(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2024 年 11 月 28 日 (28.11.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/239285 A1

(51) 国际专利分类号:
G06T 7/33 (2017.01)

(21) 国际申请号: PCT/CN2023/096110

(22) 国际申请日: 2023 年 5 月 24 日 (24.05.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 深圳华大生命科学研究院 (**BGI SHENZHEN**) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

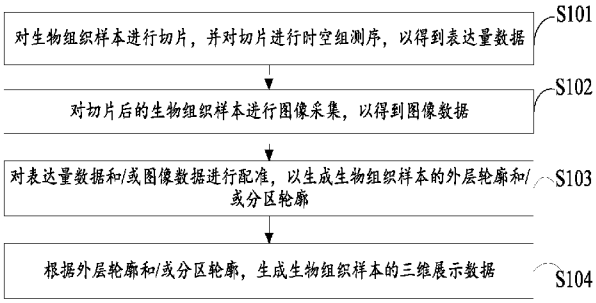
(72) 发明人: 刘超 (**LIU, Chao**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。夏天祎 (**XIA, Tianyi**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。史昶 (**SHI, Chang**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。郭俊甫 (**GUO, Junfu**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。彭勇 (**PENG, Yong**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。张博涵 (**ZHANG, Bohan**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。陆芹 (**LU, Qin**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。李美 (**LI, Mei**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。张勇 (**ZHANG, Yong**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。黎宇翔 (**LI, Yuxiang**); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(74) 代理人: 北京法胜知识产权代理有限公司 (**FASHENG INTELLECTUAL PROPERTY COMPANY, LTD.**); 中国北京市海淀区悦秀路99号1单元4层401, Beijing 100085 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) **Title:** METHOD, APPARATUS, AND SYSTEM FOR DATA ACQUISITION

(54) 发明名称: 数据采集方法、装置及系统



S101 Section a biological tissue sample, and perform time-and-space omics sequencing on the sections to obtain expression level data

S102 Perform image acquisition on the sectioned biological tissue sample to obtain image data

S103 Perform registration on the expression level data and/or the image data to generate an outer-layer contour and/or a partition contour of the biological tissue sample

S104 Generate three-dimensional display data for the biological tissue sample according to the outer-layer contour and/or the partition contour

图 1

(57) **Abstract:** Embodiments of the present disclosure provide a method, apparatus, and system for data acquisition, which relate to the technical field of data processing. The method comprises: sectioning a biological tissue sample, and performing time-and-space omics sequencing on the sections to obtain expression level data; performing image acquisition on the sectioned biological tissue sample to obtain image data; performing registration on the expression level data and/or the image data to generate an outer-layer contour and/or a partition contour of the biological tissue sample; and generating three-dimensional display data for the biological tissue sample according to the outer-layer contour and/or the partition contour. The present disclosure provides a whole set of apparatus and method for data acquisition, and a registration method for expression level data and tissue section image data of time-and-space omics sequencing. Aiming at in situ capture time-and-space omics technology, combined with optical imaging information and assistance of rigid and elastic registration of continuous tissue sections, three-dimensional reconstruction is achieved; therefore, auxiliary visual information positioning of inner and outer tissue contours is obtained, and time-and-space omics data with large size and high resolution is effectively processed and displayed.

WO 2024/239285 A1

[见续页]

CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本公开实施例提供一种数据采集方法、装置及系统，涉及数据处理技术领域。对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；对切片后的生物组织样本进行图像采集，以得到图像数据；对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓；根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据。本公开提出一整套的数据采集装置与方法，以及时空测序的表达量数据与组织切面图像数据的配准方法，针对原位捕获的时空组学技术，结合光学成像信息，辅助连续组织切片的刚性、弹性配准，实现三维重构，进而得到内外组织轮廓辅助可视化信息定位，对大尺寸高分辨率时空组数据进行有效处理与展示。

数据采集方法、装置及系统

技术领域

本公开涉及数据处理技术领域，特别涉及一种数据采集方法、装置及系统。

背景技术

时空组学技术自发展以来，已在神经科学、病理学、发育生物学等方向得到初步应用。二维平面的时空组学分析难以得到三维空间的形态结构及准确的位置信息，如何突破二维平面的研究实现三维时空组学研究具有极大的技术挑战性和重要的科学意义。目前主要研究时空组学三维重构方法，以实现对组织三维水平上不同细胞的空间分布、相互作用、基因蛋白质等表达情况以及整体形态结构的全面解析。当前的时空组学三维重构仅依据时空组捕获的表达量丰度信息，将测量于平面上的表达量数据进行刚性配准，存在数据重构回原位表达水平准确性低、可视化效果有限、对于大数据量的组织样本数据处理和后期展示能力不足等问题。

发明内容

本公开提出了一种数据采集方法、装置及系统，针对原位捕获的时空组学技术，结合光学成像信息得到内外组织轮廓辅助可视化信息定位，每个信息采集位点实现细胞精度的位置配准，而且可以对大尺寸高分辨率时空组数据进行有效处理与展示。

本公开的第一方面实施例提供了一种数据采集方法，该方法包括：对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；对切片后的生物组织样本进行图像采集，以得到图像数据；对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓；根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据。

在本公开的一些实施例中，对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据包括：对生物组织样本进行连续切片，每切第一预定数量的切片进行一次时空组测序，以得到切片的表达量数据，其中，第一预定数量为正整数。

在本公开的一些实施例中，该方法还包括：对表达量数据进行预处理，其中，预处理包括聚合处理、质量控制、对数归一化、找高变基因、降维处理中的至少一项。

在本公开的一些实施例中，对切片后的生物组织样本进行图像采集，以得到图像数据包括：每切第二预定数量的切片，对切片后的生物组织样本进行一次图像采集，以得到图像数据，其中，第二预定数量为正整数。

在本公开的一些实施例中，对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓包括：对图像数据进行图像分割处理，以生成分割结果图；对分割结果图进行配准，以生成配准结果图；基于配准结果图，生成生物组织样本的外层轮廓。

在本公开的一些实施例中，对图像数据进行图像分割处理包括：将图像数据转换为 HSV 颜色空间，提取 H 通道的图像数据并转换为灰度图；通过预设图像处理算法，对图像数据进行二值化，以确定二值化结果轮廓图；保留符合预设条件的轮廓图，与图像数据进行合并。

在本公开的一些实施例中，该方法还包括：获取图像数据的灰度化图像和直方图；对图像数据进行最大熵阈值二值化处理；将处理后的图像与图像数据进行合并。

在本公开的一些实施例中，对分割结果图进行配准，以生成配准结果图包括：对分割结果图进行空间变换处理和线性插值算法处理；对经处理的分割结果图以及配准参考图进行均方差相似度测量；将测量最优解确定为配准结果图。

在本公开的一些实施例中，基于配准结果图，生成生物组织样本的外层轮廓包括：将所有配准结果图堆叠为预设格式的数组文件；对数组文件进行处理，以得到生物组织样本的外层轮廓。

在本公开的一些实施例中，对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓包括：对表达量数据进行计算，得到表达量图；将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；对图像对依次进行刚性配准和弹性配准，以生成分区轮廓。

在本公开的一些实施例中，对图像对进行刚性配准包括：对图像对分别进行下采样；提取图像对的特征信息；对图像对进行相似性估计仿射变换，得到刚性配准结果。

在本公开的一些实施例中，对图像对进行弹性配准包括：对刚性配准结果进行直方图均衡化；将刚性配准结果输入深度学习模型，输出深度学习模型的中间结果，得到形变场，其中，深度学习模型是通过表达量图-图像数据对进行训练得到的；根据形变场，对刚性配准结果进行最近邻插值变换，得到弹性配准结果。

在本公开的一些实施例中，对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓包括：对两个相邻或相近切片中的表达量数据进行配准，以得到配准表达量数据；根据配准表达量数据，生成生物组织样本的外层轮廓和/或分区轮廓。

在本公开的一些实施例中，根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据包括：对表达量数据和/或图像数据进行配准得到的配准结果进行生物信息学分析，得到聚类和注释结果；基于聚类和注释结果，构造三维数组，并增加空间尺度信息；利用三维渲染工具，渲染得到生物组织样本的三维展示数据。

本公开实施例的第二方面提供一种数据处理方法，该方法包括：获取生物组织样本的表达量数据和图像数据，其中表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，图像数据是对切片后的生物组织样本进行图像采集得到的；对表达量数据进行计算，得到表达量图；将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据。

在本公开的一些实施例中，对图像对进行刚性配准包括：对图像对分别进行下采样；提取图像对的特征信息；对图像对进行相似性估计仿射变换，得到刚性配准结果。

在本公开的一些实施例中，对图像对进行弹性配准包括：对刚性配准结果进行直方图均衡化；将刚性配准结果输入深度学习模型，输出深度学习模型的中间结果，得到形变场，其中，深度学习模型是通过表达量图-图像数据对进行训练得到的；根据形变场，对刚性配准结果进行最近邻插值变换，得到弹性配准结果。

本公开第三方面实施例提供一种数据采集装置，该装置包括：测序模块，用于对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；采集模块，用于对切片后的生物组织样本进行图像采集，以得到图像数据；配准模块，用于对表达量数据和/或图像数据进行配准，以生成生物组

织样本的外层轮廓和/或分区轮廓；生成模块，用于根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据。

本公开第四方面实施例提供一种数据处理装置，该装置包括：获取模块，用于获取生物组织样本的表达量数据和图像数据，其中表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，图像数据是对切片后的生物组织样本进行图像采集得到的；计算模块，用于对表达量数据进行计算，得到表达量图；均衡化模块，用于将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；配准模块，对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据。

本公开第五方面实施例提供一种电子设备，包括：至少一个处理器；以及与至少一个处理器通信连接的存储器；其中，存储器存储有可被至少一个处理器执行的指令，指令被至少一个处理器执行，以使至少一个处理器能够执行本公开第一方面实施例或第二方面实施例中描述的方法。

本公开第六方面实施例提供一种计算机存储介质，其中，计算机存储介质存储有计算机可执行指令；计算机可执行指令被处理器执行后，能够实现本公开第一方面或第二方面实施例中描述的方法。

根据本公开的数据采集方法，对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；对切片后的生物组织样本进行图像采集，以得到图像数据；对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓；根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据。本公开的方法针对原位捕获的时空组学技术，结合光学成像信息，辅助连续组织切片的刚性、弹性配准，实现三维重构，进而得到内外组织轮廓辅助可视化信息定位，对大尺寸高分辨率时空组数据进行有效处理与展示。

本公开附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本公开的实践了解到。

附图说明

本公开上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

图 1 为根据本公开实施例的一种数据采集方法的流程示意图；

图 2 为根据本公开实施例的一种数据采集方法的流程示意图；

图 3 为根据本公开实施例的一种数据采集方法的示意图；

图 4 为根据本公开实施例的一种切片机的示意图；

图 5 为根据本公开实施例的一种拍照装置滑动位置的示意图；

图 6 为根据本公开实施例的一种表达量数据预处理方法的示意图；

图 7 为根据本公开实施例的一种图像采集和时空测序规则的示意图；

图 8 为根据本公开实施例的一种拍照装置拍摄样例的示意图；

图 9 为根据本公开实施例的一种图像分割处理策略的示意图；

图 10 为根据本公开实施例的一种图像分割处理策略的示意图；

图 11 为根据本公开实施例的一种刚性配准算法的示意图；

图 12 为根据本公开实施例的一种图像分割和配准结果的示意图；

图 13 为根据本公开实施例的一种组织外轮廓生成流程的示意图；

图 14 为根据本公开实施例的一种图像渲染的效果示意图；

图 15 为根据本公开实施例的一种图像分割和配准效果的示意图；

图 16 为根据本公开实施例的一种表达量数据配准流程的示意图；

图 17 为根据本公开实施例的一种直方图均衡化的效果示意图；

图 18 为根据本公开实施例的一种刚性配准流程的示意图；

图 19 为根据本公开实施例的一种刚性配准结果的示意图；

图 20 为根据本公开实施例的一种弹性配准流程的示意图；

图 21 为根据本公开实施例的一种弹性配准结果的示意图；

图 22 为根据本公开实施例的一种表达量数据连续配准的示意图；

图 23 为根据本公开实施例的一种分区轮廓模型生成流程的示意图；

图 24 为根据本公开实施例的一种统领文件基本数据结构的示意图；

图 25 为根据本公开实施例的一种统领文件内单基因分层分块数据结构的示意图；

图 26 为根据本公开实施例的一种统领文件内细胞类型分层分块数据的示意图；

图 27 为根据本公开实施例的一种鼠脑重构渲染效果的示意图；

图 28 为根据本公开实施例的一种鼠脑可展示单基因列表的示意图；

图 29 为根据本公开实施例的一种生成组织外轮廓模型的效果示意图；

图 30 为根据本公开实施例的一种生成组织分区轮廓的效果示意图；

图 31 为根据本公开实施例的一种弹性配准结果的示意图；

图 32 为根据本公开实施例的一种数据采集装置的示意框图；

图 33 为根据本公开实施例的一种数据处理装置的示意框图；

图 34 为根据本公开实施例的一种电子设备的结构示意图。

具体实施方式

下面详细描述本公开的实施例，实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，旨在用于解释本公开，而不能理解为对本公开的限制。

时空组学技术自发展以来，已在神经科学、病理学、发育生物学等方向得到初步应用。基于原位捕获的方法，通常为依靠芯片和组织切片间的物理接触和探针捕获来获取原位空间信息及各种组学信息，对组织的破坏性切割以及平面捕获相应信息限制其只能用于组织的二维平面时空组学分析，难以得到三维空间的形态结构及准确的位置信息。而不同器官和组织具有独特的三维细胞分布和基因表达图谱，对组织进行三维时空组学研究具有极大的技术挑战性和重要的科学意义。

三维时空组学的关键点在于如何突破二维平面的研究，实现组织三维水平上不同细胞的空间分布、相互作用、基因蛋白质等表达情况以及整体形态结构的全面解析。

当前，基于时空组的三维重构流程主要是将相邻或相近切片做切片之间的刚性配准，然后基于聚类或注释信息，生成分区轮廓，再基于切片堆生成整个组织/生物体的轮廓。三维重构的结果通常是以文章中的图表或者代码库的 demo 图片形式进行展示。存在以下不足：

（1）仅依据时空组捕获的表达量丰度信息，将测量于平面上的表达量数据进行刚性配准，无法校正实验过程中因人员操作和生化过程等带来的表达量数据的形变，从而没有将数据重构回原位表达水平；

（2）采用基于表达量丰度信息的刚性配准的方法，需要捕获到的表达量丰度信息来自相邻或相近切片。如果切片距离较远，刚性配准的准确性会有所下降；

（3）还原后的三维空间信息需要投影到如电脑屏幕，纸质期刊文献等二维平面来可视化，在无其他辅助信息如边界线和面的情况下，二维平面内的散点可提供的三维效果有限，尤其当散点数量较大且无明显组织轮廓信息时，不利于人眼识别三维空间信息；

（4）受技术和样本等限制，单张组织切片上信息点数量较少，当样本物理尺寸和芯片分辨率提高后，下游分析和展示时受硬件条件限制，对于大数据量的组织样本数据处理和后期展示能力不足。

为此，本公开提出了一种数据采集方法、装置及系统，基于原位捕获的时空组学技术，结合原位光学影像得到原始组织轮廓，辅助内部采集到的生物信息的定位，以还原组织切割前的三维位置信息，减少实验操作对于组织形态的影响，对于芯片上每个信息采集位点实现细胞精度的位置配准，对于较大尺寸和数据量的组织样本进行有效的处理和展示。

下面结合附图对本申请所提供的传输配置方案进行详细介绍。

图 1 示出了根据本公开实施例的一种数据采集方法的流程示意图。如图 1 所示，该方法可以包括以下步骤。

S101，对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据。

具体地，对生物组织样本进行连续切片，根据实际情况选取指定位置切片进行时空组测序。

在本公开的实施例中，生物组织样本取自生物样品，生物样品通常是指动物(包括人)的体液(如尿血、唾液、胆汁、胃液、淋巴液及生物体的其他分泌液等)、毛发、肌肉和一些组织器官(如胸腺、胰腺、肝、肺、脑、胃、肾等)以及各种微生物等。

在本公开的实施例中，对生物组织样本进行切片的设备不予限制，优选地采用生物组织切片机对生物组织样本进行切片，精度和稳定性更高。

在本公开的实施例中，基于原位捕获的方法获取表达量数据，可以采用捕获测序芯片信号的方法获取表达量数据，其中芯片的性能决定了精度和分辨率等技术效果，优选地基于 DNA 纳米球测序芯片进行组织的原位捕获，芯片大、信号密度高实现更优的分辨率和视野。另外，也可以应用目前处于开发中的蛋白组学等其他基于原位捕获测序的时空组数据，本公开不予限制。

本公开的实施例中，在每次选取切片进行空组测序时，记录切片对应位置，切片的对应位置用于后续三维重构。

S102，对切片后的生物组织样本进行图像采集，以得到图像数据。

本公开的实施例中，可以采用光学成像对切片后的生物组织样本进行图像采集，也可采取其他不影响时空组学的原位成像技术手段，本公开不予限制。

本公开的实施例中，对所有选取切片进行时空组测序对应的切片后的组织进行图像采集，主要采集切片截面的图像，用于后续配准，同时也可以选取其他切片后的生物组织样本进行图像采集，具体选取规则本公开不予限制。举例而言，每切 m 次对切片后的生物组织样本进行图像采集，每切 n 次取切下的切片做时空测序，其中 m、n 为正整数，可以根据实际情况调整，n 可以为 m 的倍数。

本公开的实施例中，在每次选取切片进行图像采集时，记录切片对应位置，切片的对应位置用于后续三维重构。

S103，对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓。

可选地，在匹配之前，还包括对表达量数据进行预处理，对图像数据进行图像分割处理，其中预处理可以包括聚合处理、质量控制、对数归一化、找高变基因、降维处理中的至少一项。其中图像分割处理主要指的是对使用图像分割算法对其进行组织分割，判别删除不合格的分割样本，保留合格的分割结果作为后续配准的图像数据。

在本公开的实施例中，该方法可以包括对图像数据进行配准，以生成生物组织样本的外层轮廓。

在本公开的一种实施例中，通过对图像数据进行图像分割处理，生成分割结果图，对分割结果图进行配准，生成配准结果图，最后基于配准结果图，生成生物组织样本的外层轮廓实现。其中，选取第一张分割结果图对其进行旋转校正，校正后作为第一张参考的图对其他分割图进行配准，具体可以采用刚性配准的方式或其他配准方法，本公开不予限制。

在本公开的实施例中，该方法可以包括对表达量数据和图像数据进行配准，以生成生物组织样本的分区轮廓。

在本公开的一种实施例中，通过对表达量数据进行计算，得到表达量图，将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对，对图像对依次进行刚性配准和弹性配准，生成分区轮廓实现。

在本公开的实施例中，不使用图像数据时，例如当没有采集图像或其他原因，该方法还可以包括对表达量数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓，但是需要参与配准的时空组切片相邻或相近。

在本公开的一种实施例中，通过对两个相邻或相近切片中的表达量数据进行配准，以得到配准表达量数据，根据配准表达量数据，生成生物组织样本的外层轮廓和/或分区轮廓实现。

应当的理解的是，通过上述任一方法可以生成生物组织样本的外层轮廓和/或分区轮廓，根据实际研究或应用的需求选择，本公开不予限制。

在本公开的实施例中，基于配准后的时空组数据生成生物组织样本的轮廓数据，具体地，对表达量数据和/或图像数据进行配准得到的配准结果进行生物信息学分析，得到聚类和注释结果，基于聚类和注释结果，构造三维数组，并增加空间尺度信息，对每个聚类或分区结果进行分割，生成样本内部结构多个组织或者区域的平滑轮廓面。其中，可以采用类似 3D slicer 中的 segment editor 等工具，对每个聚类或分区结果进行分割，最终生成的轮廓图以 vtk、obj、stl 等常见的 3D 矢量模型数据存储。

S104，根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据。

在本公开的实施例中，利用三维渲染工具，渲染得到生物组织样本的三维展示数据，所采用的渲染工具本公开不与限制

在本公开的一种实施例中，外层轮廓和/或分区轮廓数据整合其他图像渲染数据、配准后时空组数据等组成项目统领文件，其中综合考虑传输效率和渲染效果，生成项目统领文件时对时空组数据按单基因表达量、细胞类型等角度分别进行数据的分层分块处理。将生成好的统领文件，通过基于 flask 的后端程序加载到渲染器，可以实现对基因表达分布、组织轮廓等信息在三维空间内的情况的直观展示。

综上，根据本公开实施例提供的传输配置方法，对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；对切片后的生物组织样本进行图像采集，以得到图像数据；对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓；根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据。本公开针对原位捕获的时空组学技术，结合光学成像信息，进行配准，得到内外组织轮廓辅助可视化信息定位，可以对大尺寸高分辨率时空组数据进行有效的处理与展示，应用于三维时空组学技术可以实现基因与影像同时分析。

图 2 示出了根据本公开实施例的一种数据采集方法的流程示意图。基于图 1 所示实施例，如图 2 所示，该方法可以包括以下步骤。为方便理解，图 3 示出了根据本公开实施例的一种数据采集方法的示意图。

在本公开中，为了便于理解方案，还详细描述了本方法的一种实施方式、过程以及效果，以获取猕猴脑数据为例，将在每个步骤的实施方式中具体展开描述，应该理解的是，该示例不是对方法的限制。

基于图 1 所示实施例，步骤 S101 包括了图 2 所示的步骤 S201。

S201，对生物组织样本进行连续切片，每切第一预定数量的切片进行一次时空组测序，以得到切片的表达量数据。

其中，第一预定数量为正整数，例如每切 10 张取一张切片做时空组测序。

本公开的实施例中，优选采用生物组织切片机对生物组织样本进行切片，精度和稳定性更高。具体地，将生物组织样本固定在切片机的冷冻头位置，根据后续拍照位置的要求，大致调整好动头的位置和角度，首先对组织切片进行修片操作以去掉多余的 OCT 包埋剂，直至露出组织区域，对准组织进行连续切片。

在本公开的一种实施方式中，以获取猕猴脑数据为例，采用切片机进行切片，如图 4 所示，首先实验人员准备猴脑组织样品，并将猴脑组织样品固定在切片机冻头的中心位置。其次将含待切猴脑组织的 OCT（冰冻切片包埋剂）包埋块固定在样本托上，调节切面至合适角度，在图 4 所示指示屏上设置为修片模式，调整修片厚度，依据客户提供的样本包埋状态信息，摇动右侧手轮进行空转粗修切面，使 OCT 表面被切平整；调焦前，需确保样本头的前进位移显示刻度指针在 15mm 左右，因此在预估包埋块可能跳出此范围时，需要回退切片机冻头，此时要按回退按钮，回退到多少刻度需要根据猴脑组织的剩余厚度情况决定，接下来需要调整刀架使其与组织面贴平，具体为：当包埋块厚度在 10mm 左右厚度时，此时不需要调刀架；当猴脑组织大于 10mm 时，比如说为 15mm 厚的包埋块，由于冻头只能前进 10mm，所以当冻头不能前进时需要将冻头回退至少 5mm，具体移动的数值可以从右图显示屏读出，此时刀架也要随之前移到与组织面大致相贴，此时第一刀作为平整度调整应去掉，从第二刀开始新的一轮实验。

本公开的实施例中，基于原位捕获的方法获取表达量数据，可以采用捕获测序芯片信号的方法获取表达量数据，其中芯片的性能决定了精度和分辨率等技术效果，优选地基于 DNA 纳米球测序芯片进行组织的原位捕获，芯片大、信号密度高实现更优的分辨率和视野。另外，也可以应用目前处于开发中的蛋白组学等其他基于原位捕获测序的时空组数据，本公开不予限制。

在本公开的一些实施例中，得到切片的表达量数据后，还包括对表达量数据进行预处理，其中，预处理包括聚合处理、质量控制、对数归一化、找高变基因、降维处理中的至少一项。

在本公开中，对数据进行预处理有利于减少表达量数据所带的噪音，减少因实验条件和操作而带来的数据规模效应，并由此得到更准确的注释结果、更能体现组织结构的表达量图、以及渲染器中更准确的展示结果。

在本公开中，如图 6 所示，还包括对数据进行注释，基于预处理的注释结果将用于分区轮廓的生成，以及渲染器中数据的展示。

应当说明的是，具体预处理和注释方法由流程的使用者自行决定，可以采用既有的处理流程，例如 binset 聚合、质量控制、对数归一化（又称为'测序深度标准化'）、找高变基因、降维、聚类的流程，也可以采用自行设计的流程。

在本公开的一种实施方式中，以获取猕猴脑数据为例，如图 6 所示，上部示出了猕猴脑切片的原始表达量数据，下部示出了预处理和注释后对应的表达量数据。

基于图 1 所示实施例，步骤 S102 包括了图 2 所示的步骤 S202。

S202，每切第二预定数量的切片，对切片后的生物组织样本进行一次图像采集，以得到图像数据。

其中，第二预定数量为正整数。

本公开的实施例中，测序和图像采集按照一定规则进行，其中第一预定数量为第二预定数量的倍数，时空组测序必然与拍照一起发生，有时拍照但不测序，根据实际情况调整。举例而言，如图 7 所示，第一预定数量为 10，第二预定数量为 5，将生物组织样本切为 13 片，第一次切片后，对切片进行时空组测序、对切片后的生物样本进行图像采集，连续切 5 片后，对切片后的生物样本进行图像采集，再连续切 5 片后，对切片进行时空组测序、对切片后的生物样本进行图像采集，直到切完为止。

本公开的实施例中，可以采用光学成像对切片后的生物组织样本进行图像采集，也可采取其他不影响时空组学的原位成像技术手段，本公开不予限制。

本公开的一种实施例中，采用拍照设备对切片后的生物组织样本进行图像采集，具体地，如图 5 所示，将拍照设备对准组织切片，系统校准，使样本截面在焦，通过滑台软件控制面板记录当前相机位置。按照计划对组织进行切片，并记录切刀流水。取第一张切片进行时空组测序，对切完剩下的组织切面进行 BlockFace 拍照，具体拍照和时空组测序规则按图 7 所示进行。

在本公开的一种实施方式中，图像采集方式采用原位光学成像，以获取猕猴脑数据为例，采用的拍照设备如图 4 所示，首先，连接拍照设备和电脑。具体地，选取合适地镜头，调整镜头上白色圆点对应的刻度值大小表示此时镜头的实际放大倍率，连接好相机、滑台的电源线和 USB 信号线，其中 USB 信号线要接到电脑的 USB3.0 接口，在电脑桌面打开相机拍照软件，选择相机的“MV-CH250-90UC”，使相机正常连接。其次，通过操作滑台软件和相机软件，以控制拍照装置进行图像采集。具体地，启动滑台软件，选择 Easy 模式，接着确认伺服，出现滑台的操作界面，设置滑台前进和后退的数值大小，以 100 为例，实际过程中一般在 80~100mm 之间，具体大小以初始焦面为准(拍照装置还没移动到切片机拍照的位置时，只设置数值，不点击“+”)。如图 5 所示，将拍照装置从初始端，沿滑轨滑动至限位端，并打开舱门的安全位置。选择相机软件的视频键，此时在相机的视频显示窗口会出现视频实时显示画面，然后点击滑台“+”（输入值为上述的设置即 100)，控制滑台会伸入到切片机内，则在显示框会出现组织截面，如果组织面不在焦面的，设置滑动单位，可在滑台软件框输入框单

位“1”，根据实际情况选择“-”和“+”操作(选择“-”则表示滑台向外伸出，选择“+”则表示滑台向内伸入)，通过这种微调，使 OCT 上的黑色 Mark 点（用于标记微调）能在相机显示框清晰的成像(Mark 点在标记前需保证 OCT 已被切片机切平)，确定 OCT 处于焦面后，在 Mark 点聚焦的基础上连续切片 OCT 直至微落出需要的组织面，此时就可以按照切片流程正式进行切片了。最后，拍照保存图像，将拍照装置复位。具体地，对于需要拍照保存的 Block Face 图片，点击相机拍照界面的“抓拍图像按键”，则此时的图像被保存。执行完成一次拍照后，点击滑台软件的“原点复位”，如图 5 所示，使拍照装置从机箱内，安全退出机箱外，并关闭舱门，这时实验人员再将拍照装置移动至初始位置，则单次拍照完成，拍照效果如图 8 所示。

图 2 所示步骤 S203-205，是对图 1 实施例中步骤 S103：对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓的进一步公开。

S203，对图像数据进行图像分割处理，以生成分割结果图。

在本公开地一种实施例中，具体地，将光学成像拍的 BlockFace 图像，使用图像分割算法对其进行组织分割，判别删除不合格的分割样本，保留合格的 BlockFace 图像分割结果进行后续处理。

在本公开的一些实施例中，对图像数据进行图像分割处理包括：将图像数据转换为 HSV 颜色空间，提取 H 通道的图像数据并转换为灰度图；通过预设图像处理算法，对图像数据进行二值化，以确定二值化结果轮廓图；保留符合预设条件的轮廓图，与图像数据进行合并。

具体的，如图 9 所示，首先对原始 BlockFace 图进行分割处理，去掉大面积的背景，将图像转换为 HSV 颜色空间格式，提取 H 通道转换为灰度图，通过 Li 算法对图像进行二值化，寻找二值化结果轮廓图，通过轮廓一阶矩，保留大于阈值的轮廓，与原图取并获得目标分割区域。

可选地，对图像数据进行图像分割处理还可以包括：获取图像数据的灰度化图像和直方图；对图像数据进行最大熵阈值二值化处理；将处理后的图像与图像数据进行合并。

在本公开中，优先采用如图 9 所示策略进行图像分割处理，对于不适合采用该策略分割的图像，或者采用该策略分割错误的图像，可选的，可以采用如图 10 所示策略进行图像分割。例如，由于组织区域的 BlockFace 图本身的多样性，如果 BlockFace 图中背景像素和待分割的组织目标像素值很接近，那么使用 Li 算法得到的阈值就不可能适用于所有情况。

具体地，如图 10 所示，首先对原始 BlockFace 图进行分割处理，去掉大面积的背景，然后对去掉大面积背景后的图进行灰度化处理，统计灰度化后图的直方图（灰度直方图是将数字图像中的所有像素），按照灰度值的大小，统计其出现的频率。基于对象和背景之间信息度量的最大化得到最大熵二值化阈值，使用它对灰度化的图像进行分割。根据分割后的 mask 图与去掉大面积背景后的图取并获取目标分割区域。

S204，对分割结果图进行配准，以生成配准结果图。

在本公开的实施例中，在进行配准之前，取第一张 BlockFace 分割结果图对其进行旋转校正，校正后作为第一张参考的图对其他分割图进行配准。

在本公开的一些实施例中，对分割结果图进行配准，以生成配准结果图包括：对分割结果图进行空间变换处理和线性插值算法处理；对经处理的分割结果图以及配准参考图进行均方差相似度测量；将测量最优解确定为配准结果图。

在本公开的一种实施例中，BlockFace 图像分割结果配准逻辑如图 11 所示，设置刚性配准算法的固定步长的梯度下降优化算法的最大迭代次数为 K 和学习率为 r，步长为 t，（K、r、t 的值根据实际情况调整），实现配准。具体地，首先对 BlockFace 图像分割结果图进行空间变换、线性插值算法，第一张 BlockFace 分割结果图作为参考图，与参考图进行均方差相似度测试，需要优化则进行固定步长 r 的下降优化算法，再次进行空间变换、线性插值算法，直到达到最大迭代次数 K，获得最优结果输出，即获得配准结果图。

对于步骤 S203 和步骤 S204，以鼠心组织的 BlockFace 图像为例，展示对鼠心组织的 BlockFace 图进行分割和对其进行配准的效果，如图 12 所示。

S205，基于配准结果图，生成生物组织样本的外层轮廓。

在本公开的一些实施例中，基于配准结果图，生成生物组织样本的外层轮廓包括：将所有配准结果图堆叠为预设格式的数组文件；对数组文件进行处理，以得到生物组织样本的外层轮廓。

在本公开的一种实施例中，如图 13 所示，对配准结果图进行 Volume 和 Mesh 的生成，具体地，将配准后的 BlockFace 分割图堆叠成为一个.npy 格式的文件。得到的 npy 数组文件，其中非 0 值的元素所占据的空间位置为样本组织，0 值的元素为空间背景，因此可以以医学图像格式 NIfTI（Neuroimaging Informatics Technology Initiative）来进行存储。通过 SimpleITK、Nibabel 等图像处理工具，辅以切片的三维空间尺度，存储为 nii 格式的文件，然后通过 3Dslicer 等相关工具，进行 segment-make surface 的操作，形成平滑的组织外部轮廓，以 vtk、obj、stl 等常见的 3D 矢量模型数据存储。

在本公开的实施例中，还包括处理配准后的光学成像结果得到渲染所需格式。

具体地，将配准后的 BlockFace 分割图堆叠成为一个.npy 格式或.tif 的文件,通过 python 的工具包 CloudVolume 和 Igneous 将 Blockface 分割图像处理成渲染器支持的带分层分块结构的二进制渲染文件，最终渲染效果如图 14 所示。

在本公开的一种实施方式中，如图 15 所示，以获取猕猴脑数据为例，首先利用 ImageJ 对 BlockFace 拍照图进行旋转校正并选择感兴趣的 ROI 分析区域，接着对 BlockFace ROI 图像进行组织目标对象和背景的分割，并将 BlockFace 分割图进行配准，形成空间中对齐的 BlockFace 图像堆，算法依据 BlockFace 分割图生成对象的最外层轮廓，依据 BlockFace 分割图生成对象的最外层轮廓。

在本公开的一些实施例中，在步骤 S203-205 生成生物组织样本的外层轮廓之后，还包括：对表达量数据进行计算，得到表达量图；将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；对图像对依次进行刚性配准和弹性配准，以生成分区轮廓。

具体地，如图 16 所示，将时空组的表达量数据处理成为反映不同区域表达量丰度的表达量图，分别对表达量图和 BlockFace 分割图做直方图均衡化处理，直方图均衡化处理的效果如图 17 所示。然后在切片位置相互近邻的 BlockFace 分割图-表达量图图像对中，做如下操作：以 BlockFace 分割图为参考图像，使用表达量图与之进行刚性配准和弹性配准。完成两种配准后，将配准中决定的变换关系（刚性配准的变换关系形如一个 3*3 的变换矩阵，弹性配准的变换关系形如两个形变场，分别表示 x 和 y 方向的形变的分布）复用在表达量数据上，实现表达量数据与 BlockFace 分割图的配准。

可选地，对图像对进行刚性配准包括：对图像对分别进行下采样；提取图像对的特征信息；对图像对进行相似性估计仿射变换，得到刚性配准结果。

在本公开的一种实施例中，如图 18 所示，图像对分别进行下采样，并进行噪声过滤去除伪影，分别通过深度学习模型提取两种不同模态的图像中的目标组织的边缘特征和形状特征和区域分割信息，然后在模型输出处最大化目标组织最外层轮廓的相似性估计仿射变换，最后将仿射变换应用在待配准的表达量数据原始图像上，将处理完的一对表达量数据和 BlockFace 分割图结果送至弹性配准环节处理。其中刚性配准的结果图如图 19 所示。

可选地，对图像对进行弹性配准包括：对刚性配准结果进行直方图均衡化；将刚性配准结果输入深度学习模型，输出模型的中间结果，得到形变场，其中，深度学习模型是通过表达量图-图像数据对进行训练得到的；根据形变场，对刚性配准结果进行最近邻插值变换，得到弹性配准结果。

在本公开的一种实施例中，弹性配准的实现基于的是深度学习模型的推导，如 20 图。对应的深度学习模型要经过如下过程获得：首先将表达量数据处理成为表达量图后，分别对表达量图和对应的 BlockFace 分割图进行直方图均衡化。然后在软件中对表达量图-BlockFace 分割图进行刚性配准，作为模型训练的输入图像，再进行弹性配准，作为模型训练的标签。再基于手动弹性配准时软件输出的控制点坐标，加入随机参数，生成随机变形的表达量图，实现数据合成。最后，使用合成的和真实的数据训练有监督模型，在训练指标达到要求后停止，保存模型。获得深度学习模型之后，就可以使用此模型进行弹性形变的推导。将实际问题中的表达量数据和 BlockFace 分割图按照上述的处理方法，获得直方图均衡化的图像。将两个图像输入模型，输出模型的中间结果：形变场。根据形变场，获取对应位移，对表达量数据进行最近邻插值变换，计算得到新的表达量数据。图 21 所示，以猴脑为例，展示弹性配准前后，表达量图和 BlockFace 图的对应情况变换：其中红色的为 BlockFace 图像，绿色的为表达量图像。弹性配准后，可视的表达量图的形变得到了校正。

在另一种可选的方式中，也可以使用表达量数据直接进行配准生成生物组织样本的外层轮廓和/或分区轮廓，方法包括：对两个相邻或相近切片中的表达量数据进行配准，以得到配准表达量数据；根据配准表达量数据，生成生物组织样本的外层轮廓和/或分区轮廓。

应当的理解的是，在没有进行图像采集的情况下，比如出于用户的习惯不进行拍照、没有拍照设备等情况下，执行该方法也可以生成生物组织样本的外层轮廓和/或分区轮廓，但是需要参与配准的时空组切片相邻或相近。

在本公开的一种实施例中，如图 22 所示，依据两张相邻或相近切片中的表达量数据计算 bin set 或 cell bin，经过对数归一化等预处理后，解最优传输问题，得到最优传输的转移概率矩阵（简称 pi 矩阵）；对 pi 矩阵进行后处理后，可以计算刚性，或刚性+弹性的变换关系；最后将此变换关系应用到原始表达量数据上。

S206，对表达量数据和/或图像数据进行配准得到的配准结果进行生物信息学分析，得到聚类和注释结果。

S207，基于聚类和注释结果，构造三维数组，并增加空间尺度信息。

在本公开中，基于表达量数据和/或图像数据进行配准得到的配准结果生成外层轮廓和/或分区轮廓模型。

在本公开的一种实施例中，根据基于图像数据进行配准得到的配准结果生成外层轮廓，具体地，如图 13 所示，参照步骤 S205。

在本公开的一种实施例中，根据表达量数据和/或图像数据进行配准得到的配准结果生成分区轮廓，具体与步骤 S205 相似，如图 23 所示，对配准好的表达量数据或，每一个 bin set 或者 cell bin 在通过常规的生物信息学分析，得到聚类和注释结果，以每个 bin set 或 cell bin 的注释或者聚类结果的值为单位，构造三维 npy 数组，同时增加空间尺度信息，得到基于注释聚类结果的 NIfTI 文件。再运用类似 3D slicer 中的 segment editor 等工具，对每个聚类或分区结果进行分割，生成样本内部结构多个组织或者区域的平滑轮廓面，以 vtk、obj、stl 等常见的 3D 矢量模型数据存储。对应于上述步骤 S205。

在本公开中，生成后的外层轮廓和/或分区轮廓模型，通过 python 的工具包 multiresolution-mesh-creator，将文件处理成与分区标签绑定后的二进制 3D 模型文件。

在本公开中，整合上述图像渲染数据、配准后时空组数据、组织轮廓数据、分区轮廓数据等成项目统领文件，文件内容结构根据研究或应用需求决定。

在本公开的一种实施例中，将配准后的时空组表达量数据通过 geftools 生成 cgef 格式文件，再与前述步骤中生成的二进制图像数据、二进制分区轮廓模型数据、组织外轮廓模型数据等通过 gef3d 生成统领文件，文件内容结构如图 24 所示，主要展示了统领文件的前三级目录。

在本公开的一些实施例中，综合考虑传输效率和渲染效果，可以在生成项目统领文件时对时空组数据按单基因表达量、细胞类型等角度分别进行数据的分层分块处理。

其中，分层分块是指对大量数据点云按照空间位置进行类似金字塔结构的抽样，并保证所有层的所有块内数据点并集为全部数据。该处理在 gef3d 生成统领文件时自动进行，需指定每块的 bin/cellbin 个数的阈值。

在本公开的一种实施例中，分层分块方法可以包括：首先遍历所有 cgef 获取到整体的空间大小，设定一个最大层数，根据数据点三维空间各自的长度，计算好划分尺度，从上到下按照 8 分法预先划分出每层的块数，每块近似立方体。默认第一层只有一块。经过第一步的遍历可以获取到每个 gene 在各个切片中的数据，也就是在整个空间的数据点的属性值，包括坐标及表达量丰度等。对单个 gene 的数据进行分层分块，根据第一步计算的结果，从顶层到下层，依次从 gene 的整体数据中去填充每一层的每一小块，直至小块填满（由阈值控制）。剩下的部分继续填充下一层，当下一层某一块数据太少（阈值控制），会将本块数据添加到上一层的父块中去。依次向下分配，直至 gene 的数据分配完。对于 celltype 等细胞类型注释结果的分层分块与单基因类似。

分层分块后的表达量或细胞类型等数据在统领文件中存储结构如图 25、图 26 所示，在 L0 目录下，其中，blk 表示块，为 block 的简写，该目录用于储存分块数据情况，在 blk 目录下，offset 表示偏移量，储存该分块数据的起始索引偏移量，count 表示数目，储存该分块数据的点数量；cell 表示细胞，该目录用于存储该层级下所有 bin/cellbin 数据相关信息，在 cell 目录下，x、y、z 分别储存该 bin/cellbin 空间坐标的 x、y、z 坐标值，umicnt 储存该 bin/cellbin 表达量，cellid 表示细胞序号，储存该 bin/cellbin 序号。在 attrs 目录下，其中，ctype_list 表示细胞类型列表，储存细胞类型序号、名称和颜色的列表；maxlevelnum 表示最大分层数，所有细胞类型数据中分层数的最大值；blkcnt 表示分块数，储存不同层级的各轴向分块数量；maxumi 表示表达量最大值，储存表达量最大值。

S208，利用三维渲染工具，渲染得到生物组织样本的三维展示数据。

具体地，将经上述步骤生成好的统领文件，通过基于 flask 的后端程序加载到渲染器，可以实现对基因表达分布、组织轮廓等信息在三维空间内的情况的直观展示。在网页端进行展示，效果如图 27 所示，其中分别加载了整体表达量数据，组织外轮廓模型。

综上，根据本公开实施例提供的数据采集方法，针对原位捕获的时空组学技术，结合光学成像信息，辅助连续组织切片的刚性、弹性配准，实现三维重构，进而得到内外组织轮廓辅助可视化信息定位，对大尺寸高分辨率时空组数据进行有效处理与展示。

本公开的方法效果展示如下：

1. 通过步骤 S204 和步骤 S205 对采集的样本图像数据进行图像分割处理方法和图像配准，以鼠心组织的 BlockFace 图像为例，展示对鼠心组织的 BlockFace 图进行分割和对其进行配准的效果，如图 12 所示。
2. 参照上述图 18 对应方法，以猴脑为例，展示弹性配准前后，表达量图和采集图像的对应情况变换，效果如图 19 所示，弹性配准后，可视的表达量图的形变得到了校正。
3. 生成组织外轮廓模型，与转录组等空间散点信息结合进行展示，如图 28 所示，从左到右依次为，鼠脑组织轮廓及细胞类型展示、鼠脑组织轮廓及单基因表达量展示以及鼠脑组织轮廓模型及基因总表达量展示。
4. 生成组织分区轮廓模型，如图 29 所示，左边为鼠脑组织分区轮廓模型展示，右边为鼠脑组织分区轮廓模型及对应的细胞类型 bin 展示。
5. 对于较大尺寸和数据量的组织样本的处理和展示如图 30 所示，选取的展示例为鼠脑的 100 张以上时空组芯片数据。芯片大小为 1cm x 1cm，所展示的数据约 900K 个坐标点，包括约 22K 个基因。

因此，本公开的方法可以更好地还原组织样本切片前的组织轮廓以及相关表达量数据，应用于时空组学三维重构，将有利于推动三维时空组学技术的研究和应用。

图 31 为根据本公开实施例的一种数据处理方法的流程示意图。如图 31 所示，该方法可以包括以下步骤。

S301，获取生物组织样本的表达量数据和图像数据。

其中，表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，图像数据是对切片后的生物组织样本进行图像采集得到的。

在本公开的实施例中，对生物组织样本进行连续切片，根据实际情况选取指定位置切片进行时空组测，获取表达量数据。

在本公开的实施例中，可以采用光学成像对切片后的生物组织样本进行图像采集，获取图像数据，也可采取其他不影响时空组学的原位成像技术手段，本公开不予限制。

在本公开的实施例中，生物组织样本取自生物样品，生物样品通常是指动物(包括人)的体液(如尿血、唾液、胆汁、胃液、淋巴液及生物体的其他分泌液等)、毛发、肌肉和一些组织器官(如胸腺、胰腺、肝、肺、脑、胃、肾等)以及各种微生物等。

在本公开的实施例中，对生物组织样本进行切片的设备不予限制，优选地采用生物组织切片机对生物组织样本进行切片，精度和稳定性更高。

在本公开的实施例中，基于原位捕获的方法获取表达量数据，可以采用捕获测序芯片信号的方法获取表达量数据，其中芯片的性能决定了精度和分辨率等技术效果，优选地基于 DNA 纳米球测序芯片进行组织的原位捕获，芯片大、信号密度高实现更优的分辨率和视野。另外，也可以应用目前处于开发中的蛋白组学等其他基于原位捕获测序的时空组数据，本公开不予限制。

本公开的实施例中，对所有选取切片进行时空组测序对应的切片后的组织进行图像采集，主要采集切片截面的图像，用于后续配准，同时也可以选取其他切片后的生物组织样本进行图像采集，具体选取规则本公开不予限制。

本公开的实施例中，在每次选取切片进行时空组测序或图像采集时，记录切片对应位置，切片的对应位置用于后续三维重构。

在本公开中，获取表达量数据后，还可以包括对表达量数据进行预处理，其中，预处理包括聚合处理、质量控制、对数归一化、找高变基因、降维处理中的至少一项。对数据进行预处理有利于减少表达量数据所带的噪音，减少因实验条件和操作而带来的数据规模效应，并由此得到更准确的注释结果、更能体现组织结构的表达量图、以及渲染器中更准确的展示结果。

在本公开中，获取图像数据后，还可以包括对图像数据进行分割处理，具体地，将图像数据转换为 HSV 颜色空间，提取 H 通道的图像数据并转换为灰度图，通过预设图像处理算法，对图像数据进行二值化，以确定二值化结果轮廓图，保留符合预设条件的轮廓图，与图像数据进行合并。

可以参照步骤 S201 和步骤 202 的具体实施方式，本公开中不再赘述。

S302，对表达量数据进行计算，得到表达量图。

本公开的实施例中，根据实际研究需求，对时空组的表达量数据的相关系数进行计算，得到可以反映不同区域表达量丰度的表达量图，具体的计算、统计及画图方式本公开不予限制。

S303，将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对。

其中，直方图均衡化是一种简单有效的图像增强技术，通过改变图像的直方图来改变图像中各像素的灰度，主要用于增强动态范围偏小的图像的对比度。原始图像由于其灰度分布可能集中在较窄的区间，造成图像不够清晰。例如，过曝光图像的灰度级集中在高亮度范围内，而曝光不足将使图像灰度级集中在低亮度范围内。采用直方图均衡化，可以把原始图像的直方图变换为均匀分布（均衡）的形式，这样就增加了像素之间灰度值差别的动态范围，从而达到增强图像整体对比度的效果。

在本公开的一种实施例中，经过步骤 S301 采集 BlockFace 图像作为图像数据，并对图像进行分割处理获取 BlockFace 分割图，取切片位置相互近邻的 BlockFace 分割图-表达量图像对中作为表达量图-图像数据的图像对。

S304，对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据。

在本公开的一种实施例中，如图 16 所示，以 BlockFace 分割图为参考图像，使用表达量图与之进行刚性配准和弹性配准。完成两种配准后，将配准中决定的变换关系（刚性配准的变换关系形如一个 3*3 的变换矩阵，弹性配准的变换关系形如两个形变场，分别表示 x 和 y 方向的形变的分布）复用在表达量数据上，实现表达量数据与 BlockFace 分割图的配准。

在本公开的可选实施例中，对图像对进行刚性配准包括：对图像对分别进行下采样；提取图像对的特征信息；对图像对进行相似性估计仿射变换，得到刚性配准结果。

在本公开的一种实施例中，如图 18 所示，图像对分别进行下采样，并进行噪声过滤去除伪影，分别通过深度学习模型提取两种不同模态的图像中的目标组织的边缘特征和形状特征和区域分割信息，然后在模型输出处最大化目标组织最外层轮廓的相似性估计仿射变换，最后将仿射变换应用在待配准的表达量数据原始图像上，将处理完的一对表达量数据和 BlockFace 分割图结果送至弹性配准环节处理。其中刚性配准的结果图如图 19 所示。

在本公开的可选实施例中，对图像对进行弹性配准包括：对刚性配准结果进行直方图均衡化；将刚性配准结果输入深度学习模型，输出深度学习模型的中间结果，得到形变场，其中，深度学习模型是通过表达量图-图像数据对进行训练得到的；根据形变场，对刚性配准结果进行最近邻插值变换，得到弹性配准结果。

在本公开的一种实施例中，弹性配准的实现基于的是深度学习模型的推导，如 20 图。对应的深度学习模型要经过如下过程获得：首先将表达量数据处理成为表达量图后，分别对表达量图和对应的 BlockFace 分割图进行直方图均衡化。然后在软件中对表达量图-BlockFace 分割图进行刚性配准，作为模型训练的输入图像，再进行弹性配准，作为模型训练的标签。再基于手动弹性配准时软件输出的控制点坐标，加入随机参数，生成随机变形的表达量图，实现数据合成。最后，使用合成的和真实的数据训练有监督模型，在训练指标达到要求后停止，保存模型。获得深度学习模型之后，就可以使用此模型进行弹性形变的推导。将实际问题中的表达量数据和 BlockFace 分割图按照上述的处理方法，获得直方图均衡化的图像。将两个图像输入模型，输出模型的中间结果：形变场。根据形变场，获取对应位移，对表达量数据进行最近邻插值，计算得到新的表达量数据。图 21 所示，以猴脑为例，展示弹性配准前后，表达量图和 BlockFace 图的对应情况变换：其中红色的为 BlockFace 图像，绿色的为表达量图像。弹性配准后，可视的表达量图的形变得到了校正。

在本公开中，基于配准后的时空组数据生成生物组织样本的轮廓数据，具体地，基于已经配准好的表达量时空数据，每一个 bin set 或者 cell bin 在通过常规的生物信息学分析后，得到聚类和注释的结果。以每个 bin set 或 cell bin 的注释或者聚类结果的值为单位，构造三维 npy 数组，同时增加空间尺度信息，得到基于注释聚类结果的 NIfTI 文件。再运用类似 3D slicer 中的 segment editor 等工具，对每个聚类或分区结果进行分割，生成样本内部结构多个组织或者区域的平滑轮廓面，以 vtk、obj、stl 等常见的 3D 矢量模型数据存储。

在本公开中，轮廓数据用于生成生物组织样本的三维展示数据，具体地，轮廓数据整合其他图像渲染数据、配准后时空组数据等组成项目统领文件，其中综合考虑传输效率和渲染效果，生成项目统领文件时对时空组数据按单基因表达量、细胞类型等角度分别进行数据的分层分块处理。将生成好的统领文件，通过基于 flask 的后端程序加载到渲染器，可以实现对基因表达分布、组织轮廓等信息在三维空间内的情况的直观展示。

可以参照基于图 2 中步骤 S203-208 对应内容的具体实施方式，本公开中不再赘述。

综上，根据本公开实施例提供的数据处理方法，获取生物组织样本的表达量数据和图像数据，其中表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，图像数据是对切片后的生物组织样本进行图像采集得到的；对表达量数据进行计算，得到表达量图；将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据，提供了时空测序的表达

量数据与组织切面图像数据的配准方法，应用本方法可以得到内外组织轮廓辅助可视化信息定位，对大尺寸高分辨率时空组数据进行有效处理与展示，推动三维时空组学技术的研究与应用。

为了实现上述本申请实施例提供的方法中的各功能，网络设备和终端设备可以包括硬件结构、软件模块，以硬件结构、软件模块、或硬件结构加软件模块的形式来实现上述各功能。上述各功能中的某个功能可以以硬件结构、软件模块、或者硬件结构加软件模块的方式来执行。

与上述几种实施例提供的数据采集方法相对应，本公开还提供一种数据采集装置，由于本公开实施例提供的数据采集装置与上述几种实施例提供的数据采集方法相对应，因此数据采集方法的实施方式也适用于本实施例提供的数据采集装置，在本实施例中不再详细描述。

图 32 为本公开实施例提供的一种数据采集装置 400 的结构示意图，如图 32 所示，该装置 400 可以包括：测序模块 410，用于对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；采集模块 420，用于对切片后的生物组织样本进行图像采集，以得到图像数据；配准模块 430，用于对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓；生成模块 440，用于根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据。

根据本公开实施例提供的数据采集装置，对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；对切片后的生物组织样本进行图像采集，以得到图像数据；对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓；根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据，针对原位捕获的时空组学技术，结合光学成像信息，配准生成内外组织轮廓辅助可视化信息定位，可以对大尺寸高分辨率时空组数据进行有效的处理与展示。

在一些实施例中，所述传输配置信息还包括推荐 OAM 模态数量和/或推荐 OAM 模态集合，所述推荐 OAM 模态数量和/或所述推荐 OAM 模态集合为所述第二设备基于所述推荐 PMI 得到。

在一些实施例中，测序模块 410 具体用于：对生物组织样本进行连续切片，每切第一预定数量的切片进行一次时空组测序，以得到切片的表达量数据，其中，第一预定数量为正整数。

在一些实施例中，还包括预处理模块，用于对表达量数据进行预处理，其中，预处理包括聚合处理、质量控制、对数归一化、找高变基因、降维处理中的至少一项。

在一些实施例中，采集模块 420 具体用于：每切第二预定数量的切片，对切片后的生物组织样本进行一次图像采集，以得到图像数据，其中，第二预定数量为正整数。

在一些实施例中，配准模块 430 具体用于：对图像数据进行图像分割处理，以生成分割结果图；对分割结果图进行配准，以生成配准结果图；基于配准结果图，生成生物组织样本的外层轮廓。

在一些实施例中，对图像数据进行图像分割处理包括：将图像数据转换为 HSV 颜色空间，提取 H 通道的图像数据并转换为灰度图；通过预设图像处理算法，对图像数据进行二值化，以确定二值化结果轮廓图；保留符合预设条件的轮廓图，与图像数据进行合并。

在一些实施例中，还包括：获取图像数据的灰度化图像和直方图；对图像数据进行最大熵阈值二值化处理；将处理后的图像与图像数据进行合并。

在一些实施例中，对分割结果图进行配准，以生成配准结果图包括：对分割结果图进行空间变换处理和线性插值算法处理；对经处理的分割结果图以及配准参考图进行均方差相似度测量；将测量最优解确定为配准结果图。

在一些实施例中，基于配准结果图，生成生物组织样本的外层轮廓包括：将所有配准结果图堆叠为预设格式的数组文件；对数组文件进行处理，以得到生物组织样本的外层轮廓。

在一些实施例中，配准模块430还用于：对表达量数据进行计算，得到表达量图；将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；对图像对依次进行刚性配准和弹性配准，以生成分区轮廓。

在一些实施例中，对图像对进行刚性配准包括：对图像对分别进行下采样；提取图像对的特征信息；对图像对进行相似性估计仿射变换，得到刚性配准结果。

在一些实施例中，对图像对进行弹性配准包括：对刚性配准结果进行直方图均衡化；将刚性配准结果输入深度学习模型，输出深度学习模型的中间结果，得到形变场，其中，深度学习模型是通过表达量图-图像数据对进行训练得到的；根据形变场，对刚性配准结果进行最近邻插值变换，得到弹性配准结果。

在一些实施例中，配准模块430还用于：对两个相邻或相近切片中的表达量数据进行配准，以得到配准表达量数据；根据配准表达量数据，生成生物组织样本的外层轮廓和/或分区轮廓。

在一些实施例中，生成模块440具体用于：对表达量数据和/或图像数据进行配准得到的配准结果进行生物信息学分析，得到聚类和注释结果；基于聚类和注释结果，构造三维数组，并增加空间尺度信息；利用三维渲染工具，渲染得到生物组织样本的三维展示数据。

综上，根据本公开实施例提供的数据采集方法，对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；对切片后的生物组织样本进行图像采集，以得到图像数据；对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓；根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据，针对原位捕获的时空组学技术，结合光学成像信息，辅助连续组织切片的刚性、弹性配准，实现三维重构，进而得到内外组织轮廓辅助可视化信息定位，对大尺寸高分辨率时空组数据进行有效处理与展示，应用于三维时空组学技术可以实现基因与影像同时分析。

与上述几种实施例提供的数据处理方法相对应，本公开还提供一种数据处理装置，由于本公开实施例提供的数据处理装置与上述几种实施例提供的数据处理方法相对应，因此数据处理方法的实施方式也适用于本实施例提供的数据处理装置，在本实施例中不再详细描述。

图33为本公开实施例提供的一种数据处理装置500的结构示意图。如图33所示，该装置500可以包括：获取模块510，用于获取生物组织样本的表达量数据和图像数据，其中表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，图像数据是对切片后的生物组织样本进行图像采集得到的；计算模块520，用于对表达量数据进行计算，得到表达量图；均衡化模块530，用于将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；配准模块540，对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据。

根据本公开实施例提供的数据处理装置，获取生物组织样本的表达量数据和图像数据，其中表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，图像数据是对切片后的生物组织样本进行图像采集得到的；对表达量数据进行计算，得到表达量图；将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据，实现时空测序的表达量数据与

组织切面图像数据的配准，可以得到内外组织轮廓辅助可视化信息定位，对大尺寸高分辨率时空组数据进行有效处理与展示，推动三维时空组学技术的研究与应用。

在一些实施例中，配准模块 540 用于对图像对进行刚性配准包括：对图像对分别进行下采样；提取图像对的特征信息；对图像对进行相似性估计仿射变换，得到刚性配准结果。

在一些实施例中，配准模块 540 用于对图像对进行弹性配准包括：对刚性配准结果进行直方图均衡化；将刚性配准结果输入深度学习模型，输出深度学习模型的中间结果，得到形变场，其中，深度学习模型是通过表达量图-图像数据对进行训练得到的；根据形变场，对刚性配准结果进行最近邻插值变换，得到弹性配准结果。

综上，根据本公开实施例提供的数据处理方法，获取生物组织样本的表达量数据和图像数据，其中表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，图像数据是对切片后的生物组织样本进行图像采集得到的；对表达量数据进行计算，得到表达量图；将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据，提供了时空测序的表达量数据与组织切面图像数据的配准方法，应用本方法可以得到内外组织轮廓辅助可视化信息定位，对大尺寸高分辨率时空组数据进行有效处理与展示，推动三维时空组学技术的研究与应用。

上述本申请提供的实施例中，对本申请实施例提供的方法及装置进行了介绍。为了实现上述本申请实施例提供的方法中的各功能，电子设备可以包括硬件结构、软件模块，以硬件结构、软件模块、或硬件结构加软件模块的形式来实现上述各功能。上述各功能中的某个功能可以以硬件结构、软件模块、或者硬件结构加软件模块的方式来执行。

图 34 是根据一示例性实施例示出的一种用于实现上述数据采集或数据处理方法的电子设备 600 的框图。例如，电子设备 600 可以是移动电话，计算机，消息收发设备，游戏控制台，平板设备，医疗设备，健身设备，个人数字助理等。

参照图 34，电子设备 600 可以包括以下一个或多个组件：处理组件 602，存储器 604，电源组件 606，多媒体组件 606，音频组件 610，输入/输出（I/O）的接口 612，传感器组件 614，以及通信组件 616。

处理组件 602 通常控制电子设备 600 的整体操作，诸如与显示，电话呼叫，数据通信，相机操作和记录操作相关联的操作。处理组件 602 可以包括一个或多个处理器 620 来执行指令，以完成上述的方法的全部或部分步骤。此外，处理组件 602 可以包括一个或多个模块，便于处理组件 602 和其他组件之间的交互。例如，处理组件 602 可以包括多媒体模块，以方便多媒体组件 608 和处理组件 602 之间的交互。

存储器 604 被配置为存储各种类型的数据以支持在电子设备 600 的操作。这些数据的示例包括用于在电子设备 600 上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。存储器 604 可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器（SRAM），电可擦除可编程只读存储器（EEPROM），可擦除可编程只读存储器

（EPROM），可编程只读存储器（PROM），只读存储器（ROM），磁存储器，快闪存储器，磁盘或光盘。

电源组件 606 为电子设备 600 的各种组件提供电力。电源组件 606 可以包括电源管理系统，一个或多个电源，及其他与为电子设备 600 生成、管理和分配电力相关联的组件。

多媒体组件 606 包括在电子设备 600 和用户之间的提供一个输出接口的屏幕。在一些实施例中，屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。触摸传感器可以不仅感测触摸或滑动动作的边界，而且还检测与触摸或滑动操作相关的持续时间和压力。在一些实施例中，多媒体组件 608 包括一个前置摄像头和/或后置摄像头。当电子设备 600 处于操作模式，如拍摄模式或视频模式时，前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

音频组件 610 被配置为输出和/或输入音频信号。例如，音频组件 610 包括一个麦克风（MIC），当电子设备 600 处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器 604 或经由通信组件 616 发送。在一些实施例中，音频组件 610 还包括一个扬声器，用于输出音频信号。

I/O 接口 612 为处理组件 602 和外围接口模块之间提供接口，上述外围接口模块可以是键盘，点击轮，按钮等。这些按钮可包括但不限于：主页按钮、音量按钮、启动按钮和锁定按钮。

传感器组件 614 包括一个或多个传感器，用于为电子设备 600 提供各个方面的状态评估。例如，传感器组件 614 可以检测到电子设备 600 的打开/关闭状态，组件的相对定位，例如组件为电子设备 600 的显示器和小键盘，传感器组件 614 还可以检测电子设备 600 或电子设备 600 一个组件的位置改变，用户与电子设备 600 接触的存在或不存在，电子设备 600 方位或加速/减速和电子设备 600 的温度变化。传感器组件 614 可以包括接近传感器，被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件 614 还可以包括光传感器，如 CMOS 或 CCD 图像传感器，用于在成像应用中使用。在一些实施例中，该传感器组件 614 还可以包括加速度传感器，陀螺仪传感器，磁传感器，压力传感器或温度传感器。

通信组件 616 被配置为便于电子设备 600 和其他设备之间有线或无线方式的通信。电子设备 600 可以接入基于通信标准的无线网络，如 WiFi，2G 或 3G，4G LTE、5G NR（New Radio）或它们的组合。在一个示例性实施例中，通信组件 616 经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，通信组件 616 还包括近场通信（NFC）模块，以促进短程通信。例如，在 NFC 模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

在示例性实施例中，电子设备 600 可以被一个或多个应用专用集成电路（ASIC）、数字信号处理器（DSP）、数字信号处理设备（DSPD）、可编程逻辑器件（PLD）、现场可编程门阵列（FPGA）、控制器、微控制器、微处理器或其他电子元件实现，用于执行上述方法。

在示例性实施例中，还提供了一种包括指令的非临时性计算机可读存储介质，例如包括指令的存储器 604，上述指令可由电子设备 600 的处理器 620 执行以完成上述方法。例如，非临时性计算机可读存储介质可以是 ROM、随机存取存储器（RAM）、CD-ROM、磁带、软盘和光数据存储设备等。

本公开的实施例还提出了一种存储有计算机指令的非瞬时计算机可读存储介质，其中，计算机指令用于使计算机执行本公开上述实施例中描述的数据采集或数据处理方法。

本公开的实施例还提出一种计算机程序产品，包括计算机程序，计算机程序在被处理器执行本公开上述实施例中描述的数据采集或数据处理方法。

本公开的实施例还提出了一种芯片，该芯片包括一个或多个接口电路和一个或多个处理器；接口电路用于从电子设备的存储器接收信号，并向处理器发送信号，信号包括存储器中存储的计算机指令，当处理器执行计算机指令时，使得电子设备执行本公开上述实施例中描述的数据采集或数据处理方法。

需要说明的是，本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

在本说明书的描述中，参考术语“一个实施方式”、“一些实施方式”、“示意性实施方式”、“示例”、“具体示例”或“一些示例”等的描述意指结合实施方式或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施方式或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施方式或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施方式或示例中以合适的方式结合。

流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本发明的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本发明的实施例所属技术领域的技术人员所理解。

在流程图中表示或在此以其他方式描述的逻辑和/或步骤，例如，可以被认为是用于实现逻辑功能的可执行指令的定序列表，可以具体实现在任何计算机可读介质中，以供指令执行系统、装置或设备(如基于计算机的系统、包括处理模块的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用，或结合这些指令执行系统、装置或设备而使用。就本说明书而言，"计算机可读介质"可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例(非穷尽性列表)包括以下：具有一个或多个布线的电连接部(控制方法)，便携式计算机盘盒(磁装置)，随机存取存储器(RAM)，只读存储器(ROM)，可擦除可编辑只读存储器(EPROM 或闪速存储器)，光纤装置，以及便携式光盘只读存储器(CDROM)。另外，计算机可读介质甚至可以是可在其上打印程序的纸或其他合适的介质，因为可以例如通过对纸或其他介质进行光学扫描，接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得程序，然后将其存储在计算机存储器中。

应当理解，本发明的实施方式的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列(PGA)，现场可编程门阵列(FPGA)等。

本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成，程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

此外，在本发明的各个实施例中的各功能单元可以集成在一个处理模块中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，也可以存储在一个计算机可读取存储介质中。上述提到的存储介质可以是只读存储器，磁盘或光盘等。

尽管上面已经示出和描述了本发明的实施方式，可以理解的是，上述实施方式是示例性的，不能理解为对本发明的限制，本领域的普通技术人员在本发明的范围内可以对上述实施实施进行变化、修改、替换和变型。

权利要求书

1、一种数据采集方法，其特征在于，所述方法包括：

对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；

对切片后的生物组织样本进行图像采集，以得到图像数据；

对所述表达量数据和/或所述图像数据进行配准，以生成所述生物组织样本的外层轮廓和/或分区轮廓；

根据所述外层轮廓和/或所述分区轮廓，生成所述生物组织样本的三维展示数据。

2、根据权利要求 1 所述的方法，其特征在于，所述对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据包括：

对生物组织样本进行连续切片，每切第一预定数量的切片进行一次时空组测序，以得到所述切片的表达量数据，

其中，所述第一预定数量为正整数。

3、根据权利要求 2 所述的方法，其特征在于，所述方法还包括：

对所述表达量数据进行预处理，

其中，所述预处理包括聚合处理、质量控制、对数归一化、找高变基因、降维处理中的至少一项。

4、根据权利要求 1 至 3 中任一项所述的方法，其特征在于，所述对切片后的生物组织样本进行图像采集，以得到图像数据包括：

每切第二预定数量的切片，对切片后的生物组织样本进行一次图像采集，以得到所述图像数据，

其中，所述第二预定数量为正整数。

5、根据权利要求 1 至 4 中任一项所述的方法，其特征在于，所述对所述表达量数据和/或所述图像数据进行配准，以生成所述生物组织样本的外层轮廓和/或分区轮廓包括：

对所述图像数据进行图像分割处理，以生成分割结果图；

对所述分割结果图进行配准，以生成配准结果图；

基于所述配准结果图，生成所述生物组织样本的外层轮廓。

6、根据权利要求 5 所述的方法，其特征在于，所述对所述图像数据进行图像分割处理包括：

将所述图像数据转换为 HSV 颜色空间，提取 H 通道的图像数据并转换为灰度图；

通过预设图像处理算法，对所述图像数据进行二值化，以确定二值化结果轮廓图；

保留符合预设条件的轮廓图，与所述图像数据进行合并。

7、根据权利要求 6 所述的方法，其特征在于，所述方法还包括：

获取所述图像数据的灰度化图像和直方图；

对所述图像数据进行最大熵阈值二值化处理；

将处理后的图像与所述图像数据进行合并。

8、根据权利要求 5 至 7 中任一项所述的方法，其特征在于，所述对所述分割结果图进行配准，以生成配准结果图包括：

对所述分割结果图进行空间变换处理和线性插值算法处理；

对经处理的分割结果图以及配准参考图进行均方差相似度测量；

将测量最优解确定为配准结果图。

9、根据权利要求 8 所述的方法，其特征在于，所述基于所述配准结果图，生成所述生物组织样本的外层轮廓包括：

将所有配准结果图堆叠为预设格式的数组文件；

对所述数组文件进行处理，以得到所述生物组织样本的外层轮廓。

10、根据权利要求 5 至 9 中任一项所述的方法，其特征在于，所述对所述表达量数据和/或所述图像数据进行配准，以生成所述生物组织样本的外层轮廓和/或分区轮廓包括：

对所述表达量数据进行计算，得到表达量图；

将所述表达量图和所述图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；

对所述图像对依次进行刚性配准和弹性配准，以生成所述分区轮廓。

11、根据权利要求 10 所述的方法，其特征在于，所述对所述图像对进行刚性配准包括：

对所述图像对分别进行下采样；

提取所述图像对的特征信息；

对所述图像对进行相似性估计仿射变换，得到刚性配准结果。

12、根据权利要求 11 所述的方法，其特征在于，所述对所述图像对进行弹性配准包括：

对所述刚性配准结果进行直方图均衡化；

将所述刚性配准结果输入深度学习模型，输出所述深度学习模型的中间结果，得到形变场，其中，所述深度学习模型是通过表达量图-图像数据对进行训练得到的；

根据所述形变场，对所述刚性配准结果进行最近邻插值变换，得到弹性配准结果。

13、根据权利要求 5 至 12 中任一项所述的方法，其特征在于，所述对所述表达量数据和/或所述图像数据进行配准，以生成所述生物组织样本的外层轮廓和/或分区轮廓包括：

对两个相邻或相近切片中的表达量数据进行配准，以得到配准表达量数据；

根据所述配准表达量数据，生成所述生物组织样本的外层轮廓和/或分区轮廓。

14、根据权利要求 1 至 13 中任一项所述的方法，其特征在于，所述根据所述外层轮廓和/或所述分区轮廓，生成所述生物组织样本的三维展示数据包括：

对所述表达量数据和/或所述图像数据进行配准得到的配准结果进行生物信息学分析，得到聚类和注释结果；

基于所述聚类和注释结果，构造三维数组，并增加空间尺度信息；

利用三维渲染工具，渲染得到所述生物组织样本的三维展示数据。

15、一种数据处理方法，其特征在于，所述方法包括：

获取生物组织样本的表达量数据和图像数据，其中所述表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，所述图像数据是对切片后的生物组织样本进行图像采集得到的；

对所述表达量数据进行计算，得到表达量图；

将所述表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；

对所述图像对依次进行刚性配准和弹性配准，以生成所述生物组织样本的轮廓数据，所述轮廓数据用于生成所述生物组织样本的三维展示数据。

16、根据权利要求 15 所述的方法，其特征在于，所述对所述图像对进行刚性配准包括：

对所述图像对分别进行下采样；

提取所述图像对的特征信息；

对所述图像对进行相似性估计仿射变换，得到刚性配准结果。

17、根据权利要求 16 所述的方法，其特征在于，所述对所述图像对进行弹性配准包括：

对所述刚性配准结果进行直方图均衡化；

将所述刚性配准结果输入深度学习模型，输出所述深度学习模型的中间结果，得到形变场，其中，所述深度学习模型是通过表达量图-图像数据对进行训练得到的；

根据所述形变场，对所述刚性配准结果进行最近邻插值变换，得到弹性配准结果。

18、一种数据采集装置，其特征在于，所述装置包括：

测序模块，用于对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据；

采集模块，用于对切片后的生物组织样本进行图像采集，以得到图像数据；

配准模块，用于对所述表达量数据和/或所述图像数据进行配准，以生成所述生物组织样本的外层轮廓和/或分区轮廓；

生成模块，用于根据所述外层轮廓和/或所述分区轮廓，生成所述生物组织样本的三维展示数据。

19、一种数据处理装置，其特征在于，所述装置包括：

获取模块，用于获取生物组织样本的表达量数据和图像数据，其中所述表达量数据是对生物组织样本进行切片并对切片进行时空组测序得到的，所述图像数据是对切片后的生物组织样本进行图像采集得到的；

计算模块，用于对所述表达量数据进行计算，得到表达量图；

均衡化模块，用于将所述表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对；

配准模块，对所述图像对依次进行刚性配准和弹性配准，以生成所述生物组织样本的轮廓数据，所述轮廓数据用于生成所述生物组织样本的三维展示数据。

20、一种电子设备，其特征在于，包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求 1-17 中任一项所述的方法。

21、一种存储有计算机指令的非瞬时计算机可读存储介质，其特征在于，所述计算机指令用于使所述计算机执行根据权利要求 1-17 中任一项所述的方法。

22、一种计算机程序产品，其特征在于，包括计算机程序，所述计算机程序在被处理器执行时实

现根据权利要求 1-17 中任一项所述的方法。

23、一种芯片，其特征在于，包括一个或多个接口电路和一个或多个处理器；所述接口电路用于从电子设备的存储器接收信号，并向所述处理器发送所述信号，所述信号包括存储器中存储的计算机指令，当所述处理器执行所述计算机指令时，使得所述电子设备执行权利要求 1-17 中任一项所述的方法。

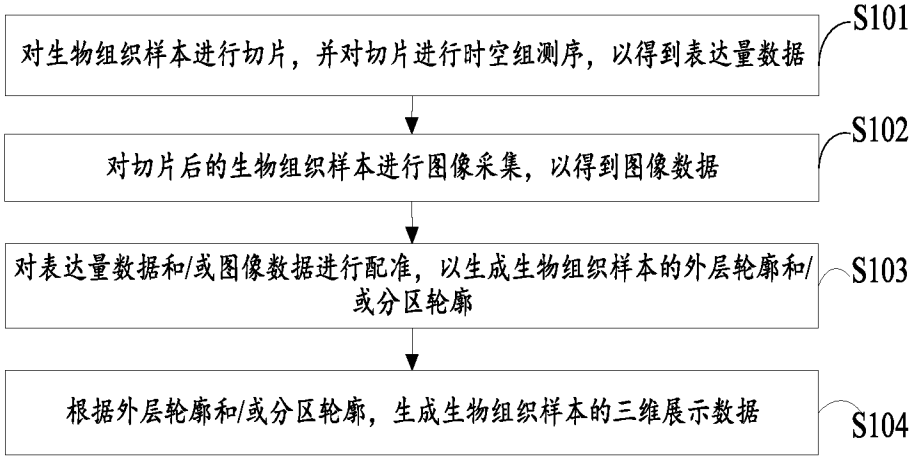

对生物组织样本进行切片，并对切片进行时空组测序，以得到表达量数据
S101
对切片后的生物组织样本进行图像采集，以得到图像数据
S102
对表达量数据和/或图像数据进行配准，以生成生物组织样本的外层轮廓和/或分区轮廓
S103
根据外层轮廓和/或分区轮廓，生成生物组织样本的三维展示数据
S104

图 1

对生物组织样本进行连续切片，每切第一预定数量的切片进行一次时空组测序，以得到切片的表达量数据
S201
每切第二预定数量的切片，对切片后的生物组织样本进行一次图像采集，以得到图像数据
S202
对图像数据进行图像分割处理，以生成分割结果图
S203
对分割结果图进行配准，以生成配准结果图
S204
基于所述配准结果图，生成所述生物组织样本的外层轮廓
S205
对表达量数据和/或图像数据进行配准得到的配准结果进行生物信息学分析，得到聚类和注释结果
S206
基于聚类和注释结果，构造三维数组，并增加空间尺度信息
S207
利用三维渲染工具，渲染得到生物组织样本的三维展示数据
S208

图 2

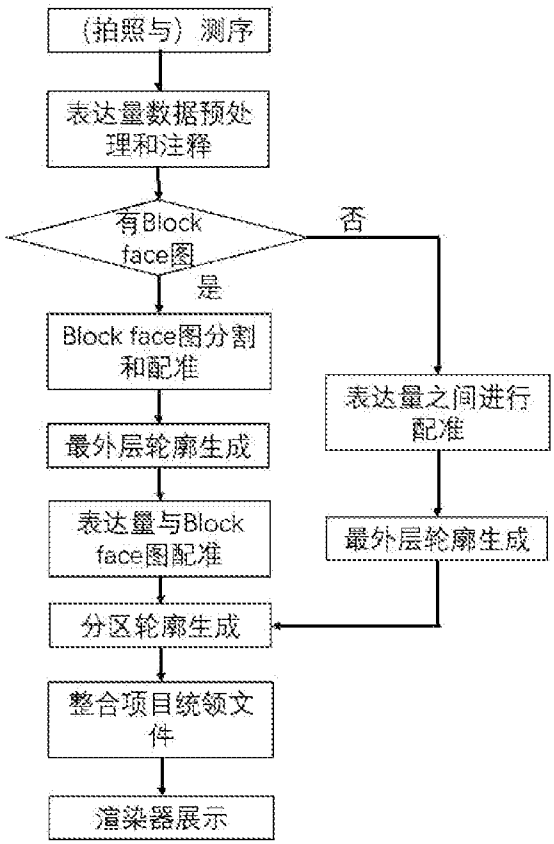

（拍照与）测序
表达量数据预处理和注释
有Block face图
否
是
Block face图分割和配准
表达量之间进行配准
最外层轮廓生成
表达量与Block face图配准
最外层轮廓生成
分区轮廓生成
整合项目统领文件
渲染器展示

图 3

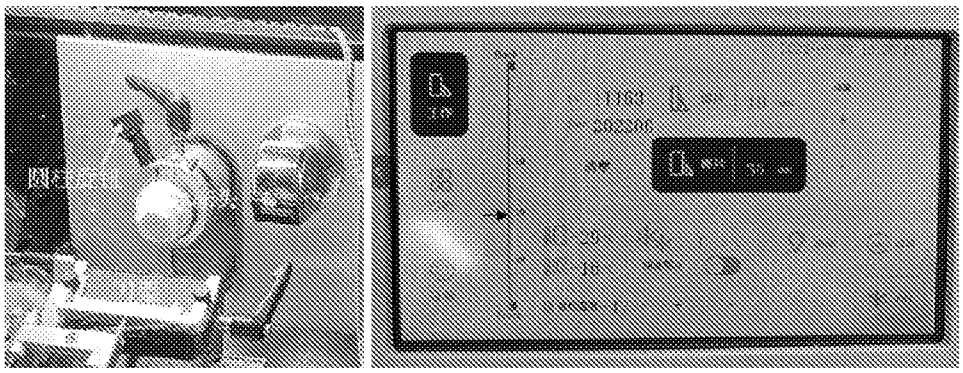


图 4

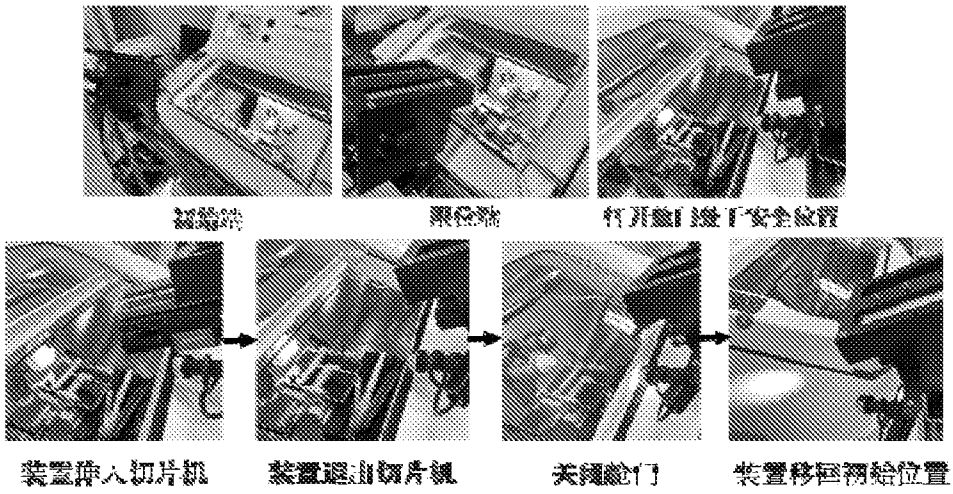


图 5

原始表达量数据：

geneID	x	y	UMICount
Gm42418	12765	24043	1
Gm42418	13327	22767	1
Socs5	14890	18298	1
Rpl35a	14890	18298	1
Eif2s1	14890	18298	1
Arl6	17890	20562	2

预处理和注释后的表达量数据：

spotID	geneID	xPos	yPos	UMICount		label	tag
0	Gm42418	12765	24043	1	60291.0	adjust	
0	Gm42418	13327	22767	1	54209.0	adjust	
0	Socs5	14890	18298	1	23901.0	adjust	
0	Rpl35a	14890	18298	1	23901.0	adjust	

图 6

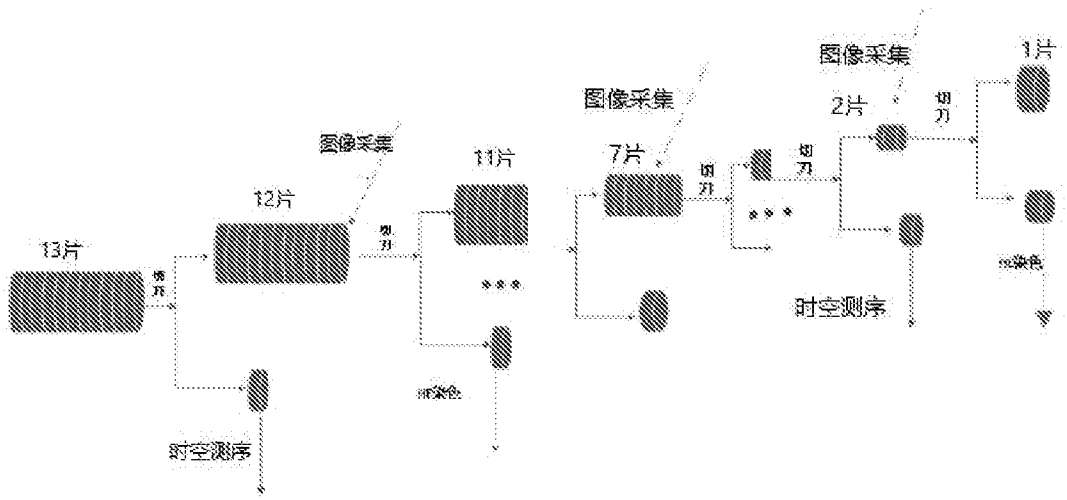


图 7

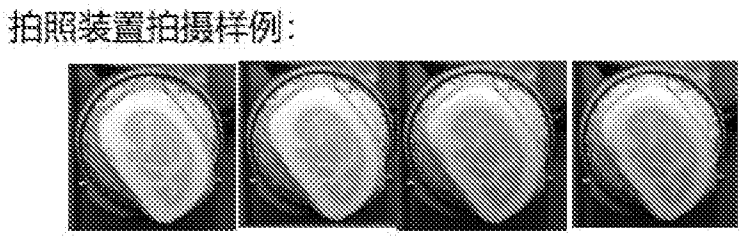

拍照装置拍摄样例：

图 8

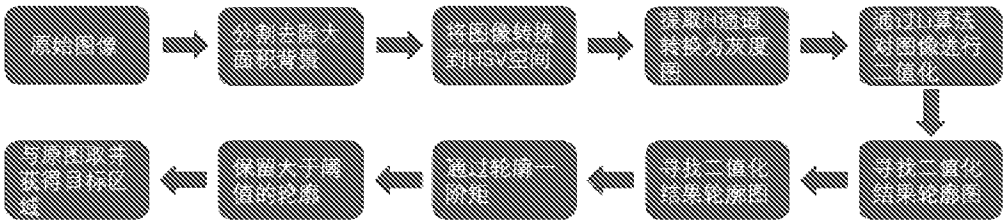


图 9

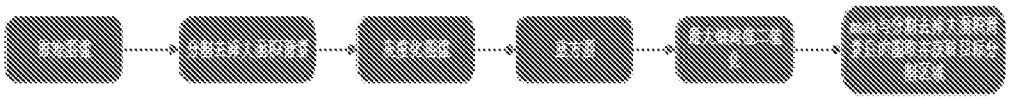


图 10

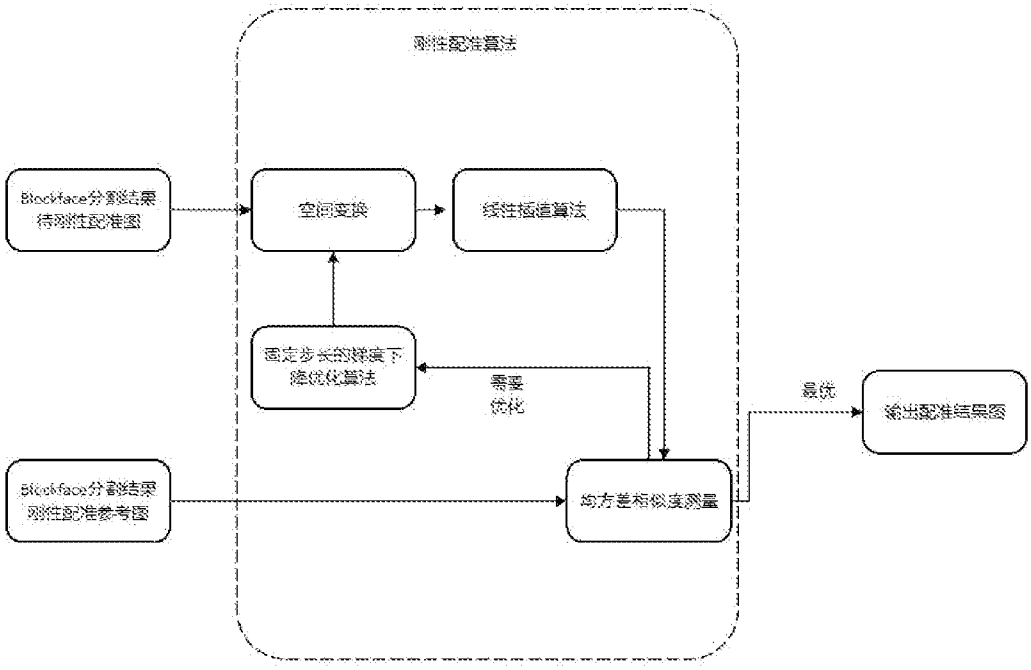

刚性配准算法
Blockface分割结果
待刚性配准图
空间变换
线性插值算法
固定步长的梯度下
降优化算法
需要
优化
最优
输出配准结果图
Blockface分割结果
刚性配准参考图
均方差相似度测量

图 11

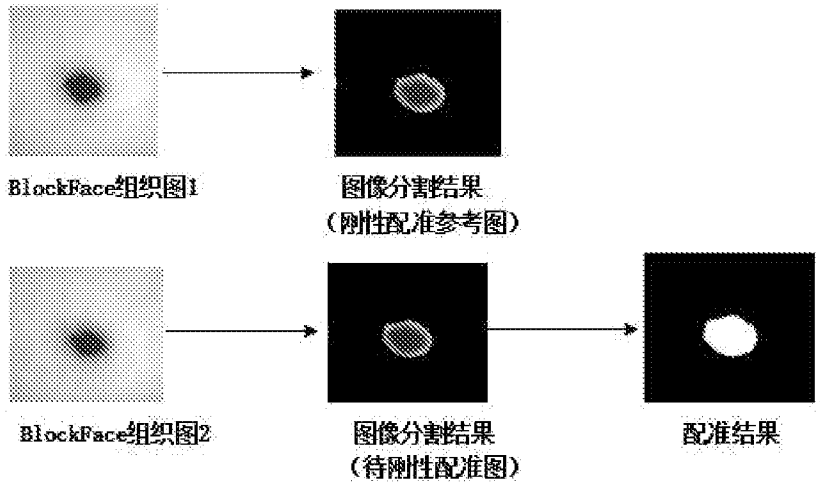

BlockFace组织图1
图像分割结果
（刚性配准参考图）
BlockFace组织图2
图像分割结果
（待刚性配准图）
配准结果

图 12

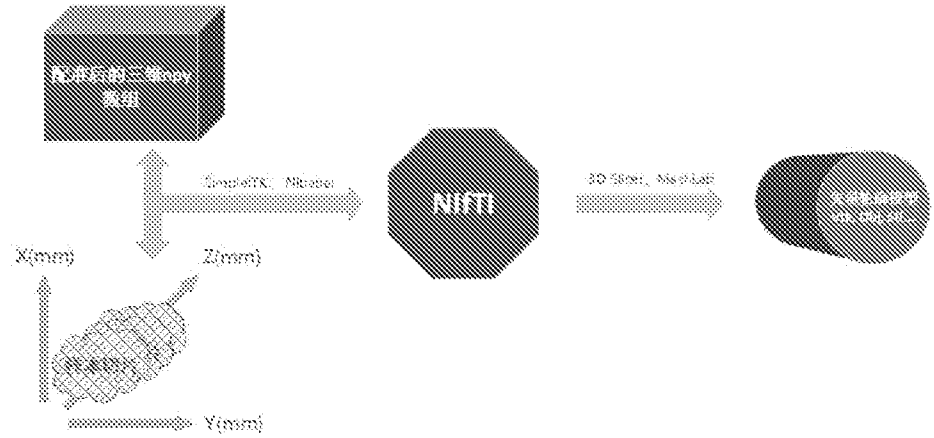

NIfTI
X(mm)
Z(mm)
Y(mm)

图 13

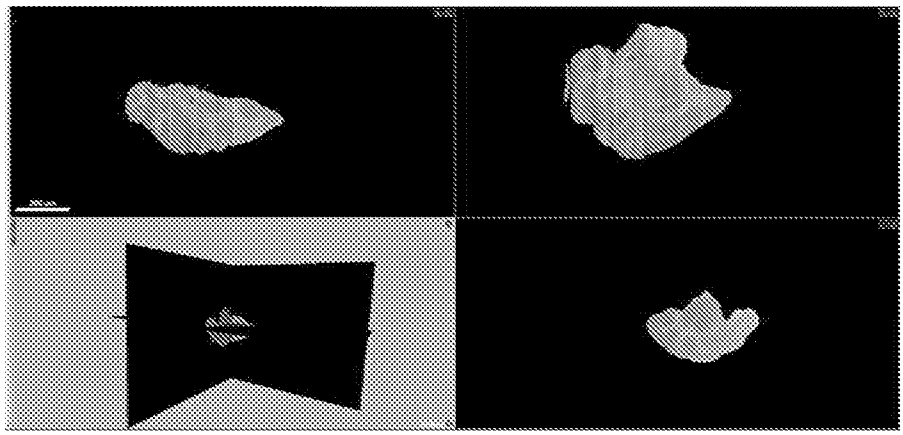


图 14

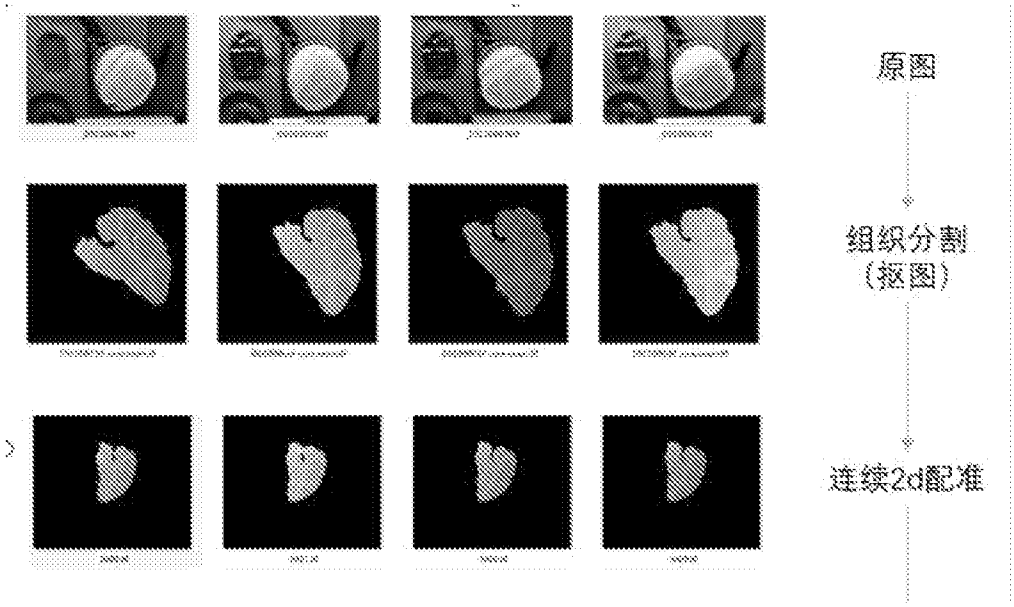

原图
组织分割
（抠图）
连续2d配准

图 15

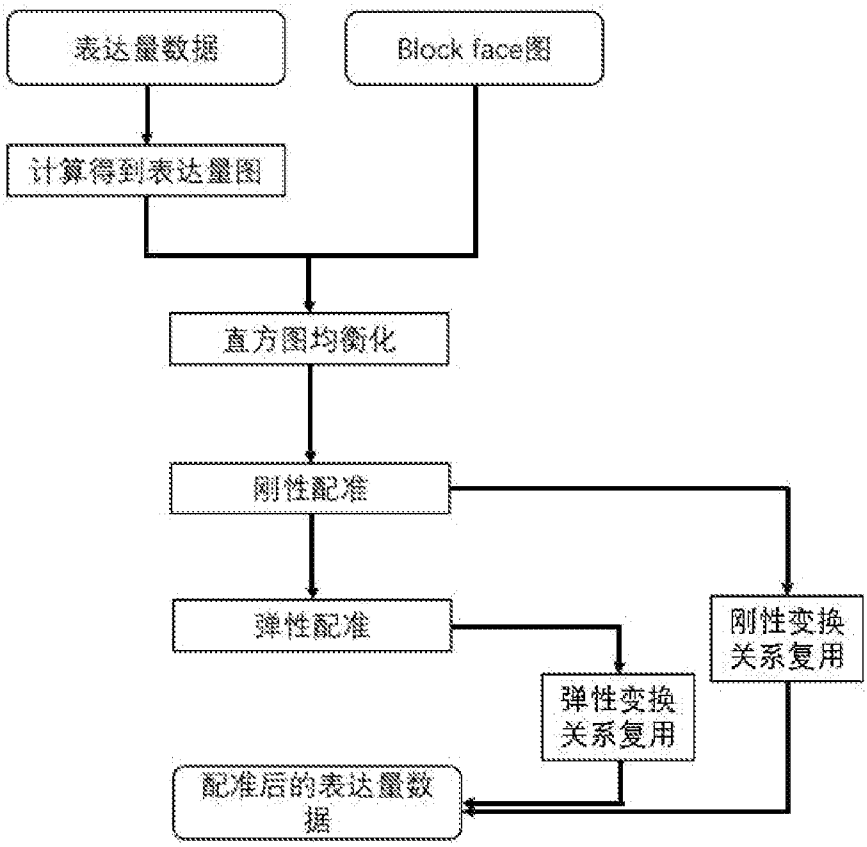

表达量数据
Block face图
计算得到表达量图
直方图均衡化
刚性配准
弹性配准
刚性变换
关系复用
弹性变换
关系复用
配准后的表达量数
据

图 16

BlockFace图和表达量图直方图均衡化：

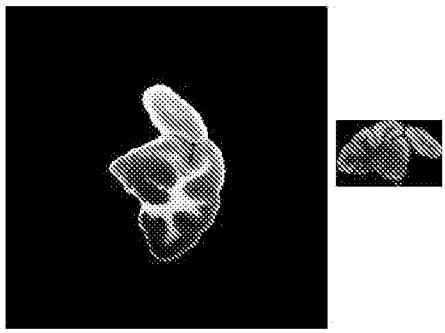


图 17

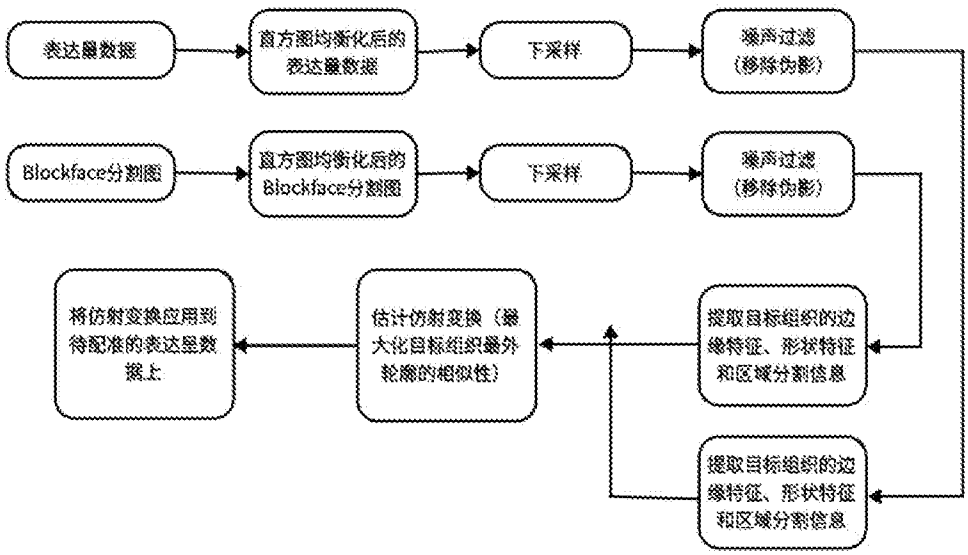


图 18

BlockFace图和表达量图刚性配准结果：

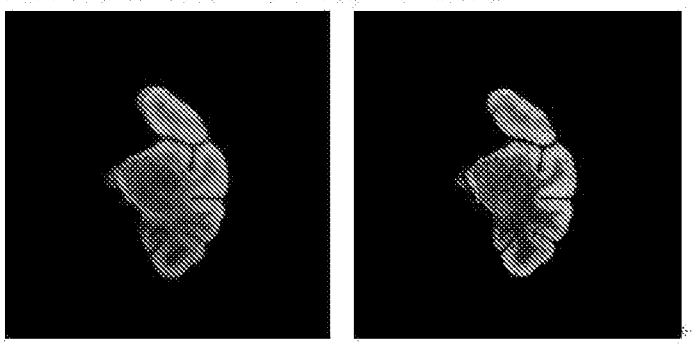


图 19

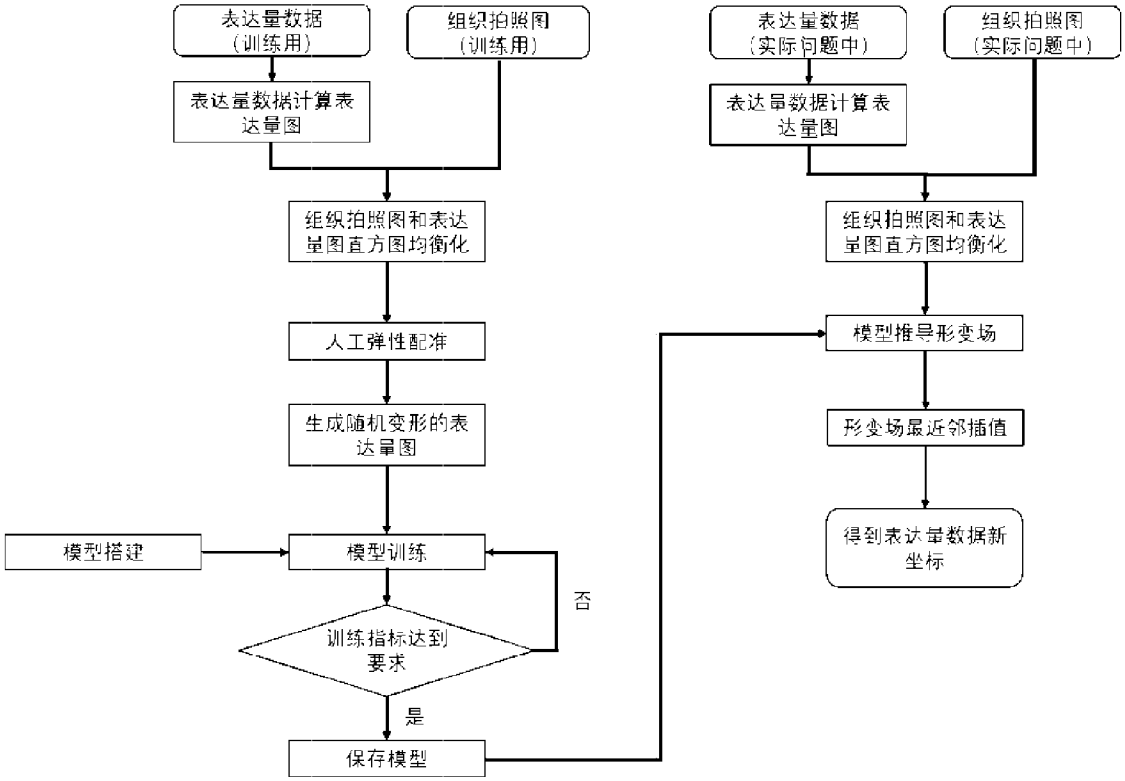

表达量数据（训练用）
组织拍照图（训练用）
表达量数据计算表达量图
组织拍照图和表达量图直方图均衡化
人工弹性配准
生成随机变形的表达量图
模型搭建
模型训练
训练指标达到要求
否
是
保存模型
表达量数据（实际问题中）
组织拍照图（实际问题中）
表达量数据计算表达量图
组织拍照图和表达量图直方图均衡化
模型推导形变场
形变场最近邻插值
得到表达量数据新坐标

图 20

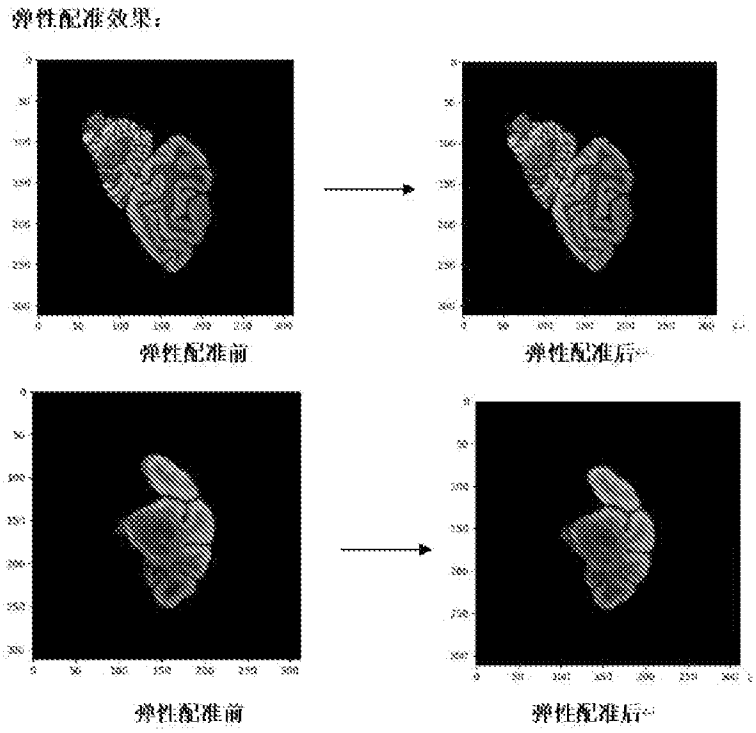

弹性配准效果：
弹性配准前
弹性配准后
弹性配准前
弹性配准后

图 21

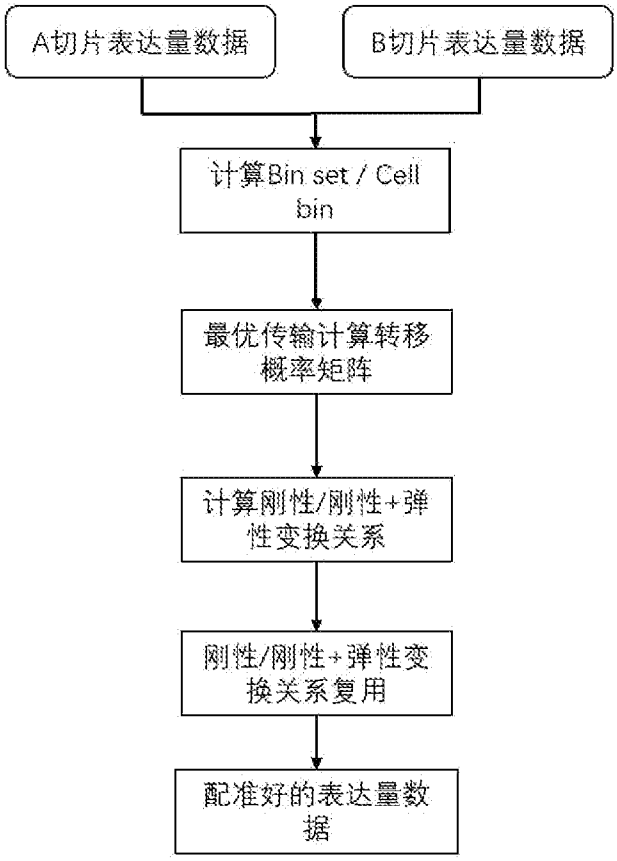

A切片表达量数据
B切片表达量数据
计算Bin set / Cell bin
最优传输计算转移概率矩阵
计算刚性/刚性+弹性变换关系
刚性/刚性+弹性变换关系复用
配准好的表达量数据

图 22

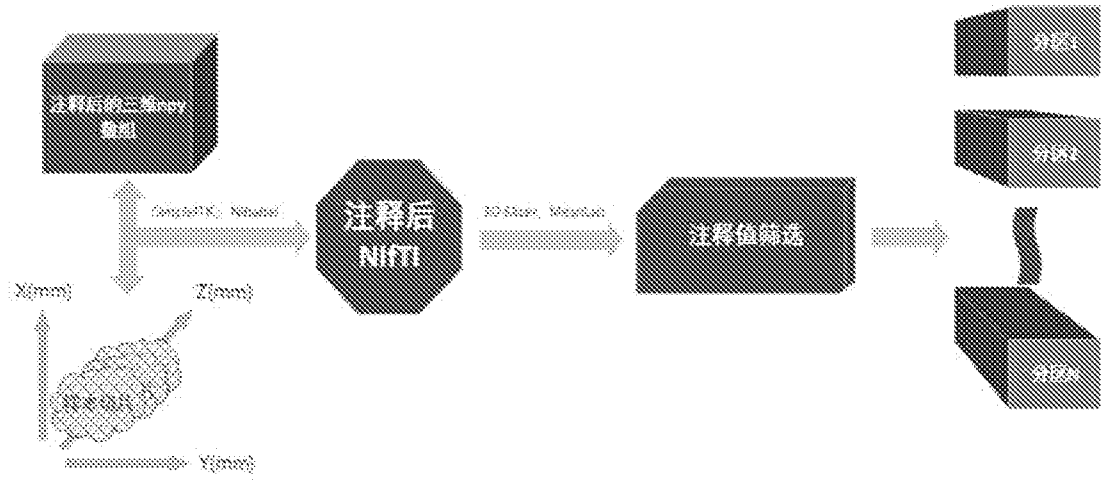

注释后
NIfTI
注释细胞迁
X(mm)
Z(mm)
Y(mm)

图 23

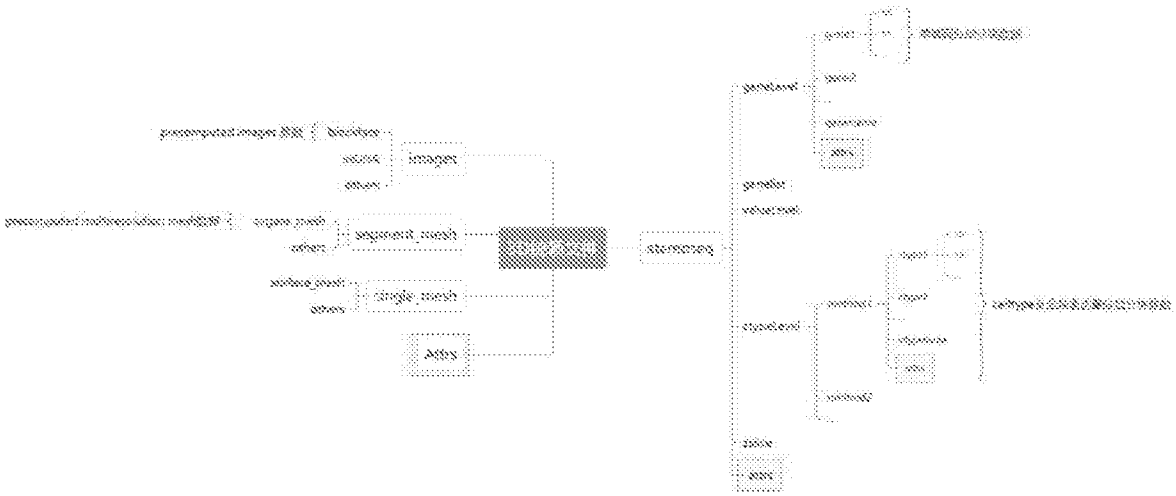


图 24

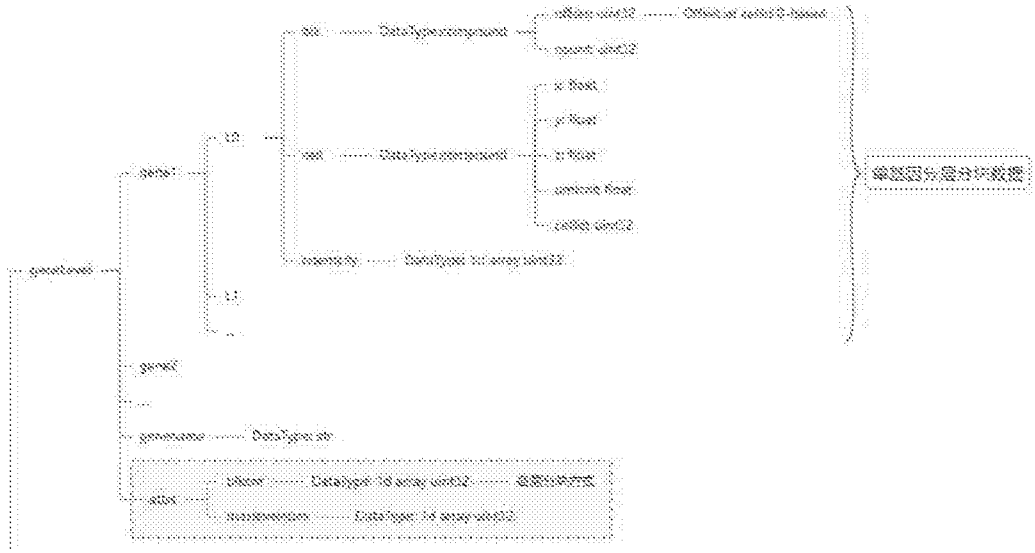


图 25

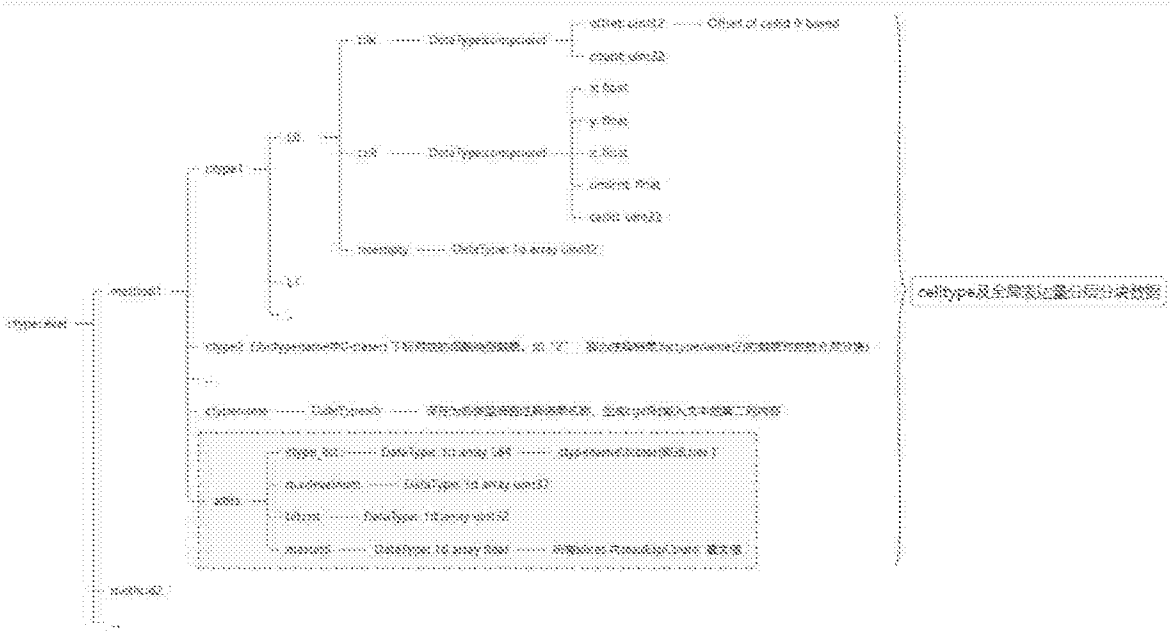


图 26

渲染效果：

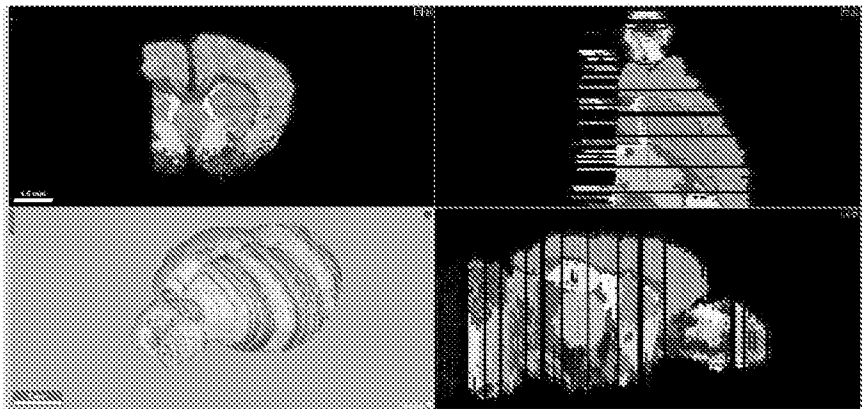


图 27

生成组织外轮廓模型：

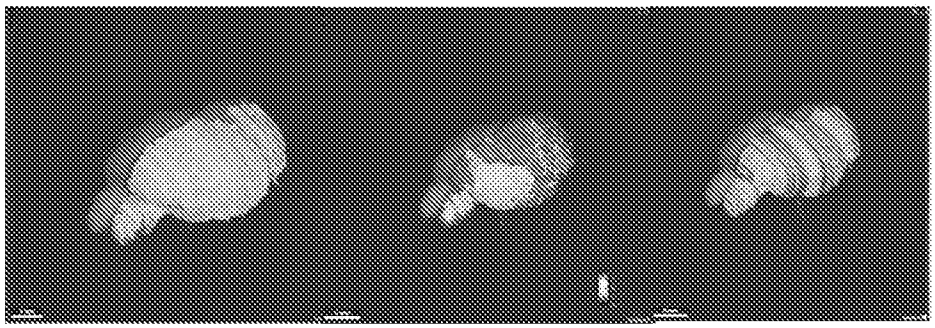


图 28

生成组织分区轮廓模型：

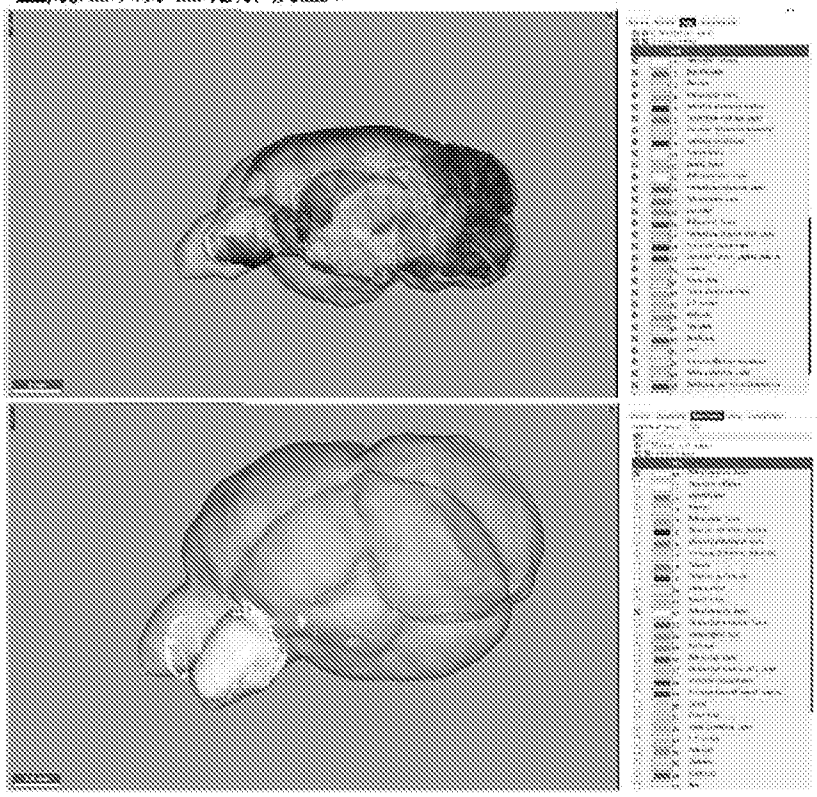


图 29

鼠脑可展示单基因列表示意图：

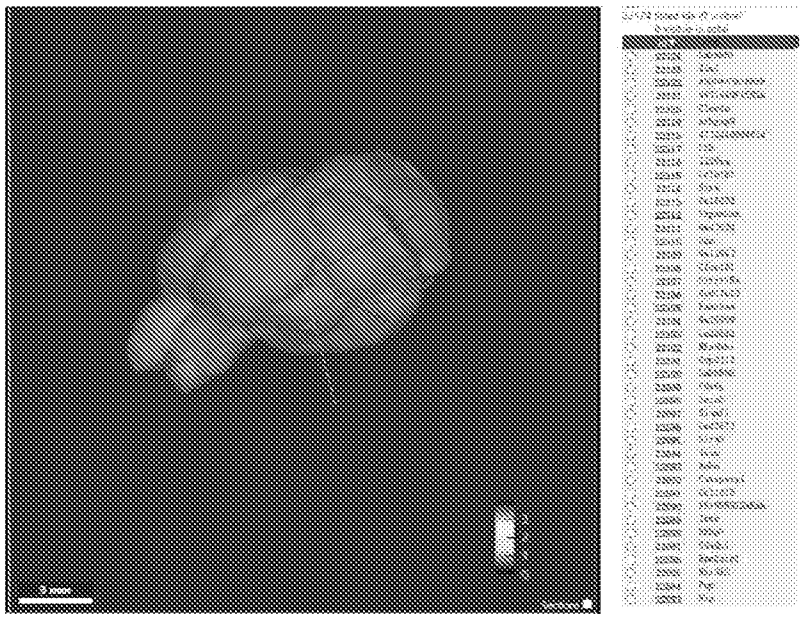


图 30

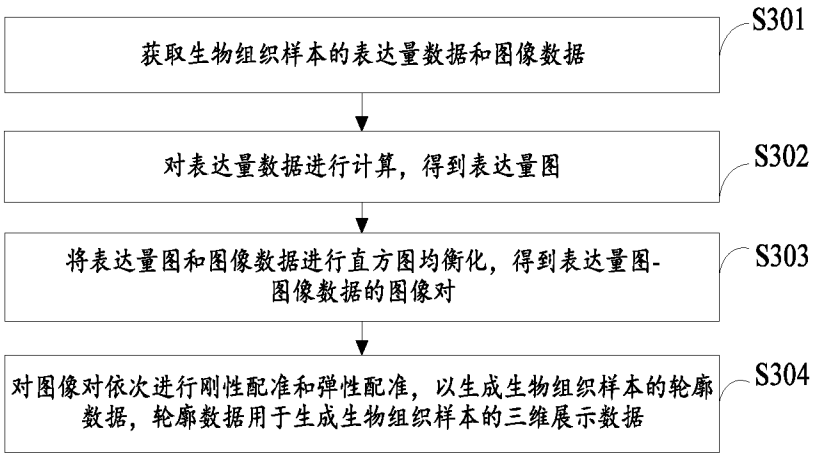

获取生物组织样本的表达量数据和图像数据
S301
对表达量数据进行计算，得到表达量图
S302
将表达量图和图像数据进行直方图均衡化，得到表达量图-图像数据的图像对
S303
对图像对依次进行刚性配准和弹性配准，以生成生物组织样本的轮廓数据，轮廓数据用于生成生物组织样本的三维展示数据
S304

图 31

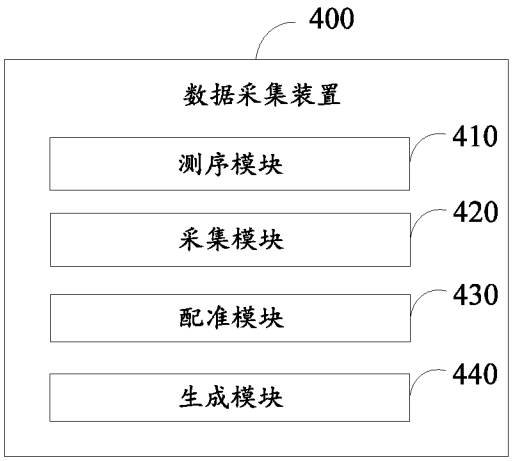

400
数据采集装置
测序模块
410
采集模块
420
配准模块
430
生成模块
440

图 32

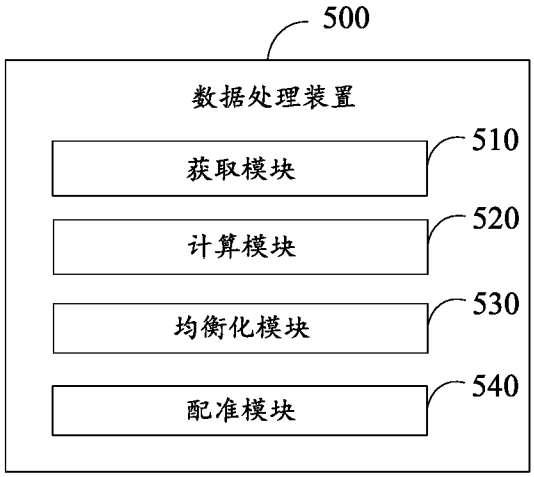

500
数据处理装置
获取模块
510
计算模块
520
均衡化模块
530
配准模块
540

图 33

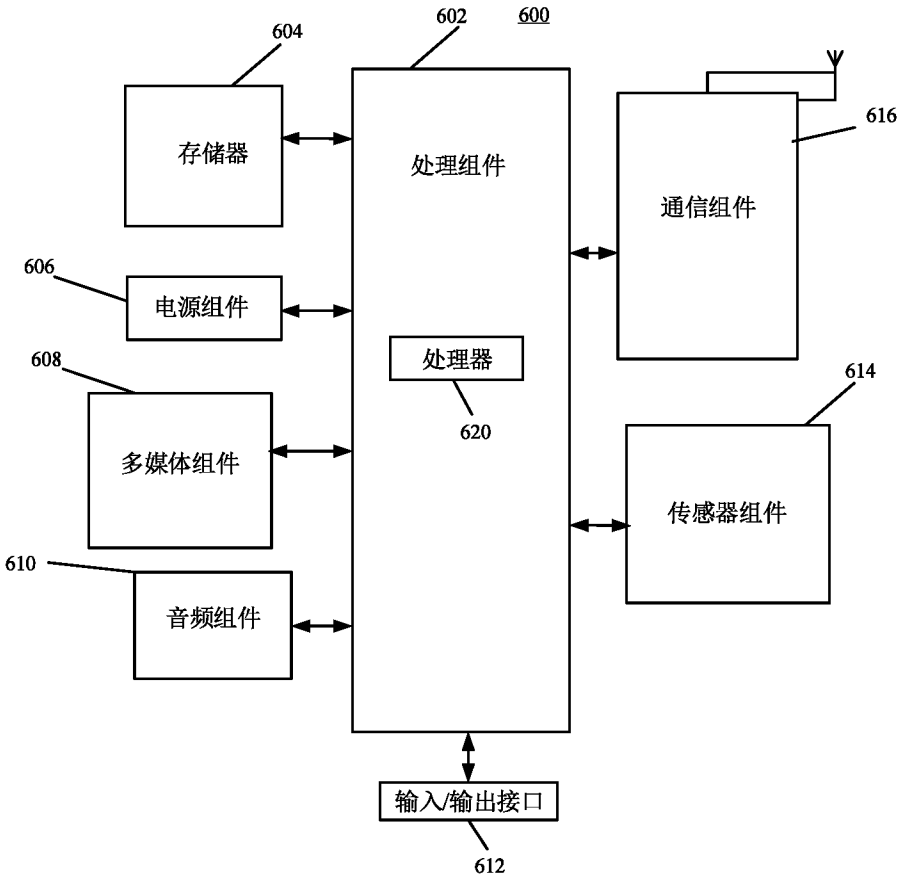

604
602
600
存储器
处理组件
616
通信组件
606
电源组件
处理器
620
608
多媒体组件
614
传感器组件
610
音频组件
输入/输出接口
612

图 34

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/096110

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/33(2017.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNTXT; CNABS; EPTXT; USTXT; WOTXT; VEN; CNKI; IEEE: 组织, 切片, 图像, 表达量, 配准, 轮廓, 三维, tissue, slice, image, expression level, registration, profile, three dimensional

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010150414 A1 (YOKOTA, Hideo et al.) 17 June 2010 (2010-06-17) description, paragraphs 60-125	1-23
Y	CN 113592925 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 02 November 2021 (2021-11-02) description, paragraph 5	1-23
Y	CN 112581457 A (WUHAN UNIVERSITY OF TECHNOLOGY) 30 March 2021 (2021-03-30) description, paragraphs 5-14	1-23
A	US 2018137689 A1 (MICROBRIGHTFIELD, INC.) 17 May 2018 (2018-05-17) entire document	1-23

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 January 2024	**30 January 2024**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/096110

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 2010150414 A1	17 June 2010	JP 2007064689 A	15 March 2007
		JP 4724497 B2	13 July 2011
		US 8189898 B2	29 May 2012
		WO 2007026705 A1	08 March 2007
		EP 1935988 A1	25 June 2008
		EP 1935988 A4	25 June 2014
CN 113592925 A	02 November 2021	None	
CN 112581457 A	30 March 2021	CN 112581457 B	12 December 2023
US 2018137689 A1	17 May 2018	US 10643396 B2	05 May 2020

国际检索报告

国际申请号
PCT/CN2023/096110

A. 主题的分类

G06T 7/33(2017.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:G06T

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNTXT;CNABS;EPTXT;USTXT;WOTXT;VEN;CNKI;IEEE:组织,切片,图像,表达量,配准,轮廓,三维, tissue, slice, image, expression level, registration, profile, three dimensional

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	US 2010150414 A1 (YOKOTA,Hideo 等) 2010年6月17日 (2010 - 06 - 17) 说明书第60-125段	1-23
Y	CN 113592925 A (华中科技大学) 2021年11月2日 (2021 - 11 - 02) 说明书第5段	1-23
Y	CN 112581457 A (武汉理工大学) 2021年3月30日 (2021 - 03 - 30) 说明书第5-14段	1-23
A	US 2018137689 A1 (MICROBRIGHTFIELD INC) 2018年5月17日 (2018 - 05 - 17) 全文	1-23

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2024年1月26日	2024年1月30日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	受权官员 黄长霞 电话号码 (+86) 020-28958625

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/096110

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
US	2010150414	A1	2010年6月17日	JP	2007064689	A	2007年3月15日
				JP	4724497	B2	2011年7月13日
				US	8189898	B2	2012年5月29日
				WO	2007026705	A1	2007年3月8日
				EP	1935988	A1	2008年6月25日
				EP	1935988	A4	2014年6月25日
CN	113592925	A	2021年11月2日		无		
CN	112581457	A	2021年3月30日	CN	112581457	B	2023年12月12日
US	2018137689	A1	2018年5月17日	US	10643396	B2	2020年5月5日