



# (12)发明专利申请

(10)申请公布号 CN 111448569 A

(43)申请公布日 2020.07.24

(21)申请号 201880079402.1

(22)申请日 2018.12.04

(30)优先权数据

62/595,143 2017.12.06 US

(85)PCT国际申请进入国家阶段日

2020.06.08

(86)PCT国际申请的申请数据

PCT/EP2018/083434 2018.12.04

(87)PCT国际申请的公布数据

WO2019/110561 EN 2019.06.13

(71)申请人 文塔纳医疗系统公司

地址 美国亚利桑那州

(72)发明人 J·布雷德诺 A·洛萨库勒

(74)专利代理机构 北京坤瑞律师事务所 11494

代理人 岑晓东

(51)Int.Cl.

G06K 9/00(2006.01)

G06K 9/46(2006.01)

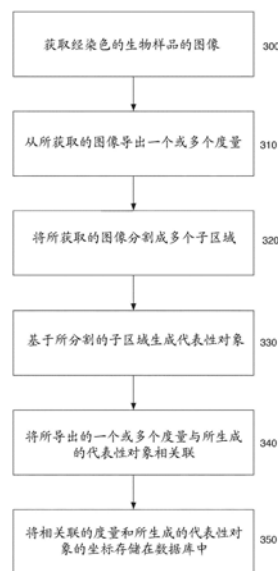
权利要求书3页 说明书30页 附图17页

## (54)发明名称

存储和检索数字病理学分析方法

## (57)摘要

本公开文本主要涉及用于分析、存储和/或检索与具有不规则形状的生物对象相关联的信息的自动化系统和方法。在一些实施方案中,所述系统和方法基于输入图像中的局部颜色、纹理和/或强度将输入图像划分成多个子区域,其中每个子区域代表有生物学意义的数据。



1. 一种存储图像分析数据的方法,所述图像分析数据从具有至少一种染色剂的生物样本的图像中导出,所述方法包括:

(a) 从所述图像中导出一个或多个特征度量;

(b) 将所述图像分割成多个子区域,每个子区域包括在染色存在、染色强度、或局部纹理中的至少一个方面基本一致的像素;

(c) 基于所述多个分割子区域生成多个代表性对象;

(d) 将所述多个代表性对象中的每一个与所导出的特征度量相关联;以及

(e) 将每个代表性对象的坐标与所述相关联的导出特征度量一起存储在数据库中。

2. 根据权利要求1所述的方法,其中将所述图像分割成所述多个子区域包括导出超像素。

3. 根据权利要求2所述的方法,其中所述超像素是通过以下步骤来导出的:(i) 用局部k均值聚类对像素进行分组;以及(ii) 使用连通分量算法将小的孤立区域合并到最近的大的超像素中。

4. 根据权利要求1至3中任一项所述的方法,其中将所述图像分割成所述多个子区域包括将采样网格覆盖在所述图像上,所述采样网格定义具有预定大小和形状的非重叠区。

5. 根据权利要求1至4中任一项所述的方法,其中所述子区域具有MxN大小,其中M的范围在50像素到100像素,并且其中N的范围在50像素到大约100像素。

6. 根据权利要求1至5中任一项所述的方法,其中所述代表性对象包括满足预定染色强度阈值的子区域的轮廓。

7. 根据权利要求1至6中任一项所述的方法,其中所述代表性对象包括种子点。

8. 根据权利要求7所述的方法,其中通过计算所述多个子区域中每一个的形心来导出所述种子点。

9. 根据权利要求6所述的方法,其中所导出的特征度量包括染色强度,并且其中计算每个生成的代表性对象轮廓内的所有像素的平均染色强度。

10. 根据权利要求1至7所述的方法,其中所导出的特征度量包括表达得分,并且其中对应于每个生成的子区域内的区域的平均表达得分与所生成的多个代表性对象相关联。

11. 根据权利要求1至7中任一项所述的方法,其进一步包括从所述数据库中检索所存储的坐标和相关联的特征度量数据,并将所述检索到的数据投影到所述图像上。

12. 一种用于从包含至少一种染色剂的生物样品的图像中导出对应于不规则形状细胞的数据的系统,所述系统包括:(i) 一个或多个处理器、和(ii) 联接到所述一个或多个处理器的存储器,所述存储器用于存储计算机可执行指令,当所述指令被所述一个或多个处理器执行时,使得所述系统执行操作,所述操作包括:

(a) 从所述图像中导出一个或多个特征度量;

(b) 在所述图像内生成多个子区域,每个子区域具有特性相似的像素,所述特性选自颜色、亮度和/或纹理;

(c) 基于所生成的多个子区域计算一系列代表性对象;以及

(d) 将从所述图像中导出的所述一个或多个特征度量与所述一系列计算出的代表性对象中的每一个的计算所得坐标相关联。

13. 根据权利要求12所述的系统,其中将所述图像分割成所述多个子区域包括导出超

像素。

14. 根据权利要求12至13中任一项所述的系统,其中使用基于图形的方法或基于梯度上升的方法中的一种方法来导出所述超像素。

15. 根据权利要求12至14中任一项所述的系统,其中所述超像素是通过以下步骤来导出的:(i)用局部k均值聚类对像素进行分组;以及(ii)使用连通分量算法将小的孤立区域合并到最近的大超像素中。

16. 根据权利要求12至15中任一项所述的系统,其中所述代表性对象包括满足预定染色强度阈值的子区域的轮廓。

17. 根据权利要求12至16中任一项所述的系统,其中所述代表性对象包括种子点。

18. 根据权利要求12至17中任一项所述的系统,其中所述操作进一步包括将所导出的一个或多个特征度和相关联的计算出的代表性对象坐标存储在数据库中。

19. 根据权利要求12至18中任一项所述的系统,其中所述一个或多个导出的特征度量包括选自阳性百分比、H得分和染色强度的至少一个表达得分。

20. 根据权利要求12至19中任一项所述的系统,其中为所述图像内的感兴趣区域导出对应于不规则形状细胞的数据。

21. 根据权利要求20所述的系统,其中所述感兴趣区域是由医学专业人员标注的所述图像的区域。

22. 一种存储用于分析与具有不规则形状的生物对象相关联的数据的指令的非暂时性计算机可读介质,所述指令包括:

(a) 用于从生物样品的图像中导出一个或多个特征度量的指令,所述生物样品包括至少一种染色剂;

(b) 用于通过将具有相似特性的像素分组来将所述图像划分成一系列子区域的指令,所述特性选自颜色、亮度、和/或纹理;

(c) 用于基于所述一系列划分的子区域计算多个代表性对象的指令;和

(d) 用于将从所述图像中导出的所述一个或多个特征度量与所述多个计算出的代表性对象中的每一个的计算所得坐标相关联的指令。

23. 根据权利要求22所述的非暂时性计算机可读介质,其中,将所述图像划分成所述一系列子区域包括计算超像素。

24. 根据权利要求22至23中任一项所述的非暂时性计算机可读介质,其中使用归一化切割算法、凝聚聚类算法、快速移位算法、涡轮像素(turbopixel)算法、或简单线性迭代聚类算法中的一种来计算所述超像素。

25. 根据权利要求22至24中任一项所述的非暂时性计算机可读介质,其中使用简单的迭代聚类来生成所述超像素,并且其中超像素大小参数被设置在大约40像素到大约400像素之间,并且其中紧密度参数被设置在大约10到大约100之间。

26. 根据权利要求22至25中任一项所述的非暂时性计算机可读介质,其中所述超像素是通过以下步骤来计算的:(i)用局部k均值聚类对像素进行分组;以及(ii)使用连通分量算法将小的孤立区域合并到最近的大超像素中。

27. 根据权利要求22至26所述的非暂时性计算机可读介质,其中所述生物样品至少用FAP染色,并且其中所导出的一个或多个特征度量包括FAP染色强度或FAP阳性百分比中的至少

一者。

28. 根据权利要求27所述的非暂态计算机可读介质,其中计算子区域内所有像素的平均FAP阳性百分比。

29. 根据权利要求27所述的非暂态计算机可读介质,其中计算子区域内所有像素的平均FAP染色强度。

30. 根据权利要求22至26中任一项所述的非暂时性计算机可读介质,其中所述代表性对象包括多边形轮廓和种子点中的至少一者。

31. 根据权利要求22至26中任一项所述的非暂时性计算机可读介质,其进一步包括用于将所导出的一个或多个特征度量和相关联的计算出的代表性对象坐标存储在数据库中的指令。

32. 根据权利要求31所述的非暂时性计算机可读介质,其进一步包括用于将存储的信息投影到所述生物样品图像上的指令。

## 存储和检索数字病理学分析结果的方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2017年12月6日提交的美国临时专利申请No.62/595,143的申请日的权益,其披露内容通过引用整体并入本文中。

### 背景技术

[0003] 数字病理学涉及将整个组织病理学或细胞病理学玻片扫描成可在计算机屏幕上解释的数字图像。这些图像随后将由成像算法处理或由病理学家解释。为了检查组织切片(实际上是透明的),使用选择性结合细胞成分的有色组织化学染色剂制备组织切片。临床医生或计算机辅助诊断(CAD)算法使用颜色增强或染色的细胞结构来识别疾病的形态学标记,并相应地进行治疗。观察该测定可以实现多种过程,包括疾病诊断、对治疗反应的评估、以及研发抗击疾病的新药物。

[0004] 免疫组织化学(IHC)玻片染色可以用于识别组织切片的细胞中的蛋白质,并且因此广泛地用于对诸如生物组织中的癌性细胞和免疫细胞等不同类型的细胞的研究中。因此,可以在研究中使用IHC染色以理解癌组织中免疫细胞(诸如,T细胞或B细胞)的差异表达的生物标记的分布和位置以用于免疫应答研究。例如,肿瘤经常包含免疫细胞的浸润液,该浸润液可以防止肿瘤的发展或有利于肿瘤的向外生长。

[0005] 原位杂交(ISH)可用于寻找遗传异常或病状的存在,例如在显微镜下观察时在形态学上表现为恶性的细胞中特异性地致癌基因扩增。ISH使用与靶基因序列或转录物反义的标记的DNA或RNA探针分子来检测或定位细胞或组织样品中的靶核酸靶基因。通过将固定在玻片上的细胞或组织样品暴露于标记的核酸探针来进行ISH,所述核酸探针能够与细胞或组织样品中的给定靶基因特异性杂交。通过将细胞或组织样品暴露于已经用多个不同核酸标签标记的多个核酸探针,可以同时分析几个靶基因。通过利用具有不同发射波长的标记,可以在单个步骤中对单个靶细胞或组织样品进行同时多色分析。

### 发明内容

[0006] 本公开文本主要涉及用于分析和存储与具有不规则形状的生物对象(例如成纤维细胞或巨噬细胞)相关联的数据的自动化系统和方法。本公开还涉及使用中分辨率分析(或中分辨率分析)法(即,将具有相似属性(例如,染色强度、染色存在、和/或纹理)的像素分组为“子区域”的方法)来分析和存储与生物对象相关联的数据的自动化系统和方法。

[0007] 在数字病理学中,图像是从安装在玻片上并被染色以识别生物标记的生物样本(例如,组织样本)获取的。可以在高倍显微镜下评估生物样品,或者用对感兴趣的生物对象进行检测和分类的数字病理学算法对其进行自动分析。例如,感兴趣的对象可以是细胞、血管、腺体、组织区域等。任何导出的信息都可以存储在数据库中以供以后检索,并且该数据库可以包括感兴趣的生物结构的存在、缺失、空间关系和/或染色属性的统计。本领域技术人员将会理解,清晰区分的细胞(例如,肿瘤细胞或免疫细胞)的分析结果的存储和检索相对简单,因为这样的细胞可以由每个细胞中心位置的点表示并存储在数据库中(例如,参见

图4)。类似地,具有明确定义的尺寸和形状的生物对象(例如,血管)可以由简单的轮廓表示,其中轮廓的坐标可以存储在数据库中,用于以后的检索和/或进一步的分析(本文中也称为“多边形”或“多边形轮廓”)。

[0008] 另一方面,一些感兴趣的生物结构,例如成纤维细胞或巨噬细胞,具有不规则的形状。这些类型的细胞组可以围绕彼此或其他细胞延伸(见图5)。因此,通常很难通过观察者或自动算法来精确地单独识别这些不规则形状的细胞。相反,这些细胞通常仅通过其被染色的细胞质或膜的局部存在来识别,而不识别单个细胞。

[0009] 虽然使用高分辨率分析来分析和存储这种不规则形状的结构是可能的,但是这种方法通常需要大量的计算机资源(计算时间和/或存储资源)。事实上,存储感兴趣的生物结构的所有像素信息(例如,每个像素的分析结果)的高分辨率分析法被认为消耗了太多的软件和硬件资源(例如,用于处理或显示信息的存储器和处理器),并且最终可能不会为某些生物对象提供有意义的结果。

[0010] 还可以使用低分辨率分析来分析这种不规则结构,其中这种低分辨率数据表示可以将几个单独的细胞“聚团”成单个对象,以存储在数据库中。作为示例,图6A和图6B示出了对肿瘤(黄色,620)和成纤维细胞(紫色,610)染色的IHC图像的示例,其由围绕一组相关细胞的大多边形轮廓(红色,630)表示,对于不期望的区域具有排除“孔”(青色,640)。在该示例中,分析结果在可能包含大量具有不同特征(例如,形状、大小、染色强度等)的单个细胞的大区域(红色轮廓,630)上进行平均。例如,关于图6B,画出轮廓的成纤维细胞活化蛋白(FAP)阳性面积为 $928.16\mu\text{m}^2$ ,计算的FAP阳性平均强度为0.26。给定如此大的像素面积区域中的平均强度,0.26的平均强度是相当粗略的,以指示并代表该图像中的整个FAP阳性。不希望受任何特定理论的约束,认为当存储的结果随后被用于下游处理时,这种低分辨率分析法可能导致准确度的损失。因此,认为由于染色细胞的这种异质性,该方法不能局部呈现这种感兴趣的生物结构区域的实际细节。

[0011] 与上述高分辨率和低分辨率分析方法相比,本公开文本提供了通过将图像分割成多个子区域来使用中分辨率分析法导出对应于不规则形状细胞的数据的系统和方法,所述子区域具有相似的图像属性(例如,纹理、强度、或颜色中的至少一个)。

[0012] 鉴于前述内容,在当前公开文本的一个方面中,是一种存储从具有至少一种染色剂的生物样本的图像导出的图像分析数据的方法,包括:(a)从所述图像中导出一个或多个特征度量;(b)将所述图像分割成多个子区域,每个子区域包括在染色存在、染色强度、或局部纹理中的至少一个方面基本一致的像素;(c)基于所述多个分割子区域生成多个代表性对象;(d)将所述多个代表性对象中的每一个与所导出的特征度量相关联;以及(e)将每个代表性对象的坐标与所述相关联的导出特征度量一起存储在数据库中。本领域技术人员将理解,至少步骤(a)和(b)可以以任何顺序执行。在一些实施方案中,将所述图像分割成所述多个子区域包括导出超像素。在一些实施方案中,所述超像素是通过以下步骤来导出的:(i)用局部k均值聚类对像素进行分组;以及(ii)使用连通分量算法将小的孤立区域合并到最近的大的超像素中。不希望受任何特定理论的约束,认为超像素(作为子区域)具有感知意义,使得每个超像素是感知上一致的单元,即超像素中的所有像素在颜色和纹理上可能是一致的。在一些实施方案中,连通分量标记扫描图像,并基于像素连通性将其像素分组为分量,即,连通分量中的所有像素共享相似的像素强度值,并以某种方式彼此连通。

[0013] 在一些实施方案中,将所述图像分割成所述多个子区域包括将采样网格覆盖在所述图像上,所述采样网格定义具有预定大小和形状的非重叠区。在一些实施方案中,所述子区域具有MxN大小,其中M的范围在50像素到100像素,并且其中N的范围在50像素到大约100像素。

[0014] 在一些实施方案中,所述代表性对象包括满足预定染色强度阈值的子区域的轮廓。在一些实施方案中,代表性对象包括种子点。在一些实施方案中,通过计算所述多个子区域中每一个的形心来导出所述种子点。在一些实施方案中,所导出的特征度量是染色强度,并且其中计算每个生成的代表性对象轮廓内的所有像素的平均染色强度。在一些实施方案中,所导出的特征度量是表达得分,并且其中对应于每个生成的子区域内的区域的平均表达得分与所生成的多个代表性对象相关联。在一些实施方案中,所述方法还包括从所述数据库中检索所存储的坐标和相关联的特征度量数据,并将所述检索到的数据投影到所述图像上。在一些实施方案中,相应子区域内的分析结果(例如,强度、面积)可以以代表该子区域的像素数据的平均像素测量值的形式存储。

[0015] 在一些实施方案中,所述生物样品被两种细胞膜染色剂染色。在一些实施方案中,生物样品用细胞膜染色剂和细胞核染色剂中的至少一种染色。在一些实施方案中,所述生物样品至少用FAP染色,并且其中所导出的一个或多个特征度量包括FAP染色强度或FAP阳性百分比中的至少一者。在一些实施方案中,计算子区域内所有像素的平均FAP阳性百分比。在一些实施方案中,计算子区域内所有像素的平均FAP染色强度。在一些实施方案中,样品用FAP和H&E染色。在一些实施方案中,样品用FAP和另一种细胞核或细胞膜染色剂染色。

[0016] 在一些实施方案中,作为输入接收的图像首先被解混合成图像通道图像,例如特定染色剂的图像通道图像。在一些实施方案中,在图像分析之前选择感兴趣区域。

[0017] 在本公开文本的另一方面,是一种用于从包含至少一种染色剂的生物样品的图像中导出对应于不规则形状细胞的数据的系统,所述系统包括:(i)一个或多个处理器,和(ii)联接到所述一个或多个处理器的存储器,所述存储器用于存储计算机可执行指令,当所述指令被所述一个或多个处理器执行时,使得所述一个或多个处理器执行操作,所述操作包括:(a)从所述图像中导出一个或多个特征度量;(b)在所述图像内生成多个子区域,每个子区域具有特性相似的像素,所述特性选自颜色、亮度和/或纹理;(c)基于所生成的多个子区域计算一系列代表性对象;以及(d)将从所述图像中导出的所述一个或多个特征度量与所述一系列计算出的代表性对象中的每一个的计算所得坐标相关联。在一些实施方案中,通过对像素进行分组来形成子区域,这些像素(i)相邻,(ii)具有相似的感知上有意义的属性(例如颜色、亮度、和/或纹理),并且(iii)在生物属性(例如生物结构、生物结构的染色属性、细胞特征、细胞组)方面足够均匀。在一些实施方案中,子区域中的像素对于感兴趣的生物对象具有相似的属性和描述性统计,例如不规则形状的细胞,包括但不限于成纤维细胞和巨噬细胞。

[0018] 在一些实施方案中,将所述图像分割成所述多个子区域包括导出超像素。在一些实施方案中,使用基于图形的方法或基于梯度上升的方法中的一种方法来导出所述超像素。在一些实施方案中,所述超像素是通过以下步骤来导出的:(i)用局部k均值聚类对像素进行分组;以及(ii)使用连通分量算法将小的孤立区域合并到最近的大的超像素中。

[0019] 在一些实施方案中,所述代表性对象包括满足预定染色强度阈值的子区域的轮

廓。在一些实施方案中,所述代表性对象包括种子点。在一些实施方案中,所述系统还包括用于将所导出的一个或多个特征度量和相关联的计算出的代表性对象坐标存储在数据库中的指令。在一些实施方案中,所述一个或多个导出的特征度量包括选自阳性百分比、H得分或染色强度的至少一个表达得分。在一些实施方案中,为所述图像内的感兴趣区域导出对应于不规则形状细胞的数据。在一些实施方案中,所述感兴趣区域是由医学专业人员标注的所述图像的区域。

[0020] 在本公开文本的另一方面,是一种存储用于分析与具有不规则形状的生物对象相关联的数据的指令的非暂时性计算机可读介质,所述指令包括:(a)用于从生物样品的图像中导出一个或多个特征度量的指令,所述生物样品包括至少一种染色剂;(b)用于通过将具有相似特性的像素分组来将所述图像划分成一系列子区域的指令,所述特性选自颜色、亮度、和/或纹理;(c)用于基于所述一系列划分的子区域计算多个代表性对象的指令;和(d)用于将从所述图像中导出的所述一个或多个特征度量与所述多个计算出的代表性对象中的每一个的计算所得坐标相关联的指令。

[0021] 在一些实施方案中,将所述图像划分成所述一系列子区域包括计算超像素。在一些实施方案中,使用归一化切割算法、凝聚聚类算法、快速移位算法、涡轮像素(turbopixel)算法、或简单线性迭代聚类算法中的一种来计算所述超像素。在一些实施方案中,使用简单的迭代聚类来生成所述超像素,并且其中超像素大小参数被设置在大约40像素到大约400像素之间,并且其中紧密度参数被设置在大约10到大约100之间。在一些实施方案中,所述超像素是通过以下步骤来计算的:(i)用局部k均值聚类对像素进行分组;以及(ii)使用连通分量算法将小的孤立区域合并到最近的大的超像素中。

[0022] 在一些实施方案中,所述生物样品至少用FAP染色,并且其中所导出的一个或多个特征度量包括FAP染色强度或FAP阳性百分比中的至少一者。在一些实施方案中,计算子区域内所有像素的平均FAP阳性百分比。在一些实施方案中,计算子区域内所有像素的平均FAP染色强度。在一些实施方案中,所述代表性对象包括多边形轮廓和种子点中的至少一者。在一些实施方案中,所述存储器包括用于将所导出的一个或多个特征度量和相关联的计算出的代表性对象坐标存储在数据库中的指令。在一些实施方案中,所述存储器包括用于将存储的信息投影到所述生物样品图像上的指令。

[0023] 申请人已经表明,本文描述的系统和方法提供了一种用于存储不能由每个感兴趣的对象的单个位置或轮廓来限定的生物对象的分析结果的改进的解决方案。此外,申请人认为,与像素级高分辨率分析法相比,本文公开的系统和方法允许用于存储分析结果的存储空间减小,因为特定像素及其周围像素的分析结果一起存储在子区域中,子区域中的像素具有相似的属性或特性(例如颜色、亮度、纹理)。申请人还认为,该系统和方法在计算上是有效的,因为生成的子区域允许将图像的复杂性从几千个像素降低到更小、更易管理的子区域数量,从而允许更快地进一步检索和报告分析结果。申请人还认为子区域具有代表性,因为它们对于存储和表示分析结果来说不是太小或太大。最后,申请人提出,本文公开的系统和方法允许提高精度,特别是与低分辨率分析方法相比,因为与存储来自较大区域表示的信息相比,所生成的子区域描述了感兴趣的生物相关对象的属性或统计信息(即,子区域包括在染色存在、染色强度和纹理方面尽可能一致的像素)。本文进一步描述了这些和其他优点。



## 附图说明

[0024] 为了全面理解本公开的特征,参考了附图。在附图中,相同的附图标记始终用于标识相同的元件。

[0025] 根据一些实施方案,图1示出了包括图像获取装置和计算机系统的代表性数字病理学系统。

[0026] 图2给出了根据一些实施方案的可以在数字病理学系统中或者在数字病理学工作流程中使用的各种模块。

[0027] 图3给出了根据一些实施方案的流程图,该流程图示出了导出图像分析数据并将这种图像分析数据与生成的子区域相关联的各个步骤。

[0028] 图4提供了根据一些实施方案的高水平分辨率的肝癌细胞的数字病理学图像的示例。在图像分析处理和分类之后,可以存储分析结果并从数据库中检索以供显示(例如,作为位于细胞中心的标注点(红色=阳性染色肿瘤细胞(410),绿色=阴性染色肿瘤细胞(420))。每个标注点可以包含读出的信息,例如感兴趣的生物结构的存在、缺失、空间关系和染色属性的描述性统计。

[0029] 图5A至图5D示出了具有不同外观(例如,不规则的大小、形状和细胞边界)的形态不均匀的成纤维细胞的外观。在该图中,正常和活化的成纤维细胞分别在(A)和(C)中示出。(C)和(D)分别给出了正常成纤维细胞和活化成纤维细胞的苏木精和曙红染色(H&E)图像。

[0030] 图6A给出了与肿瘤细胞相关联的成纤维细胞的免疫组织化学(IHC)的例子,其中成纤维细胞(610)被染成紫色,肿瘤(620)被染成黄色。如图所示,成纤维细胞可以接触并具有非常不规则的形状,延伸到其他细胞之外或周围。

[0031] 图6B给出了成纤维细胞表达阳性的区域的低分辨率多边形轮廓(红色,630)和青色的排除区域(孔,640)的例子。

[0032] 图7示出了具有简单形状(例如,圆形)的子区域(710),其可以使用本文描述的中分辨率法与图像数据相关联。

[0033] 图8A给出了在IHC图像上的成纤维细胞区域中使用SLIC生成的超像素的示例。

[0034] 图8B提供了高倍放大的原始IHC图像,其中肿瘤细胞(830)被染成黄色,成纤维细胞(840)被染成紫色。

[0035] 图8C示出了根据一些实施方案的超像素的初始形状,其在调整正则化参数之前看起来类似于正方形。

[0036] 图8D示出了根据一些实施方案的超像素的最终表示,其中调整了SLIC算法中的正则化参数。

[0037] 图9A示出了根据一些实施方案的属于感兴趣区域(成纤维细胞区域)的子区域(这里是超像素)的多边形轮廓(黑色,910)。

[0038] 图9B根据一些实施方案给出了属于感兴趣的生物对象(成纤维细胞)的子区域(超像素)的多边形轮廓(黑色,920)和中心种子(绿色点,930)。

[0039] 图10A提供了将成纤维细胞(1010)的成纤维细胞活化蛋白(FAP)染成紫色和将上皮肿瘤(1020)的泛细胞角蛋白(PanCK)染成黄色的头颈癌组织的全玻片IHC图像的例子。

[0040] 图10B给出了多边形轮廓的例子,该多边形轮廓附有属于成纤维细胞区域的超像素(蓝色,1030)的分析结果,该结果可以存储在数据库中。

[0041] 图11给出了中心种子的例子,该中心种子附有它们对属于成纤维细胞区域的超像素(红色,1140)的分析结果,其可以存储在数据库中。

[0042] 图12提供了从全玻片超像素检索的FAP强度的直方图的例子。

[0043] 图13提供了说明根据一些实施方案的区域选择步骤的流程图。

[0044] 图14给出了生物样品图像内的六种不同标注形状和区域。

[0045] 图15示出了在(i)使用高分辨率分析法确定的FAP+区域和(ii)使用本文描述的示例性中分辨率(子区域)法之间的FAP阳性面积的百分比的一致性。

## 具体实施方式

[0046] 还应该理解,除非明确指出相反的情况,否则在本文要求保护的包括多于一个步骤或动作的任何方法中,该方法的步骤或动作的顺序不一定限于该方法的步骤或动作被叙述的顺序。

[0047] 如本文所使用的,单数术语“一个”、“一种”以及“该”包括复数个指示物,除非上下文中另外明确指示。类似地,词语“或”旨在包括“和”,除非上下文中另外明确指示。术语“包括”被定义为包含性的,使得“包括A或B”是指包括A、B或A和B。

[0048] 如本说明书和权利要求书中所用的,“或”应被理解为与如上所定义的“和/或”具有相同含义。例如,在将所列项目分开时,“或”或“和/或”应解释为包容性的,即包括所列元素中的多个元素或至少一个元素,但也包括一个以上元素,以及(可选地)其他未列出的项目。只有明确指示相互矛盾,否则诸如“只有一个”或“恰好一个”或者在权利要求中使用“由……组成”将指代恰好包括许多元件或元件列表中的一个元件。一般而言,如本文中所使用的术语“或”之后有诸如“两者之一”、“中的一个”、“中的仅一个”或“中的恰好一个”之类的排他性术语时仅应被解释为指示排他性备选方案(即,“一个或另一个但不是两个”)。“基本上由……组成”在权利要求中使用时的普通意义如同在专利法领域中使用的那样。

[0049] 术语“包括”、“包含”、“具有”等可互换地使用并且具有相同的含义。类似地,术语“包括”、“包含”、“具有”等可互换地使用并且具有相同的意思。具体而言,每个术语的定义与美国专利法中“包括”的一般定义一致,因此被解释为一个开放式术语,意思是“至少以下”,并且也被解释为不排除附加的特征、限制、方面等。因此,例如,“具有部件a、b和c的装置”意味着该装置至少包括部件a、b和c。类似地,短语“涉及步骤a、b和c的方法”意味着该方法至少包括步骤a、b和c。此外,虽然步骤和过程可以在本文中以特定的顺序概述,但是本领域技术人员将认识到顺序步骤和过程可以变化。

[0050] 如本说明书和权利要求书中所使用的,关于一个或多个元件的列表,短语“至少一个”应被理解为表示选自元件列表中的任何一个或多个元件的至少一个元件,但不一定包括元件列表中具体列出的每个元件中的至少一个元件,并且不排除元件列表中元件的任何组合。该定义还允许可选地存在除在短语“至少一个”所指代的元件列表内具体表示的元件之外的元件,而无论是与具体表示的那些元件相关还是不相关。因此,作为非限制性示例,“A和B中的至少一者”(或等同地,“A或B中的至少一者”,或等效地“A和/或B中的至少一者”)在一个实施方案中可以指代至少一个A,可选地包括一个以上A,而不存在B(并且可选地包括除B之外的元件);在另一个实施方案中,指代至少一个B,可选地包括一个以上B,而不存在A(并且可选地包括除A之外的元件);在又一个实施方案中,指代至少一个A,可选地包括

一个以上A和至少一个B,可选地包括一个以上B(和可选地包括其他元件);等等。

[0051] 如本文所用,术语“生物样品”(在本文中可与术语“生物样本”或“样本”互换使用)或“组织样品”(在本文中可与术语“组织样本”互换使用)是指从包括病毒在内的任何生物体获取的包括生物分子(如蛋白质、肽、核酸、脂质、碳水化合物或其组合)的任何样品。生物的其他示例包括哺乳动物(诸如人类;兽类,诸如猫、狗、马、牛和猪;以及实验动物,诸如小鼠、大鼠和灵长类动物)、昆虫、环节动物、蛛形纲动物、有袋动物、爬行动物、两栖动物、细菌和真菌。生物样品包括组织样品(例如组织切片和组织的针活检),细胞样品(例如细胞学涂片,例如巴氏涂片或血液涂片或通过显微切割获取的细胞样品),或细胞组分、片段或细胞器(例如通过裂解细胞并通过离心或其他方式分离它们的组分获取的)。生物样品的其他示例包括血液、血清、尿液、精液、粪便、脑脊液、间质液、粘液、泪液、汗液、脓、活检组织(例如,通过外科活组织检查或针活组织检查获取的)、乳头抽吸物、耳垢、乳汁、阴道液、唾液、拭子(例如口腔拭子),或任何包含来自第一生物样品的生物分子的材料。在某些实施方案中,本文使用的术语“生物样品”指从肿瘤制备的样品(例如均质或液化样品)或从受试者获取的肿瘤的一部分。

[0052] 如本文所用,术语“生物标记”或“标记”是指某些生物状态或病状的可测量的指标。具体地,生物标记可以是可以被特定地染色并且指示细胞的生物特征(例如,细胞类型或细胞的生理状态)的蛋白质或肽(诸如,表面蛋白质)。免疫细胞标记是选择性地指示与哺乳动物的免疫应答相关的特征的生物标记。生物标志物可以用于确定身体对疾病或病状的治疗反应如何或者受试者是否易患疾病或病状。在癌症的情况下,生物标记是指指示体内癌症存在的生物物质。生物标记可以是肿瘤分泌的分子或机体对癌症存在的特定反应。遗传、表观遗传学、蛋白质组学、糖组学和成像生物标记可用于癌症诊断、预后和流行病学。这种生物标记可以在非侵入性收集的生物流体如血液或血清中进行测定。几种基于基因和蛋白质的生物标志物已经用于患者护理,包括但不限于AFP(肝癌)、BCR-ABL(慢性髓性白血病)、BRCA1/BRCA2(乳腺癌/卵巢癌)、BRAF V600E(黑色素瘤/结肠直肠癌)、CA-125(卵巢癌)、CA19.9(胰腺癌)、CEA(结肠直肠癌)、EGFR(非小细胞肺癌)、HER-2(乳腺癌)、KIT(胃肠间质瘤)、PSA(前列腺特异性抗原)、S100(黑色素瘤)和许多其他生物标志物。生物标记可用作诊断(标识早期癌症)和/或预测(预测癌症的侵袭性和/或预测受试者对特定治疗的反应和/或癌症复发的可能性)。

[0053] 如本文所使用的,术语“图像数据”如本文所理解的涵盖从生物样品获取(诸如借助于光学传感器或传感器阵列)的原始图像数据或经过预处理的图像数据。具体地,图像数据可以包括像素矩阵。如本文所使用的,术语“免疫组织化学”是指一种通过检测抗原与特定结合药剂(诸如,抗体)的相互作用来确定样品中抗原的存在或分布的方法。在允许抗体-抗原结合的条件下,样品与抗体接触。抗体-抗原结合可以通过与抗体结合的可检测标记(直接检测)或通过特异性结合第一抗体的第二抗体结合的可检测标记(间接检测)来检测。本文使用的“掩码”是数字图像的派生物,其中掩码中的每个像素被表示为二进制值,例如“1”或“0”(或“真”或“假”)。通过使用所述掩码来覆盖数字图像,在对数字图像施加的进一步处理步骤中隐藏、移除或以他方式忽略或过滤掉映射到二进制值中的特定一个值的掩码像素的数字图像的所有像素。例如,可以通过将具有高于阈值的强度值的原始图像的所有像素指定为真并且在其他情况下指定为假来从原始数字图像中生成掩码,从而创建可以

过滤掉将被“假”掩码像素覆盖的所有像素的掩码。如本文所理解的，“多通道图像”涵盖从生物组织样品获得的数字图像，其中，不同的生物结构（诸如核和组织结构）同时被特定的荧光染料、量子点、色原等染色，其中的每一种发荧光或在不同光谱带中以其他方式可检测到，从而构成了多通道图像的通道之一。

#### [0054] 概览

[0055] 申请人已经开发了一种将具有不规则形状的生物对象的分析结果存储在数据库或其他非暂时性存储器中的系统和方法，所述生物对象包括例如成纤维细胞或巨噬细胞。分析结果可随后从数据库或存储器中检索，用于进一步分析或用于其他下游过程。分析结果也可以投影到输入图像或其他导出图像上；或者通过其他方式可视化。此外，本公开文本还允许调整生成的子区域的大小的能力（例如，通过增加或减少简单形状的大小；或者调整超像素算法的参数），便于以可调整的细节水平存储和报告分析结果。与本文描述的低分辨率分析方法相比，这被认为允许提高效率和准确度，其中保存了来自全局感兴趣区域的平均分析结果。

[0056] 如本文进一步描述的，所公开的系统和方法基于使用局部相似小区域（子区域）来存储分析结果的中分辨率分析法。子区域可以是简单的形状（例如，圆形、正方形）或复杂的形状（例如，超像素），并且用于存储整个玻片中每个小区域的局部分析结果。由本公开的中分辨率法定义的子区域将具有相似（或同质）属性（例如，染色存在（即，特定染色剂的存在或不存在）、染色强度（即，染色剂的相对强度（或量））、局部纹理（即，关于图像或图像的选定区域中颜色或强度的空间排列的信息））的像素进行分组，从而允许识别不规则形状的对象。在一些实施方案中，中分辨率法内的子区域具有从大约50到大约100像素的大小；或者在大约2,500像素<sup>2</sup>和大约10,000像素<sup>2</sup>之间的像素区域。当然，子区域可以具有任何大小，并且所述大小可以基于正在进行的分析的类型和/或正在研究的细胞的类型。

[0057] 本领域技术人员将会理解，中等水平的方法落在本文所述的高分辨率分析法和低分辨率分析法之间，使得在子区域水平上收集数据，子区域在比例上小于低分辨率分析中的感兴趣区域，并且明显大于高分辨率分析方法中的像素。“高分辨率分析”是指在像素级或基本上在像素级捕获的图像数据。另一方面，“低分辨率分析”指的是区域级分析，例如具有至少500像素乘500像素的大小的区域或者具有大于250,000像素<sup>2</sup>的大小的区域。本领域技术人员将理解，低分辨率分析方法将包括许多生物对象，例如多个不规则形状的细胞。

[0058] 本公开文本可以适用于具有不规则形状和/或大小的生物对象（包括成纤维细胞或巨噬细胞）的分析和存储。应当理解，本公开文本不限于成纤维细胞或巨噬细胞，而是可以扩展到具有未明确限定的大小或形状的任何生物对象。

[0059] 就成纤维细胞而言，成纤维细胞是构成由动物组织中的细胞外基质和胶原组成的结构框架或基质的细胞。这些细胞是动物体内最常见的结缔组织，对伤口愈合很重要。成纤维细胞有各种形状和大小，也有激活和未激活的形式（例如见图5A至图5D）。成纤维细胞（Fibroblast）是激活的形式（后缀“blast”指的是代谢活性细胞），而纤维细胞被认为活性较低。然而，有时成纤维细胞和纤维细胞并没有被指定为不同的，只是被称为成纤维细胞。形态学上，成纤维细胞可以通过其丰富的粗面内质网和相对较大的尺寸与纤维细胞区分开来。此外，成纤维细胞被认为与它们的相邻细胞接触，并且这种接触被认为是可能扭曲孤立细胞的形式粘连。本文提供的中分辨率分析法能够解释这些形态学差异，并且被认为非

常适合于存储关于成纤维细胞、巨噬细胞和其他不规则生物对象的信息。

[0060] 根据一些实施方案,用于成像和分析样本的数字病理学系统200在图1中示出。数字病理学系统200可以包括成像设备12(例如,具有用于扫描承载样本的显微镜切片的装置的设备)和计算机14,由此成像设备12和计算机可以通信地耦合在一起(例如,直接地或间接地通过网络20)。计算机系统14可以包括台式计算机、膝上型计算机、平板电脑等、数字电子电路、固件、硬件、存储器、计算机存储介质、计算机程序或指令集(例如,其中程序存储在存储器或存储介质中)、一个或多个处理器(包括编程处理器)、以及任何其他硬件、软件或固件模块或其组合。例如,图1中展示的计算系统14可以具有显示装置16和外壳18的计算机。计算机可以以二进制形式存储数字图像(本地地诸如存储在存储器、服务器或另一个网络连接装置中)。还可以将数字图像分成像素矩阵。像素可以包括由位深定义的具有一个或多个位的数字值。技术人员将了解到,可以利用其他计算机装置或系统,并且本文所描述的计算机系统可以通信地耦合到另外的部件,例如样本分析仪、显微镜、其他成像系统、自动玻片制备装备等。本文将进一步描述这些附加部件中的一些以及可以使用的各种计算机、网络等。

[0061] 通常,成像设备12(或包括存储在存储器中的预扫描图像的其他图像源)可以包括但不限于一个或多个图像捕获装置。图像捕获装置可以包括但不限于相机(例如,模拟相机、数字相机等)、光学器件(例如,一个或多个透镜、传感器聚焦透镜组,显微镜物镜等)、成像传感器(例如,电荷耦合装置(CCD)、互补金属氧化物半导体(CMOS)图像传感器等)、胶片等。在数字实施方案中,图像捕获装置可以包括协作以证明即时聚焦的多个透镜。图像传感器(例如,CCD传感器)可以捕获样本的数字图像。在一些实施方案中,成像设备12是明场成像系统、多光谱成像(MSI)系统或荧光显微镜系统。数字化的组织数据可以例如由图像扫描系统生成,诸如VENTANA MEDICAL SYSTEMS, Inc. (Tucson, Arizona)的VENTANA iSCAN HT扫描仪、或其他适合的成像装备。本文还描述了另外的成像装置和系统。本领域技术人员将理解,由成像装置12采集的数字彩色图像可以传统地由基本彩色像素组成。每个彩色像素可以在三个数字分量上编码,每个数字分量包括相同数量的位,每个分量对应于原色,通常是红色、绿色或蓝色,也由术语“RGB”分量表示。

[0062] 图2提供了当前公开的数字病理学系统中使用的各种模块的概述。在一些实施方案中,数字病理学系统采用具有一个或多个处理器203和至少一个存储器201的计算机装置200或计算机实施的方法,所述至少一个存储器201存储非暂态计算机可读指令以由所述一个或多个处理器执行从而使所述一个或多个处理器执行一个或多个模块(例如,模块202和205至209)中的指令(或存储的数据)。

[0063] 参考图2和图3,本公开文本提供了一种对具有不规则形状的生物对象的分析结果进行分析和/或将其存储在数据库或其他非暂时性存储器中的计算机实现的方法,所述生物对象包括例如成纤维细胞或巨噬细胞。该方法可以包括,例如:(a)运行图像获取模块202以生成或接收多通道图像数据,例如用一种或多种染色剂染色的生物样品的获取图像(步骤300);(b)运行图像分析模块205以从所获取的图像内的特征导出一个或多个度量(步骤310);(c)运行分割模块206以将获取的图像分割成多个子区域(步骤320);(d)运行代表性对象生成模块207以生成多边形、中心种子或标识子区域的其他对象(步骤330);(e)运行标记模块208以将导出的一个或多个度量与生成的表示对象相关联(步骤340);以及(f)将代

表性对象与相关联的度量存储在数据库209中(步骤350)。本领域技术人员还将理解,附加的模块或数据库可以被合并到工作流程中。例如,可以运行图像处理模块来将某些滤波器应用于所采集的图像,或者标识组织样品内的某些组织和/或形态结构。此外,感兴趣区域选择模块可用于选择图像的特定部分进行分析。同样,可以运行解混模块来提供对应于特定染色剂或生物标记的图像通道图像。

#### [0064] 图像获取模块

[0065] 在一些实施方案中,作为初始步骤并且参考图2,数字病理学系统200运行图像获取模块202以捕获具有一种或多种染色剂的生物样品的图像或图像数据(步骤300)。在一些实施方案中,接收或获取的图像是RGB图像或多光谱图像(例如,多路明场和/或暗场图像)。在一些实施方案中,捕获的图像被存储在存储器201中。

[0066] 图像或图像数据(在本文中可互换使用)可以使用成像设备12获取(诸如实时地)。在一些实施方案中,图像是从显微镜或能够捕获承载样本的显微镜玻片的图像数据的其他仪器获取的,如本文所指出的。在一些实施方案中,图像是使用2D扫描仪采集的,例如能够扫描图像块的扫描仪,或者能够以逐行方式扫描图像的行扫描仪,例如VENTANA DP 200扫描仪。可替代地,图像可以是先前已经获取(例如,扫描)并且存储在存储器201中(或者就此而言,经由网络20从服务器中检索到)的图像。

[0067] 生物样品可以通过应用一种或多种染色剂被染色,并且所得图像或图像数据包括对应于一种或多种染色剂中每一种的信号。这样,虽然本文描述的系统和方法可以估计或归一化为单一的染色剂,例如苏木精,但是对生物样品中染色剂的数量没有限制。实际上,除了或包括任何复染色剂,生物样品可能已经在复用测定中被两种或多种染色剂染色。

[0068] 如本领域技术人员将理解的,可以针对不同类型的细胞核和/或细胞膜生物标记来对生物样品进行染色。例如在“Sambrook等人,Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory Press(1989)”和“Ausubel等人,Current Protocols in Molecular Biology,Greenlee Publishing Associates and Wiley-Intersciences(1987)”中讨论了用于染色组织结构的方法和选择适于各种目的的染色的指南,其披露内容通过引用并入本文。

[0069] 作为一个非限制性示例,在一些实施方案中,组织样品在IHC测定中被染色,以检测一种或多种生物标记物的存在,所述生物标记物包括成纤维细胞活化蛋白(FAP)。FAP在成纤维细胞系中的过度表达被认为促进了恶性行为。已有研究表明,基质成纤维细胞是肿瘤微环境的重要组成部分,通常被称为癌相关成纤维细胞(CAF),可通过多种机制促进肿瘤的发生和发展,包括增殖、血管生成、侵袭、存活和免疫抑制。不希望受任何特定理论的束缚,认为癌细胞激活基质成纤维细胞并诱导FAP的表达,这反过来影响癌细胞的增殖、侵袭和迁移。认为FAP在90%的人类上皮癌(包括乳腺、肺、结肠直肠、卵巢、胰腺和头颈癌)的反应性基质成纤维细胞中大量表达。因此,FAP的量最有可能呈现对肿瘤临床行为的重要预测(这是一种类型的度量的例子,这种类型的度量可以被导出并随后与生成的子区域或代表性对象相关联)。

[0070] 显色染色剂可以包括苏木精、曙红、固红或3,3'-二氨基联苯胺(DAB)。当然,本领域技术人员将理解,也可以用一种或多种荧光团染色任何生物样品。在一些实施方案中,组织样品用初染剂(例如苏木精)染色。在一些实施方案中,组织样品在特定生物标记的IHC测

定中被染色。样品也可以用一种或多种荧光染料染色。

[0071] 典型的生物样品在向样品施加染色剂的自动染色/测定平台上进行加工。市场上有各种适合用作染色/测定平台的商业产品,其中一个示例是Ventana Medical Systems, Inc. (Tucson, AZ) 的Discovery™。相机平台还可以包括明视场显微镜,例如Ventana Medical Systems, Inc. 的VENTANA iScan HT或VENTANA DP 200扫描仪,或者具有一个或多个物镜和数字成像器的任何显微镜。可以使用用于捕获不同波长的图像的其他技术。适用于对染色生物样本成像的其他相机平台在本领域是已知的,并且可从诸如Zeiss、Canon、Applied Spectral Imaging等公司购得,并且这种平台易于适用于本主题公开的系统、方法和设备。

[0072] 在一些实施方案中,输入图像被掩码成使得仅组织区域存在于图像中。在一些实施方案中,生成组织区域掩码以由组织区域来掩码非组织区域。在一些实施方案中,可以通过识别组织区域并且自动或半自动地(即,以最少的用户输入)排除背景区域(例如,对应于没有样品的玻璃的全玻片图像的区域,例如仅存在来自成像源的白光的区域)来创建组织区域掩码。如本领域技术人员将理解的,除了由组织区域来掩码非组织区域之外,组织掩码模块还可以根据需要进行掩码其他兴趣区域,诸如,被识别为属于某一组织类型或属于疑似肿瘤区的组织的一部分。在一些实施方案中,使用分割技术通过在输入图像中由非组织区掩码组织区来生成组织区掩码图像。适当的分割技术是如本领域已知的这种技术(参见《数字图像处理》,第三版,Rafael C.Gonzalez,Richard E.Woods,第10章,第689页和医学成像手册,处理与分析,Isaac N.Bankman Academic Press,2000,第2章)。在一些实施方案中,利用图像分割技术在图像中的数字化组织数据与玻片之间进行区分,该组织与前景相对应并且该玻片与背景相对应。在一些实施方案中,所述部件计算全玻片图像中的兴趣区(AOI),以便检测在AOI中的所有组织区域同时限制分析的背景非组织区的数量。可以使用各种图像分割技术(例如,基于HSV彩色的图像分割、实验室图像分割、均值平移颜色图像分割、区域生长、水平设置方法、快速行进方法等)来确定例如组织数据和非组织或背景数据的边界。至少部分地基于分割,所述部件还可以生成可以用于识别与组织数据相对应的数字玻片数据的这些部分的组织前景掩码。替代性地,所述部件可以生成用于识别与组织数据不对应的数字化玻片数据的这些部分的背景掩码。

[0073] 这种识别可以通过诸如边缘检测等图像分析操作来实现。组织区域掩码可用于去除图像中的非组织背景噪声,例如非组织区域。在一些实施方案中,组织区域掩码的生成包括以下操作中的一个或多个(但不限于以下操作):计算低分辨率分析输入图像的亮度,产生亮度图像,对亮度图像应用标准偏差滤波器,产生滤波后的亮度图像,以及对滤波后的照度图像应用阈值,使得照度高于给定阈值的像素被设置为1,而低于阈值的像素被设置为0,产生组织区域掩码。在标题为“An Image Processing Method and System for Analyzing a Multi-Channel Image Obtained from a Biological Tissue Sample Being Stained by Multiple Stains (用于分析从由多个染色剂染色的生物组织样品中获得多通道图像的图像处理方法和系统)”的PCT/EP/2015/062015中披露了与生成组织区域掩码相关的附加信息和示例,其披露内容通过引用整体并入本文中。

[0074] 在一些实施方案中,感兴趣区域识别模块可以用于选择生物样品的一部分,应当获取所述部分的图像或所述部分的图像数据,例如具有高浓度成纤维细胞的感兴趣区域。

图13提供了说明根据一些实施方案的区域选择步骤的流程图。在步骤420中,区域选择模块接收识别的感兴趣区域或视场。在一些实施方案中,感兴趣区域由本公开文本的系统或通信地耦合到本公开文本的系统的另一个系统的用户识别。可替代地并且在其他实施方案中,区域选择模块从存储设备/存储器中检索感兴趣区域的位置或标识。在一些实施方案中,如步骤430所示,区域选择模块例如经由PCT/EP2015/062015中描述的方法自动生成视场(FOV)或感兴趣区域(ROI),其披露内容通过引用整体并入本文中。在一些实施方案中,感兴趣区域由系统基于图像中或图像的某些预定标准或特性自动地确定(例如,对于被多于两种染色剂染色的生物样品,识别图像的仅包括两种染色剂的区域)。在步骤440中,区域选择模块输出ROI。

#### [0075] 图像分析模块

[0076] 在一些实施方案中,从作为输入接收的图像内的特征中导出某些度量,例如FAP阳性面积、FAP阳性强度(步骤300)(见图3)。导出的度量可以与在此生成的子区域相关联(步骤320、330和340),并且度量(或其平均值、标准差等)和子区域位置可以一起存储在数据库中(步骤350),用于以后的检索和/或下游处理。本文描述的过程和算法可适用于从各种类型的细胞或细胞核获得度量和/或对其进行分类,包括从成纤维细胞和/或巨噬细胞获得度量。

[0077] 在一些实施方案中,通过检测输入图像内的细胞核和/或通过从检测到的细胞核(例如从检测到的细胞核周围的图像块)和/或从细胞膜(当然,取决于输入图像内使用的生物标记)提取特征来导出度量。在其他实施方案中,通过分析细胞膜染色、细胞质染色、和/或点状染色(例如,区分膜染色区域和非膜染色区域)来获得度量。如本文所用,术语“细胞质染色”是指以带有细胞的细胞质区域形态特征的图案排列的一组像素。如本文所用,术语“细胞膜染色”是指以带有细胞膜形态特征的图案排列的一组像素。如本文所用,术语“点状染色”是指一组具有强局部染色强度的像素,表现为分散在细胞膜区域上的斑点/点。本领域技术人员将会理解,细胞核、细胞质和细胞膜具有不同的特征,并且不同染色的组织样品可以揭示不同的生物学特征。实际上,本领域技术人员将理解,某些细胞表面受体可以具有定位于细胞膜或定位于细胞质的染色模式。因此,“细胞膜”染色模式在分析上不同于“细胞质”染色模式。同样,“细胞质”染色模式和“细胞核”染色模式在分析上是不同的。例如,基质细胞可能被FAP染色强烈,而肿瘤上皮细胞可能被EpCAM染色强烈,而细胞角蛋白可能被panCK染色。因此,通过利用不同的染色,不同的细胞类型可以在图像分析期间被区分和区分,并且可以导出不同的度量。

[0078] 美国专利号7,760,927(“’927专利”)中描述了对在具有一种或多种染色剂的生物样品的图像中的细胞核、细胞膜和细胞质进行识别和/或评分的方法,其披露内容通过引用整体并入本文中。例如,’927专利描述了一种用于同时识别用生物标记染色的生物组织的输入图像中的多个像素的自动化方法,包括考虑输入图像的前景中的多个像素的第一颜色平面,用于同时识别细胞质和细胞膜像素,其中输入图像已经被处理以去除输入图像的背景部分并去除输入图像的复染色成分;确定数字图像前景中细胞质和细胞膜像素之间的阈值水平;以及使用所确定的阈值水平,与所选择的像素及其前景的八个相邻像素同时确定所选择的像素是数字图像中的细胞质像素、细胞膜像素还是过渡像素。’927专利进一步描述了与选定像素及其八个相邻像素同时确定的步骤包括:确定选定像素与其八个相邻像素



的乘积的平方根；将乘积与确定的阈值水平进行比较；基于所述比较，递增细胞膜的第一计数器、细胞质的第二计数器或过渡像素的第三计数器；确定第一计数器、第二计数器或第三计数器是否超过预定最大值，如果是，则基于超过预定最大值的计数器对所选像素进行分类。除了对细胞核进行评分之外，'927专利还提供了对细胞质和膜进行评分的例子，例如分别基于计算的细胞质像素体积指数、细胞质像素中值强度、细胞膜像素体积和细胞膜像素中值强度。

[0079] 对细胞膜、细胞核和其他感兴趣的细胞特征进行识别和/或评分的其他方法在PCT公开号W0 2017/037180（“'180公开”）中有所描述，其披露内容通过引用整体并入本文中。'180公开文件进一步描述了对生物样品中感兴趣的分析物的细胞膜染色进行定量的方法，其中细胞膜染色区域与细胞质染色和/或点状染色混合在一起。为了实现这一点，'180公开文件描述了一种定量区域中生物区室的分析物染色的方法，在所述区域中，染色与分析上不同的不同生物区室的分析物染色混合在一起（例如（i）扩散细胞膜染色与细胞质染色混合在一起的区域，或者（ii）其中扩散细胞膜染色与点状染色混合的区域），所述方法包括以下步骤：（A）基于分析物染色模式将组织或细胞学样品的数字图像分割成多个不同的区域，所述多个区域包括至少一个复合染色区域，即在其中第一生物隔室中的分析物阳性染色与至少第二生物隔室中的分析物阳性染色混合的图像区域，其中所述第一生物区室和所述至少第二生物区室在分析上是不同的；（B）与（A）分开，识别候选生物区室，即数字图像中至少对应于第一生物区室的像素簇；（C）与（A）和（B）分开，通过将对应于分析物染色的像素簇分割成高强度区、低强度区和背景强度区来生成分析物强度图；（D）通过将化合物染色区域内的候选生物区室与来自分析物强度图的适当区相匹配，识别每个化合物染色区域的分析相关部分；以及（E）定量化合物染色区域的分析相关部分中的分析物染色。然后可以量化任何识别的区室中的像素，从而可以确定区室的面积或染色强度的量化。'180公开文件还描述了对细胞膜特异性表达水平进行评分。

[0080] 在一些实施方案中，对分类的细胞核进行评分，得到特定生物标记的阳性百分比度量或H评分度量。通过识别细胞核，可以识别相应的细胞。在其他实施方案中，通过将各自的细胞核与它们周围的被染色的细胞膜相关联来对细胞进行评分。基于细胞核周围染色细胞膜的存在，细胞可被分类，例如未染色（细胞核周围未发现染色细胞膜）、部分染色（细胞核部分被染色细胞膜包围）或完全染色（细胞核完全被染色细胞膜包围）。

[0081] 在一些实施方案中，通过首先识别候选细胞核，然后自动区分肿瘤细胞核和非肿瘤细胞核，来自动识别肿瘤细胞核。本领域中已知许多识别组织图像中候选细胞核的方法。例如，自动候选核检测可以通过应用径向对称基方法、Parvin等人的径向对称基方法来执行，如本文所述，例如在苏木精图像通道或生物标记图像通道上，所述通道使用如Ruifrok等人所述的颜色反卷积来获得，也如本文所述。在一个示例性实施方案中，如共同转让和共同未决的专利申请W0 2014140085A1中所述，使用基于径向对称的核检测操作，其全部内容通过引用整体并入本文。美国专利公开号2017/0140246中讨论了其他方法，其披露内容通过引用并入本文中。

[0082] 在候选细胞核被识别后，它们被进一步分析以将肿瘤细胞核与其他候选细胞核区分开来。其他候选核可以进一步分类（例如，通过识别淋巴细胞核和基质细胞核）。在一些实施方案中，学习的监督分类器被应用于识别肿瘤细胞核。例如，学习的监督分类器在核特征

上被训练以识别肿瘤核,然后被用于将测试图像中的候选核分类为肿瘤核或非肿瘤核。可选地,学习的监督分类器可以被进一步训练以区分不同类别的非肿瘤细胞核,例如淋巴细胞核和基质细胞核。在一些实施方案中,用于识别肿瘤细胞核的学习监督分类器是随机森林分类器。例如,可以通过以下方式训练随机森林分类器:(i) 创建肿瘤和非肿瘤细胞核的训练集,(ii) 提取每个细胞核的特征,以及(iii) 训练随机森林分类器,以基于提取的特征区分肿瘤细胞核和非肿瘤细胞核。然后,训练后的随机森林分类器可用于将测试图像中的细胞核分类为肿瘤细胞核和非肿瘤细胞核。可选地,随机森林分类器可以被进一步训练以区分不同类别的非肿瘤细胞核,例如淋巴细胞核和基质细胞核。

[0083] 在一些实施方案中,处理作为输入接收的图像处以诸如检测核中心(种子)和/或分割核。例如,可以使用本领域普通技术人员公知的技术提供指令来基于径向对称投票检测核中心(参见Parvin、Bahram等人,“Iterative voting for inference of structural saliency and characterization of subcellular events.”Image Processing,IEEE Transactions on 16.3(2007):615-623,其披露内容通过引用整体并入本文中)。在一些实施方案中,使用径向对称性检测细胞核以检测细胞核的中心,然后基于细胞中心周围的染色强度对细胞核进行分类。例如,可以在图像内计算图像大小,并且通过将所选区域内的大小相加来累积每个像素处的一个或多个投票。可以使用均值平移聚类来找出区域中的局部中心,其中,局部中心表示实际核位置。基于径向对称投票的细胞核检测是在彩色图像强度数据上执行的,并且明确地利用了细胞核是具有不同大小和偏心率的椭圆形斑点的先验域知识。为了实现这一点,与输入图像中的颜色强度一起,图像梯度信息也用于径向对称投票并且与自适应分割过程组合以精确地检测和定位细胞核。如本文使用的“梯度”是例如通过考虑围绕所述特定像素的一组像素的强度值梯度来针对特定像素计算的像素的强度梯度。每个梯度可以具有相对于由数字图像的两个正交边缘定义x和y轴的坐标系的特定“取向”。例如,细胞核种子检测涉及将种子定义为假定位于细胞核内的点,并且用作定位细胞核的起始点。第一步是使用基于径向对称的高度稳健的方法检测与每个核相关联的种子点,以检测椭圆形斑点、类似于细胞核的结构。径向对称方法使用基于内核的投票过程来对梯度图像进行操作。通过处理通过投票内核积累投票的每个像素来创建投票应答矩阵。内核基于在特定像素处计算的梯度方向以及预期的最小和最大核大小以及投票内核角的范围(通常在 $[\pi/4, \pi/8]$ 范围内)。在所得投票空间中,可以求解出具有高于预定阈值的投票值的局部极大值位置作为种子点。稍后可以在后续分割或分类过程期间丢弃外来种子。

[0084] 可以使用本领域普通技术人员已知的其他技术来识别核。例如,可以根据H&E或IHC图像之一的特定图像通道来计算图像大小,并且在指定大小周围的每个像素可以被分配有基于像素周围区域内的大小总和的多个投票。替代性地,可以执行均值偏移聚类操作以找出投票图像内的表示核的实际位置的局部中心。在其他实施方案中,可以使用核分割来基于现在已知的核中心经由形态操作和局部阈值化来分割整个核。在又其他实施方案中,可以利用基于模型的分割来检测核(即,从训练数据集中学习核的形状模型,并将其作为先验知识来分割测试图像中的核)。

[0085] 在一些实施方案中,然后随后使用针对每个核单独计算的阈值来分割核。例如,Otsu方法可以用于在经识别核周围区域中进行分割,由于认为核区域中的像素强度是变化的。如本领域普通技术人员将理解的,Otsu方法用于通过使类内方差最小化来确定最优阈

值并且对于本领域技术人员而言是已知的。更具体地，Otsu方法用于自动地执行基于聚类的图像阈值或将灰度图像减小为二进制图像。算法假定图像包含遵循双模态直方图的两类像素（前背景像素和后背景像素）。然后，计算使两个类分类的最佳阈值，使得它们的组合扩展（类内方差）最小或等效（因为成对平方距离的总和是恒定的），因此它们的类间方差是最大的。

[0086] 在一些实施方案中，所述系统和方法还包括自动分析图像中所识别的细胞核的光谱和/或形状特征，用于识别非肿瘤细胞的细胞核。例如，可以在第一步骤中在第一数字图像中识别斑点。本文使用的“斑点”可以是例如数字图像的区域，其中一些属性，例如强度或灰度值，是恒定的或者在规定的值范围内变化。在某种意义上，斑点中的所有像素可以被认为是彼此相似的。例如，可以使用基于数字图像上位置函数的导数的微分方法和基于局部极值的方法来识别斑点。核斑点是这样一种斑点，其像素和/或轮廓形状表明该斑点可能是由第一染色剂染色的核产生的。例如，可以评估斑点的径向对称性，以确定该斑点是否应该被识别为核斑点或任何其他结构，例如染色伪像。例如，在斑点具有长的形状并且不是径向对称的情况下，所述斑点可以不被识别为核斑点，而是被识别为染色伪像。取决于实施方案，被识别为“核斑点”的斑点可以代表被识别为候选核的一组像素，并且可以进一步分析这些像素以确定所述核斑点是否代表细胞核。在一些实施方案中，任何种类的核斑点被直接用作“识别的细胞核”。在一些实施方案中，对所识别的核或核斑点应用过滤操作，用于识别不属于生物标记阳性肿瘤细胞的核，并用于从已经识别的核的列表中移除所述识别的非肿瘤核，或者从开始就不将所述核添加到所识别的核的列表中。例如，可以分析所识别的核斑点的附加光谱和/或形状特征，以确定该核或核斑点是否是肿瘤细胞的核。例如，淋巴细胞的细胞核大于其他组织细胞的细胞核，例如肺细胞的细胞核。在肿瘤细胞来源于肺组织的情况下，通过识别所有最小尺寸或直径明显大于正常肺细胞核平均尺寸或直径的核斑点来鉴定淋巴细胞的细胞核。可以从一组已经识别的细胞核中去除（即，“过滤掉”）与淋巴细胞的细胞核相关的已识别的核斑点。通过过滤掉非肿瘤细胞的细胞核，可以提高该方法的准确性。取决于生物标记，非肿瘤细胞也可以在一定程度上表达生物标记，并且因此可以在第一数字图像中产生不源自肿瘤细胞的强度信号。通过从所有已经鉴定的细胞核中鉴定和过滤出不属于肿瘤细胞的细胞核，可以提高鉴定生物标记阳性肿瘤细胞的准确性。美国专利公开2017/0103521中描述了这些和其他方法，其披露内容通过引用整体并入本文中。在一些实施方案中，一旦检测到种子，可以使用局部自适应阈值方法并且创建检测到的中心周围的斑点。在一些实施方案中，还可以结合其他方法，诸如，还可以使用基于标记的流域算法来识别检测到的核中心周围的核斑点。在被公开为W02016/120442的共同未决的申请PCT/EP2016/051906中描述的这些或其他方法，其披露内容通过引用整体并入本文中。

[0087] 该系统可以使用至少一个图像特征度量和至少一个形态学度量来确定图像内的特征是否对应于感兴趣的结构（统称为“特征度量”）。图像特性度量（源自图像内的特征）可以包括例如颜色、色平衡、强度等。形态学度量（从图像内的特征导出）可以包括例如特征大小、特征颜色、特征方向、特征形状、特征之间的关系或距离（例如，相邻特征）、特征相对于另一解剖结构的关系或距离等。图像特征度量、形态度量和其他度量可用于训练如本文所述的分类器。从图像特征导出的度量的具体示例如下：

[0088] (A) 从形态特征导出的度量

[0089] 如本文中使用的“形态特征”是例如指示核的形状或尺寸的特征。在不希望受任何特定理论的束缚的情况下,认为形态特征提供关于细胞或其核的大小和形状的一些重要信息。例如,可以通过对包含在核斑点或种子中或周围的像素应用各种图像分析算术来计算形态特征。在一些实施方案中,形态特征包括面积、短轴和长轴长度、周长、半径、固态性等。

[0090] (B) 从表观特征导出度量

[0091] 例如,如本文使用的“表观特征”是针对特定核通过对包含在用于识别核的核斑点或种子中或其周围的像素的像素强度值进行比较来计算的特征,由此所比较的像素密度是从不同图像通道中导出的(例如,背景通道、用于对生物标记进行染色的通道等)。在一些实施方案中,可以根据像素强度的百分数值(例如,第10位、第50位和第95位百分数值)以及根据不同图像通道计算的梯度大小的百分数值来计算从表观特征导出的度量。例如,首先,识别表示兴趣核的核斑点内的多个图像通道IC(例如,三个通道:HTX、DAB、亮度)中的每一个的像素值的多个(P个) $\times$ 百分位值( $X=10, 50, 95$ )。计算外观特征度量可能是有利的,因为导出的度量可以描述核区域的属性以及描述核周围的膜区域。

[0092] (C) 从背景特征导出的度量

[0093] 例如,“背景特征”是指示细胞质中的表观和/或染色存在的特征,以及包括从图像中提取背景特征的核的细胞的细胞膜特征。可以例如通过以下各种操作来计算数字图像中描绘的核和相应细胞的背景特征和相应度量:识别表示核的核斑点或种子;分析与所计算的经识别细胞组直接相邻的像素面积(例如,在核边界周围的20个像素(大约9微米)厚的带状结构),因此,在具有这种核的细胞的细胞质和膜以及与该细胞直接相邻的面积中捕获表观和染色存在。这些度量类似于核表观特征,但是在每个核边界周围大约20像素(大约9微米)厚的带状结构中计算,因此捕获具有经识别核的细胞的细胞质和细胞膜以及直接与该细胞相邻的面积中捕获表观和染色存在。在不希望受任何特定理论的束缚的情况下,选择带状结构大小,因为认为其捕获在核周围的可以用于提供用于核区分的有用信息的足够量的背景组织区。这些特征与由“J.Kong等人,“A comprehensive framework for classification of nuclei in digital microscopy imaging:An application to diffuse gliomas,”ISBI,2011,第2128-2131页”披露的特征类似,其披露内容通过引用整体并入本文中。认为可以使用这些特征来判定周围组织是基质还是上皮细胞(诸如,在H&E染色组织样品中)。在不希望受任何特定理论的束缚的情况下,认为这些背景特征还捕获膜染色模式,该膜染色模式在使用适当的膜染色剂染色组织样品时是有用的。

[0094] (D) 从颜色导出的度量。

[0095] 在一些实施方案中,从颜色导出的度量包括颜色比率( $R/(R+G+B)$ )或颜色主要成分。在其他实施方案中,从颜色导出的度量包括对颜色中每种颜色的局部统计(均值/中值/方差/标准偏差)和/或局部图像窗口中的颜色强度相关性。

[0096] (E) 从强度特征导出的度量

[0097] 将具有某些特定属性值的该组相邻细胞设置在以组织病理学玻片图像表示的灰色细胞的黑色阴影与白色阴影之间。颜色特征的相关性定义了大小类的实例,因此这种颜色细胞的强度从其周围的暗细胞簇中确定受影响的细胞。在PCT公开号W0/2016/075095中描述了纹理特征的示例,其披露内容通过引用整体并入本文中。

[0098] (六) 空间特征

[0099] 在一些实施方案中,空间特征包括:细胞的局部密度;两个相邻检测细胞之间的平均距离;和/或从细胞到分割区域的距离。

[0100] (G)从核特征导出的度量

[0101] 技术人员还将理解的是,度量也可以从核特征中导出。Xing等人,“Robust Nucleus/Cell Detection and Segmentation in Digital Pathology and Microscopy Images:A Comprehensive Review,”IEEE Rev Biomed Eng 9,234-263,2016年1月,描述了这种核特征的计算,其披露内容通过引用整体并入本文中。当然,可以考虑本领域普通技术人员已知的其他特征并将其用作特征计算的基础。

[0102] 在导出特征度量之后,该特征可以单独使用或者与训练数据结合使用(例如,在训练期间,根据本领域普通技术人员已知的程序,示例细胞与专家观察者提供的基本事实识别一起被呈现)来对细胞核或细胞分类。在一些实施方案中,所述系统可以包括分类器,该分类器至少部分地基于每个生物标记的一组训练或参考玻片而被训练。本领域技术人员将理解,不同组的玻片可用于训练每种生物标记的分类器。因此,对于单个生物标记,在训练后获得单个分类器。本领域技术人员还将意识到,由于从不同生物标记获得的图像数据之间存在差异性,可以为每个不同的生物标记训练不同的分类器,以确保对未知测试数据的更好性能,其中测试数据的生物标记类型将是已知的。可以至少部分地基于如何最好地处理训练数据差异性来选择训练的分类器,例如,在组织类型、染色方案和其他感兴趣的特征方面,用于玻片解释。

[0103] 在一些实施方案中,分类模块是支持向量机(“SVM”)。一般而言,SVM是一种分类技术,其基于非线性输入数据集经由用于非线性情况的内核被转换为高维线性特征空间的统计学习理论。不希望受任何特定理论的约束,认为支持向量机通过核函数K将表示两个不同类的一组训练数据E投影到高维空间中。在这个变换的数据空间中,非线性数据被变换,使得可以生成一条平坦的线(鉴别超平面)来分离类,以便最大化类分离。然后经由K将测试数据投射到高维空间中,并且在该测试数据相对于超平面下降位置的基础上对测试数据进行分类。内核函数K定义了将数据投影到高维空间的方法。

[0104] 在其他实施方案中,使用AdaBoost算法执行分类。AdaBoost是一种结合多个弱分类器来生成强分类器自适应算法。由病理学家在训练阶段识别的图像像素(例如具有特定染色或属于特定组织类型的图像像素)用于为每个单独的纹理特征 $\Phi_j$ 生成概率密度函数,对于 $j \in \{1, \dots, K\}$ 被认为是弱分类器。贝叶斯定理然后用于为构成弱学习者的每个 $\Phi_j$ 生成可能场景 $L_j = (C_j, l_j \in \{1, \dots, K\})$ 。这些由AdaBoost算法组合成强分类器 $\Pi_j = \sum T_i = 1 \alpha_{ji} l_{ji}$ ,其中对于每个像素 $c_j \in C_j$ , $\Pi_j(c_j)$ 是像素 $c_j$ 属于类 $\omega_T$ 的组合可能性,其中 $\alpha_{ji}$ 是在特征 $\Phi_i$ 的训练期间确定的权重,并且T是迭代次数。

[0105] 在一些实施方案中,可使用导出的染色强度值、特定细胞核的计数或其他分类结果来确定各种标记表达得分(在本文中可与术语“表达得分”互换使用),例如阳性百分比或H得分(即,根据分类特征,可计算表达得分)。在2013年12月19日提交的共同转让和共同未决的申请W0/2014/102130A1“Image analysis for breast cancer prognosis”和2014年3月12日提交的申请W0/2014/140085A1“Tissue object-based machine learning system for automated scoring of digital whole slides”中进一步详细描述了评分方法,每一个的内容通过引用整体并入本文中。例如,至少部分基于生物标记阳性肿瘤细胞/生物标记

阳性非肿瘤细胞的数量,可以确定得分(例如,全玻片得分)。在一些实施方案中,对于每个检测到的细胞核斑点,可以计算平均斑点强度、颜色和几何特征,例如检测到的细胞核斑点的面积和形状,并且细胞核斑点被分类为肿瘤细胞核和非肿瘤细胞的细胞核。鉴定的细胞核输出的数量对应于在FOV中检测到的生物标记阳性肿瘤细胞的总数,由计数的肿瘤细胞核的数量证明。

[0106] 在一些实施方案中,同样在用FAP染色的情况下,导出特征度量并训练分类器,使得可以阐明FAP阳性或阴性细胞的百分比(例如阳性表达百分比得分),例如阳性或阴性染色的基质细胞。在一些实施方案中,得分0可被分配给肿瘤细胞 $\leq 10\%$ 的染色区域,1分配给肿瘤细胞 $>11\%$ 至 $\leq 25\%$ 的区域,2分配给肿瘤细胞 $>26\%$ 至 $\leq 50\%$ 的区域,3分配给肿瘤细胞 $>51\%$ 的区域。对于染色强度,缺失/弱染色(阴性对照)得0分,明显强于阴性对照水平的弱染色得1分,中等强度染色得2分,强染色得3分。在一些实施方案中, $\geq 3$ 的最终得分可被识别为指示FAP的阳性表达。

#### [0107] 分割模块

[0108] 中分辨率分析方法采用分割算法来生成输入图像内的子区域,这些子区域被定义为捕获有生物学意义的感兴趣区域。在用图像分析模块205从输入图像导出度量之后(步骤310),利用分割生成模块206将输入图像分割成多个子区域(步骤320)。

[0109] 在一些实施方案中,对单通道图像执行分割,例如解混合的FAP图像中的“紫色”通道。解混的方法是本领域普通技术人员已知的(例如,线性解混描述于例如‘Zimmermann “Spectral Imaging and Linear Unmixing in Light Microscopy” Adv Biochem Engin/ Biotechnol (2005) 95:245-265’以及C.L.Lawson和R.J.Hanson,“Solving least squares Problems”,Prentice Hall,1974,第23章,第161页’,所述文献的披露内容通过引用整体并入本文)。本文公开了其他的解混方法。另见Ruifok等人,“Quantification of histochemical staining by color deconvolution,”Anal Quant Cytol Histol.2001年8月;23(4):291-9,其披露内容通过引用整体并入本文中。

[0110] 在一些实施方案中,子区域在输入图像的区域中生成捕获信息,该区域具有预定的尺寸或者在图像处理算法(例如,如本文所述的SLIC超像素生成算法的参数)中设定的范围内的尺寸。

[0111] 在一些实施方案中,输入图像被分割成具有预定形状、大小、面积和/或间距的子区域。例如,子区域(710)可以是椭圆形、圆形、正方形、矩形等,如图7所示。在一些实施方案中,椭圆形、圆形、正方形或矩形子区域可以具有从50个像素到大约100个像素的大小,或者一些其他大小,使得选择具有相似属性或特性(例如颜色、亮度和/或纹理)的像素组。在一些实施方案中,子区域是非重叠的,并且可以通过采样网格生成。如本文所使用的,术语“采样网格”涉及水平和垂直线的网络,这些线均匀间隔并叠加在图像上,最终用于定位图像内不重叠的点。在一些实施方案中,由水平线和垂直线建立的任意数量的相邻位置可用于定义图像分割。在一些实施方案中,子区域以捕获相关区域的代表性样品用于分析的方式分布在图像上,例如不规则形状的区域是主要特征的区域。

[0112] 在其他实施方案中,通过对图像应用一系列算法来分割输入图像,包括全局阈值滤波器、局部自适应阈值滤波器、形态学运算和分水岭变换。滤波器可以顺序运行或者以本领域普通技术人员认为必要的任何顺序运行。当然,任何滤波器都可以反复应用,直到达到

期望的结果。在一些实施方案中,将第一滤波器应用于输入图像,以去除不太可能具有细胞核的区域,例如去除那些白色的图像区域(对应于组织样品中未染色或几乎未染色的区域)。在一些实施方案中,这是通过应用全局阈值滤波器来实现的。在一些实施方案中,全局阈值化基于在第一主成分通道上计算的中值和/或标准偏差,例如类似于灰度通道。通过获得全局阈值,认为可以丢弃任何白色图像区域,其代表细胞核可能不存在的未染色或接近未染色的区域。然后对图像应用滤波器,以选择性地去除伪影,例如小斑点、小间断、其他小物体和/或填充孔洞。在一些实施方案中,形态算子被应用于去除伪影和/或填充孔洞。在一些实施方案中,基于作为输入引入的二值图像(例如,由先前的滤波步骤产生的二值图像),应用基于距离的分水岭。

[0113] 在一些实施方案中,输入图像被分割成超像素。认为超像素算法将图像分割成代表具有感知意义的实体的多个片段(像素组)。每个超像素通过低级分组过程获得,并且具有感知上一致的单位,即,包含在超像素中的生物对象中的所有像素在染色存在(例如,存在于超像素中的像素是特定类型的染色)、染色强度(例如,像素具有特定的相对强度值或值范围)和纹理(例如,像素具有颜色或强度的特定空间排列)方面尽可能一致。可以存储和报告每个超像素的局部分析结果,以在数字病理图像上表示分析结果。

[0114] 超像素是具有相似特性(如颜色、亮度和纹理)的像素的集合。图像可以由一定数量的超像素组成,这些超像素包含像素的多个组合特征,并且可以保留原始图像的边信息。与单个像素相比,超像素包含丰富的特征信息,可以大大降低图像后处理的复杂度,显著提高图像分割的速度。超像素对于估计概率和用小邻域模型做决策也很有用。

[0115] 超像素算法是将像素分成大小相似的有意义的原子区域的方法。不希望被任何特定的理论所束缚,相信超像素是强大的,因为它们经常落在图像内的重要边界上,并且当它们包含显著的对象特征时倾向于呈现异常或独特的形状。与以中等分辨率分析来获取和存储信息的愿望相一致,超像素位于像素和对象级别之间:它们通过表示具有感知意义的像素组来携带比像素更多的信息,而不是全面地表示图像对象。超像素可以理解为图像分割的一种形式,在短的计算时间内对图像进行过分割。超像素的轮廓已经显示出很好地粘附到自然图像边界,因为图像中的大多数结构都被保留。通过为每个超像素而不是每个像素计算图像特征,随后的处理任务在复杂度和计算时间上得以降低。因此,超像素被认为是有用的预处理步骤,用于对象级分析,如图像分割。

[0116] 不希望受任何特定理论的约束,认为超像素通过形成在例如颜色或几何形状上具有相似特征的紧凑且均匀的像素组来过度分割图像。过去,已经开发了多种超像素方法。它们可以分为(一)基于图形的方法和(二)基于梯度上升的方法。在基于图的方法中,每个像素被认为是图中的一个节点。在所有节点对之间定义一个与其相似性成比例的边权重。然后,在图上定义的成本函数被公式化并最小化,以便提取超像素片段。在基于梯度上升的方法中,像素被迭代地映射到特征空间,以描绘代表簇的更密集的区域。每次迭代细化每个簇以获得更好的分割,直到收敛。

[0117] 已经开发了许多超像素算法,包括归一化切割、凝聚聚类、快速移位和Turbopixel算法。归一化切割算法使用轮廓和纹理线索递归地分割图像中所有像素的图形,全局最小化在分割边界的边缘上定义的成本函数。它产生非常规则的、美观的超像素(参见Jianbo Shi和Jitendra Malik.Normalized cuts and image segmentation,IEEE Transactions

on Pattern Analysis and Machine Intelligence, (PAMI), 22 (8):888-905, 2000年8月, 其披露内容通过引用整体并入本文中)。Alastair Moore、Simon Prince、Jonathan Warrell、Umar Mohammed和Graham Jones在Superpixel Lattices. IEEE Computer Vision and Pattern Recognition (CVPR), 2008中描述了一种方法, 通过寻找最佳路径或接缝, 将图像分割成更小的竖直或水平区域, 生成符合网格的超像素。最佳路径是使用图切割方法找到的(参见Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. ACM Transactions on Graphics (SIGGRAPH), 26 (3), 2007, 其披露内容通过引用并入本文中)。快速移位(见维A. Vedaldi和S. Soatto. Quick shift and kernel methods for mode seeking. In European Conference on Computer Vision (ECCV), 2008, 其披露内容通过引用并入本文中)使用模式搜索分割方案。它使用中心点移位 (medoid shift) 过程初始化分割。然后, 它将特征空间中的每个点移动到最近的邻居, 这增加了Parzen密度估计。涡轮像素方法使用基于水平集的几何流逐步扩展一组种子位置(见A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi). Turbopixels: Fast superpixels using geometric flows. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2009, 其披露内容通过引用并入本文中)。几何流依赖于局部图像梯度, 目的是在图像平面上有规律地分布超像素。与其他方法不同, Turbopixel超像素被限制为具有一致的大小、紧密度和边界粘附性。Radhakrishna Achanta在“SLIC Superpixels Compared to State-of-the-art”, Journal of Latex Class Files, Vol. 6, No. 1, 2011年12月中描述了生成超像素的其他方法, 其披露内容通过引用整体并入本文中。

[0118] 引入了一种称为简单线性迭代聚类 (SLIC) 的超像素算法, 与目前最先进的超像素方法相比, 它在边界粘附性和效率方面都是优越的。SLIC具有两个步骤。首先, 它通过用局部k-均值聚类 (KMC) 方法对像素进行分组来生成超像素, 其中距离被测量为与数据和空间距离相结合的欧几里德距离。其次, 使用连通分量算法 (CCA) 通过将生成的小孤立区域合并到最近的大超像素中来去除它们。

[0119] K均值聚类旨在将n个观测值划分为k个簇, 其中每个观测值属于具有最近均值的簇, 作为该簇的原型。连通分量标记通过逐像素扫描图像(从上到下, 从左到右)来识别连通像素区域, 即共享同一组强度值V的相邻像素区域。(对于二值图像 $V = \{1\}$ ; 然而, 在灰度图像中, V将呈现一系列值, 例如:  $V = \{51, 52, 53, \dots, 77, 78, 79, 80\}$ 。)连通分量标记适用于二值或灰度图像, 不同的连通度量是可能的。然而, 下面我们假设二进制输入图像和8-连通性。连通分量标记算子通过沿一行移动来扫描图像, 直到到达点p(其中p表示在扫描过程的任何阶段要标记的像素), 对于该点,  $V = \{1\}$ 。当这为真时, 它检查已经在扫描中遇到的p的四个相邻像素(即, (i) p左边的相邻像素, (ii) p上面的相邻像素, 以及(iii)和(iv)两个上对角线项)。基于该信息, p的标记如下: 如果所有四个相邻像素都为0, 则为p分配一个新的标记, 否则如果只有一个相邻像素的值为 $V = \{1\}$ , 则将其标记分配给p, 否则, 如果有多个相邻像素的值为 $V = \{1\}$ , 则为p分配一个标记, 并记下等价项。

[0120] 完成扫描后, 等效标签对被分类到等效类别中, 并为每个类别分配一个唯一的标签。作为最后一步, 对图像进行第二次扫描, 在此期间, 每个标签被分配给其等价类的标签替换。对于显示, 标签可以是不同的灰度或颜色。

[0121] SLIC是超像素生成的k-均值的一种改编, 具有两个重要的区别: (i) 通过将搜索空



间限制在与超像素大小成比例的区域,优化中的距离计算的数量显著减少(这被认为减少了像素数量的线性复杂度,并且与超像素数量 $k$ 无关);以及(ii)加权距离度量结合了颜色和空间邻近性,同时提供对超像素的大小和紧凑性的控制。(参见Achanta等人,“SLIC Superpixels Compared to State-of-the-Art Superpixel Methods”,IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.34, No.11, 2012年11月,其公开内容在此全文引入作为参考)。

[0122] SLIC考虑5D空间中的图像像素,由CIELAB颜色空间的 $L*a*b$ 值以及它们的 $x$ 和 $y$ 坐标定义。5D空间中的像素基于整合了图像平面中的颜色相似性和邻近性的自适应 $k$ 均值聚类来聚类。聚类是基于距离度量 $D$ 的,距离度量 $D$ 度量 $L*a*b$ 空间中的颜色相似性( $d_c$ )和 $x, y$ 空间中的像素接近度( $d_s$ )。后者通过定义图像像素总数的平方根除以超像素数( $k$ )的网格间隔( $S$ )来归一化。超像素的紧密度和规律性由常数 $m$ 控制。该参数用作空间距离( $d_c$ )和谱距离( $d_s$ )之间的加权标准。较大的 $m$ 增加了空间邻近度的权重,这导致更紧凑的超像素,其边界与图像中的谱轮廓粘附较少。

$$[0123] \quad D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \quad (\text{等式1})$$

[0124] SLIC算法可以应用如下。设 $N_p$ 为给定图像(或其感兴趣的部分或区域)中的像素数, $k$ 为要生成的超像素数。接下来,SLIC算法的主要步骤如下:

[0125] (1) 初始化簇中心。将 $k$ 个初始簇中心设置在间隔开 $S = \sqrt{N_p/k}$ 像素的规则网格上,然后将这些簇中心移动到 $3 \times 3$ 邻域中梯度最低的位置。不希望受任何特定理论的约束,认为这样做是为了避免将超像素居中在边缘上,并减少用噪声像素播种超像素的机会。

[0126] (2) 分配像素。通过局部KMC将每个像素指定到局部搜索空间中最近的簇中心。

[0127] (3) 更新簇中心。将每个簇中心设置为相应簇中所有像素的均值。

[0128] (4) 重复步骤(2)-(3),直到簇不变或满足另一个给定标准。

[0129] (5) 后处理。如果孤立区域的大小小于最小大小 $S_{\min}$ ,则使用CCA将孤立区域重新分配给附近的超像素。

[0130] 在SLIC方法的步骤(2)中应用局部KMC,其中每个像素与搜索区域覆盖其位置的最近簇中心相关联。在传统的KMC中,每个簇中心的搜索区域是整个图像,然后计算从每个簇中心到图像中每个像素的距离。然而,在局部KMC中,簇中心的搜索空间被限制在局部 $2S \times 2S$ 正方形区域。因此,SLIC仅计算从每个簇中心到其搜索区域内的像素的距离。

[0131] 在局部KMC,欧几里德距离用于聚类。设 $z_i$ 为第 $i$ 个簇中心的数据,其空间位置为 $(x_i, y_i)$ 。设 $z_j$ 为中心搜索区域内像素的强度。然后,该像素和中心之间的积分距离为:

$$[0132] \quad D_I = \sqrt{\left(d_f/m\right)^2 + (d_s/S)^2} \quad (\text{等式2})$$

[0133] 其中 $d_f = |z_i - z_j|$ 和 $d_s = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ 分别是像素和中心之间的强度和空间距离, $m$ 是对 $d_f$ 和 $d_s$ 对积分距离 $D_I$ 的相对贡献进行加权的正则化参数。较大的 $m$ 表示 $d_s$ 比 $d_f$ 更重要。直接描述两个距离的贡献的等效积分距离 $D_I$ 可以由下式给出:

$$[0134] \quad D_I = \sqrt{w \left( d_f / N_f \right)^2 + (1 - w) (d_s / S)^2} \quad (\text{等式3})$$

[0135] 其中 $N_f$ 是整个图像的平均强度, $w \in [0, 1]$ 是正则化参数。在这种情况下, $w$ 和 $(1-w)$ 分别是归一化强度和空间距离 $D_I$ 的比值。

[0136] 在一些实施方案中,SLIC算法的参数 $k$ 指定近似相等大小的超像素的数量。在一些实施方案中,紧密度参数 $m$ 可以被设置为控制超像素的同质性和边界粘附性之间的折衷。不希望受任何特定理论的约束,认为通过改变紧密度参数,规则形状的超像素可以在未纹理化区域中生成,并且高度不规则的超像素可以在纹理化区域中生成。同样,不希望被任何特定的理论所束缚,相信参数 $m$ 也允许颜色相似性和空间接近性之间的相对重要性的加权。当 $m$ 较大时,空间邻近性更重要,并且产生的超像素更紧密(即,它们具有较低的面积与周长比)。当 $m$ 较小时,产生的超像素更紧密地附着在图像边界上,但是具有不太规则的尺寸和形状。

[0137] 在一些实施方案中,调整超像素大小和紧密度参数两者。在一些实施方案中,使用范围在大约40像素到大约400像素之间的超像素大小。在其他实施方案中,使用范围在大约60像素到大约300像素之间的超像素大小。在另外的实施方案中,使用范围在大约70像素到大约250像素之间的超像素大小。在进一步的实施方案中,使用范围在大约80像素到大约200像素之间的超像素大小。

[0138] 在一些实施方案中,紧密度参数的范围在大约10到大约100。在其他实施方案中,紧密度参数的范围在大约20到大约90。在其他实施方案中,紧密度参数的范围在大约40到大约80。在其他实施方案中,紧密度参数的范围在大约50到大约80。

[0139] 图8A示出了使用本文提到的SLIC生成的超像素的例子,其中超像素被分割成适合感兴趣区域的局部特性而不重叠,并且在它们之间没有间隙。此外,每个超像素子区域根据其局部强度(810)和生物标记表达的存在方向(820)具有特定的最终形状。因此,超像素对于这种感兴趣的生物结构具有感知意义。图8B、图8C和图8D分别示出了高放大率下的原始IHC图像、超像素生成过程的初始化以及具有局部均匀性的最终超像素,并且其中它们的形状的规律性已经由SLIC算法的技术参数进行了调整,如上所述。

#### [0140] 代表性对象生成模块

[0141] 在用子区域生成模块生成子区域之后(步骤320),使用模块207为每个子区域确定代表性对象或兴趣点(步骤330)。在一些实施方案中,代表性对象是属于感兴趣的细胞或细胞组(例如成纤维细胞或巨噬细胞)的子区域或超像素的轮廓。在其他实施方案中,代表性对象是种子点。如本文所述,本公开的目的是基于具有相似染色存在、染色强度和/或局部纹理的子区域来表征感兴趣的细胞(例如,不规则形状的细胞),并在数据库中自动保存那些同质属性子区域。代表性对象或其坐标是存储生成的子区域的一种方法。图9A和图9B提供了包含感兴趣的生物对象的那些超像素的多边形轮廓和中心种子的示例。

[0142] 在一些实施方案中,利用算法来创建边界,该边界将具有不同颜色或纹理的子区域分开,并且与图像中的主要边缘对齐,从而生成代表感兴趣的生物对象(例如,不规则大小或形状的细胞,例如成纤维细胞或巨噬细胞)的边界。在一些实施方案中,阈值算法(例如,Outsu、均值聚类等)可以被应用于染色剂通道图像,使得不具有染色剂的子区域被排除,并且只有那些包括阈值量染色剂的子区域被提供作为代表性对象。在一些实施方案中,

可以使用阈值参数(例如,由专家病理学家提供的阈值染色参数)来生成子区域的二元掩码。在一些实施方案中,通过应用一系列被设计为增强图像的滤波器来实现分割,使得(i)不太可能表示感兴趣对象的子区域与(ii)表示具有感兴趣对象的细胞的子区域分离。可以选择性地应用额外的滤波器来去除伪影、去除小斑点、去除小间断、填充孔洞以及分割更大的斑点。

[0143] 在一些实施方案中,不太可能具有识别不规则形状细胞的子区域的区域被移除,例如通过移除染色通道的二值图像中的白色图像区域(对应于组织样品中未染色或几乎未染色的区域)。在一些实施方案中,这是通过应用全局阈值滤波器来实现的。阈值处理是一种用于将强度图像(I)转换成二值图像(I')的方法,如果各像素的强度高于或低于某个阈值(这里是全局阈值),则给各像素赋值1或0。换句话说,全局阈值根据像素的强度值来划分像素。在一些实施方案中,全局阈值化基于在第一主成分通道上计算的中值和/或标准偏差,例如类似于灰度通道。通过获得全局阈值,认为可以丢弃任何白色图像区域,其代表不规则形状细胞可能不存在的未染色或接近未染色的区域。

[0144] 在一些实施方案中,就FAP染色剂而言,边界可以通过以下方式创建:1)解混紫色通道,2)对紫色通道进行阈值处理以识别FAP阳性面积,3)在紫色通道上应用超像素分割,以及4)将特征度量附加到超像素对象。在一些实施方案中,可以使用监督生成规则来识别FAP阳性面积的存在,该规则是基于从病理学家获得的基础事实来训练的。在一些实施方案中,FAP阳性阈值参数可以由病理学家提供,例如通过识别训练图像集上的阈值。然后可以使用阈值参数生成二进制掩码。这些方法在Auranuch Lorsakul等人“Automated whole-slide analysis of multiplex-brightfield IHC images for cancer cells and carcinoma-associated fibroblasts,”Proc.SPIE 10140,Medical Imaging 2017: Digital Pathology,1014007(2017/03/01)中有进一步描述,其披露内容通过引用整体并入本文中。

[0145] 在一些实施方案中,追踪子区域的边界。例如,可以提供一种算法,该算法跟踪子区域的外部边界,以及子区域内部或之间的那些“孔”的边界。在一些实施方案中,子区域的边界是通过使用称为bwboundaries的matlab函数创建边界轨迹来生成的(<https://www.mathworks.com/help/images/ref/bwboundaries.html>)。

[0146] 边界创建后,边界轨迹被转换成x,y坐标的多边形轮廓。被跟踪边界的x,y坐标可以存储在存储器或数据库中,例如,可以确定和存储子区域对象的被跟踪边界的所有像素的行和列坐标。

[0147] 在一些实施方案中,通过运算或计算每个子区域的形心或质心来导出种子点。本领域普通技术人员已知确定不规则物体形心的方法。一旦计算,子区域的形心被标记和/或种子的x,y坐标被存储在存储器或数据库中。在一些实施方案中,形心或质心的位置可以叠加在输入图像上。

#### [0148] 标记模块

[0149] 在使用分割模块206生成子区域并且使用模块207计算代表性对象之后,使用标记模块208对代表性对象进行标注、标记或与例如从图像分析模块202导出的度量(步骤310)的数据相关联(步骤330)。标记模块208可以创建数据库209,该数据库是存储本文提到的数据的非暂时性存储器。在一些实施方案中,数据库209存储作为输入接收的图像、任何多边

形和/或种子点的坐标以及来自图像分析的任何相关数据或标签(见图11)。

[0150] 在这点上,可以为图像的每个分割的子区域存储数据向量。例如,可以为每个子区域存储数据向量,包括任何代表性对象的坐标和相关联的图像分析数据。举例来说,如果数据点“a”、“b”和“c”是代表对象的坐标,并且“x”、“y”和“z”是从图像分析中导出的度量(或对应于特定子区域的度量的平均值),则数据库将存储数据的以下向量 $[a,b,c,x,y,z]^1$ 、 $[a,b,c,x,y,z]^2$ 、 $[a,b,c,x,y,z]^N$ ,其中N是通过分割模块206生成的子区域的数量。

[0151] 在一些实施方案中,来自图像分析模块的数据描述了图像中的各个像素。技术人员将理解,特定子区域内的所有像素的数据可以被平均,以提供子区域内的像素数据的平均值。例如,各个像素可以各自具有特定的强度。可以对特定子区域中所有像素的强度进行平均,以提供该子区域的平均像素强度。该子区域的平均像素可以与该子区域的代表性对象相关联,并且数据可以一起存储在存储器中。

[0152] 在用FAP染色的情况下,FAP阳性面积可以是附加到超像素对象的另一个特征/测量值。FAP阳性面积是指FAP强度高于设定阈值的像素的总和。Auranuch Lorsakul等人在“Automated whole-slide analysis of multiplex-brightfield IHC images for cancer cells and carcinoma-associated fibroblasts,”Proc.SPIE 10140,Medical Imaging 2017:Digital Pathology,1014007 (2017/03/01)中描述了阈值的选择,其披露内容通过引用整体并入本文中。

[0153] 作为由标记模块存储的数据的例子,并且在使用FAP生物标记对生物样品染色的情况下,可以通过对特定子区域的图像分析来导出子区域内FAP染色的平均强度,并且可以将FAP染色强度与该子区域的任何代表性对象的坐标一起存储在数据库中。同样,可以使用图像分析来导出子区域的特定表达得分,例如FAP表达得分,且该子区域的FAP表达得分可以与该特定子区域的表示对象一起存储。除了任何子区域内的图像部分的平均强度得分和平均表达得分之外,可以存储其他参数,包括但不限于种子点之间的距离、所识别的肿瘤细胞和不规则形状细胞之间的距离(例如,肿瘤细胞和成纤维细胞之间的距离)以及FAP阳性面积。

[0154] 在一些实施方案中,作为示例,在相应的超像素内计算的分析结果,例如平均局部强度、阳性染色面积,被附加到它们相应的多边形轮廓和种子。对于全玻片图像,这些表示对象(例如,多边形轮廓和种子)及其分析结果随后以x-y坐标存储在数据库中。图10A示出了将成纤维细胞(1010)的成纤维细胞活化蛋白(FAP)染成紫色和将上皮肿瘤(1020)的泛细胞角蛋白(PanCK)染成黄色的头颈癌组织的全玻片IHC图像的例子。图10B和图11示出了多边形轮廓和种子的例子,它们附有属于成纤维细胞区域的超像素的分析结果,这些结果可以分别存储在数据库中。

#### [0155] 数据检索或投影模块

[0156] 本领域技术人员将会理解,存储的分析结果和相关的生物特征可以稍后检索,并且数据可以以各种格式报告或可视化,例如分析结果的直方图。更具体地,可以从数据库209中检索表示对象坐标数据和相关联的图像分析数据,并用于进一步分析。在一些实施方案中,作为示例,可以从数据库中检索表示对象,用于在全玻片图像中或在用户标注的区域中可视化或报告分析结果。如图12所示,相关或附加的图像分析结果可以通过绘制从整个玻片超像素中检索的FAP强度的直方图来报告。可选地,数据可以在全玻片图像、视场图像

或由医学专业人员标注的图像的一部分上可视化,以供进一步审查。

[0157] 用于实践本公开文本的实施方案的其他部件

[0158] 本公开文本的计算机系统200可以绑定到可以对组织样本执行一个或多个制备过程的样本处理设备。制备过程可以包括但不限于对样本进行脱蜡、对样本进行调节(例如,细胞调节)、对样本进行染色、执行抗原修复、执行免疫组织化学染色(包括标记)或其他反应和/或执行原位杂交(例如,SISH、FISH等)染色(包括标记)或其他反应、以及用于制备用于显微术、微量分析、质谱法或其他分析方法的样本的其他过程。

[0159] 处理设备可以将固定剂应用于样本。固定剂可以包括交联剂(诸如醛类(例如甲醛、多聚甲醛和戊二醛)以及非醛类交联剂)、氧化剂(例如,金属离子和复合物,如四氧化铬和铬酸)、蛋白质变性剂(例如,乙酸、甲醇和乙醇)、未知机制的固定剂(例如,氯化汞、丙酮和苦味酸)、组合试剂(例如,卡诺氏固定剂(Carnoy's fixative)、methacarn、波恩氏流体(Bouin's fluid)、B5固定剂、罗斯曼氏流体(Rossmann's fluid)、詹德莱氏流体(Gendreau's fluid))、微波和混杂固定剂(例如,排出体积固定和蒸气固定)。

[0160] 如果样本是嵌入石蜡中的样品,则可以使用(多种)适当的去石蜡流体对样品进行脱石蜡。除去石蜡后,任何数量的物质都可以连续施加到样本上。物质可以用于预处理(例如,用于反转蛋白质交联、暴露核酸等)、变性、杂交、洗涤(例如,严格洗涤)、检测(例如,将视觉或标记分子与探针链接)、扩增(例如,扩增蛋白质、基因等)、复染、盖玻等。

[0161] 样本处理设备可以向样本施加各种物质。物质包括但不限于染色剂、探针、试剂、冲洗剂和/或调节剂。物质可以是流体(例如,气体、液体或气体/液体混合物)等。流体可以是溶剂(例如,极性溶剂、非极性溶剂等)、溶液(例如,水溶液或其他类型的溶液)等。试剂可以包括但不限于染色剂、润湿剂、抗体(例如,单克隆抗体、多克隆抗体等)、抗原回收流体(例如,基于水性或非水性的抗原修复溶液、抗原回收缓冲液等)等。探针可以是与可检测的标记或报告分子附接的分离的核酸或分离的合成寡核苷酸。标记可以包括放射性同位素、酶底物、辅因子、配体、化学发光或荧光剂、半抗原和酶。

[0162] 样本处理设备可以是自动化设备,例如Ventana Medical Systems, Inc.出售的BENCHMARK XT仪器和SYMPHONY仪器。Ventana Medical Systems, Inc.是许多美国专利的受让人,这些专利公开了用于执行自动分析的系统和方法,包括美国专利第5,650,327号、第5,654,200号、第6,296,809号、第6,352,861号、第6,827,901号和第6,943,029号以及美国公开专利申请第20030211630号和第20040052685号,这些专利申请的全部内容通过引用结合于此。可替代地,可以手动处理样本。

[0163] 在处理样本之后,用户可以将带有样本的玻片运送到成像设备。在一些实施方案中,成像设备是明场成像器玻片扫描仪。一种明场成像器是Ventana Medical Systems, Inc.出售的iScan超线程和DP200(Griffin)明场扫描仪。在自动化实施方案中,成像设备是数字病理学装置,如题为IMAGING SYSTEM AND TECHNIQUES(成像系统和技术)的国际专利申请号PCT/US2010/002772(专利公开号为WO/2011/049608)所公开的或于2011年9月9日提交的题为IMAGING SYSTEMS, CASSETTES, AND METHODS OF USING THE SAME(成像系统、暗盒和其使用方法)的美国专利申请号61/533,114所公开的。国际专利申请号PCT/US2010/002772和美国专利申请号61/533,114通过引用整体并入本文中。

[0164] 成像系统或设备可以是多光谱成像(MSI)系统或荧光显微镜系统。这里使用的成

像系统是MSI。MSI通常通过提供对像素级图像的光谱分布的访问用基于计算机化显微镜的成像系统来配备病理样本的分析。虽然存在各种多光谱成像系统,但是所有这些系统共有的操作方面是形成多光谱图像的能力。多光谱图像是捕获特定波长或跨电磁波谱的特定光谱带宽的图像数据的图像。可以通过光学滤波器或通过使用能够选择预定光谱分量的其他仪器来挑选这些波长,所述预定光谱分量包括在可见光范围之外的波长处的电磁辐射,如例如红外(IR)。

[0165] MSI系统可以包括光学成像系统,该光学成像系统的一部分包含光谱选择性系统,该光谱选择性系统可调谐以定义预定数量的N个离散光学带。光学系统可以适于对组织样品进行成像、用宽带光源在透射中照射到光学检测器上。在一个实施方案中可以包括放大系统(如例如显微镜)的光学成像系统具有通常在空间上与光学系统的单个光学输出对准的单个光轴。当调整或调谐光谱选择系统(例如用计算机处理器)时,系统形成组织的一系列图像,如以确保在不同的离散光谱带中获取图像。设备可以另外包含显示器,所述显示器中出现来自所获取的图像的序列中的至少一个视觉上可感知的组织图像。光谱选择系统可以包括光学色散元件(如衍射光栅)、光学滤波器(如薄膜干涉滤光器)的集合、或适于响应于用户输入或预编程处理器的命令从光源通过样品朝向检测器透射的光谱中选择特定通带的任何其他系统。

[0166] 在替代性实施方式中,光谱选择系统定义了对应于N个离散光谱带的若干个光学输出。这种类型的系统从光学系统摄入透射光输出,并且沿着N个空间上不同的光路在空间上重定向这个光输出的至少一部分,其方式为将识别的光谱带中的样品沿着对应于这个识别的光谱带的光路成像到检测器系统上。

[0167] 在本说明书中描述的主题和操作的实施方案可以在数字电子电路中或在计算机软件、固件、或硬件(包括在本说明书中公开的结构及其结构等同物)、或它们中的一个或多个的组合中实施。可以将本说明书中描述的主题的实施方案实施为一个或多个计算机程序,即在计算机存储介质上编码以用于由数据处理设备来执行或者用于控制数据处理设备的操作的计算机程序指令的一个或多个模块。本文所描述的任何模块可以包括由一个或多个处理器执行的逻辑。如本文所使用的,“逻辑”是指具有可以应用于影响处理器操作的指令信号和/或数据形式的任何信息。软件是逻辑的示例。

[0168] 计算机存储介质可以是机器可读存储装置、机器可读储存基板、随机或串行存取存储器阵列或装置、或其中的一项或多项的组合。此外,虽然计算机存储介质不是传播信号,但是计算机存储介质可以是以人工生成的传播信号编码的计算机程序指令的来源或目的地。计算机存储介质还可以是或者可以包括在一个或多个单独的物理部件或介质(例如,多个CD、磁盘或其他存储装置)中。可以将本说明书中描述的操作实施为由数据处理设备对存储在一个或多个计算机可读存储装置上或从其他来源接收的数据执行的操作。

[0169] 术语“编程处理器”包括用于处理数据的所有种类的设备、装置和机器,包括例如可编程微处理器、计算机、芯片上系统或多个芯片上系统、或前述项的组合。设备可以包括专用逻辑电路系统,例如FPGA(现场可编程门阵列)或ASIC(专用集成电路)。除了硬件之外,设备还可包括为所讨论的计算机程序创造执行环境的代码,例如,组成处理器固件、协议栈、数据库管理系统、操作系统、跨平台运行时环境、虚拟机、或其中的一个或多个的组合的代码。设备和执行环境可以实现各种不同的计算模型基础结构,诸如web服务、分布式计算

和网格计算基础结构。

[0170] 计算机程序(也称为程序、软件、软件应用、脚本或代码)可以以任何形式的编程语言书写,包括编译或解释语言、说明性或者过程性语言,并且计算机程序可以以任何形式部署,包括作为独立程序或者作为模块、部件、子例程、对象或适用于计算环境的其他单元。计算机程序可以但不需要对应于文件系统中的文件。可以将程序存储在保持其他程序或数据的文件的一部分(例如,存储在标记语言文档中的一个或多个脚本)中、专用于所讨论的程序的单个文件中、或者多个协调文件(例如,存储一个或多个模块、子程序、或代码的各部分的文件)中。计算机程序可以被部署成在一个计算机上或者在位于一个站点或跨多个站点分布并且通过通信网络互连的多个计算机上执行。

[0171] 本说明书中描述的过程和逻辑流程可以由一个或多个可编程处理器实行,所述一个或多个可编程处理器执行一个或多个计算机程序以便通过对输入数据进行操作并且生成输出来执行动作。过程和逻辑流程还可以由设备执行,并且设备还可以被实施为专用逻辑电路系统,例如FPGA(现场可编程门阵列)或ASIC(专用集成电路)。

[0172] 举例来说,适合于执行计算机程序的处理器包括通用和专用微处理器、以及任何类型的数字计算机的任何一个或多个处理器。通常来说,处理器将从只读存储器或随机存取存储器或二者中接收指令和数据。计算机的必不可少的元件是用于根据指令执行动作的处理器和用于存储指令和数据的一个或多个存储器装置。通常,计算机还将包括用于存储数据的一个或多个大容量存储装置(例如,磁盘、磁光盘或光盘),或者被操作性地耦合以从大容量存储装置接收数据或向大容量存储装置传递数据或两者。然而,计算机不需要有这种装置。此外,计算机可以嵌入另一个装置中,仅举几例,例如移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏控制台、全球定位系统(GPS)接收器或便携式存储装置(例如,通用串行总线(USB)闪存驱动器)。适用于存储计算机程序指令和数据的装置包括所有形式的非易失性存储器、介质和存储器装置,举例来讲,包括半导体存储器装置(例如,EPROM、EEPROM、以及闪存存储器装置)、磁盘(例如,内置硬盘或可移除盘)、磁光盘、以及CD ROM和DVD-ROM盘。处理器和存储器可以由专用逻辑电路补充或结合在其中。

[0173] 为了提供与用户的交互,本说明书中描述的主题的实施方案可以实施在具有用于向用户显示信息的显示装置(例如,LCD(液晶显示器)、LED(发光二极管)显示器或OLED(有机发光二极管)显示器)以及通过其用户可以向计算机提供输入的键盘和定点装置(例如鼠标或轨迹球)的计算机上。在一些实施方案中,触摸屏可以用于显示信息并接收来自用户的输入。还可以使用其他种类的装置来提供与用户的交互;例如,提供给用户的反馈可以是任何形式的感官反馈,例如,视觉反馈、听觉反馈或触觉反馈;并且可以以任何形式接收来自用户的输入,包括声音、语音或触觉输入。另外,计算机可以通过向用户使用的装置发送文档和从用户使用的装置接收文档(例如,通过响应于从用户的客户端装置上的web浏览器接收的请求将网页发送到web浏览器)来与用户交互。

[0174] 本说明书中描述的主题的实施方案可以实施在包括以下的计算系统中:后端部件(例如,作为数据服务器)、或中间件部件(例如,应用服务器)、或前端部件(例如,具有图形用户界面或Web浏览器的客户端计算机,用户可以通过所述图形用户界面或所述Web浏览器与本说明书中描述的主题的实施方式交互)、或者一个或多个这种后端、中间件或前端部件的任何组合。系统的部件可以通过数字数据通信的任何形式或介质(例如,通信网络)进行

互连。通信网络的示例包括局域网 (“LAN”) 和广域网 (“WAN”)、互联网络 (例如, 互联网) 以及对等网络 (例如, 自组织对等网络)。例如, 图1的网络20可以包括一个或多个局域网。

[0175] 计算系统可以包括任何数量的客户端和服务端。客户端和服务端通常远离彼此并且通常通过通信网络进行交互。客户端与服务端的关系借助于在相应计算机上运行并且彼此具有客户端-服务端关系的计算机程序产生。在一些实施方案中, 服务端将数据 (例如, HTML页面) 传输到客户端装置 (例如, 目的是向与客户端装置交互的用户显示数据和从与客户端装置交互的用户接收用户输入)。可以从服务端处的客户端装置接收在客户端装置处生成的数据 (例如, 用户交互的结果)。

[0176] 解混的附加方法/可选解混模块

[0177] 解混是将混合像素的测量光谱分解成一组组成光谱或端元以及一组相应的分数或丰度的过程, 这些分数或丰度表示像素中存在的每个端元的比例。具体而言, 解混过程可以提取染色剂特异性通道, 以使用标准类型的组织和染色剂组合所熟知的参考光谱来确定单个染色剂的局部浓度。解混可以使用从对照图像检索的或者从观察图像估计的参考光谱。解混每个输入像素的分量信号使得能够检索和分析染色特异性通道, 例如H&E图像中的苏木精通道和曙红通道, 或者IHC图像中的二氨基联苯胺 (DAB) 通道和复染色 (例如苏木精) 通道。术语“解混”和“颜色反卷积” (或“反卷积”) 等 (例如, “反卷积”、“解混”) 在本领域中可以互换使用。在一些实施方案中, 使用线性解混通过解混模块解混多路图像。线性解混描述于例如 ‘Zimmermann “Spectral Imaging and Linear Unmixing in Light Microscopy” Adv Biochem Engin/Biotechnol (2005) 95:245-265’ 以及 C.L.Lawson 和 R.J.Hanson, “Solving least squares Problems”, PrenticeHall, 1974, 第23章, 第161页’, 所述文献的披露内容通过引用整体并入本文。在线性染色剂解混中, 任何像素处的测量光谱 ( $S(\lambda)$ ) 被认为是染色剂光谱成分的线性混合, 并且等于在像素处表示的每个单独染色剂的颜色参考 ( $R(\lambda)$ ) 的比例或权重 ( $A$ ) 的总和

[0178]  $S(\lambda) = A_1 \cdot R_1(\lambda) + A_2 \cdot R_2(\lambda) + A_3 \cdot R_3(\lambda) \dots \dots A_i \cdot R_i(\lambda)$

[0179] 这可以更一般地表示为矩阵形式

[0180]  $S(\lambda) = \sum A_i \cdot R_i(\lambda)$  或  $S = R \cdot A$

[0181] 如果获取了  $M$  个信道图像并且存在  $N$  种单独的染色剂, 则  $M \times N$  矩阵  $R$  的列是如本文导出的最佳颜色系统、 $N \times 1$  向量  $A$  是单独染色剂的未知比例并且  $M \times 1$  向量  $S$  是像素处测量的多通道光谱向量。在这些方程中, 每个像素中的信号 ( $S$ ) 在获取多路图像和参考光谱期间进行测量, 即最佳颜色系统如本文所描述的那样导出。各种染色剂的贡献 ( $A_i$ ) 可以通过计算它们对测量的光谱中的每个点的贡献来确定。在一些实施方案中, 使用最小二乘逆拟合方法来获取解决方案, 所述方法通过求解以下方程组来最小化测量光谱和计算光谱之间的平方差,

[0182]  $[\partial \sum_j \{S(\lambda_j) - \sum_i A_i \cdot R_i(\lambda_j)\}^2] / \partial A_i = 0$

[0183] 在这个等式中,  $j$  代表检测通道的数量,  $i$  等于染色剂的数量。线性方程解通常包括允许受约束的解混来强制权重 ( $A$ ) 求和为 1。

[0184] 在其他实施方案中, 使用在 2014 年 5 月 28 日提交的题为 “Image Adaptive Physiologically Plausible Color Separation (图像自适应生理上似然颜色分离)” 的 W02014/195193 中描述的方法来完成解混, 其披露内容通过引用整体并入本文中。一般而



言,W02014/195193描述了一种通过使用迭代优化的参考向量分离输入图像的分量信号来进行解混的方法。在一些实施方案中,来自测定的图像数据与特定于测定特征的预期或理想结果相关,以确定质量度量。在低质量图像或与理想结果相关性差的情况下,调整矩阵R中的一个或多个参考列向量,并且使用调整后的参考向量迭代地重复解混,直到相关性显示出匹配生理和解剖要求的良好质量图像。解剖、生理和测定信息可用于定义应用于测量图像数据的规则,以确定质量度量。这些信息包括组织是如何染色的,组织内的哪些结构是打算染色的或不打算染色的,以及结构、染色剂和特定于正在处理的测定的标记之间的关系。迭代过程产生特定于染色的向量,该向量可以生成精确标识感兴趣结构和生物学相关信息的图像,没有任何噪声或不想要的光谱,因此适于分析。参考向量被调整到搜索空间内。搜索空间定义了参考向量可以用来表示染色剂的值的范围。搜索空间可以通过扫描包括已知或常见问题在内的各种代表性训练测定,并确定训练测定的高质量参考向量集来确定。

[0185] 在其他实施方案中,使用在2015年2月23日提交的题为“Group Sparsity Model for Image Unmixing (用于图像解混的群稀疏模型)”的W02015/124772中描述的方法来完成解混,其披露内容通过引用整体并入本文中。总的来说,W02015/124772描述了使用组稀疏性框架来解混,其中在“相同的组”内对来自多个共存标记的染色贡献的分数进行建模,并且在不同的组中对来自多个非共存标记的染色贡献的分数进行建模,向建模的组稀疏性框架提供多个共存标记的共同定位信息,使用组套索求解建模的框架以在每个组内产生最小二乘解,其中最小二乘解对应于共存标记的解混,并且在对应于非共存标记的解混的组中产生稀疏解。此外,W0 2015124772描述了一种通过输入从生物组织样品获取的图像数据、从电子存储器读取参考数据、从电子存储器读取共存数据来解混的方法,所述参考数据描述多种染色剂中每一种染色剂的染色剂颜色,所述共存数据描述染色剂的组,每个组包括可以在生物组织样品中并置的染色剂,并且每个组形成用于组套索标准的组,至少一个组具有二或更大的大小,并且使用参考数据作为参考矩阵来计算用于获得解混合图像的组套索标准的解。在一些实施方案中,用于解混图像的方法可以包括生成组稀疏模型,其中来自共定位标记的一部分染色贡献被分配在单个组内,来自非共定位标记的一部分染色贡献被分配在单独的组内,并且使用解混算法求解组稀疏模型以在每个组内产生最小二乘解。

[0186] 示例-高分辨率和中分辨率分析方法之间FAP阳性面积的比较

[0187] 进行该实验是为了比较FAP阳性面积结果的准确性,使用:

[0188] 1) FAP阳性高分辨率分析。对于该测量,以0.465微米像素大小的空间分辨率累积在高放大率(20X)下阈值化后的所有FAP阳性像素。然后,从预先标注的区域中选择的报告面积被获得作为感兴趣区域的逐像素FAP阳性面积。

[0189] 2) 在预先标注的区域内,使用本文描述的中分辨率分析法测量的FAP阳性面积,其通过对FAP超像素对象、种子或多边形轮廓的FAP阳性面积求和来计算。

[0190] 六个不同的标注区域(见图14)根据两种方法进行分析,每个区域具有不同的形状(大、小、圆或奇怪的形状等)。如图15和下表所示,使用这两种方法测量的FAP阳性面积的比较结果之间没有显著差异( $R^2=0.99, p<0.001$ )。

玻片 15-020499 中不同标注的 FAP 阳性面积测量			
标注	方法#1	方法#2	NID
[0191] 1	1.68E+05	1.68E+05	40958
2	2.00E+05	2.02E+05	229850
3	1.33E+05	1.35E+05	229865
4	1.94E+04	1.97E+04	229873
5	3.38E+03	3.63E+03	229886
6	6.12E+05	6.13E+05	229890

[0192] 方法#1-使用FAP+高分辨率结果图像进行测量

[0193] 方法#2-通过FAP种子的FAP+面积的附加特征的总和进行测量

[0194] 总之,当我们在特定标注中对超像素内计算的面积特征求和时,面积的总和等于我们在该标注中使用高分辨率分析方法直接计算的面积。FAP阳性面积的结果显示,在具有不同形状的标注区域的两种方法(利用和不利用超像素)之间计算没有显著差异。

[0195] 本说明书中提到的和/或在申请数据表中列出的所有美国专利、美国专利申请公开、美国专利申请、外国专利、外国专利申请和非专利出版物通过引用整体并入本文。如果需要,可以修改实施方案的方面,以采用各种专利、申请和出版物的概念来提供进一步的实施方案。

[0196] 尽管已经参考多个说明性实施方案描述了本公开,但是应当理解,本领域技术人员可以设计出在本公开的原理的精神和范围内的许多其他修改和实施方案。更具体地,在不脱离本公开的精神的情况下,在前述公开、附图和所附权利要求的范围内,主题组合布置的组成部分和/或布置中的合理变化和修改是可能的。除了部件和/或布置的变化和修改之外,替代用途对于本领域技术人员也是显而易见的。

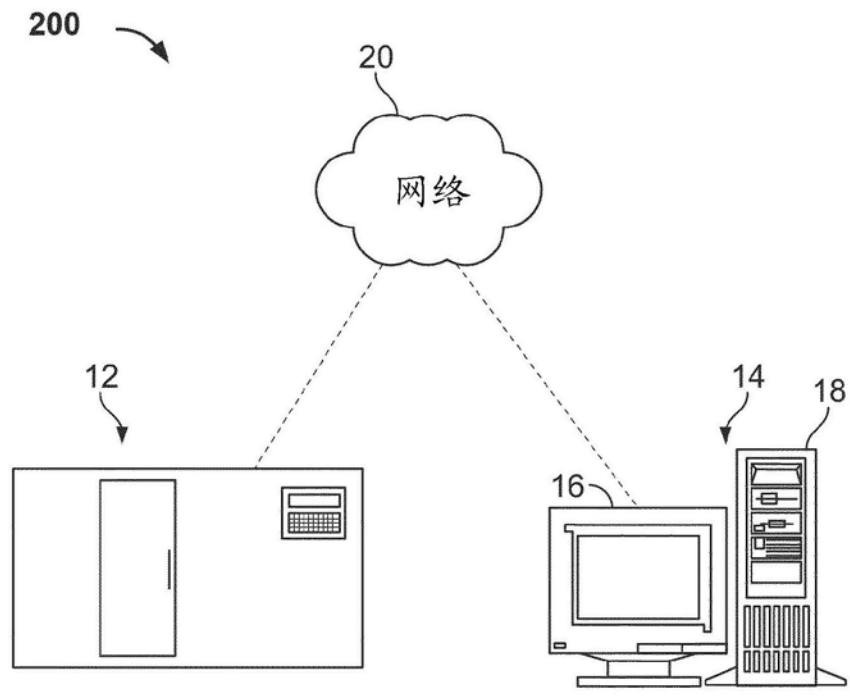


图1

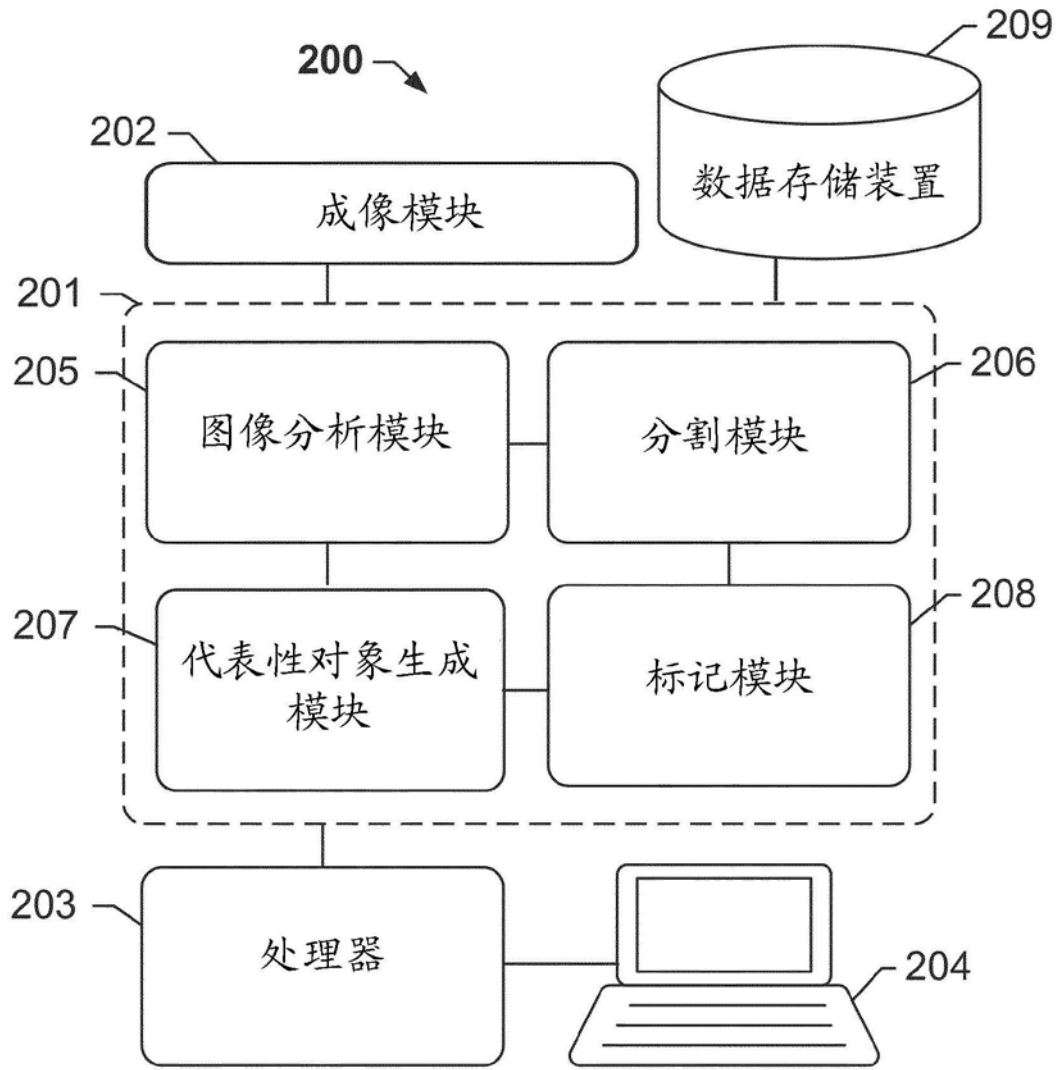


图2

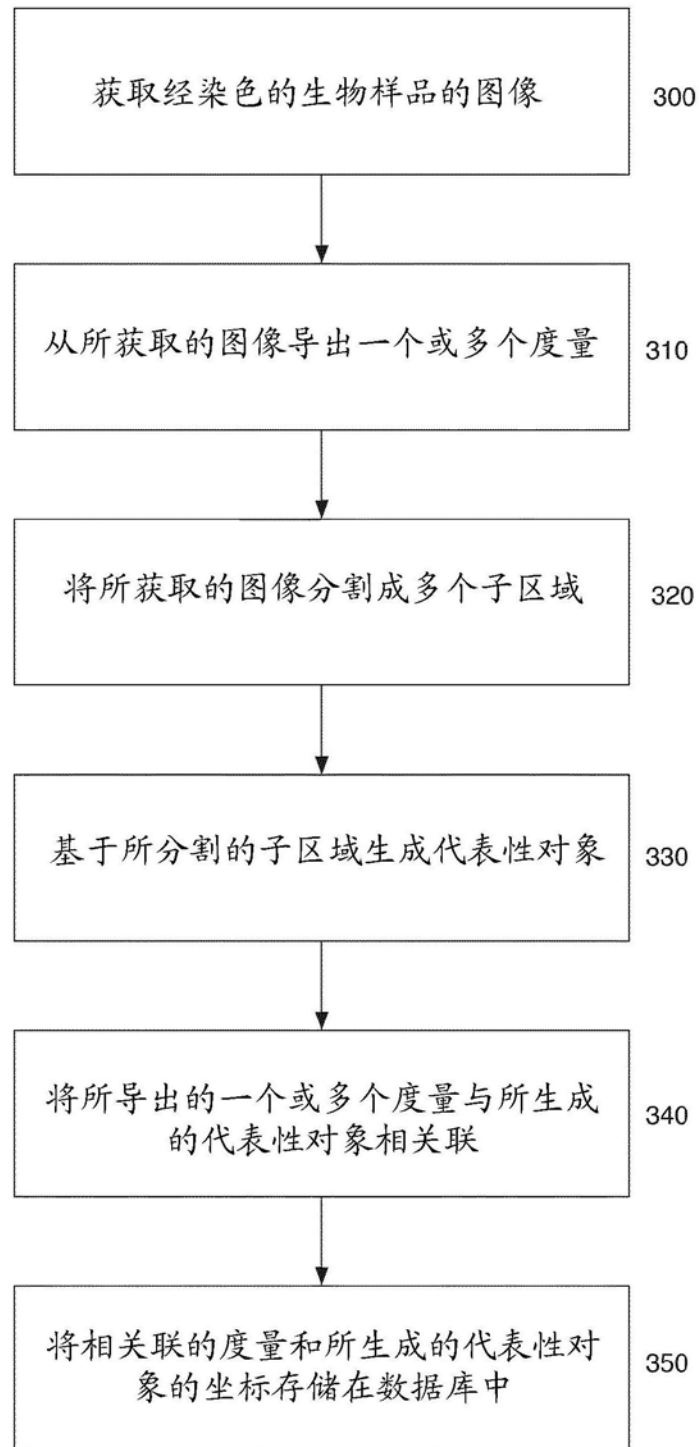


图3

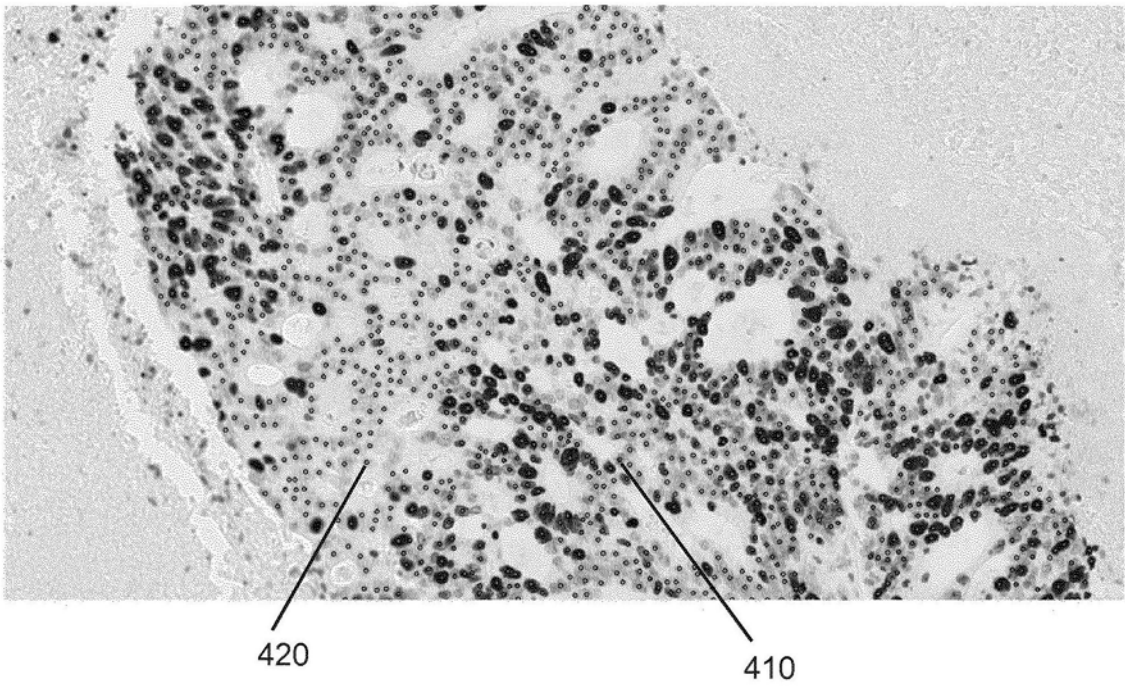


图4

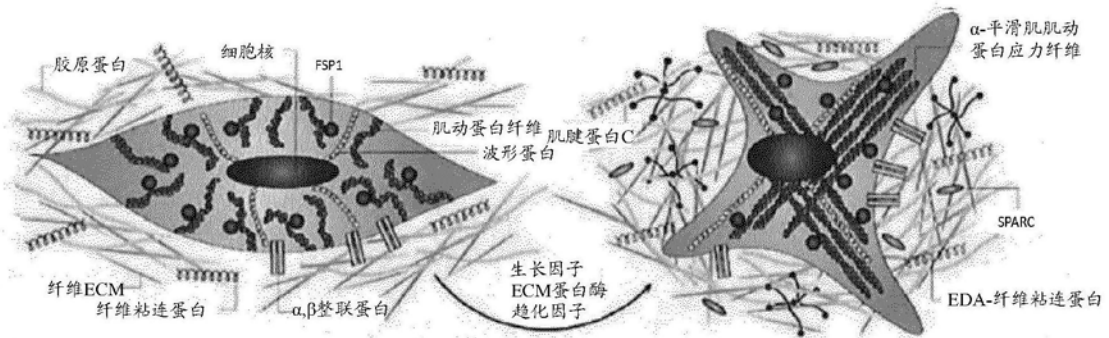


图5A

图5B

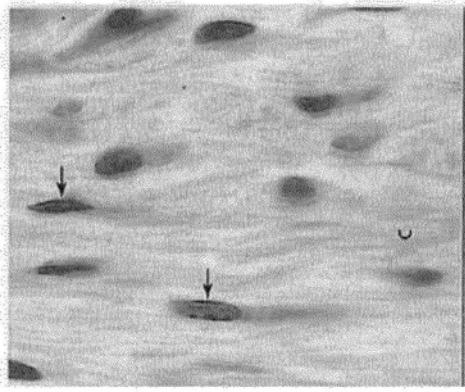


图5C

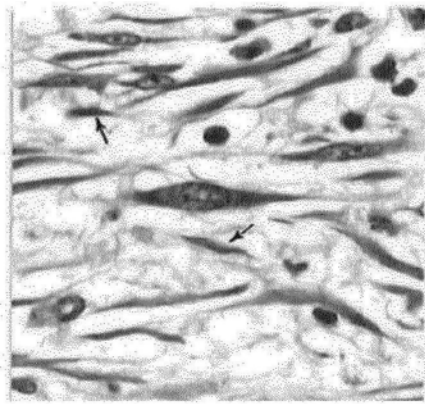


图5D

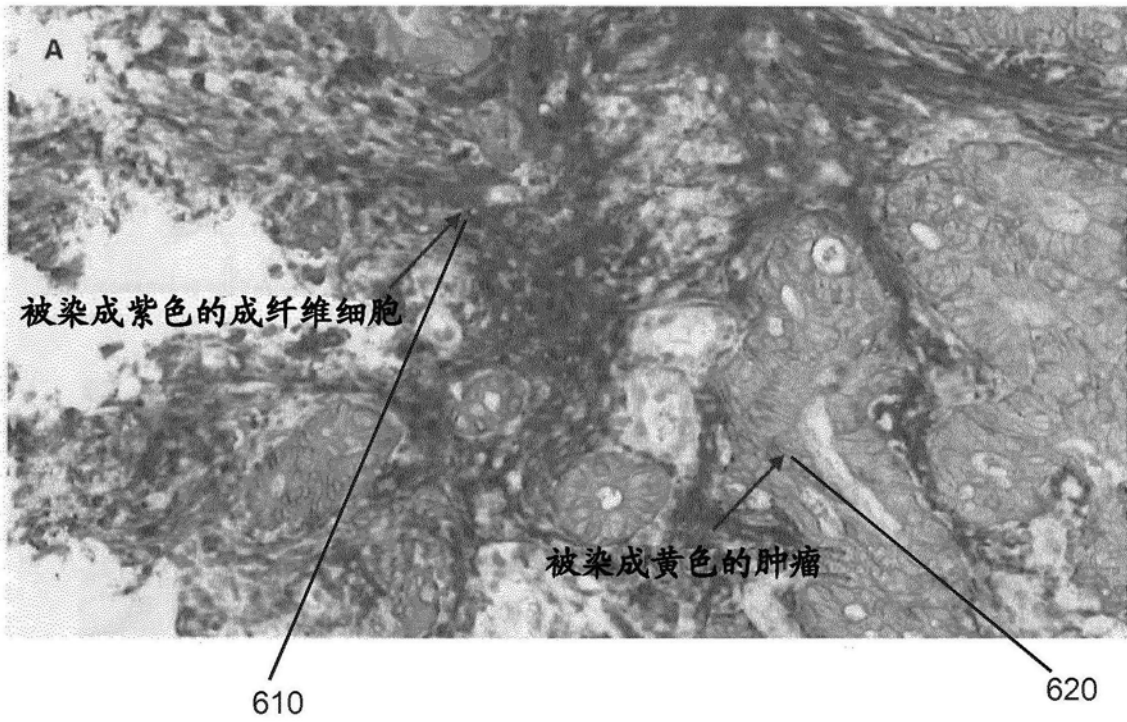


图6A

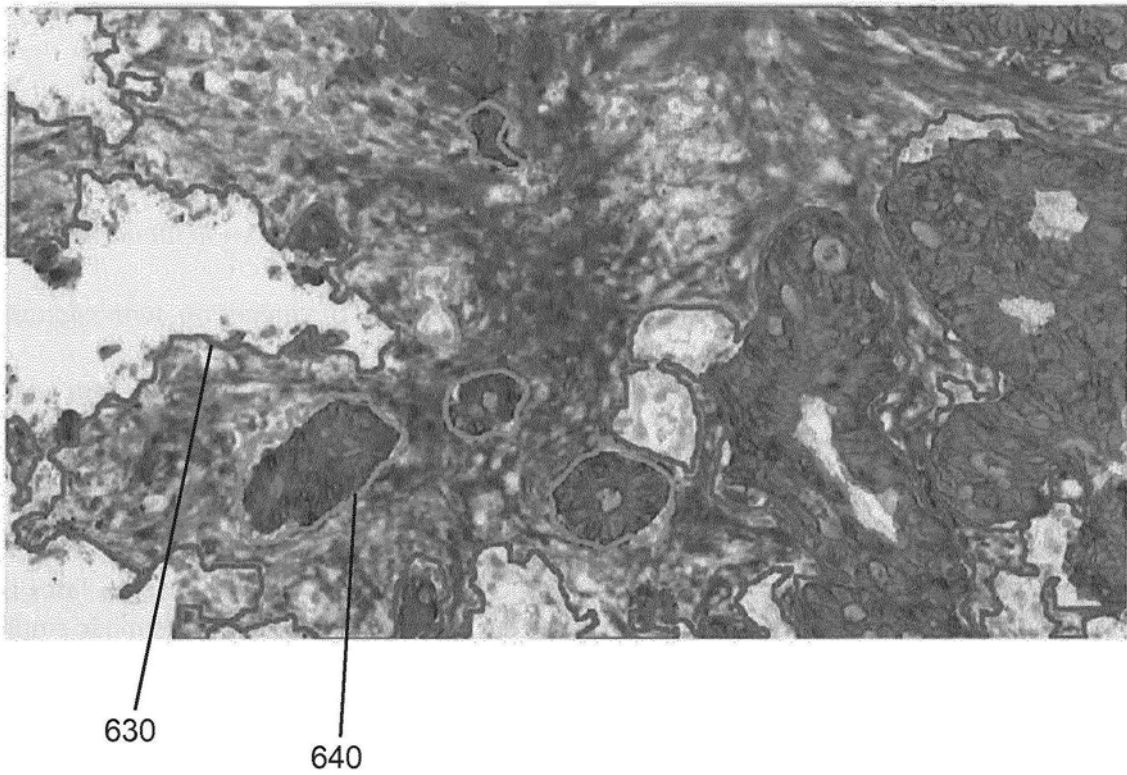


图6B



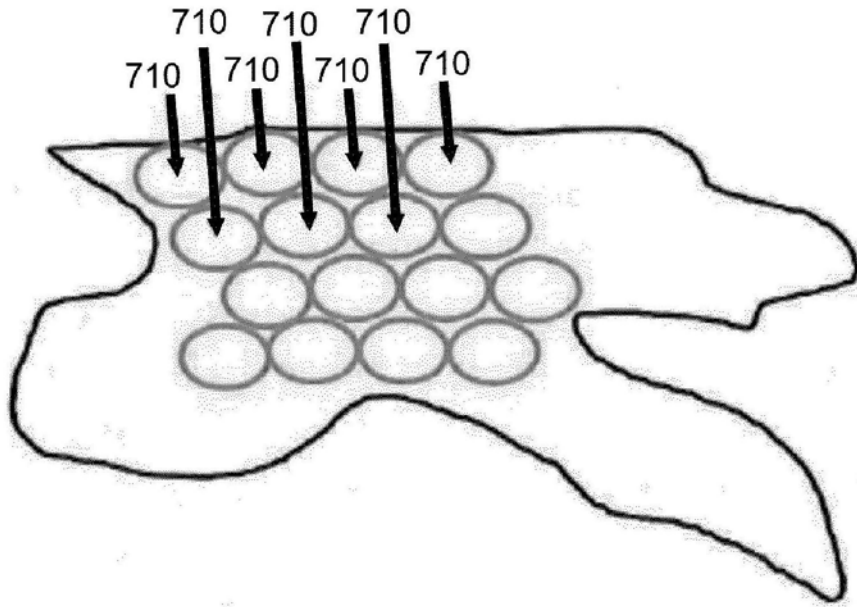
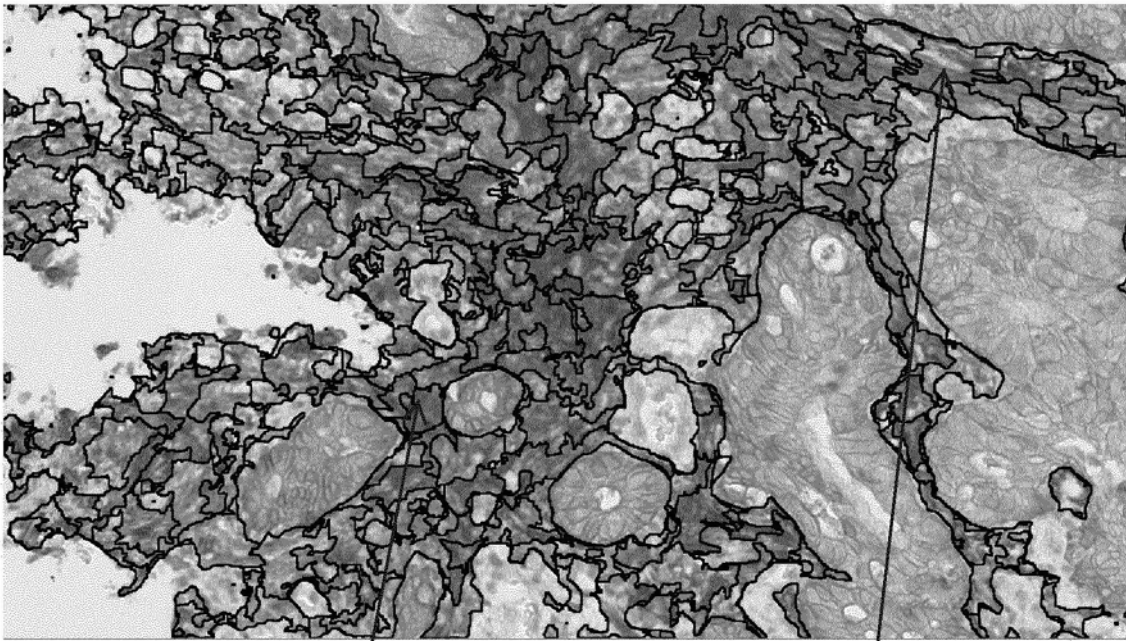


图7



810

820

图8A

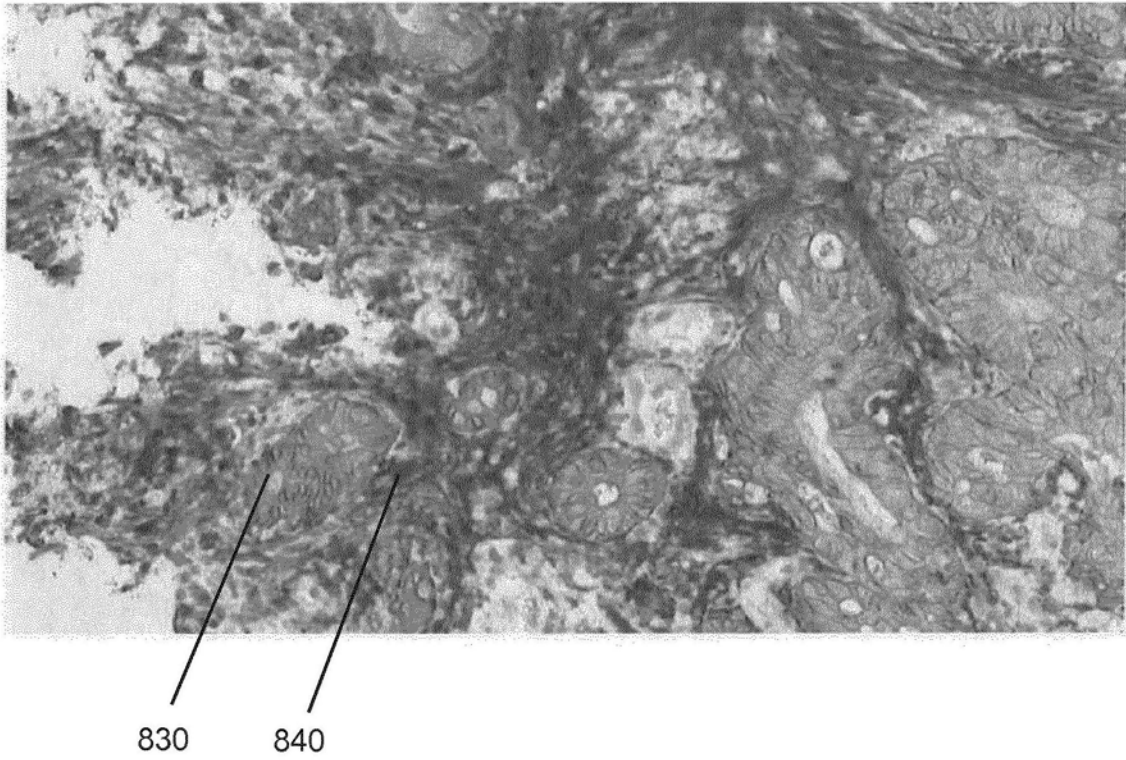


图8B

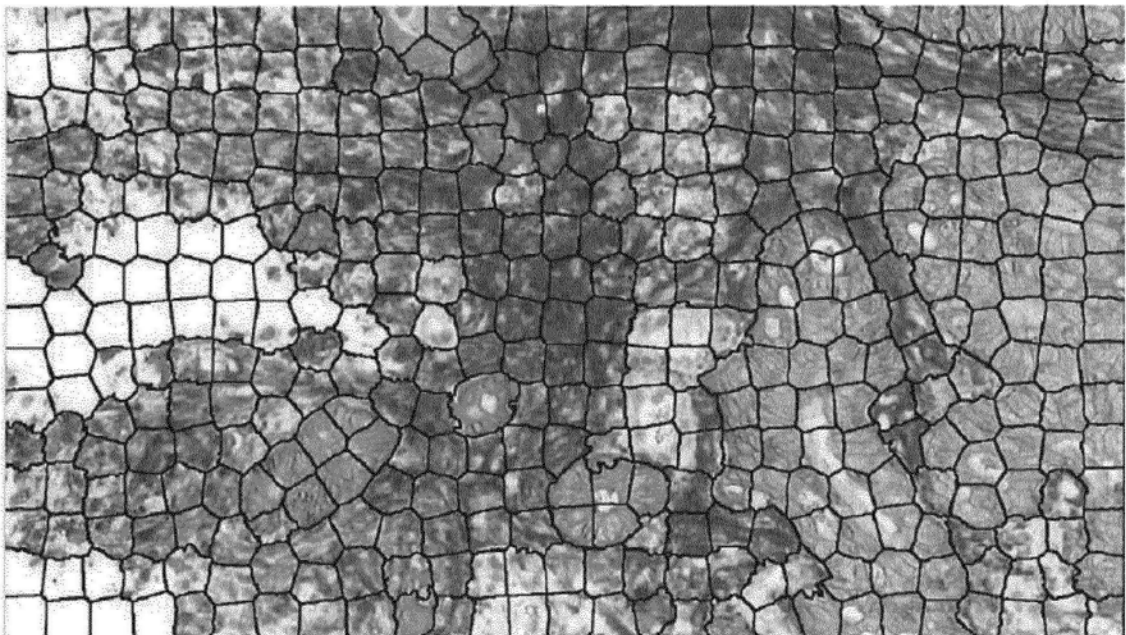


图8C

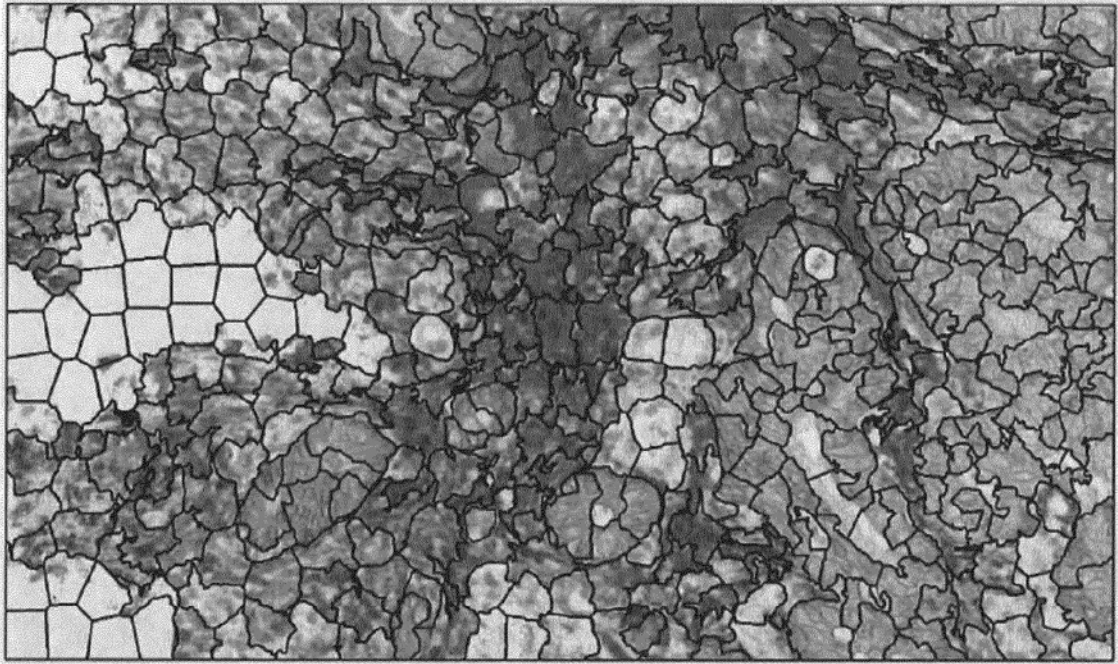
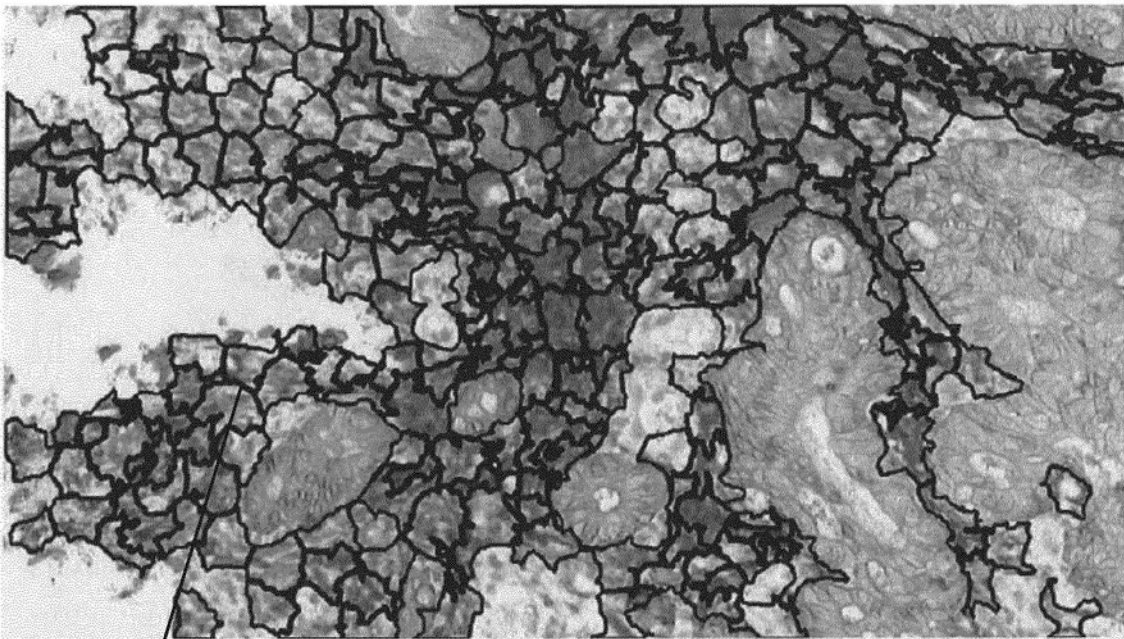


图8D



910

图9A

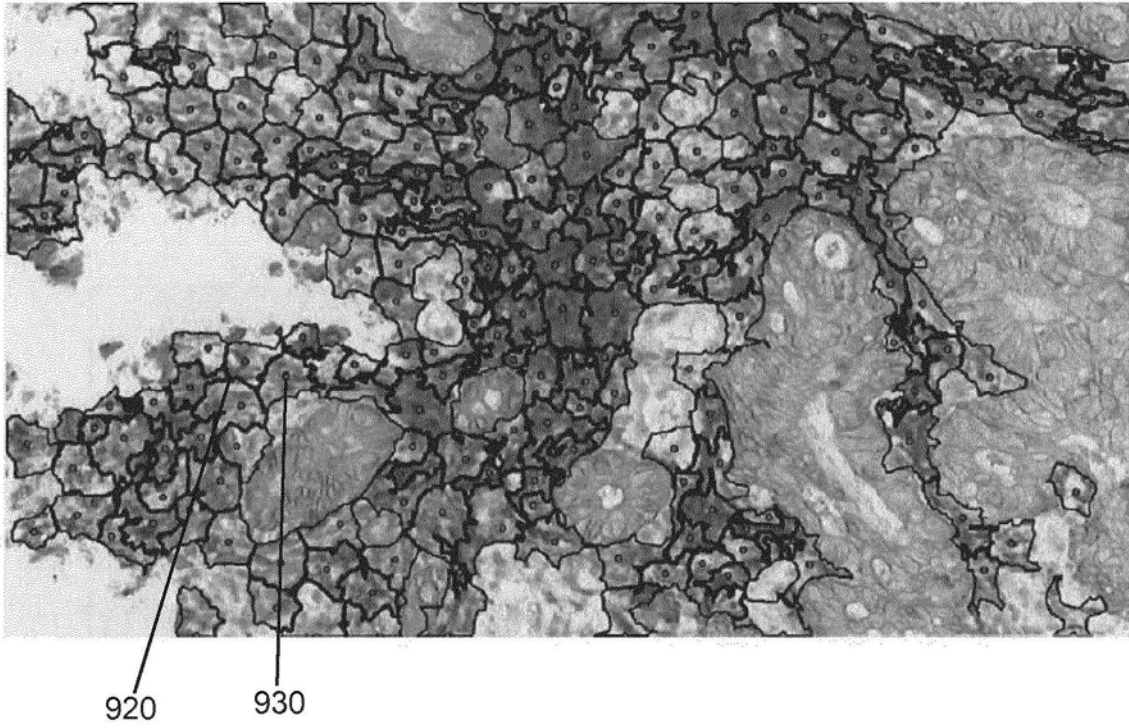


图9B

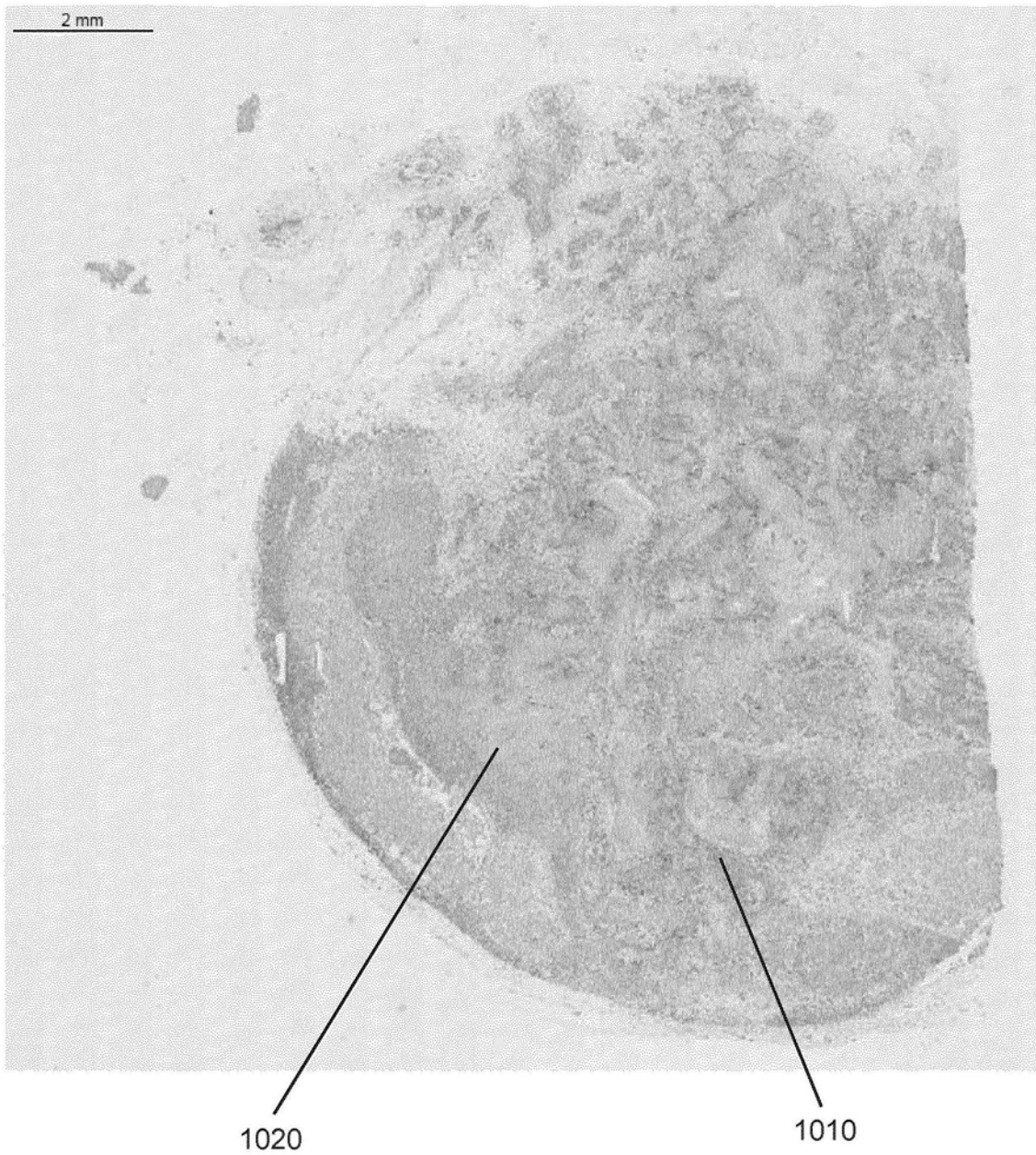


图10A

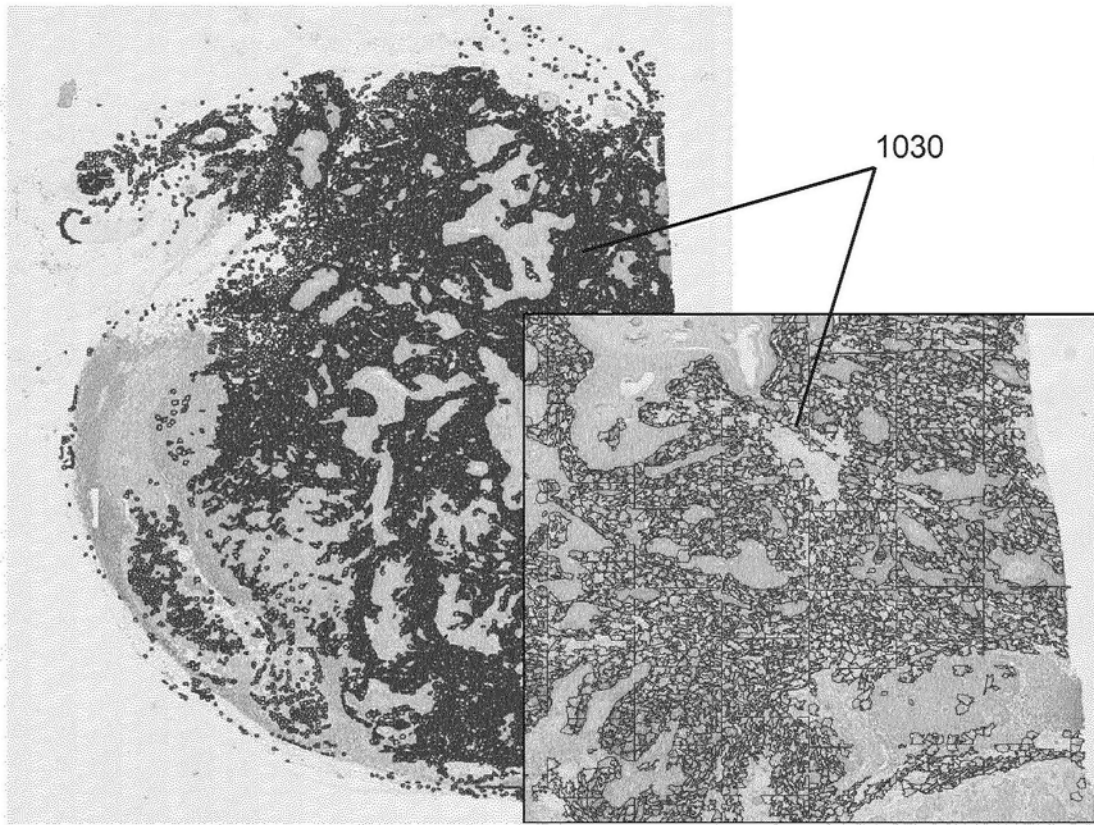


图10B

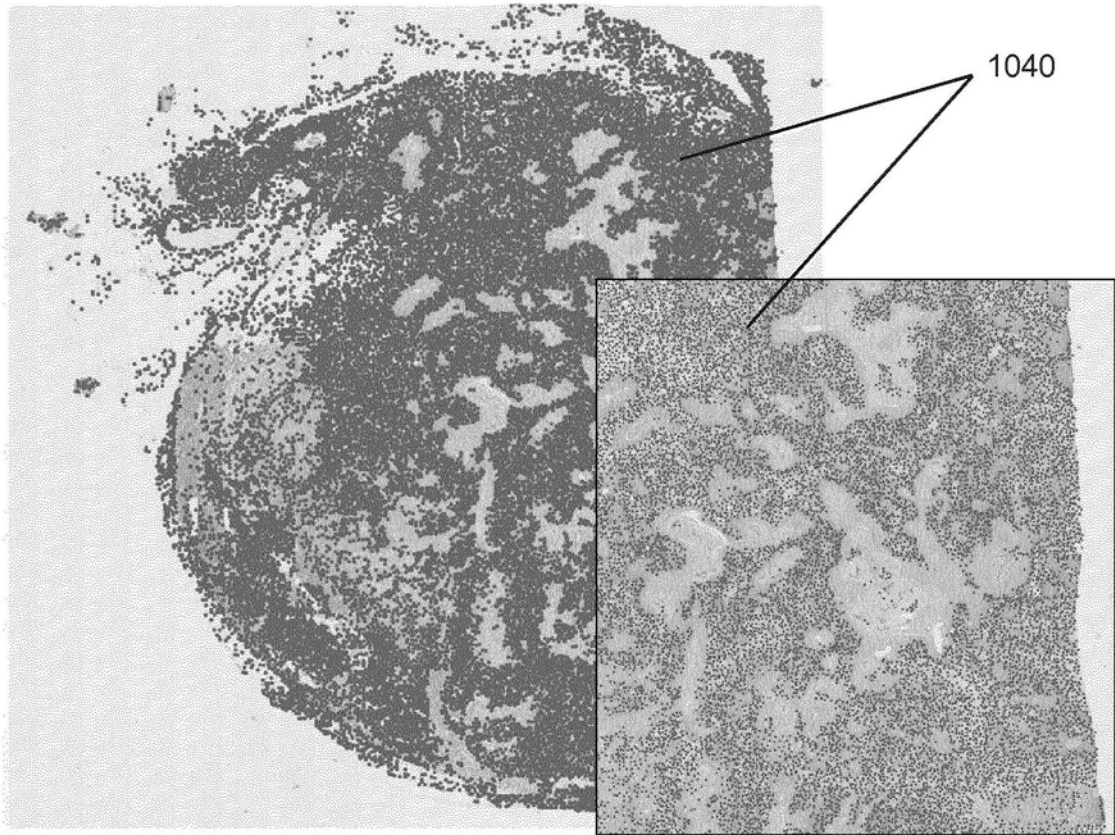


图11

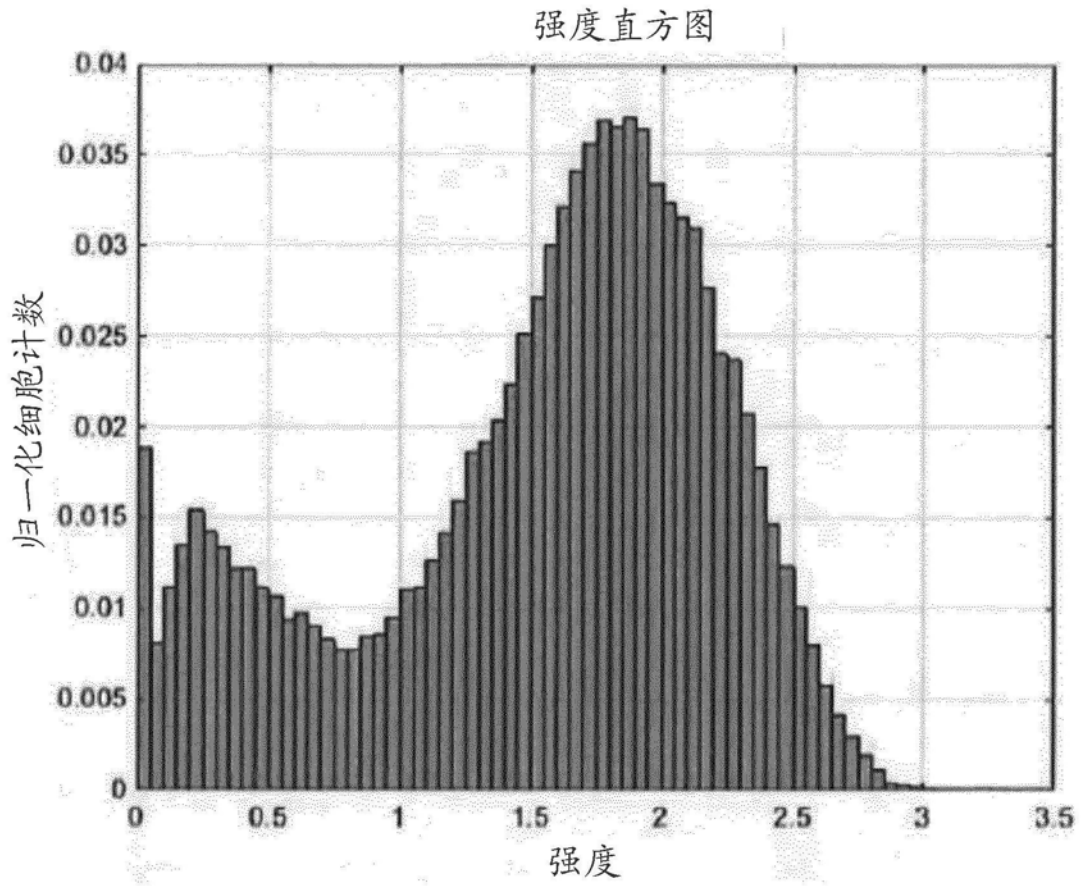


图12



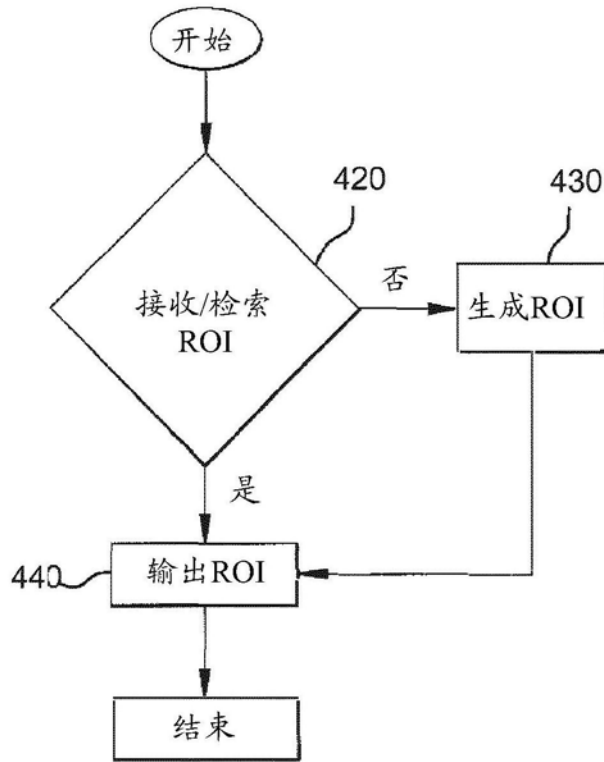


图13

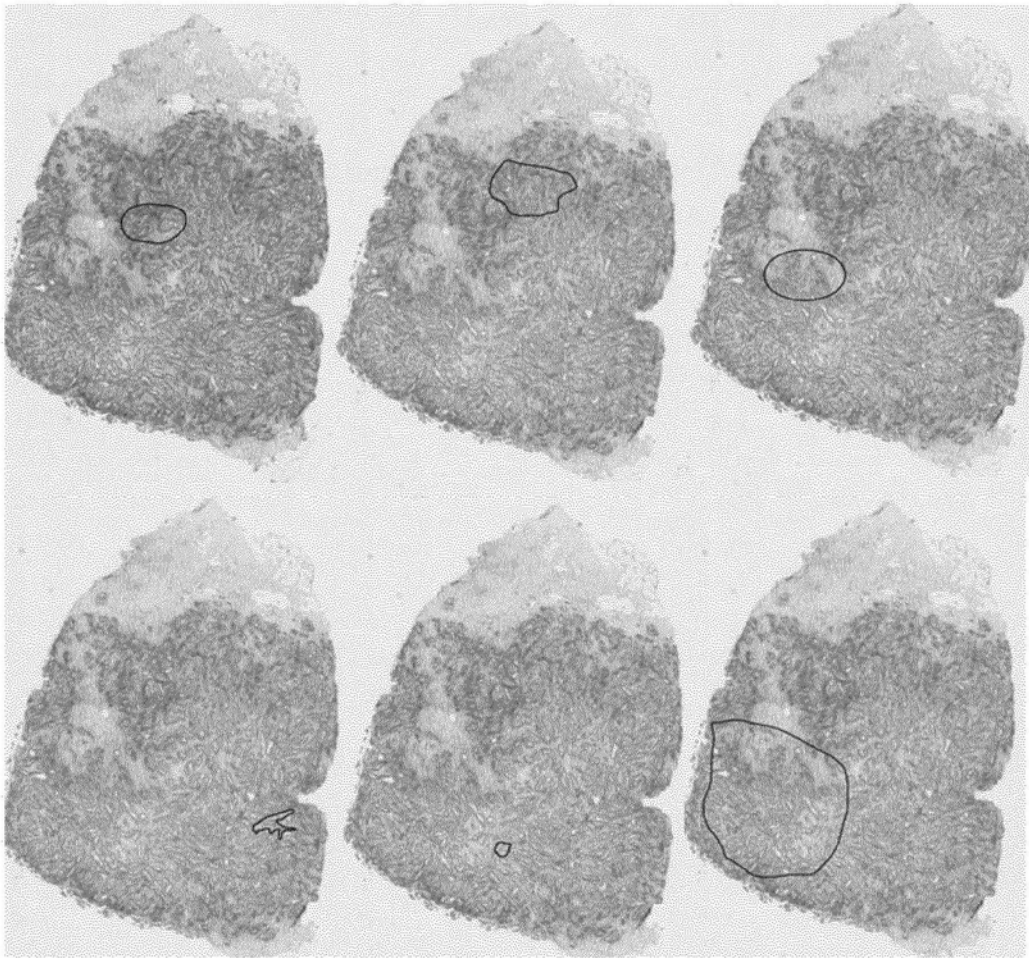


图14

两种不同测量之间的FAP阳性区域的百分比的一致性

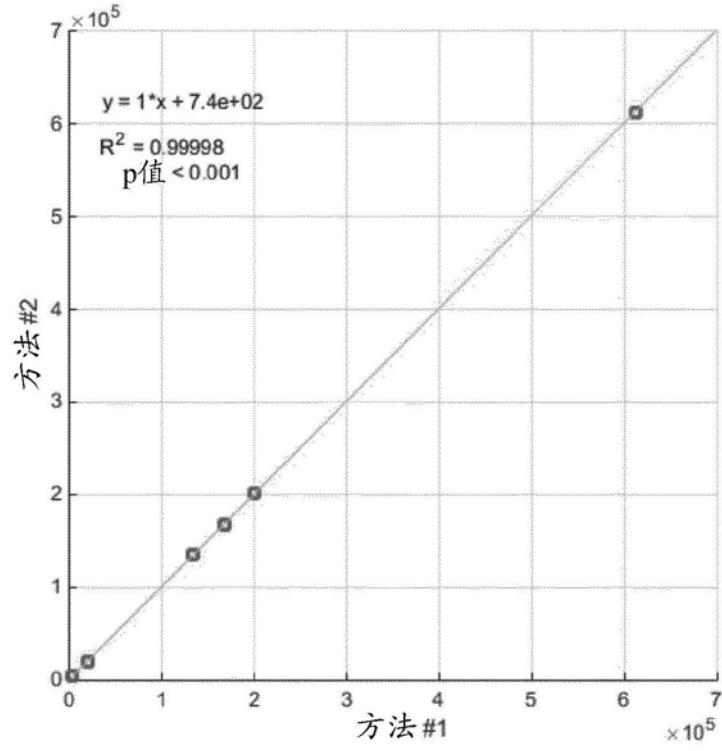


图15