NO.24:CSGYPRSWFVYYTPWKLFGK;SEQ ID NO.25:CSGSHMQDIYRTVRS LGKSM;SEQ ID NO.26:CSGVQLSSYTLKLGKVYQER;SEQ ID NO.27:CSGKTMTTQWRSSLFKFAGM;SEQ ID NO.28:CSGMKYNPFPKYKSYLQYVN;SEQ ID NO.29:CSGISTKFWWKRNSIVFPKL;SEQ ID NO.30:CSGTRGRWYDRRSPSKFLGY;SEQ ID NO.31:CSGQNVSAKYVKGRSVQSWI;SEQ ID NO.32:CSGHIMGRKRHWPMSTSYGV;SEQ ID NO.33:CSGFNKPYVLKYKMDTIHYN;SEQ ID NO.34:CSGYAQQVRYATRFWNKGKY;SEQ ID NO.35:CSGWKHKYHKAAYFHKPFW;和SEQ ID NO.36:CSGWSKPHPKMIARNFFRHL。

[0267] 图22为示出了在图20中表征的受试者在2011年12月11日被症状后诊断为流感的热图。图23为示出了接受肝炎疫苗治疗以及在其后3个月接受第一次加强治疗的受试者的免疫特征分析结合模式的热图。

[0268] 多种传染病的同时鉴定

[0269] 图24显示了采用本发明的方法和阵列对多种传染病的鉴定。图24示出了多种传染病的分类的汇总。A图为示出了登革热、西尼罗病毒(WNV)、梅毒、乙型肝炎病毒(HBV)、正常血液、溪谷热和丙型肝炎病毒的聚类免疫特征分析结合谱的热图。B图为PCA分类的图形表示。

[0270] 实施例4:免疫特征分析系统

[0271] 下述实施例描述了用于免疫特征分析的自动化系统。

[0272] 该自动化系统包括几个组件:1)用于接收、记录和稀释来自受试者的生物样品(如

血液或唾液样品)的自动化系统。该自动化系统使生物样品与本发明的肽微阵列接触。

[0273] 可在对受试者的免疫特征测定中获得受试者的免疫特征分析,该测定由以下自动化步骤组成:a)将稀释的样品施加至肽阵列;b)孵育特定时间;c)除去样品并洗涤该阵列;d)施加第二抗体溶液特定时间;e)用洗涤步骤除去未结合的和/或过量的第二抗体;以及f)干燥并扫描该阵列以测定斑点的荧光。图25为本发明的免疫特征分析系统的组件的示意图。

[0274] 数据采集和分析

[0275] 将阵列进行比对,并相对于标准特征确定特征。标准特征可以是健康受试者的特征或未结合的肽的参考信号。

[0276] 基于利用本发明的系统获得的免疫特征,可以提供诊断。

[0277] 图26的A图示出了噬菌体展示文库。A图示出了以下步骤:a)采用组合合成产生噬菌体文库,b)针对噬菌体展示的随机抗原淘选血清;以及c)选择及测序。B图示出了肽微阵列。

[0278] 实施例5:肽阵列的设计与制造

[0279] 用于肽阵列产生的一组掩模被设计成满足以下标准:

[0280] • 使用18种不同的氨基酸;

[0281] • 阵列中有331,000个肽;

[0282] • 每个肽为10至16个氨基酸的长度;

[0283] • 优化肽序列,以使所呈现的不同五聚体的总数最大化(作为使序列多样性最大化的方式,在阵列上的肽序列内呈现尽可能多的不同的5个氨基酸的序列);

[0284] • 使不超过6%的肽在N-末端具有18种氨基酸中的任意一种;

[0285] • 文库必须能够使用90个掩模(90个光刻步骤)产生。

[0286] 进行下列步骤:

[0287] 用随机数发生器产生一大组(约 $10^{10}$ 个)16个残基的肽序列。

[0288] 使用先前所概述的方法的计算机模拟(参见上文的“制造阵列”),仅使用90个光刻步骤尽可能多地产生 $10^{10}$ 肽组中的各个肽的序列。

[0289] 从由模拟合成产生的肽序列中,仅选择那些长度为10至16个氨基酸的肽。

[0290] 从经过长度选择的肽中,选择为了包括尽可能多的独特五聚体序列(长度为5的氨基酸序列)而优化的肽的亚组。

[0291] 从经过五聚体选择的肽中,选择其中N末端氨基酸组成包含不超过6%的任何特定氨基酸的肽。

[0292] 总计,满足所有上述标准的所选择的最终组的肽序列为331,000个。

[0293] 图31中的图显示了如上所述选择的肽序列的长度分布。Y轴为具有特定长度的肽的数目。该轴从0延伸至100,000。X轴显示了以氨基酸表示的肽的长度。如上述标准所要求的,所有肽为10至16个氨基酸。平均长度为约11.5个氨基酸。

[0294] 图32中示出的图为3个、4个或5个氨基酸长的可能序列的分布。最上方的两幅图显示了三聚体序列(3个氨基酸长的肽)的分布。存在 $18 \times 18 \times 18 = 5832$ 个不同的可能的三聚体序列。左侧显示了如上所述选择的肽的这些三聚体序列的群体分布。右侧显示了使用随机数发生器所产生的肽序列的文库的分布。对于每幅图,X轴描绘了特定三聚体序列在文库

中出现的次数。Y轴描绘了以X轴上表示的次数出现的三聚体序列的数目。因此,可以看出:对于使用随机数发生器产生的肽序列,几乎所有5832个三聚体序列在文库中都呈现400至600次。相比之下,对于左侧的选定的肽,三聚体序列的分布更广,其中一些三聚体序列仅出现约100次而另一些三聚体序列出现超过1000次。所有可能的三聚体序列在文库中呈现多次。

[0295] 中间的图是四聚体序列(4个氨基酸的序列)的图。其中的轴与针对三聚体所描述的轴相似。可以看出:大多数四聚体序列在使用随机数发生器产生的这样大小的文库(右图)中出现约30次。在如上所述选择的肽文库(左图)中,该分布的峰值为约20且宽度大于从用随机数发生器产生的序列中所看到的宽度。存在共计 $18^4=104976$ 个可能的四聚体序列。所有可能的四聚体序列中有99.99%在如上所述选择的肽文库中呈现。

[0296] 最下方的图是五聚体序列(5个氨基酸的序列)的图。其中的轴如针对三聚体序列所述。存在1,889,568个可能的五聚体序列。注意到:对于如上所述选择的肽序列(左图),所有可能的五聚体序列中有14%未呈现(这是图中的第一个柱)。大多数五聚体序列呈现一次(第二个柱),而少数五聚体序列呈现多于一次。总计,所有可能的五聚体序列中有86%在选定的文库中呈现。相比之下,所有五聚体中仅有约75%在使用随机数发生器产生的肽序列的文库中呈现(右图)。可以看出:关于随机产生的序列的右侧图中的第一个柱(代表文库中未呈现的序列)比关于如上所述选择的肽的左侧图中的第一个柱大。

[0297] 图33示出了对于如上所述选择的肽文库,作为肽中的位置的函数的氨基酸组成。N末端在位置1处而C-末端在10至16间的位置处(如上所述,存在该文库中肽长度的分布)。这显示于X轴上。Y轴显示了在X轴上显示的位置处包含特定氨基酸的肽的部分。每条线(每种颜色)代表18种氨基酸中的一种。可以看出:在N-末端,两种氨基酸略微呈现不足(少于5%的肽在其N-末端含有这些氨基酸)。在大多数序列中,氨基酸组成基本上是恒定的并且平均稍大于5%。这是人们对于18种氨基酸的均匀分布( $1/18 \approx 0.056$ )所期望的。由于肽的数目随着该区域中长度的增加而降低,因此部分地发生靠近C末端的发散。

#### [0298] 实施方案

[0299] 以下非限制性实施方案提供了本发明的说明性实例,但并不限制本发明的范围。

[0300] 实施方案1.一种健康监测方法,该方法包括:a)使复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述生物样品中的至少一种抗体的不同的肽;b)测定所述抗体与所述肽阵列中的多个不同的肽的脱靶结合以形成免疫特征;以及c)将所述免疫特征与健康状态相关联。

[0301] 实施方案2.实施方案1的方法,其中所述肽阵列上的不同的肽为8至35个残基的长度。

[0302] 实施方案3.实施方案1和2中任一个的方法,其中所述肽阵列上的不同的肽为15至25个残基的长度。

[0303] 实施方案4.实施方案1-3中任一个的方法,其中所述肽阵列上的不同的肽具有范围为2-4nm的平均间距。

[0304] 实施方案5.实施方案1-4中任一个的方法,其中所述肽阵列上的不同的肽具有范围为3-6nm的平均间距。

[0305] 实施方案6.实施方案1-5中任一个的方法,其中所述不同的肽以约 $10^3\text{M}^{-1}$ 的缔合常

数与分子结合。

[0306] 实施方案7.实施方案1-6中任一个的方法,其中所述不同的肽以 $10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0307] 实施方案8.实施方案1-7中任一个的方法,其中所述不同的肽以 $2 \times 10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0308] 实施方案9.实施方案1-8中任一个的方法,其中所述不同的肽以 $10^4$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0309] 实施方案10.实施方案1-9中任一个的方法,其中所述不同的肽包含肽模拟物。

[0310] 实施方案11.实施方案1-10中任一个的方法,其中所述不同的肽具有随机氨基酸序列。

[0311] 实施方案12.实施方案1-11中任一个的方法,其中所述不同的肽包含非天然氨基酸。

[0312] 实施方案13.一种提供治疗的方法,该方法包括:a)接收来自受试者的复杂生物样品;b)使所述复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述生物样品中的至少一种抗体的不同的肽;c)测定所述抗体与多个不同的肽的脱靶结合以形成免疫特征;d)将所述免疫特征与状况相关联;以及e)提供针对所述状况的治疗。

[0313] 实施方案14.实施方案13的方法,其中所述肽阵列上的不同的肽为8至35个残基的长度。

[0314] 实施方案15.实施方案13和14中任一个的方法,其中所述肽阵列上的不同的肽为15至25个残基的长度。

[0315] 实施方案16.实施方案13-15中任一个的方法,其中所述肽阵列上的不同的肽具有范围为2-4nm的平均间距。

[0316] 实施方案17.实施方案13-16中任一个的方法,其中所述肽阵列上的不同的肽具有范围为3-6nm的平均间距。

[0317] 实施方案18.实施方案13-17中任一个的方法,其中所述不同的肽以约 $10^3\text{M}^{-1}$ 的缔合常数与分子结合。

[0318] 实施方案19.实施方案13-18中任一个的方法,其中所述不同的肽以 $10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0319] 实施方案20.实施方案13-19中任一个的方法,其中所述不同的肽以 $2 \times 10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0320] 实施方案21.实施方案13-20中任一个的方法,其中所述不同的肽以 $10^4$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0321] 实施方案22.实施方案13-21中任一个的方法,其中所述不同的肽包含肽模拟物。

[0322] 实施方案23.实施方案13-22中任一个的方法,其中所述不同的肽具有随机氨基酸序列。

[0323] 实施方案24.实施方案13-23中任一个的方法,其中所述不同的肽包含非天然氨基酸。

[0324] 实施方案25.一种预防状况的方法,该方法包括:a)提供来自受试者的复杂生物样品;b)使所述复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述复杂生

物样品中的至少一种抗体的不同的肽;c)测定所述复杂生物样品与多个不同的肽的脱靶结合以形成免疫特征;d)将所述免疫特征与状况相关联;以及e)接受针对所述状况的处理。

[0325] 实施方案26.实施方案25的方法,其中所述肽阵列上的不同的肽为8至35个残基的长度。

[0326] 实施方案27.实施方案25和26中任一个的方法,其中所述肽阵列上的不同的肽为15至25个残基的长度。

[0327] 实施方案28.实施方案25-27中任一个的方法,其中所述肽阵列上的不同的肽具有范围为2-4nm的平均间距。

[0328] 实施方案29.实施方案25-28中任一个的方法,其中所述肽阵列上的不同的肽具有范围为3-6nm的平均间距。

[0329] 实施方案30.实施方案25-29中任一个的方法,其中所述不同的肽以约 $10^3\text{M}^{-1}$ 的缔合常数与分子结合。

[0330] 实施方案31.实施方案25-30中任一个的方法,其中所述不同的肽以 $10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0331] 实施方案32.实施方案25-31中任一个的方法,其中所述不同的肽以 $2 \times 10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0332] 实施方案33.实施方案25-32中任一个的方法,其中所述不同的肽以 $10^4$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

[0333] 实施方案34.实施方案25-33中任一个的方法,其中所述不同的肽包含肽模拟物。

[0334] 实施方案35.实施方案25-34中任一个的方法,其中所述不同的肽具有随机氨基酸序列。

[0335] 实施方案36.实施方案25-35中任一个的方法,其中所述不同的肽包含非天然氨基酸。

[0336] 实施方案37.一种诊断方法,该方法包括:a)接收来自受试者的复杂生物样品;b)使所述复杂生物样品与肽阵列接触,其中所述肽阵列包含能够脱靶结合所述生物样品中的至少一种抗体的不同的肽;c)测定所述抗体与所述肽阵列中的一组不同的肽的脱靶结合以形成免疫特征;以及d)基于所述免疫特征诊断状况。

[0337] 实施方案38.实施方案37的方法,其中所述肽阵列上的不同的肽为8至35个残基的长度。

[0338] 实施方案39.实施方案37和38中任一个的方法,其中所述肽阵列上的不同的肽为15至25个残基的长度。

[0339] 实施方案40.实施方案37-39中任一个的方法,其中所述肽阵列上的不同的肽具有范围为2-4nm的平均间距。

[0340] 实施方案41.实施方案37-40中任一个的方法,其中所述肽阵列上的不同的肽具有范围为3-6nm的平均间距。

[0341] 实施方案42.实施方案37-41中任一个的方法,其中所述不同的肽以约 $10^3\text{M}^{-1}$ 的缔合常数与分子结合。

[0342] 实施方案43.实施方案37-42中任一个的方法,其中所述不同的肽以 $10^3$ 至 $10^6\text{M}^{-1}$ 范围内的缔合常数与分子结合。

- [0343] 实施方案44.实施方案37-43中任一个的方法,其中所述不同的肽以 $2 \times 10^3$ 至 $10^6 \text{M}^{-1}$ 范围内的缔合常数与分子结合。
- [0344] 实施方案45.实施方案37-44中任一个的方法,其中所述不同的肽以 $10^4$ 至 $10^6 \text{M}^{-1}$ 范围内的缔合常数与分子结合。
- [0345] 实施方案46.实施方案37-45中任一个的方法,其中所述不同的肽包含肽模拟物。
- [0346] 实施方案47.实施方案37-46中任一个的方法,其中所述不同的肽具有随机氨基酸序列。
- [0347] 实施方案48.实施方案37-47中任一个的方法,其中所述不同的肽结合互补位。
- [0348] 实施方案49.一种阵列,其包含多个固定至固体支持体上不同位置的、原位合成的具有可变长度的聚合物,其中聚合物的原位合成包括以下步骤:
- [0349] a.将第一单体添加至所述固体支持体上预定部分的位置;
- [0350] b.将第二单体添加至所述固体支持体上预定部分的位置,其中用于所述第二单体的预定部分的位置包括含有第一单体的位置和不含单体的位置;
- [0351] c.将第三单体添加至所述固体支持体上预定部分的位置,其中用于所述第二单体的预定部分的位置包括含有第一和第二单体的位置、含有第二单体的位置以及不含单体的位置;以及
- [0352] d.用限定的一组单体重复步骤a-c,直到所述聚合物达到所期望的平均长度并且所述部分的总和总计至少为100%。
- [0353] 实施方案50.实施方案49的阵列,其中所述阵列是伪随机阵列。
- [0354] 实施方案51.实施方案49的阵列,其中所述阵列是随机阵列。
- [0355] 实施方案52.实施方案49的阵列,其中所述单体选自氨基酸、核酸和肽核酸。
- [0356] 实施方案53.实施方案49的阵列,其中所述限定的单体组中的单体出现一次或多次。
- [0357] 实施方案54.实施方案49的阵列,其中所述限定的单体组中的独特单体的数目至少为2。
- [0358] 实施方案55.实施方案49的阵列,其中所述聚合物具有至少10个残基的平均长度。
- [0359] 实施方案56.实施方案49的阵列,其中所述聚合物具有至少12个残基的平均长度。
- [0360] 实施方案57.实施方案49的阵列,其中所述聚合物具有不少于5个残基的平均长度。
- [0361] 实施方案58.实施方案49的阵列,其中至少5%的聚合物具有至少12个残基的长度。
- [0362] 实施方案59.实施方案49的阵列,其中所述聚合物可与样品中的成分结合。
- [0363] 实施方案60.实施方案49的阵列,其中所述部分的总和总计为100%。
- [0364] 实施方案61.实施方案49的阵列,其中所述部分的总和大于100%。
- [0365] 实施方案62.实施方案49的阵列,其中所述聚合物的数目大于3,000。
- [0366] 实施方案63.实施方案49的阵列,其中所述聚合物的数目大于10,000。
- [0367] 实施方案64.实施方案49的阵列,其中所述聚合物的数目大于100,000。
- [0368] 实施方案65.实施方案49的阵列,其中所述聚合物的数目大于330,000。
- [0369] 实施方案66.一种制造阵列的方法,该阵列包含多个固定至固体支持体上不同位

置的、原位合成的具有可变长度的聚合物,该方法包括以下步骤:

- [0370] a. 提供基底作为在其上合成聚合物的固体支持体;
- [0371] b. 将第一单体添加至所述固体支持体上预定部分的位置;
- [0372] c. 将第二单体添加至所述固体支持体上预定部分的位置,其中用于所述第二单体的预定部分的位置包括含有第一单体的位置和不含单体的位置;
- [0373] d. 将第三单体添加至所述固体支持体上预定部分的位置,其中用于所述第二单体的预定部分的位置包括含有第一和第二单体的位置、含有第二单体的位置以及不含单体的位置;以及
- [0374] e. 用限定的一组单体重复步骤b-d,直到所述聚合物达到所期望的平均长度并且所述部分的总和总计至少为100%。
- [0375] 实施方案67. 实施方案66的方法,其中所述阵列是伪随机阵列。
- [0376] 实施方案68. 实施方案66的方法,其中所述阵列是伪随机阵列。
- [0377] 实施方案69. 实施方案66的方法,其中所述阵列是随机阵列。
- [0378] 实施方案70. 实施方案66的方法,其中所述单体选自氨基酸、核酸和肽核酸。
- [0379] 实施方案71. 实施方案66的方法,其中所述限定的单体组中的单体出现一次或多次。
- [0380] 实施方案72. 实施方案66的方法,其中所述限定的单体组中的独特单体的数目至少为2。
- [0381] 实施方案73. 实施方案66的方法,其中所述聚合物具有至少10个残基的平均长度。
- [0382] 实施方案74. 实施方案66的方法,其中所述聚合物具有至少12个残基的平均长度。
- [0383] 实施方案75. 实施方案66的方法,其中所述聚合物具有不少于5个残基的平均长度。
- [0384] 实施方案76. 实施方案66的方法,其中至少5%的聚合物具有至少12个残基的长度。
- [0385] 实施方案77. 实施方案66的方法,其中所述聚合物可与样品中的成分结合。
- [0386] 实施方案78. 实施方案66的方法,其中所述部分的总和总计为100%。
- [0387] 实施方案79. 实施方案66的方法,其中所述部分的总和大于100%。
- [0388] 实施方案80. 实施方案66的方法,其中所述聚合物的数目大于3,000。
- [0389] 实施方案81. 实施方案66的方法,其中所述聚合物的数目大于10,000。
- [0390] 实施方案82. 实施方案66的方法,其中所述聚合物的数目大于100,000。
- [0391] 实施方案83. 实施方案66的方法,其中所述聚合物的数目大于330,000。
- [0392] 实施方案89. 一种使用阵列来监测受试者的健康状态的方法,该方法包括以下步骤:
- [0393] a) 使复杂生物样品与权利要求49至65中任意一项所述的肽阵列接触,其中所述肽阵列包含能够脱靶结合所述生物样品中的至少一种抗体的不同的肽;
- [0394] b) 测定所述抗体与所述肽阵列中的多个不同的肽的脱靶结合以形成免疫特征;以及
- [0395] c) 将所述免疫特征与健康状态相关联。

<110> 亚利桑那州评议委员会, 亚利桑那州法人团体, 代理和代表亚利桑那州立大学

<120> 免疫特征分析: 通向早期诊断和健康监测的途径

<130> 43638-707.601

<140> PCT/US2013/057373

<141> 2013-08-29

<150> 61/694,598

<151> 2012-08-29

<160> 36

<170> PatentIn version 3.5

<210> 1

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

[0001]

<400> 1

Cys Ser Gly Ser Tyr Asn Met Asp Lys Tyr Phe Thr Tyr Ser Trp Tyr  
1                   5                   10                   15

Arg Glu Glu Arg  
                  20

<210> 2

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 2

Cys Ser Gly Trp Asp Ser Phe Arg His Tyr Glu Arg Ile Thr Asp Arg  
1                   5                   10                   15

His Gln Gly Asp  
                  20

<210> 3  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述：合成肽

<400> 3  
 Cys Ser Gly Arg Tyr Phe Met His Met Glu Pro Thr Ile Asn His Tyr  
 1                   5                   10                   15

Tyr Glu Gly Met  
 20

<210> 4  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述：合成肽

[0002]

<400> 4  
 Cys Ser Gly Val Met Met Pro Asp Tyr Arg Ile His Val His Trp Ser  
 1                   5                   10                   15

Asn Trp Thr Gly  
 20

<210> 5  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述：合成肽

<400> 5  
 Cys Ser Gly Leu Arg His Tyr Asn Val Tyr Asp Phe Arg Ser Asn Asp  
 1                   5                   10                   15

Arg His Trp Ala  
 20

<210> 6  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述: 合成肽

<400> 6  
 Cys Ser Gly Val Met Ala His Thr Gly His Ser Gly Glu Met Gly Pro  
 1                   5                   10                   15

Pro Asp Phe Gln  
                   20

<210> 7  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述: 合成肽

[0003]

<400> 7  
 Cys Ser Gly Asn Asp His Ser Gln His Asp Phe Ala Pro Val Glu Ser  
 1                   5                   10                   15

Tyr Ile Met Met  
                   20

<210> 8  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述: 合成肽

<400> 8  
 Cys Ser Gly Ile Leu Phe Phe Thr Arg Glu Thr Asp Val His Tyr Pro  
 1                   5                   10                   15

Ala Asn Glu Gly  
                   20

<210> 9

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 9

Cys Ser Gly Val Asp Pro Trp Arg Ser His Ala Asn Gln Arg Glu Tyr  
1                   5                   10                   15

Ala Ile Ala Asn  
                  20

<210> 10

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 10

[0004]

Cys Ser Gly Asn Gly Val His Glu Phe Ser Ala Met Leu Ile Met Asp  
1                   5                   10                   15

Met Ile Ile Phe  
                  20

<210> 11

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 11

Cys Ser Gly Ile Gly Asp His Met Pro Leu Asn Glu Pro Asn Pro Leu  
1                   5                   10                   15

Arg Asp Leu Lys  
                  20

<210> 12

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 12

Cys Ser Gly Thr His Ile Ala Thr Asn Pro Leu Asn Val Gln Tyr Val  
1                   5                   10                   15

Met Val Gln Ser  
                  20

<210> 13

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 13

Cys Ser Gly Thr Arg Lys Glu His Tyr Leu Glu His Val Ala Lys His  
1                   5                   10                   15

[0005]

Met Glu Val Trp  
                  20

<210> 14

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 14

Cys Ser Gly Pro Thr Asp Ile Thr Glu Leu Met Met Arg Pro Lys Tyr  
1                   5                   10                   15

Ser Arg Ile Asn  
                  20

<210> 15

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 15

Cys Ser Gly Asp Gln Gln Gly Thr Trp Gly Arg Val Asp Met Trp Ser  
1                   5                   10                   15

Asn Arg Met His  
                  20

<210> 16

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 16

Cys Ser Gly Ile Met Lys Arg Ile His Ala Gln Thr Met Trp Tyr Ser  
1                   5                   10                   15

[0006]

Pro Ile Thr Asp  
                  20

<210> 17

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 17

Cys Ser Gly Ser Phe Phe Tyr Val Asn Lys Gln Val Asn Asn Lys Asn  
1                   5                   10                   15

Tyr Gln Thr Ile  
                  20

<210> 18

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 18

Cys Ser Gly Leu Tyr Ala Lys Gln Val Ala Ala Gln Arg Pro Ile Lys  
1                   5                   10                   15

Tyr Trp Asp His  
                  20

<210> 19

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 19

Cys Ser Gly Met Met Trp Tyr His Gly Tyr Pro His Val His Ala Asn  
1                   5                   10                   15

[0007]

Asp Ala His Trp  
                  20

<210> 20

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 20

Cys Ser Gly Arg Tyr His Pro Asn Tyr Gly Asp Ala Lys Lys His Glu  
1                   5                   10                   15

Met Ser Arg Phe  
                  20

<210> 21

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 21

Cys Ser Gly His Trp Lys Gly Asp Leu Arg Ser Gly Arg His Tyr His  
1                   5                   10                   15

His Gln Glu Phe  
                  20

<210> 22

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 22

Cys Ser Gly Glu Asp Thr Arg Arg Gly His Ala Trp Lys Phe Ser Glu  
1                   5                   10                   15

[0008]

Ile Ser Pro His  
                  20

<210> 23

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 23

Cys Ser Gly Trp Val Arg Lys Ile Leu Lys Lys Arg Ile Trp Thr Asp  
1                   5                   10                   15

Pro Thr Asn Tyr  
                  20

<210> 24

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 24

Cys Ser Gly Tyr Pro Arg Ser Trp Phe Val Tyr Tyr Thr Pro Trp Lys  
1                   5                   10                   15

Leu Phe Lys Gly  
                  20

<210> 25

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 25

Cys Ser Gly Ser His Met Gln Asp Ile Tyr Arg Thr Val Arg Ser Leu  
1                   5                   10                   15

Gly Lys Ser Met  
                  20

[0009]

<210> 26

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 26

Cys Ser Gly Val Gln Leu Ser Ser Tyr Thr Leu Lys Leu Gly Lys Val  
1                   5                   10                   15

Tyr Gln Glu Arg  
                  20

<210> 27

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 27

Cys Ser Gly Lys Thr Met Thr Thr Gln Trp Arg Ser Ser Leu Phe Lys  
1                   5                   10                   15

Phe Ala Gly Met  
                  20

<210> 28

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 28

Cys Ser Gly Met Lys Tyr Asn Pro Phe Pro Lys Tyr Lys Ser Tyr Leu  
1                   5                   10                   15

Gln Tyr Val Asn  
                  20

[0010]

<210> 29

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 29

Cys Ser Gly Ile Ser Thr Lys Phe Trp Trp Lys Arg Asn Ser Ile Val  
1                   5                   10                   15

Phe Pro Lys Leu  
                  20

<210> 30

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述: 合成肽

<400> 30  
 Cys Ser Gly Thr Arg Gly Arg Trp Tyr Asp Arg Arg Ser Pro Ser Lys  
 1                   5                   10                   15

Phe Leu Gly Tyr  
                   20

<210> 31  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述: 合成肽

<400> 31  
 Cys Ser Gly Gln Asn Val Ser Ala Lys Tyr Val Lys Gly Arg Ser Val  
 1                   5                   10                   15

Gln Ser Trp Ile  
                   20

[0011]

<210> 32  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述: 合成肽

<400> 32  
 Cys Ser Gly His Ile Met Gly Arg Lys Arg His Trp Pro Met Ser Thr  
 1                   5                   10                   15

Ser Tyr Gly Val  
                   20

<210> 33  
 <211> 20  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 人工序列的描述: 合成肽

<400> 33



1

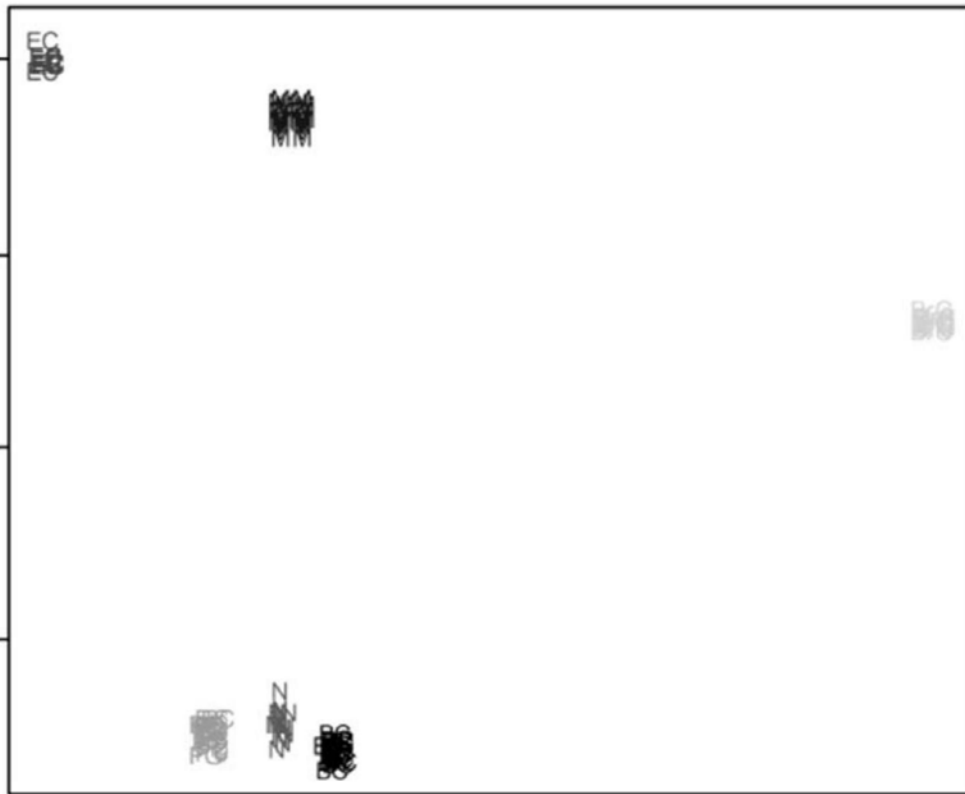
5

10

15

[0013]

Phe Arg His Leu  
20



支持向量机 (SVM)

图1

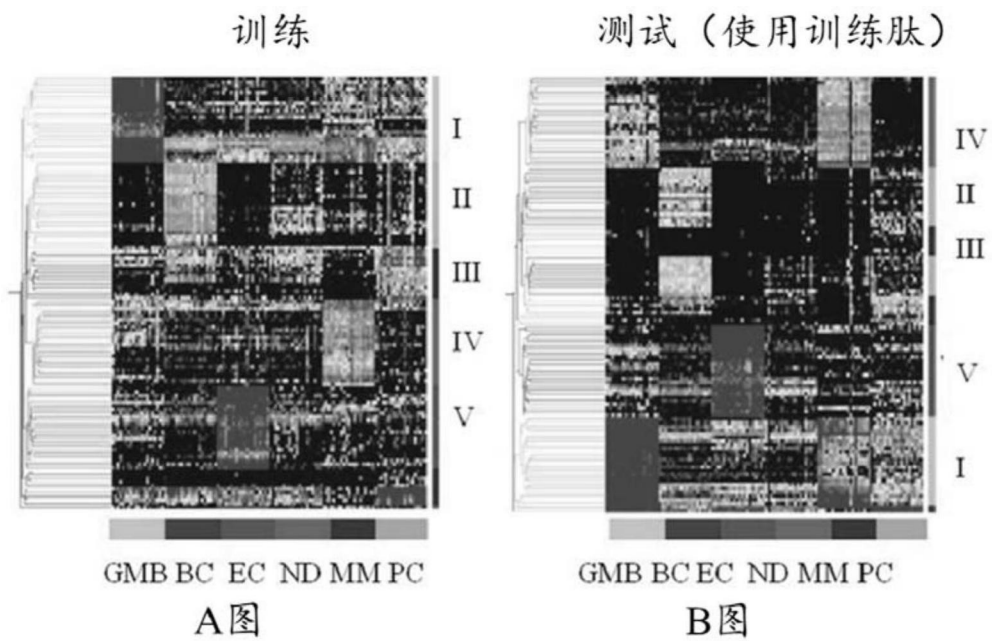


图2

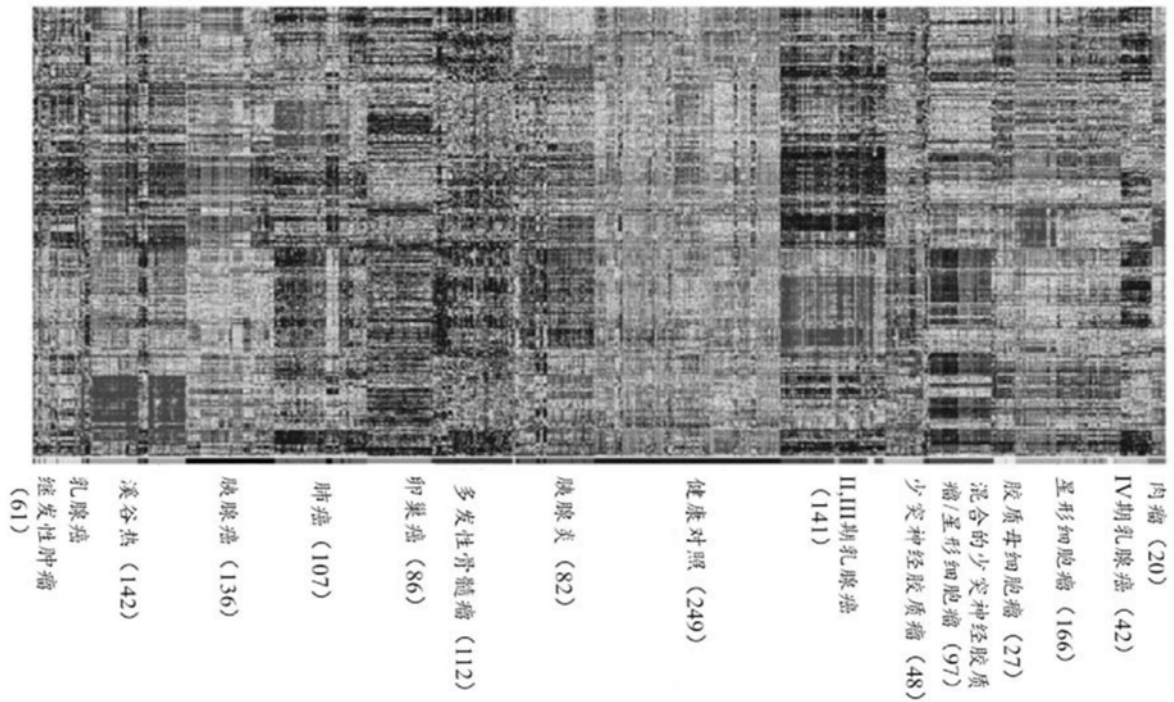


图3

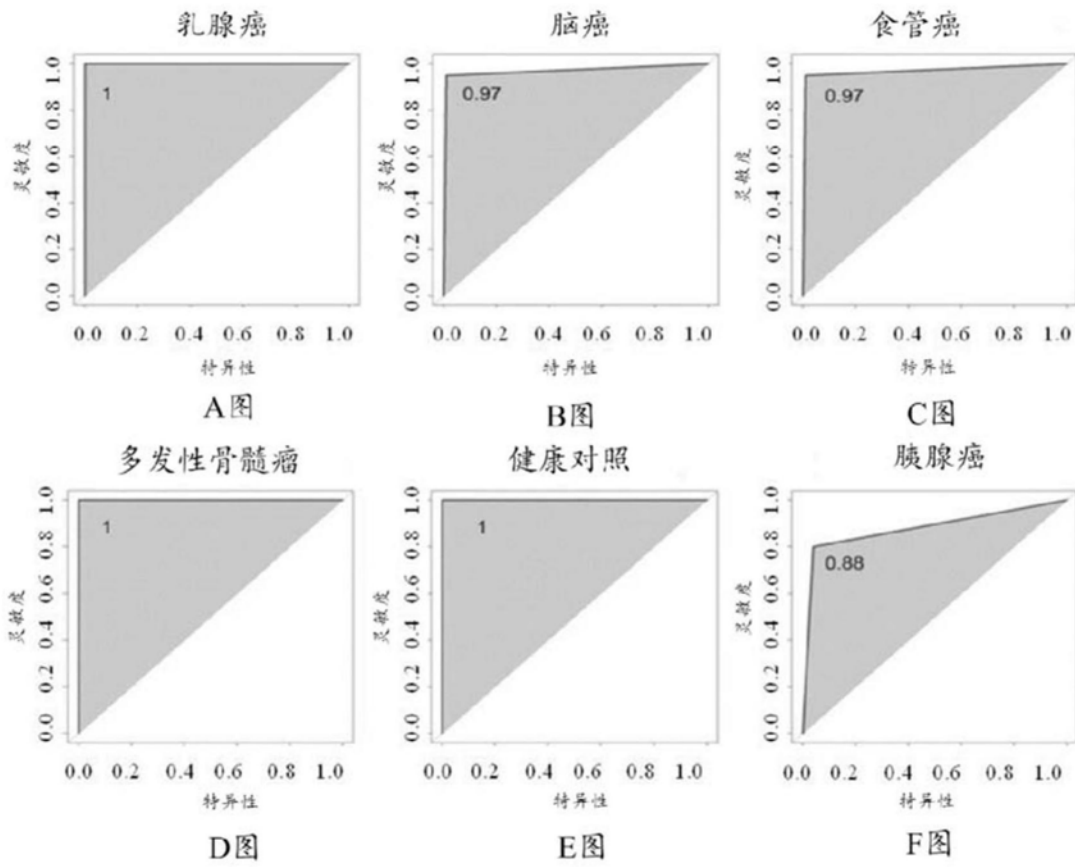


图4

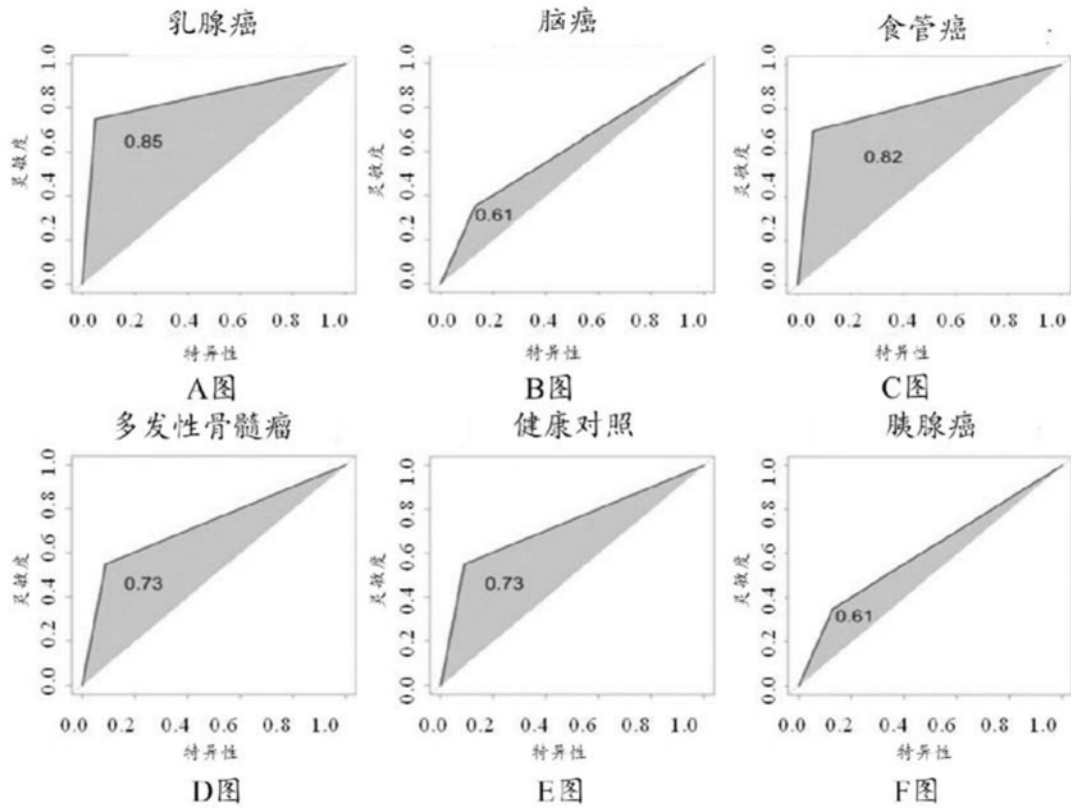


图5

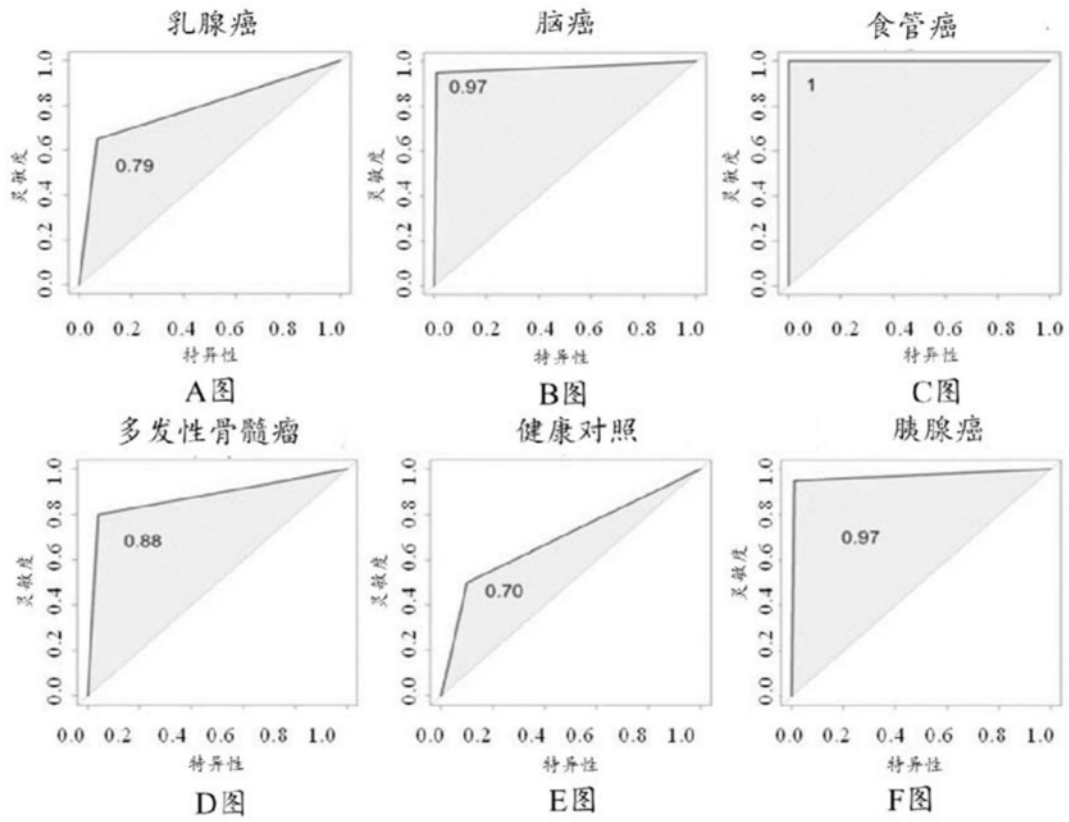


图6

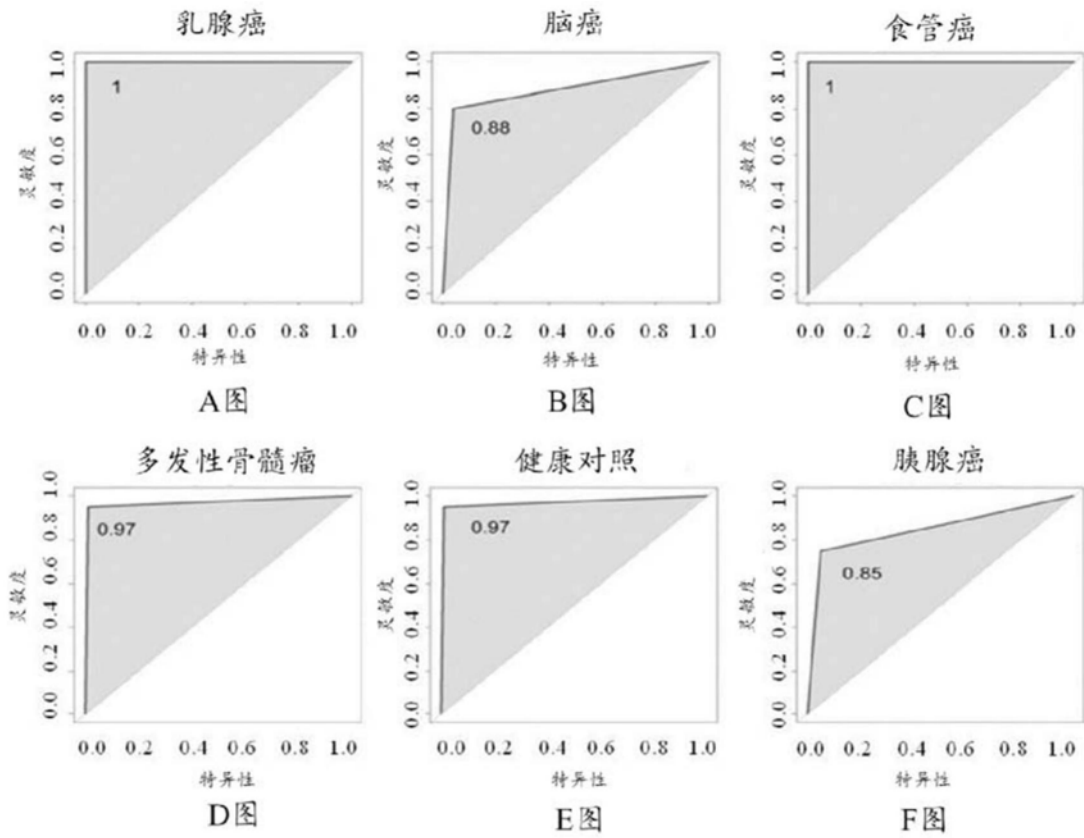


图7

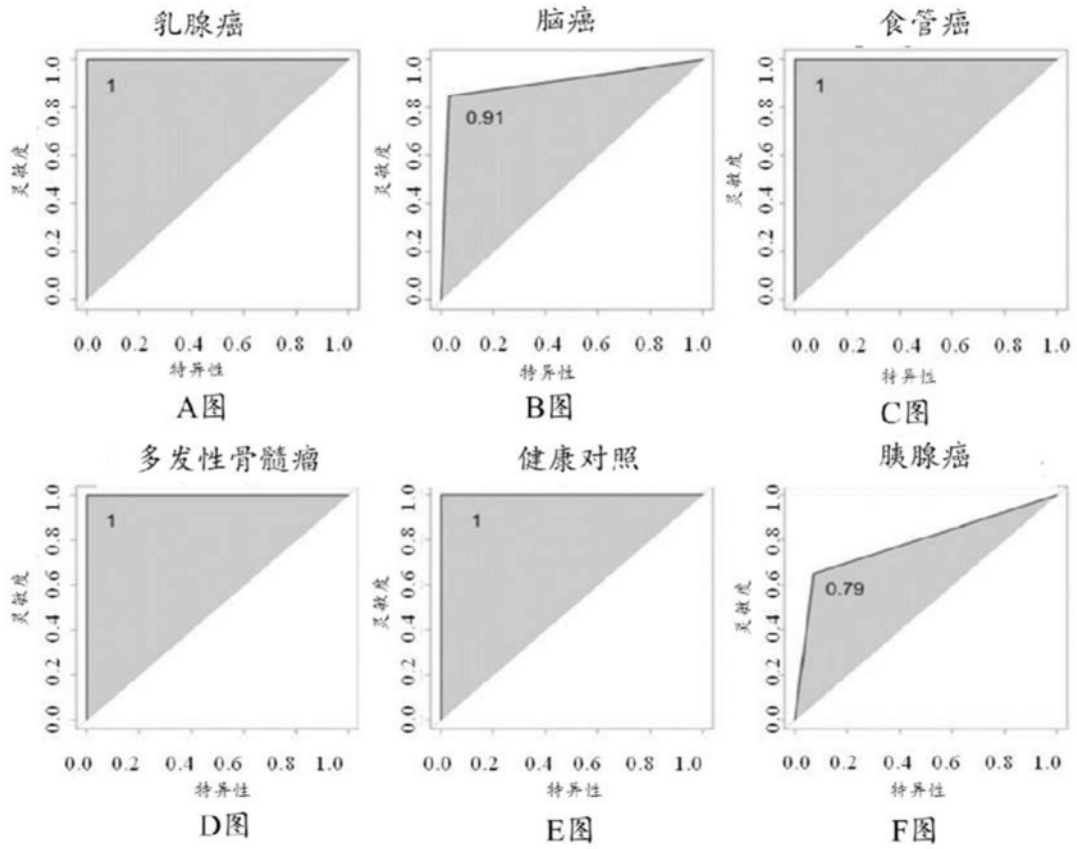


图8

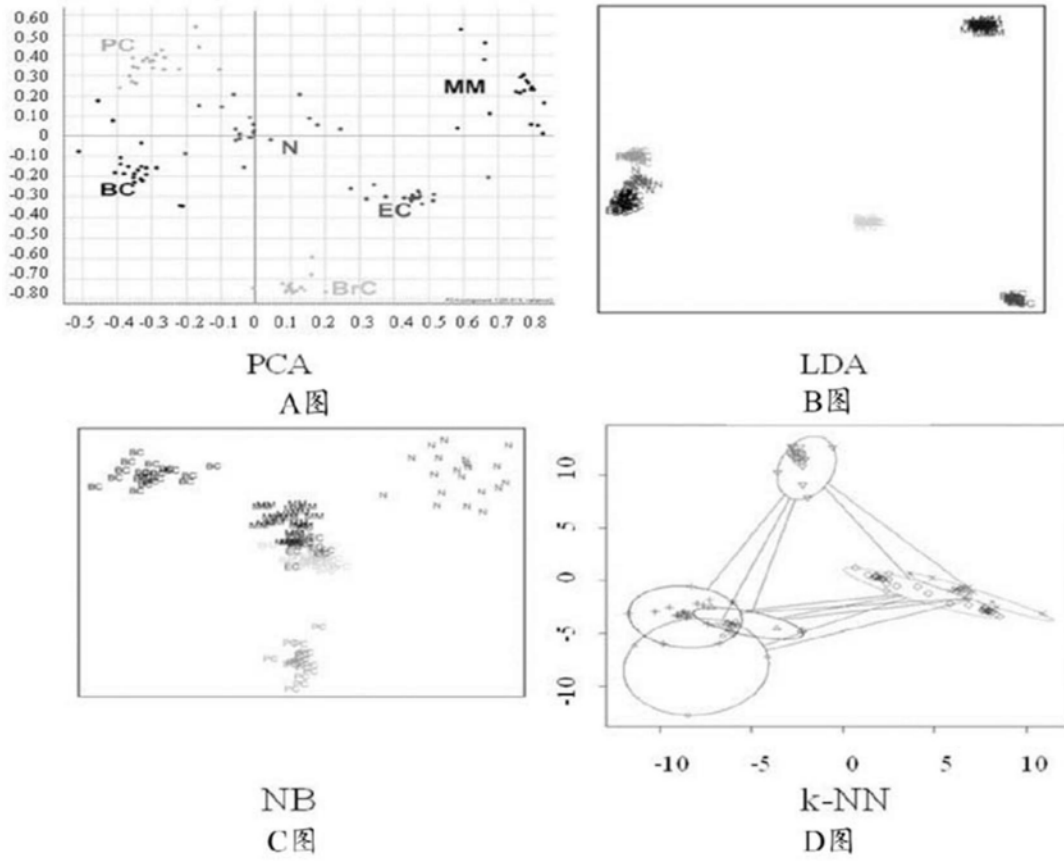


图9

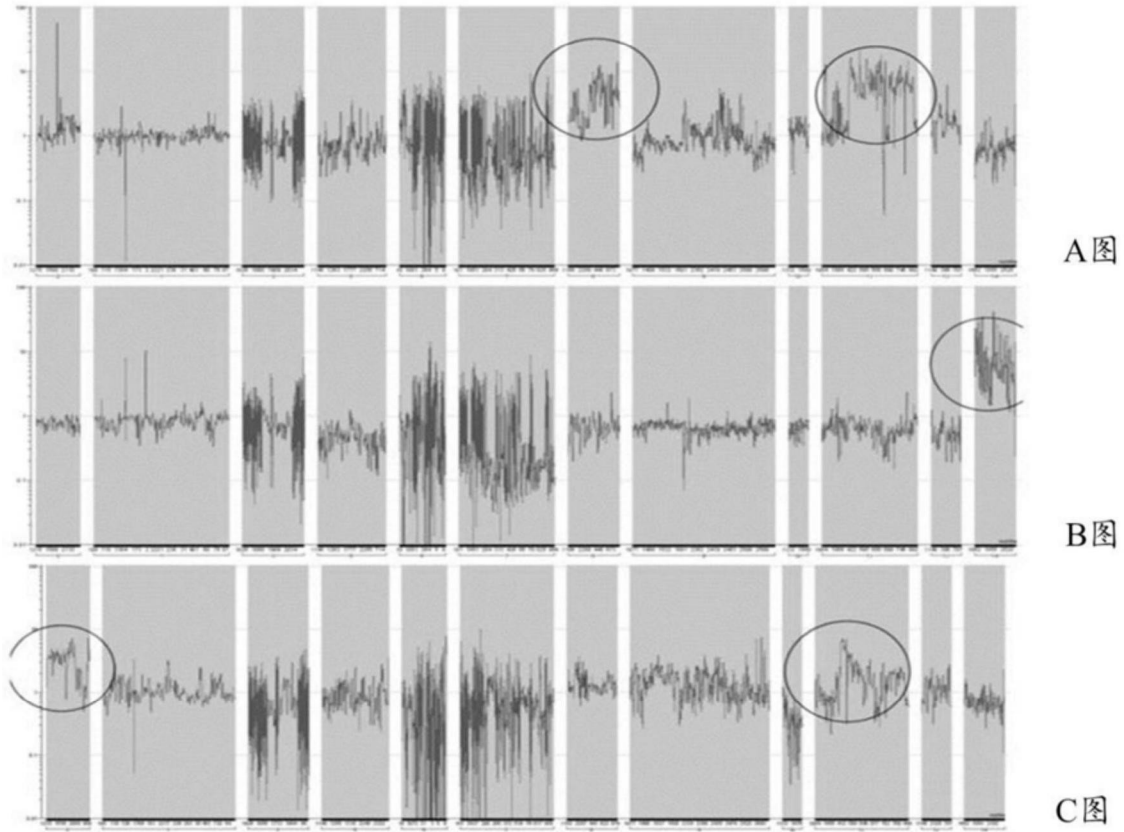


图10

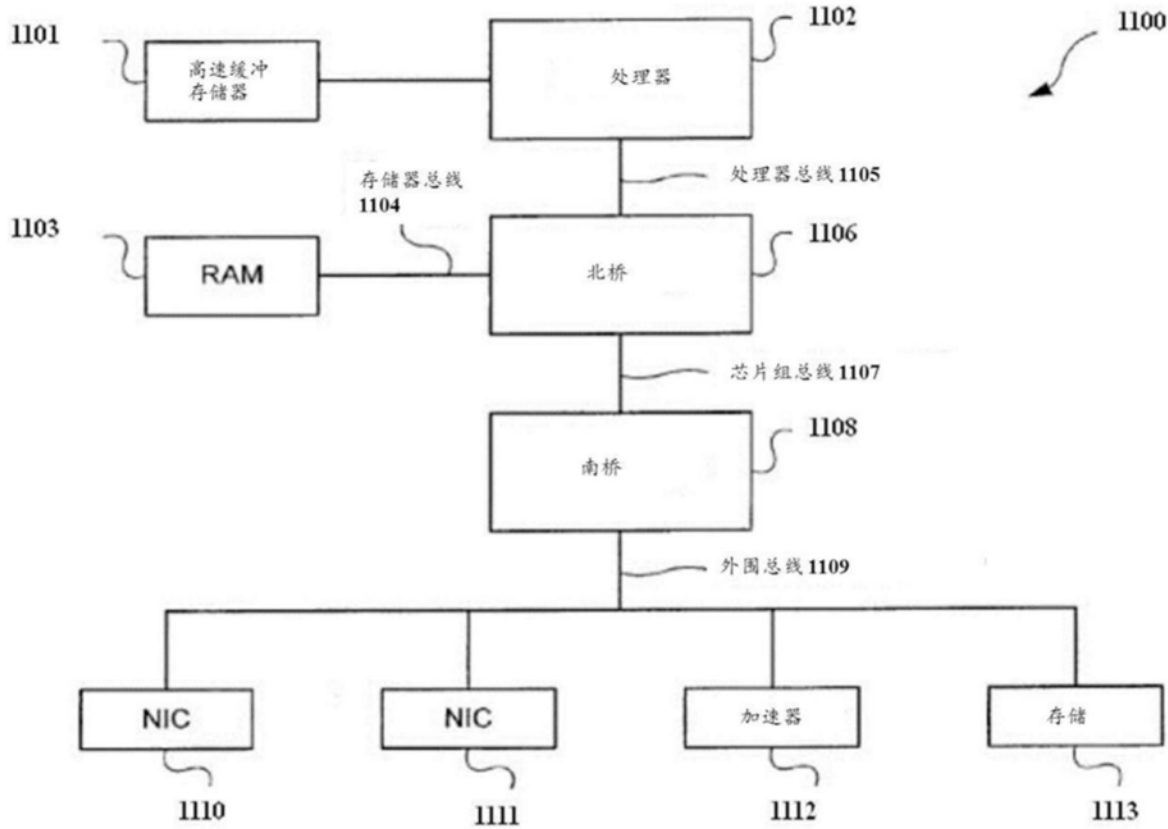


图11

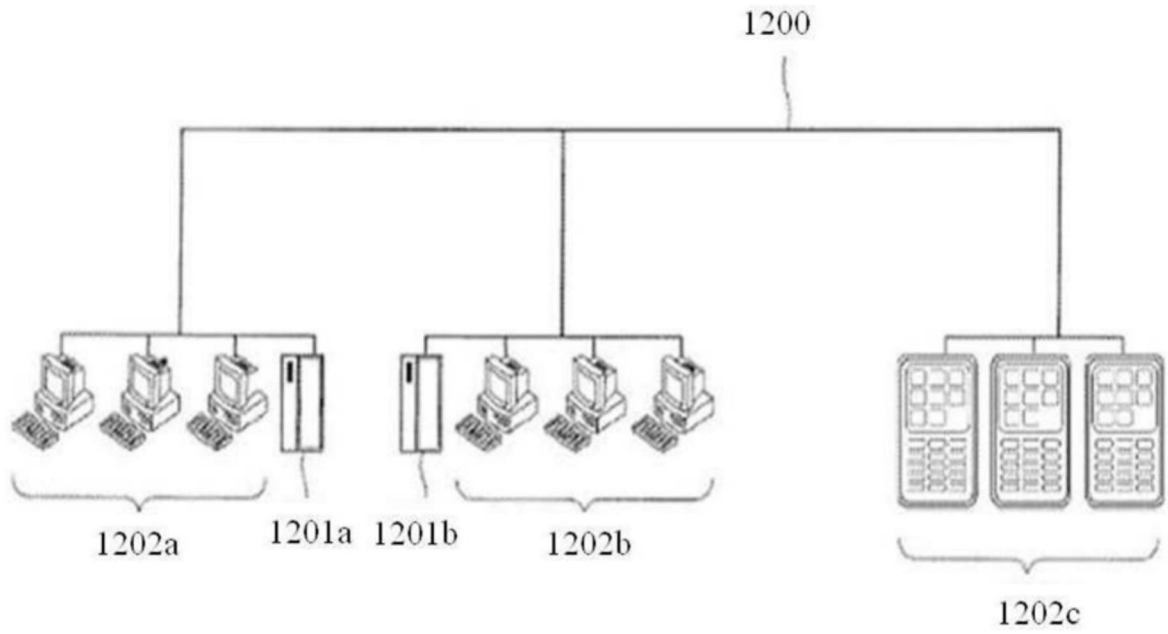


图12

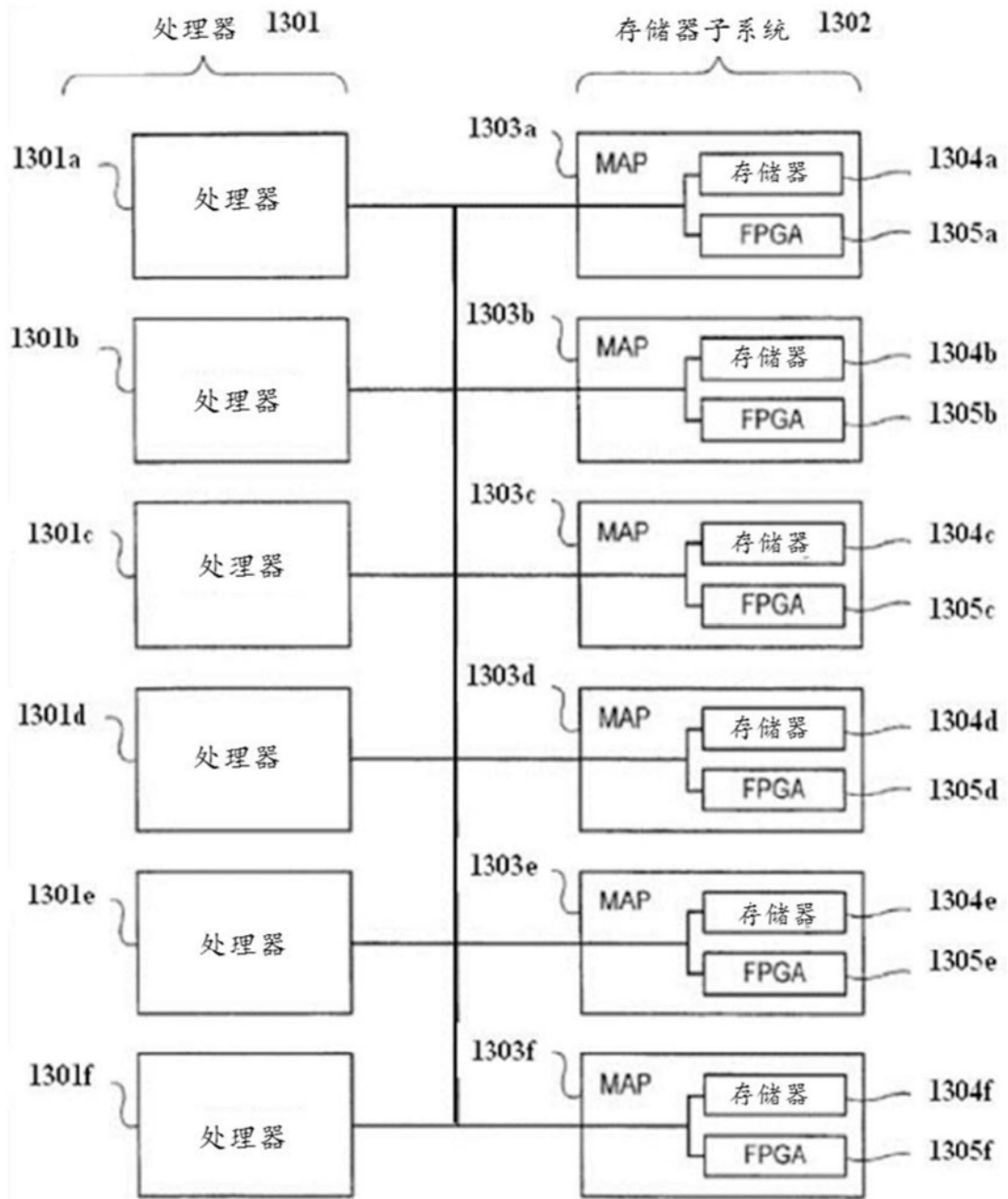


图13

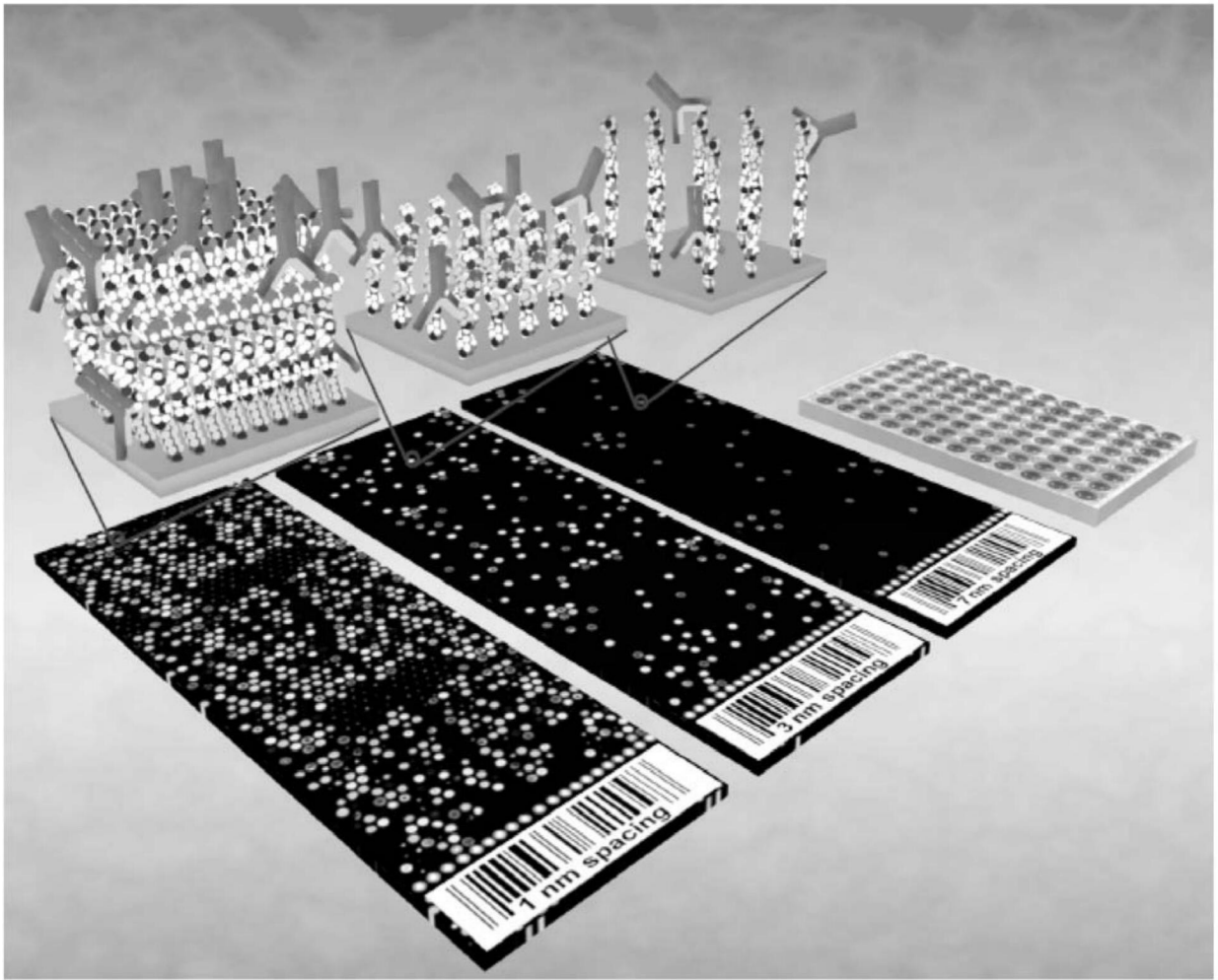


图14

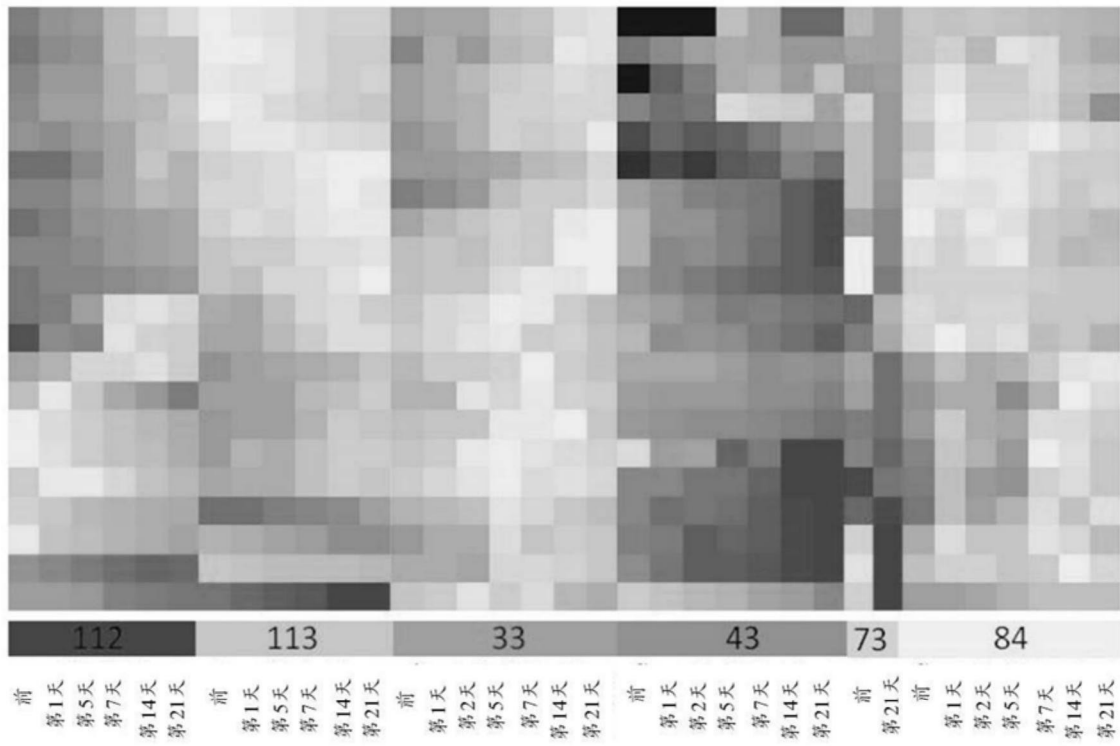
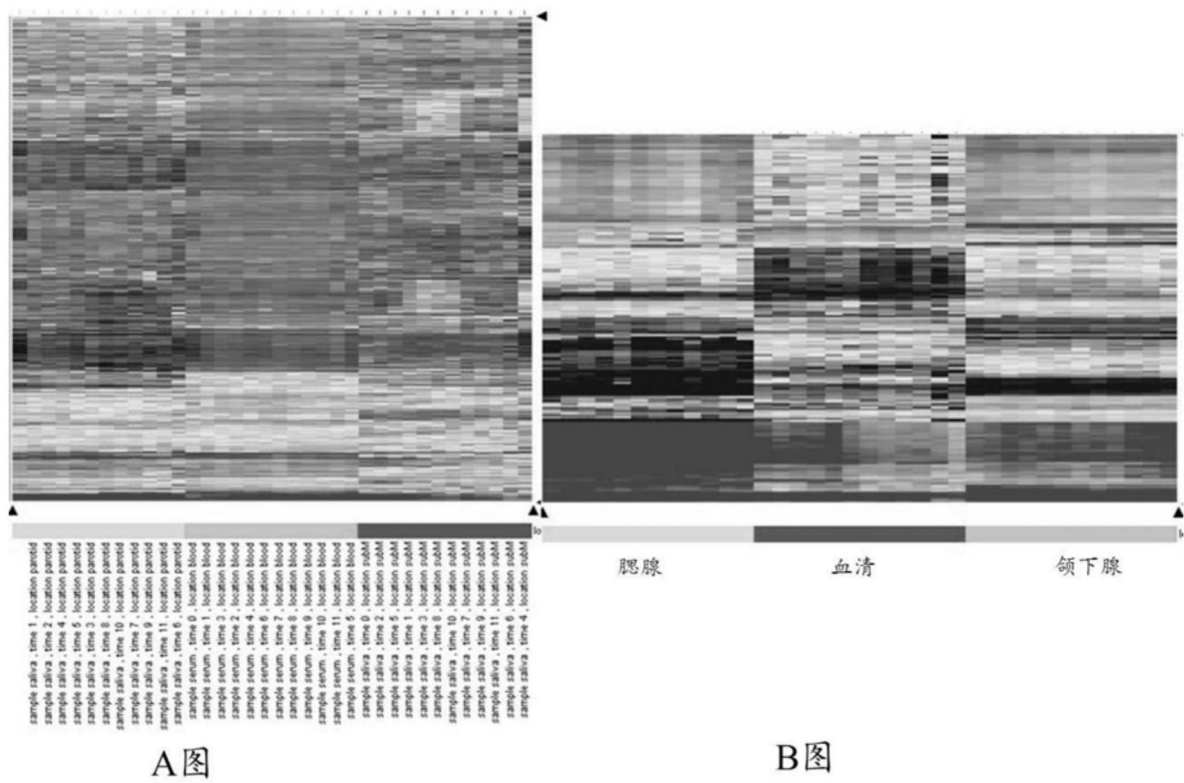


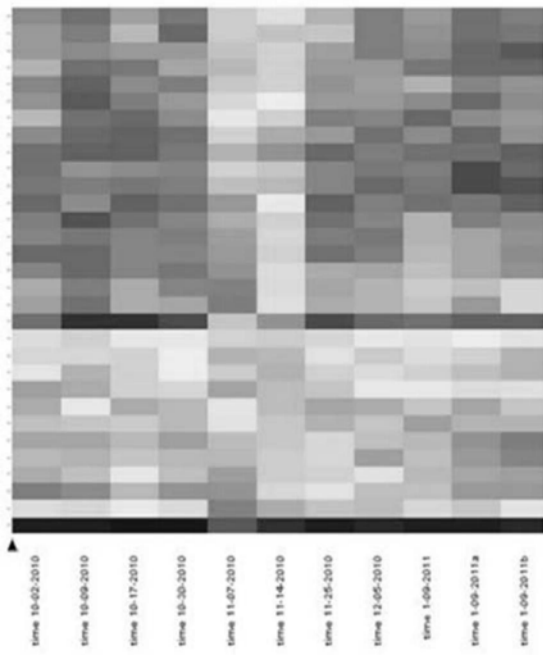
图15



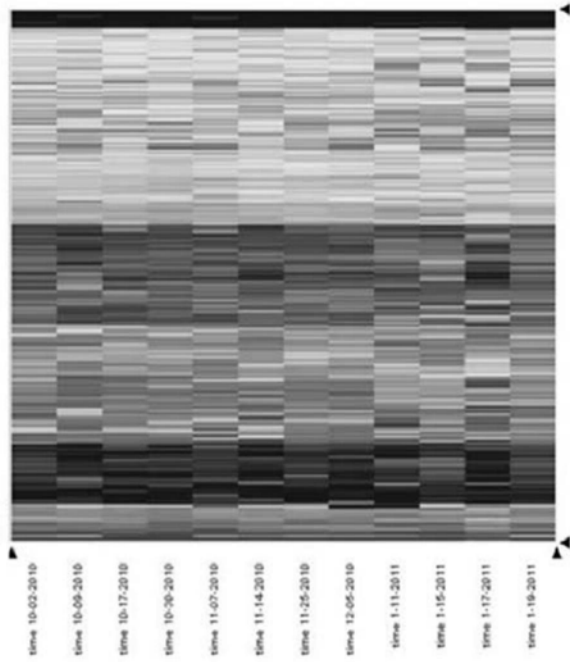
A图

B图

图16



A图



B图

图17

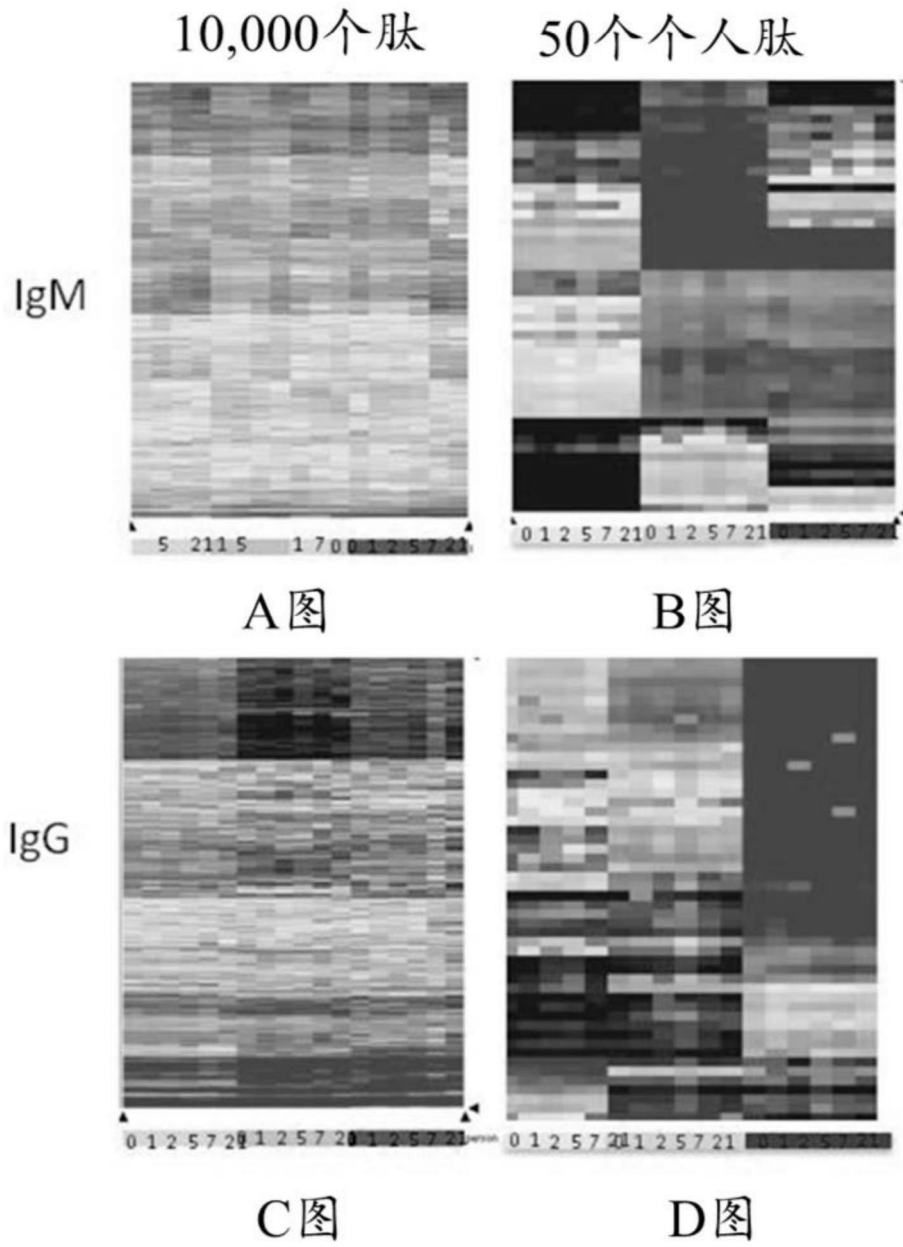


图18

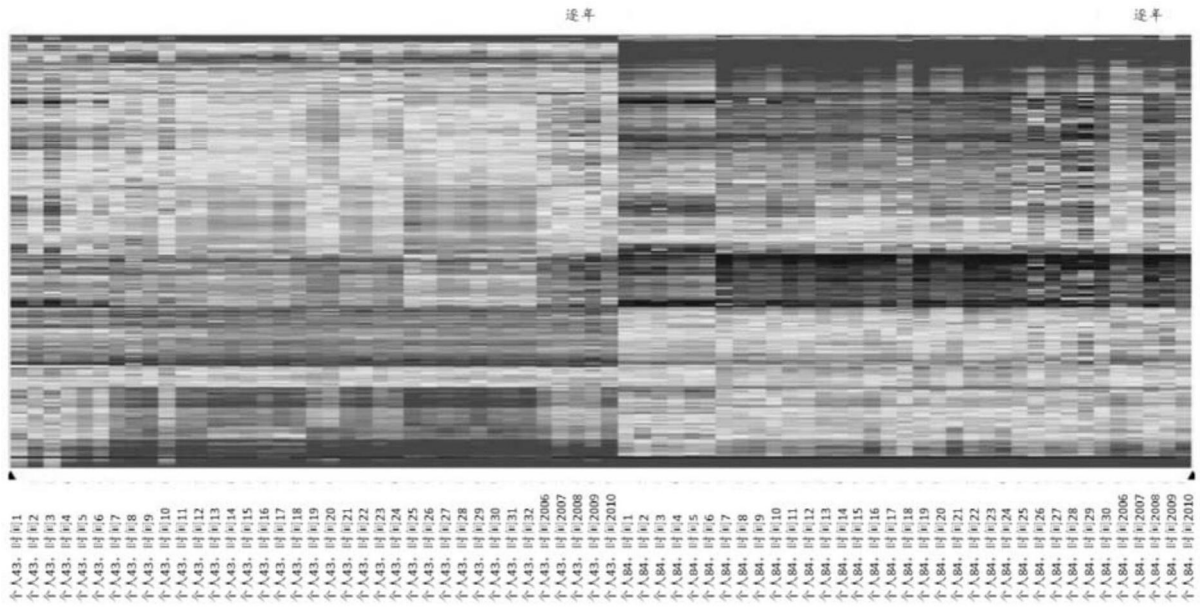


图19

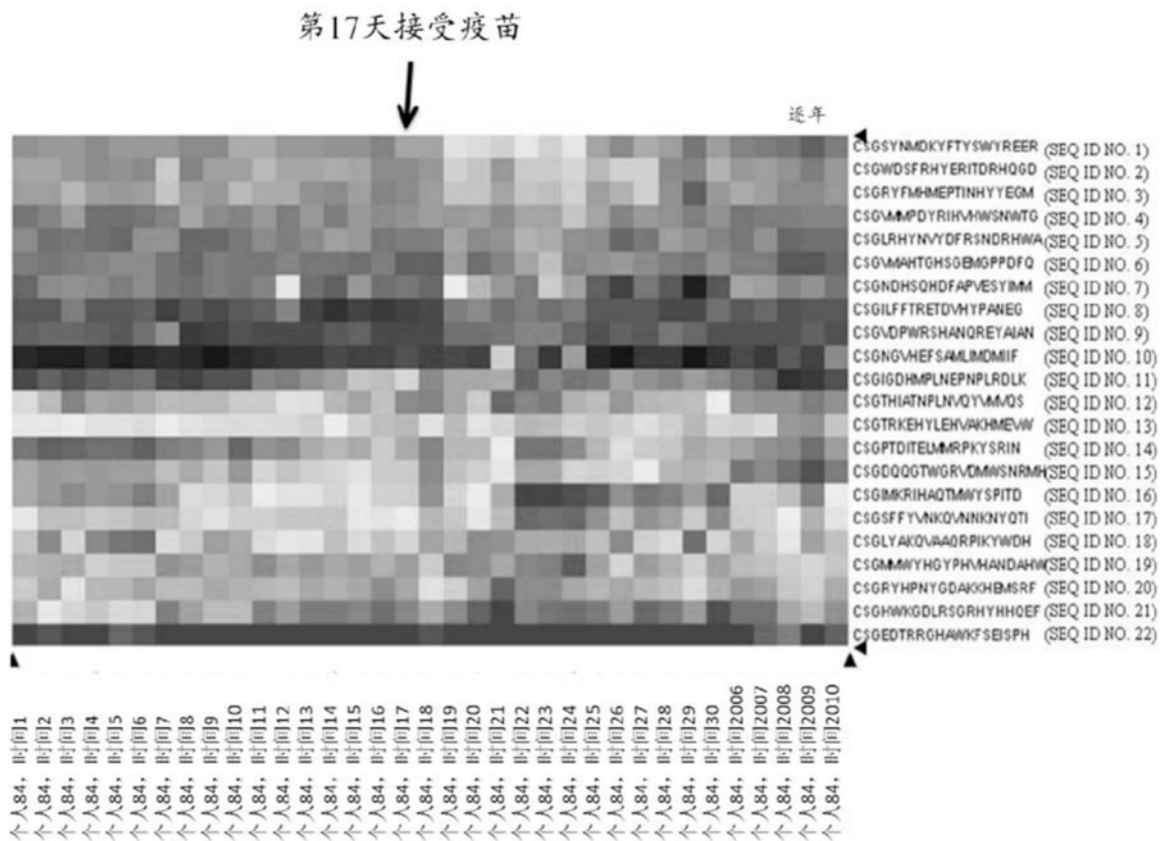


图20

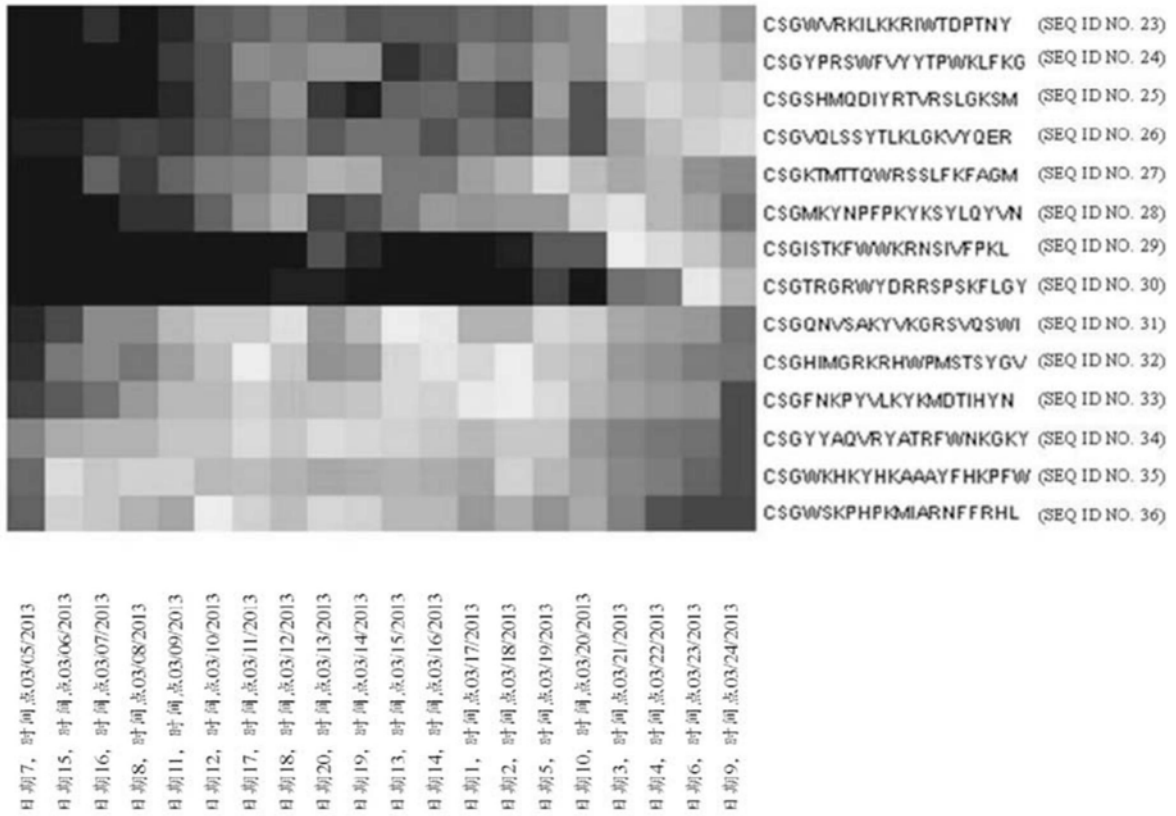


图21

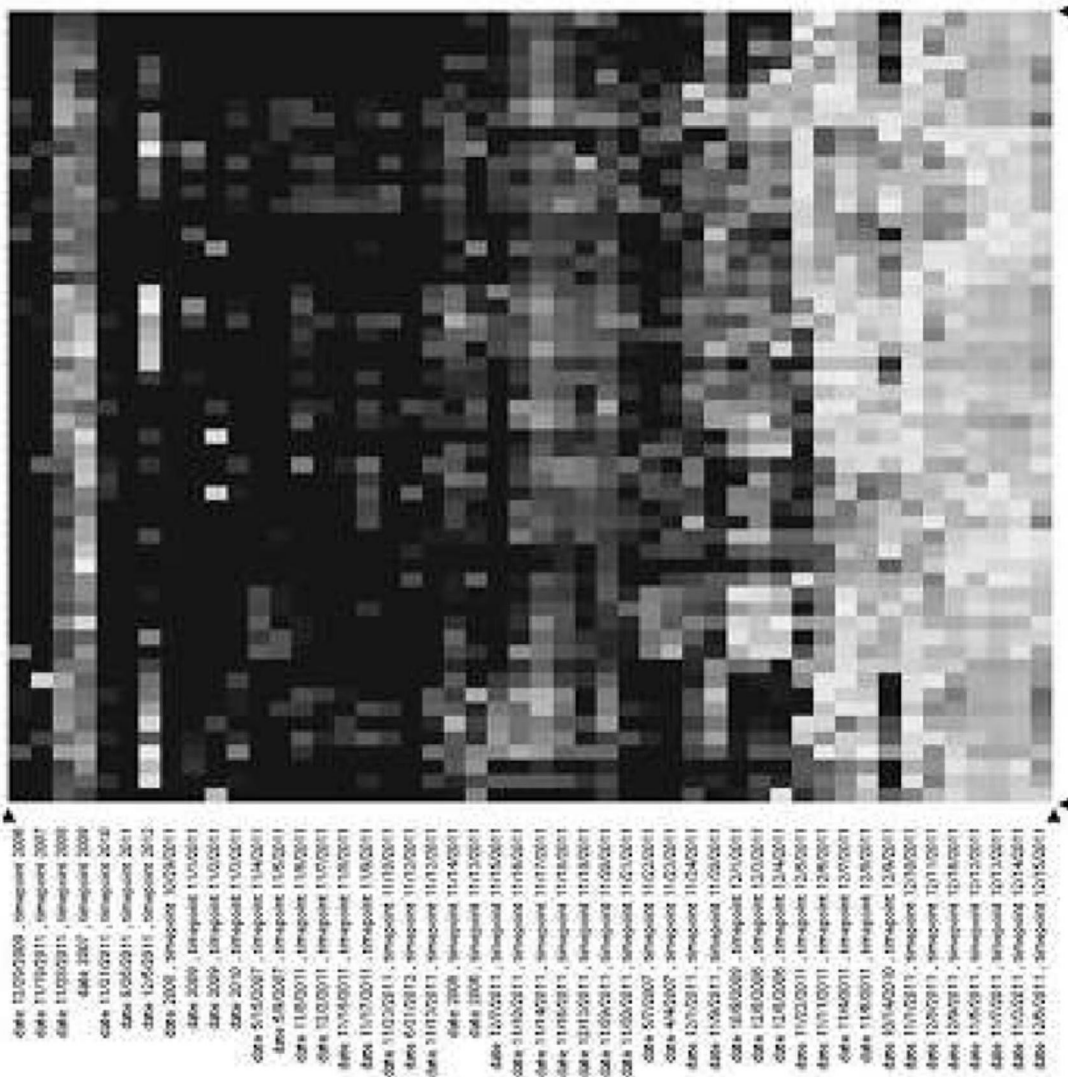


图22

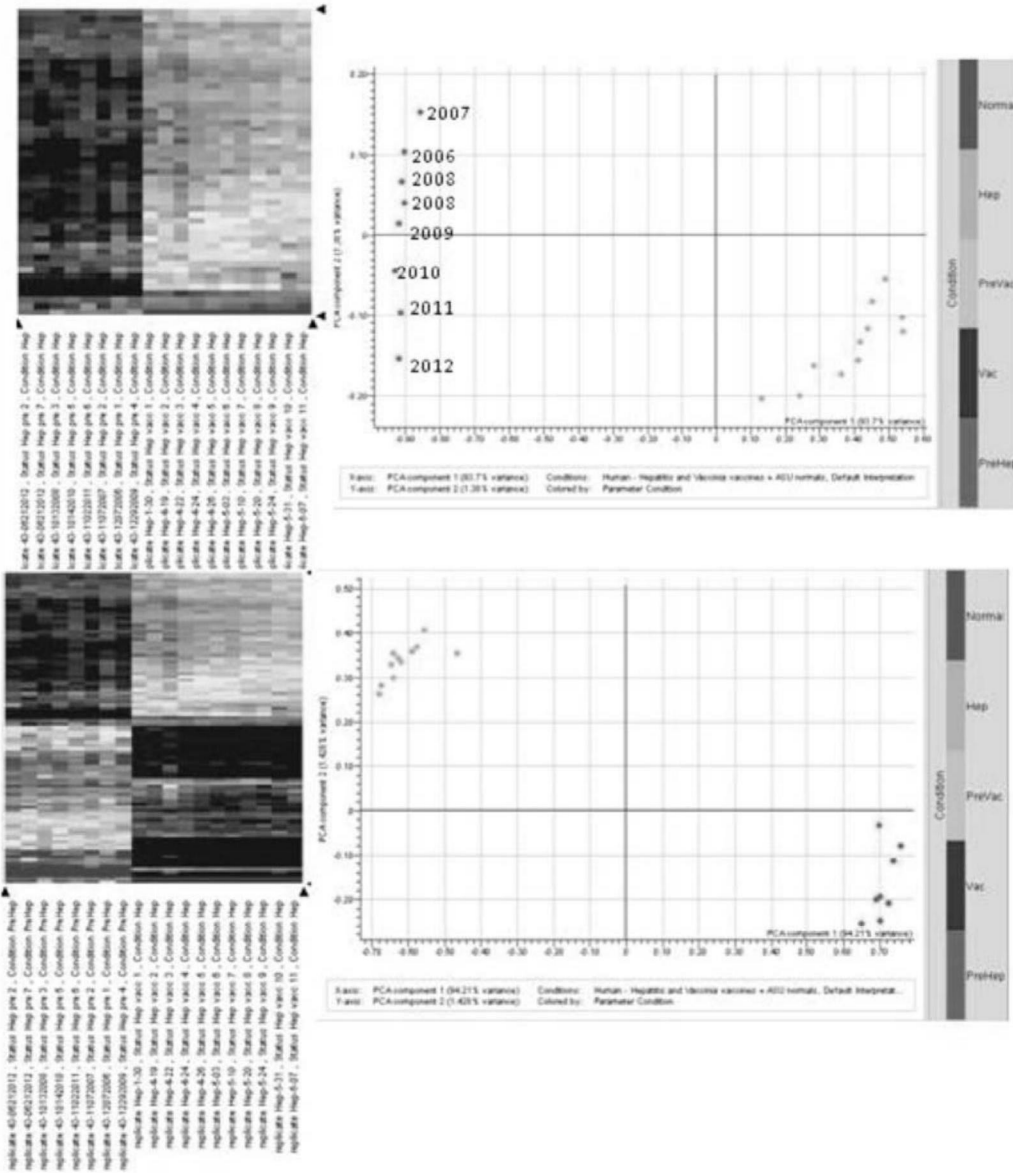
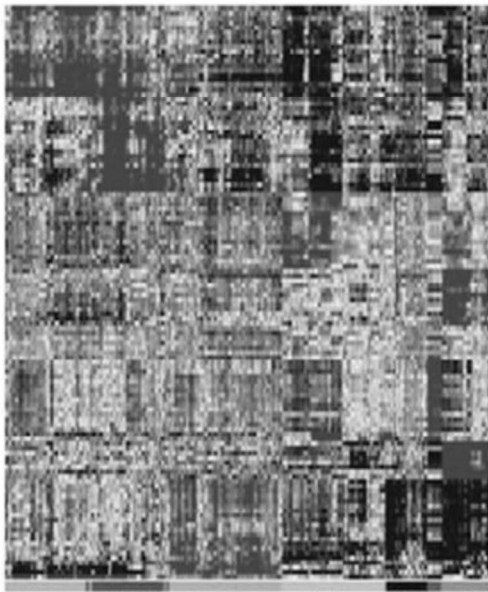


图23

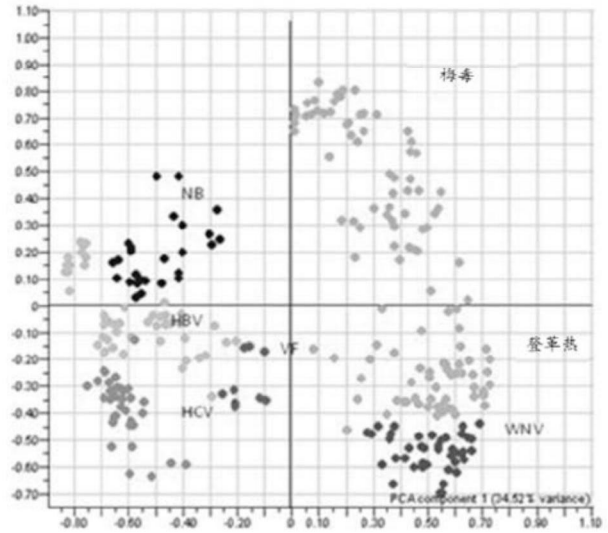


登革热 WNV 梅毒 HBV NB VF HCV

NB LOOCV: 255个正确的, 19个不正确的

SVM LOOCV: 260个正确的, 14个不正确的

A图



B图

图24

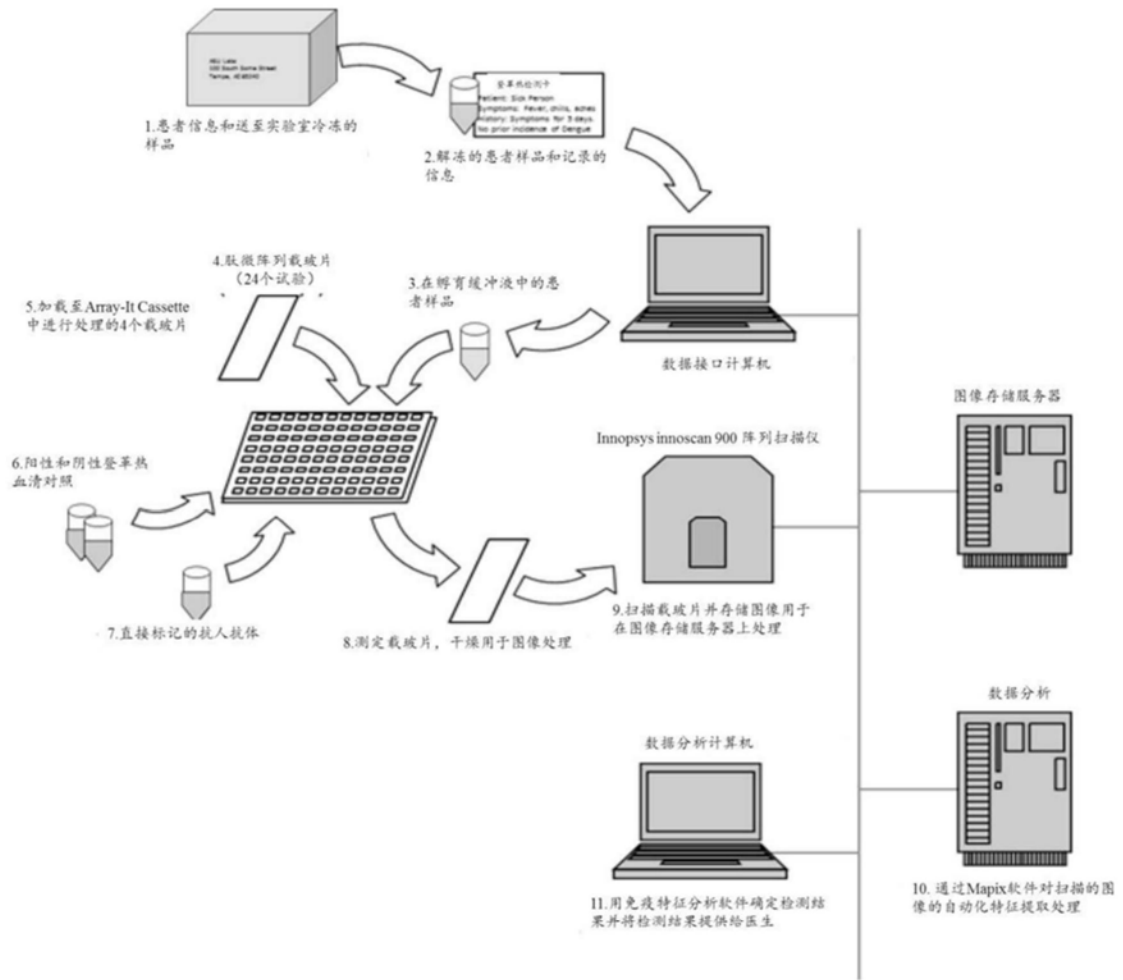


图25

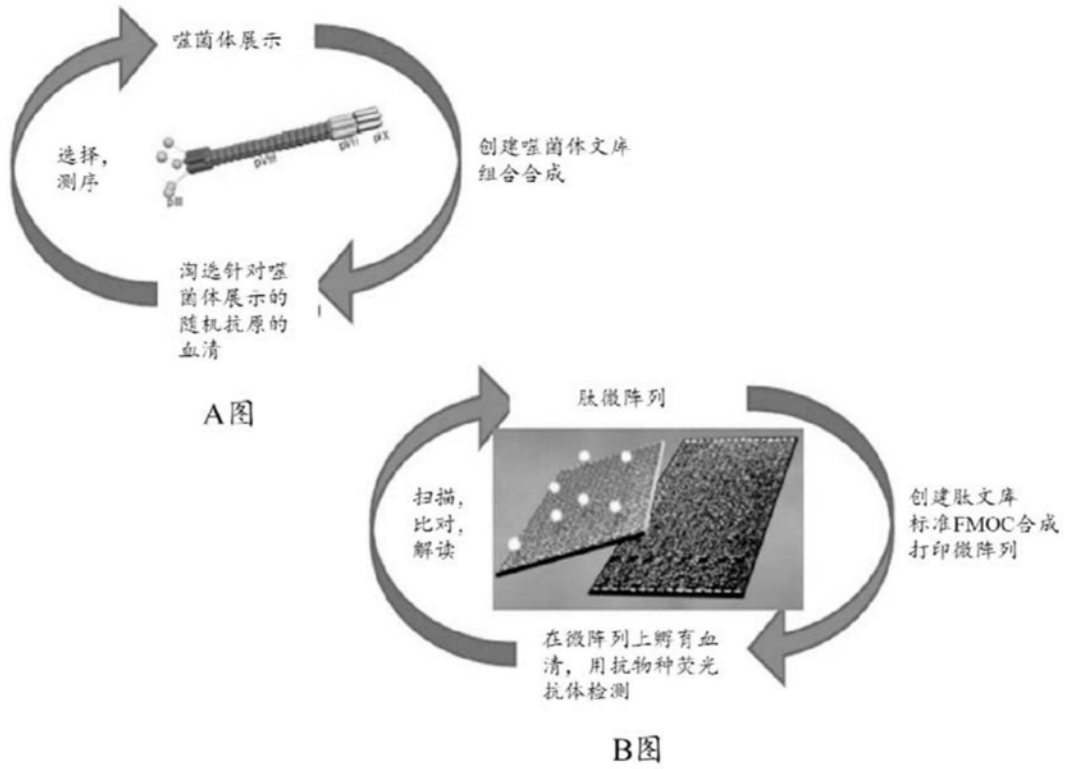


图26

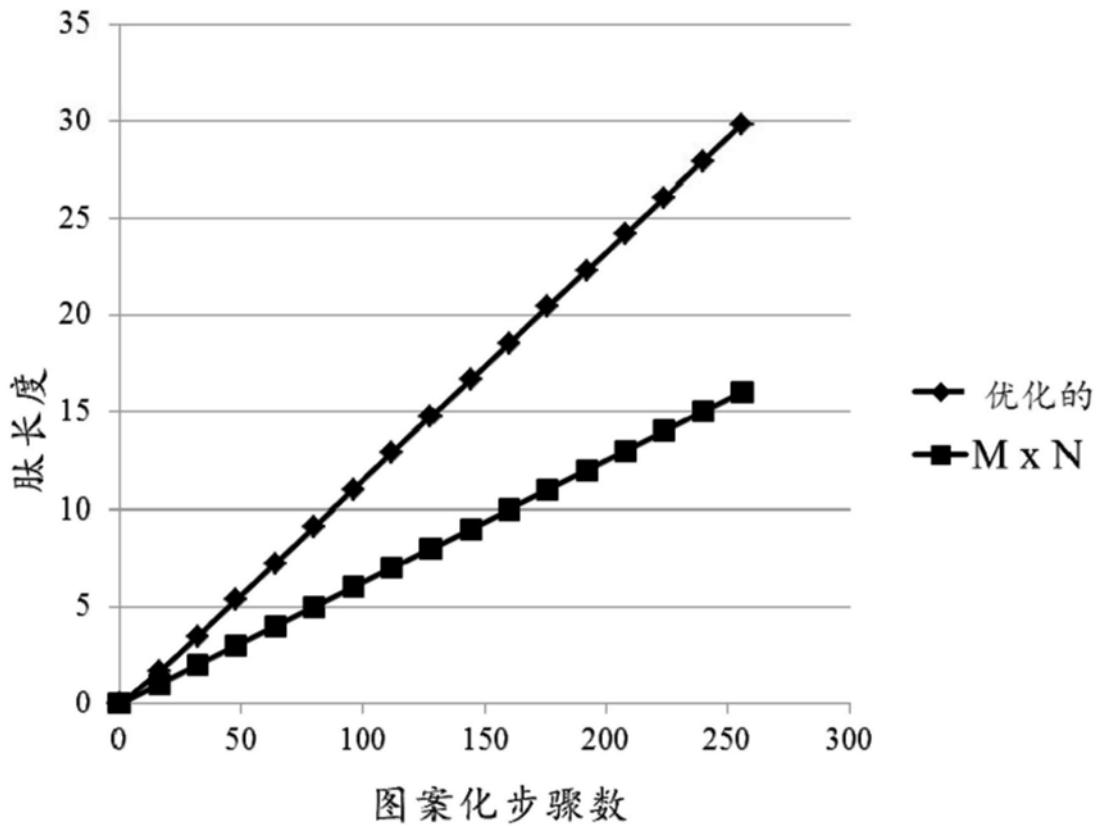


图27

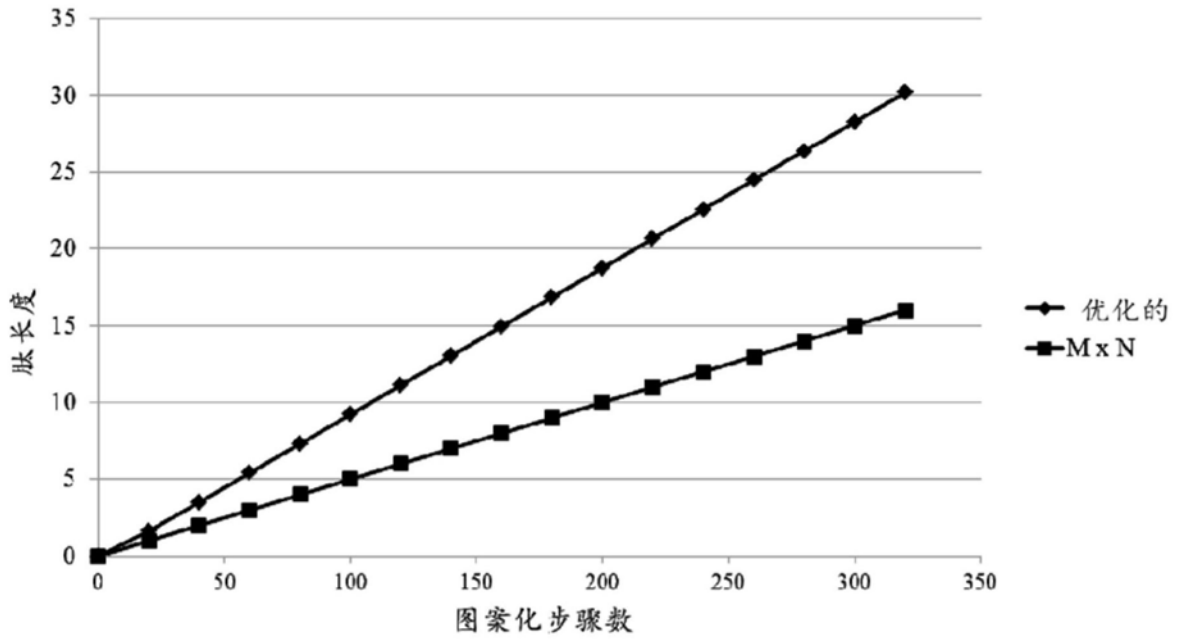


图28

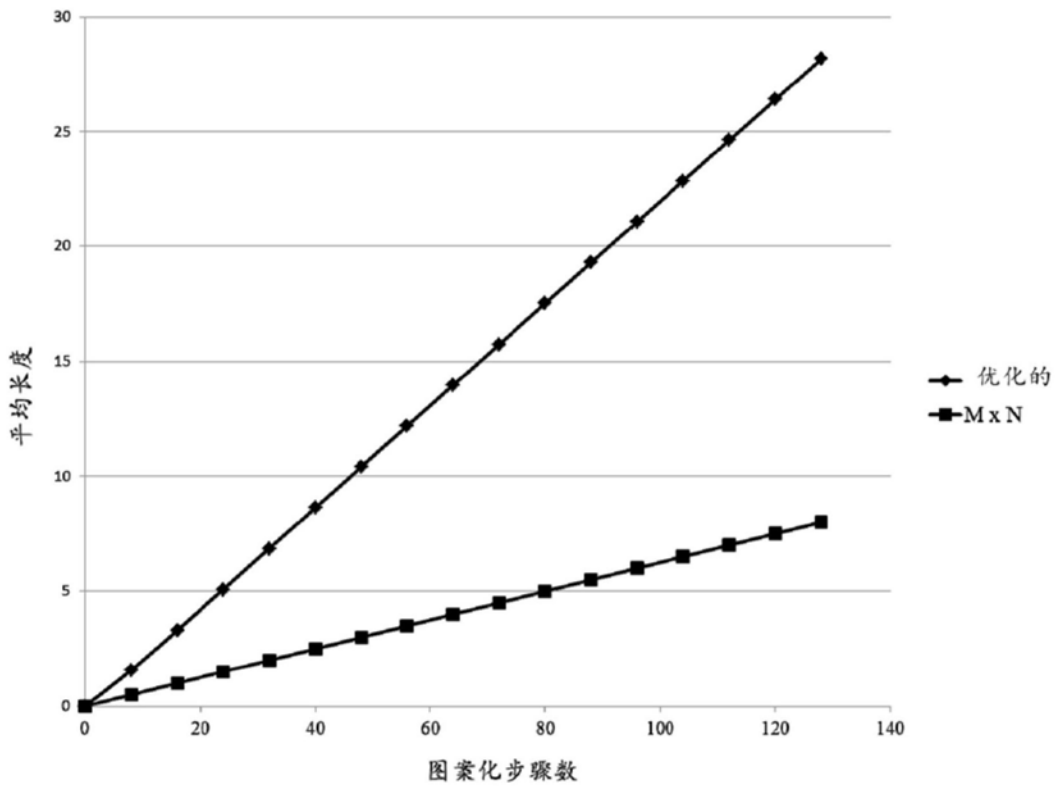


图29

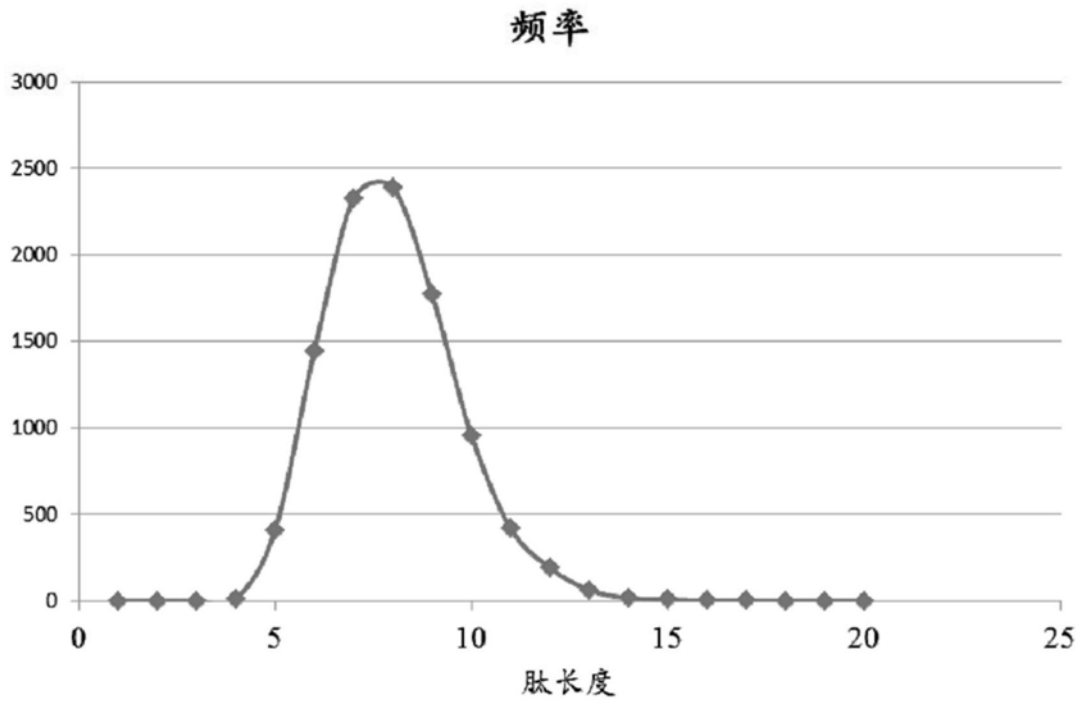


图30

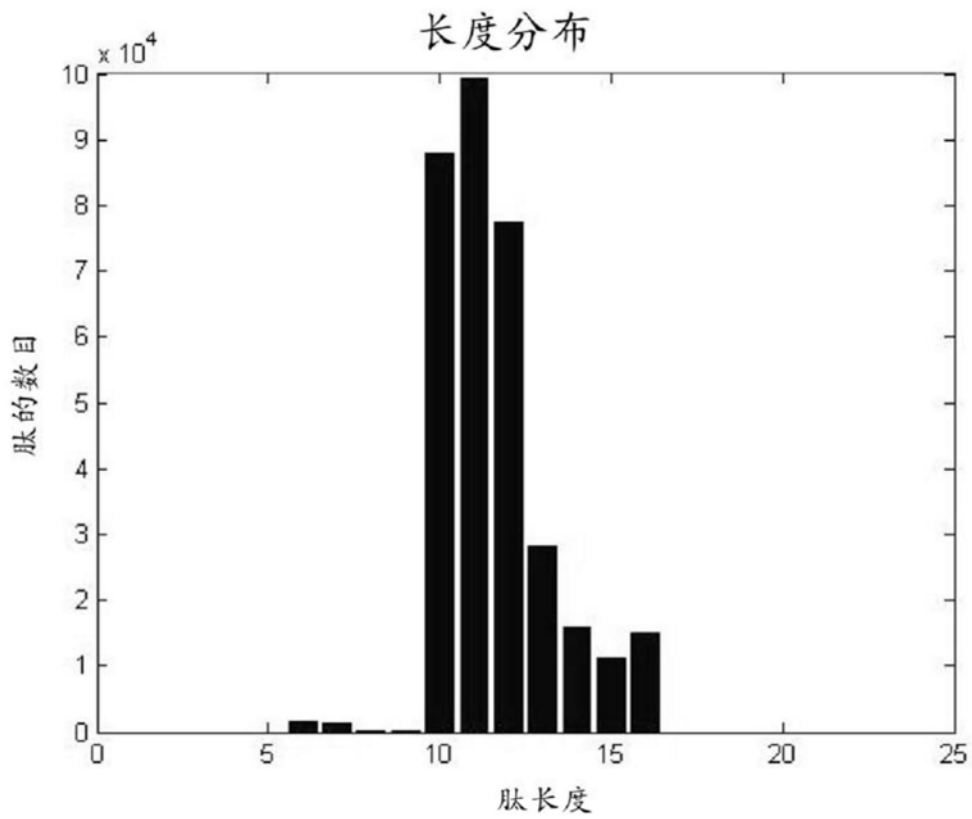


图31

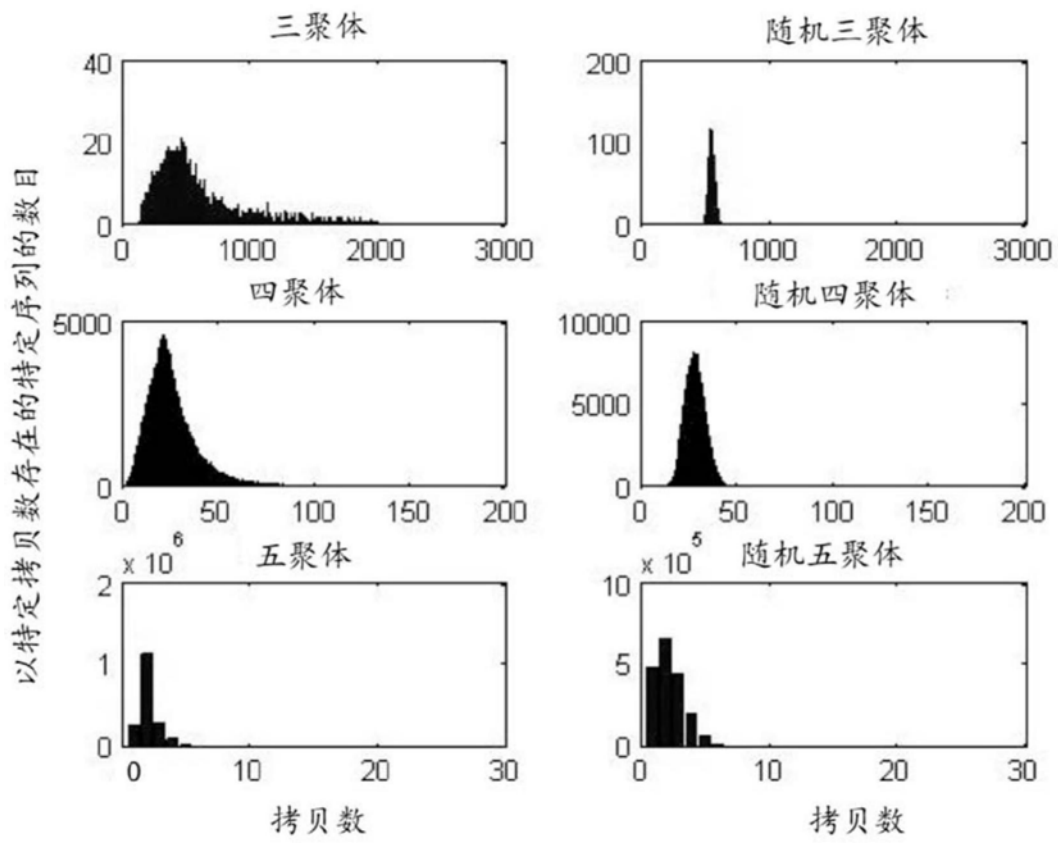


图32

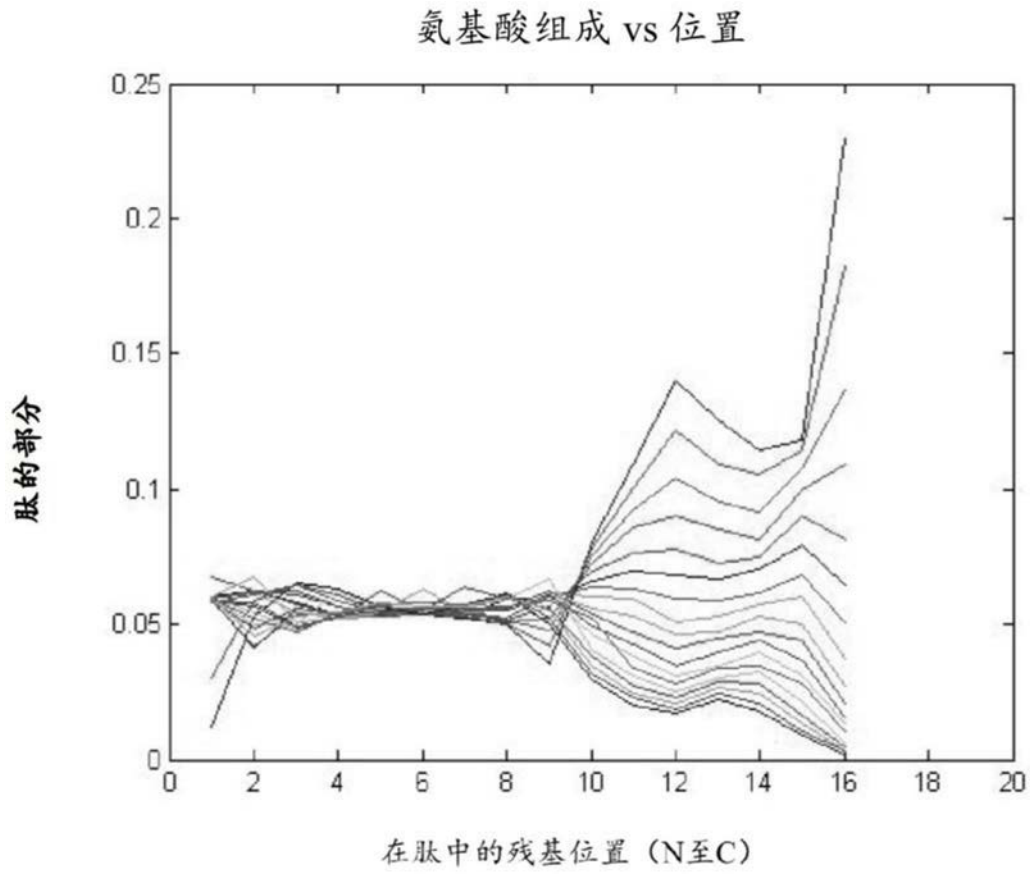


图33

专利名称(译)	免疫特征分析：通向早期诊断和健康监测的途径		
公开(公告)号	<a href="#">CN104781668B</a>	公开(公告)日	2018-07-24
申请号	CN201380056857.9	申请日	2013-08-29
[标]发明人	斯蒂芬阿尔伯特约翰斯顿 菲利普斯塔福德 尼尔伍德伯里		
发明人	斯蒂芬·阿尔伯特·约翰斯顿 菲利普·斯塔福德 尼尔·伍德伯里		
IPC分类号	G01N33/53 G01N33/00 G01N33/567 A61K38/00 C07K1/00		
CPC分类号	G01N33/54306 B01J19/0046 B01J2219/00659 B01J2219/00722 B01J2219/00725 B01J2219/00729 B82Y5/00 B82Y15/00 B82Y30/00 G01N33/6845 G01N33/6854 G01N33/6893 G01N2410/00 G01N2800 /52 G01N2800/60		
代理人(译)	刘小立 郑霞		
审查员(译)	贾静		
优先权	61/694598 2012-08-29 US		
其他公开文献	CN104781668A		
外部链接	<a href="#">SIPO</a>		

摘要(译)

健康是一种复杂的状态，其代表了几乎所有人类活动和互动的不断变化的结果。本发明提供了用于健康监测、诊断、治疗和预防性保健的有效方法和阵列。本发明监测来自受试者的广泛的标识分子，如循环抗体，并且本发明评估这些分子与肽阵列的结合的模式。用本发明的方法表征此类分子与肽阵列的结合模式提供了对受试者健康状态的稳健测量。

