

(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
2020年2月27日 (27.02.2020)



(10) 国际公布号
WO 2020/037574 A1

- (51) 国际专利分类号:
G06T 7/00 (2017.01)
- (21) 国际申请号: PCT/CN2018/101819
- (22) 国际申请日: 2018年8月22日 (22.08.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 深圳市真迈生物科技有限公司 (GENEMIND BIOSCIENCES COMPANY LIMITED) [CN/CN]; 中国广东省深圳市罗湖区清水河街道清水河一路116号深业进元大厦2座5楼、6楼, Guangdong 518000 (CN)。
- (72) 发明人: 李林森(LI, Linsen); 中国广东省深圳市罗湖区莲塘街道国威路72号高新技术产业第一园区111栋一楼, Guangdong 518000 (CN)。徐伟彬(XU, Weibin); 中国广东省深圳市罗湖区莲塘街道国威路72号高新技术产业第一园区111栋一楼, Guangdong 518000 (CN)。金欢(JIN, Huan); 中国广东省深圳市罗湖区莲塘街道国威路72号高新技术产业第一园区111栋一楼, Guangdong 518000 (CN)。姜泽飞(JIANG, Zefei); 中国广东省深圳市罗湖区莲塘街道国威路72号高新技术产业第一园区111栋一楼, Guangdong 518000 (CN)。周志良(ZHOU, Zhiliang); 中国广

东省深圳市罗湖区莲塘街道国威路72号高新技术产业第一园区111栋一楼, Guangdong 518000 (CN)。颜钦(YAN, Qin); 中国广东省深圳市罗湖区莲塘街道国威路72号高新技术产业第一园区111栋一楼, Guangdong 518000 (CN)。

- (74) 代理人: 深圳鼎合诚知识产权代理有限公司 (DHC IP ATTORNEYS); 中国广东省深圳市福田区金田路与福华路交汇处现代商务大厦2201, Guangdong 518048 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,

(54) Title: METHOD FOR CONSTRUCTING SEQUENCING TEMPLATE BASED ON IMAGE, AND BASE RECOGNITION METHOD AND DEVICE

(54) 发明名称: 基于图像构建测序模板的方法、碱基识别方法和装置

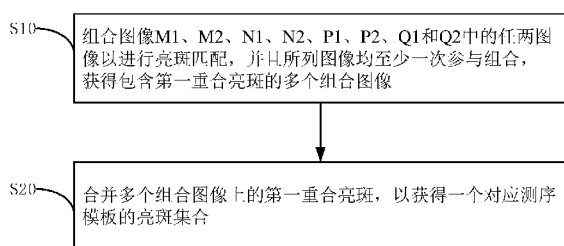


图 1

- S10 Combine any two of images M1, M2, N1, N2, P1, P2, Q1, and Q2 to perform bright spot matching, and enable the images to participate in the combination for at least one time to obtain a plurality of combined images comprising first coincident bright spots
- S20 Merge the first coincident bright spots on the plurality of combined images to obtain a bright spot set corresponding to the sequencing template

(57) Abstract: A method for constructing a sequencing template based on an image, a device, and a system. The image comprises first, second, third and fourth images of one same field of view corresponding to base extension reactions of A/U, T, G, and C respectively; the first, second, third and fourth images respectively comprise images M1 and M2, images N1 and N2, images P1 and P2, and images Q1 and Q2; the method for constructing the sequencing template comprises: combining any two of the images M1, M2, N1, N2, P1, P2, Q1, and Q2 to perform bright spot matching, and enabling the images M1, N1, N2, P1, P2, Q1, and Q2 to participate in the combination for at least one time to obtain a plurality of combined images comprising first coincident bright spots (S10), two or more bright spots whose distances are less than that of a first predetermined pixel on the combined images being one first coincident bright spot; and merging the first coincident bright spots on the plurality of combined images to obtain a bright spot set corresponding to the sequencing template (S20). The method can effectively obtain the bright spot set corresponding to a nucleic acid template.

CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, 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：一种基于图像构建测序模板的方法、装置和系统。所称的图像包括分别对应A/U、T、G和C四种碱基延伸反应时的一个相同视野的第一、第二、第三和第四图像，第一、第二、第三和第四图像分别包括图像M1和M2、图像N1和N2、图像P1和P2、图像Q1和Q2，该构建测序模板的方法包括：组合图像M1、M2、N1、N2、P1、P2、Q1和Q2中的任两图像以进行亮斑匹配，并且使图像M1、N1、N2、P1、P2、Q1和Q2均至少一次参与该组合，获得包含第一重合亮斑的多个组合图像(S10)，在组合图像上距离小于第一预定像素的两个或多个亮斑为一个第一重合亮斑；合并多个组合图像上的第一重合亮斑，以获得一个对应测序模板的亮斑集合(S20)。该方法能够有效地获取对应核酸模板的亮斑集合。

基于图像构建测序模板的方法、碱基识别方法和装置

技术领域

本发明涉及图像处理和信息识别领域，特别地，涉及一种基于图像构建测序模板的方法、一种碱基识别方法、一种基于图像构建测序模板的装置、一种碱基识别装置和一种计算机程序产品。

背景技术

在相关技术中，包括在基于成像系统多次对生化反应中的核酸分子（模板）进行图像采集以测定该核酸分子的核苷酸顺序的测序平台中，如何处理以及关联多次不同时间点所采集的图像包括图像上的信息，以有效且准确地获得核酸模板的至少一部分的核苷酸组成和顺序，是值得关注的问题。

发明内容

本发明实施方式旨在至少解决相关技术中存在的技术问题之一或者至少提供一种可选择的实用方案。

依据本发明的一个实施方式，提供一种基于图像构建测序模板的方法，所称的图像包括分别对应 A/ U、T、G 和 C 四种碱基延伸反应时的一个相同视野的第一图像、第二图像、第三图像和第四图像，碱基延伸反应时的该视野存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在图像上表现为亮斑，定义顺序和/或同时实现一次四种类型碱基延伸反应为一轮测序反应，第一图像包括图像 M1 和图像 M2，第二图像包括图像 N1 和图像 N2，第三图像包括图像 P1 和图像 P2，第四图像包括图像 Q1 和图像 Q2，图像 M1 和图像 M2 分别来自两轮测序反应，图像 N1 和图像 N2 分别来自两轮测序反应，图像 P1 和图像 P2 分别来自两轮测序反应，图像 Q1 和图像 Q2 分别来自两轮测序反应，该方法包括：组合图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 中的任两图像以进行亮斑匹配，并且使图像 M1、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 均至少一次参与组合，获得包含第一重合亮斑的多个组合图像，在组合图像上距离小于第一预定像素的两个或多个亮斑为一个第一重合亮斑；以及合并多个组合图像上的第一重合亮斑，以获得一个对应测序模板的亮斑集合。

依据本发明的一个实施方式，提供一种基于图像构建测序模板的装置，该装置用以实施上述本发明实施方式中的基于图像构建测序模板的方法的全部或部分步骤。所称的图像包括分别对应 A/ U、T、G 和 C 四种碱基延伸反应时的一个相同视野的第一图像、第二图像、第三图像和第四图像，碱基延伸反应时的该视野存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在图像上表现为亮斑，定义顺序和/或同时实现一次四种类型碱基延伸反应为一轮测序反应，第一图像包括图像 M1 和图像 M2，第二图像包括图像 N1 和图像 N2，第三图像包括图像 P1 和图像 P2，第四图像包括图像 Q1 和图像 Q2，图像 M1 和图像 M2 分别来自两轮测序反应，图像 N1 和图像 N2 分别来自两轮测序反应，图像 P1 和图像 P2 分别来自两轮测序反应，图像 Q1 和图像 Q2 分别来自两轮测序反应，该装置包括：组合单元，用于组合图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 中的任两图像以进行亮斑匹配，并且使图像 M1、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 均至少一次参与该组合，获得包含第一重合亮斑的多个组合图像，在组合图像上距离小于第一预定像素的两个或多个亮斑为一个第一重合亮斑；以及合并单元，用于合并来自所述组合单元的多个组合图像上的第一重合亮斑，以获得一个对应测序模板的亮斑集合。

依据本发明的一个实施方式，提供一种计算机可读存储介质，用于存储供计算机执行的程序，执行所述程序包括完成上述任一实施方式中的基于图像构建测序模板的方法。计算机可读存储介质包括但不限于只读存储器、随机存储器、磁盘或光盘等。

依据本发明的一个实施方式，还提供一种终端，一种计算机程序产品，该产品包括指令，该指令在计算机执行所称的程序时，使该计算机执行上述本发明实施方式中的基于图像的构建测序模板的方法。

利用上述基于图像构建测序模板的方法、装置、计算机可读存储介质和/或计算机程序产品构建获得的测序模板，是一个对应测序模板的亮斑集合，该亮斑集合能有效、准确且全面的反映测序模板的信息，利于进一步的碱基的准确识别（base call），即准确识别获取模板核酸的至少一部分的核苷酸序列。

依据本发明的另一个实施方式，提供一种碱基识别方法，该方法包括将获自碱基延伸反应的图像上的亮斑匹配到对应测序模板的亮斑集合，依据匹配上的亮斑进行碱基识别，获自碱基延伸反应的图像对应的视野中存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在获自碱基延伸反应的图像上表现为亮斑，对应测序模板的亮斑集合通过上述本发明实施方式中的基于图像构建测序模板的方法、装置、计算机可读存储介质和/或计算机程序产品来构建获取。

依据本发明的一个实施方式，提供一种碱基识别装置，该装置用于实施上述本发明实施方式中的碱基识别方法，该装置用于将获自碱基延伸反应的图像上的亮斑匹配到对应测序模板的亮斑集合，依据匹配上的亮斑进行碱基识别，获自碱基延伸反应的图像对应的视野中存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在获自碱基延伸反应的图像上表现为亮斑，对应测序模板的亮斑集合通过上述本发明实施方式中的基于图像构建测序模板的方法和/或装置构建获得。

依据本发明的一个实施方式，提供一种计算机可读存储介质，用于存储供计算机执行的程序，执行所述程序包括完成上述任一实施方式中的碱基识别方法。计算机可读存储介质包括但不限于只读存储器、随机存储器、磁盘或光盘等。

依据本发明的一个实施方式，还提供一种计算机程序产品，该产品包括实现碱基识别的指令，该指令在计算机执行所称的程序时，使该计算机执行上述本发明实施方式中的碱基识别的方法。

利用该碱基识别方法、装置、计算机可读存储介质和/或计算机程序产品，基于构建得的对应测序模板的亮斑集合，能够识别碱基延伸反应时与模板核酸结合的碱基的类型，能够用于实现模板核酸序列的准确测定。

本发明实施方式的附加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，或通过本发明实施方式的实践了解到。

附图说明

图 1 是本发明具体实施方式中的基于图像构建测序模板的方法的流程示意图。

图 2 是本发明具体实施方式中的基于亮斑对图像 Repeat1、Repeat5、Repeat6 和 Repeat7 进行组合和合并以构建测序模板的示意图。

图 3 是本发明的具体实施方式中的纠偏过程和纠偏结果的示意图。

图 4 是本发明具体实施方式中的候选亮斑的对应的矩阵以及连同像素示意图。

图 5 是本发明具体实施方式中的以像素点矩阵的中心像素点为中心的 $m1*m2$ 范围的像素值示意图。

图 6 是本发明具体实施方式中的依据第二亮斑检测阈值进行判定之前和之后的亮斑检测结果对比示意图。

图 7 是本发明具体实施方式中的基于图像构建测序模板的装置的结构示意图。

具体实施方式

下面详细描述本发明的实施方式，实施方式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方式是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。

在本发明的描述中，术语“第一”、“第二”、“第三”、“第四”等仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量或者顺序。在本发明的描述中，“多个”的含义是两个或两个以上，除非另有明确具体的限定。

请参阅图 1，本发明实施方式提供一种基于图像构建测序模板的方法，所称的图像采集自一个相同视野，包括分别采集于 A/U、T、G 和 C 四种碱基延伸反应时的第一图像、第二图像、第三图像和第四图像，碱基延伸反应时的该视野存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在图像上表现为亮斑，定义顺序和/或同时实现一次四种类型碱基延伸反应为一轮测序反应，第一图像包括图像 M1 和图像 M2，第二图像包括图像 N1 和图像 N2，第三图像包括图像 P1 和图像 P2，第四图像包括图像 Q1 和图像 Q2，图像 M1 和图像 M2 分别来自两轮测序反应，图像 N1 和图像 N2 分别来自两轮测序反应，图像 P1 和图像 P2 分别来自两轮测序反应，图像 Q1 和图像 Q2 分别来自两轮测序反应，该方法包括：S10 组合图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 中的任两图像以进行亮斑匹配，并且使图像 M1、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 均至少一次参与组合，获得包含第一重合亮斑的多个组合图像，在组合图像上距离小于第一预定像素的两个或多个亮斑为一个第一重合亮斑；以及 S20 合并多个组合图像上的第一重合亮斑，以获得一个对应测序模板的亮斑集合。所称的“亮斑”也称为“亮点”（spots 或 peaks），指图像上的发光点，一个发光点占有至少一个像素点。所称“像素点”同“像素”。

该方法通过对多个图像上的亮斑先取交集再取并集，能够获得与模板核酸分子对应的亮斑集合。利用该方法获得的测序模板，是一个对应测序模板的亮斑集合，该亮斑集合能有效、准确且全面的反映测序模板的信息，获得的亮斑集合能够进一步用于碱基的准确识别（base call），即用于准确获取模板核酸的至少一部分的核苷酸序列。

所称的一轮测序反应，顺序和/或同时实现一次四种类型碱基延伸反应，可以是四种类型碱基反应底物（例如核苷酸类似物/碱基类似物）同时于一个碱基延伸反应体系中实现一轮测序反应，可以是两种类型碱基类似物于一个碱基延伸反应体系中、另外两种类型反应底物于下一个碱基延伸反应体系以实现一轮测序反应，也可以是一种类型碱基类似物于一个碱基延伸反应体系中、依次在四个连续的碱基延伸反应体系中加入该四种类型碱基类似物以实现一轮测序反应。可知，第一图像、第二图像、第三图像和第四图像可以采集自两次碱基延伸反应或者更多次的碱基延伸反应。另外，一个碱基延伸反应可能包含一次图像采集，也可能包含多次图像采集。

在一个示例中，一轮测序反应包括多次碱基延伸反应，例如单色测序，利用的四种类型碱基对应的反应底物（核苷酸类似物）均带有同一种荧光染料，一轮测序反应包括四次碱基延伸反应（4 repeats），对于一个视野来说，一次碱基延伸反应包含一次图像采集，图像 M1、图像 N1、图像 P1 和图像 Q1 分别为来自一轮测序反应的四次碱基延伸反应的同一视野。

在另一个示例中，例如单分子双色测序反应，利用的四种类型碱基对应的反应底物（核苷酸类似物）中的两种带有一种荧光染料、另两种带有另一种不同激发波长的荧光染料，一轮测序反应包括两次碱基延伸反应，带有不同染料的两种类型碱基反应底物于一次碱基延伸反应中进行结合反应，对于一个视野，一次碱基延伸反应包括两次于不同激发波长下的图像采集，图像 M1、图像 N1、图像 P1 和图像 Q1 分别来自一轮测序反应的两次碱基延伸反应的两种激发波长下的同一视野。

在又一个示例中，一轮测序反应包括一次碱基延伸反应，例如二代测序平台的双色测序反应，四种类型碱基反应底物（例如核苷酸类似物）分别带有染料 a、带有染料 b、带有染料 a 和染料 b 以及不带任何染料，染料 a 和染料 b 的激发波长不一样；四种类型反应底物于同一次碱基延伸反应中实现一轮测序反应，一次碱基延伸反应包括两次于不同激发波长下的图像采集，第一图像同第三图像、第二图像同第四图像，图像 M1 和图像 N1 分别来自不同轮测序反应或者同一轮测序反应中的不同激发波长下的同一视野。

在某些具体实施方式中，S20 合并多个组合图像上的第一重合亮斑，包括对不同组合图像中的第一重合亮斑进行一次或多次匹配，以获得对应测序模板的亮斑集合。如此，利于获得准确的、与模板核酸分子一一对应的亮斑的集合，利于基于图像信息构建准确的模板。

在某些具体实施方式中，图像 M1、图像 N1、图像 P1 和图像 Q1 为顺序获得，图像 M2、图像 N2、图像 P2 和图像 Q2 为顺序获得，即图像 M1、图像 N1、图像 P1 和图像 Q1 于一轮测序反应中获得，图像 M2、图像 N2、图像 P2 和图像 Q2 于另一轮测序反应中获得，S10 包括：间隔 S 个图像对图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 进行两两组合，获得 K 个组合图像以及对组合图像上的亮斑进行匹配，弃组合图像上的非重合亮斑，S 为整数， $0 \leq S \leq S_{\max}$ ， $S_{\max} = \text{参与组合的图像总数} - 4$ 。可以计算出， $K = [(\text{参与组合的图像总数} - S - 1) + 1] * (\text{参与组合的图像总数} - S - 1) / 2$ ，即 $K = C_{(B-S)}^2$ ，例如，S=2 时，K=15。如此，能充分的利用尽可能少的图像信息构建完整的测序模板。

对于一轮测序反应包含四次碱基延伸反应，即每次碱基延伸反应只含一种类型核苷酸类似物，较佳地，S 大于 1，更佳地，S 大于 2，有利于避免或减少生化试验因素带来的噪声对基于图像构建测序模板带来的干扰，利于有效且准确地确定模板。

在一个实施例中，参与组合的图像的总数为 12，S=2，如此，能够获得较完整的测序模板，减少下机数据（读段）的损失。

在另一个实施例中，参与组合的图像的总数为 8，S=2。当 Repeat=5 的时候，将图像 Repeat1-5 (Repeat1 和 Repeat5 的图像)，Repeat2-5 分别进行重合亮斑匹配，然后将匹配结果合并至模板容器 (Template；初始为空) 中；在一个示例中，因 Repeat4 图像用于参考图像的构建，为减少计算量，模板构建从图像 Repeat5 开始；Repeat=6 时，分别对图像 Repeat1-6、Repeat2-6、Repeat3-6 进行重合亮斑匹配，再合并匹配结果至 Template 中；Repeat=7 时，分别对图像 Repeat1-7、Repeat2-7、Repeat3-7、Repeat4-7 进行重合亮斑匹配，再合并匹配结果至 Template 中；Repeat=8 时，分别对图像 Repeat1-8、Repeat2-8、Repeat3-8、Repeat4-8、Repeat5-8 进行重合亮斑匹配，再合并匹配结果至 Template 中；最后统计所有 Template 容器中的亮斑，输出，每个亮斑坐标表示一条链，即一条 reads。模板构建成功后，能知道 reads 总数 TotalRead。图 2 为该过程的示意图，图 2 的上面四个图像依次为 Repeat1、Repeat5、Repeat6 和 Repeat7，中间的图像变化示意 Repeat1 和 Repeat5 重合亮斑匹配的过程和结果，下图示意图像 Repeat1、Repeat5、Repeat6 和 Repeat7 重合亮斑匹配的结果。

在一个示例中，成像系统中，电子传感器的尺寸为 $6.5\mu\text{m}$ ，显微镜放大倍率 60 倍，能看到的最小尺寸就是 $0.1\mu\text{m}$ 。对应核酸分子的亮斑的大小一般为小于 $10*10$ 像素。

所称的第一预定像素，在一个示例中，为 1.05 像素。

在一个示例中，设置距离大于 1.85 像素的两个第一重合亮斑为两个第一重合亮斑。

在一个示例中，舍弃距离一个重合亮斑大于 1.05 像素但小于另一个重合亮斑 1.85 像素的重合亮斑。如此，有利于构建准确的测序模板。

在某些具体实施方式中，图像为经过配准的图像。如此，利于准确地获取对应与测序模板的亮斑集合。

本发明实施方式对实现图像配准即纠偏的方式不作限制。在一些示例中，利用如下方法进行图像配准，包括：基于参考图像对待配准图像进行第一配准，参考图像和待配准图像对应相同对象，参考图像和待配准图像均包含多个亮斑，包括确定待配准图像上的预定区域和参考图像上的相应预定区域的第一偏移量，基于第一偏移量移动待配准图像上的所有亮斑，获得第一配准后的待配准图像；基于参考图像对第一配准后的待配准图像进行第二配准，包括合并第一配准后的待配准图像和参考图像，获得合并图像，计算合并图像上的预定区域的所有重合亮斑的偏移量，以确定第二偏移量，距离小于预定像素的两个或多个亮斑为一个重合亮斑，基于该第二偏移量移动第一配准后的待配准图像上的所有亮斑，以实现对待配准图像的配准。该图像配准方法通过两次关联配准，可相对称为粗配准和细配准，包括利用图像上的亮斑进行细配准，能够基于少量数据信息快速地实现图像的高精度纠偏，特别适于高精度图像纠偏要求的场景。例如，单分子级别的图像检测，比如来自第三代测序平台的测序反应的图像。所称单分子级别指分辨率为单个或少数几个分子的大小，例如 10 个、8 个、5 个、4 个或 3 个以下分子。

在某些具体实施方式中，待配准图像即构建测序模板的图像来自利用光学成像原理进行序列测定的测序平台。所称的测序，也称为序列测定，指核酸序列测定，包括 DNA 测序和/或 RNA 测序，包括长片段测序和/或短片段测序，测序生化反应包括碱基的延伸。测序可以通过测序平台进行，测序平台可选择但不限于 Illumina 公司的 Hisq/Miseq/Nextseq 测序平台、Thermo Fisher/ Life Technologies 公司的 Ion Torrent 平台、华大基因的 BGISEQ 平台和单分子测序平台；测序方式可以选择单端测序，也可以选择双末端测序；获得的测序结果/数据即测读出来的片段，称为读段 (reads)，读段的长度称为读长。所称的“亮斑”对应延伸碱基或碱基簇的光学信号。

所称的图像上的预定区域，可以是整个图像，也可以是图像的一部分。在一个示例中，图像上的预定区域为图像的一部分，例如为图像中心的 512*512 区域。所称的图像中心，为该视野的中心，成像系统的光轴与成像平面的交点可称为图像中心点，以该中心点为中心的区域可视为图像中心区域。

在某些具体实施方式中，待配准图像来自核酸测序平台，该平台包括成像系统和核酸样本承载系统，带有光学检测标记的待测核酸分子固定于反应器中，该反应器也称为芯片，芯片装载在一个可移动台子上，通过该移动台子带动芯片运动来实现对位于芯片不同位置（不同视野）的待测核酸分子进行图像采集。一般地，光学系统和/或移动台子的运动存在精度限制，例如，指令指定运动至某个位置和该机械结构实际运动达到的位置存在偏差，特别是在对精度高要求的应用情景，由此，在依据指令移动硬件以对不同时间点的同一位置（视野）进行多次图像采集的过程中，不同时间点采集的同一视野的多个图像难以完全对齐，对该些图像进行纠偏对齐，有利于基于该多个时间点采集的多个图像中的信息的变化来准确确定核酸分子核苷酸顺序。

在某些具体实施方式中，所称的参考图像是通过构建获得的，参考图像可以在对待配准图像进行配准时构建，也可以预先构建保存需要时调用。

在一些示例中，构建参考图像包括：获取第五图像和第六图像，第五图像和第六图像与待配准图像对应相同对象；基于第五图像对第六图像进行粗配准，包括确定第六图像和第五图像的偏移量，基于该偏移量移动第六图像，获得粗配准后的第六图像；合并第五图像和粗配准后的第六图像，以获得参考图像，第五图像和第六图像均包含多个亮斑。如此，利用构建获得包含更多或相对更完整的信息的图像，利用该图像作为纠偏的基准，利于实现更准确的图像配准。对于核酸序列测定得到的图像，利用多个图像进行参考图像构建，利于使得该参考图像获得完整的对应核酸分子的亮斑信息，利于基于亮斑的图像纠偏。

在一些实施例中，第五图像、第六图像分别来自核酸序列测定反应（测序反应）的不同时刻的同一个视野。在一个示例中，一轮测序反应包括多次碱基延伸反应，例如单色测序，利用的四种类型碱基对

应的反应底物(核苷酸类似物)均带有同一种荧光染料,一轮测序反应包括四次碱基延伸反应(4 repeats),对于一个视野来说,一次碱基延伸反应包含一次图像采集,第五图像和第六图像分别来自不同次的碱基延伸反应的同一视野。如此,通过处理以及集合第五图像和第六图像的信息获得的参考图像作为纠偏的基准,利于进行更准确的图像纠偏。

在另一个示例中,单分子双色测序反应,利用的四种类型碱基对应的反应底物(核苷酸类似物)中的两种带有一种荧光染料、另两种带有另一种不同激发波长的荧光染料,一轮测序反应包括两次碱基延伸反应,带有不同染料的两种类型碱基反应底物于一次碱基延伸反应中进行结合反应,对于一个视野,一次碱基延伸反应包括两次于不同激发波长下的图像采集,第五图像和第六图像分别来自不同次的碱基延伸反应或者同一次碱基延伸反应中的不同激发波长下的同一视野。如此,通过处理以及集合第五图像和第六图像的信息获得的参考图像作为纠偏的基准,利于进行更准确的图像纠偏。

在又一个示例中,一轮测序反应包括一次碱基延伸反应,例如二代测序平台的双色测序反应,四种类型碱基反应底物(例如核苷酸类似物)分别带有染料 a、带有染料 b、带有染料 a 和染料 b 以及不带任何染料,染料 a 和染料 b 的激发波长不一样;四种类型反应底物于同一次碱基延伸反应中实现一轮测序反应,第五图像和第六图像分别来自不同轮测序反应或者同一轮测序反应中的不同激发波长下的同一视野。如此,通过处理以及集合第五图像和第六图像的信息获得的参考图像作为纠偏的基准,利于进行更准确的图像纠偏。

第五图像和/或第六图像,可以是一个图像也可以是多个图像。在一个示例中,第五图像为第一图像,第六图像为第二图像。进一步地,在一些具体实施方式中,还包括利用第七图像和第八图像构建所称的参考图像,待配准图像、第五图像、第六图像、第七图像和第八图像来自测序反应的相同视野,第五图像、第六图像、第七图像和第八图像分别对应 A/U、T、G 和 C 四种类型碱基延伸反应时的视野,碱基延伸反应时的该视野存在多个带有光学可检测标记的核酸分子,至少一部分核酸分子在图像上表现为亮斑,构建参考图像还包括:基于第五图像对第七图像进行粗配准,包括确定第七图像和第五图像的偏移量,基于该偏移量移动第七图像,获得粗配准后的第七图像;基于第五图像对第八图像进行粗配准,包括确定第八图像和第五图像的偏移量,基于该偏移量移动第八图像,获得粗配准后的第八图像;合并第五图像和粗配准后的第六图像、粗配准后的第七图像以及粗配准后的第八图像,以获得参考图像。

本发明实施方式对第一配准的实现方式不作限制,例如可利用傅里叶变换,使用频域配准,确定第一偏移量。具体地,例如可参考 Kenji TAKITA et al, IEICE TRANS. FUNDAMENTALS, VOL. E86-A, NO.8 AUGUST 2003.中的纯相位相关函数(Phase-Only Correlation Function)中的二维离散傅里叶变换确定第一偏移量、第六图像和第五图像的偏移量、第七图像和第五图像的偏移量和/或第八图像和第五图像的偏移量。第一配准/粗配准可达到 1 像素(1 pixel)的精度。如此,可快速准确地确定第一偏移量和/或构建利于精确纠偏的参考图像。

在某些具体实施方式中,参考图像和待配准图像为二值化图像。如此,利于减少运算量快速纠偏。

在一个示例中,待纠偏图像和参考图像均为二值化图像,即图像中的各个像素非 a 即 b,例如 a 为 1, b 为 0,像素标记为 1 的较像素标记为 0 的亮,或者说强度大;参考图像是利用一轮测序反应的四次碱基延伸反应的图像 repeat1、repeat2、repeat3 和 repeat4 构建的,第五图像、第六图像选自图像 repeat1-4 中的任一个、两个或三个。

在一个示例中,第五图像为图像 repeat1,图像 repeat2、repeat3 和 repeat4 为第六图像,基于图像图像 repeat1 依次对图像 repeat2-4 进行粗配准,分别获得粗配准后的图像 repeat2-4;合并图像 repeat1 和粗配准后的图像 repeat2-4,获得参考图像。所称的合并图像为合并图像中的重合亮斑。主要基于对应核酸分子的亮斑的大小和成像系统分辨率,在一个示例中,设定两个图像上距离不大于 1.5 个像素的两个亮

斑为重合亮斑。这里，采用 4 个 repeat 的合成的图像中心区域作为参考图像，一来利于使得参考图像具有足够量的亮斑，利于后续配准，二来检测及定位出的图像中心区域中的亮斑，亮斑信息是相对更准确的，利于准确配准。

在一个示例中，进行如下步骤对图像进行纠偏：1) 对采集自另一轮测序反应的一次碱基延伸反应的某个视野的图像 repeat5 进行粗纠偏，repeat5 为二值化后的图像，取该图像中心例如 512*512 区域，与 repeat1-4 合成的中心图像（相应参考图像的中心 512*512 区域），进行二维离散傅里叶变换，使用频域配准，得到偏移量 $offset(x_0, y_0)$ ，即实现图像粗配准， x_0 、 y_0 能达到 1 pixel 的精度；2) 将上述粗配准后的图像和参考图像基于图像上的亮斑进行合并（merge），包括计算 repeat5 图像的中心区域内与参考图像相应区域内的重合亮斑的偏移量 $offset(x_1, y_1) = \text{待纠偏图像的该亮斑的坐标位置} - \text{参考图像上的相应亮斑的坐标位置}$ ，可表示为 $offset(x_1, y_1) = \text{curRepeatPoints} - \text{basePoints}$ ；求取所有重合亮斑的平均偏移量，从而得到 [0,0] 到 [1,1] 范围内的细偏移量。在一个示例中，设定两个图像上距离不大于 1.5 个像素的两个亮斑为重合亮斑；3) 综上，得到一个视野图像（fov）不同 cycle 的偏移量 $(x_0, y_0) - (x_1, y_1)$ ，对于一个亮斑（peak）可表示为： $\text{curRepeatPoints} + (x_0, y_0) - (x_1, y_1)$ ，curRepeatPoints 表示该亮斑原始坐标，即在纠偏前的图像中的坐标。上述图像纠偏获得的纠偏结果具有较高的准确性，且纠偏精度小于或等于 0.1 像素。图 3 示意纠偏过程及结果，图 3 中，基于图像 A 对图像 C 进行纠偏，图像 A 和图像 C 中的圆圈表示亮斑、相同数字标记的亮斑为重合亮斑，图像 C->A 表示纠偏结果，即图像 C 对齐至图像 A 的结果。

本发明的实施方式对图像上亮斑的识别检测方式不作限定。在某些具体实施方式中，进行图像配准还包括识别亮斑，包括利用 $k_1 * k_2$ 矩阵对图像进行亮斑检测，判定矩阵的中心像素值不小于矩阵非中心任一像素值的矩阵对应一个候选亮斑，以及确定候选亮斑是否为亮斑， k_1 和 k_2 均为大于 1 的奇数， $k_1 * k_2$ 矩阵包含 $k_1 * k_2$ 个像素点。所称的图像选自待配准图像、构建参考图像的图像中的至少一个。利用该方式检测图像上的亮斑，能够快速有效地实现图像上的亮斑（spots 或 peaks）的检测，特别是对采集自核酸序列测定反应的图像。该方法对待检测图像即原始输入数据没有特别的限制，适用于任何利用光学检测原理进行核酸序列测定的平台所产生的图像的处理分析，包括但不限于二代和三代测序，具有高准确性和高效的特点，能从图像中获取更多的代表序列的信息。特别是对于随机图像及高准确度要求的信号识别，尤其具有优势。

在一些实施例中，图像来自核酸序列测定反应，核酸分子上带有光学可检测标记，利如荧光标记，荧光分子在特定波长激光照射下能够被激发发出荧光，通过成像系统采集图像。采集到的图像包括可能与荧光分子所在位置相对应的光斑/亮斑。可以理解地，当处于焦面位置时，所采集到的图像中的与荧光分子所在位置相对应的亮斑的尺寸较小且亮度较高；当位于非焦面位置时，所采集到的图像中的与荧光分子所在位置相对应的亮斑的尺寸较大且亮度较低。另外，视野中的可能存在其它非目标或者后续难以利用的物质/信息，比如杂质等；进一步地，在对单分子视野进行拍照中，大量分子聚集（簇）等也会干扰目标单分子信息采集。所称的单分子为一个少数几个分子，例如分子数目不大于 10，例如为一个、两个、三个、四个、五个、六个、八个或者十个。

在一些示例中，矩阵的中心像素值大于第一预设值，矩阵非中心任一像素值大于第二预设值，第一预设值和第二预设值与图像的平均像素值相关。

在一些实施例中，可以利用 $k_1 * k_2$ 矩阵对图像进行遍历检测，所称的第一预设值和/或第二预设值的设置与该图像的平均像素值相关。对于灰度图像，像素值同灰度值。 $k_1 * k_2$ 矩阵， k_1 、 k_2 可以相等也可以不相等。在一个示例中，成像系统相关参数为：物镜 60 倍，电子传感器的尺寸为 $6.5 \mu\text{m}$ ，经过显微镜成的像再经过电子传感器，能看到的最小尺寸为 $0.1 \mu\text{m}$ ，获得的图像或者输入的图像可为 512*512、

1024*1024 或 2048*2048 的 16 位的灰度或彩色图像, k_1 和 k_2 的取值范围均为大于 1 且小于 10。在一个示例中, $k_1=k_2=3$; 在另一个示例中, $k_1=k_2=5$ 。若图像是彩色图像, 彩色图像的一个像素点具有三个像素值, 可以将彩色图像转化为灰度图像, 再进行亮斑检测, 以降低图像检测过程的计算量和复杂度。可选择但不限于利用浮点算法、整数方法、移位方法或平均值法等将非灰度图像转换成灰度图像。

在一个示例中, 发明人经过大量图像处理统计, 取第一预设值为该图像的平均像素的 1.4 倍, 取第二预设值为该图像的平均像素值的 1.1 倍, 能够排除干扰、获得来自于光学检测标记的亮斑检测结果。

可利用大小、与理想亮斑的相似程度和/或强度来对候选亮斑进一步进行筛选判断。在某些具体实施方式中, 利用候选亮斑对应的连通域的大小来定量反映比较图像上候选亮斑的大小, 以此来筛选判断候选亮斑是否为要的亮斑。

在一个示例中, 确定候选亮斑是否为亮斑包括: 计算一个候选亮斑对应的连通域的大小 $Area=A*B$, 判定对应的连通域的大小大于第三预设值的候选亮斑为一个亮斑, A 表示以该候选亮斑对应的矩阵的中心的所在行的相连像素/连通像素的大小, B 表示以该候选亮斑对应的矩阵的中心的所在列的相连像素/连通像素的大小, 定义一个 k_1*k_2 矩阵中大于平均像素值的相连像素为一个所称的候选亮斑对应的连通域。如此, 能够能够有效获得对应标记分子且符合后续序列识别的亮斑, 获得核酸序列信息。

在一个例子中, 以该图像的平均像素值作为基准, 相邻的不小于平均像素值的两个或多个像素为所称的相连像素/连通像素 (pixel connectivity), 如图 4 所示, 加粗加大的表示候选亮斑对应的矩阵的中心, 粗线框表示候选亮斑对应的 $3*3$ 矩阵, 标记为 1 的像素为不小于该图像的平均像素值的像素点, 标记为 0 的像素为小于平均像素值的像素点, 可看出 $A=3$, $B=6$, 该候选亮斑对应的连通域的大小为 $A*B=3*6$ 。

所称的第三预设值可依据该图像上所有候选亮斑对应的连通域的大小这一信息来确定。例如通过计算该图上各候选亮斑对应的连通域的大小, 取亮斑的连通域大小的平均值代表该图像一个特性, 作为第三预设值; 又例如, 可将该图像上各个候选亮斑对应的连通域大小按从小到大排序, 取第 50、第 60、第 70、第 80 或第 90 百分位数连通域大小作为该第三预设值。如此, 可有效获得亮斑信息, 利于后续识别核酸序列。

在某些示例中, 通过统计设置参数来定量反映比较候选亮斑的强度特征, 以此来筛选候选亮斑。在一个示例中, 确定候选亮斑是否为亮斑包括: 计算一个候选亮斑的分值 $Score=((k_1*k_2-1)CV-EV)/((CV+EV)/(k_1*k_2))$, 判定分值大于第四预设值的候选亮斑为一个亮斑, CV 表示候选亮斑对应的矩阵的中心像素值, EV 表示亮斑对应的矩阵的非中心像素值的总和。如此, 能够能够有效获得对应标记分子且符合后续序列识别的亮斑, 获得核酸序列信息。

所称的第四预设值可依据该图像上所有候选亮斑的分值的大小这一信息来确定。例如, 当该图像上的候选亮斑的数量大于一定数目符合统计上对量的要求, 例如该图像上候选亮斑的数目大于 30, 可计算且将该图像的所有候选亮斑的 $Score$ 值按升序排序, 第四预设值可设置为第 50、第 60、第 70、第 80 或 90 分位数 $Score$ 值, 如此, 可排除掉小于第 50、第 60、第 70、第 80 或第 90 分位数 $Score$ 值的候选亮斑, 利于有效获得目标亮斑, 利于后续碱基序列准确识别。进行该处理或者说该筛选设置的依据是, 一般地, 认为中心与边缘强度/像素值差异大且汇聚的亮斑为与待检分子所在位置相对应的亮斑。一般情况下, 图像上的候选亮斑的数量大于 50、大于 100 或大于 1000。

在某些示例中, 结合形态和强度/亮度对候选亮斑进行筛选。在一个示例中, 确定候选亮斑是否为亮斑包括: 计算一个候选亮斑对应的连通域的大小 $Area=A*B$, 以及计算一个候选亮斑的分值 $Score=((k_1*k_2-1)CV-EV)/((CV+EV)/(k_1*k_2))$, A 表示以该候选亮斑对应的矩阵的中心的所在行的相连像素/连通像素的大小, B 表示以该候选亮斑对应的矩阵的中心的所在列的相连像素/连通像素的大小, 定义一个 k_1*k_2 矩阵中大于平均像素值的相连像素为一个所称的候选亮斑对应的连通域, CV 表示候选亮

斑对应的矩阵的中心像素值，EV 表示亮斑对应的矩阵的非中心像素值的总和；判定对应的连通域的大小大于第三预设值且分值大于第四预设值的候选亮斑为一个亮斑。如此，能够有效地获得对应核酸分子且利于后续序列识别的亮斑信息。对于所称的第三预设值和/或第四预设值，可以参照前面具体实施方式进行考虑和设置。

在某些具体实施方式中，图像配准方法还包括亮斑识别检测，包括：预处理图像，获得预处理后的图像，所称的图像选自第一图像、第二图像、第三图像、第四图像、第五图像、第六图像、第七图像和第八图像中的至少一个；确定临界值以简化预处理后的图像，包括对小于临界值的预处理后的图像上的像素点的像素值赋值为第一预设值，对不小于临界值的预处理后的图像上的像素点的像素值赋值为第二预设值，以获得简化图像；基于预处理后的图像确定第一亮斑检测阈值 $c1$ ；基于预处理后的图像和简化图像识别图像上的候选亮斑，包括判定满足以下 a) -c) 中至少两个条件的像素点矩阵为一个候选亮斑，a) 在预处理后的图像中，像素点矩阵的中心像素点的像素值为最大，像素点矩阵可表示为 $r1*r2$ ， $r1$ 和 $r2$ 均为大于 1 的奇数， $r1*r2$ 像素点矩阵包含 $r1*r2$ 个像素点，b) 在简化图像中，像素点矩阵的中心像素点的像素值为第二预设值并且像素点矩阵的连通像素大于 $2/3*r1*r2$ ，以及 c) 在预处理后的图像中的像素点矩阵的中心像素点的像素值大于第三预设值，并且满足 $g1*g2 > c1$ ， $g1$ 为以像素点矩阵的中心像素点为中心的 $m1*m2$ 范围的二维高斯分布的相关系数， $g2$ 为该 $m1*m2$ 范围的像素， $m1$ 和 $m2$ 均为大于 1 的奇数， $m1*m2$ 范围包含 $m1*m2$ 个像素点；以及确定候选亮斑是否为亮斑。利用该方式检测图像上的亮斑，包括利用发明人通过大量数据训练确定的判断条件或判断条件的组合，能够快速有效地实现图像上的亮斑的检测，特别是对采集自核酸序列测定反应的图像。该方法对待检测图像即原始输入数据没有特别的限制，适用于任何利用光学检测原理进行核酸序列测定的平台所产生的图像的处理分析，包括但不限于二代和三代测序，具有高准确性和高效的特点，能从图像中获取更多的代表序列的信息。特别是对于随机图像及高准确度要求的信号识别，尤其具有优势。

对于灰度图像，像素值同灰度值。若图像是彩色图像，彩色图像的一个像素点具有三个像素值，可以将彩色图像转化为灰度图像，再进行亮斑检测，以降低图像检测过程的计算量和复杂度。可选择但不限于利用浮点算法、整数方法、移位方法或平均值法等将非灰度图像转换成灰度图像。

在一些实施例中，预处理图像包括：利用开运算确定图像的背景；基于背景，利用顶帽运算将图像转化为第一图像；对第一图像进行高斯模糊处理，获得第二图像；对第二图像进行锐化，以获得所称的预处理后的图像。如此，能对图像进行有效的降噪或者说提高图像的信噪比，利于亮斑的准确检测。

开运算是一种形态学处理，即先膨胀后腐蚀的过程，腐蚀操作会使得前景（感兴趣的部分）变小，而膨胀会使得前景变大；开运算可以用来消除小物体，在纤细点处分离物体，并且在平滑较大物体的边界的同时不明显改变其面积。该实施方式对图像做开运算的结构元 $p1*p2$ （用来处理图像的基本模板）的大小不作特别限制， $p1$ 和 $p2$ 为奇数。在一个示例中，结构元 $p1*p2$ 可以为 $15*15$ 、 $31*31$ 等，最终都能够获得利于后续处理分析的预处理后的图像。

顶帽运算往往用来分离比临近点（亮点/亮斑）亮一些的斑块，在一幅图像具有大幅的背景，而微小物品比较有规律的情况下，可以使用顶帽运算进行背景提取。在一个示例中，对图像进行顶帽变换包括先对图像做开运算，进而利用原图像减去开运算结果，获得第一图像即顶帽变换后的图像。顶帽变换的数学表达式为 $dst = tophat(src, element) = src - open(src, element)$ 。发明人认为，开运算的结果放大了裂缝或者局部低亮度的区域，因此从原图中减去开运算后的图，得到的图像突出了比原图轮廓周围的区域更明亮的区域，这一操作与选择的核的大小相关，可以认为与亮点/亮斑的预期大小相关，若亮点不是预期大小，处理后的效果会使得整张图产生许多小凸起，具体可以参考虚焦图片，即亮点/亮斑晕染成一团。在一个示例中，亮点的预期大小即选择的核的大小为 $3*3$ ，得到的顶帽变换后的图像利于后续进一步去噪处理。

高斯模糊 (Gaussian Blur) 也称为高斯滤波, 是一种线性平滑滤波, 适用于消除高斯噪声, 广泛应用于图像处理的减噪过程。通俗的讲, 高斯滤波就是对整幅图像进行加权平均的过程, 每一个像素点的值, 都由其本身和邻域内的其他像素值经过加权平均后得到。高斯滤波的具体操作是: 用一个模板 (或称卷积、掩模) 扫描图像中的每一个像素, 用模板确定的邻域内像素的加权平均灰度值去替代模板中心像素点的值。在一个示例中, 对第一图像进行高斯模糊处理, 在 OpenCV 中使用高斯滤波 GaussianBlur 函数进行, 高斯分布参数 Sigma 取 0.9, 所使用的二维滤波器矩阵 (卷积核) 是 3*3, 从图像角度看经过该高斯模糊处理后, 第一图像上的小突起被抹平, 图像边缘光滑。进一步地, 对第二图像即高斯过滤后的图像进行锐化, 例如进行二维拉普拉斯锐化, 从图像角度看经过处理后, 边缘被锐化, 高斯模糊后的图像得以恢复。

在一些实施例中, 简化预处理后的图像包括: 基于背景和预处理后的图像, 确定临界值; 比较预处理后的图像上的像素点的像素值与临界值, 对小于临界值的预处理后的图像上的像素点的像素值赋值为第一预设值, 对不小于临界值的预处理后的图像上的像素点的像素值赋值为第二预设值, 获得简化图像。如此, 根据发明人大量测试数据总结的确定临界值的方式以及确定的临界值, 据此将预处理后的图像简化, 例如二值化, 利于后续亮斑准确检测, 利于后续碱基准确识别、获得高质量数据等。

具体地, 在一些示例中, 获得简化图像包括: 将预处理后获得的锐化后的结果除以开运算结果, 获得和图像像素点对应的一组数值; 通过该组数值, 确定二值化预处理后的图像的临界值。例如, 可将该组数值按大小升序排列, 取该组数值中第 20、30 或 40 百分位数对应的数值作为二值化临界值/阈值。如此, 获得的二值化图像利于后续亮斑的准确检测识别。

在一个示例中, 图像预处理时的开运算的结构元为 $p1*p2$, 所称的将预处理后的图像 (锐化后的结果) 除以开运算结果, 获得一组和结构元一样大小的数组/矩阵 $p1*p2$, 在每个数组中, 将该数组包含的 $p1*p2$ 个数值按大小升序排列, 取该数组中第三十百分位数对应的数值作为该区域 (数值矩阵) 的二值化临界值/阈值, 如此, 分别确定阈值对图像上的各个区域进行二值化, 最终获得的二值化结果在去噪的同时更加突出所需信息, 利于后续亮斑的准确检测。

在一些示例中, 利用大津法进行第一亮斑检测阈值的确定。大津法 (OTSU 算法) 也可称为最大类间方差法, 大津法利用类间方差最大来分割图像, 意味着错分概率小, 准确性高。假设预处理后的图像的前景和背景的分割阈值为 T ($c1$), 属于前景的像素点数占整幅图像的比例为 w_0 , 其平均灰度为 μ_0 ; 属于背景的像素点数占整幅图像的比例为 w_1 , 其平均灰度为 μ_1 。待处理图像的总平均灰度记为 μ , 类间方差记为 var , 则有:

$\mu = \omega_0 * \mu_0 + \omega_1 * \mu_1$; $var = \omega_0 (\mu_0 - \mu)^2 + \omega_1 (\mu_1 - \mu)^2$, 将后者代入前者, 得到等价公式:
 $var = \omega_0 \omega_1 (\mu_1 - \mu_0)^2$ 。采用遍历的方法得到使类间方差最大的分割阈值 T , 即为所求的第一亮斑检测阈值 $c1$ 。

在一些实施例中, 基于预处理后的图像和简化图像识别图像上的候选亮斑, 包括判断同时满足 a) -c) 三个条件的像素点矩阵为一个候选亮斑。如此, 能有效地提高后续基于亮斑信息确定核酸序列的准确性和下机数据的质量。

具体地, 在一个示例中, 候选亮斑的判定需要满足的条件包括 a), $k1$ 、 $k2$ 可以相等也可以不相等。在一个示例中, 成像系统相关参数为: 物镜 60 倍, 电子传感器的尺寸为 $6.5\mu m$, 经过显微镜成的像再经过电子传感器, 能看到的最小尺寸为 $0.1\mu m$, 获得的图像或者输入的图像可为 $512*512$ 、 $1024*1024$ 或 $2048*2048$ 的 16 位的灰度或彩色图像, $k1$ 和 $k2$ 的取值范围均为大于 1 且小于 10。在一个示例中, 在一个预处理后的图像中, 依据亮斑的预期大小设置 $k1=k2=3$; 在另一个示例中, 设置 $k1=k2=5$ 。

在一个示例中，候选亮斑的判定需要满足的条件包括 b)，在简化图像中，像素点矩阵的中心像素点的像素值为第二预设值，并且该像素点矩阵的连通像素大于 $2/3 * k1 * k2$ ，即中心像素点的像素值大于临界值且连通像素大于矩阵的三分之二。这里，相邻的像素值都为第二预设值的两个或多个像素为所称的相连像素/连通像素 (pixel connectivity)，例如，简化图像为二值化图像，第一预设值为 0，第二预设值为 1，如图 4 所示，加粗加大的表示所称的像素点矩阵的中心，粗线框表示像素点矩阵 $3 * 3$ ，即 $k1 = k2 = 3$ ，该矩阵的中心像素点的像素值为 1，连通像素为 4，小于 $2/3 * k1 * k2 = 6$ ，该像素点矩阵不满足条件 b)，非候选亮斑。

在一个示例中，候选亮斑的判定需要满足的条件包括 c)，在预处理图像中， g_2 为修正后的 $m1 * m2$ 范围的像素，即为修正后的 $m1 * m2$ 范围像素总和。在一个例子中，依据简化图像相应 $m1 * m2$ 范围中像素值为第二预设值的像素点所占的比例进行修正，例如，如图 5 所示， $m1 = m2 = 5$ ，所称的简化图像相应 $m1 * m2$ 范围中像素值为第二预设值的像素点所占的比例为 $13/25$ (13 个“1”)，修正后的 g_2 为原来的 $13/25$ 。如此，利于更准确的检测识别亮斑，利于后续亮斑信息的分析读取。

在一些示例中，所称的判定候选亮斑是否为亮斑还包括：基于预处理后的图像确定第二亮斑检测阈值，以及判定像素值不小于第二亮斑检测阈值的候选亮斑为亮斑。在具体示例中，以候选亮斑的坐标所在的像素点的像素值作为该候选亮斑的像素值。通过利用基于预处理后的图像确定的第二亮斑检测阈值对候选亮斑的进一步筛选，能够排除掉至少一部分更可能是图像背景但亮度(强度)和/或形状表现为“亮斑”的亮斑，利于后续基于亮斑的序列的准确识别，提高下机数据的质量。

在一个示例中，可利用重心法获取候选亮斑的坐标，包括亚像素级坐标。利用双线性插值法计算候选亮斑的坐标位置的灰度值。

在某些具体示例中，判定候选亮斑是否为亮斑包括：将预处理后的图像划分为预定大小的一组区域 (block)，对该区域中的像素点的像素值进行排序，以确定该区域对应的第二亮斑检测阈值；对于位于区域的候选亮斑，判定像素值不小于该区域对应的第二亮斑检测阈值的候选亮斑为亮斑。如此，区分图像的不同区域的差异比如光强的整体落差，分开进行亮斑的进一步检测识别，利于准确识别亮斑并且获得更多的亮斑。

所称的将预处理后的图像划分为预定大小的一组区域 (block)，block 之间可以有重叠也可以没有重叠。在一个示例中，block 之间没有重叠。在一些实施例中，预处理后的图像的大小不小于 $512 * 512$ ，例如为 $512 * 512$ 、 $1024 * 1024$ 、 $1800 * 1800$ 或者 $2056 * 2056$ 等，所称预定大小的区域可以设置为 $200 * 200$ 。如此，利于快速计算判断识别亮斑。

在一些实施例中，确定该区域对应的第二亮斑检测阈值时，对每个 block 中的像素点的像素值按大小进行升序排列，取 $p10 + (p10 - p1) * 4.1$ 作为该 block 对应的第二亮斑检测阈值，即该 block 的背景， $p1$ 表示第百分之一分位的像素值， $p10$ 表示第百分之十分位的像素值。该阈值是发明人通过大量数据训练测试得出的较为稳定的阈值，能够消除大量背景上的亮斑。可以理解地，当光学系统调整，图像整体像素分布发生改变时，此阈值可能需要适当调整。图 6 为进行该处理之前和之后的亮斑检测结果对比示意图，即排除掉区域背景前后的亮斑检测结果示意图，图 6 的上半部分为作该处理后的亮斑检测结果、下半部分为不作该处理的亮斑检测结果，十字标记的为候选亮斑或亮斑。

本发明的实施方式还提供一种碱基识别方法，包括将获自碱基延伸反应的图像上的亮斑匹配到对应测序模板的亮斑集合，依据匹配上的亮斑进行碱基识别，获自碱基延伸反应的图像对应的视野中存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在获自碱基延伸反应的图像上表现为亮斑，对应测序模板的亮斑集合通过上述任一实施例中的基于图像构建测序模板的方法获得。

上述对任一实施方式中的基于图像构建测序模板的方法的技术特征和优点的描述，同样适用本发明这一实施方式中的碱基识别方法，在此不再赘述。

具体地，可以利用遍历的方式将获自碱基延伸反应的图像上的亮斑与构建的亮斑集合进行匹配。在某些具体实施方式中，对应测序模板的亮斑集合中存在与获自碱基延伸反应的图像上的任一亮斑的距离小于第三预定像素，则判定获自碱基延伸反应的图像上的该亮斑匹配上对应测序模板的亮斑集合。在一个示例中，所称的第三预定像素为 2。如此，能够实现碱基的准确识别，获得模板的部分碱基序列（读段）。

上述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序列列表，可以具体实现在任何计算机可读存储介质中，以供指令执行系统、装置或设备（如基于计算机的系统、包括处理器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，“计算机可读存储介质”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。计算机可读存储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（电子装置），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），可擦除可编程只读存储器（EPROM 或闪存存储器），光纤装置，以及便携式光盘只读存储器（CDROM）。另外，计算机可读存储介质甚至可以是可在其上打印所述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所述程序，然后将其存储在计算机存储器中。

本发明的实施方式还提供一种基于图像构建测序模板的装置 100，如图 7 所示，用于实施上述本发明任一实施例中的基于图像构建测序模板的方法，所称的图像包括分别对应 A/U、T、G 和 C 四种碱基延伸反应时的一个相同视野的第一图像、第二图像、第三图像和第四图像，碱基延伸反应时的该视野存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在图像上表现为亮斑，定义顺序和/或同时实现一次四种类型碱基延伸反应为一轮测序反应，第一图像包括图像 M1 和图像 M2，第二图像包括图像 N1 和图像 N2，第三图像包括图像 P1 和图像 P2，第四图像包括图像 Q1 和图像 Q2，图像 M1 和图像 M2 分别来自两轮测序反应，图像 N1 和图像 N2 分别来自两轮测序反应，图像 P1 和图像 P2 分别来自两轮测序反应，图像 Q1 和图像 Q2 分别来自两轮测序反应，该装置包括：组合单元 110，用于组合图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 中的任两图像以进行亮斑匹配，并且使图像 M1、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 均至少一次参与该组合，获得包含第一重合亮斑的多个组合图像，在组合图像上距离小于第一预定像素的两个或多个亮斑为一个第一重合亮斑；合并单元 130，用于合并来自所述组合单元的多个组合图像上的第一重合亮斑，以获得一个对应测序模板的亮斑集合。

上述对本发明任一实施例中的基于图像构建测序模板的方法的技术特征和优点的描述，同样适用本发明这一实施方式中的装置 100，在此不再赘述。

例如，在合并单元 130 中，合并多个组合图像上的第一重合亮斑，包括对不同组合图像中的第一重合亮斑进行一次或多次匹配，以获得对应测序模板的亮斑集合。

在一些示例中，图像 M1、图像 N1、图像 P1 和图像 Q1 为顺序获得，图像 M2、图像 N2、图像 P2 和图像 Q2 为顺序获得，组合单元 130 用于：间隔 S 个图像对图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 进行两两组合，获得 K 个组合图像以及对组合图像上的亮斑进行匹配，弃组合图像上的非重合亮斑，S 为整数， $0 \leq S \leq S_{\max}$ ， $S_{\max} = \text{参与组合的图像总数} - 4$ ， $K = [(\text{参与组合的图像总数} - S - 1) + 1] * (\text{参与组合的图像总数} - S - 1) / 2$ 。

在一些示例中，图像为经过配准的图像。

具体地，装置 100 还包括配准单元 108，配准单元用于图像配准，配准单元包括第一配准模块和第二配准模块，第一配准模块用于基于参考图像对待配准图像进行第一配准，参考图像和待配准图像对应相同视野，包括确定待配准图像上的预定区域和参考图像上的相应预定区域的第一偏移量，基于第一偏移量移动待配准图像上的所有亮斑，获得第一配准后的待配准图像；第二配准模块用于基于参考图像对第一配准后的待配准图像进行第二配准，包括合并第一配准后的待配准图像和参考图像，获得合并图像，计算合并图像上的预定区域的所有第二重合亮斑的偏移量，以确定第二偏移量，在合并图像上的距离小于第二预定像素的两个或多个亮斑为一个第二重合亮斑，以及基于该第二偏移量移动第一配准后的待配准图像上的所有亮斑，以实现对待配准图像的配准。

在一些示例中，参考图像通过构建获得，配准单元 108 还包括参考图像构建模块，参考图像构建模块用于：获取第五图像和第六图像，第五图像和第六图像与待配准图像对应相同视野；基于第五图像对第六图像进行粗配准，包括确定第六图像和第五图像的偏移量，基于该偏移量移动第六图像，获得粗配准后的第六图像；合并第五图像和粗配准后的第六图像，以获得参考图像。

在一些示例中，在利用参考图像构建模块构建参考图像时，还包括利用第七图像和第八图像，第七图像和第八图像与待配准图像来自测序反应的不同视野，第五图像、第六图像、第七图像和第八图像分别对应 A/U、T、G 和 C 四种类型碱基延伸反应时的视野，构建参考图像还包括：基于第五图像对第七图像进行粗配准，包括确定第七图像和第五图像的偏移量，基于该偏移量移动第七图像，获得粗配准后的第七图像；基于第五图像对第八图像进行粗配准，包括确定第八图像和第五图像的偏移量，基于该偏移量移动第八图像，获得粗配准后的第八图像；合并第五图像和粗配准后的第六图像、粗配准后的第七图像以及粗配准后的第八图像，以获得参考图像。

在一些示例中，参考图像和待配准图像为二值化图像。

在一些示例中，利用二维离散傅里叶变换确定第一偏移量、第六图像和第五图像的偏移量、第七图像和第五图像的偏移量和/或第八图像和第五图像的偏移量。

在一些示例中，装置 100 还包括亮斑检测单元 106，亮斑检测单元 106 用于：预处理图像，获得预处理后的图像；确定临界值以简化预处理后的图像，包括对小于临界值的预处理后的图像上的像素点的像素值赋值为第一预设值，对不小于临界值的预处理后的图像上的像素点的像素值赋值为第二预设值，以获得简化图像；基于预处理后的图像确定第一亮斑检测阈值 c_1 ；基于预处理后的图像和简化图像识别图像上的候选亮斑，包括判定满足以下 a) -c) 中至少两个条件的像素点矩阵为一个候选亮斑，a) 在预处理后的图像中，像素点矩阵的中心像素点的像素值为最大，像素点矩阵可表示为 $k_1 * k_2$ ， k_1 和 k_2 均为大于 1 的奇数， $k_1 * k_2$ 像素点矩阵包含 $k_1 * k_2$ 个像素点，b) 在简化图像中，像素点矩阵的中心像素点的像素值为第二预设值并且像素点矩阵的连通像素大于 $2/3 * k_1 * k_2$ ，以及 c) 在预处理后的图像中的像素点矩阵的中心像素点的像素值大于第三预设值，并且满足 $g_1 * g_2 > c_1$ ， g_1 为以像素点矩阵的中心像素点为中心的 $m_1 * m_2$ 范围的二维高斯分布的相关系数， g_2 为该 $m_1 * m_2$ 范围的像素， m_1 和 m_2 均为大于 1 的奇数， $m_1 * m_2$ 范围包含 $m_1 * m_2$ 个像素点。

在一些示例中，亮斑检测单元 106 还包括用于判定候选亮斑是否为亮斑，包括：基于预处理后的图像确定第二亮斑检测阈值，以及判定像素值不小于第二亮斑检测阈值的候选亮斑为亮斑。

在一些示例中，候选亮斑的像素值为该候选亮斑的坐标所在的像素点的像素值。

在一些示例中，在亮斑检测单元 106 中判定候选亮斑是否为亮斑包括：将预处理后的图像划分为预定大小的一组区域，对该区域中的像素点的像素值进行排序，以确定该区域对应的第二亮斑检测阈值，对于位于区域的候选亮斑，判定像素值不小于该区域对应的第二亮斑检测阈值的候选亮斑为亮斑。

在一些示例中，在亮斑检测单元 106 中预处理图像，包括：利用开运算确定图像的背景，基于背景，利用顶帽运算将图像转化为第一图像，对第一图像进行高斯模糊处理，获得第二图像，对第二图像进行锐化，获得预处理后的图像。

在一些示例中，在亮斑检测单元 106 中确定临界值以简化预处理后的图像，获得简化图像，包括：基于背景和预处理后的图像，确定临界值，比较预处理后的图像上的像素点的像素值与临界值，以获得简化图像。

在一些示例中， g_2 为修正后的 m_1*m_2 范围的像素，依据简化图像相应 m_1*m_2 范围中像素值为第二预设值的像素点所占的比例进行修正。

本发明的实施方式还提供一种碱基识别装置 1000，该装置用以实现上述本发明任一具体实施方式中的碱基识别方法，该装置 1000 用于将获自碱基延伸反应的图像上的亮斑匹配到对应测序模板的亮斑集合，依据匹配上的亮斑进行碱基识别，获自碱基延伸反应的图像对应的视野中存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在获自碱基延伸反应的图像上表现为亮斑，对应测序模板的亮斑集合通过上述任一实施例中的基于图像构建测序模板的方法和/或基于图像构建测序模板的装置构建。

具体地，在碱基识别装置 1000 中，对应测序模板的亮斑集合中存在与获自碱基延伸反应的图像上的任一亮斑的距离小于第三预定像素，则判定获自碱基延伸反应的图像上的该亮斑匹配上对应测序模板的亮斑集合。

依据本发明的实施方式，还提供一种计算机程序产品，该产品包括实现基于图像构建测序模板的指令，指令在计算机执行程序时，使计算机执行上述本发明任一具体实施方式中的基于图像构建测序模板的方法。

依据本发明的实施方式，还提供另一种计算机程序产品，该产品包括实现碱基识别的指令，指令在计算机执行程序时，使计算机执行上述本发明任一具体实施方式中的碱基识别方法。

本领域技术人员知晓，除了以纯计算机可读程序代码方式实现控制器/处理器外，完全可以通过将方法步骤进行逻辑变过来使得控制器以逻辑门、开关、专用集成电路、可编辑逻辑控制器和嵌入微控制器等的形式来实现相同的功能。因此，这种控制器/处理器可以被认为是一种硬件部件，而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

在本说明书的描述中，一个实施方式、一些实施方式、一个或一些具体实施方式、一个或一些实施例、示例等的描述意指结合该实施方式或示例描述的具体特征、结构或者特点包含于本发明的至少一个实施例或示例中。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例或示例。而且，描述的具体特征、结构等特点可以在任何一个或多个实施例或示例中以合适的方式结合。

尽管已经示出和描述了本发明的实施例，本领域的普通技术人员可以理解：在不脱离本发明的原理和宗旨的情况下可以对这些实施例进行多种变化、修改、替换和变型，本发明的范围由权利要求及其等同限定。

权 利 要 求

1.一种基于图像构建测序模板的方法，其特征在于，图像包括分别对应 A/ U、T、G 和 C 四种碱基延伸反应时的一个相同视野的第一图像、第二图像、第三图像和第四图像，碱基延伸反应时的该视野存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在图像上表现为亮斑，定义顺序和/或同时实现一次四种类型碱基延伸反应为一轮测序反应，

第一图像包括图像 M1 和图像 M2，第二图像包括图像 N1 和图像 N2，第三图像包括图像 P1 和图像 P2，第四图像包括图像 Q1 和图像 Q2，

图像 M1 和图像 M2 分别来自两轮测序反应，图像 N1 和图像 N2 分别来自两轮测序反应，图像 P1 和图像 P2 分别来自两轮测序反应，图像 Q1 和图像 Q2 分别来自两轮测序反应，方法包括：

组合图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 中的任两图像以进行亮斑匹配，并且使图像 M1、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 均至少一次参与组合，获得包含第一重合亮斑的多个组合图像，在组合图像上距离小于第一预定像素的两个或多个亮斑为一个第一重合亮斑；

合并多个组合图像上的第一重合亮斑，以获得一个对应测序模板的亮斑集合。

2.权利要求 1 的方法，其特征在于，合并多个组合图像上的第一重合亮斑，包括对不同组合图像中的第一重合亮斑进行一次或多次匹配，以获得对应测序模板的亮斑集合。

3.权利要求 1 的方法，其特征在于，图像 M1、图像 N1、图像 P1 和图像 Q1 为顺序获得，图像 M2、图像 N2、图像 P2 和图像 Q2 为顺序获得，

组合图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 中的任两图像以进行亮斑匹配，并且使图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 均至少一次参与组合，获得包含第一重合亮斑的多个组合图像，包括：

间隔 S 个图像对图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 进行两两组合，获得 K 个组合图像以及对组合图像上的亮斑进行匹配，弃组合图像上的非重合亮斑，S 为整数， $0 \leq S \leq S_{\max}$ ， $S_{\max} = \text{参与组合的图像总数} - 4$ 。

4.权利要求 1-3 任一方法，其特征在于，图像为经过配准的图像。

5.权利要求 4 的方法，其特征在于，配准图像，包括：

基于参考图像对待配准图像进行第一配准，参考图像和待配准图像对应相同视野，包括，

确定待配准图像上的预定区域和参考图像上的相应预定区域的第一偏移量，基于第一偏移量移动待配准图像上的所有亮斑，获得第一配准后的待配准图像；

基于参考图像对第一配准后的待配准图像进行第二配准，包括，

合并第一配准后的待配准图像和参考图像，获得合并图像，

计算合并图像上的预定区域的所有第二重合亮斑的偏移量，以确定第二偏移量，在合并图像上的距离小于第二预定像素的两个或多个亮斑为一个第二重合亮斑，

基于该第二偏移量移动第一配准后的待配准图像上的所有亮斑，以实现对待配准图像的配准。

6.权利要求 5 的方法，其特征在于，参考图像通过构建获得，构建参考图像包括：

获取第五图像和第六图像，第五图像和第六图像与待配准图像对应相同视野；

基于第五图像对第六图像进行粗配准，包括确定第六图像和第五图像的偏移量，基于该偏移量移动第六图像，获得粗配准后的第六图像；

合并第五图像和粗配准后的第六图像，以获得参考图像。

7.权利要求 6 的方法，其特征在于，构建参考图像还包括利用第七图像和第八图像，第七图像和第八图像与待配准图像来自测序反应的相同视野，第五图像、第六图像、第七图像和第八图像分别对应 A/U、T、G 和 C 四种类型碱基延伸反应时的视野，构建参考图像还包括：

基于第五图像对第七图像进行粗配准，包括确定第七图像和第五图像的偏移量，基于该偏移量移动第七图像，获得粗配准后的第七图像；

基于第五图像对第八图像进行粗配准，包括确定第八图像和第五图像的偏移量，基于该偏移量移动第八图像，获得粗配准后的第八图像；

合并第五图像和粗配准后的第六图像、粗配准后的第七图像以及粗配准后的第八图像，以获得参考图像。

8.权利要求 5-7 任一方法，其特征在于，参考图像和待配准图像为二值化图像。

9.权利要求 5-8 任一方法，其特征在于，利用二维离散傅里叶变换确定第一偏移量、第六图像和第五图像的偏移量、第七图像和第五图像的偏移量和/或第八图像和第五图像的偏移量。

10.权利要求 1-9 任一方法，其特征在于，还包括检测图像上的亮斑，包括：

预处理图像，获得预处理后的图像；

确定临界值以简化预处理后的图像，包括对小于临界值的预处理后的图像上的像素点的像素值赋值为第一预设值，对不小于临界值的预处理后的图像上的像素点的像素值赋值为第二预设值，以获得简化图像；

基于预处理后的图像确定第一亮斑检测阈值 c_1 ；

基于预处理后的图像和简化图像识别图像上的候选亮斑，包括判定满足以下 a) -c) 中至少两个条件的像素点矩阵为一个候选亮斑，

a) 在预处理后的图像中，像素点矩阵的中心像素点的像素值为最大，像素点矩阵可表示为 $r_1 * r_2$ ， r_1 和 r_2 均为大于 1 的奇数， $r_1 * r_2$ 像素点矩阵包含 $r_1 * r_2$ 个像素点，

b) 在简化图像中，像素点矩阵的中心像素点的像素值为第二预设值并且像素点矩阵的连通像素大于 $2/3 * r_1 * r_2$ ，以及

c) 在预处理后的图像中的像素点矩阵的中心像素点的像素值大于第三预设值，并且满足 $g_1 * g_2 > c_1$ ， g_1 为以像素点矩阵的中心像素点为中心的 $m_1 * m_2$ 范围的二维高斯分布的相关系数， g_2 为该 $m_1 * m_2$ 范围的像素， m_1 和 m_2 均为大于 1 的奇数， $m_1 * m_2$ 范围包含 $m_1 * m_2$ 个像素点。

11.权利要求 10 的方法，其特征在于，还包括判定候选亮斑是否为亮斑，包括：

基于预处理后的图像确定第二亮斑检测阈值，以及

判定像素值不小于第二亮斑检测阈值的候选亮斑为亮斑。

12.权利要求 11 的方法，其特征在于，候选亮斑的像素值为该候选亮斑的坐标所在的像素点的像素值。

13.权利要求 11 或 12 的方法，其特征在于，判定候选亮斑是否为亮斑包括：

将预处理后的图像划分为预定大小的一组区域，

对该区域中的像素点的像素值进行排序，以确定该区域对应的第二亮斑检测阈值，

对于位于区域的候选亮斑，判定像素值不小于该区域对应的第二亮斑检测阈值的候选亮斑为亮斑。

14.权利要求 10-13 任一方法，其特征在于，预处理图像，包括：

利用开运算确定图像的背景，

基于背景，利用顶帽运算将图像转化为第一图像，

对第一图像进行高斯模糊处理，获得第二图像，

对第二图像进行锐化，获得预处理后的图像。

15.权利要求 14 的方法，其特征在于，确定临界值以简化预处理后的图像，获得简化图像，包括：

基于背景和预处理后的图像，确定临界值，

比较预处理后的图像上的像素点的像素值与临界值，以获得简化图像。

16.权利要求 10-15 任一方法，其特征在于， g_2 为修正后的 m_1*m_2 范围的像素，依据简化图像相应 m_1*m_2 范围中像素值为第二预设值的像素点所占的比例进行修正。

17.一种碱基识别方法，其特征在于，包括将获自碱基延伸反应的图像上的亮斑匹配到对应测序模板的亮斑集合，依据匹配上的亮斑进行碱基识别，获自碱基延伸反应的图像对应的视野中存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在获自碱基延伸反应的图像上表现为亮斑，对应测序模板的亮斑集合通过权利要求 1-16 任一方法构建。

18.权利要求 17 的方法，其特征在于，对应测序模板的亮斑集合中存在与获自碱基延伸反应的图像上的任一亮斑的距离小于第三预定像素，则判定获自碱基延伸反应的图像上的该亮斑匹配上对应测序模板的亮斑集合。

19.一种基于图像构建测序模板的装置，其特征在于，所述图像包括分别对应 A/ U、T、G 和 C 四种碱基延伸反应时的一个相同视野的第一图像、第二图像、第三图像和第四图像，碱基延伸反应时的该视野存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在图像上表现为亮斑，定义顺序和/或同时实现一次四种类型碱基延伸反应为一轮测序反应，

第一图像包括图像 M1 和图像 M2，第二图像包括图像 N1 和图像 N2，第三图像包括图像 P1 和图像 P2，第四图像包括图像 Q1 和图像 Q2，

图像 M1 和图像 M2 分别来自两轮测序反应，图像 N1 和图像 N2 分别来自两轮测序反应，图像 P1 和图像 P2 分别来自两轮测序反应，图像 Q1 和图像 Q2 分别来自两轮测序反应，该装置包括：

组合单元，用于组合图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 中的任两图像以进行亮斑匹配，并且使图像 M1、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 均至少一次参与该组合，获得包含第一重合亮斑的多个组合图像，在组合图像上距离小于第一预定像素的两个或多个亮斑为一个第一重合亮斑；

合并单元，用于合并来自所述组合单元的多个组合图像上的第一重合亮斑，以获得一个对应测序模板的亮斑集合。

20.权利要求 19 的装置，其特征在于，在所述合并单元中，合并多个组合图像上的第一重合亮斑，包括对不同组合图像中的第一重合亮斑进行一次或多次匹配，以获得对应测序模板的亮斑集合。

21.权利要求 19 的装置，其特征在于，图像 M1、图像 N1、图像 P1 和图像 Q1 为顺序获得，图像 M2、图像 N2、图像 P2 和图像 Q2 为顺序获得，所述组合单元用于：

间隔 S 个图像对图像 M1、图像 M2、图像 N1、图像 N2、图像 P1、图像 P2、图像 Q1 和图像 Q2 进行两两组合，获得 K 个组合图像以及对组合图像上的亮斑进行匹配，弃组合图像上的非重合亮斑， S 为整数， $0 \leq S \leq S_{\max}$ ， S_{\max} =参与组合的图像总数-4。

22.权利要求 19-21 任一装置，其特征在于，所述图像为经过配准的图像。

23.权利要求 22 的装置，其特征在于，还包括配准单元，所述配准单元用于图像配准，所述配准单元包括第一配准模块和第二配准模块，

所述第一配准模块用于基于参考图像对待配准图像进行第一配准，参考图像和待配准图像对应相同视野，包括确定待配准图像上的预定区域和参考图像上的相应预定区域的第一偏移量，基于第一偏移量移动待配准图像上的所有亮斑，获得第一配准后的待配准图像；

所述第二配准模块用于基于参考图像对第一配准后的待配准图像进行第二配准，包括合并第一配准后的待配准图像和参考图像，获得合并图像，计算合并图像上的预定区域的所有第二重合亮斑的偏移量，以确定第二偏移量，在合并图像上的距离小于第二预定像素的两个或多个亮斑为一个第二重合亮斑，以及基于该第二偏移量移动第一配准后的待配准图像上的所有亮斑，以实现对待配准图像的配准。

24.权利要求 22 的装置，其特征在于，所述参考图像通过构建获得，所述配准单元还包括参考图像构建模块，所述参考图像构建模块用于：

获取第五图像和第六图像，第五图像和第六图像与待配准图像对应相同视野；

基于第五图像对第六图像进行粗配准，包括确定第六图像和第五图像的偏移量，基于该偏移量移动第六图像，获得粗配准后的第六图像；

合并第五图像和粗配准后的第六图像，以获得参考图像。

25.权利要求 24 的装置，其特征在于，在利用所述参考图像构建模块构建参考图像时，还包括利用第七图像和第八图像，第七图像和第八图像与待配准图像来自测序反应的相同视野，第五图像、第六图像、第七图像和第八图像分别对应 A/ U、T、G 和 C 四种类型碱基延伸反应时的视野，构建参考图像还包括：

基于第五图像对第七图像进行粗配准，包括确定第七图像和第五图像的偏移量，基于该偏移量移动第七图像，获得粗配准后的第七图像；

基于第五图像对第八图像进行粗配准，包括确定第八图像和第五图像的偏移量，基于该偏移量移动第八图像，获得粗配准后的第八图像；

合并第五图像和粗配准后的第六图像、粗配准后的第七图像以及粗配准后的第八图像，以获得参考图像。

26.权利要求 23-25 任一装置，其特征在于，参考图像和待配准图像为二值化图像。

27.权利要求 23-26 任一装置，其特征在于，利用二维离散傅里叶变换确定第一偏移量、第六图像和第五图像的偏移量、第七图像和第五图像的偏移量和/或第八图像和第五图像的偏移量。

28.权利要求 19-27 任一装置，其特征在于，还包括亮斑检测单元，所述亮斑检测单元用于：

预处理图像，获得预处理后的图像；

确定临界值以简化预处理后的图像，包括对小于临界值的预处理后的图像上的像素点的像素值赋值为第一预设值，对不小于临界值的预处理后的图像上的像素点的像素值赋值为第二预设值，以获得简化图像；

基于预处理后的图像确定第一亮斑检测阈值 c_1 ；

基于预处理后的图像和简化图像识别图像上的候选亮斑，包括判定满足以下 a) -c) 中至少两个条件的像素点矩阵为一个候选亮斑，

a) 在预处理后的图像中，像素点矩阵的中心像素点的像素值为最大，像素点矩阵可表示为 $r_1 * r_2$ ， r_1 和 r_2 均为大于 1 的奇数， $r_1 * r_2$ 像素点矩阵包含 $r_1 * r_2$ 个像素点，

b) 在简化图像中，像素点矩阵的中心像素点的像素值为第二预设值并且像素点矩阵的连通像素大于 $2/3 * r_1 * r_2$ ，以及

c) 在预处理后的图像中的像素点矩阵的中心像素点的像素值大于第三预设值，并且满足 $g_1 * g_2 > c_1$ ， g_1 为以像素点矩阵的中心像素点为中心的 $m_1 * m_2$ 范围的二维高斯分布的相关系数， g_2 为该 $m_1 * m_2$ 范围的像素， m_1 和 m_2 均为大于 1 的奇数， $m_1 * m_2$ 范围包含 $m_1 * m_2$ 个像素点。

29.权利要求 28 的装置，其特征在于，所述亮斑检测单元还包括用于判定候选亮斑是否为亮斑，包括：

基于预处理后的图像确定第二亮斑检测阈值，以及
判定像素值不小于第二亮斑检测阈值的候选亮斑为亮斑。

30. 权利要求 29 的装置，其特征在于，候选亮斑的像素值为该候选亮斑的坐标所在的像素点的像素值。

31. 权利要求 29 或 30 的装置，其特征在于，在所述亮斑检测单元中判定候选亮斑是否为亮斑包括：

将预处理后的图像划分为预定大小的一组区域，

对该区域中的像素点的像素值进行排序，以确定该区域对应的第二亮斑检测阈值，

对于位于区域的候选亮斑，判定像素值不小于该区域对应的第二亮斑检测阈值的候选亮斑为亮斑。

32. 权利要求 28-31 任一装置，其特征在于，在所述亮斑检测单元中预处理图像，包括：

利用开运算确定图像的背景，

基于背景，利用顶帽运算将图像转化为第一图像，

对第一图像进行高斯模糊处理，获得第二图像，

对第二图像进行锐化，获得预处理后的图像。

33. 权利要求 28 的装置，其特征在于，在所述亮斑检测单元中确定临界值以简化预处理后的图像，获得简化图像，包括：

基于背景和预处理后的图像，确定临界值，

比较预处理后的图像上的像素点的像素值与临界值，以获得简化图像。

34. 权利要求 28-33 任一装置，其特征在于， g_2 为修正后的 $m_1 * m_2$ 范围的像素，依据简化图像相应 $m_1 * m_2$ 范围中像素值为第二预设值的像素点所占的比例进行修正。

35. 一种碱基识别装置，其特征在于，该装置用于将获自碱基延伸反应的图像上的亮斑匹配到对应测序模板的亮斑集合，依据匹配上的亮斑进行碱基识别，获自碱基延伸反应的图像对应的视野中存在多个带有光学可检测标记的核酸分子，至少一部分核酸分子在获自碱基延伸反应的图像上表现为亮斑，对应测序模板的亮斑集合通过权利要求 19-34 任一装置构建。

36. 权利要求 35 的装置，其特征在于，对应测序模板的亮斑集合中存在与获自碱基延伸反应的图像上的任一亮斑的距离小于第三预定像素，则判定获自碱基延伸反应的图像上的该亮斑匹配上对应测序模板的亮斑集合。

37. 一种计算机程序产品，包括指令，所述指令在所述计算机执行所述程序时，使所述计算机执行权利要求 1-16 任一方法。

38. 一种计算机程序产品，包括指令，所述指令在所述计算机执行所述程序时，使所述计算机执行权利要求 17 或 18 的方法。

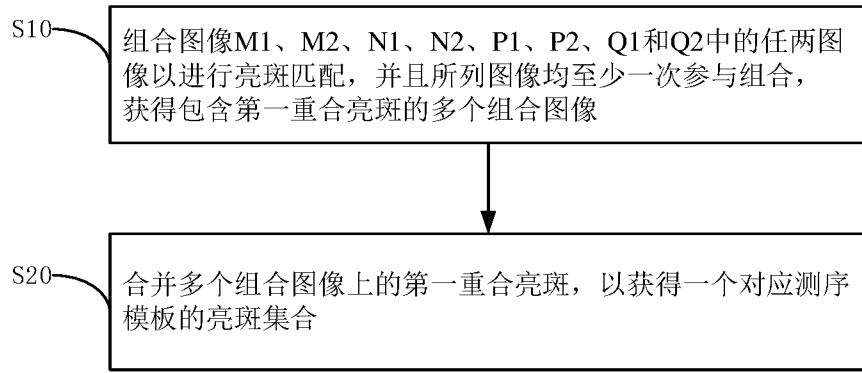


图 1

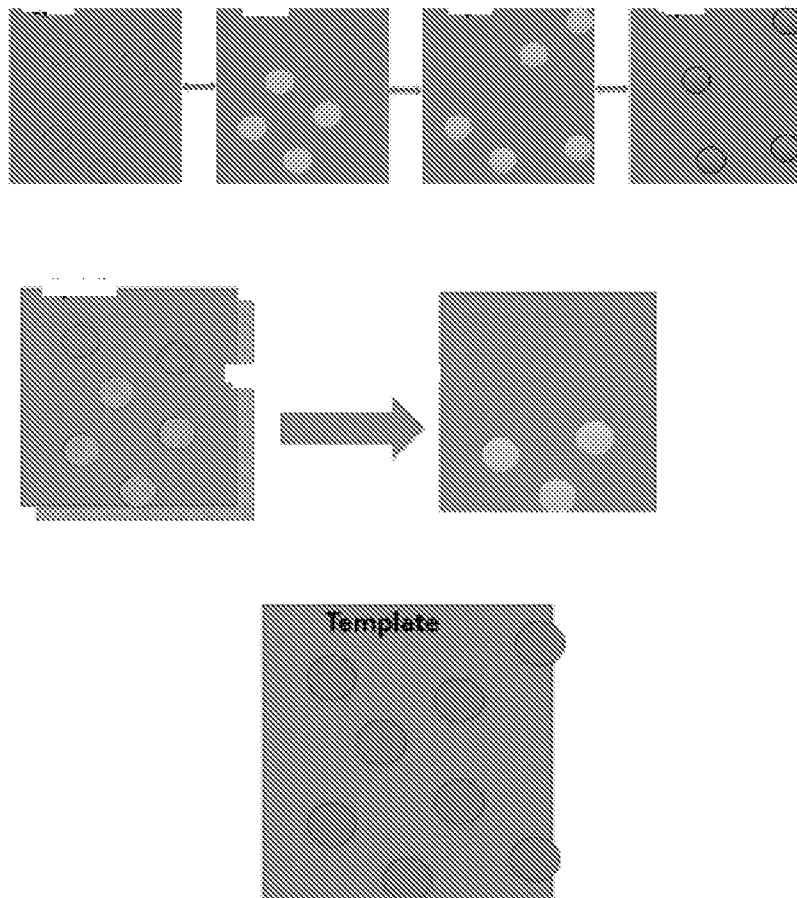


图 2

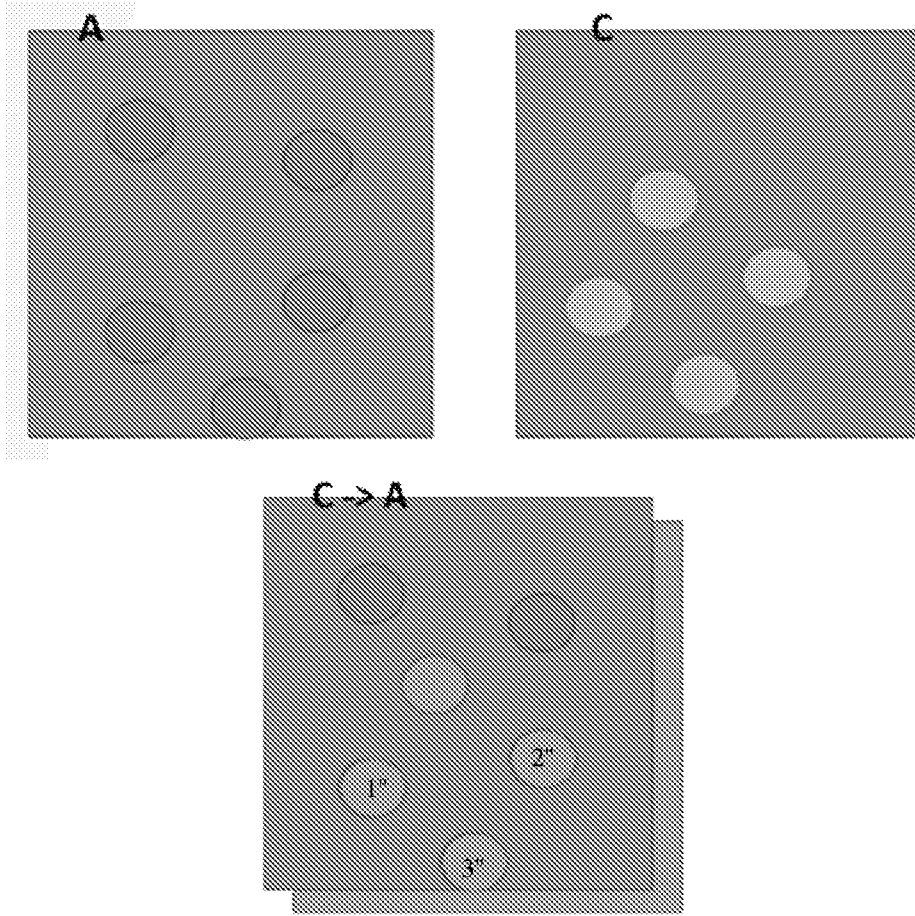


图 3

0	1	0	0	0	0
0	1	0	0	0	0
1	1	1	1	1	1
0	0	0	0	0	0

图 4

1	0	1	0	0
0	0	1	1	0
0	1	1	0	1
1	1	0	1	0
1	0	0	1	1

图 5

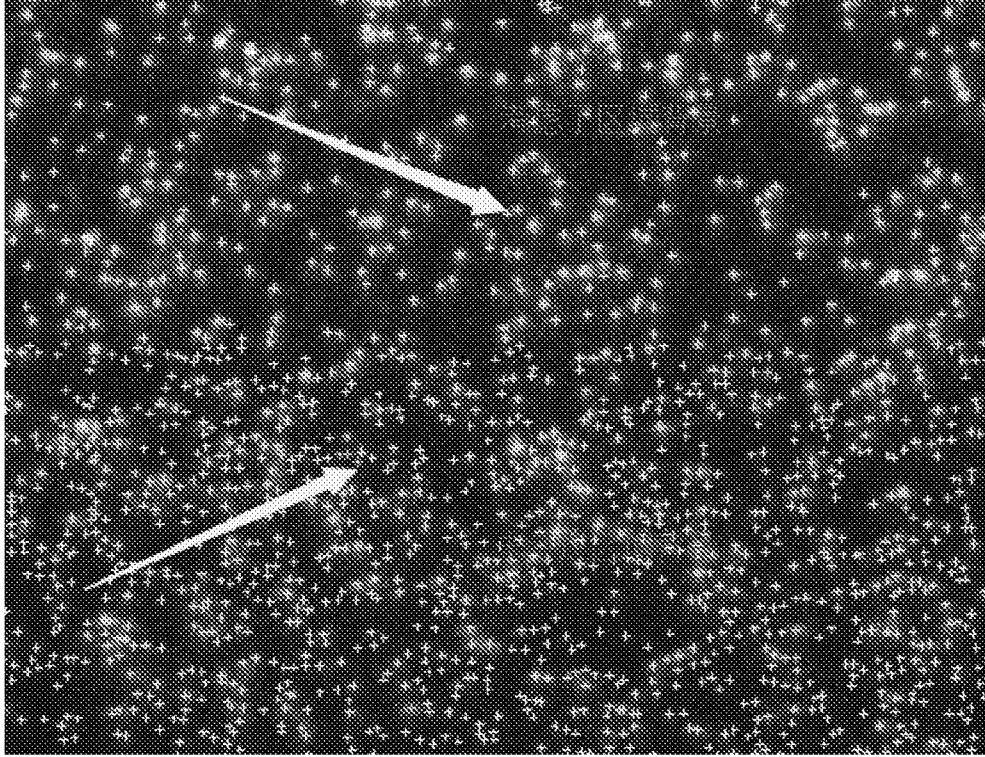


图 6

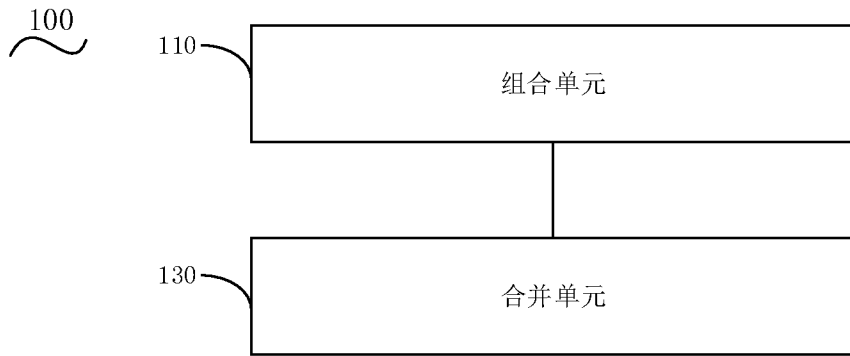


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/101819

A. CLASSIFICATION OF SUBJECT MATTER		
G06T 7/00(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) G06T; G06K; G06F		
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) CNPAT, EPODOC, WPI, CNKI, IEEE: 基因, 核酸, 碱基, 测序, 图像, 亮斑, 亮点, 合并, 融合, 配准, 组合, 识别, gene, DNA, base, sequence, image, spot, peak, registration, combine, merge, recognition, identify		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 108229098 A (DIRECT GENOMICS CO., LTD.) 29 June 2018 (2018-06-29) description, paragraphs [0062]-[0075]	1-38
A	CN 107945150 A (DIRECT GENOMICS BIOTECHNOLOGY CO., LTD.) 20 April 2018 (2018-04-20) entire document	1-38
A	CN 106295124 A (GUANGZHOU MELOW INFORMATION TECHNOLOGY CO., LTD.) 04 January 2017 (2017-01-04) entire document	1-38
A	CN 105551034 A (BEIJING ZHONGKEZIXIN TECHNOLOGY CO., LTD.) 04 May 2016 (2016-05-04) entire document	1-38
A	CN 105524827 A (BEIJING ZHONGKEZIXIN TECHNOLOGY CO., LTD.) 27 April 2016 (2016-04-27) entire document	1-38
A	US 2015317433 A1 (COMPLETE GENOMICS, INC.) 05 November 2015 (2015-11-05) entire document	1-38
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 "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 30 April 2019		Date of mailing of the international search report 21 May 2019
Name and mailing address of the ISA/CN National Intellectual Property Administration, PRC (ISA/ CN) No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China		Authorized officer
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/101819

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108229098	A	29 June 2018	WO	2018103373	A1	14 June 2018
				EP	3336802	A1	20 June 2018
				US	2018165409	A1	14 June 2018
				HK	1250816	A0	11 January 2019

CN	107945150	A	20 April 2018	WO	2018068599	A1	19 April 2018
				WO	2018068511	A1	19 April 2018
				CN	107918931	A	17 April 2018
				WO	2018068600	A1	19 April 2018
				HK	1247720	A0	28 September 2018
				HK	1247722	A0	28 September 2018
				HK	1247724	A0	28 September 2018
				EP	3306566	A1	11 April 2018
				US	2018101951	A1	12 April 2018

CN	106295124	A	04 January 2017	None			

CN	105551034	A	04 May 2016	None			

CN	105524827	A	27 April 2016	None			

US	2015317433	A1	05 November 2015	None			

<p>A. 主题的分类</p> <p>G06T 7/00 (2017.01) i</p> <p>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</p>																							
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06T; G06K; G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNPAT, EPDOC, WPI, CNKI, IEEE: 基因, 核酸, 碱基, 测序, 图像, 亮斑, 亮点, 合并, 融合, 配准, 组合, 识别, gene, DNA, base, sequence, image, spot, peak, registration, combine, merge, recognition, identify</p>																							
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 108229098 A (深圳市瀚海基因生物科技有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第[0062]-[0075]段</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>CN 107945150 A (深圳市瀚海基因生物科技有限公司) 2018年 4月 20日 (2018 - 04 - 20) 全文</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>CN 106295124 A (广州麦仑信息科技有限公司) 2017年 1月 4日 (2017 - 01 - 04) 全文</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>CN 105551034 A (北京中科紫鑫科技有限责任公司) 2016年 5月 4日 (2016 - 05 - 04) 全文</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>CN 105524827 A (北京中科紫鑫科技有限责任公司) 2016年 4月 27日 (2016 - 04 - 27) 全文</td> <td>1-38</td> </tr> <tr> <td>A</td> <td>US 2015317433 A1 (COMPLETE GENOMICS, INC.) 2015年 11月 5日 (2015 - 11 - 05) 全文</td> <td>1-38</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 108229098 A (深圳市瀚海基因生物科技有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第[0062]-[0075]段	1-38	A	CN 107945150 A (深圳市瀚海基因生物科技有限公司) 2018年 4月 20日 (2018 - 04 - 20) 全文	1-38	A	CN 106295124 A (广州麦仑信息科技有限公司) 2017年 1月 4日 (2017 - 01 - 04) 全文	1-38	A	CN 105551034 A (北京中科紫鑫科技有限责任公司) 2016年 5月 4日 (2016 - 05 - 04) 全文	1-38	A	CN 105524827 A (北京中科紫鑫科技有限责任公司) 2016年 4月 27日 (2016 - 04 - 27) 全文	1-38	A	US 2015317433 A1 (COMPLETE GENOMICS, INC.) 2015年 11月 5日 (2015 - 11 - 05) 全文	1-38
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
A	CN 108229098 A (深圳市瀚海基因生物科技有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第[0062]-[0075]段	1-38																					
A	CN 107945150 A (深圳市瀚海基因生物科技有限公司) 2018年 4月 20日 (2018 - 04 - 20) 全文	1-38																					
A	CN 106295124 A (广州麦仑信息科技有限公司) 2017年 1月 4日 (2017 - 01 - 04) 全文	1-38																					
A	CN 105551034 A (北京中科紫鑫科技有限责任公司) 2016年 5月 4日 (2016 - 05 - 04) 全文	1-38																					
A	CN 105524827 A (北京中科紫鑫科技有限责任公司) 2016年 4月 27日 (2016 - 04 - 27) 全文	1-38																					
A	US 2015317433 A1 (COMPLETE GENOMICS, INC.) 2015年 11月 5日 (2015 - 11 - 05) 全文	1-38																					
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																							
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																							
<p>国际检索实际完成的日期</p> <p>2019年 4月 30日</p>		<p>国际检索报告邮寄日期</p> <p>2019年 5月 21日</p>																					
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>受权官员</p> <p>宋朝</p> <p>电话号码 86-(10)-53961349</p>																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/101819

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)
CN	108229098	A	2018年 6月 29日	WO	2018103373	A1	2018年 6月 14日
				EP	3336802	A1	2018年 6月 20日
				US	2018165409	A1	2018年 6月 14日
				HK	1250816	A0	2019年 1月 11日

CN	107945150	A	2018年 4月 20日	WO	2018068599	A1	2018年 4月 19日
				WO	2018068511	A1	2018年 4月 19日
				CN	107918931	A	2018年 4月 17日
				WO	2018068600	A1	2018年 4月 19日
				HK	1247720	A0	2018年 9月 28日
				HK	1247722	A0	2018年 9月 28日
				HK	1247724	A0	2018年 9月 28日
				EP	3306566	A1	2018年 4月 11日
				US	2018101951	A1	2018年 4月 12日

CN	106295124	A	2017年 1月 4日	无			

CN	105551034	A	2016年 5月 4日	无			

CN	105524827	A	2016年 4月 27日	无			

US	2015317433	A1	2015年 11月 5日	无			

表 PCT/ISA/210 (同族专利附件) (2015年1月)