



(21) 申请号 202080087114.8

(22) 申请日 2020.12.18

(30) 优先权数据

62/954,720 2019.12.30 US

63/060,947 2020.08.04 US

(85) PCT国际申请进入国家阶段日

2022.06.15

(86) PCT国际申请的申请数据

PCT/US2020/065797 2020.12.18

(87) PCT国际申请的公布数据

W02021/138078 EN 2021.07.08

(71) 申请人 哈佛学院院长及董事

地址 美国马萨诸塞

(72) 发明人 庄小威 B·宾图 S·S·金若特

郑璞 苏俊翰

(74) 专利代理机构 中国贸促会专利商标事务所

有限公司 11038

专利代理师 李程达

(51) Int.Cl.

C12Q 1/6816 (2018.01)

C12Q 1/6837 (2018.01)

C12Q 1/6841 (2018.01)

G40B 70/00 (2006.01)

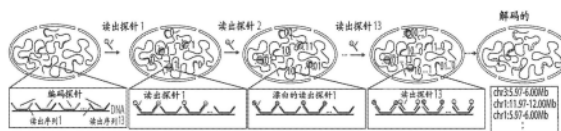
权利要求书13页 说明书80页 附图98页

(54) 发明名称

染色质的3D组织和转录活动的基因组规模成像

(57) 摘要

本发明一般涉及基因组学。一些实施方案涉及在序列空间中以高通量对基因组或基因组的一部分的3D组织进行成像。一些实施方案涉及在转录活动和核结构的背景下对基因组或基因组的一部分的3D组织进行成像。此外,某些实施方案涉及染色质结构、3D染色质组织、跨染色体相互作用和染色质-核结构相互作用以及它们与转录的关系等。此外,各种实施方案涉及允许在核结构和转录活动的背景下对基因组或基因组的一部分的3D组织进行作图的成像方法。一些实施方案涉及用于在染色体或基因组规模上对染色质基因座和/或新生RNA转录物进行成像的大规模多重荧光原位杂交方法。



1. 一种方法, 包括:
将基因组的多个核酸靶标与多个代码字相关联, 其中代码字包括多个位置和每个位置的值;
将含有基因组的样品暴露于多个核酸探针;
对于多个核酸探针中的每个核酸探针, 确定样品内的核酸探针的结合;
产生对应于样品内多个核酸探针的结合的代码字; 和
基于分配的代码字确定核酸靶标的身份。
2. 权利要求1的方法, 进一步确定所鉴定的核酸靶标的空间位置。
3. 权利要求1或2中任一项的方法, 进一步基于所鉴定的核酸靶标的空间位置确定染色质或基因组的三维组织。
4. 权利要求1-3中任一项的方法, 其中在代码字中形成错误检查和/或错误校正代码空间。
5. 权利要求4的方法, 其中错误检查和/或错误校正检测技术包括MERFISH。
6. 权利要求1-5中任一项的方法, 其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分, 其中每个读出序列代表代码字内的位置的值。
7. 权利要求6的方法, 还包括将样品暴露于能够结合读出序列的读出探针。
8. 权利要求7的方法, 其中读出探针包含信号传导实体。
9. 权利要求8的方法, 其中信号传导实体是荧光分子。
10. 权利要求7-9中任一项的方法, 还包括将样品依次暴露于多个读出探针。
11. 权利要求7-10中任一项的方法, 还包括创建对应于多个核酸探针在细胞核内的结合的代码字, 其中代码字的数字的值基于存在于核酸探针上的读出序列。
12. 权利要求7-11中任一项的方法, 其中对于代码字中的至少一些, 将代码字与有效代码字匹配, 其中如果未找到匹配, 则丢弃代码字或对代码字应用错误校正以形成有效代码字, 有效代码字是分配给多个核酸靶标的多个代码字。
13. 权利要求1-12中任一项的方法, 其中多个核酸靶标在基因组内被至少100,000个核苷酸分开。
14. 权利要求1-13中任一项的方法, 其中多个核酸靶标在基因组内被至少300,000个核苷酸分开。
15. 权利要求1-14中任一项的方法, 其中多个核酸靶标在基因组内被至少1,000,000个核苷酸分开。
16. 权利要求1-15中任一项的方法, 其中多个核酸靶标在基因组内被至少10,000,000个核苷酸分开。
17. 权利要求1-16中任一项的方法, 其中基因组的多个核酸靶标分布为使得基因组的每个染色体包含不超过10,000个核酸靶标。
18. 权利要求1-17中任一项的方法, 其中基因组具有10至100,000个核酸靶标。
19. 权利要求1-18中任一项的方法, 还包括确定细胞核内的新生RNA。
20. 权利要求1-19中任一项的方法, 还包括确定细胞核内的核斑点。
21. 权利要求1-20中任一项的方法, 还包括确定细胞核内的核仁。

22. 权利要求1-21中任一项的方法,还包括确定细胞核内的核纤层。
23. 权利要求1-22中任一项的方法,还包括确定细胞核内的其他蛋白质和核酸种类。
24. 权利要求1-23中任一项的方法,还包括确定样品中的其他蛋白质和核酸种类。
25. 权利要求1-24中任一项的方法,包括将样品暴露于至少10个核酸探针。
26. 权利要求1-25中任一项的方法,包括将细胞核暴露于至少30个核酸探针。
27. 权利要求1-26中任一项的方法,包括将细胞核暴露于至少100个核酸探针。
28. 权利要求1-27中任一项的方法,包括将细胞核暴露于至少1,000个核酸探针。
29. 权利要求1-28中任一项的方法,包括将细胞核暴露于至少10,000个核酸探针。
30. 权利要求1-29中任一项的方法,包括将细胞核暴露于至少100,000个核酸探针。
31. 权利要求1-30中任一项的方法,包括将细胞核暴露于至少1,000,000个核酸探针。
32. 权利要求1-31中任一项的方法,包括将样品依次暴露于核酸探针。
33. 权利要求1-32中任一项的方法,其中多个核酸探针包括具有不同序列的核酸探针的组合性组合。
34. 权利要求33的方法,其中多个核酸探针包含至少8个读出序列。
35. 权利要求33或34中任一项的方法,其中多个核酸探针包含至少32个读出序列。
36. 权利要求33-35中任一项的方法,其中多个核酸探针包含不超过32个可能的读出序列。
37. 权利要求33的方法,其中多个核酸探针包含不超过8个可能的读出序列。
38. 权利要求33-37中任一项的方法,其中多个读出序列分布在多个核酸探针上以便定义错误检查代码。
39. 权利要求33-38中任一项的方法,其中多个读出序列分布在多个核酸探针上以便定义错误校正代码。
40. 权利要求33-39中任一项的方法,其中多个读出序列具有5个核苷酸至50个核苷酸的平均长度。
41. 权利要求33-40中任一项的方法,其中多个核酸探针中的至少一些包含不超过50个读出序列。
42. 权利要求33-41中任一项的方法,其中多个核酸探针中的至少一些包含不超过10个读出序列。
43. 权利要求33-42中任一项的方法,其中确定核酸探针在细胞核内的结合包括:
将细胞核暴露于包含第一信号传导实体的第一读出探针,所述第一读出探针能够结合核酸探针的一个或多个读出序列;和
通过确定细胞核内的第一信号传导实体来确定核酸探针的结合。
44. 权利要求43的方法,其中第一信号传导实体是荧光的。
45. 权利要求43或44中任一项的方法,其中第一信号传导实体是蛋白质。
46. 权利要求43-45中任一项的方法,其中第一信号传导实体是染料。
47. 权利要求43-46中任一项的方法,其中第一信号传导实体是纳米颗粒。
48. 权利要求43-47中任一项的方法,还包括:
将细胞核暴露于包含第二信号传导实体的第二读出探针,所述第二读出探针能够结合核酸探针的一些读出序列;和

通过确定细胞核内的第二信号传导实体来确定核酸探针的结合。

49. 权利要求48的方法, 其中第一信号传导实体和第二信号传导实体是相同的。

50. 权利要求48或49中任一项的方法, 其中第一信号传导实体和第二信号传导实体不相同。

51. 权利要求48-50中任一项的方法, 还包括在将细胞核暴露于第二次级探针之前使第一信号传导实体灭活。

52. 权利要求51的方法, 包括通过光漂白至少一些第一信号传导实体来使第一信号传导实体灭活。

53. 权利要求51或52中任一项的方法, 包括通过化学漂白至少一些第一信号传导实体来使第一信号传导实体灭活。

54. 权利要求51-53中任一项的方法, 包括通过将第一信号传导实体暴露于能够改变信号传导实体的结构的反应物来使第一信号传导实体灭活。

55. 权利要求51-54中任一项的方法, 包括通过去除至少一些第一信号传导实体来使第一信号传导实体灭活。

56. 权利要求51-55中任一项的方法, 包括通过将第一信号传导实体与第一读出探针分离来使第一信号传导实体灭活。

57. 权利要求51-56中任一项的方法, 包括通过从样品中解离包含第一信号传导实体的第一读出探针来使第一信号传导实体灭活。

58. 权利要求51-57中任一项的方法, 包括通过从第一读出探针化学裂解第一信号传导实体来使第一信号传导实体灭活。

59. 权利要求51-58中任一项的方法, 其包括通过从第一读出探针酶促裂解第一信号传导实体来使第一信号传导实体灭活。

60. 权利要求51-59中任一项的方法, 包括通过将信号传导实体或第一读出探针暴露于酶来使第一信号传导实体灭活。

61. 权利要求43-60中任一项的方法, 包括使用用于确定非重叠单个发射器的算法来确定第一信号传导实体的质心。

62. 权利要求43-61中任一项的方法, 包括使用用于确定部分重叠单个发射器的算法来确定第一信号传导实体的质心。

63. 权利要求43-62中任一项的方法, 包括使用最大似然算法确定第一信号传导实体的质心。

64. 权利要求43-63中任一项的方法, 包括使用最小二乘算法确定第一信号传导实体的质心。

65. 权利要求43-64中任一项的方法, 包括使用贝叶斯算法确定第一信号传导实体的质心。

66. 权利要求43-65中任一项的方法, 包括使用压缩感测算法来确定第一信号传导实体的质心。

67. 权利要求1-66中任一项的方法, 其中多个核酸探针中的至少一些包含DNA。

68. 权利要求1-67中任一项的方法, 其中多个核酸探针中的至少一些包含RNA。

69. 权利要求1-68中任一项的方法, 其中多个核酸探针中的至少一些包含PNA。

70. 权利要求1-69中任一项的方法,其中多个核酸探针中的至少一些包含LNA。
71. 权利要求1-70中任一项的方法,其中多个核酸探针具有10至300个核苷酸的平均长度。
72. 权利要求1-71中任一项的方法,其中核酸探针在细胞核内的至少一些结合是特异性结合。
73. 权利要求1-72中任一项的方法,其中核酸探针在细胞核内的至少一些结合是通过Watson-Crick碱基配对进行的。
74. 权利要求1-73中任一项的方法,包括以优于300nm的分辨率确定核酸探针在样品中的结合。
75. 权利要求1-74中任一项的方法,包括以优于100nm的分辨率确定核酸探针在样品中的结合。
76. 权利要求1-75中任一项的方法,包括以优于80nm的分辨率确定核酸探针在样品中的结合。
77. 权利要求1-76中任一项的方法,包括以优于50nm的分辨率确定核酸探针在样品中的结合。
78. 权利要求1-77中任一项的方法,其中样品是细胞。
79. 权利要求78的方法,其中细胞是固定的。
80. 权利要求1-79中任一项的方法,包括通过对样品的至少一部分进行成像来确定核酸探针的结合。
81. 权利要求1-80中任一项的方法,包括使用光学成像技术确定核酸探针的结合。
82. 权利要求1-81中任一项的方法,包括使用荧光成像技术确定核酸探针的结合。
83. 权利要求1-82中任一项的方法,包括使用多色荧光成像技术确定核酸探针的结合。
84. 权利要求1-83中任一项的方法,包括使用超分辨率荧光成像技术确定核酸探针的结合。
85. 权利要求1-84中任一项的方法,包括使用随机光学重建显微术 (STORM) 确定核酸探针的结合。
86. 一种方法,包括:
确定新生RNA在细胞核内的位置;
将核糖核酸酶应用于细胞核;和
确定DNA在细胞核内的位置。
87. 一种方法,包括:
确定新生RNA在细胞核内的位置;
确定DNA在细胞核内的位置;和
确定蛋白质在细胞核内的位置。
88. 一种方法,包括:
确定新生RNA在细胞核内的位置;
确定DNA在细胞核内的位置;和
确定核酸在细胞核内的位置,其中核酸不是新生RNA或DNA。
89. 一种方法,包括:

使用MERFISH对细胞中的染色质进行成像。

90. 权利要求89的方法, 包括对染色质进行3维成像。

91. 权利要求89或90中任一项的方法, 还包括确定细胞核结构。

92. 权利要求89-91中任一项的方法, 还包括确定细胞中的转录活动。

93. 权利要求92的方法, 还包括确定细胞内的至少100个不同的转录位点。

94. 权利要求89-93中任一项的方法, 还包括确定细胞内的至少100个不同的基因组基因座。

95. 一种方法, 包括:

对单个细胞中至少100个不同的基因组基因座进行成像。

96. 一种方法, 包括:

将基因组的多个核酸靶标与多个代码字相关联;

将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针, 其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分, 其中每个读出序列代表多个代码字内的位置的值;

将样品暴露于一个轮次的一个或多个衔接子, 其中每个衔接子包含与读出序列之一基本上互补的第一部分和包含一个鉴定序列的第二部分;

将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列, 其中每个读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分和包含信号传导实体的第二部分;

确定样品中至少一些位置中的信号传导实体; 和

灭活样品中至少一些位置中的信号传导实体;

重复将样品暴露于一个轮次的一个或多个衔接子和一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤, 其中在每个轮次中使用一个或多个不同的信号传导实体;

基于确定样品中的信号传导实体确定位置处的代码字; 和

基于代码字确定样品中的核酸靶标。

97. 权利要求96的方法, 其中确定核酸靶标包括确定核酸靶标的空间位置。

98. 权利要求97的方法, 还包括从核酸靶标的空间位置确定染色质或基因组的三维组织。

99. 权利要求96-98中任一项的方法, 其中在所有轮次中使用不超过10个信号传导实体。

100. 权利要求96-99中任一项的方法, 其中在所有轮次中使用不超过5个信号传导实体。

101. 权利要求96-100中任一项的方法, 其中在所有轮次中使用不超过3个信号传导实体。

102. 权利要求96-101中任一项的方法, 其中将样品暴露于一个轮次的一个或多个读出探针还包括将样品暴露于封闭探针, 所述封闭探针包含与在前一轮次中确定的鉴定序列之一基本上互补的序列。

103. 权利要求96-102中任一项的方法, 包括使用不超过10个信号传导实体来确定至少

20个鉴定序列。

104. 权利要求96-103中任一项的方法, 包括使用不超过10个信号传导实体来确定至少100个鉴定序列。

105. 权利要求96-104中任一项的方法, 包括使用不超过10个信号传导实体来确定至少1,000个鉴定序列。

106. 权利要求96-105中任一项的方法, 包括使用不超过3个信号传导实体来确定至少100个鉴定序列。

107. 权利要求96-106中任一项的方法, 包括使用不超过3个信号传导实体来确定至少1,000个鉴定序列。

108. 权利要求96-107中任一项的方法, 包括在将细胞核暴露于多个核酸探针之后固定样品。

109. 权利要求96-108中任一项的方法, 包括在将样品暴露于一个或多个读出探针的轮次之间固定样品。

110. 权利要求96-109中任一项的方法, 包括将样品固定至少5次。

111. 权利要求96-110中任一项的方法, 包括将样品至少每4天固定1次。

112. 权利要求96-111中任一项的方法, 包括将样品至少每2天固定1次。

113. 权利要求96-112中任一项的方法, 包括将样品至少每24小时固定1次。

114. 权利要求96-113中任一项的方法, 包括将样品至少每12小时固定1次。

115. 权利要求96-114中任一项的方法, 包括将样品至少每6小时固定1次。

116. 权利要求96-115中任一项的方法, 包括使用甲醛固定样品。

117. 权利要求96-116中任一项的方法, 包括将重复步骤重复至少10次。

118. 权利要求96-117中任一项的方法, 包括将重复步骤重复至少50次。

119. 权利要求96-118中任一项的方法, 包括将重复步骤重复至少100次。

120. 权利要求96-119中任一项的方法, 包括将重复步骤重复至少200次。

121. 权利要求96-120中任一项的方法, 包括将样品暴露于一个轮次的与前一轮次的一个或多个读出探针相同的一个或多个读出探针。

122. 权利要求121的方法, 还包括基于两个相同轮次的一个或多个读出探针来确定样品的降解。

123. 权利要求96-122中任一项的方法, 其中读出序列不表现出针对基因组的特异性结合。

124. 权利要求96-123中任一项的方法, 其中读出序列不表现出针对彼此的特异性结合。

125. 权利要求96-124中任一项的方法, 其中鉴定序列不表现出针对彼此的特异性结合。

126. 权利要求96-125中任一项的方法, 其中鉴定序列不表现出针对基因组的特异性结合。

127. 权利要求96-126中任一项的方法, 包括将样品暴露于至少50个可区分的核酸探针。

128. 权利要求96-127中任一项的方法, 包括将样品暴露于至少100个可区分的核酸探

针。

129. 权利要求96-128中任一项的方法, 包括将样品暴露于至少1,000个可区分的核酸探针。

130. 权利要求96-129中任一项的方法, 包括将样品暴露于至少10,000个可区分的核酸探针。

131. 权利要求96-130中任一项的方法, 包括将样品暴露于至少100,000个可区分的核酸探针。

132. 权利要求96-131中任一项的方法, 包括将样品暴露于至少1,000,000个可区分的核酸探针。

133. 权利要求96-132中任一项的方法, 还包括确定细胞核内的新生RNA。

134. 权利要求96-133中任一项的方法, 还包括确定细胞核内的核斑点。

135. 权利要求96-134中任一项的方法, 还包括确定细胞核内的核仁。

136. 权利要求96-135中任一项的方法, 还包括确定细胞核内的核纤层。

137. 权利要求96-136中任一项的方法, 还包括确定细胞核内的其他蛋白质和核酸种类。

138. 权利要求96-137中任一项的方法, 还包括确定样品中的其他蛋白质和核酸种类。

139. 权利要求96-138中任一项的方法, 其中至少一些信号传导实体是荧光的。

140. 权利要求96-139中任一项的方法, 包括对样品进行成像以确定在样品中的至少一些位置中的信号传导实体。

141. 一种方法, 包括:

将基因组的多个核酸靶标与多个代码字相关联;

将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针, 其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分, 其中每个读出序列代表多个代码字内的位置的值;

将样品暴露于一个轮次的一个或多个衔接子, 其中每个衔接子包含与读出序列之一基本上互补的第一部分和包含一个鉴定序列的第二部分;

将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列, 其中每个读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分和包含信号传导实体的第二部分;

确定样品中至少一些位置中的信号传导实体; 和

灭活样品中至少一些位置中的信号传导实体;

重复将样品暴露于一个轮次的一个或多个衔接子和一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤, 其中信号传导实体中的至少一个用于多于一个轮次;

基于确定样品中的信号传导实体确定位置处的代码字; 和

基于代码字确定样品中的核酸靶标。

142. 一种方法, 包括:

将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针, 其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分;

将样品暴露于一个轮次的一个或多个衔接子, 其中每个衔接子包含与读出序列之一基

本上互补的第一部分,以及包含一个鉴定序列的第二部分;

将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列,其中每个读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分和包含信号传导实体的第二部分;

确定样品中至少一些位置中的信号传导实体;和

灭活样品中至少一些位置中的信号传导实体;

重复将样品暴露于一个轮次的一个或多个衔接子和一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤,其中在每个轮次中使用一个或多个不同的信号传导实体;

基于在每个轮次中确定的信号传导实体确定样品中的核酸靶标。

143. 权利要求142的方法,其中确定核酸靶标包括确定核酸靶标的空间位置。

144. 权利要求143的方法,还包括从核酸靶标的空间位置确定染色质或基因组的三维组织。

145. 权利要求142-144中任一项的方法,其中在所有轮次中使用不超过10个信号传导实体。

146. 权利要求142-145中任一项的方法,其中在所有轮次中使用不超过5个信号传导实体。

147. 权利要求142-146中任一项的方法,其中在所有轮次中使用不超过3个信号传导实体。

148. 权利要求142-147中任一项的方法,其中在所有轮次中使用不超过2个信号传导实体。

149. 权利要求142-148中任一项的方法,其中在所有轮次中使用不超过1个信号传导实体。

150. 权利要求142-149中任一项的方法,其中将样品暴露于一个轮次的一个或多个读出探针还包括将样品暴露于封闭探针,所述封闭探针包含与在前一轮次中确定的鉴定序列之一基本上互补的序列。

151. 权利要求142-150中任一项的方法,包括使用不超过10个信号传导实体来确定至少20个鉴定序列。

152. 权利要求142-151中任一项的方法,包括使用不超过10个信号传导实体来确定至少100个鉴定序列。

153. 权利要求142-152中任一项的方法,包括使用不超过10个信号传导实体来确定至少1,000个鉴定序列。

154. 权利要求142-153中任一项的方法,包括使用不超过3个信号传导实体来确定至少100个鉴定序列。

155. 权利要求142-154中任一项的方法,包括使用不超过3个信号传导实体来确定至少1,000个鉴定序列。

156. 权利要求142-155中任一项的方法,包括在将细胞核暴露于多个核酸探针之后固定样品。

157. 权利要求142-156中任一项的方法,包括在将样品暴露于一个或多个读出探针的

轮次之间固定样品。

158. 权利要求142-157中任一项的方法,包括将样品固定至少5次。

159. 权利要求142-158中任一项的方法,包括将样品至少每4天固定1次。

160. 权利要求142-159中任一项的方法,包括将样品至少每2天固定1次。

161. 权利要求142-160中任一项的方法,包括将样品至少每24小时固定1次。

162. 权利要求142-161中任一项的方法,包括将样品至少每12小时固定1次。

163. 权利要求142-162中任一项的方法,包括将样品至少每6小时固定1次。

164. 权利要求142-163中任一项的方法,包括使用甲醛固定样品。

165. 权利要求142-164中任一项的方法,包括将重复步骤重复至少10次。

166. 权利要求142-165中任一项的方法,包括将重复步骤重复至少50次。

167. 权利要求142-166中任一项的方法,包括将重复步骤重复至少100次。

168. 权利要求142-167中任一项的方法,包括将重复步骤重复至少200次。

169. 权利要求142-168中任一项的方法,包括将样品暴露于一个轮次的与前一轮次的一个或多个读出探针相同的一个或多个读出探针。

170. 权利要求169的方法,其还包括基于两个相同轮次的一个或多个读出探针来确定样品的降解。

171. 权利要求142-170中任一项的方法,其中读出序列不表现出针对基因组的特异性结合。

172. 权利要求142-171中任一项的方法,其中读出序列不表现出针对彼此的特异性结合。

173. 权利要求142-172中任一项的方法,其中鉴定序列不表现出针对彼此的特异性结合。

174. 权利要求142-173中任一项的方法,其中鉴定序列不表现出针对基因组的特异性结合。

175. 权利要求142-174中任一项的方法,包括将样品暴露于至少50个可区分的核酸探针。

176. 权利要求142-175中任一项的方法,包括将样品暴露于至少100个可区分的核酸探针。

177. 权利要求142-176中任一项的方法,包括将样品暴露于至少1,000个可区分的核酸探针。

178. 权利要求142-177中任一项的方法,包括将样品暴露于至少10,000个可区分的核酸探针。

179. 权利要求142-178中任一项的方法,包括将样品暴露于至少100,000个可区分的核酸探针。

180. 权利要求142-179中任一项的方法,包括将样品暴露于至少1,000,000个可区分的核酸探针。

181. 权利要求142-180中任一项的方法,其中至少一些信号传导实体是荧光的。

182. 权利要求142-181中任一项的方法,包括对样品进行成像以确定在样品中的至少一些位置中的信号传导实体。

183. 权利要求142-182中任一项的方法,还包括确定细胞核内的新生RNA。

184. 权利要求142-183中任一项的方法,还包括确定细胞核内的核斑点。

185. 权利要求142-184中任一项的方法,还包括确定细胞核内的核仁。

186. 权利要求142-185中任一项的方法,还包括确定细胞核内的核纤层。

187. 权利要求142-186中任一项的方法,还包括确定细胞核内的其他蛋白质和核酸种类。

188. 权利要求142-187中任一项的方法,还包括确定样品中的其他蛋白质和核酸种类。

189. 一种方法,包括:

将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分;

将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个读出序列,其中每个读出探针含有包含与读出序列之一基本上互补的序列的第一部分和包含信号传导实体的第二部分;

确定样品中至少一些位置中的信号传导实体;和

灭活样品中至少一些位置中的信号传导实体;

重复将样品暴露于一个轮次的一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤,其中在每个轮次中使用一个或多个不同的信号传导实体;

基于在每个轮次中确定的信号传导实体确定样品中的核酸靶标。

190. 权利要求189的方法,其中确定核酸靶标包括确定核酸靶标的空间位置。

191. 权利要求190的方法,还包括从核酸靶标的空间位置确定染色质或基因组的三维组织。

192. 权利要求189-191中任一项的方法,其中在所有轮次中使用不超过10个信号传导实体。

193. 权利要求189-192中任一项的方法,其中在所有轮次中使用不超过5个信号传导实体。

194. 权利要求189-193中任一项的方法,其中在所有轮次中使用不超过3个信号传导实体。

195. 权利要求189-194中任一项的方法,其中在所有轮次中使用不超过2个信号传导实体。

196. 权利要求189-195中任一项的方法,其中在所有轮次中使用不超过1个信号传导实体。

197. 权利要求189-196中任一项的方法,其中将样品暴露于一个轮次的一个或多个读出探针还包括将样品暴露于封闭探针,所述封闭探针包含与在前一轮次中确定的鉴定序列之一基本上互补的序列。

198. 权利要求189-197中任一项的方法,包括在将细胞核暴露于多个核酸探针之后固定样品。

199. 权利要求189-198中任一项的方法,包括在将样品暴露于一个或多个读出探针的轮次之间固定样品。

200. 权利要求189-199中任一项的方法,包括将样品固定至少5次。

201. 权利要求189-200中任一项的方法,包括将样品至少每4天固定1次。
202. 权利要求189-201中任一项的方法,包括将样品至少每2天固定1次。
203. 权利要求189-202中任一项的方法,包括将样品至少每24小时固定1次。
204. 权利要求189-203中任一项的方法,包括将样品至少每12小时固定1次。
205. 权利要求189-204中任一项的方法,包括将样品至少每6小时固定1次。
206. 权利要求189-205中任一项的方法,包括使用甲醛固定样品。
207. 权利要求189-206中任一项的方法,包括将重复步骤重复至少10次。
208. 权利要求189-207中任一项的方法,包括将重复步骤重复至少50次。
209. 权利要求189-208中任一项的方法,包括将重复步骤重复至少100次。
210. 权利要求189-209中任一项的方法,包括将重复步骤重复至少200次。
211. 权利要求189-210中任一项的方法,包括将样品暴露于一个轮次的与前一轮次的一个或多个读出探针相同的一个或多个读出探针。
212. 权利要求211的方法,还包括基于两个相同轮次的一个或多个读出探针来确定样品的降解。
213. 权利要求189-212中任一项的方法,其中读出序列不表现出针对基因组的特异性结合。
214. 权利要求189-213中任一项的方法,其中读出序列不表现出针对彼此的特异性结合。
215. 权利要求189-214中任一项的方法,包括将样品暴露于至少50个可区分的核酸探针。
216. 权利要求189-215中任一项的方法,包括将样品暴露于至少100个可区分的核酸探针。
217. 权利要求189-216中任一项的方法,包括将样品暴露于至少1,000个可区分的核酸探针。
218. 权利要求189-217中任一项的方法,包括将样品暴露于至少10,000个可区分的核酸探针。
219. 权利要求189-218中任一项的方法,包括将样品暴露于至少100,000个可区分的核酸探针。
220. 权利要求189-219中任一项的方法,包括将样品暴露于至少1,000,000个可区分的核酸探针。
221. 权利要求189-220中任一项的方法,其中至少一些信号传导实体是荧光的。
222. 权利要求189-221中任一项的方法,包括对样品进行成像以确定在样品中的至少一些位置中的信号传导实体。
223. 权利要求189-222中任一项的方法,还包括确定细胞核内的新生RNA。
224. 权利要求189-223中任一项的方法,还包括确定细胞核内的核斑点。
225. 权利要求189-224中任一项的方法,还包括确定细胞核内的核仁。
226. 权利要求189-225中任一项的方法,还包括确定细胞核内的核纤层。
227. 权利要求189-226中任一项的方法,还包括确定细胞核内的其他蛋白质和核酸种类。

228. 权利要求189-227中任一项的方法,还包括确定样品中的其他蛋白质和核酸种类。

229. 一种方法,包括:

将含有怀疑含有基因组的细胞的样品暴露于一个轮次的多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含信号传导实体的第二部分;

确定样品中至少一些位置中的信号传导实体;和

灭活样品中至少一些位置中的信号传导实体;

重复将样品暴露于一个轮次的多个核酸探针、确定信号传导实体和灭活信号传导实体的步骤,其中在每个轮次中使用一个或多个不同的信号传导实体;

基于在每个轮次中确定的信号传导实体确定样品中的核酸靶标。

230. 权利要求229的方法,其中确定核酸靶标包括确定核酸靶标的空间位置。

231. 权利要求230的方法,还包括从核酸靶标的空间位置确定染色质或基因组的三维组织。

232. 权利要求229-231中任一项的方法,其中在所有轮次中使用不超过10个信号传导实体。

233. 权利要求229-232中任一项的方法,其中在所有轮次中使用不超过5个信号传导实体。

234. 权利要求229-233中任一项的方法,其中在所有轮次中使用不超过3个信号传导实体。

235. 权利要求229-234中任一项的方法,其中在所有轮次中使用不超过2个信号传导实体。

236. 权利要求229-235中任一项的方法,其中在所有轮次中使用不超过1个信号传导实体。

237. 权利要求229-236中任一项的方法,包括在将细胞核暴露于多个核酸探针之后固定样品。

238. 权利要求229-237中任一项的方法,包括在将样品暴露于一个或多个读出探针的轮次之间固定样品。

239. 权利要求229-238中任一项的方法,包括将样品固定至少5次。

240. 权利要求229-239中任一项的方法,包括将样品至少每4天固定1次。

241. 权利要求229-240中任一项的方法,包括将样品至少每2天固定1次。

242. 权利要求229-241中任一项的方法,包括将样品至少每24小时固定1次。

243. 权利要求229-242中任一项的方法,包括将样品至少每12小时固定1次。

244. 权利要求229-243中任一项的方法,包括将样品至少每6小时固定1次。

245. 权利要求229-244中任一项的方法,包括使用甲醛固定样品。

246. 权利要求229-245中任一项的方法,包括将重复步骤重复至少10次。

247. 权利要求229-246中任一项的方法,包括将重复步骤重复至少50次。

248. 权利要求229-247中任一项的方法,包括将重复步骤重复至少100次。

249. 权利要求229-248中任一项的方法,包括将重复步骤重复至少200次。

250. 权利要求229-249中任一项的方法,包括将样品暴露于至少50个可区分的核酸探针。

251. 权利要求229-250中任一项的方法,包括将样品暴露于至少100个可区分的核酸探针。

252. 权利要求229-251中任一项的方法,包括将样品暴露于至少1,000个可区分的核酸探针。

253. 权利要求229-252中任一项的方法,包括将样品暴露于至少10,000个可区分的核酸探针。

254. 权利要求229-253中任一项的方法,包括将样品暴露于至少100,000个可区分的核酸探针。

255. 权利要求229-254中任一项的方法,包括将样品暴露于至少1,000,000个可区分的核酸探针。

256. 权利要求229-255中任一项的方法,其中至少一些信号传导实体是荧光的。

257. 权利要求229-256中任一项的方法,包括对样品进行成像以确定在样品中的至少一些位置中的信号传导实体。

258. 权利要求229-257中任一项的方法,还包括确定细胞核内的新生RNA。

259. 权利要求229-258中任一项的方法,还包括确定细胞核内的核斑点。

260. 权利要求229-259中任一项的方法,还包括确定细胞核内的核仁。

261. 权利要求229-260中任一项的方法,还包括确定细胞核内的核纤层。

262. 权利要求229-261中任一项的方法,还包括确定细胞核内的其他蛋白质和核酸种类。

263. 权利要求229-262中任一项的方法,还包括确定样品中的其他蛋白质和核酸种类。

染色质的3D组织和转录活动的基因组规模成像

[0001] 相关申请

[0002] 本申请要求Zhuang等人于2019年12月30日提交的题为“Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin”的美国临时专利申请序列号62/954,720和Zhuang等人于2020年8月4日提交的题为“Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin”的美国临时专利申请序列号63/060,947的权益。这些中的每一个都通过引用整体并入本文。

技术领域

[0003] 本发明一般涉及基因组学。一些实施方案涉及在转录活动和核结构的背景下对基因组的3D组织进行成像。此外,某些实施方案涉及染色质组织和染色质-核结构相互作用以及它们与转录的关系。

背景技术

[0004] 基因组的三维(3D)组织调节从基因表达到DNA复制的许多基本细胞功能。生化和成像测量揭示了广泛尺度内的复杂染色质结构。最近,高通量染色体构象捕获方法例如Hi-C和其他基于测序的方法极大地丰富了3D基因组组织的知识,从全基因组的角度揭示了染色质结构,例如环、结构域和隔室。这些强大的基于测序的方法也有局限性。例如,这些方法提供染色质基因座对之间的联系信息,但不提供这些基因座的直接空间位置信息。此外,大多数关于染色质组织的全基因组见解都是建立在数百万个细胞的群体平均接触图上的。尽管单细胞Hi-C方法不断改进,但单细胞中染色质接触的捕获效率和/或这些方法的细胞通量仍然相对较低,因此研究单细胞中的3D基因组组织仍然是一项具有挑战性的任务。此外,尽管已经出现了将Hi-C与其他测量方式相结合的方法,例如,在相互作用的蛋白质、核结构或DNA修饰的背景下提供染色质接触的特征,但通过测序进行多模态测量仍然具有挑战性。值得注意的是,允许对同一细胞中染色质组织和转录活动两者进行基因组规模测量的方法尚未出现,而这种方法是非常需要的,因为了解染色质组织如何调节转录以及转录如何进而影响染色质组织是至关重要的。

[0005] 另一方面,基于成像的方法以很高的检测效率提供了对个体细胞中染色质基因座的空间位置的直接测量。特别是,荧光原位杂交(FISH)允许固定的细胞中的染色质基因座的高度特异性的检测,最近,规律成簇间隔短回文重复(CRISPR)系统大大增强了我们对活细胞中特定染色质基因座成像的能力。染色质成像也可以与RNA和蛋白质成像相结合,以揭示染色质组织与转录活动或相互作用的蛋白质因子之间的相互作用。然而,目前的成像方法在序列空间中在通量方面受到限制,传统上一次只允许研究几个不同的基因组基因座。基因组规模的成像需要大幅增加在个体细胞中成像的基因组基因座的数量。因此,需要进行新的改进。

发明内容

[0006] 本发明一般涉及基因组学。一些实施方案涉及在序列空间中以高通量对基因组或基因组的一部分的3D组织进行成像。一些实施方案涉及在转录活动和核结构的背景下对基因组或基因组的一部分的3D组织进行成像。此外,某些实施方案涉及染色质结构、3D染色质组织、跨染色体相互作用和染色质-核结构相互作用以及它们与转录的关系等。本公开内容的主题在某些情况下涉及相互关联的产品,特定问题的替代解决方案,和/或一个或多个系统和/或物品的多种不同用途。

[0007] 某些方面一般涉及使用多重FISH和在某些情况下使用多重抗错FISH(MERFISH)以对例如细胞中的染色质进行成像的系统和方法。此外,某些方面一般涉及成像和/或确定单个细胞中至少100或至少500个不同基因组基因座的系统和方法。一些方面一般涉及使用FISH对例如细胞中的染色质进行成像的系统和方法。

[0008] 在一组实施方案中,该方法包括将基因组的多个核酸靶标与多个代码字相关联,其中代码字包括多个位置和每个位置的值;将含有基因组的样品暴露于多个核酸探针;对于多个核酸探针中的每个核酸探针,确定样品内的核酸探针的结合;产生对应于样品内多个核酸探针的结合的代码字;以及基于分配的代码字确定核酸靶标的身份。

[0009] 在另一组实施方案中,该方法包括确定新生RNA在细胞核内的位置;将核糖核酸酶应用于细胞核;并确定DNA在细胞核内的位置。

[0010] 在一组实施方案中,该方法包括使用MERFISH对细胞中的染色质进行成像。在另一组实施方案中,该方法包括对单个细胞中的至少100个或至少500个不同的基因组基因座进行成像。

[0011] 根据一组实施方案,该方法包括将基因组的多个核酸靶标与多个代码字相关联;将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针,其中所述多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分,其中每个读出序列代表多个代码字内的位置的值得;将样品暴露于一个轮次的一个或多个衔接子,其中每个衔接子包含与读出序列之一基本上互补的第一部分,以及包含一个鉴定序列的第二部分;将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列,其中每个读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分和包含信号传导实体的第二部分;确定样品中至少一些位置中的信号传导实体;和灭活样品中至少一些位置中的信号传导实体;重复将样品暴露于一个轮次的一个或多个衔接子和一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤,其中在每个轮次中使用一个或多个不同的信号传导实体;基于确定样品中的信号传导实体确定位置处的代码字;以及基于所述代码字确定样品中的核酸靶标。

[0012] 在另一组实施方案中,该方法包括将基因组的多个核酸靶标与多个代码字相关联;将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针,其中所述多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分,其中每个读出序列代表多个代码字内的位置的值得;将样品暴露于一个轮次的一个或多个衔接子,其中每个衔接子包含与读出序列之一基本上互补的第一部分,以及包含一个鉴定序列的第二部分;将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列,其中每个读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分和包含信号传导实体

的第二部分;确定样品中至少一些位置中的信号传导实体;和灭活样品中至少一些位置中的信号传导实体;重复将样品暴露于一个轮次的一个或多个衔接子和一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤,其中信号传导实体中的至少一个用于多于一个轮次;基于确定样品中的信号传导实体确定位置处的代码字;基于所述代码字确定样品中的核酸靶标。

[0013] 在另一组实施方案中,该方法包括将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分;将样品暴露于一个轮次的一个或多个衔接子,其中每个衔接子包含与读出序列之一基本上互补的第一部分,以及包含一个鉴定序列的第二部分;将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列,其中每个读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分和包含信号传导实体的第二部分;确定样品中至少一些位置中的信号传导实体;和灭活样品中至少一些位置中的信号传导实体;重复将样品暴露于一个轮次的一个或多个衔接子和一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤,其中在每个轮次中使用一个或多个不同的信号传导实体;基于在每个轮次中确定的信号传导实体确定样品中的核酸靶标。

[0014] 在另一组实施方案中,该方法包括将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读出序列的第二部分;将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个读出序列,其中每个读出探针含有包含与读出序列之一基本上互补的序列的第一部分和包含信号传导实体的第二部分;确定样品中至少一些位置中的信号传导实体;和灭活样品中至少一些位置中的信号传导实体;重复将样品暴露于一个轮次的一个或多个读出探针、确定信号传导实体和灭活信号传导实体的步骤,其中在每个轮次中使用一个或多个不同的信号传导实体;基于在每个轮次中确定的信号传导实体确定样品中的核酸靶标。

[0015] 在又一组实施方案中,该方法包括将含有怀疑含有基因组的细胞的样品暴露于一个轮次的多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含信号传导实体的第二部分;确定样品中至少一些位置中的信号传导实体;和灭活样品中至少一些位置中的信号传导实体;重复将样品暴露于一个轮次的多个核酸探针、确定信号传导实体和灭活信号传导实体的步骤,其中在每个轮次中使用一个或多个不同的信号传导实体;基于在每个轮次中确定的信号传导实体确定样品中的核酸靶标。

[0016] 在一组实施方案中,该方法包括将基因组的多个核酸靶标与多个代码字相关联,其中代码字包括多个位置和每个位置的值,并且代码字形成错误检查和/或错误校正代码空间,并且其中多个核酸靶标在基因组内被至少100,000个核苷酸分开;将包含基因组的细胞核暴露于多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读取序列的第二部分,其中每个读取序列表示代码字内的位置的值;对于多个核酸探针中的每个核酸探针,确定核酸探针在细胞核内的结合;创建对应于多个核酸探针在细胞核内的结合的代码字,其中代码字的数字的值基于核酸探针上存在的读取序列;对于至少一些代码字,将代码字与有效代码字匹配,其中,如果未找到匹配,则丢弃代码字或对代码字应用错误校正以形成有效代码字,有效代码字是分配给多个核酸靶标的多个代码字;以及使用对应于多个核酸探针在细胞核内的结合的有效代码字确定细胞核内的核

酸丰度和/或空间分布。

[0017] 在另一组实施方案中,该方法包括将基因组的多个核酸靶标与多个代码字相关联,其中代码字包括多个位置和每个位置的值,并且代码字形成错误检查和/或错误校正代码空间,以及其中基因组的多个核酸靶标分布为使得基因组的每个染色体包含不超过200个核酸靶标;将包含基因组的细胞核暴露于多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读取序列的第二部分,其中每个读取序列表示代码字内的位置的值;对于多个核酸探针中的每个核酸探针,确定核酸探针在细胞核内的结合;创建对应于多个核酸探针在细胞核内的结合的代码字,其中代码字的数字的值基于核酸探针上存在的读取序列;对于至少一些代码字,将代码字与有效代码字匹配,其中,如果未找到匹配,则丢弃代码字或对代码字应用错误校正以形成有效代码字,有效代码字是分配给多个核酸靶标的多个代码字;以及使用对应于多个核酸探针在细胞核内的结合的有效代码字确定细胞核内的核酸丰度和/或空间分布。

[0018] 根据另一组实施方案,该方法包括将基因组的500至1500个的多个核酸靶标与多个代码字相关联,其中代码字包括多个位置和每个位置的值,并且代码字形成错误检查和/或错误校正代码空间;将包含基因组的细胞核暴露于多个核酸探针,其中多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读取序列的第二部分,其中每个读取序列表示代码字内的位置的值;对于多个核酸探针中的每个核酸探针,确定核酸探针在细胞核内的结合;创建对应于多个核酸探针在细胞核内的结合的代码字,其中代码字的数字的值基于核酸探针上存在的读取序列;对于至少一些代码字,将代码字与有效代码字匹配,其中,如果未找到匹配,则丢弃代码字或对代码字应用错误校正以形成有效代码字,有效代码字是分配给多个核酸靶标的多个代码字;以及使用对应于多个核酸探针在细胞核内的结合的有效代码字确定细胞核内的核酸丰度和/或空间分布。

[0019] 在又一组实施方案中,该方法包括将基因组的多个核酸靶标与多个代码字相关联,其中代码字包括多个位置和每个位置的值,并且代码字形成错误检查和/或错误校正代码空间,以及其中多个核酸靶标在基因组内被至少100,000个核苷酸分开;将包含基因组的细胞核暴露于多个核酸探针;以及通过使用错误检查和/或错误校正检测技术确定多个核酸探针在细胞核内的结合来确定细胞核内的核酸丰度和/或空间分布。

[0020] 在另一组实施方案中,该方法包括将基因组的多个核酸靶标与多个代码字相关联;将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针,其中所述多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读取序列的第二部分,其中每个读取序列表示多个代码字内的位置的值;将样品暴露于多个衔接子,其中至少一些衔接子包含与一个或多个读取序列基本上互补的第一部分,以及包含一个或多个鉴定序列的第二部分;将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列,其中至少一些读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分,以及包含信号传导实体的第二部分;确定样品中至少一些位置中的信号传导实体;和灭活样品中至少一些位置中的信号传导实体;重复将样品暴露于一个轮次、确定信号传导实体和灭活信号传导实体的步骤,其中在所有轮次中使用不超过10个不同的信号传导实体;基于确定样品中的信号传导实体确定位置处的代码字;基于所述代码字确定样品中的核酸靶标。

[0021] 根据又一组实施方案,该方法包括将基因组的多个核酸靶标与多个代码字相关

联;将含有怀疑含有基因组的细胞的样品暴露于多个核酸探针,其中所述多个核酸探针中的至少一些含有包含靶序列的第一部分和包含一个或多个读取序列的第二部分,其中每个读取序列表示多个代码字内的位置的值;将样品暴露于多个衔接子,其中至少一些衔接子包含与一个或多个读取序列基本上互补的第一部分,以及包含一个或多个鉴定序列的第二部分;将样品暴露于一个轮次的一个或多个读出探针以确定一个或多个鉴定序列,其中至少一些读出探针含有包含与鉴定序列之一基本上互补的序列的第一部分,以及包含信号传导实体的第二部分;确定样品中至少一些位置中的信号传导实体;和灭活样品中至少一些位置中的信号传导实体;重复将样品暴露于一个轮次、确定信号传导实体和灭活信号传导实体的步骤,其中信号传导实体中的至少一个在多于一个轮次中使用;基于确定样品中的信号传导实体确定位置处的代码字;以及基于所述代码字确定样品中的核酸靶标。

[0022] 根据另一组实施方案,该方法包括确定新生RNA在细胞核内的位置;确定DNA在细胞核内的位置;并确定细胞核内核斑点的位置。

[0023] 在另一组实施方案中,该方法包括确定新生RNA在细胞核内的位置;确定DNA在细胞核内的位置;并确定蛋白质在细胞核内的位置。在另一组实施方案中,该方法包括确定新生RNA在细胞核内的位置;确定DNA在细胞核内的位置;以及确定核酸在细胞核内的位置,其中所述核酸不是新生RNA或DNA。

[0024] 一些方面包括制造本文所述的一个或多个实施方案的方法。此外,一些方面包括使用本文描述的一个或多个实施方案的方法。

[0025] 当结合附图考虑时,本公开内容的其他优点和新颖特征将从以下各种非限制性实施方案的详细描述中变得明显。

[0026] 附图简要说明

[0027] 将参考附图以示例的方式描述本发明的非限制性实施方案,这些附图是示意性的并且不旨在按比例绘制。在图中,所示出的每个相同或几乎相同的组件通常由单个数字表示。为清楚起见,并非在每幅图中都标明了每个组件,并且在不需要说明以使本领域的普通技术人员能够理解本公开内容时没有示出本发明的每个实施方案的每个组件。

[0028] 图1A-1I显示了根据某些实施方案的基因组规模染色质成像;

[0029] 图2A-2E显示在另一个实施方案中的跨染色体接触富集;

[0030] 图3A-3H显示在另一个实施方案中,在核结构的背景下染色质和转录活动的基因组规模成像;

[0031] 图4A-4F显示在另一个实施方案中活性染色质之间的跨染色体相互作用;

[0032] 图5A-5E示出了一个实施方案中的可饱和放大系统;

[0033] 图6A-6B显示了在又一个实施方案中的接触频率矩阵;

[0034] 图7A-7C显示了在又一个实施方案中源自基因组规模成像和与整体Hi-C数据的比较的亚染色体结构;

[0035] 图8显示在另一个实施方案中,染色质成像实验的在重复实验之间的再现性;

[0036] 图9A-9B显示在某些实施方案中单个细胞中不同的空间分布;

[0037] 图10A-10B显示了在其他实施方案中的新生RNA转录物成像;

[0038] 图11显示在某些实施方案中隔壁-B基因座与核纤层的关联;

[0039] 图12显示了在一些实施方案中隔壁-A基因座与核斑点的关联;

- [0040] 图13A-13C显示在另一个实施方案中转录抑制后核纤层和核斑点关联的变化；
- [0041] 图14显示在另一个实施方案中，每个成像的基因座附近的跨染色体A基因座的局部密度；
- [0042] 图15A-15B显示在另一个实施方案中染色质基因座之间活性-活性跨染色体相互作用的富集；
- [0043] 图16A-16B显示在另一个实施方案中活性-活性跨染色体相互作用的富集；
- [0044] 图17A-17M显示在一个实施方案中，通过依次杂交和表征单个细胞中的染色质结构域的高分辨率全染色体追踪；
- [0045] 图18A-18I显示在另一个实施方案中，单个染色体中的隔室结构以及转录活动与局部染色质含量之间的关系；
- [0046] 图19A-19H显示在又一个实施方案中结构域间相互作用对其A/B组成和基因组距离的依赖性；
- [0047] 图20A-20H显示在又一个实施方案中通过大规模多重组合FISH进行的基因组规模染色质成像；
- [0048] 图21A-21E显示根据一个实施方案的跨染色体相互作用中活性-活性染色质相互作用的富集；
- [0049] 图22A-22J显示根据另一个实施方案在核结构的背景下染色质和转录活动的多模态基因组规模成像；
- [0050] 图23A-23D显示在又一个实施方案中转录活动与跨染色体活性染色质的局部富集之间的相关性；
- [0051] 图24A-24N显示在又一个实施方案中通过依次杂交的高分辨率全染色体追踪，以及与Hi-C相比Chr21结构特征的整体统计；
- [0052] 图25A-25G显示在又一个实施方案中Chr21和Chr2的整体A/B隔室分析；
- [0053] 图26A-26J显示在又一个实施方案中对RNA和DNA FISH探针串扰的测量；
- [0054] 图27A-27J显示在一个实施方案中通过组合FISH进行的基因组规模成像：定位误差、再现性以及和Hi-C的比较；
- [0055] 图28A-28B显示根据另一个实施方案，隔室-A和隔室-B基因座在细胞核中显示出不同的空间分布；
- [0056] 图29A-29F显示在又一个实施方案中转录抑制对跨染色体染色质相互作用和染色质基因座的核体缔合率的影响；和
- [0057] 图30A-30D显示在又一个实施方案中在不同核环境中跨染色体活性染色质相互作用的富集。
- [0058] 发明详述
- [0059] 本发明一般涉及基因组学。一些实施方案涉及在序列空间中以高通量对基因组或基因组的一部分的3D组织进行成像。一些实施方案涉及在转录活动和核结构的背景下对基因组或基因组的一部分的3D组织进行成像。此外，某些实施方案涉及染色质结构、3D染色质组织、跨染色体相互作用和染色质-核结构相互作用以及它们与转录的关系等。此外，各种实施方案涉及允许在核结构和转录活动的背景下对基因组或基因组的一部分的3D组织进行作图的成像方法。一些实施方案涉及用于在染色体或基因组规模上对染色质基因座和/

或新生RNA转录物进行成像的大规模多重荧光原位杂交方法。在一些情况下,可以同时数百个基因组基因座进行成像。在一些情况下,可以在各种核结构的情况下同时对约1000个基因组基因座和/或这些基因座内约1000个基因的转录活动进行成像。在一些情况下,可以观察到染色质结构域和隔室。在一些情况下,可以观察到以转录相关方式富含活性染色质相互作用的广泛的跨染色体相互作用。在一些情况下,可以观察到在基因组中与核斑点和核纤层的转录依赖性染色质相互作用。

[0060] 染色质的三维(3D)组织调节许多基因组功能。由于缺乏允许在染色体尺度和基因组规模上在其天然背景下直接可视化染色质组织的工具,阻碍了对3D基因组组织的理解。因此,在某些实施方案中描述的是通过在多个杂交轮次上连续成像的多重FISH方法,例如,使得每个轮次使用单色或双色或三色成像靶向一个或两个或三个基因组基因座。在其他实施方案中描述的是组合FISH方法,许多染色质基因座在每个轮次中同时成像,并且它们的不同身份基于它们出现在其中的轮次的组合来确定。这通常基于MERFISH和其他方法,例如,如在题为“Systems and Methods for Determining Nucleic Acids”的国际专利申请公开号W02016/018960;和题为“Probe Library Construction”的国际专利申请公开号W02016/018963中所讨论的,每个都通过引用整体并入本文。诸如本文所讨论的方法可用于对单个细胞中不同的染色质基因座进行成像,并可用于提供对染色质结构、它们与转录的关系、与核蛋白的相互作用等的深入了解。

[0061] 一些方面一般涉及使用多重FISH或其他技术(在一些情况下使用MERFISH,包括本文所述的那些)以对例如细胞中的染色体或染色质进行成像的系统和方法。此外,某些实施方案一般涉及对单个细胞中的至少100个不同基因组基因座、至少500个不同基因组基因座或至少1,000个不同基因组基因座等进行成像和/或确定的系统和方法。在一些情况下,可以确定细胞的其他部分或细胞核,例如,存在于细胞核内的RNA例如新生RNA、核斑点、核仁、核纤层、其他核结构或蛋白质等。作为非限制性示例,对于细胞核,可以确定染色体或染色质、新生RNA、核斑点、核仁和/或核纤层的位置。

[0062] 某些实施方案涉及确定样品,其可以包括细胞培养物、细胞悬浮液、生物组织、活组织检查、生物体等。样品也可以是无细胞的,但在某些情况下仍含有核酸。如果样品含有细胞,则细胞可以是人细胞或任何其他合适的细胞,例如哺乳动物细胞、鱼细胞、昆虫细胞、植物细胞等。在某些情况下可能存在不止一个细胞。

[0063] 在样品中,待确定的靶标可以包括核酸、蛋白质等。例如,这些可能存在于样品内的细胞核内。在某些实施方案中,可以确定细胞内的染色质,例如,相对于细胞的核结构,包括核斑点、核仁、核纤层或核结构或蛋白质。在一些情况下,染色质基因座和/或RNA转录物可以在细胞内确定,例如以染色体或基因组规模确定。

[0064] 现在讨论这种方法的一个例子。然而,应该理解的是,该方法是作为解释而不是限制的方式呈现的;本文还讨论了其他方面和实施方案。在一组实施方案中,测定细胞内,例如细胞核内的核酸。这些通常包括DNA(例如,基因组DNA,它可能以染色质的形式存在,例如与蛋白质诸如组蛋白一起包装)和RNA(例如,当DNA转录成RNA时,在转录阶段开始时;核内的这种RNA有时被称为新生RNA)。与检测可能存在于细胞内任何地方的RNA的技术相比,DNA高度堆积在细胞核内,这使得确定其结构变得更加困难。例如,DNA可以作为染色体或染色质包装在细胞内,并且这种DNA通常可以在细胞核内缠绕或紧密包装在一起。因此,在某些

实施方案中,可以选择DNA靶标以在空间上分离。

[0065] 在一些情况下,样品与核酸探针进行多个轮次的杂交,其中一个或多个轮次使用单色或多色成像靶向一个或多个核酸靶标。在一些情况下,核酸靶标的身份基于它们被成像的哪个轮次和/或哪个颜色通道来确定。在一些情况下,测定核酸靶标的位置。在一些情况下,测定至少50、至少100、至少500、至少1000、至少5000或至少10,000个核酸靶标。在一些情况下,核酸靶标是基因组基因座。在一些情况下,核酸靶标是基因组基因座和/或新生RNA转录物。在一些情况下,基因组基因座的位置用于确定细胞中染色质的三维组织或基因组的三维组织。

[0066] 在一些情况下,设计了能够靶向细胞内,例如细胞核内的核酸的初级核酸探针。探针各自包含与核酸靶标之一结合的靶序列。探针还可以含有包含一个或多个“读出序列”的部分,其可用于确定初级核酸探针的身份和位置。在一些实施方案中,初级核酸探针可以包含多个读出序列。这些可以使用一个或多个轮次的次级核酸探针(称为读出探针,其能够结合初级核酸探针的读出序列)单独读取。读出探针还可以包含信号传导实体,例如荧光实体,例如,其可以使用各种显微技术来确定。在一些情况下,可以依次应用多个轮次的读出探针,使得将一种类型的读出探针应用于样品并确定样品内的荧光,然后灭活或去除读出探针或读出探针上的信号传导实体并应用下一种类型的读出探针。在一些情况下,样品内的位置可能与多个读出探针相关联,并且该信息可以被数字化以供分析。

[0067] 在一些情况下,可以依次应用多个轮次的读出探针,使得在每个轮次中将多于一种类型的读出探针应用到样品和/或使用多色成像确定样品内的荧光,然后灭活或去除读出探针和/或读出探针上的信号传导实体,并且应用下一组的多于一种类型的读出探针。在一些情况下,样品内的位置可能与多个读出探针相关联,并且该信息可以被数字化以供分析。

[0068] 在一些情况下,初级核酸探针和核酸靶标的位置可以使用一个或多个轮次的读出探针来确定。例如,可以有至少10个、至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少150个、至少200个、至少250个、至少300个、至少400个、至少500个、至少600个、至少700个、至少800个、至少900个、至少1000个等轮次的读出探针。因此,在一些情况下,可将样品暴露于多个轮次的应用读出探针、确定样品内的探针(例如,使用信号传导实体,如本文所述)和去除或灭活次级核酸探针。

[0069] 此外,应该理解,读出探针不必全部不同。在一些情况下,可以使用多于一个轮次的相同的读出探针,例如,以作为对照确定样品中是否由于供应多个轮次的核酸或其他化学品的影响例如随着时间的推移发生了任何降解和/或移动等。

[0070] 在一些情况下,样品与核酸探针进行多个轮次的杂交,每个轮次进行单色或多色成像。在一些情况下,核酸靶标的身份是基于它们被成像的轮次和/或颜色通道的组合来确定的。在一些情况下,确定核酸靶标的位置。在一些情况下,测定至少50、至少100、至少500、至少1000、至少5000或至少10,000个核酸靶标。在一些情况下,核酸靶标是基因组基因座。在一些情况下,核酸靶标是基因组基因座和/或新生RNA转录物。在一些情况下,基因组基因座的位置用于确定细胞中染色质的三维组织或基因组的三维组织。

[0071] 在一些情况下,设计了能够靶向细胞内例如细胞核内的核酸的初级核酸探针(也称为编码探针)。探针各自包含与核酸靶标之一结合的靶序列。探针还可以含有包含一个或

多个“读出序列”的部分,所述“读出序列”可用于确定初级或编码核酸探针的身份和位置。在一些实施方案中,初级或编码核酸探针可以包含多个读出序列。这些可以使用一个或多个轮次的读出探针单独读取,读出探针能够结合初级或编码核酸探针的读出序列。读出探针还可以包含信号传导实体,例如荧光实体,例如,其可以使用各种显微技术来确定。在一些情况下,可以依次应用多个轮次的读出探针,使得将一种类型的读出探针应用于样品并确定样品内的荧光,然后灭活或去除读出探针或读出探针上的信号传导实体并应用下一种类型的读出探针。在一些情况下,样品内的位置可能与多个读出探针相关联,并且该信息可以被数字化以供分析。在一些情况下,可以依次应用多个轮次的读出探针,使得在每个轮次中将多于一种类型的读出探针应用到样品并且使用多色成像确定样品内的荧光,然后灭活或去除读出探针或读出探针上的信号传导实体,并应用下一组的多于一种类型的读出探针。在一些情况下,样品内的位置可能与多个读出探针相关联,并且该信息可以被数字化以供分析。

[0072] 在一些情况下,可以使用一个或多个轮次的读出探针来确定初级或编码核酸探针和核酸靶标的位置。例如,可以有至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少12个、至少16个、至少20个、至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少200个、至少300个、至少400个、至少500个、至少600个、至少700个、至少800个、至少900个、至少1,000个等轮次的读出探针。因此,在一些情况下,可将样品暴露于多个轮次的应用读出探针,确定样品内的探针(例如,使用信号传导实体,如本文所述)和去除或灭活次级核酸探针。

[0073] 在一些实施方案中可以设计初级或编码核酸探针,使得可以使用读出序列的不同组合来确定样品内的不同靶标,而不必要求每个读出序列是唯一的。作为非限制性实例,如果靶向每个核酸靶标的初级或编码核酸探针组仅包含2个读出序列,则对于4个可能的读出序列A、B、C和D,可以鉴定多达6个不同的靶标,例如,对应于AB、AD、CB、CD、AC和DB。

[0074] 然而,在一些实施方案中,并非所有可能的读出序列组合都将被使用。相反,一些组合可能不被分配给细胞核中的任何靶标,例如,可以不使用具有这些组合的初级或编码核酸探针。在一些情况下,可排列用于初级或编码核酸探针的读出序列的有效组合以形成错误检查和/或错误校正代码空间。使用这种方法,样品中与有效初级核酸探针不对应的读出序列的测定可以使用错误检查来确定为是错误的,并且在一些情况下,甚至可以使用错误校正来校正,例如,以对应于有效的初级核酸探针。

[0075] 虽然之前已经描述了这些方法,例如在题为“Systems and Methods for Determining Nucleic Acids”的国际专利申请公开号W02016/018960;和题为“Probe Library Construction”的国际专利申请公开号W02016/018963中,但这些方法并未应用于在细胞核内更受限制的环境中对DNA进行成像。如前所述,与细胞的其他部分不同,细胞核含有非常高部分的核酸,包括几乎所有的基因组DNA,和通常高浓度的RNA(例如,新生RNA)。

[0076] 因此,为了接近细胞核内的DNA,可以选择初级或编码核酸探针的靶标,使得核内的结合以空间分离的方式发生。例如,可以选择靶标,使得它们在基因组空间中分开,例如,在基因组内分开至少10,000bp至少30,000bp至少100,000bp、至少300,000bp、至少1,000,000bp,或使得基因组空间包含不超过100、不超过200、不超过300、不超过500、不超过1000、不超过5000、不超过10,000、不超过50,000、不超过100,000个核酸靶标。在一些情况下,也

可以使用不止一种类型的荧光探针或“颜色”，例如，以允许在细胞核内测定更多靶标。

[0077] 在一些实施方案中，细胞和/或细胞核也可以被修饰以允许这样的探针到达其中的核酸。例如，细胞可以被透化或“固定”以允许核酸探针进入。此外，在一些实施方案中，DNA可以变性，例如通过加热，以允许初级或编码核酸探针更容易接近DNA。这通常不适用于RNA测定，因为RNA是单链的，而DNA通常是双链的。此外，在某些实施方案中，在可以研究DNA之前，必须去除和/或灭活细胞核内的RNA，例如，以防止靶向DNA的探针与RNA结合。例如，在一些情况下，可以将酶（例如核糖核酸酶）应用于细胞核，以防止RNA干扰DNA测定。

[0078] 此外，需要注意的是，在某些实施方案中，还可以测定细胞核内的RNA。这可能特别有价值，例如，在研究细胞核内DNA和RNA的空间位置以及它们如何相互关联时。因此，在一组实施方案中，细胞核内的RNA可以在如上所述的RNA的去除或灭活之前测定，例如，类似于上文对基因组DNA所述的那些。

[0079] 此外，在某些实施方案中，还可以测定细胞内的蛋白质，例如细胞核内的蛋白质。例子包括但不限于核斑点、核仁或组蛋白。可以使用多种测定蛋白质的方法。例如，在一组实施方案中，可以使用免疫荧光测定。在另一组实施方案中，可以使用“夹心测定法”，其中应用能够特异性结合核蛋白的一抗，然后使用能够特异性结合一抗的二抗，其中二抗含有信号传导实体，例如荧光实体。可以对与上述相同的样品或相同的细胞核进行此类蛋白质测定，例如，在测定细胞核内的核酸之前或之后。因此，在一些情况下，可以例如在空间上测定细胞核内的蛋白质和核酸。

[0080] 上述讨论是一个实施方案的非限制性实例，该实施方案可用于测定细胞核内的核酸，例如基因组DNA和/或新生RNA。然而，其他实施方案也是可能的。因此，更一般地，各个方面涉及针对核酸的各种系统和方法。

[0081] 如所提到的，在某些实施方案中，可以测定细胞例如细胞核内的DNA、RNA和蛋白质中的一种、两种或更多种。待测定的细胞核内的核酸可以包括例如DNA（例如基因组DNA）、RNA或存在于细胞（或其它样品）内的其它核酸。核酸对于细胞可以是内源的，或被添加到细胞中。例如，核酸可以是病毒的，或人工产生的。在一些情况下，待测定的核酸可以由细胞表达。在一些实施方案中，核酸是RNA。RNA可以是编码和/或非编码RNA。例如，RNA可以编码蛋白质。可在细胞内研究的RNA的非限制性实例包括mRNA、siRNA、rRNA、miRNA、tRNA、lncRNA、snoRNA、snRNA、exRNA、piRNA等。

[0082] 在一组实施方案中，可以测定细胞基因组的全部或至少重要部分。测定的基因组区段可以是连续的或散布在基因组上。例如，在一些情况下，在细胞内测定至少4个基因组区段，在一些情况下，可以在细胞内测定至少3个、至少4个、至少7个、至少8个、至少12个、至少14个、至少15个、至少16个、至少22个、至少30个、至少31个、至少32个、至少50个、至少63个、至少64个、至少72个、至少75个、至少100个、至少127个、至少128个、至少140个、至少255个、至少256个、至少500个、至少1,000个、至少1,500个、至少2,000个、至少2,500个、至少3,000个、至少4,000个、至少5,000个、至少7,500个、至少10,000个、至少12,000个、至少15,000个、至少20,000个、至少25,000个、至少30,000个、至少40,000个、至少50,000个、至少75,000个或至少100,000个基因组区段。

[0083] 在一些情况下，可以测定细胞的整个基因组。应当理解，基因组通常包括细胞内产生的所有DNA分子，而不仅仅是染色体DNA。因此，例如，在一些情况下，基因组还可以包括线

粒体DNA、叶绿体DNA、质粒DNA等,例如,除了(或代替)染色体DNA。在一些实施方案中,可以测定细胞的至少约0.01%、至少约0.1%、至少约1%、至少约5%、至少约10%、至少约15%、至少约20%、至少约25%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%或100%的基因组。

[0084] 此外,在一些实施方案中,可以研究细胞内或细胞核内的大部分核酸。例如,在一些情况下,可以测定细胞核内的RNA,例如新生RNA。此外,在一些情况下,可以测定细胞内存在的足够量的RNA,以便产生细胞的部分或完整转录组。在一些情况下,测定细胞内或细胞核内的至少4种类型的RNA(例如,mRNA、新生RNA等),和在一些情况下,可以测定细胞内或细胞核内的至少3个、至少4个、至少7个、至少8个、至少12个、至少14个、至少15个、至少16个、至少20个、至少22个、至少30个、至少31个、至少32个、至少50个、至少63个、至少64个、至少72个、至少75个、至少100个、至少127个、至少128个、至少140个、至少255个、至少256个、至少500个、至少1,000个、至少1,500个、至少2,000个、至少2,500个、至少3,000个、至少4,000个、至少5,000个、至少7,500个、至少10,000个、至少12,000个、至少15,000个、至少20,000个、至少25,000个、至少30,000个、至少40,000个、至少50,000、至少75,000或至少100,000个类型的RNA。

[0085] 在一些情况下,可以测定细胞的转录组。应该理解,转录组通常包括细胞内产生的所有RNA分子,而不仅仅是mRNA。因此,例如,在某些情况下,转录组还可以包括rRNA、tRNA、siRNA等。在一些实施方案中,可以测定细胞的至少约0.01%、至少约0.1%、至少约1%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或100%的转录组。此外,在一些情况下,可以测定细胞核的转录组。

[0086] 此外,在一些实施方案中,待测定的其他靶标可以包括与核酸、蛋白质等连接的靶标。例如,在一组实施方案中,能够识别靶标的结合实体可以与核酸探针缀合。结合实体可以是能够识别靶标(例如,特异性地或非特异性地)的任何实体。非限制性实例包括酶、抗体、受体、互补核酸链、适配体等。例如,寡核苷酸连接的抗体可用于测定靶标。靶标能够结合寡核苷酸连接的抗体,并且寡核苷酸如本文所讨论的那样测定。

[0087] 靶标例如细胞或其他样品中的核酸的测定可以是定性的和/或定量的。此外,测定还可以是空间的,例如,可以在二维或三维中确定细胞或其他样品内的核酸或其他靶标的位置。在一些实施方案中,可以确定细胞或其他样品内的核酸或其他靶标的位置、数量和/或浓度。

[0088] 如所提到的,在一组实施方案中,可以研究细胞核内的DNA,例如细胞的基因组DNA,例如,使用如本文讨论的核酸探针,例如,包括使用顺序成像或使用利用错误检测和/或错误校正代码的组合成像。

[0089] 在某些实施方案中,可以选择细胞内或细胞核内的DNA靶标或与DNA靶标相关的代码,使得靶标在每个轮次的成像中在空间上分离,例如在基因组空间中,或在基于染色质组织知识(例如,将染色体组织成紧凑的区域)的物理空间中。例如,这对于能够鉴定细胞核的细胞内的不同靶标可能是有用的,例如,使用诸如本文讨论的那些技术。

[0090] 可以使用任何合适的技术(例如随机地或具有基本上均匀的概率分布等)来选择基因组空间内的靶标。在某些实施方案中,可以单独地选择靶标以确保空间分离。此外,在

一些实施方案中,靶标可以被选择为基因组内的那些感兴趣的靶标,例如,用于特定研究。

[0091] 例如,在一些实施方案中,可以在基因组空间内选择靶标,使得细胞核将具有不超过一定数量的核酸靶标。例如,可以选择靶标使得基因组空间包含不超过100,000、不超过10,000、不超过8,000、不超过6,000、不超过5,000、不超过4,000、不超过3,000、不超过2,000、不超过1,500、不超过1,000、不超过900、不超过800、不超过700、不超过600、不超过500、不超过400、不超过300、不超过200、不超过100个核酸靶标、不超过30个核酸靶标或者不超过10个核酸靶标。此外,在一些实施方案中,可以选择靶标以使得基因组空间包含至少10、至少30、至少50、至少100、至少200、至少300、至少500、至少1,000、至少1,500、至少2,000、至少3,000、至少5,000、至少10,000、至少100,000等个核酸靶标。在某些实施方案中,这些中的任何一种的组合也是可能的,例如,可以有30至100、3,000至5,000、500至1,500等个核酸靶标。可以选择(例如,选择性地、随机地等)这样的靶标,如本文所讨论的。

[0092] 作为另一个例子,在一些实施方案中,可以选择靶标使得基因组内的染色体具有不超过一定数量的核酸靶标(例如,基因组基因座)。例如,可以选择靶标使得每个染色体具有不超过10,000、不超过1000、不超过500、不超过400、不超过300、不超过200、不超过150、不超过125、不超过100、不超过90、不超过80、不超过70、不超过60、不超过50、不超过40、不超过30、不超过20或不超过10个核酸靶标。在一些情况下,可以选择靶标以具有至少10个、至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少125个、至少150个、至少200个、至少300个、至少400个、至少1,000个、至少10,000个等核酸靶标。在一些情况下,可以选择这些的组合,例如,染色体可以具有30至50、40至100、50至60、30至80等个核酸靶标。此外,不同的染色体可以独立地具有相同或不同数量的核酸靶标,例如,包括本文所述的范围。

[0093] 可以选择(例如选择性地、随机地等)这样的靶标,例如,如本文所讨论的。作为非限制性实例,可以选择基因组内的核酸靶标以具有特定的结构或功能特性,例如启动子、增强子和由特定核结构蛋白结合的基因座。在一些情况下,一些或所有的核酸靶标可以是它们各自的染色体所特有的核酸靶标。

[0094] 在又一个实施方案中,可以选择靶标以被最少一定数量的核苷酸分开,例如,以促进空间上分开的靶标的分布。例如,可以在基因组内选择靶标,使得每个靶标相隔至少1,000、至少3,000、至少5,000、至少10,000、至少30,000、至少50,000、至少100,000、至少300,000、至少500,000、至少1,000,000、至少3,000,000、至少5,000,000、至少10,000,000等个核苷酸。此外,在某些实施方案中,可以在基因组内选择靶标,使得每个靶标相隔不超过10,000,000、不超过5,000,000、不超过3,000,000、不超过1,000,000、不超过500,000、不超过300,000、不超过100,000、不超过50,000、不超过30,000、不超过10,000个核苷酸。在某些实施方案中,这些中的任何一种的组合也是可能的,例如,靶标可以相隔30,000至100,000、3,000,000至5,000,000、500,000至1,000,000等个核苷酸。可以例如选择性地、随机地等选择这样的靶标,如本文所讨论的。

[0095] 此外,在一组实施方案中,可以研究细胞核内的RNA,例如细胞的新生RNA,例如,代替或附加于如上所述的核内DNA。在一些情况下,例如,可以测定细胞核内的RNA,然后可以测定细胞核内的DNA。

[0096] 在一些情况下,在测定RNA之后,可以在测定DNA之前去除或灭活RNA。这可以促进

DNA和RNA测定的分开,例如,因为不会出现可能使DNA测定复杂化的RNA信号。去除或灭活RNA的方法的例子包括使用核糖核酸酶,例如核糖核酸内切酶或核糖核酸外切酶。具体的非限制性实例包括核糖核酸酶A、核糖核酸酶H、核糖核酸酶III、核糖核酸酶L、核糖核酸酶P、核糖核酸酶PhyM、核糖核酸酶T1、核糖核酸酶T2、核糖核酸酶U2、核糖核酸酶V、PNPase、核糖核酸酶PH、核糖核酸酶R、核糖核酸酶D、核糖核酸酶T、寡核糖核酸酶、核糖核酸外切酶I、核糖核酸外切酶II等。

[0097] 然而,应当理解,在其他实施方案中,DNA可以在RNA之前测定,和/或两者可以同时测定。例如,可以在测定后使用诸如脱氧核糖核酸外切酶或脱氧核糖核酸内切酶的脱氧核糖核酸酶去除或灭活DNA。实例包括但不限于脱氧核糖核酸酶I (DNase I)、脱氧核糖核酸酶II (DNase II)、DNase IV、UvrABC核酸内切酶等。作为另一个例子,DNA可以通过暴露于限制性内切核酸酶而被降解。许多这样的核酸酶是可商购获得的。

[0098] 细胞核内的RNA可以使用任何合适的技术来测定,并且可以使用与用于测定细胞核内的DNA相同或不同的技术来测定。在一个实施方案中,可以使用MERFISH来测定RNA。参见,例如,题为“Systems and Methods for Determining Nucleic Acids”的国际专利申请公开号W02016/018960;和题为“Probe Library Construction”的国际专利申请公开号W02016/018963,每个都通过引用整体并入本文。在另一个实施方案中,可以使用多个核酸探针来测定RNA,例如,如本文所讨论的。例如,在一些实施方案中,可以使用核酸例如编码核酸探针、初级扩增核酸、次级扩增核酸等来测定RNA,如下文所讨论的。在一些情况下,核酸探针可以定义错误检测和/或错误校正代码,例如,如本文所讨论的。

[0099] 在一些实施方案中,DNA例如基因组DNA可以使用核酸例如编码核酸探针、初级扩增核酸、次级扩增核酸等来测定,如本文所述。在一些情况下,核酸探针可以定义错误检测和/或错误校正代码,例如,如本文所讨论的。

[0100] 此外,在一组实施方案中,例如,除了存在于细胞核内的核酸之外,还可以使用诸如上述那些技术来研究细胞核内的蛋白质。可以研究的蛋白质的例子包括但不限于核斑点、核仁、核纤层或组蛋白等。斑点是富含前信使RNA剪接因子的结构,并且可能位于哺乳动物细胞的核质的染色质间区域。核仁是围绕编码核糖体RNA (rRNA) 的高度转录的基因组基因座形成的结构,并且可能富含rRNA和与之相关联的转录机制。核纤层是与内核膜相关联的蛋白质结构,并且可能富含中间纤丝(核纤层蛋白)以及转录灭活的染色质。组蛋白是一种蛋白质,用于在细胞核内将DNA包裹或折叠成更紧凑的复合物,形成染色质。

[0101] 可以使用多种测定蛋白质的方法。例如,在一组实施方案中,可以使用免疫荧光测定。在另一组实施方案中,可以使用“夹心测定法”,其中应用能够特异性结合核蛋白的一抗,然后使用能够特异性结合一抗的二抗,其中二抗含有信号传导实体,例如荧光实体或可以检测的寡核苷酸,例如,使用与荧光实体连接的互补寡核苷酸。可以对与上述相同的样品或相同的细胞核进行此类蛋白质测定,例如,在测定细胞核内的核酸之前或之后。因此,在一些情况下,细胞核内的蛋白质和核酸可以例如在空间上测定。

[0102] 如所提到的,在诸如本文所述的各种实施方案中,多个核酸探针可用于测定细胞或其他样品内例如细胞核内的一个或多个靶标。探针可包含核酸(或可与核酸杂交(例如特异性地)的实体),例如DNA、RNA、LNA(锁核酸)、PNA(肽核酸)和/或它们的组合。核酸探针的例子包括但不限于在题为“Systems and Methods for Determining Nucleic Acids”的国

际专利申请公开号W02016/018960;和题为“Probe Library Construction”的国际专利申请公开号W02016/018963中描述的那些,每个都通过引用整体并入本文。在一些情况下,核酸探针中也可能存在额外的组分,例如,如下文所讨论的。此外,可以使用任何合适的方法将核酸探针引入细胞,例如以靶向其细胞核。

[0103] 例如,在一些实施方案中,在引入核酸探针之前固定细胞,例如以保持核酸或其他靶标在细胞内,例如在其细胞核内的位置。用于固定细胞的技术是本领域普通技术人员已知的。作为非限制性示例,可以使用诸如甲醛、多聚甲醛、戊二醛、乙醇、甲醇、丙酮、乙酸等的化学品来固定细胞。在一个实施方案中,可以使用HEPES-谷氨酸缓冲液介导的有机溶剂(HOPE)固定细胞。

[0104] 此外,在一些情况下,细胞(或其他样品)可能被固定多于一次,例如,在相对较长的实验期间。例如,可以在实验开始后重新固定样品,例如在将细胞核暴露于多个核酸探针之后。例如,可以至少每7天一次、至少每4天一次、至少每2天一次、至少每天一次、至少每12小时一次、至少每6小时一次、至少每3小时一次等固定细胞或其他样品。在一些情况下,这可以在例如暴露于核酸探针(例如,初级或次级核酸探针)等的不同轮次之间进行。在一些情况下,样品可以固定一定次数,例如2、3、4、5、6、7、8、9、10次或任何其他合适的次数。如果发生多次固定,这些固定可独立地使用相同或不同的固定技术。

[0105] 可以使用任何合适的方法将核酸探针引入细胞(或其他样品)中。在一些情况下,细胞可以被充分透化,从而可以通过在细胞周围流动含有核酸探针的流体将核酸探针引入细胞中。在一些情况下,作为固定过程的一部分,细胞可能被充分透化;在其他实施方案中,细胞可以通过暴露于某些化学物质例如乙醇、甲醇、Triton等而被透化。此外,在一些实施方案中,诸如电穿孔或显微注射的技术可用于将核酸探针引入细胞或其他样品中。

[0106] 因此,某些方面一般涉及被引入细胞(或其他样品)中的核酸探针。取决于应用,探针可以包含可以与核酸杂交(通常通过Watson-Crick碱基配对)的多种实体例如DNA、RNA、LNA、PNA等中的任一种。核酸探针通常包含能够结合靶标(例如核酸靶标)的至少一部分的靶序列。在一些情况下,结合可以是特异性结合(例如,通过互补结合)。当被引入细胞或其他系统时,靶序列可能能够结合特定靶标(例如,新生RNA、基因组DNA、mRNA或本文讨论的其他核酸)。如下文所讨论的,核酸探针还可以包含一个或多个读出序列。

[0107] 在一些情况下,可以将多于一个类型的核酸探针应用于样品,例如,依次或同时地。例如,可以有应用于样品(例如应用于细胞以靶向其细胞核)的至少2个、至少5个、至少10个、至少25个、至少50个、至少75个、至少100个、至少300个、至少1,000个、至少3,000个、至少10,000个、至少30,000个、至少100,000个、至少300,000个、至少1,000,000个可区分的核酸探针。在一些情况下,可以依次添加核酸探针。然而,在一些情况下,可以同时添加多于一种核酸探针。

[0108] 核酸探针可包括一个或多个靶序列,其可位于核酸探针内的任何位置。靶序列可以包含与靶(例如核酸靶标,其可以在细胞核内)的一部分基本上互补的区域。例如,在一些情况下,这些部分可以是至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%、至少92%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%互补的,例如,以产生特异性结合。通常,互补性是根据Watson-Crick核苷酸碱基配对确定的。

[0109] 在一些情况下,靶序列的长度可以是至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少50、至少60、至少65、至少75、至少100、至少125、至少150、至少175、至少200、至少250、至少300、至少350、至少400或至少450个核苷酸。在一些情况下,靶序列的长度可能不超过500、不超过450、不超过400、不超过350、不超过300、不超过250、不超过200、不超过175、不超过150、不超过125、不超过100、不超过75、不超过60、不超过65、不超过60、不超过55、不超过50、不超过45、不超过40、不超过35、不超过30、不超过20或不超过10个核苷酸。这些中的任何一个的组合也是可能的,例如,靶序列可以具有10至30个核苷酸、20至40个核苷酸、5至50个核苷酸、10至200个核苷酸或25至35个核苷酸、10至300个核苷酸等的长度。

[0110] 在一些实施方案中,可以选择细胞或细胞核内的核酸靶标或与核酸靶标相关的代码,使得靶标在每个轮次的成像中在空间上分离,例如在基因组空间中,或在基于染色质组织的先前知识(例如,将染色体组织成紧凑的区域)的物理空间中。

[0111] 此外,在一些情况下,核酸探针的靶序列可以参考怀疑存在于细胞或其他样品中(例如,在细胞核中)的靶标来确定。例如,针对蛋白质(例如,核斑点、核纤层等)的核酸靶标可以使用蛋白质的序列来确定,例如,通过确定被表达以形成蛋白质的核酸。在一些情况下,仅使用编码蛋白质的核酸的一部分,例如,具有如上所讨论的长度。

[0112] 根据某些实施方案,可以使用多于一种可用于鉴定特定靶标的靶序列。例如,可以依次和/或同时使用多个探针,它们能够结合或杂交到相同靶标的相同或不同区域。杂交通常指互补单链核酸通过Watson-Crick核苷酸碱基配对(例如氢键、鸟嘌呤-胞嘧啶和腺嘌呤-胸腺嘧啶)缔合以形成双链核酸的退火过程。

[0113] 在一些实施方案中,核酸探针还可以包含一个或多个“读出”序列。读出序列可用于鉴定核酸探针,例如,通过与信号传导实体缔合,如下文所讨论的。在一些实施方案中,核酸探针可包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16或更多、20或更多、24或更多、32或更多、40或更多、48或更多、50或更多、64或更多、75或更多、100或更多、128或更多、256或更多、512或更多、1024或更多、2048或更多、4096或更多、8192或更多、16384或更多、32768或更多、65536或更多、131072或更多、262144或更多、524288或更多、1048576或更多、2097152或更多、4194304或更多、8388608或更多、16777216或更多、33554432或更多、67108864或更多、134217728或更多、268435456或更多、536870912或更多、1073741824或更多、2147483648或更多、4294967296或更多、8589934592或更多、17179869184或更多、34359738368或更多、68719476736或更多、137438953472或更多、274877906944或更多、549755813888或更多、1099511627776或更多、2199023255552或更多、4398046511104或更多、8796093022208或更多、17592186044416或更多、35184372088832或更多、70368744177664或更多、140737488355328或更多、281474976710656或更多、562949953421312或更多、1125899906842624或更多、2251799813685248或更多、4503599627370496或更多、9007199254740992或更多、18014398509481984或更多、36028797018963968或更多、72057594037927936或更多、144115188075855872或更多、288230376151711744或更多、576460752303423488或更多、1152921504606846976或更多、2305843009213693952或更多、4611686018427387904或更多、9223372036854775808或更多、18446744073709551616或更多、36893488147419103232或更多、73786976294838206464或更多、147573952589676412928或更多、295147905179352825856或更多、590295810358705651712或更多、1180591620717411303424或更多、2361183241434822606848或更多、4722366482869645213696或更多、9444732965739290427392或更多、18889465931478580854784或更多、37778931862957161709568或更多、75557863725914323419136或更多、151115727451828646838272或更多、302231454903657293676544或更多、604462909807314587353088或更多、1208925819614629174706176或更多、2417851639229258349412352或更多、4835703278458516698824704或更多、9671406556917033397649408或更多、19342813113834066795298816或更多、38685626227668133590597632或更多、77371252455336267181195264或更多、154742504910672534362390528或更多、309485009821345068724781056或更多、618970019642690137449562112或更多、1237940039285380274899124224或更多、2475880078570760549798248448或更多、4951760157141521099596496896或更多、9903520314283042199192993792或更多、19807040628566084398385987584或更多、39614081257132168796771975168或更多、79228162514264337593543950336或更多、158456325028528675187087900672或更多、316912650057057350374175801344或更多、633825300114114700748351602688或更多、1267650600228229401496703205376或更多、2535301200456458802993406410752或更多、5070602400912917605986812821504或更多、10141204801825835211973625643008或更多、20282409603651670423947251286016或更多、40564819207303340847894502572032或更多、81129638414606681695789005144064或更多、162259276829213363391578010288128或更多、324518553658426726783156020576256或更多、649037107316853453566312041152512或更多、1298074214633706907132624082305024或更多、2596148429267413814265248164610048或更多、5192296858534827628530496329220096或更多、10384593717069655257060992658440192或更多、20769187434139310514121985316880384或更多、41538374868278621028243970633760768或更多、83076749736557242056487941267521536或更多、166153499473114484112975882535043072或更多、332306998946228968225951765070086144或更多、664613997892457936451903530140172288或更多、1329227995784915872903807060280344576或更多、2658455991569831745807614120560689152或更多、5316911983139663491615228241121378304或更多、10633823966279326983230456482242756608或更多、21267647932558653966460912964485513216或更多、42535295865117307932921825928971026432或更多、85070591730234615865843651857942052864或更多、170141183460469231731687303715884105728或更多、340282366920938463463374607431768211456或更多、680564733841876926926749214863536422912或更多、1361129467683753853853498429727072845824或更多、2722258935367507707706996859454145691648或更多、5444517870735015415413993718908291383296或更多、10889035741470030830827987437816582766592或更多、21778071482940061661655974875633165533184或更多、43556142965880123323311949751266331066368或更多、87112285931760246646623899502532662132736或更多、17422457186352049329324779900506524465536或更多、34844914372704098658649559801013048931072或更多、69689828745408197317299119602026097862144或更多、139379657490816394634598239204052195724288或更多、278759314981632789269196478408104391448576或更多、557518629963265578538392956816208782897152或更多、111503725992653115707678591363241756579424或更多、223007451985306231415357182726483513158848或更多、446014903970612462830714365452967026317696或更多、892029807941224925661428730905934052635392或更多、1784059615882449851322857461811868105270784或更多、3568119231764899702645714923623736210541568或更多、7136238463529799405291429847247472421083136或更多、14272476927059598810582859694494944842166272或更多、28544953854119197621165719388989889684332544或更多、57089907708238395242331438777979779368665088或更多、114179815416476790484662877555959558737330176或更多、228359630832953580969325755111919117474660352或更多、456719261665907161938651510223838234949320704或更多、913438523331814323877303020447676469898641408或更多、1826877046663628647754606040895352939797282816或更多、3653754093327257295509212081790705879594565632或更多、7307508186654514591018424163581411759189131264或更多、14615016373309029182036848327162823518378262528或更多、29230032746618058364073696654325647036756525056或更多、58460065493236116728147393308651294073513050112或更多、116920130986472233456294786617302588147026100224或更多、233840261972944466912589573234605176294052200448或更多、467680523945888933825179146469210352588104400896或更多、935361047891777867650358292938420705176208801792或更多、1870722095783555735300716585876841410352417603584或更多、3741444191567111470601433171753682820704835207168或更多、7482888383134222941202866343507365641409670414336或更多、14965776766268445882405732687014731282819340828672或更多、29931553532536891764811465374029462565638681657344或更多、59863107065073783529622930748058925131277363314688或更多、119726214130147567059245861496117850262554726629376或更多、239452428260295134118491722992235700525109453258752或更多、478904856520590268236983445984471401050218906517504或更多、957809713041180536473966891968942802100437813035008或更多、1915619426082361072947933783937885604200875626070016或更多、3831238852164722145895867567875771208401751252140032或更多、7662477704329444291791735135751542416803502504280064或更多、15324955408658888583583470271503084833607005008560128或更多、30649910817317777167166940543006169667214010017120256或更多、61299821634635554334333881086012339334428020034240512或更多、122599643269271108668667762172024678668856040068481024或更多、245199286538542217337335524344049357337712080136962048或更多、490398573077084434674671048688098714675424160273924096或更多、980797146154168869349342097376197429350848320547848192或更多、196159429230833773869868419475239485870169664109569384或更多、392318858461667547739736838950478971740339328219138768或更多、784637716923335095479473677900957943480678656438277536或更多、1569275433846670190958947355801915886961357312876555072或更多、3138550867693340381917894711603831773922714625753110144或更多、6277101735386680763835789423207663547845429251506220288或更多、12554203470773361527671578846415327095690858503012440576或更多、25108406941546723055343157692830654191381717006024881152或更多、50216813883093446110686315385661308382763434012049762304或更多、100433627766186892221372630771322616765526868024099524608或更多、200867255532373784442745261542645233531053736048199049216或更多、401734511064747568885490523085290467062107472096398098432或更多、803469022129495137770981046170580934124214944192796196864或更多、1606938044258990275541962092341161868248429888385592393728或更多、3213876088517980551083924184682323736496859776771184787456或更多、6427752177035961102167848369364647472993719553542369574912或更多、12855504354071922204335696738729294945987439107084739149824或更多、25711008708143844408671393477458589891974878214169478299648或更多、51422017416287688817342786954917179839549756428338956599296或更多、102844034832575377634685573909834359679099512856677913198592或更多、205688069665150755269371147819668719358199025713355826397184或更多、411376139330301510538742295639337438716398051426711652794368或更多、822752278660603021077484591278674877432796102853423305588736或更多、1645504557321206042154969182557349754865592205706846611177504或更多、3291009114642412084309938365114699509731184411413693222355008或更多、6582018229284824168619876730229399019462368822827386444710016或更多、13164036458569648337239753460458798038924737645654672889420032或更多、26328072917139296674479506920917596077849475291309345778840064或更多、52656145834278593348959013841835192155698950582618691557680128或更多、105312291668557186697918027683670384311397901165237383115360256或更多、210624583337114373395836055367340768622795802330474766230720512或更多、421249166674228746791672110734681537245591604660949532461441024或更多、842498333348457493583344221469363074491183209321899064922882048或更多、1684996666696914987166688442938726148982366418643798129845764096或更多、3369993333393829974333376885877452297964732837287596259691528192或更多、6739986666787659948666753771754904595929465674575192519383056384或更多、13479973333575319897333507543509809191858931349150385038766112768或更多、26959946667150639794667015087019618383717862698300770077532225536或更多、53919893334301279589334030174039236767435725396601540155064451072或更多、107839786668602559178668060348078473534871450793203080310128902144或更多、215679573337205118357336120696156947069742901586406160620257804288或更多、431359146674410236714672241392313894139485803172812321240515608576或更多、862718293348820473429344482784627788278971606345624642481031217152或更多、1725436586697640946858688965569255576557943212691249284962062434304或更多、3450873173395281893717377931138511153115886425382498569924124868608或更多、6901746346790563787434755862277022306231772850764997139848249737216或更多、13803492693581127574869511724554044612463545701529994279696499474432或更多、27606985387162255149739023449108089224927091403059988559392998948864或更多、55213970774324510299478046898216178449854182806119977118785997897728或更多、110427941548649020598956093796432356899708365612239954237571995795456或更多、220855883097298041197912187592864713799416731224479908475143991590912或更多、441711766194596082395824375185729427598833462448959816950287983181824或更多、883423532389192164791648750371458855197666924897919633900575966363648或更多、1766847064778384329583297500742917710395333849795839267801151932727296或更多、3533694129556768659166595001485835420790667699591678535602303865454592或更多、7067388259113537318333190002971670841581335399183357071204607730909184或更多、14134776518227074636666380005943341683162670798366714142409215461818368或更多、28269553036454149273332760011886683366325341596733428284818430923636736或更多、56539106072908298546665520023773366732650683193466856569636861847273472或更多、113078212145816597093331040047546733465301366386933713139273723694546944或更多、226156424291633194186662080095093466930602732773867426278547447389093888或更多、452312848583266388373324160190186933861205465547734852557094894778187776或更多、904625697166532776746648320380373867722410931095469705114189789556375552或更多、1809251394333065553493296640760747735444821862190939410228379579112751104或更多、3618502788666131106986593281521495470889643724381878820456759158225502208或更多、7237005577332262213973186563042990941779287448763757640913518316451004416或更多、14474011154664524427946373126085981883558574897527515281827036632902008832或更多、28948022309329048855892746252171963767117149795055030563654073265804017664或更多、57896044618658097711785492504343927534234299590110061127308146531608035328或更多、115792089237316195423570985008687855068468599180220122254616293063216070656或更多、231584178474632390847141970017375710136937198360440244509232586126432141312或更多、463168356949264781694283940034751420273874396720880489018465172252864282624或更多、926336713898529563388567880069502840547748793441760978036930344505728565248或更多、1852673427797059126777135760139005681095497586883521956073860689011457130496或更多、3705346855594118253554271520278011362190995173767043912147721378022914260992或更多、7410693711188236507108543040556022724381990347534087824295442756045828521984或更多、14821387422376473014217086081112045448763980695068175648590885512091657043968或更多、29642774844752946028434172162224090897527961390136351297181771024183314087936或更多、59285549689505892056868344324448181795055922780272702594363542048366628175872或更多、118571099379011784113736688648896363590111845560545405188727084096733256351744或更多、237142198758023568227473377297792727180223691121090810377454168193466512703488或更多、474284397516047136454946754595585454360447382242181620754908336386933025406976或更多、948568795

于8%、小于7%、小于6%、小于5%、小于4%、小于3%、小于2%或小于1%。在一些情况下,可能存在少于20个碱基对、少于18个碱基对、少于15个碱基对、少于14个碱基对、少于13个碱基对、少于12个碱基对、少于11个碱基对或少于10个碱基对的同源性。在一些情况下,这样的碱基对是连续的。

[0116] 此外,在一些实施方案中,可以选择一些或所有读出序列以使得它们不表现出彼此之间和/或与怀疑存在于样品中的基因组或其他核酸的特异性结合。例如,可以对一群读出序列进行“比对 (blasted)”或测试其特异性结合或互补性。在一些情况下,读出序列可能不表现出彼此之间的特异性结合,和/或使得读出序列群中没有一个读出序列与读出序列群中的另一个读出序列具有超过5、6、7、8、9、10等个核苷酸的互补性。

[0117] 在一组实施方案中,核酸探针群可以包含一定数量的读出序列,其可以与样品中待测定的核酸靶标的数量相同,例如,每个独特的读出序列对应于一个独特的靶标。在另一组实施方案中,核酸探针群可包含一定数量的读出序列,其可能少于样品中待测定的核酸靶标的数量。本领域普通技术人员将意识到,如果存在一个信号传导实体和n个读出序列,则通常可以唯一鉴定 $2^n - 1$ 个不同的核酸靶标。然而,并非需要使用所有可能的组合。例如,核酸探针群可以靶向12个不同的核酸靶标,但包含不超过8个读出序列。作为另一个例子,核酸探针群可以靶向140个不同的核酸靶标,但包含不超过16个读出序列。不同的核酸靶标可以通过在每个探针内使用不同的读出序列组合来分别鉴定。例如,核酸探针群可包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16等或更多个读出序列。在一些情况下,核酸探针群可以各自包含相同数量的读出序列,尽管在其他情况下,各种探针上可能存在不同数量的读出序列。

[0118] 作为非限制性实例,第一核酸探针可以包含第一靶序列、第一读出序列和第二读出序列,而第二不同的核酸探针可以包含第二靶序列、相同的第一读出序列和第三读出序列而不是第二读出序列。这样的探针因此可以通过确定存在的或与给定探针或位置相关联的各种读出序列来区分,如本文所讨论的。例如,可以使用“代码字”依次鉴定和编码探针,如下所讨论的。任选地,还可以对代码字进行错误检测和/或错误校正。

[0119] 作为另一个非限制性实例,第一核酸探针群可包含第一靶序列、第一读出序列和第二读出序列,而第二不同的核酸探针群可包含第二靶序列、相同的第一读出序列和第三读出序列而不是第二读出序列。这样的探针因此可以通过确定存在的或与给定探针或位置相关联的各种读出序列来区分,如本文所讨论的。例如,可以使用“代码字”依次鉴定和编码探针群,如所讨论的。任选地,还可以对代码字进行错误检测和/或错误校正。

[0120] 此外,在某些实施方案中,核酸探针群可以使用4个天然存在的核苷酸碱基中的仅2个或仅3个来制备,例如在探针群体中省去所有“G”或省去所有“C”。在某些实施方案中,缺少“G”或“C”的序列可以形成非常少的次级结构,并且可以有助于更均匀、更快的杂交。因此,在一些情况下,核酸探针可能仅包含A、T和G;仅包含A、T和C;仅包含A、C和G;或仅包含T、C和G。

[0121] 一方面,核酸探针上的读出序列可能能够结合(例如,特异性地)初级扩增核酸上的相应识别序列。因此,当核酸探针识别生物样品中的靶标例如DNA或RNA靶标时,初级扩增核酸也能够通过核酸探针与靶标缔合,其中在核酸探针的读出序列和初级扩增核酸上的相应的识别序列之间发生相互作用,例如互补结合。例如,识别序列可能能够识别靶标读出序

列,但基本上不能识别或结合其他非靶标读出序列。取决于应用,初级扩增核酸还可包含能够与核酸杂交的多种实体中的任一种,例如DNA、RNA、LNA和/或PNA等。例如,这样的实体可以形成部分或全部识别序列。

[0122] 在一些情况下,识别序列可以与靶标读出序列基本上互补。在一些情况下,序列可以是至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%、至少92%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%互补。通常,互补性是根据Watson-Crick核苷酸碱基配对确定的。靶标读出序列的结构可以包括先前描述的那些。

[0123] 在一些情况下,识别序列的长度可以是至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少50、至少60、至少65、至少75、至少100、至少125、至少150、至少175、至少200、至少250、至少300、至少350、至少400或至少450个核苷酸。在一些情况下,识别序列的长度可能不超过500、不超过450、不超过400、不超过350、不超过300、不超过250、不超过200、不超过175、不超过150、不超过125、不超过100、不超过75、不超过60、不超过65、不超过60、不超过55、不超过50、不超过45、不超过40、不超过35、不超过30、不超过20或不超过10个核苷酸。这些中的任何一种的组合也是可能的,例如,识别序列的长度可以为10至30个核苷酸、20至40个核苷酸、5至50个核苷酸、10至200个核苷酸或25至35个核苷酸、10至300个核苷酸等。

[0124] 在一些实施方案中,初级扩增核酸还可以包含一个或多个能够结合次级扩增核酸的读出序列,如下文所讨论的。例如,初级扩增核酸可包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16或更多、20或更多、32或更多、40或更多、50或更多、64或更多、75或更多、100或更多、128或更多个读出序列。读出序列可以位于初级扩增核酸内的任何位置。如果存在多于一个读出序列,则读出序列可以彼此相邻定位,和/或散布有其他序列。在一个实施方案中,初级扩增核酸在第一端包含识别序列和在第二端包含多个读出序列。

[0125] 在一些情况下,初级扩增核酸内的读出序列的长度可以是至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少50、至少60、至少65、至少75、至少100、至少125、至少150、至少175、至少200、至少250、至少300、至少350、至少400或至少450个核苷酸。在一些情况下,读出序列的长度可以不超过500、不超过450、不超过400、不超过350、不超过300、不超过250、不超过200、不超过175、不超过150、不超过125、不超过100、不超过75、不超过60、不超过65、不超过60、不超过55、不超过50、不超过45、不超过40、不超过35、不超过30、不超过20或不超过10个核苷酸。这些中的任何一种的组合也是可能的,例如,读出序列可以具有10至20个核苷酸、10至30个核苷酸、20至40个核苷酸、5至50个核苷酸、10至200个核苷酸、或25至35个核苷酸、10至300个核苷酸等的长度。

[0126] 在初级扩增核酸内可以有任意数量的读出序列。例如,可以有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个读出序列存在于初级扩增核酸中。如果初级扩增核酸内存在多于一个读取序列,则读取序列可以相同或不同。例如,在一些情况下,读出序列可以都相同。

[0127] 在一些实施方案中,初级扩增核酸群可以使用4个天然存在的核苷酸碱基中的仅2个或仅3个来制备,例如在核酸群中省略所有“G”或省略所有“C”。在某些实施方案中,缺少“G”或“C”的序列可以形成非常少的次级结构,并且可以有助于更均匀、更快的杂交。因此,

在一些情况下,初级扩增核酸可能仅包含A、T和G;仅包含A、T和C;仅包含A、C和G;或仅包含T、C和G。

[0128] 在一些情况下,可以将多于一种类型的初级扩增核酸应用于样品,例如,依次或同时地。例如,可以有至少2个、至少5个、至少10个、至少25个、至少50个、至少75个、至少100个、至少300个、至少1,000个、至少3,000个、至少10,000个、或至少30,000个可区分的被应用于样品的初级扩增核酸。在一些情况下,可以依次添加初级扩增核酸。然而,在一些情况下,可以同时添加多于一个初级扩增核酸。

[0129] 在一组实施方案中,初级扩增核酸上的读出序列可能能够结合(例如,特异性地)次级扩增核酸上的相应识别序列。因此,当核酸探针识别生物样品中的靶标,例如DNA或RNA靶标时,次级扩增核酸也能够通过初级扩增核酸与靶标缔合,其中初级扩增核酸的读出序列和次级扩增核酸上的相应识别序列之间发生相互作用,例如互补结合。例如,次级扩增核酸上的识别序列可能能够识别初级扩增核酸上的读出序列,但基本上不能识别或结合其他非靶标读出序列。取决于应用,次级扩增核酸还可包含能够与核酸杂交的多种实体中的任一种,例如DNA、RNA、LNA和/或PNA等。例如,这样的实体可以形成部分或全部识别序列。

[0130] 在一些情况下,次级扩增核酸上的识别序列可以与初级扩增核酸上的读出序列基本上互补。在一些情况下,序列可以是至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%、至少92%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或100%互补的。

[0131] 在一些情况下,次级扩增核酸上的识别序列的长度可以是至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少50、至少60、至少65、至少75、至少100、至少125、至少150、至少175、至少200、至少250、至少300、至少350、至少400或至少450个核苷酸。在一些情况下,识别序列的长度可能不超过500、不超过450、不超过400、不超过350、不超过300、不超过250、不超过200、不超过175、不超过150、不超过125、不超过100、不超过75、不超过60、不超过65、不超过60、不超过55、不超过50、不超过45不超过40、不超过35、不超过30、不超过20或不超过10个核苷酸。这些中的任何一种的组合也是可能的,例如,识别序列的长度可以为10至30个核苷酸、20至40个核苷酸、5至50个核苷酸、10至200个核苷酸或25至35个核苷酸、10至300个核苷酸等。

[0132] 在一些实施方案中,次级扩增核酸可包含信号传导实体,和/或可包含一个或多个能够结合信号传导实体的读出序列,如本文所讨论的。例如,次级扩增核酸可包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16或更多、20或更多、32或更多、40或更多、50或更多、64或更多、75或更多、100或更多、128或更多能够结合信号传导实体的读出序列。读出序列可以位于次级扩增核酸内的任何位置。如果存在多于一个读出序列,则读出序列可以彼此相邻定位,和/或散布有其他序列。在一个实施方案中,次级扩增核酸在第一端包含识别序列,在第二端包含多个读出序列。该结构也可以与初级扩增核酸的结构相同或不同。

[0133] 在一些情况下,次级扩增核酸内的读出序列的长度可以是至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少50、至少60、至少65、至少75、至少100、至少125、至少150、至少175、至少200、至少250、至少300、至少350、至少400或至少450个核苷酸。在一些情况下,读出序列的长度可能不超过500、不超过450、不超过400、不超过350、不超过300、不超过250、不超过200、不超过175、不超过150、不超过125、不超过100、不超过75、不超

过60、不超过65、不超过60、不超过55、不超过50、不超过45、不超过40、不超过35、不超过30、不超过20或不超过10个核苷酸。这些中的任何一种的组合也是可能的,例如,次级扩增核酸内的读出序列可以具有10至20个核苷酸、10至30个核苷酸、20至40个核苷酸、5至50个核苷酸、10至200个核苷酸或25至35个核苷酸、10至300个核苷酸等的长度。

[0134] 在次级扩增核酸内可以有任意数量的读出序列。例如,可以有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多个读出序列存在于次级扩增核酸内。如果次级扩增核酸内存在多于一个读出序列,则读出序列可以相同或不同。例如,在一些情况下,读出序列可以都相同。此外,初级和次级扩增核酸中可以独立地存在相同或不同数量的读出序列。

[0135] 次级扩增核酸群可以使用4个天然存在的核苷酸碱基中的仅2个或仅3个来制备,在某些实施方案中,例如在核酸群内省略所有“G”或省略所有“C”。在某些实施方案中,缺少“G”或“C”的序列可以形成非常少的次级结构,并且可以有助于更均匀、更快的杂交。因此,在一些情况下,次级扩增核酸可能仅包含A、T和G;仅包含A、T和C;仅包含A、C和G;或仅包含T、C和G。

[0136] 在一些情况下,可以将多于一种类型的次级扩增核酸应用于样品,例如,依次或同时地。例如,可以有至少2、至少5、至少10、至少25、至少50、至少75、至少100、至少300、至少1,000、至少3,000、至少10,000或至少30,000个可区分的应用于样品的次级扩增核酸。在一些情况下,可以依次添加次级扩增核酸。然而,在一些情况下,可以同时添加多于一个次级扩增核酸。

[0137] 此外,在某些实施方案中,该模式可以替代地在信号传导实体之前重复,例如使用三级扩增核酸、四级核酸等,类似于以上所讨论的。因此,信号传导实体可以与末端扩增核酸结合。因此,作为非限制性实例,编码核酸探针可以与靶标结合,初级扩增核酸结合该编码核酸探针,次级扩增核酸结合该初级扩增核酸,三级扩增核酸结合该次级扩增核酸,信号传导实体结合该三级扩增核酸,或编码核酸探针可以与靶标结合,初级扩增核酸结合该编码核酸探针,次级扩增核酸结合该初级扩增核酸,三级扩增核酸结合该次级扩增核酸,四级扩增核酸结合该三级扩增核酸,信号传导实体结合该四级扩增核酸,等。因此,在所有实施方案中,末端扩增核酸不必一定是次级扩增核酸。

[0138] 这种系统的非限制性示例在图5中示出。图5A-5E显示了可饱和系统的创建。图5A显示了编码核酸探针的一个例子,其中编码核酸探针15已与靶RNA结合。图5B显示了根据某些实施方案正在使用的初级扩增核酸。图5C显示了可以与初级扩增核酸结合的次级扩增核酸。图5D显示多个信号传导实体已与次级扩增核酸的读出序列结合。图5E显示如果不应用扩增,核酸探针可暴露于含有信号传导实体的合适的次级核酸探针。

[0139] 在一些情况下,其他成分也可能存在于核酸探针或扩增核酸中。例如,在一组实施方案中,可以存在一个或多个引物序列,例如以促进酶促扩增。本领域普通技术人员将了解适用于扩增(例如,使用PCR或其他合适的技术)等应用的引物序列。许多这样的引物序列是可商购获得的。可存在于初级或编码核酸探针内的序列的其他实例包括但不限于启动子序列、操纵子、鉴定序列、无义序列等。

[0140] 通常,引物是用作核酸合成起始点的单链或部分双链核酸(例如,DNA),允许聚合酶(例如核酸聚合酶)延伸引物并复制互补链。引物是(例如,设计成)与核酸靶标互补并杂

交的。在一些实施方案中，引物是合成引物。在一些实施方案中，引物是非天然存在的引物。引物通常具有10至50个核苷酸的长度。例如，引物可具有10至40、10至30、10至20、25至50、15至40、15至30、20至50、20至40或20至30个核苷酸的长度。在一些实施方案中，引物具有18至24个核苷酸的长度。

[0141] 在一些方面，如前所述，某些实施方案使用编码各种结合事件的代码空间，并且任选地可以使用错误检测和/或校正来确定核酸探针与其靶标的结合。在一些情况下，核酸探针群可能包含某些“读出序列”，这些“读出序列”能够结合某些扩增核酸，如上文所讨论的，并且核酸探针或靶标的位置可以使用与扩增核酸缔合的信号传导实体在样品内确定，例如，在某个代码空间内，例如，如本文所讨论的。还参见国际专利申请公开号W02016/018960和W02016/018963，各自通过引用整体并入本文。在一些情况下，核酸探针内的读出序列群可以以各种组合进行组合，例如，使得相对少量的读出序列可以用于确定相对大量的不同核酸探针，如本文所讨论的。

[0142] 因此，在一些情况下，核酸探针群可以各自包含一定数量的读出序列，其中一些在不同的核酸探针之间共有，使得核酸探针的总群体可以含有一定数量的读出序列。核酸探针群可以具有任何合适数量的读出序列。例如，核酸探针群可具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20等个读取序列。在一些实施方案中，超过20个也是可能的。此外，在一些情况下，核酸探针群可以总共具有1或更多、2或更多、3或更多、4或更多、5或更多、6或更多、7或更多、8或更多、9或更多、10或更多、11或更多、12或更多、13或更多、14或更多、15或更多、16或更多、20或更多、24或更多、32或更多、40或更多、50或更多、60或更多、64或更多、100或更多、128或更多等个可能的读出序列，尽管一些或所有探针可各自包含多于一个读出序列，如本文所讨论。此外，在一些实施方案中，核酸探针群可具有不超过100、不超过80、不超过64、不超过60、不超过50、不超过40、不超过32、不超过24、不超过20、不超过16、不超过15、不超过14、不超过13、不超过12、不超过11、不超过10、不超过9、不超过8、不超过7、不超过6、不超过5、不超过4、不超过3或不超过两个读出序列存在。这些中的任何一种的组合也是可能的，例如，核酸探针群可以总共包含10至15个读出序列。

[0143] 作为从包含在核酸探针内的相对少量的读出序列组合鉴定相对大量的核酸探针的方法的非限制性示例，在6种不同类型的核酸探针或核酸探针的6个不同组（例如，每组探针与核酸靶标结合）的群体中，每种类型或每组的核酸探针包含一个或多个读出序列，群体内的读出序列的总数可以不大于4。应当理解，虽然为了便于说明，本示例中使用了4个读出序列，但在其他实施方案中，可以实现更大数量的核酸探针，例如使用5、8、10、16、32等或更多个读出序列，或本文描述的任何其他合适数量的读出序列，取决于应用。例如，如果每个核酸探针或每组核酸探针包含两个不同的读出序列，则通过使用4个这样的读取序列（A、B、C和D），最多6个探针或6组探针可以被单独鉴定。应当注意，在该示例中，核酸探针或核酸探针组上的读出序列的排序不是必需的，即“AB”和“BA”可以被视为同义词（尽管在其他实施方案中，读取序列的排序可能是必不可少的，“AB”和“BA”可能不一定是同义词）。类似地，如果在核酸探针群中使用5个读出序列（A、B、C、D和E），则可以单独鉴定多达10个探针或10组探针（例如AB、AC、AD、AE、BC、BD、BE、CD、CE、DE）。例如，本领域普通技术人员将理解，对于在每个探针或组别的每组上具有n个读出序列的群体中的k个读出序列，可以产生多达 $\binom{n}{k}$ 个

不同的探针,假设读出序列的排序不是必需的;因为不是所有的探针或所有的探针组都需要具有相同数量的读出序列并且不是读出序列的所有组合都需要在每个实施方案中使用,所以还可以在有些实施方案中使用多于或少于这个数量的不同探针。此外,还应该理解,在一些实施方案中,每个探针或每组探针上的读出序列的数量不必相同。例如,一些探针或一些探针组可能包含2个读取序列,而其他探针或其他探针组可能包含3个读取序列。在一些实施方案中,每组探针结合核酸靶标。

[0144] 在一些方面,样品中核酸探针的读出序列和/或结合模式可用于定义错误检测和/或错误校正代码,例如,以减少或防止核酸的错误鉴定或错误。因此,例如,如果指示了结合(例如,使用信号传导实体确定),则可以用“1”标识该位置;相反,如果未指示结合,则可以用“0”标识该位置(在一些情况下,反之亦然)。多个轮次的结合确定(例如,使用与读出序列互补的不同读出探针)然后可以用于创建“代码字”,例如,用于该空间位置。在一些实施方案中,可以对代码字进行错误检测和/或校正。例如,可以组织代码字,使得如果对于给定的一组读出序列或核酸探针的结合模式没有发现匹配,则可以将匹配鉴定为错误,并且任选地,可以对序列应用错误校正以确定核酸探针的正确靶标。在一些情况下,代码字可能具有少于由代码字编码的核酸的总数的“字母”或位置,例如在每个代码字编码不同的核酸时。

[0145] 这种错误检测和/或错误校正代码可以采取多种形式。之前已经在诸如电信行业的其他环境中开发了多种这样的代码,例如Golay码或Hamming码。在一组实施方案中,分配核酸探针的读出序列或结合模式,使得并非分配每个可能的组合。

[0146] 例如,如果可能有4个读出序列并且核酸探针或核酸探针组包含2个读出序列,则最多6个核酸探针或6组核酸探针(例如,使得每组核酸探针与核酸靶标结合)可以被鉴定;但使用的核酸探针的数量或核酸探针组的数量可以少于6个。类似地,对于在每个核酸探针或每组核酸探针上具有 n 个读出序列的群体中的 k 个读出序列,可以产生 $\binom{n}{k}$ 个不同探针或不同探针组,但是所使用的核酸探针的数量或核酸探针组的数量可以是大于或小于 $\binom{n}{k}$ 的任何数量。此外,这些可以随机分配,或以特定方式分配以增加检测和/或校正错误的能力。

[0147] 作为另一个例子,如果使用多个轮次的核酸探针(例如,使得能够结合初级或编码探针上的读出序列的多个轮次的读出探针),可以任意选择轮次数。如果在每个轮次中,每个靶标都可以给出两种可能的结果,例如被检测到或未被检测到,那么对于 n 个轮次的探针,最多可以有 2^n 个不同的靶标,但实际使用的靶标数量可能是少于 2^n 的任意数量。在另一个示例中,如果在每个轮次中,每个靶标都可以给出多于两个可能的结果,例如在不同的颜色通道中被检测到,那么对于 n 个轮次的探针,可能有超过 2^n (例如 3^n 、 4^n 、...)个不同的靶标。在一些情况下,实际使用的靶标数量可能是小于此数量的任何数量。此外,这些可以随机分配,或以特定方式分配以增加检测和/或校正错误的能力。

[0148] 代码字可用于定义各种代码空间。每个核酸靶标与一个代码字相关。例如,在一组实施方案中,可以在代码空间内分配代码字,使分配被Hamming距离分开,Hamming距离测量给定模式中导致代码字或相关核酸靶标被误解为不同的有效代码字或核酸靶标的不正确“读取”的数量。在某些情况下,Hamming距离可以是至少2、至少3、至少4、至少5、至少6等。此外,在一组实施方案中,分配可以形成为Hamming码,例如,Hamming(7,4)码、Hamming(15,11)码、Hamming(31,26)码、Hamming(63,57)码、Hamming(127,120)码等。在另一组实施方案

中,分配可以形成SECDED码,例如SECDED (8,4) 码、SECDED (16,4) 码、SECDED (16,11) 码、SECDED (22,16) 码、SECDED (39,32) 码、SECDED (72,64) 码等。在又一组实施方案中,分配可以形成扩展的二进制Golay码、完美的二进制Golay码或三元Golay码。在另一组实施方案中,分配可以代表取自上述任何代码的可能值的子集。

[0149] 例如,可以通过仅使用包含固定或恒定数量的“1”位(或“0”位)的二进制字来对靶标进行编码来形成错误校正代码。例如,代码空间可能仅包括1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16等个“1”位(或“0”位),例如,所有代码具有相同数量的“1”位或“0”位等。在另一组实施方案中,分配可以代表取自上述代码的可能值的子集,用于解决不对称读出错误的目的。例如,在一些情况下,在其中对于所有使用的二进制字可以固定“1”位的数量的代码可以消除在“0”位被测量为“1”或“1”位被测量为“0”的速率不同时具有不同数量的“1”的字的偏倚测量。

[0150] 因此,在一些实施方案中,一旦确定了代码字(例如,如本文所讨论的),就可以将代码字与有效的核酸代码字进行比较。如果发现匹配,则可以鉴定或确定核酸靶标。如果未找到匹配,则可以鉴定代码字读取中的错误。在一些情况下,还可以应用错误校正来确定正确的代码字,从而导致核酸靶标的正确身份。在一些情况下,可以选择代码字,使得假设仅存在一个错误,则只有一个可能的正确代码字可用,因此,核酸靶标的唯一正确身份是可能的。在一些情况下,这也可以推广到更大的代码字间距或Hamming距离;例如,可以选择代码字,使得如果存在两个、三个或四个错误(或者在某些情况下更多),则只有一个可能的正确代码字可用,因此,核酸靶标的唯一正确身份是可能的。

[0151] 错误校正代码可以是二进制错误校正代码,或者它可以基于其他编号系统,例如三元或四元错误校正代码。例如,在一组实施方案中,可以使用多于一种类型的信号传导实体并将其分配给错误校正代码内的不同数字。因此,作为非限制性示例,可以将第一信号传导实体(或在某些情况下多于一个信号传导实体)分配为“1”,并且可以将第二信号传导实体(或在某些情况下多于一个信号传导实体)分配为“2”(“0”表示不存在信号传导实体),并且分配代码字以定义三元错误校正代码。类似地,第三信号传导实体可以另外被分配为“3”以产生四进制错误校正代码等。

[0152] 在一组实施方案中,样品中的核酸靶标各自被分配有代码字。例如,这些代码字可以从本文描述的代码空间之一中选择。在一些情况下,代码字形成错误检测和/或错误校正代码。在一些情况下,样品可以与初级或编码核酸探针群进行杂交。初级或编码探针中的一些或全部可包含可与核酸靶标之一结合的靶序列和/或还可包含一个或多个读出序列。与每个核酸靶标结合的初级或编码探针的集合上的读出序列可以形成与分配给核酸靶标的代码字相对应的唯一代码字。然后样品与读出探针进行一个或多个轮次的杂交。读出探针可能能够结合读出序列和/或可能与信号传导实体缔合。读出序列的集合可以与核酸靶标相关联,因此分配给核酸靶标的代码字然后可以被鉴定,例如通过读出探针的结合。

[0153] 在一些情况下,可以在每个轮次中使用多色成像,以允许同时成像和确定与不同信号传导实体缔合的多个读出探针。在一些情况下,确定核酸靶标的位置。在一些情况下,以这种方式确定至少50、至少100、至少500、至少1000、至少5000或至少10,000个核酸靶标。在一些情况下,核酸靶标是基因组基因座。在一些情况下,核酸靶标是基因组基因座和/或新生RNA转录物。在一些情况下,基因组基因座的位置用于确定细胞中染色质的三维组织或

基因组的三维组织。在一些情况下,初级扩增核酸和/或次级扩增核酸和/或三级扩增核酸和/或四级扩增核酸用于扩增来自每个读出序列的信号。在一些情况下,如下文所述使用衔接子。

[0154] 一方面,可以使用多个衔接子来促进样品内靶标的检测。这样的衔接子可以用于例如允许使用相对少量的可区分信号传导实体,同时仍然允许在样品中测定相对大量的靶标。例如,可以测定样品中的至少3、至少4、至少7、至少8、至少12、至少14、至少15、至少16、至少20、至少22、至少30、至少31、至少32、至少50、至少63、至少64、至少72、至少75、至少100、至少127、至少128、至少140、至少255、至少256、至少500、至少1,000、至少1,500、至少2,000、至少2,500、至少3,000、至少4,000、至少5,000、至少7,500、至少10,000、至少12,000、至少15,000、至少20,000、至少25,000、至少30,000、至少40,000、至少50,000、至少75,000或至少100,000个等靶标,同时使用较少数量的信号传导实体,例如,不超过20个、不超过15个、不超过10个、不超过5个、不超过4个、不超过3个、或不超过2个信号传导实体。

[0155] 在一组实施方案中,可以使用多个衔接子。衔接子可包含与核酸探针(例如初级核酸探针)上的一个或多个读出序列基本上互补的第一部分,和包含一个或多个鉴定序列的第二部分。因此,衔接子序列能够结合特定的核酸探针,该探针能够结合样品中的靶标。然后鉴定序列可用于结合,例如通过读出探针或次级核酸探针,例如本文讨论的那些。因此,在一些情况下,衔接子可以存在于初级核酸探针和次级核酸探针之间。图24A中示出了这种情况的一个非限制性示例。

[0156] 在一些情况下,可以选择衔接子以允许使用相对少量的信号传导实体,如上所述。例如,鉴定序列可以充当次级核酸探针能够结合的读出序列。在一轮检测中,可以使用相对少量的次级核酸探针,例如,包含信号传导实体和与鉴定序列之一基本上互补的序列,并测定信号传导实体,例如,如本文所讨论的。然后可以在下一轮检测之前去除和/或灭活次级核酸探针,例如,如本文所述。下一个和随后的轮次可以使用相同或不同的信号传导实体,例如在包含与不同鉴定序列基本上互补的序列的次级核酸探针上。

[0157] 此外,在一些实施方案中,为了减少污染或“串扰”,可以以某种方式灭活前一轮次中使用的衔接子。例如,可以添加包含与先前的鉴定序列基本上互补的序列的封闭核酸探针,使得它们能够结合先前的衔接子,但是因为它们通常在没有信号传导实体存在的情况下是不可检测的。因此,在随后的检测轮次中,由于先前轮次的信号可以被最小化。

[0158] 因此,在一些情况下,可以使用不多于相对少量的信号传导实体来确定相对大量的鉴定序列。例如,可以使用不超过20个、不超过15个、不超过10个、不超过5个、不超过4个、不超过3个或不超过2个信号传导实体来确定至少3、至少4、至少7、至少8、至少12、至少14、至少15、至少16、至少20、至少22、至少30、至少31、至少32、至少50、至少63、至少64、至少72、至少75、至少100、至少127、至少128、至少140、至少255、至少256、至少500、至少1,000、至少1,500、至少2,000、至少2,500、至少3,000、至少4,000、至少5,000、至少7,500、至少10,000、至少12,000、至少15,000、至少20,000、至少25,000、至少30,000、至少40,000、至少50,000、至少75,000或至少100,000个等鉴定序列。

[0159] 鉴定序列可以是任何长度。如果使用多于一个鉴定序列,则鉴定序列可以独立地具有相同或不同的长度。例如,鉴定序列的长度可以是至少5、至少10、至少15、至少20、至少25、至少30、至少35、至少40、至少50、至少60、至少65、至少75、至少100、至少125、至少150、

至少175、至少200、至少250、至少300、至少350、至少400或至少450个核苷酸。在一些情况下,鉴定序列的长度可能不超过500、不超过450、不超过400、不超过350、不超过300、不超过250、不超过200、不超过175、不超过150、不超过125、不超过100、不超过75、不超过60、不超过65、不超过60、不超过55、不超过50、不超过45、不超过40、不超过35、不超过30、不超过20或不超过10个核苷酸。这些中的任何一种的组合也是可能的,例如,鉴定序列可以具有10至30个核苷酸、20至40个核苷酸、5至50个核苷酸、10至200个核苷酸或25至35个核苷酸、10至300个核苷酸等的长度。

[0160] 在一些实施方案中,鉴定序列可以是任意的或随机的。在某些情况下,选择鉴定序列以减少或最小化与细胞或其他样品的其他成分的同源性,例如,使得鉴定序列本身不与怀疑存在于细胞或其他样品内的其他核酸结合或杂交。在一些情况下,同源性可以小于10%、小于8%、小于7%、小于6%、小于5%、小于4%、小于3%、小于2%或小于1%。在一些情况下,可能存在少于20个碱基对、少于18个碱基对、少于15个碱基对、少于14个碱基对、少于13个碱基对、少于12个碱基对、少于11个碱基对或少于10个碱基对的同源性。在一些情况下,这样的碱基对是连续的。

[0161] 此外,在一些实施方案中,可以选择一些或所有的鉴定序列,使得它们不表现出针对彼此和/或针对样品中怀疑存在的基因组或其他核酸例如读出序列的特异性结合。例如,可以对鉴定序列群进行“比对”或测试其特异性结合或互补性。在一些情况下,鉴定序列可能不表现出针对彼此的特异性结合,和/或使得鉴定序列群中没有鉴定序列具有与鉴定序列群内和/或读出序列群内的另一个序列的超过5、6、7、8、9、10等个核苷酸的互补性。

[0162] 在一些实施方案中,首先使样品与初级或编码核酸探针群杂交。初级或编码探针中的一个或多个包含可结合核酸靶标之一的靶序列,并且还可包含一个或多个读出序列。然后将样品与衔接子探针和读出探针进行多个轮次的杂交。衔接子探针可以包含能够结合读出序列的序列并且还包含一个或多个鉴定序列。读出探针可能能够结合鉴定序列并且还与信号传导实体缔合。在一些情况下,可以在每个轮次中使用多色成像,以允许同时成像和确定与不同信号传导实体缔合的多个读出探针。

[0163] 如本文所讨论的,在某些方面,例如通过成像测定信号传导实体以确定核酸探针和/或产生代码字。信号传导实体的示例包括这里讨论的那些。在一些情况下,可以使用多种技术例如在空间上确定样品内的信号传导实体。在一些实施方案中,信号传导实体可以是荧光的,并且用于测定样品内荧光的技术例如荧光显微术或共焦显微术可用于在空间上鉴定信号传导实体在细胞内的位置。在一些情况下,样品中实体的位置可以在两个甚至三个维度上确定。此外,在一些实施方案中,可以一次和/或依次确定超过一个信号传导实体(例如,具有不同颜色或发射的信号传导实体)。

[0164] 此外,在一些实施方案中,可以确定鉴定的靶标例如核酸靶标的置信水平。例如,可以使用精确匹配的数量与具有一个或多个一位错误的匹配的数量的比率来确定置信水平。在一些情况下,可以仅使用具有大于某个值的置信比的匹配。例如,在某些实施方案中,只有当匹配的置信比大于约0.01、大于约0.03、大于约0.05、大于约0.1、大于约0.3、大于约0.5、大于约1、大于约3、大于约5、大于约10、大于约30、大于约50、大于约100、大于约300、大于约500、大于约1000或任何其他合适的值时,才可接受匹配。此外,在一些实施方案中,只有当所鉴定的靶标的置信比大于内标或假阳性对照约0.01、约0.03、约0.05、约0.1、约0.3、

约0.5、约1、约3、约5、约10、约30、约50、约100、约300、约500、约1000或任何其他合适的值时,才可接受匹配。

[0165] 在一些实施方案中,可以以相对较高的分辨率确定实体(以及因此可能与实体结合的核酸探针)的空间位置。例如,可以以优于约100微米、优于约30微米、优于约10微米、优于约3微米、优于约1微米、优于约800nm、优于约600nm、优于约500nm、优于约400nm、优于约300nm、优于约200nm、优于约100nm、优于约90nm、优于约80nm、优于约70nm、优于约60nm、优于约50nm、优于约40nm、优于约30nm、优于约20nm或优于约10nm等的空间分辨率确定位置。

[0166] 有多种技术能够在光学上例如使用荧光显微术确定或成像实体的空间位置。在一些实施方案中可以使用多于一种颜色。在一些情况下,可以以超分辨率或比光波长或衍射极限更好的分辨率确定空间位置。非限制性示例包括STORM(随机光学重建显微术)、STED(受激发射耗尽显微术)、NSOM(近场扫描光学显微术)、4Pi显微术、SIM(结构化照明显微术)、SMI(空间调制照明)显微术、RESOLFT(可逆饱和光学线性荧光跃迁显微术)、GSD(基态耗尽显微术)、SSIM(饱和结构照明显微术)、SPDM(光谱精密距离显微术)、光激活定位显微术(PALM)、荧光光激活定位显微术(FPALM)、LIMON(3D光显微纳米尺寸显微术)、超分辨率光学波动成像(SOFI)、膨胀显微术等。参见,例如,Zhuang等人于2010年11月23日发布的题为“Sub-Diffraction Limit Image Resolution and Other Imaging Techniques”的美国专利号7,838,302;Zhuang等人于2013年10月22日发布的题为“Sub-diffraction Limit Image Resolution in Three Dimensions”的美国专利号8,564,792,;或Zhuang等人于2013年6月20日公开的题为“High Resolution Dual-Objective Microscopy”的国际专利申请公开号W02013/090360,其各自通过引用整体并入本文。

[0167] 作为说明性的非限制性示例,在一组实施方案中,可以使用高数值孔径、具有100X放大率的油浸物镜和在电子倍增CCD相机上收集的光对样品进行成像。在另一个示例中,可以使用高数值孔径、具有40X放大率的油浸透镜和使用广域科学CMOS相机收集的光对样品进行成像。在各种非限制性实施方案中,使用物镜和相机的不同组合,单个视野可能对应于不少于1x1微米、10x10微米、40x40微米、80x80微米、120x120微米、240x240微米、340x340微米或500x500微米等。类似地,在一些实施方案中,单个相机像素可以对应于不小于10x10nm、20x20nm、40x40nm、80x80nm、120x120nm、160x160nm、240x240nm或300x300nm等的样品区域。在另一个示例中,可以使用低数值孔径、具有10X放大率的空气透镜和使用sCMOS相机收集的光对样品进行成像。在另外的实施方案中,样品可以通过经由扫描镜或旋转圆盘产生的单个或多个扫描衍射限制焦点照明而进行光学切片,并且使收集的样品通过单个或多个针孔。在另一个实施方案中,样品还可以通过经由本领域技术人员已知的多种方法中的任何一种产生的薄光片来照明。

[0168] 在一个实施方案中,样品可以通过单高斯模式激光线照明。在一些实施方案中,可以通过使这些激光线穿过通过压电或其他机械装置振动的多模光纤来使照明轮廓变平。在一些实施方案中,可以通过使单模高斯光束穿过各种折射光束整形器例如piShaper或一系列堆叠的Powell透镜来使照明轮廓变平。在又一组实施方案中,高斯光束可以穿过各种不同的漫射元件,例如毛玻璃或工程漫射器,在一些情况下,这些漫射元件可以高速旋转以去除残留的激光斑点。在又一个实施方案中,激光照明可以通过一系列小透镜阵列以产生接

近平坦照明场的照明的重叠图像。

[0169] 在一些实施方案中,可以确定实体的空间位置的质心。例如,可以使用本领域普通技术人员已知的图像分析算法在图像或图像系列内确定信号传导实体的质心。在一些情况下,可以选择算法以确定样品中的非重叠单个发射器和/或部分重叠单个发射器。合适技术的非限制性示例包括最大似然算法、最小二乘算法、贝叶斯算法、压缩感知算法等。在某些情况下也可以使用这些技术的组合。

[0170] 在一些实施方案中,可以测定一个或多个信号传导实体。例如,信号传导实体可以与次级扩增核酸(或其他末端扩增核酸)上的读出探针或识别实体结合。信号传导实体的非限制性实例包括荧光实体(荧光团)或磷光实体,例如,如本文所讨论的。然后可以测定信号传导实体,例如以测定核酸探针或靶标。在一些情况下,该测定可以是空间的,例如,在两个或三个维度中。此外,在一些情况下,测定可以是定量的,例如,可以测定信号传导实体和/或靶标的量或浓度。

[0171] 在一组实施方案中,信号传导实体可以附接到次级扩增核酸(或其他末端扩增核酸)。信号传导实体可以在次级扩增核酸与样品内的靶标缔合之前或之后附接至次级扩增核酸(或其他末端扩增核酸)。例如,信号传导实体可以最初附接至次级扩增核酸,或在次级扩增核酸已应用于样品之后附接至次级扩增核酸。在一些情况下,添加信号传导实体,然后进行反应以将它们附接到扩增核酸。

[0172] 在一组实施方案中,信号传导实体可以通过可以被裂解以释放信号传导实体的键附接至核苷酸序列。例如,在确定样品内核酸探针的分布之后,可以在另一轮核酸探针和/或扩增核酸之前释放或灭活信号传导实体。因此,在一些实施方案中,该键可以是可裂解键,例如二硫键或可光裂解键。本文详细讨论了可光裂解键的实例。在一些情况下,这种键可以例如在暴露于还原剂或光(例如,紫外光)时被裂解。有关其他详细信息,参见下文。在一些情况下,信号传导实体通过光漂白而灭活。本文更详细地讨论了用于灭活和/或去除信号传导实体的系统和方法的其他示例。

[0173] 在某些实施方案中,初级和次级扩增核酸的使用可用于产生最大数量的可与给定核酸探针结合的信号传导实体。例如,可能存在最大数量的能够结合核酸探针的信号传导实体,例如,由于最大数量的能够结合有限数量的次级扩增核酸的具有信号传导实体的读出探针,由于最大数量的能够结合有限数量的初级扩增核酸的次级扩增核酸,和/或由于最大数量的能够结合核酸探针上的有限数量的读出序列的初级扩增核酸。虽然每个潜在位置实际上不需要填充信号传导实体,但该结构表明存在信号传导实体的饱和极限,超出该极限可能碰巧存在的任何其他信号传导实体都不能与核酸探针或其靶标缔合。

[0174] 因此,某些实施方案一般涉及扩增指示可饱和的核酸探针或其靶标(即,使得存在有多少信号传导实体可以与核酸探针或其靶标缔合的饱和上限)的信号的系统和方法。通常,该数量大于1。例如,信号传导实体的上限可以是至少2个、至少3个、至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少40个、至少50个、至少75个、至少100个、至少125个、至少150个、至少175个、至少200个、至少250个、至少300个、至少400个、至少500个等。在一些情况下,上限可能小于500、小于400、小于300、小于250、小于200、小于175、小于150、小于125、小于100、小于75、小于50、小于40、小于30、小于25、小于20、小于15、小于10、小于5等。在一些情况下,上限可以被确定为可与次级扩增核酸结合的信号传导实体的最大

数量,乘以可与初级扩增核酸结合的次级扩增核酸的最大数量,乘以可与结合靶标的核酸探针结合的初级扩增核酸的最大数量。相比之下,诸如滚环扩增或发夹展开的技术允许以不受控制的方式扩增信号,即,当存在足够的试剂时,扩增可以在没有预定终点或饱和限制的情况下继续进行。因此,此类技术对于可与核酸探针或其靶标缔合的信号传导实体的数量没有理论上的上限。

[0175] 然而,应该理解的是,实际上与核酸探针或其靶标结合的信号传导实体的平均数量实际上不需要与其上限相同,即,信号传导实体可能实际上并不处于完全饱和(尽管它们可以这样)。例如,饱和量(或结合的信号传导实体的数量,相对于能够结合的最大数量)可能小于97%、小于95%、小于90%、小于85%、小于80%、小于75%等,和/或至少50%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%等。在一些情况下,允许更多时间发生结合和/或增加试剂浓度可能增加饱和量。

[0176] 由于与核酸探针或其靶标实际结合的信号传导实体数量的潜在上限,与不受控制的扩增相比,样品内分布(例如,空间上)的结合事件可能呈现基本一致的尺寸和/或亮度,例如上面讨论的那些。例如,由于可以与初级扩增核酸结合的次级扩增核酸的特定数量,不能发现大于与核酸探针或其靶标的固定距离的次级扩增核酸,这可能会限制指示结合的来自信号传导实体的荧光的“斑点尺寸”或直径。

[0177] 在某些实施方案中,至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%或至少95%的结合事件可以表现出基本上相同的亮度、尺寸(例如,表观直径)、颜色等,这可以使得更容易将结合事件与其他事件(例如非特异性结合、噪声等)区分开来。

[0178] 此外,在一些情况下,信号传导实体可能被灭活。例如,在一些实施方案中,可将可与信号传导实体缔合(例如,使用扩增核酸)的第一次级核酸探针或读出探针应用于可识别第一读出序列(例如,在初级或编码核酸探针上)的样品,则可在将第二次级核酸探针或读出探针(例如,其可与信号传导实体缔合(例如,使用扩增核酸))应用于样品之前使信号传导实体灭活。如果使用多个信号传导实体,则可以使用相同或不同的技术来灭活信号传导实体,并且可以例如依次或同时地灭活多个信号传导实体中的一些或全部。

[0179] 灭活可能是由于信号传导实体的去除(例如,从样品中,或从核酸探针中等),和/或通过以某种方式化学改变信号传导实体(例如,通过光漂白信号传导实体,漂白或化学改变信号传导实体的结构,例如通过还原等)来引起。例如,在一组实施方案中,荧光信号传导实体可通过化学或光学技术灭活,例如通过氧化、光漂白、化学漂白、严格洗涤或酶促消化或通过暴露于酶的反应、将信号传导实体与其他组分(例如,探针)解离、信号传导实体的化学反应(例如,与能够改变信号传导实体的结构的反应物)等。例如,漂白可通过暴露于氧气、还原剂发生或信号传导实体可从核酸探针化学裂解并通过流体流动冲走。

[0180] 在一些实施方案中,各种核酸探针可以与一种或多种信号传导实体缔合,例如,使用本文讨论的扩增核酸。如果使用多于一种核酸探针(或次级核酸探针或读出探针),则信号传导实体可各自相同或不同。在某些实施方案中,信号传导实体是能够发光的任何实体。例如,在一个实施方案中,信号传导实体是荧光的。在其他实施方案中,信号传导实体可以是磷光的、放射性的、吸收性的等。在一些情况下,信号传导实体是可以在样品内以相对高的分辨率(例如,以比可见光的波长或衍射极限更好的分辨率)确定的任何实体。信号传导

实体可以是例如染料、小分子、肽或蛋白质等。在一些情况下,信号传导实体可以是单个分子。如果使用多个次级核酸探针或读出探针,则核酸探针可以与相同或不同的信号传导实体缔合。

[0181] 信号传导实体的非限制性实例包括荧光实体(荧光团)或磷光实体,例如花青染料(例如Cy2、Cy3、Cy3B、Cy5、Cy5.5、Cy7等)、Alexa Fluor染料、Atto染料、可光切换染料、可光活化染料、荧光染料、金属纳米颗粒、半导体纳米颗粒或“量子点”,

[0182] 在一组实施方案中,信号传导实体可以通过可以被裂解以释放信号传导实体的键附接至寡核苷酸序列。在一组实施方案中,荧光团可以通过可裂解键例如可光裂解键与寡核苷酸缀合。可光裂解键的非限制性实例包括但不限于1-(2-硝基苄基)乙基、2-硝基苄基、生物素亚磷酰胺、丙烯酸亚磷酰胺、二乙基氨基香豆素、1-(4,5-二甲氧基-2-硝基苄基)乙基、环十二烷基(二甲氧基-2-硝基苄基)乙基、4-氨基甲基-3-硝基苄基、(4-硝基-3-(1-氯羰基氧基乙基)苄基)甲基-S-乙酰硫代酸酯、(4-硝基-3-(1-硫代羰基氧基乙基(thlorocarbonyloxyethyl))苄基)甲基-3-(2-吡啶二硫代丙酸)酯、3-(4,4'-二甲氧三苯甲基)-1-(2-硝基苄基)-丙烷-1,3-二醇-[2-氰基乙基-(N,N-二异丙基)]-亚磷酰胺、1-[2-硝基-5-(6-三氟乙酰基己酰胺甲基)苄基]-乙基-[2-氰基-乙基-(N,N-二异丙基)]-亚磷酰胺、1-[2-硝基-5-(6-(4,4'-二甲氧基三苯甲基氧基)丁酰胺甲基)苄基]-乙基-[2-氰基-(N,N-二异丙基)]-亚磷酰胺、1-[2-硝基-5-(6-(N-(4,4'-二甲氧基三苯甲基))-生物素氨基己酰氨基-甲基)苄基]-乙基-[2-氰基-(N,N-二异丙基)]-亚磷酰胺或类似的接头。寡核苷酸序列可以是例如初级或次级(或其他)扩增核酸,例如本文讨论的那些。

[0183] 在另一组实施方案中,荧光团可以通过二硫键与寡核苷酸缀合。二硫键可以被多种还原剂裂解,例如但不限于二硫苏糖醇、二硫赤藓糖醇、β-巯基乙醇、硼氢化钠、硫氧还蛋白、谷氧还蛋白、胰蛋白酶原、胍、二异丁基氢化铝、草酸、甲酸、抗坏血酸、亚磷酸、氯化锡、谷胱甘肽、巯基乙酸盐、2,3-二巯基丙醇、2-巯基乙醇、2-氨基乙醇、三(2-羧乙基)膦、双(2-巯基乙基)砷、N,N'-二甲基-N,N'-双(巯基乙酰基)胍、3-巯基丙酸酯、二甲基甲酰胺、硫丙基-琼脂糖、三正丁基膦、半胱氨酸、硫酸铁、亚硫酸钠、亚磷酸盐、次磷酸盐、硫代磷酸盐等,和/或这些的任何一种的组合。寡核苷酸可以是例如初级核酸探针、编码核酸探针、读出探针、初级或次级(或其他)扩增核酸,例如本文所讨论的那些。

[0184] 在另一个实施方案中,荧光团可以通过一个或多个硫代磷酸酯修饰的核苷酸与寡核苷酸缀合,其中硫修饰取代桥接和/或非桥接氧。在某些实施方案中,可以通过添加诸如但不限于碘乙醇、在乙醇中混合的碘、硝酸银或氯化汞的化合物从寡核苷酸裂解荧光团。在又一组实施方案中,信号传导实体可以通过还原或氧化而被化学灭活。例如,在一个实施方案中,可以使用硼氢化钠将诸如Cy5或Cy7的发色团还原到稳定的非荧光状态。在另一组实施方案中,荧光团可以通过偶氮键与寡核苷酸缀合,并且偶氮键可以用2-[(2-N-芳基氨基)苄基偶氮]吡啶裂解。在又一组实施方案中,荧光团可以通过合适的核酸区段与寡核苷酸缀合,该核酸区段可以在适当暴露于DNA酶(例如外切脱氧核糖核酸酶或内切脱氧核糖核酸酶)时被裂解。实例包括但不限于脱氧核糖核酸酶I或脱氧核糖核酸酶II。在一组实施方案中,裂解可以通过限制性核酸内切酶发生。可能合适的限制性内切核酸酶的非限制性实例包括BamHI、BsrI、NotI、XmaI、PspAI、DpnI、MboI、MnII、Eco57I、Ksp632I、DraIII、AhaII、SmaI、MluI、HpaI、ApaI、BclI、BstEII、TaqI、EcoRI、SacI、HindII、HaeII、DraII、Tsp509I、

Sau3AI、PacI等。已详细研究了3000多种限制酶,其中600多种可商购获得。在又一组实施方案中,荧光团可以与生物素缀合,并且寡核苷酸可以与抗生物素蛋白或链霉抗生物素蛋白缀合。生物素和抗生物素蛋白或链霉抗生物素蛋白之间的相互作用允许荧光团与寡核苷酸结合,而在充分暴露于过量的添加时,游离生物素可能“胜过”连接,从而导致荧光团与寡核苷酸解除结合。此外,在另一组实施方案中,可以使用相应的“toe-hold-probe”去除探针,该“toe-hold-probe”包含与次级或读出探针相同的序列,以及额外数量的与初级或编码探针同源的碱基(例如,1-20个额外的碱基,例如,5个额外的碱基)。这些探针可以通过链置换相互作用去除标记的次级或读出探针。寡核苷酸可以是例如初级核酸探针、编码核酸探针、读出探针、初级或次级(或其他)扩增核酸,例如本文所讨论的那些。

[0185] 如本文所用,术语“光”通常是指具有任何合适波长(或等效地,频率)的电磁辐射。例如,在一些实施方案中,光可以包括光学或视觉范围内的波长(例如,具有在约400nm和约700nm之间的波长,即“可见光”)、红外波长(例如,具有约300微米至700纳米之间的波长)、紫外波长(例如,具有约400纳米至约10纳米之间的波长)等。在某些情况下,如本文所讨论的,可以使用多于一种实体,即化学上不同或有区别(例如,在结构上)的实体。然而,在其他情况下,实体可以在化学上相同或至少在化学上基本相同。

[0186] 在一组实施方案中,信号传导实体是“可切换的”,即,该实体可以在两种或更多种状态之间切换,其中至少一种状态发射具有所需波长的光。在其他状态下,实体可能不发射光,或发射不同波长的光。例如,实体可以被“激活”到能够产生具有所需波长的光的第一状态,并且“灭活”到不能发射相同波长的光的第二状态。如果一个实体可以被合适波长的入射光激活,那么它就是“可光激活的”。作为非限制性示例,Cy5或Alexa 647可以通过不同波长的光以可控和可逆的方式在荧光和暗状态之间切换,即633nm(或642nm、647nm、656nm)红光可将Cy5或Alexa 647切换或灭活到稳定的暗状态,而405nm绿光可将Cy5或Alexa 647切换或激活回荧光状态。在一些情况下,实体可以在两种或更多种状态之间可逆地切换,例如,在暴露于适当的刺激时。例如,第一刺激(例如,第一波长的光)可用于激活可切换实体,而第二刺激(例如,第二波长的光)可用于灭活可切换实体至例如非发光状态。可以使用任何合适的方法来激活实体。例如,在一个实施方案中,可以使用合适波长的入射光来激活实体以发射光,即,实体是“可光切换的”。因此,可光切换实体可以通过例如不同波长的入射光在不同的发光或不发光状态之间切换。光可以是单色的(例如,使用激光产生的)或多色的。在另一个实施方案中,实体可以在受到电场和/或磁场刺激时被激活。在其他实施方案中,实体可在暴露于合适的化学环境时(例如通过调节pH或诱导涉及实体的可逆化学反应等)被激活。类似地,任何合适的方法可用于使实体灭活,并且激活和灭活实体的方法不必相同。例如,实体可以在暴露于合适波长的入射光时被灭活,或者实体可以通过等待足够的时间来被灭活。

[0187] 通常,本领域普通技术人员可以通过确定其中第一状态的实体在暴露于激发波长时可以发光的条件,将实体从第一状态切换到第二状态(例如,在暴露于切换波长的光时),然后显示实体在处于第二状态时在暴露于激发波长时不再能够发光(或以大大降低的强度发光)来鉴定“可切换”实体。

[0188] 在一组实施方案中,如所讨论的,可切换实体可在暴露于光时切换。在一些情况下,用于激活可切换实体的光可以来自外部源例如光源诸如激光光源、靠近可切换实体的

另一个发光实体等。在一些情况下,第二发光实体可以是荧光实体,并且在某些实施方案中,第二发光实体本身也可以是可切换实体。

[0189] 在一些实施方案中,可切换实体包括第一发光部分(例如,荧光团)和激活或“切换”第一部分的第二部分。例如,在暴露于光时,可切换实体的第二部分可以激活第一部分,导致第一部分发光。激活剂部分的例子包括但不限于Alexa Fluor 405 (Invitrogen)、Alexa Fluor488 (Invitrogen)、Cy2 (GE Healthcare)、Cy3 (GE Healthcare)、Cy3B (GE Healthcare)、Cy3.5 (GE Healthcare),或其他合适的染料。发光部分的示例包括但不限于Cy3B (GE Healthcare)、Cy5、Cy5.5 (GE Healthcare)、Cy7 (GE Healthcare)、Alexa Fluor 647 (Invitrogen)、Alexa Fluor 680 (Invitrogen)、Alexa Fluor 700 (Invitrogen)、Alexa Fluor750 (Invitrogen)、Alexa Fluor 790 (Invitrogen)、DiD、DiR、YOYO-3 (Invitrogen)、YO-PRO-3 (Invitrogen)、TOT-3 (Invitrogen)、TO-PRO-3 (Invitrogen)或其他合适的染料。参见,例如,美国专利号7,838,302,通过引用整体并入本文。在一些情况下,随后可以通过任何合适的技术(例如,通过将647nm红光引导至分子的Cy5部分)使第一发光部分灭活。

[0190] 在一些实施方案中,使用具有不同序列的多个核酸探针,并且基于每个核酸探针的结合模式,依次分析每个核酸探针的分布并将其用于为每个位置创建“代码字”。通过选择定义合适代码空间的核酸探针,可以鉴定和/或丢弃和/或校正观察到的结合模式中的明显错误以鉴定正确的代码字,从而鉴定样品内核酸探针的正确靶标。这种抗错和错误校正系统最初是为多重抗错荧光原位杂交(MERFISH)引入的,随后也被用于各种相关技术。参见,例如,国际专利申请公开号W02016/018960和W02016/018963,各自通过引用整体并入本文。

[0191] 如所提到的,在某些实施方案中,这种技术可以与错误校正相结合,例如,如在MERFISH或其他类似技术中使用的那样。例如,代码字可以基于能够结合初级或编码核酸探针上的读出序列的多个读出探针的结合(或非结合),并且在一些情况下,代码字可以定义错误校正代码以帮助减少或防止核酸探针的错误鉴定。在一些情况下,可以使用相对少量的读出探针来鉴定相对大量的不同靶标,例如,通过使用各种组合方法。荧光显微术、宽视野荧光显微术、落射荧光显微术、共聚焦显微术或光片显微术可用于图像采集。诸如STORM或其他超分辨率成像方法的图像采集技术也可用于对此类样品进行成像并促进核酸探针的确定。关于诸如MERFISH的技术的更多细节参见例如美国专利号9,712,805或10,073,035,或国际专利申请公开号W02008/091296或W02009/085218,每一个都通过引用整体并入本文。在一些情况下,也可以使用膨胀显微术,其中样品在成像之前膨胀。参见,例如,题为“Matrix Imprinting and Clearing”的国际专利申请公开号W02018/089445或题为“Multiplexed Imaging Using MERFISH and Expansion Microscopy”的国际专利申请公开号W02018/089438,每个都通过引用整体并入本文。

[0192] 另一方面涉及一种计算机实现的方法。例如,可以提供能够自动和/或重复执行本文描述的任何方法的计算机和/或自动化系统。如本文所用,“自动化”设备是指能够在没有人指导的情况下操作的设备,即,自动化设备可以在任何人完成采取任何行动以促进功能(例如通过在向计算机输入指令来启动该过程)之后的一段时间内执行功能。通常,自动化设备可以在此时间点之后执行重复功能。在一些情况下,处理步骤也可以记录在机器可读介质上。

[0193] 例如,在一些情况下,计算机可用于控制样品的成像,例如,使用荧光显微术、宽视野荧光显微术、落射荧光显微术、共聚焦显微术、光片显微术、衍射极限光学显微术、STORM或其他超分辨率技术,例如本文所述的那些。在一些情况下,计算机还可以控制操作诸如漂移校正、物理配准、图像分析中的杂交和簇对齐、簇解码(例如,荧光簇解码)、错误检测或校正(例如,如本文所讨论的)、降噪、从背景特征(例如图像中的噪声或碎片)中鉴定前景特征等。作为示例,计算机可用于控制样品内信号传导实体的激活和/或激发和/或灭活,和/或信号传导实体的图像的采集。在一组实施方案中,可以使用具有各种波长和/或强度的光来激发样品,并且可以使用计算机将用于激发样品的光的波长序列与包含信号传导实体的样品的获得的图像相关联。例如,计算机可以将具有各种波长和/或强度的光施加到样品以在每个感兴趣区域中产生不同平均数量的信号传导实体(例如,每个位置一个激活的实体,每个位置两个激活的实体等)。在一些情况下,该信息可用于构建图像和/或确定信号传导实体的位置,在一些情况下以高分辨率实现,如上所述。

[0194] 在一些方面,样品位于显微镜上。在一些情况下,显微镜可以包含一个或多个通道,例如流体或微流体通道,以引导或控制流体流入或流出样品。例如,在一个实施方案中,可以通过使流体流经一个或多个通道而流入或流出样品来从样品中引入和/或去除核酸探针(例如本文所讨论的那些)。在一些情况下,还可以有一个或多个腔室或储存器用于保持流体,例如,与通道和/或与样品流体连通。本领域普通技术人员将熟悉用于将流体移入或移出样品的通道,包括流体或微流体通道。

[0195] 以下文件通过引用并入本文:题为“Systems and Methods for High-Throughput Image-Based Screening”的国际专利申请公开号W0 2018/218150;题为“Systems and Methods for Determining Nucleic Acids”的W0 2016/018960;题为“Probe Library Construction”的W0 2016/018963;题为“Matrix Imprinting and Clearing”的W0 2018/089445;题为“Multiplexed Imaging Using MERFISH and Expansion Microscopy”的W0 2018/089438;和题为“Imaging-Based Pooled CRISPR Screening”的美国专利申请系列号62/836,578和题为“Amplification Methods and Systems for MERFISH and Other Applications”的62/779,333。以下文件也通过引用整体并入本文:美国专利号2017/0220733和2017/0212986。

[0196] 此外,题为“Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin”的美国专利申请公开号62/954,720和题为“Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin”的美国专利申请公开号63/060,947各自通过引用整体并入本文。

[0197] 以下实施例旨在说明本发明的某些实施方案,但并不示例说明本发明的全部范围。

实施例

[0198] 实施例1

[0199] 以下实施例显示了用于在单个细胞中对染色质的3D组织进行成像的大规模多重FISH方法,并进一步证明了通过将染色质和新生转录物成像(两者都在基因组规模上)与核结构鉴定组合将3D基因组组织置于其天然结构和功能环境中的能力。

[0200] 第一个实施例报告了允许对单细胞中的染色质组织进行基因组规模成像的大规模多重FISH方法。使用这种方法,证实了在单个细胞中对人类基因组中>1,000个不同的基因组基因座(约2,000个染色质基因座,计算了染色体的同源对)进行成像和鉴定。此外,使用在各种核结构(包括核斑点、核仁和核纤层)的背景下存在于这些基因座中的>1,000个基因的新生RNA转录物证实了这些基因组基因座的同時成像。该方法用于探索单细胞中染色质组织、转录活动和核环境之间的关系。

[0201] 为了实现基因组规模的染色质成像,设计了一种组合FISH方法,其灵感来自先前为转录组成像开发的多重抗错FISH方法,但通过考虑染色质的聚合性质(即基因组序列中的相邻基因座在空间上接近)和染色体的区域组织(即不同的染色体倾向于占据不同的空间区域)进行了为染色质成像专门设计的重大修改。参见,例如,题为“Systems and Methods for Determining Nucleic Acids”的W02016/018960;题为“Probe Library Construction”的W02016/018963,每个都通过引用整体并入本文。为了允许组合成像,每个基因组基因座被分配一个唯一的100位二进制代码,Hamming权重为2,即每个条形码包含两个“1”位和98个“0”位(图1A)。这些条形码中的位值确定了在成像的连续轮次中每个基因座的信号的存在(1)或不存在(0)。为了避免在相同位中同时成像空间上接近的染色质区域,从这些100位Hamming权重2条形码中,进一步选择了一个子集来编码靶基因组基因座并优化这些条形码的分配,使得在相同条形码位置中具有“1”位的基因座在基因组空间中最大程度地分离。这种策略可以最大限度地减少由来自附近染色质基因座的重叠信号引起的检测错误。此外,由于绝大多数可能的100位二进制代码是无效的(即未分配给任何靶基因座),这种设计允许鉴定和丢弃检测错误并进一步提高测量精度。

[0202] 使用高度多样性的编码探针库将条形码物理印在靶基因组基因座上,每个编码探针库包含用于与靶基因座之一结合的40-nt靶区域和从100个预先设计的读出序列中选择的20-nt读出序列(图1A)。每个读出序列对应于100个位中的一个,并且每个基因组基因座的编码探针组(每个基因座约400个探针)仅包含两个不同的读出序列,对应于分配给该基因座的条形码中读取“1”的两个位。在编码探针结合后,通过荧光标记的读出探针的依次杂交检测印在染色质基因座上的条形码,每个探针与100个读出序列之一互补(图1A)。每个杂交轮次引入两个不同的读出探针,并在两个颜色通道中成像,这样所有约1000个基因组基因座在50轮杂交后都被成像和鉴定(图1A-1C)。相比之下,对1000个基因组基因座进行成像的直接顺序方法将需要使用双色成像的500轮杂交。由于二倍体细胞中的每个染色体都有两个同源物,因此使用聚类算法进一步分配成像的基因座的同源物身份,利用染色体在每个细胞核中占据不同区域的倾向。

[0203] 在本实施例中,选择了1,041个基因组基因座进行成像,每个基因座尺寸为约30-kb,均匀覆盖人肺成纤维细胞(IMR90)细胞中的22个常染色体和X染色体。还要求每个染色体至少包含30个靶基因座,因此每个染色体同源物成像的基因座数量为30至80,取决于染色体的长度。在5个生物复制中约5,400个个体细胞中的这1,041个基因组基因座以每个基因座约80%的检测效率成像,考虑到每个染色体有两个同源物,在每个细胞中检测到约1700个染色质基因座(图1D-1E)。

[0204] 为了获得染色质组织的群体平均视图,在每个细胞中,计算每对成像染色质基因座之间的空间距离,然后在所有成像的细胞中确定每对基因座之间的中位距离和接触频率

(图1F和图6A)。从成像数据确定的同一染色体内染色质基因座对之间的接触频率显示出与整体Hi-C检测到的接触频率高度相关,Pearson相关系数为0.91(图6B)。成像数据捕获了多个尺度的染色质结构,从染色体组织成区域(图1F和图6A)到染色体臂内A和B隔室的形成(图7A),这也与通过整体Hi-C测量鉴定的隔室一致(图7B-7C)。此外,成像结果显示独立生物学重复之间的高再现性(图8)。

[0205] 通过探索个体细胞中的染色质组织,染色体在占据每个细胞内不同区域的同时(图1F-1G)也显示出彼此之间的显著重叠(图1G-1H)。平均而言,任何给定染色体占据的凸包体积的约80%与同一细胞中的其他染色体共有(图1I),表明高度的跨染色体相互作用。由于这些相互作用未被充分探索,下面的分析集中在这些跨染色体相互作用上。

[0206] 图1A-1I显示了基因组规模的染色质成像。图1A显示了成像方案。靶基因组基因座被分配了抗错条形码,例如Hamming权重为2的100位二进制条形码(即100位中的两个读取“1”)。条形码用编码寡核苷酸探针印在基因组基因座上,该探针识别基因座并将两个不同的读出序列与每个基因座相关联,对应于分配给基因座的条形码中读取“1”的两个位。每个基因座被总共400个编码探针标记,但只显示了4个。依次添加与读出序列互补的荧光读出探针并成像,从而允许在每个基因座处读取“1”的位,从而确定该基因座的条形码身份。对约1000个基因组基因座进行了成像。图1B显示了单个细胞的核中多个成像轮次的代表性图像。来自读出探针的染色质基因座的荧光信号以较浅的阴影显示,而用作核标记的4',6-二脒基-2-苯基吲哚(DAPI)的信号以较暗的阴影显示。比例尺:5微米。图1C是在所有成像轮次中以一个染色质基因座为中心的小空间区域(图1B中的框)的放大图像。基于发出信号的两个读出探针(1和13)确定基因座身份。比例尺:300nm。图1D是单个细胞中所有检测到的染色质基因座的3D渲染,根据它们所属的染色体进行灰度化。基因组序列中的相邻基因座由细线连接。图1E显示了与图1D中相同细胞的染色质基因座,但所示染色体的两个同源物以与其他基因座不同的灰度显示。图1F是从约5,400个单细胞计算得出的中值距离矩阵。对于每对基因座,显示基因座之间的所有观察到的3D空间距离的中值。图1G显示了显示单个细胞中多个染色体区域位置的示例图像。阴影区域代表每个染色体周围的凸包,它被用作染色体区域的操作定义。图1H显示了图1G中所示相同细胞距离矩阵。显示了每对染色质基因座之间的空间距离。染色体顺序如热图下方所述,其中分别显示每个染色体的两个同源物。图1I是同一细胞中至少一个其他染色体共有的每个染色体区域的体积分数的量化。显示了中位数(中心线)、第25至第75个百分位数(方框)和第5至第95个百分位数(须线)。n=10,910个染色体拷贝(5,455个细胞,对于每个染色体每个细胞有两个同源拷贝)。

[0207] 图6A-6B显示了源自基因组规模成像的接触频率矩阵以及与整体Hi-C数据的比较。图6A显示了在该实施例中成像的所有1041个基因组基因座的接触频率矩阵。一对基因座之间的接触频率计算为基因座之间测量的距离小于500nm的发生次数除以两个基因座之间测量的距离的总数。图6B显示了源自成像数据的染色体内的基因座对与源自整体Hi-C实验的那些之间的接触频率的相关图,以500kb分箱并以靶基因座为中心。Pearson相关系数为0.91。

[0208] 图7A-7C显示了源自基因组规模成像的亚染色体结构以及与整体Hi-C数据的比较。图7A显示了从22号染色体的一个臂的成像数据生成的接触频率矩阵。矩阵下方的条中显示了基于该矩阵将每个基因座分配给A或B隔室。图7B显示了22号染色体的同一臂的接触

频率矩阵,从Hi-C数据计算,以500kb分箱并以靶基因座为中心。矩阵下方的条中显示了基于该矩阵的每个基因座的A和B隔室分配,使用与图7A相同的程序进行分配。从成像数据和Hi-C数据得出的A/B隔室分配是相同的。图7C显示了源自成像数据的22号染色体中的基因座对与源自整体Hi-C实验的那些之间的接触频率的相关图。Pearson相关系数为0.91。

[0209] 图8显示了重复实验之间染色质成像实验的再现性。图中显示的是在1041个基因组基因座成像实验的两个独立生物学重复中观察到的染色质基因座之间成对距离的相关性。重复实验之间的Pearson相关系数为0.98。右上方的云表示跨染色体成对距离,左下方的云表示染色体内存对距离。

[0210] 实施例2

[0211] 在这个实施例中研究了跨染色体相互作用如何依赖于染色质的表观遗传特性。先前通过Hi-C和成像分析表明,染色质被分离为A和B隔室,分别富含活性和非活性染色质。不同的机制可以介导活性-活性和非活性-非活性染色质相互作用,例如HP1介导的异染色质浓缩和转录因子和辅因子介导的活性染色质浓缩。在此实施例中,每个成像的基因组基因座都使用基于已发布的整体Hi-C数据的已建立的调用方法被分类为A和B隔室。38%的成像基因座属于A隔室,倾向于相对富含基因并富含活性染色质标记,如H3K27Ac,而62%属于B隔室,倾向于富含非活性染色质标记,如H3K9me3。为了检查活性和非活性染色质的跨染色体相互作用的程度是否不同,基因组基因座在跨染色体接触频率矩阵中排序,将所有A基因座彼此相邻放置,同样将所有B基因座放在一起。该矩阵显示,与B隔室基因座相比,A隔室基因座与A隔室基因座进行跨染色体相互作用的趋势明显更强(图2A-2B)。相比之下,B隔室基因座彼此之间没有显示出相似的跨染色体亲和力,而是显示出与A隔室染色质跨染色体相互作用的略高的概率(图2A-2B)。换言之,跨染色体A-A相互作用看起来具有明显强于A-B相互作用的趋势,而A-B相互作用又看起来具有比B-B相互作用略强的趋势。这与同一染色体内存的顺式相互作用形成鲜明对比,其中A和B隔室倾向于分离,导致A-A和B-B相互作用相对于A-B相互作用的富集。

[0212] 接下来,在单细胞水平上检查了跨染色体相互作用的表观遗传依赖性。在个体细胞中,A隔室和B隔室基因座采用不同的空间分布,其中A基因座在细胞核中表现出比B基因座更中心地定位的趋势(图2C和图9A-9B)。A和B基因座之间也存在相当程度的混合(图2C和图9A-9B)。对于每个染色体中的每个成像的基因座,计算其来自所有其他染色体的A基因座和B基因座的局部密度,并确定这两个密度的比率(以下称为跨染色体A/B密度比)(图2C)。该数量提供了对基因座附近跨染色体活性染色质的局部富集的测量。大部分(62%)的成像的基因座属于B隔室,导致A/B比率的总体偏差小于1。为了控制这种偏差,将针对A基因座和B基因座的观察到的跨染色体A/B密度比的分布与随机对照中获得的分布进行比较,随机对照中成像的基因座的A和B身份在成像的基因座之间随机打乱,同时保持A和B基因座的数量不变。值得注意的是,观察到的A基因座的跨染色体A/B密度比显著高于观察到的B基因座的值,而B基因座的值又高于随机对照得出的值(图2D),并且在大多数单细胞中观察到了这种趋势(图2E)。这些单细胞分析再次支持了跨染色体相互作用优先富集活性染色质之间的相互作用的概念。

[0213] 图2A-2E显示跨染色体接触优先富集活性染色质之间的相互作用。图2A显示了标准化的跨染色体接触频率矩阵。显示了每个跨染色体基因座对(不同染色体上的基因座对)

之间的接触频率。基因座被重新排序,使得A隔室基因座首先出现,然后是B隔室基因座,因此左上角的块代表A基因座对之间的相互作用,右下角代表B基因座对之间的相互作用。矩阵中的每个条目都通过源自同一对染色体的所有基因座对的中值接触频率进行标准化,以解释染色体对之间相互作用的不同基础水平。图2B显示了A基因座对(AA,右; $n=72,771$ 个基因座对)、B基因座对(BB,左; $n=193,753$ 个基因座对)和由一个A和一个B基因座组成的对(AB, $n=237,986$ 个基因座对)的跨染色体接触频率分布,源自图2A中所示的矩阵。分布在顶部图中表示为直方图,在底部图中表示为箱线图,显示中位数(中心线)、第25至第75个百分位数(框)和第5至第95个百分位数(须线)。图2C显示了单细胞中A隔室和B隔室基因座的分布。左图表示单个核中单个z平面内所有检测到的基因座的位置。A隔室基因座显示在标尺的顶部,而B隔室基因座显示在底部。在右图中,每个基因座的阴影表示跨染色体A和B基因座的局部密度的比率,与右侧显示的阴影比例尺一致。图2D显示了成像的基因组基因座的局部跨染色体A/B密度比的分布。对于每个基因座,确定了所有细胞的中值A/B密度比,对于A隔室基因座($n=382$ 个基因座)和B隔室基因座($n=623$ 个基因座)显示了不同基因座的分布。由于本研究中使用不同形式的基因组组装和用于隔室调用的Hi-C数据集,1041个成像的基因座中的36个未分配A/B身份。深灰色直方图表示随机对照,其中A和B隔室身份被随机打乱,同时保持A基因座的总数和B基因座的总数不变。图2E显示了相对于随机对照的跨染色体A/B密度比的富集的分布。对于每个成像的细胞,所有A基因座的中值A/B密度比除以随机对照的中值A/B密度比,如图2D中所述,并且呈现了该值在所有成像的细胞($n=5,455$ 个细胞)中的分布。显示了B基因座的相同富集的分布($n=5,455$ 个细胞)。线条标记了1的值,即没有富集。

[0214] 图9A-9B显示A隔室和B隔室基因座在单个细胞中显示不同的空间分布。图9A,左图显示了在单个细胞的单个z平面中显示A隔室基因座和B隔室基因座的示例图像。右图显示了这些单细胞中A隔室基因座和B隔室基因座到核外围的距离的分布。核外围被鉴定为围绕所有检测到的染色质基因座的凸包。直方图显示了在检测到的染色质基因座周围的凸包内均匀采样的点与核外围的距离的分布。图9B显示了A隔室基因座和B隔室基因座到核外围的距离的群体平均分布。 $n=382$ 个A基因座; $n=623$ 个B基因座。

[0215] 实施例3

[0216] 为了将染色质的3D组织置于其功能活动和其他核结构的背景下,在此实施例中扩展了成像方法,以允许同时测量染色质组织以及许多基因组基因座的转录活动以及核内的标志性结构。具体来说,1,041个基因组基因座与从位于这些基因座的1,137个基因中的每一个转录的新生RNA一起成像,同时与重要的核结构(包括核斑点和核仁)一起成像(图3A)。

[0217] 为了允许在同一细胞内进行DNA、RNA和核结构成像,通过采用与上文针对染色质所述相似的组合成像策略对1,137个基因的内含子RNA进行多重成像(图3A)。考虑到并非所有基因都会在每个个体细胞中转录,因此转录灶的密度不应像染色质基因座的密度那样高,因此RNA采用54位Hamming权重2代码进行编码,并选择了1,137个可能条形码来编码基因,其方式类似于选择用于染色质成像的条形码以最小化在相同位中成像空间邻近的基因的机会。完成RNA成像后,对RNA转录物进行酶消化(在我们的单模态染色质成像实验中也进行了这一步骤),并如上所述进行多重DNA FISH以对1,041个基因组基因座进行成像(图3A)。基因组基因座和新生RNA转录物的解码在很大程度上是独立进行的,另外还有一个限

制条件,即转录物与其所在的基因组位点共定位。该程序进一步提高了转录RNA的检测准确性,并允许估计每个基因组基因座的转录爆发的检测效率($\sim 90\%$)。最后,使用针对这些结构的已知分子成分的免疫荧光对核斑点和核仁进行成像(图3A)。通过计算包含所有成像的基因组基因座的凸包并确定凸包的边界来估计核纤层的位置。总之,这些多模态测量允许3D基因组结构、转录活动和核组织的集成单细胞视图(图3B)。这些多模态成像实验是在约3700个个体细胞上在两个生物复制中进行的。来自这些多模态实验的染色质成像数据也包含在上述5个重复(约5,400个细胞)中,用于3D基因组组织分析。

[0218] 从这些多模态实验的新生RNA转录物测量中,针对每个基因量化了作为活性转录基因的细胞的分数的转录爆发频率(图3C)和来自RNA内含子信号亮度的中值爆发尺寸(图3D)。这些测量在重复实验中显示出高度相关性(图10A-10B)。爆发频率表现出双峰行为,其中高爆发频率基因主要存在于A隔室中,而低爆发频率基因存在于两个隔室中(图3C)。此外,使用250nm的截止空间距离估计特定染色质基因座是否与核体相关,并发现B隔室基因座与核纤层的更高的关联频率(图11)和A隔室基因座与核斑点的更高的关联频率(图12)。这些结果与先前观察到的非活性和活性染色质分别与核纤层和核斑点优先关联的观察结果一致。对于个体基因座,它们的局部跨染色体A/B密度比与核纤层关联频率呈负相关(图3E),与核斑点关联频率呈正相关(图3F)。最后,核仁显示出与着丝粒、某些染色体的端粒以及含有核糖体编码基因的染色体的优先关联(图3G)。这些生物学结果为多模态测量提供了进一步的验证。

[0219] 值得注意的是,对于基本上所有成像的基因座,核纤层关联降低了观察到的转录活动(图3H),而核斑点关联与大多数成像的基因座的更高转录活动相关(图3H)。此外,在用 α -鹅膏蕈素处理以抑制转录后,几乎所有基因座与核纤层的关联率总体上增加,而与核斑点的关联率总体上降低(图13A-13C)。这些结果扩展了先前关于转录激活或抑制后单个或几个基因组区域的核重新定位的成像研究,并提供了转录活动和与核结构的相互作用之间的关系基因组规模视图。

[0220] 图3A-3H显示了核结构背景下染色质和转录活动的基因组规模成像。图3A是结合染色质(左图)、新生RNA转录物(中图)和核体(右图)成像以生成核结构的背景下的染色质组织和功能活动的集成视图的多模态成像方案的图示。对约1000个基因组基因座、靶基因座中约1100个基因的新生RNA转录物和两种类型的核体(核斑点和核仁)进行成像。以下是每种成像模式的代表性原始图像:多个成像轮次的染色质基因座(左)、多个成像轮次的新生RNA转录物(中)和核体(右:核斑点;左:核仁)。比例尺:5微米。图3B是单个细胞中染色质基因座、转录爆发和核体的3D渲染。左:所有检测到的染色质基因座,按染色体灰度化(基于下面显示的染色体索引)。中间:显示为球体的所有检测到的内含子RNA,阴影表示成像基因的身份,球体尺寸表示转录爆发尺寸。染色质基因座显示在背景中。右:检测到的核体的体积填充表示。核仁和核斑点以不同的阴影显示。核纤层被鉴定为围绕所有检测到的染色质基因座的凸包表面。图3C-3D是位于A隔室基因座($n=494$ 个基因)和B隔室基因座($n=625$ 个基因)中的基因的转录爆发频率(图3C)和爆发尺寸(图3D)的分布。图3E-3F是每个成像的基因组基因座的局部跨染色体A/B密度比作为发现与核纤层(图3E)和核斑点(图3F)相关的基因座的频率的函数的散点图。如果基因座与核结构的测量的距离小于250nm,则认为该基因座与核结构相关联。图中显示的跨染色体A/B密度比的值是所有成像的细胞的中值。图3G显

示了所有成像的基因组基因座与核仁的关联频率,按基因组位置排序。垂直线是着丝粒的位置,括号突出显示含有核糖体编码基因(rDNA)的染色体。图3H显示了核结构缔合对转录的影响。当比较其中基因座是核纤层关联相对非核纤层关联(左)和斑点关联相对非斑点关联(右)的细胞群时,圆圈是每个基因座的转录爆发频率的变化倍数。虚线突出显示没有变化,实线表示每种情况下的中值变化倍数。

[0221] 图10A-10B显示了重复实验之间的新生RNA转录物成像实验的再现性。图10A-10B显示了每个基因的爆发频率(图10A)和爆发尺寸(图10B)的RNA成像的重复实验之间的相关性。Pearson相关系数分别为0.94和0.81。

[0222] 图11显示了B隔壁基因座与核纤层的优先关联。显示了A隔壁基因座($n=382$ 个基因座)和B隔壁基因座($n=623$ 个基因座)与核纤层的关联率的分布。如果基因座与核外围的距离小于250nm,则该基因座在操作上被定义为与核纤层关联。

[0223] 图12显示了A隔壁基因座与核斑点的优先关联。显示了A隔壁基因座($n=382$ 个基因座)和B隔壁基因座($n=623$ 个基因座)与核斑点的关联率的分布。如果基因座与最近的斑点的距离小于250nm,则基因座在操作上被定义为与斑点关联。

[0224] 图13A-13C显示了转录抑制后核纤层和核斑点关联的变化。图13A-13B显示了未处理细胞(图13A)和用转录抑制剂 α -鹅膏蕈素处理的细胞(图13B)所显示的具有成像的染色质基因座、核仁和核斑点的个体核的代表性图像。图13C显示了在通过 α -鹅膏蕈素抑制转录时每个基因座与核纤层(左)和核斑点(右)的关联率的变化倍数。每个基因组基因座的数据点以圆圈显示,实线是每种情况下所有基因座的中值变化倍数,虚线表示没有变化。

[0225] 实施例4

[0226] 在此实施例中,这些多模态单细胞测量用于进一步表征转录活动和核结构背景下的跨染色体相互作用。鉴于观察到跨染色体相互作用优先富集A隔壁基因座之间的相互作用,测试了这些相互作用是否与染色质的转录活动相关。为此,计算了每个细胞中每个基因座的来自其他染色体的A和B染色质的局部密度和跨染色体A/B密度比,并确定了两个细胞群的这些数量的中值:(i)所考虑的基因座表现出转录活动的细胞(即RNA爆发信号),和(ii)基因座至少暂时出现转录沉默的细胞(图4A)。值得注意的是,除了观察到A隔壁基因座显示出比B隔壁基因座更高的局部跨染色体A/B密度比(图2D-2E),即使对于相同的基因座,也观察到当基因座活性转录时更高的跨染色体A密度和A/B密度比的一致趋势(图4B和图14)。在转录活动和跨染色体相互作用之间观察到的这种相关性与以下两种解释一致:染色质基因座的较高表观遗传或转录活动增加其跨染色体相互作用的速率,或基因座在富含活性染色质的环境中的定位增强其转录活动。

[0227] 图4A-4F显示活性染色质之间的优先跨染色体相互作用与转录相关并且在干扰凝聚物形成的处理时被破坏。图4A是染色质基因座和转录活动的单细胞图像。左:来自单个核的单个z平面中所有成像的A隔壁(标尺的上方)和B(下方)基因座的位置。中间:基于灰度比例尺的相同基因座的局部跨染色体A/B密度比。右:与中间图相同,其中检测到的转录爆发叠加并显示为圆圈。图4B显示了转录状态与沉默状态下每个基因座的局部跨染色体A/B密度比的比较。对于包含至少一个成像的基因的每个基因组基因座,计算其活性转录(称为转录)的细胞和未转录(称为沉默)的细胞的跨染色体A/B密度比。针对每个状态显示跨细胞的中值。基因座按它们在沉默状态下的A/B密度比排序,并为沉默和转录状态绘制A/B密度比。

图4C显示了用 α -鹅膏蕈素处理以抑制转录的细胞的标准化跨染色体接触频率矩阵。如图2A中所述,矩阵是有序和标准化的。图4D显示了显示为箱线图的AA、BB和AB接触频率的分布,如在图2B中所描述的。对于AA的 $n=72,771$ 个基因座对,对于BB的 $n=193,753$ 个基因座对,对于AB的 $n=237,986$ 个基因座对。图4E-4F与图4C-4D相同,但用1,6-己二醇处理细胞。

[0228] 图14显示了当基因座处于活性转录状态或沉默状态时,每个成像基因座附近的跨染色体A基因座的局部密度。对于每个基因座,取决于基因座是活性转录还是沉默的,将细胞分为两组。显示了这两组细胞(转录和沉默)的A基因座的中值局部密度。基因座基于它们在沉默状态下的局部跨染色体A基因座密度进行排序。

[0229] 因为核斑点是集中了活性转录的基因座的最突出的核体之一,所以不知道与核斑点的关联是否可以为跨染色体的活性-活性染色质相互作用的观察到的优先发生提供简单的解释。有趣的是,当分析限于与核斑点无关的基因座时,观察到跨染色体接触中A-A相对于A-B和B-B相互作用富集的相同趋势(图15A)和活性转录基因座表现出相对于沉默基因座更高的局部A/B密度比的相同趋势(图15B)。值得注意的是,即使仅考虑与核纤层关联并因此在富含B隔室染色质的环境中的基因座,也观察到这些趋势(图16A-16B)。后一种结果还表明,观察到的活性-活性跨染色体相互作用的富集不能简单地解释活性染色质更集中于细胞核中心的事实。

[0230] 图15A-15B显示与核斑点无关的染色质基因座中活性-活性跨染色体相互作用的富集。图15A显示了A基因座对(AA)、B基因座对(BB)和由一个A和一个B基因座组成的对(AB)之间的跨染色体接触频率,仅考虑其中两个基因座不与核斑点关联的细胞。接触频率如图2A所述被标准化。分布表示为箱线图,如图2B中所述。对于AA的 $n=72,771$ 个基因座对(左),对于BB的 $n=193,753$ 个基因座对(右),对于AB的 $n=237,986$ 个基因座对(中)。为了比较,所有数据的中值,无论斑点关联状态如何,都显示为三角形。图15B显示了与核斑点无关的基因座在转录状态和沉默状态之间的局部跨染色体A/B密度比的变化倍数。对于每个基因组基因座,计算基因座转录状态和沉默状态之间局部跨染色体A/B密度比的变化倍数,仅考虑其中该基因座不与核斑点关联的细胞。确定每个基因座的每个状态(转录或沉默)的中值A/B密度比,左侧显示两种状态之间的变化倍数(每个圆圈对应于基因组基因座)。无论基因座的核斑点关联状态如何,从所有数据得出的相应变化倍数都显示在右侧以进行比较。虚线表示没有变化,实线表示每种情况下所有基因座的中值变化倍数。

[0231] 图16A-16B显示了与核纤层关联的染色质基因座中活性-活性跨染色体相互作用的富集。图16A显示了A基因座对(AA)、B基因座对(BB)和由一个A和一个B基因座组成的对(AB)之间的跨染色体接触频率,仅考虑与核纤层关联的基因座对(在250nm内)。接触频率如图2A所述被标准化。分布表示为箱线图,如图2B中所述。对于AA的 $n=72,771$ 个基因座对(左),对于BB的 $n=193,753$ 个基因座对(右),对于AB的 $n=237,986$ 个基因座对(中)。为了比较,所有数据的中值,无论核纤层关联状态如何,都显示为三角形。在这些情况下,方差相对较大,因为只有相对较小部分的基因座与核纤层关联,但不同类型的AA、BB和AB对之间的差异具有统计学意义($P < 10^{-10}$)。图16B显示了与核纤层关联的基因座在转录状态和沉默状态之间的局部跨染色体A/B密度比的变化倍数。对于每个基因组基因座,计算基因座转录状态和沉默状态之间局部跨染色体A/B密度比的变化倍数,仅考虑其中该基因座与核纤层关联的细胞。确定每个基因座的每个状态(转录或沉默)的中值A/B密度比,两个状态之间的变

化倍数显示在左侧(每个圆圈对应一个基因座)。省略了异常值(所呈现的标尺以上的33个基因座和以下的18个基因座),以允许更清晰地可视化中值变化倍数。无论核纤层关联状态如何,从所有数据得出的变化倍数都显示在右侧以进行比较。虚线表示没有倍数变化,实线表示每种情况下所有基因座的中值变化倍数。

[0232] 结果表明,跨染色体相互作用优先发生在活性染色质基因座之间,并且在多个不同的核环境中始终观察到这种行为。接下来探索的是是什么可能导致这种优先的、广泛的活性-活性染色质相互作用。由于RNA聚合酶II(Pol II)包含低复杂性结构域(LCD)并且可以形成凝聚物,因此通过使用转录抑制药物 α -鹅膏蕈素(其导致Pol II解离和降解)来测试Pol II的转录是否可以导致这些优先的跨染色体相互作用。尽管消除了转录并改变了核结构及其与染色质的关联(图13A-13C),但用 α -鹅膏蕈素处理并没有显著降低活性染色质之间的优先跨染色体相互作用(图4C-4D),这表明额外的或其他活性染色质结合因子参与了这些跨染色体相互作用。结果表明,与活性染色质关联的多种其他蛋白质含有可能潜在介导凝聚物形成的LCD。因此,目的是通过使用1,6-己二醇(一种已知破坏LCD之间的疏水相互作用的药物)来更普遍地干扰凝聚物的形成。值得注意的是,在用2%1,6-己二醇处理细胞45分钟后,跨染色体接触中活性染色质相互作用的优先富集在很大程度上被消除(图4E-4F),这表明在建立或维持这些相互作用过程中凝聚物形成的潜在作用。

[0233] 总之,这些实施例开发了用于在单细胞中在基因组规模上对染色质的3D组织进行成像的大规模多重FISH方法,并进一步证明了通过将染色质和新生转录物成像(两者都在基因组规模上)与核结构鉴定组合将3D基因组组织置于其天然结构和功能环境中的能力。这提供了单个细胞中核组织的综合视图。虽然此处在所有染色体中一致地选择靶基因座以提供整体3D基因组组织的无偏视图,但该方法也可用于靶向具有特定结构和功能特性的基因组基因座,例如启动子、增强子和特定核结构蛋白结合的基因座,以研究这些基因座之间的相互作用及其与转录和其他染色质功能的关系。这种方法在与基因组组织相关的广泛问题中的广泛应用可以阐明控制染色质组织的机制和染色质结构在调节基因组功能中的作用。

[0234] 实施例5

[0235] 此实施例说明了可用于上述实施例的各种材料和方法。

[0236] 靶基因组区域。对于染色质成像,以下列方式选择基因组基因座进行成像。对于每个人类常染色体和X染色体,每约3Mb间距选择一个30kb片段。如果该间距导致给定染色体上的选定基因座少于30个,则该染色体的间距减小,直到所有染色体都至少选择了30个基因座。这导致总共1,041个靶基因组基因座用于成像,单个染色体中的基因座数量范围为30-80。然后为每个30kb片段设计编码探针(约400个寡核苷酸探针),用于组合FISH成像。

[0237] 对于新生RNA转录物的成像,选择了与靶基因组基因座完全或部分重叠的所有包含内含子的基因。然后为所有这些RNA的内含子设计编码探针,使每个RNA具有约20个编码探针,并且编码探针的靶向序列尽可能靠近转录起始位点(TSS)。靶向了总共1,137个基因。

[0238] 用于组合FISH成像的条形码设计。以下列方式选择用于对1,041个基因组基因座进行成像的二进制条形码。首先,生成所有可能的Hamming权重为2的100位二进制条形码(即每个条形码包含两个“1”位和98个“0”位),并从该列表中随机选择1,041个条形码。然后首先将选定的条形码任意分配到1041个基因组基因座。接下来,在使用和未使用的代码库

之间以及来自不同染色体的基因座之间随机交换条形码,以最小化每个染色体在不同位中出现(即读取“1”)的基因座数量的差异。这导致每个染色体的每个位成像的基因座数量大致相等。为了优化条形码与每个染色体内基因座的关联,同一染色体内的基因座被允许交换条形码并针对具有在相同的代码位置读取“1”的条形码的基因座之间的最大最小基因组距离进行优化。当比较具有相同最小基因组距离的代码分配时,选择最小化基因组距离的变异系数的代码分配(这样基因组距离具有更大的平均值和更小的标准偏差)。

[0239] 类似地选择了用于对1,137个基因的新生RNA转录物进行成像的条形码,但使用54位Hamming距离2代码而不是100位Hamming距离2代码。

[0240] 编码探针设计。用于染色质成像的编码探针是从购买自Twist Biosciences的寡核苷酸库合成的。该库中的每个寡核苷酸使用以下子序列(从5'到3'):

[0241] 1. 用于PCR扩增和逆转录(RT)的20个核苷酸(nt)正向引发区域

[0242] 2. 20-nt读出序列,对应于其中将对探针所靶向的基因组基因座进行成像的位之一

[0243] 3. 40-nt靶序列,设计为与单个靶基因组基因座唯一结合

[0244] 4. 上述20-nt读出序列的附加拷贝

[0245] 5. 用于PCR扩增的20-nt反向启动序列

[0246] 从先前生成的针对PCR优化的随机20-nt序列列表中选择正向和反向引发序列。

[0247] 通过以下过程选择读出序列。首先,创建了与人类基因组具有最小同源性的30-nt序列列表。然后,根据观察到的信噪比(SNR)对这些序列的子集进行排序,并选择前100个作为DNA读出探针。最后,通过反向互补这些序列各自的最后20nt来选择读出序列。

[0248] 类似地选择了40-nt靶序列。简而言之,对每个感兴趣的基因组区域重复以下程序(参见上面的“靶基因组区域”讨论)。首先,创建了与感兴趣的基因组区域互补的所有40-nt序列的列表(从靶区域中的每个可能碱基开始)。然后,通过要求它们处于定义的熔解温度和GC含量范围内来过滤序列。然后通过限制与人类基因组、人类转录组和包含重复序列的数据库的允许同源程度来进一步过滤剩余的序列。同源性是通过创建所有可能的17-nt序列及其在靶数据库(例如人类基因组、人类转录组)中出现的次数的表,并计算其与给定候选序列的精确17-nt匹配的总数来确定的。最后,在最后的过滤步骤之后从剩余的序列中选择靶序列,使得任何一对靶序列之间不存在基因组重叠。

[0249] 为了生成全长探针,为每个靶基因组基因座选择的每个40-nt靶序列被交替分配到跨越整个靶基因座的2个组。这些组中的每一个都与单个读出序列相关联,对应于将在其中对基因座进行成像的两个位中的一个。然后,将每个靶序列连接到分配给其组的读出序列的两个相同拷贝,然后连接到正向和反向PCR引物。

[0250] 用于RNA成像的探针类似地设计,除了它们在每个探针上包含相同读出序列的3个拷贝,一个在靶区域的5'端,两个在靶区域的3'端。用于RNA成像的读出序列与用于DNA成像的那些是正交的,并且是从相同排序的测试读出序列列表中选择。

[0251] 编码探针合成。从上述模板库扩增编码探针(参见上文“编码探针设计”)。这是使用涉及以下步骤的扩增方案完成的:

[0252] 1. 最初的寡核苷酸库使用有限循环PCR扩增约20个循环。此步骤中使用的反向引物还通过引物延伸引入了T7启动子序列。

[0253] 2. 所得产物通过柱纯化进行纯化, 并通过高产率体外转录反应进行进一步扩增和转化为RNA。

[0254] 3. RNA产物通过逆转录反应转化为单链DNA。

[0255] 4. 将上一步的产物进行碱解(以去除残留的RNA和引物DNA)并进行柱纯化。

[0256] 5. 如有必要, 前一步的产物在真空中干燥并重新悬浮在水中, 以达到所需的初级探针浓度。

[0257] 所有引物均购自Integrated DNA Technologies (IDT)。

[0258] 细胞培养和编码探针杂交。细胞制备如下进行。IMR-90细胞购自美国典型培养物保藏中心(ATCC, CCL-186), 并根据推荐的方案进行培养。为避免潜在的染色质结构改变, 本研究中的所有细胞均在培养开始后6周内以下述密度接种。

[0259] 为了准备DNA成像, 将细胞铺在40毫米、圆形#1.5盖玻片(Bioprotechs, 0420-0323-2)上, 每个盖玻片的密度为约500,000个细胞。让细胞在37℃和5%CO₂条件下生长约2天直到汇合。在转录抑制实验中, 在细胞固定前6小时, 将细胞培养基更换为含有100微克/mLα-鹅膏蕈素(Sigma-Aldrich, A2263)的新鲜培养基。对于使用1,6-己二醇(Sigma-Aldrich, 240117)的实验, 在细胞铺板前用10微克/mL纤连蛋白(Sigma-Aldrich, F1141)包被盖玻片, 并用含有2%w/v的1,6-己二醇的新鲜培养基替换培养基, 持续45分钟。然后将培养物使用在PBS中的4%多聚甲醛(PFA)在室温下固定10分钟, 并在PBS中洗涤2-3次。然后在两个步骤中使细胞透化: 首先, 在室温下用PBS中的0.5%v/v Triton-X(Sigma-Aldrich, T8787)处理细胞10分钟。然后, 在室温下用0.1M盐酸(HCl)处理细胞5分钟并在PBS中洗涤2-3次。在HCl处理后, 细胞用溶解在PBS中的0.1mg/mL核糖核酸酶A(ThermoFisher, EN0531)的溶液在37℃下处理30-45分钟, 以去除与RNA脱靶结合的潜在来源。在此处理之后, 将细胞在预杂交缓冲液中孵育约10分钟, 该缓冲液由2x盐水-柠檬酸钠缓冲液(SSC; Ambion, AM9763)和50%甲酰胺(Ambion, AM9342)组成。接下来, 将细胞盖玻片倒置并置于60毫米培养皿中的一滴50微升杂交缓冲液(2xSSC、50%甲酰胺、10%硫酸葡聚糖(Sigma-Aldrich, D8906), 含有具有或不具有10微克人Cot-1 DNA(ThermoFisher, 15279011)的总浓度为约25微摩尔的编码探针混合物)上。将培养皿部分浸入约90℃的水浴中3分钟, 然后在47℃的加湿室中孵育16-36小时。与编码探针孵育后, 将样品在2xSSC中洗涤并在40%甲酰胺中30分钟, 然后在室温下用2xSSC中的4%PFA进行固定后处理10分钟。然后将样品与基准珠(ThermoFisher F8805或ThermoFisher F8792)在2xSSC中孵育2-3分钟, 并在2xSSC中用1微摩尔4',6-二脒基-2-苯基吲哚(DAPI; ThermoFisher D1306)染色5-10分钟, 然后存储在2xSSC中直到成像。

[0260] 对于包括RNA成像的实验, 从固定细胞开始使用的所有缓冲液都含有1:10-1:1,000稀释的核糖核酸酶抑制剂(NEB M0314或Fisher Scientific N2615)。RNA染色的处理与上述方案相同, 直到用HCl处理。在此步骤之后, 将细胞在预杂交缓冲液中孵育10分钟, 然后将细胞盖玻片倒置并置于一滴杂交缓冲液中, 该缓冲液含有约1微摩尔总浓度的靶向RNA内含子的编码探针, 如针对DNA染色所述的。然而, 在这种情况下, 没有进行90℃的热变性, 细胞立即在47℃的加湿室中孵育16-36小时。在与编码探针一起孵育后, 将样品在甲酰胺溶液中洗涤, 并如上文针对DNA所述用PFA进行固定后处理。然后将其与基准珠一起孵育并用1微摩尔DAPI染色, 然后储存在2xSSC中直到成像。RNA成像后, 将样品从显微镜中取出, 用核糖核酸酶A处理细胞, 然后DNA杂交以与上文对没有RNA成像的DNA成像所述相同的方式进行。

[0261] 用于FISH成像的读出探针的依次杂交。方案的这一部分中的所有流体交换都是通过使用定制的流体系统实现的,其中盖玻片安装在FCS2流动室(Bioptechs,060319-2)中。流体系统使用3-4个计算机控制的八通阀(Hamilton,MVP和HVXM8-5)和一个计算机控制的蠕动泵(Gilson,MINIPLUS3)。放在一起,这些组件允许在任何给定时间控制流体流动的速率和流动的流体类型。

[0262] 每个轮次的杂交使用以下步骤:

[0263] 1.如下文所述的,在具有一组特定于每个轮次的寡核苷酸探针的杂交缓冲液中流动

[0264] 2.室温孵育10分钟

[0265] 3.流动洗涤缓冲液

[0266] 4.孵育约200秒

[0267] 5.流动成像缓冲液

[0268] 制备成像缓冲液,其包括60mM Tris pH8.0、10%w/v葡萄糖、1%葡萄糖氧化酶除氧剂溶液(含有~100mg/mL葡萄糖氧化酶(Sigma-Aldrich,G2133)和1:3稀释的过氧化氢酶(Sigma-Aldrich,C3155))、0.5mg/mL 6-羟基-2,5,7,8-四甲基色烷-2-羧酸(Trolox; Sigma-Aldrich,238813)和50微摩尔Trolox Quinone(通过Trolox溶液的UV辐照产生)。将Trolox溶解在甲醇中,然后添加到溶液中。制备后,成像缓冲液被一层约0.5厘米厚的矿物油覆盖,以防止暴露于氧气中。

[0269] 杂交缓冲液和洗涤缓冲液分别由2xSSC中的35%和30%甲酰胺组成,杂交缓冲液还含有0.01%v/v Triton-X。杂交缓冲液在每个杂交轮次中单独保存,并包含两或三组(分别用于DNA和RNA成像)读出探针。荧光信号以两种方式之一引入:

[0270] 1.对于DNA成像,每个轮次的杂交缓冲液包括两种荧光读出探针,一种用Cy5或Alexa647标记,另一种用Alexa750标记。荧光读出探针使用:1)荧光标记的寡核苷酸,其与在给定位成像的所有编码探针共有的读出序列互补,以100nM浓度添加,或2)具有与读出序列互补的序列、连接到所有衔接子共有(更准确地说,每个颜色通道中的所有衔接子共有)并且与所有其他使用的读出序列正交的附加读出序列(称为次级读出序列)的衔接子寡核苷酸和与这个次级读出序列互补的荧光标记的寡核苷酸探针的组合。衔接子和次级读出探针以1:1.5的比例预混合,并添加到最终浓度为~100nM。对于一些实验,衔接子和读出探针依次与样品杂交。这允许使用较低浓度的更昂贵的次级读出探针。

[0271] 2.对于RNA成像,每个轮次的杂交缓冲液包含三个衔接子寡核苷酸(将在三个不同的颜色通道中检测),每个与不同的读出序列结合,并且每个包含附加的次级读出序列。对应于相同颜色通道的所有衔接子共有相同的次级读出序列。每个轮次包括两个离散的杂交步骤:首先将衔接子流入、杂交和洗涤多余的材料。然后依次流入与衔接子上的次级读出序列互补的三种荧光读出探针(分别用Cy3、Cy5和Alexa750标记)。用于RNA成像的荧光读出探针包含将荧光团与次级寡核苷酸连接的二硫键,以允许在轮次之间有效去除信号。在荧光读出杂交后,流入成像缓冲液并收集信号。

[0272] 在下一轮的读出探针或衔接子探针杂交之前,来自当前轮次中读出探针或次级读出探针的荧光信号被去除,如下文“杂交轮次之间的信号去除”中所述。

[0273] 在第一轮杂交之前,进行一轮成像以获取DAPI信号并鉴定核边界。然后,在50轮杂

交和每个轮次2个颜色通道中对整组1,041个基因组基因座进行成像。在每个轮次中,基因组基因座通过在z维度上步进以3D形式成像。1,137个基因的新生RNA转录物同样以3D形式在18轮中以3种颜色进行成像。额外的轮次用于重新标记基因组基因座集并评估颜色通道之间的色差和渗色。对包含总共约1,000-2,000个细胞的约60个视野进行成像需要约3天。

[0274] 3-4阀系统允许加载多达20-30种不同的杂交溶液。结果,在耗尽所有流体系统的通道后,样品室被绕过,并且所有用于杂交的通道都用30%的甲酰胺水溶液洗涤。接下来,重新连接腔室并进行下一组杂交和成像轮次。

[0275] 抗体标记和成像。在RNA或DNA成像后立即进行抗体成像。通过上述方案完成成像后,样品经历了以下步骤:

[0276] 1.将样品与封闭溶液(含0.1%v/v Tween-20(Sigma-Aldrich P9416)和1%w/v牛血清白蛋白(BSA;Jackson ImmunoResearch 001-000-162)的PBS)孵育30分钟。

[0277] 2.将样品与在封闭液中稀释的一抗孵育1小时

[0278] 3.样品在含0.1%Tween-20的PBS中洗涤3次,每次5分钟

[0279] 4.对荧光标记的二抗重复步骤2和3

[0280] 使用上述微流体系统在显微镜上完成所有缓冲液交换。Cy5颜色通道用于成像并使用光漂白消除连续抗体标记之间的信号。

[0281] 使用了以下几组一抗和二抗:

[0282] 1.对于核斑点成像,使用以1:200的比例从原液中稀释的针对SC35(Abcam, ab11826)(一种通常用作核斑点标记物的剪接因子)的一抗,和使用以1:1,000的比例从原液浓度中稀释的Cy5染料标记的驴抗小鼠二抗(Jackson ImmunoResearch,715-175-150)。

[0283] 2.对于核仁成像,使用以1:200的比例从原液中稀释的抗纤维蛋白抗体(Abcam, ab5821),以及以1:1,000的比例从原液浓度中稀释的用Alexa 657染料标记的驴抗兔二抗(Jackson ImmunoResearch,711-605-152)。

[0284] 杂交轮次之间的信号去除。在每个轮次的成像之前,来自前一轮(或内源性背景,在第一轮的情况下)的信号被熄灭。这是通过信号的光漂白来实现的。通过将缓冲液更改为2xSSC并使用647和750激光(以及成像RNA时的560激光)的最大可用功率照明每个视野10秒来进行光漂白。在RNA成像实验中,用于漂白的缓冲液还含有50mM三(2-羧乙基)膦(TCEP;Sigma-Aldrich,C4706),以裂解连接荧光团和读出探针的二硫键。由于杂交和洗涤缓冲液中的高甲酰胺浓度,DAPI信号消失。

[0285] 图像采集。使用定制的显微镜系统进行图像采集。该系统围绕带有1.4NA的Nikon CFI Plan Apo Lambda 60x油浸物镜的Nikon Ti-U显微镜主体构建。照明基于以下两种选择之一:

[0286] 1.具有以下波长的固态单模激光器:405nm(Coherent,Obis 405nm LX 200mW)、560nm(MPB Communications,2RU-VFL-P-2000-560-B1R)、647nm(MPB Communication,2RU-VFL-P-1500-647-B1R)和750nm(MPB Communication,2RU-VFL-P-500-750-B1R)。在这种情况下,560-nm、647-nm和750-nm激光器的输出由声光可调谐过滤器(AOTF)控制,而405-nm激光器通过其激光控制箱直接控制。使用定制的二向色过滤器(Chroma,zy405/488/561/647/752RP-UF1)和发射过滤器(Chroma,ZET405/488/461/647-656/752m)来分离激发和发射照明。

[0287] 2.Lumencor CELESTA光引擎(基于光纤耦合的固态激光的照明系统),具有以下波长:405nm、446nm、477nm、520nm、546nm、638nm和749nm。该系统与五带通二向色过滤器(IDEX,FF421/491/567/659/776-Di01-25x36)和五带通过滤器(IDEX,FF01-441/511/593/684/817-25)一起使用。

[0288] 使用科学CMOS相机(针对单分子成像进行了工厂校准的Hamamatsu FLASH4.0或Hamamatsu C13440)进行图像采集。使用XYZ载物台(Ludl)控制三个维度的样品位置。定制的自动对焦系统用于长时间保持恒定的焦平面。这是通过比较从玻璃-流体界面反射并在单独的CMOS相机(Thorlabs,uc480)上成像的两个IR激光器(Thorlabs,LP980-SF15)光束的相对位置来实现的。

[0289] 对于每个实验,选择约60个视野(FOV)进行成像,避开细胞稀疏的区域(我们通常确定每个FOV 10-50个细胞)。每个相机FOV有1,000x1,000像素(相机像素对应于成像平面中每个维度中的153nm),或2048x2048像素(相机像素对应于成像平面中每个维度中的108nm)。

[0290] 在每个轮次的杂交后(参见上文“用于FISH成像的读出探针的依次杂交”),以3或4种颜色获取每个FOV的z堆栈图像:647nm和750nm照明(或在组合DNA和RNA成像的情况下用于RNA成像的560nm、647nm和750nm照明)用于获取FISH图像,使用560nm照明(或在组合RNA和DNA成像的情况下405nm照明)对基准珠进行成像。对于第一轮成像,使用405nm照明对DAPI信号进行成像,而对于抗体成像,在RNA或DNA成像后使用647nm激发通道。连续的z切片相隔85、100或150nm,覆盖所有成像细胞的整个核体积。在每个z位置,在移动载物台之前在所有通道中获取图像,并以~10Hz的速率获取图像。

[0291] DNA和RNA成像的图像分析和点拟合。以下分析管道应用于每个成像的FOV,以获得所有感兴趣基因座的三维(3D)位置:

[0292] 1.在所有轮次的成像中都拟合了基准点并用于图像对齐

[0293] 2.在第一轮成像中(第一轮杂交之前),DAPI信号用于鉴定个体细胞核的边界,以及用于RNA和DNA成像之间的图像配准

[0294] 3.将每个鉴定的核内的衍射限制点拟合到3D高斯函数,以鉴定它们的质心和高于局部背景的亮度

[0295] 4.使用自定义算法和软件(在“拟合的DNA点的解码算法”和“拟合的RNA点的解码算法”部分中有详细描述),将拟合的点与所有轮次的杂交中的相同核内的其他定位进行比较,以鉴定它们源自的基因座。

[0296] DNA和RNA成像的点拟合。使用3D高斯拟合来自个体FISH成像轮次的信号。为了使分析更易于管理,将保留用于解码的每个图像的拟合点数固定为125个(比预期无噪声的不同基因座的数量大约3倍)。

[0297] 漂移校正。以与上述相同的方式进行基准珠点拟合。然后在多个轮次的杂交之间比较一组基准珠位置,并应用刚性变换以最小化珠子相对位置的平方差之和。

[0298] 校正色彩效果。通过独立标记每个成像通道中的同一组基因组基因座并分别比较不同通道中相同基因座的信号来执行多色成像的渗色和色差。

[0299] 核分割。来自第一轮成像的DAPI图像用于鉴定个体细胞核的体积并允许进行细胞分割。这是通过构建和训练的卷积神经网络实现的,该网络将DAPI图像在xy平面上的最大

投影作为输入。

[0300] 附加图像分析:DNA和RNA成像之间的图像配准。在包括DNA和RNA成像的实验中,DAPI信号首先用于通过2D图像相关性在两组图像之间进行粗略的图像配准(达到相机像素精度)(每组中的所有图像都使用基准珠与DAPI图像对齐)。在执行了第一轮RNA解码后(参见下面的“拟合的RNA点的解码算法”),通过假设新生RNA定位与含有其的DNA基因座之间的位移在考虑视野中的所有成像的基因和细胞时应平均至零来计算更精细的比对。相应地,计算了额外的刚性变换以最小化成像的新生RNA与其相应DNA基因座之间的平均位移,并将其用作最终比对。

[0301] 从免疫荧光成像鉴定核体。通过对免疫荧光信号的强度应用阈值从免疫荧光信号提取核体(核斑点和核仁)的位置,从而产生鉴定高免疫荧光信号的像素化掩模。然后将其视为“包含”核体的一组像素化位置。

[0302] 拟合的DNA点的解码算法。每个基因组基因座的鉴定和3D定位是通过以下步骤实现的:

[0303] 1.为每个轮次的成像中所有鉴定的点的漂移和像差校正的位置生成一个列表。

[0304] 2.对于每个成像轮次中的每个检测到的点,发现来自其他轮次的所有点都在距其位置的设定截止距离(x 、 y 和 z 中约150nm)内。所有这些点对都被保留用于进一步分析,无论由点对产生的条形码(基于它们出现在哪个圆形和颜色通道中)是否对应于有效的条形码(分配给基因组基因座的条形码)。

[0305] 3.对于每对点,计算三个质量度量:

[0306] A.两个点的3D定位之间的位移

[0307] B.两个点之间的亮度差异

[0308] C.两个点的平均亮度

[0309] 每个点的亮度通过相应位中所有点的中值亮度进行标准化。

[0310] 4.然后基于点对是否对应于有效的条形码(因此可能对应于基因组基因座),将它们分成两组。在每组内,计算质量度量的分布。为方便起见,来自无效条形码的点对质量度量的分布称为“无效分布”,来自所有有效条形码的分布称为“有效分布”。

[0311] 5.对于每个点对,通过针对步骤4中的“有效分布”计算每个候选点对的组合Fisher p 值,将步骤3中的三个质量度量组合成一个度量。这被认为是每个点对的整体质量评分,并按以下方式对每个对进行计算:对于三个度量中的每一个,“有效分布”中其他点对的分数用较低质量度量计算,并将这些分数相乘。然后使用期望最大化程序依次选择与每个靶标染色质基因座相对应的具有最高质量评分的两个点对,并重新更新“有效分布”,并重复此优化程序直到收敛。收敛后,最后一组点对(各自对应于一个染色质基因座)用于确定基因座的3D空间位置。

[0312] 6.在步骤5之后,使用改进的K均值算法将属于同一染色体的染色质基因座分成两个同源物。与将点分成两组并最小化每组内的回转半径的标准K均值聚类算法相反,点在组之间逐步切换,以首先最大化每个同源物中分配的点的分数,然后最小化每个同源物的回转半径。

[0313] 7.分离两个同源物后,它们的质心和每个点对与步骤2的距离被计算为其亲本染色体的质心。除了步骤3中考虑的3个度量以外还添加了距染色体中心的距离作为另一个质

量度量和重复步骤3-6。

[0314] 8.最后,对来自步骤7的点対进行过滤,以去除质量评分与“无效分布”相似的点対。

[0315] 步骤8后剩余的点対用于确定染色质基因座的最终位置并追踪染色质结构。

[0316] 拟合的RNA点的解码算法。使用以下程序解码来自RNA成像轮次的信号:

[0317] 1.为每个轮次的成像中所有鉴定的点的漂移和像差校正的位置生成一个列表。

[0318] 2.对于每个成像轮次中的每个检测到的点,发现来自在距其位置的设定截止距离内的其他轮次的所有点,如果这些点対形成有效的条形码,则将它们保留为候选RNA爆发。

[0319] 3.在初始图像配准(基于DAPI图像)和漂移和像差校正后,将这些候选RNA爆发中的每一个的位置与含有相关基因的DNA基因座的位置进行比较,如果它们在设定的阈值距离内,则保留它们。

[0320] 4.基于初始解码的RNA定位(来自步骤3)与包含它们的DNA基因座的位置之间的位移来精细化DNA和RNA成像之间的配准,如上文“DNA和RNA成像之间的图像配准”部分所述。

[0321] 5.再次将所有候选RNA爆发的位置与包含它们解码的基因的DNA基因座的位置进行比较,这次使用经过精细化的图像配准。如果在这个阶段新生RNA定位在与含有其的DNA基因座的截止距离内,则它被认为是检测到的转录爆发。

[0322] 附加分析:核纤层的鉴定。通过生成围绕给定细胞中所有解码的染色质基因座位置的最小3D凸包(使用Python的scipy包)来估计核纤层的位置。

[0323] 空间距离。任何一对基因座之间的空间距离简单地计算为它们的拟合的3D高斯中心之间的欧几里得距离,乘以涉及相机像素和z步长与物理距离的适当比率。在距核体的距离的情况下,计算了距离所有已鉴定的核体“位置”的最小欧几里得距离或距离定义核纤层的凸包表面的最小距离。

[0324] 来自成像的接触频率矩阵。为了计算任何给定基因座对之间的接触频率,计算了该基因座对之间小于设定阈值的测量距离的数量。然后将该数字除以对于该对基因座测量的距离总数。

[0325] 局部密度分析。为了计算每个解码位置的A隔室和B隔室基因座的跨染色体局部密度,计算了每个细胞的每对染色质基因座之间的空间距离。对于每个基因座,按以下方式计算局部A/B密度比:

[0326] 1.通过评估两个基因座之间距离的具有500nm标准偏差(经调整以考虑细胞尺寸的变异性)的高斯函数值,从不同的染色体计算每个其他基因座的密度贡献。

[0327] 2.然后将基因座处的总A密度计算为所有跨染色体A基因座的该高斯函数值的总和,并以类似的方式计算总B密度。

[0328] 3.跨染色体A隔室基因座的总密度除以跨染色体B隔室基因座的密度,得到该基因座的A/B密度比。

[0329] 估计多重RNA成像中的检测效率。转录爆发事件的检测效率的估计以下列方式进行:

[0330] 1.考虑了包含其RNA内含子被成像的基因的所有靶向的基因组基因座。对于这些基因组基因座中的任何一个,如果基因被转录,它们相应的RNA信号应该出现在两个预定义的位中。知道这两个位中的每一个未被检测到的速率(p)允许推导RNA的检测效率。鉴定了

在其相应基因的两个预期的位中的至少一个中与RNA信号共定位(在约150nm内)的一组基因组基因座。

[0331] 2.从步骤1中鉴定的全部染色质基因座集合中,与RNA信号共定位的基因座的分数(f)是从其基因的对应位中的仅一个(而不是两个位)确定的。根据测量的f(8.4%),其应该

等于 $\frac{2p(1-p)}{1-p^2}$, 估计p(4.4%)。

[0332] 3.使用等式 $\eta = (1-p)^2$ 计算在两个位中检测共定位信号的总体检测效率,发现为约92%。

[0333] Hi-C数据分析。IMR-90细胞的Hi-C数据是使用吸管(straw)获取和加载的。为了鉴定个体染色体中的A/B隔室,遵循已建立的公开的方案。为了将源自成像数据的接触频率与Hi-C进行比较,创建了以靶区域为中心的箱,并通过将更高分辨率Hi-C数据中的读取数量相加来获得这些箱的Hi-C数据。

[0334] 实施例6

[0335] 染色质的三维(3D)组织调节许多基因组功能。然而,对3D基因组组织的理解由于缺乏在其天然环境中直接可视化染色质构象的工具而受到阻碍。本文报道的是用于以高基因组通量在单细胞中跨多个尺度可视化染色质组织的成像平台。首先,展示了通过依次杂交对数百个基因组基因座进行了多重成像,这允许整个染色体的高分辨率构象追踪。接下来,开发了用于基因组规模染色质追踪的组合成像方法,并证实了>1000个基因组基因座和>1000个基因的新生转录物以及标志性核结构的同时成像。使用该平台,对染色质结构域、隔室和跨染色体相互作用以及它们与单细胞转录的关系进行了表征。这种高通量、多尺度和多模态成像技术有广泛的应用,它提供了染色质组织在其天然结构和功能背景下的综合视图。

[0336] 基因组的3D组织调节从基因表达到DNA复制的许多基本细胞功能。生化和成像测量揭示了广泛尺度范围的复杂染色质结构。特别地,高通量染色体构象捕获方法(例如Hi-C和其他基于测序的方法)已经从全基因组的角度揭示了染色质结构,例如结构域和隔室。值得注意的是,染色质被划分为具有增强的自我相互作用的基因组区域,称为拓扑相关结构域(TAD),其在Hi-C接触图上显示为块状结构。这些TAD的尺寸从数百千碱基(kb)到几兆碱基(Mb)不等,通常含有共同调控的基因,并且具有与调控表观遗传元件一致的边界。在更大的尺度上,染色质分为两个主要的隔室,称为A和B隔室,其分别富含活性和非活性染色质,如Hi-C图中的交替“格子”图案所示,与之前的染色质的富含基因和缺乏基因的区段倾向于在空间上分离的基于成像的观察结果一致。最近的成像实验表明,A隔室和B隔室染色质确实倾向于在单个细胞中空间分离。A/B隔室化的生理意义与其在发育过程中和细胞类型之间的变化有关。

[0337] 总体而言,基于高通量测序的方法极大地丰富了3D基因组组织的知识。尽管如此,这些强大的方法也有局限性。例如,这些方法提供染色质基因座对的接触信息,但不提供这些基因座的直接空间位置信息。此外,关于染色质组织的大多数见解都建立在数百万个细胞的群体平均接触图上。尽管单细胞Hi-C方法不断改进,但单细胞中染色质接触的捕获效率和/或这些方法的细胞通量仍然相对较低,因此研究单细胞中的3D基因组组织仍然是一项具有挑战性的任务。此外,尽管已经出现了将Hi-C与其他测量方式相结合例如以在相互

作用蛋白质、核结构或DNA修饰的背景下提供染色质接触的表征的方法,但通过测序进行多模态测量仍然具有挑战性。值得注意的是,尚未出现允许在同一细胞中对染色质组织和转录活动进行基因组规模测量的方法,尽管需要这种方法来进一步了解染色质组织如何调节转录以及转录如何进而影响染色质组织。

[0338] 另一方面,基于成像的方法提供了对个体细胞中染色质基因座的空间位置的直接测量,具有很高的检测效率。特别是,荧光原位杂交 (FISH) 允许对固定的细胞中的染色质基因座进行高度特异性检测,最近,规律成簇间隔短回文重复 (CRISPR) 系统显著增强了对活细胞中特定染色质基因座成像的能力。染色质成像也可以与RNA和蛋白质成像相结合,以研究染色质组织与转录活动或相互作用蛋白质因子之间的相互作用。然而,目前的成像方法在基因组(序列)空间中的通量有限,传统上一次只允许研究几个不同的基因组基因座。最近开发的是通过连续几轮FISH成像的染色质追踪方法,每个轮次使用单色或双色成像靶向一个或两个基因组基因座。这种方法允许对单个细胞中数十个不同的染色质基因座进行成像,并已用于提供对染色质结构及其与转录的关系的见解。然而,由于可以在个体细胞中同时成像的基因组基因座数量仍然有限,因此仍然缺少单个细胞中整个染色体的高分辨率视图,更不用说个体细胞中染色质组织的基因组规模视图。

[0339] 本文报道的是一个多尺度、多重FISH成像平台,它允许在单个细胞中以不同的分辨率和基因组覆盖率同时成像数百到>1,000个不同的基因组基因座。首先,顺序成像方法得到了显著进步,以允许对数百个基因组基因座进行成像,并且该方法被用于提供整个染色体的高分辨率视图,阐明染色质结构域和隔室结构,它们之间的关系,以及单细胞中染色质组织与转录的关系。接下来,基于组合标记和成像开发了一种大规模多重FISH方法,该方法允许用更少的杂交轮次对更多的基因组基因座进行成像。使用这种方法,表明可以同时成像个体细胞中的>1,000个基因组基因座,以及同时成像这些基因组基因座以及位于这些基因座和标志性核结构(包括核斑点和核仁)中的>1,000个基因的新生RNA转录物,这允许将染色质组织置于其天然结构和功能环境中。该方法用于探索单细胞中跨染色体相互作用、转录活动和核结构之间的关系。

[0340] 为了允许跨多个尺度系统地观察染色质结构,开发了使用定制显微镜和流体装置(参见实施例19)的成像平台,用于在序列空间中以极高的通量直接可视化染色质,大至基因组规模。该平台包括两种互补的方法(图17A)。首先,对于相对较小的染色质结构的成像,使得其中包含的不同基因座在任何单个图像中都难以分辨,先前报道的顺序成像策略被扩展为允许追踪单个细胞中的数百个染色质基因座。在这种方法中,染色质在多个成像轮次中一次成像一个基因座(或一次2-3个基因座,使用2-3次彩色成像)(图17A,左)。这种方法通过使用它以高分辨率追踪单个细胞中整个染色体的构象来证明。其次,为了对分散在明显大于衍射极限分辨率的区域上的染色质结构(例如遍布整个细胞核的结构)进行成像,开发了一种更有效的组合策略,其中在每个轮次中同时对许多染色质基因座进行成像并且它们的不同身份是根据它们在其中出现的不同轮次组合确定的(图17A,右)。后一种方法允许在少得多的成像轮次中对大量基因组基因座进行成像。这种方法用于在单细胞的转录活动和重要核结构的背景下提供染色质组织的基因组规模视图。

[0341] 图17A-17M显示了通过单细胞中染色质结构域的依次杂交和表征追踪的高分辨率全染色体。图17A显示了多尺度染色质追踪平台的示意图。左图:通过依次杂交和成像对整

个染色体进行染色质追踪的示意图。当靶染色质结构与衍射极限分辨率相当或更小时,每个成像轮次在每个颜色通道中成像单个染色质基因座。在所有轮次成像之后,可以为靶向的染色体的每个拷贝生成3D染色质迹线。右图:通过组合FISH的基因组规模成像的示意图。当靶基因座预计散布在远大于衍射极限分辨率的空间中时,例如当基因座分散在整个细胞核中时,可以在每个轮次中对多个基因座进行成像和解析,并可以基于在其中检测基因座的成像轮次的组合从条形码导出每个基因座的身份。与顺序成像方法相比,这种方法显著减少了对相同数量的基因座进行成像所需的轮次数量。

[0342] 实施例7

[0343] 整个染色体的高分辨率染色质追踪。在本节中,描述了通过顺序成像方法进行的高分辨率全染色体追踪(图17A,左;图24A)。首先关注人类染色体21(Chr21),染色体的非重复部分(Chr21:10.4-46.7Mb)被分成>600个连续区段(即>600个基因组基因座),每个区段长度为50-kb。设计了初级寡核苷酸探针的文库,每个探针包含用于与染色体杂交的可变靶序列和每个50-kb基因座独有的读出序列(图24A)。与特定50-kb基因座结合的所有初级探针共有相同的读出序列,因此读出序列可用于通过用荧光染料标记的互补读出探针的杂交来鉴定每个基因座(图24A)。然而,由于染料标记的寡核苷酸的高成本,使用>600个不同的荧光读出探针来鉴定这些基因组基因座将非常昂贵。为了克服这一挑战,设计了一种两步标记策略,以使用三种染料标记的寡核苷酸探针的共同组(称为读出探针,每个颜色通道一种读出探针)检测不同的读出序列,由将每个基因座特异性读出序列转换为三种常见读出序列之一的未标记的衔接子探针介导(图24A)。使用这种策略,人肺成纤维细胞(IMR-90)中Chr21中>600个染色质基因座被依次成像,使用衔接子和读出探针的>200个轮次的杂交,在每个轮次中使用三色成像。为了在如此大量的杂交轮次中实现稳定成像,成像方案以下列方式进一步优化:(i)以维持样品完整性和初级探针结合稳定性,样品在成像过程中用甲醛定期重新固定;(ii)以确保在每个成像轮次后完全去除荧光信号,并最小化数百个标记轮次中残留信号的积累,使用组合的化学裂解和光漂白方法去除读出探针的荧光信号,并在每个成像轮次后添加未标记的读出探针以阻止衔接子探针上的任何未占用的结合位点;(iii)以最小化干扰并优化了实验时间、每个杂交轮次的持续时间和流体系统的流速。

[0344] 成像后,在3D中确定每个染色质基因座的质心位置,并重建每个细胞中Chr21的每个同源拷贝的构象(图17B)。为了估计样品和成像仪器在多个杂交轮次中的稳定性,在不同数量的杂交/成像轮次后重新成像相同的染色质基因座,并将原始基因座位置与相应重新成像的基因座位置之间的位移用作测量精度的测量。原始和重新成像的基因座之间的中位位移从当11个杂交轮次将两个成像实例分开时的约70nm增加到当初始和重新成像实例被约250个杂交轮次分开时的约120nm(图24B-24C),其中在重新成像时显示较大位移的基因座也具有较低的荧光信号强度(图24D)。值得注意的是,即使在>250轮杂交之后,中位位移误差也明显小于相邻染色质基因座之间的中位距离(~250nm)(图24B)。此外,成像的基因座之间的中值成对距离在生物学重复之间具有高度可重复性(图24E)。这些实验中染色质基因座的检测效率>90%(即在每个染色体中检测到>90%的靶染色质基因座)。

[0345] 为了获得Chr21的染色质构象的群体平均视图,成像的基因座之间的成对相互作用通过计算它们的中位空间距离和基因座在约3,500个成像的细胞中接近的概率来量化(图17C和图24F-24I)。观察到来自成像数据的中值成对距离与先前发表的跨Chr21中存在

的所有长度尺度的Hi-C数据之间的高度相关性 (Pearson相关性为0.89;图24F-24G),对于较短的基因组距离,一致性特别高 (Pearson相关系数为0.97;图24H-24I)。为了选择低于其就认为两个基因座接近的截止距离,计算了跨一系列截止距离的Hi-C数据和源自成像数据的邻近频率(即其中两个基因座接近的实例的分数)之间的Pearson相关系数。Pearson相关系数在很宽的截止距离范围内保持较高,但在截止距离为约400-500nm时达到峰值0.88(图24J)。因此,选择500nm作为截止距离以在整个工作中生成邻近频率图(对于截止距离的选择的更详细原理,参见实施例19)。

[0346] 中值距离和邻近频率图均显示块状TAD结构(图17C和图24F和24H)。从成像数据的距离和邻近频率图鉴定的TAD边界与从整体Hi-C数据确定的那些高度相似(图24K)。此外,证实染色质迹线中的基因座定位误差($\sim 100\text{nm}$)对结构域边界鉴定及其准确性影响不大(图24K)。

[0347] 图17B显示了通过依次杂交方法成像的单个IMR90细胞中Chr21的两个拷贝的3D结构渲染和空间距离矩阵。左图:单个细胞中Chr21的两个拷贝覆盖在细胞核的DAPI图像上。比例尺:5微米。右上图:根据它们沿染色体的基因组坐标的两个Chr21拷贝中所有检测到的染色质基因座(球体)的3D渲染(基因组位置显示在右侧)。柔性线连接相邻的染色质基因座。比例尺:1微米。右下图:与上面显示的染色体拷贝相对应的成对空间距离矩阵(没有对无适当参考基因组或包含高度重复序列的基因组区域进行成像)。

[0348] 图17C显示了Chr21的整体邻近频率矩阵和CTCF/RAD21结合位点的单细胞结构域边界的优先定位。上图:来源于成像数据的Chr21的邻近频率矩阵。每个矩阵元素定义为一对基因座之间的测量距离小于500nm的截止距离的频率。中间图:Chr21的10Mb部分的邻近频率矩阵的放大版本。下图:在每个成像的50-kb基因组区段上单细胞结构域边界形成的概率。三角形显示CTCF和RAD21 ChIP-seq峰。

[0349] 图24A-24N显示了通过依次杂交追踪的高分辨率全染色体,以及与Hi-C相比Chr21结构特征的整体统计学。图24A显示了用衔接子探针依次杂交的标记和成像方案。首先,样品与初级探针杂交,每个初级探针都包含允许特异性结合靶基因组基因座的靶序列和读出序列。每个基因座由总共350-500个初级探针标记,但只显示了一个。每个靶基因组基因座都被分配了独特的读出序列(以各种颜色显示),其对所有与该基因座结合的初级探针都是共有的。然后使用连续轮次的杂交检测读出序列。在每个轮次的杂交的过程中,对应于靶基因座的读出序列(一个用于三个颜色通道Alexa750、Alexa647和Cy3中的每一个)用寡核苷酸衔接子探针标记,每个探针由两部分组成:与基因座特异性的读出序列互补的区段和包含颜色通道特异性的共有读出序列的区段。每个颜色通道包含独特的共有读出序列,其由在同一颜色通道中可视化的所有衔接子共有。然后将共同的读出序列与相应颜色通道中的染料缀合的互补读出探针杂交。该程序允许在每个轮次的杂交期间在三个颜色通道中对三个基因组基因座进行成像。在每个轮次成像之后,通过二硫键连接到读出探针的荧光染料被TCEP从共有读出探针上裂解下来,并且衔接子上任何未占用的读出序列都被未标记的共有读出探针封闭,以防止杂交轮次之间的串扰。该过程迭代数百个轮次,直到完成所有读出序列的检测,从而完成所有靶向的基因组基因座的检测。

[0350] 图24B显示了在单个实验过程中基因座的位移。在实验开始和结束时(间隔 >250 轮杂交)对Chr21(chr21:32.45-33.35Mb)中900kb区域内的连续50kb区段进行成像。显示了重

新成像的点与其原始成像对应物之间的位移分布。为了比较,显示了在原始成像轮次中测量的同一900kb区域中相邻50kb区段之间的距离分布。

[0351] 图24C显示了由不同数量的杂交轮次分开的原始轮次和重新成像轮次之间染色质基因座位移的箱线图。显示了中位数(中心线)、第25-75个百分位数(方框)和第10-90个百分位数(须线)。

[0352] 图24D显示了具有在间隔>250轮杂交的原始和重新成像实验之间的低(<500nm)和高(>500nm)位移误差的染色质基因座的荧光信号的箱线图。显示了中位数(中心线)、第25-75个百分位数(方框)和第10-90个百分位数(须线)。

[0353] 图24E显示了两个重复实验之间的中位基因座间空间距离的比较。对于Chr21的两个生物复制实验,分别计算了成像的染色质基因座对之间的中位空间距离,并针对彼此作图。在两次重复中测量的数据之间的Pearson相关系数为 $\rho=0.98$ 。

[0354] 图24F显示了Chr21的从成像导出的中值空间距离矩阵(左)、从成像导出的邻近频率矩阵(中)和整体Hi-C接触矩阵(右)的比较。对于成像数据,当两个基因座之间的空间距离小于500nm的截止距离时,两个染色质基因座被认为是接近的。Hi-C接触矩阵以50kb分箱并以靶区域为中心。

[0355] 图24G显示了从整体Hi-C导出的接触数量和从染色质基因座的个体对的成像导出的中值成对距离的重对数散点图。线表示数据的线性回归(斜率=-4.43)。成像和Hi-C数据之间的Pearson相关系数为 $\rho=0.89$ 。

[0356] 图24H与图24F相同,但针对Chr21中的3-Mb区域(chr21:30.30-33.38Mb)。TAD边界用线条标记。

[0357] 图24I与图24H相同,但针对(H)中所示的3-Mb区域。斜率=-4.51。 $\rho=0.97$ 。

[0358] 图24J显示了Hi-C接触图和使用不同截止距离生成的成像衍生的邻近频率图之间的Pearson相关性。为了生成邻近频率图,选择截止距离,并且距离小于该截止值的两个基因座被认为是接近的。然后,将一对基因座之间的邻近频率计算为其中基因座之间的测量距离小于截止距离的发生次数除以两个基因座之间的测量距离的总数。

[0359] 图24K显示了作为Chr21上基因组位置的函数的标准化绝缘评分,其如下计算:1)来自成像的中值成对距离(上图),2)来自成像的邻近频率(中间图),和3)Hi-C接触读数(下图)。为了计算绝缘评分,首先选择上游的固定长度(250kb)基因组区段和感兴趣的位置下游的相同长度区段。标准化绝缘评分然后定义为中值区段间成对距离和中值区段内成对距离之间的差异,通过这两个中值距离的和进行标准化。TAD边界定义为沿染色体的标准化绝缘评分的局部最大值,由标准峰值调用算法鉴定(参见实施例19)。垂直虚线是从Hi-C数据调用的整体TAD边界。在上图和下图中还显示了在用标准偏差为100nm的3D高斯噪声项扰乱基因座位置后的中值距离(黑线,上图)和邻近频率(黑线,中间图),与估计的定位测量误差相当。

[0360] 图24L显示了两个示例单细胞中Chr21(chr21:28.2-38.1Mb)的10-Mb区域中的染色质结构域。显示了单个细胞中Chr21的两个单独拷贝的成对距离(上图,中间图)以及来自所有成像的细胞的人群中值成对距离(下图)。

[0361] 实施例8

[0362] 单个染色体中的染色质结构域。在单细胞水平,观察到染色体被划分为在单细胞

空间距离矩阵中表现为块状特征的结构域(图24L)。这些结构域和基因座间距离显示出细胞与细胞之间的高变异性(图24L-24M),这与在单细胞Hi-C数据中观察到的染色质接触中的显著细胞间变异一致。当以相似的分辨率对染色体的小($\sim 2\text{Mb}$)区域进行成像时,之前观察到单细胞中的类似结构域结构。然而,在这些先前测量的小区域中,相当一部分细胞没有显示出清晰的单细胞结构域边界,并且仍然不确定那些细胞内是否没有形成结构域,或者整个成像区域是否在单个结构域内。此外,由于这些先前成像区域的尺寸很小,许多结构域在成像的基因组区域的末端被人为截断,因此无法准确表征某些基本结构域特性,例如它们的物理和基因组尺寸。本研究中的高基因组通量提供了这些单细胞结构域结构的全染色体视图,揭示了它们在基本上所有成像的细胞中的整个染色体中的显著存在,从而允许以更系统的方式表征它们的特性。

[0363] 首先鉴定这些单细胞结构域边界的基因组位置,并量化每个50-kb基因组基因座的边界形成的概率。虽然在所有成像的基因组基因座处观察到边界形成的非零概率,但结构域边界优先位于CTCF和黏连蛋白的结合位点附近(图17C-17D)。

[0364] 除了结构域边界位置中的细胞间变异外,还观察到这些单细胞结构域的其他特征中的显著异质性,从结构域的物理尺寸到结构域之间的绝缘或相互作用程度(图17E-17H)。具体而言,观察到单细胞结构域在其基因组尺寸(图17I)和通过回转半径测量的物理尺寸(图17E和17J)中都是可变的。基因组尺寸的分布和这些结构域的物理尺寸的分布都对约100nm的估计的基因座定位误差不敏感(图17I-17J)。值得注意的是,由相同基因组区域界定的结构域或具有相同基因组尺寸的结构域在其物理尺寸上在细胞与细胞之间波动很大(图17E和24N)。有趣的是,由相互作用的CTCF/黏连蛋白结合位点界定的结构域在物理尺寸上往往比不受这种基因组基因座界定的结构域更小(图17K)。此外,相邻结构域之间的物理分离程度也有很大差异(图17F和17L),其中一些相邻结构域完全分离并且仅通过接头区域连接,而其他显示部分重叠和不太清晰的边界(图17F)。此外,即使是与其相邻结构域完全分离的结构域也可能在空间上与由小或大的基因组距离分开的非相邻结构域部分重叠(图17G)。最后,观察到在这些单细胞结构域末端的两个染色质基因座也表现出彼此之间的可变距离,并且与在结构域的内部由相似基因组距离分开的染色质基因座相比,没有表现出彼此更靠近的趋势,无论结构域是否由CTCF/黏连蛋白结合位点界定(图17H和17M)。

[0365] 图17D在以CTCF/Rad21结合位点或整体TAD边界为中心的基因组位置处单个细胞中结构域边界形成的平均概率(灰色)。

[0366] 图17E是在物理空间中占据较大(上图)或较小(下图)体积的具有相同基因组坐标的两个单细胞染色质结构域的实例。左图:染色质结构域的3D渲染,其中绿色球代表结构域内成像的基因组基因座,柔性接头连接基因组序列中的相邻基因座。灰色斑点代表染色体其余部分中成像的基因座。比例尺:1微米。右图:左侧显示染色质结构域的成对距离矩阵(用线标记),带有侧翼区域。

[0367] 图17F显示了具有高(上图)和低(下图)绝缘评分的两对染色质结构域的实例。左图:染色质结构域的3D渲染,如图17E所示。比例尺:250nm。右图:左侧显示染色质结构域的成对距离矩阵,其中渲染的结构域以相应的颜色标记。

[0368] 图17G显示了具有部分重叠体积的染色质结构域之间的远程接触的两个实例。左图:染色质结构域的3D渲染,如(E)所示。阴影代表不同的结构域。比例尺:250nm。右图:染色

质结构域的成对距离矩阵,其中渲染的结构域以相应的颜色标记。灰色空间表示22.85Mb的基因组距离中的间隔。

[0369] 图17H显示了CTCF结合位点两侧的染色质结构域的实例,显示了CTCF位点之间的小(上图)和大(下图)距离。左图:染色质结构域的3D渲染,如(E)所示,但CTCF位点的基因座位于结构域末端。比例尺:250nm。右图:染色质结构域的成对距离矩阵,对应地标记了结构域和边界CTCF。

[0370] 图17I显示了单细胞中Chr21中染色质结构域的测量的基因组尺寸的分布。黑线显示的是源自考虑到100nm的定位误差的模拟数据的单细胞中Chr21中染色质结构域的基因组尺寸的分布。在该模拟中,成像的基因座的位置受到标准偏差为100nm的3D高斯噪声的干扰,类似于我们的测量误差。

[0371] 图17J显示了单细胞中Chr21中染色质结构域的测量的物理尺寸(如由回转半径定义的)的分布。黑线显示的是源自考虑到100nm的定位误差的模拟数据的单细胞中Chr21中染色质结构域的物理尺寸的分布,如图17I所示。

[0372] 图17K显示了作为染色质结构域的基因组尺寸的函数的中值回转半径,其中边界基因座包含相互作用的CTCF/Rad21位点并且边界基因座都不包含CTCF/Rad21位点。误差线表示通过重采样得出的95%置信区间。

[0373] 图17L显示了相邻结构域之间的绝缘评分的分布,其中结构域边界出现在CTCF/Rad21结合位点和非CTCF/Rad21结合位点。

[0374] 图17M显示了作为染色质结构域的基因组尺寸的函数的结构域的标准化端到端距离的中值,其中边界基因座包含相互作用CTCF/Rad21位点并且边界基因座都不包含CTCF/Rad21位点。标准化的端到端距离定义为结构域的端到端距离除以由相同基因组距离分开但位于单个结构域内部中的相似基因座对之间的中值距离。误差线表示通过重采样得出的95%置信区间。

[0375] 图24M显示了Chr21的基因座间空间距离的标准偏差矩阵。对于每对区域,显示了所有单染色体拷贝中相应基因座对之间距离的标准偏差。

[0376] 图24N显示了通过回转半径测量的Chr21中不同基因组尺寸的染色质结构域的物理尺寸的箱线图。对于每个基因组尺寸,显示了中位数(中心线)、第25-75个百分位数(框)和第10-90个百分位数(须线)。

[0377] 实施例9

[0378] 单个染色体中的染色质隔室。接下来,全染色体的高分辨率视图用于检查A和B隔室中的染色质基因座如何在单个细胞中排列。首先,使用先前描述的算法,使用源自成像数据的Chr21的邻近频率图的Pearson相关矩阵的主成分分析(PCA)确定整体A/B隔室边界(图18A;图25A)。从成像数据获得的隔室边界与从先前公布的整体Hi-C数据确定的那些高度相似(图25A)。下面,从整体邻近频率图获得的隔室边界用于为个体细胞中的个体基因座分配A/B身份。

[0379] 与之前的研究相比分辨率的>10倍提高允许详细查看单个染色体中的A隔室基因座和B隔室基因座组织。在从细胞到细胞的个体染色体拷贝之间观察到A和B基因座排列的高度变异(图18B)。虽然在一些染色体中,A和B基因座被分离成基本上不重叠的空间区域,但其他染色体在A和B基因座之间表现出大量的空间重叠。有趣的是,同一染色体中的A隔室

基因座有时被分成多个“微隔室”(图18B)。

[0380] 为了量化个体染色体中A和B基因座的空间分离程度,设计了基于局部密度的方法,并且对于每个成像的基因座,计算了其他A和B基因座的局部密度(图25B)。正如预期的那样,A隔室基因座平均倾向于被A基因座包围,对于B基因座也是如此(图25C)。基于在包含大多数A或B基因座的空间体积中观察到的基因座的纯度,进一步定义了每个个体染色体的A/B分离评分(图18C)。预期A和B基因座的完全物理分离会导致1的分离评分,并且通过A和B基因座的彻底混合导致0.5的分离评分(参见实施例19)。观察到,对于细胞中的绝大多数Chr21拷贝,分离评分远高于以0.5为中心的随机对照的评分(通过沿基因组轴随机移动隔室边界同时保持隔室尺寸不变来获得)(图18C),表明A和B基因座在单个细胞中空间分离的趋势。还值得注意的是,A和B基因座的空间分离通常是不完整的(图18C)。这可能反映了活性和非活性染色质的不完全空间分离,但也可能部分是由表观遗传修饰中的细胞间变异性引起的,这可能使整体A/B隔室身份不完全代表单个细胞中活性/非活性染色质的描绘。值得注意的是,发现A/B分离的程度依赖于细胞周期:与G1期的细胞相比,G2/S期的细胞的A/B分离更强(图25D),这与之前的在细胞周期中逐渐建立A/B隔室的发现一致。

[0381] 值得注意的是,Chr21是最小的染色体之一,尺寸仅为约48Mb,并且仅划分为少量的连续A和B区域。为了将研究结果扩展到更大的染色体并研究它们的普遍性,对最大的染色体之一,染色体2(Chr2)进行了成像,该染色体沿其基因组序列在A和B隔室之间表现出大量的转换(约50个转换)。具体而言,通过沿其基因组序列以250-kb的间隔标记和成像50-kb区段来追踪Chr2。基于成像数据(图18D;图25E),使用上述相同的方法来调用染色体p和q臂中的A和B隔室,并观察到与从先前公布的整体Hi-C数据确定的A和B隔室的定量一致性(图25F-25G)。在单染色体水平,再次观察到A和B基因座的各种空间排列,从几乎完全的空间分离到A和B基因座之间的大量空间重叠(图18E)。有趣的是,一些染色体显示出“三明治”结构,其中A基因座位于两层B基因座之间,这可能是由于B基因座与靠近细胞核顶部和底部的核纤层的优先缔合。定量地,与随机对照相比,Chr2的个体拷贝中的A/B分离评分分布再次显示出A和B基因座在个体染色体中分离的总体趋势(图18F)。Chr2中的空间分离程度似乎比Chr21中的小(图18C和18F)。

[0382] 图18A-18I显示了单个染色体中的隔室结构以及转录活动与局部染色质含量之间的关系。图18A显示了源自我们的成像数据的Chr21的基因组距离标准化邻近频率的Pearson相关矩阵。如果两个基因座的距离小于500nm的截止距离,则认为两个基因座接近。底部的两个条显示源自邻近频率矩阵的A/B调用(显示为A隔室和B隔室)和染色体中每个基因组基因座的G显带。

[0383] 图18B显示了单个细胞中Chr21的个体拷贝的3D渲染,其中A和B基因座显示为球。柔性线连接基因组序列中的相邻基因座。底部的条显示了源自邻近频率矩阵的染色体中每个基因组基因座的A/B调用。比例尺:1微米。

[0384] 图18C显示了Chr21的个体拷贝的A/B分离评分的分布。为了计算A/B分离评分,通过对局部A(或B)密度进行阈值化使得2/3的A(或B)基因座包含在体积内来为每个染色体定义A(或B)密集体积(注意对于显示A和B基因座之间的空间重叠的染色体,A和B密集体积可以重叠)。染色体的A(或B)密集体积中基因座的纯度定义为体积内作为A(或B)基因座的所有基因座的分数,染色体拷贝的A/B分离评分定义为A和B体积的平均纯度。直方图表示随机

对照的A/B分离评分的分布,其中连续A和B区域之间的边界沿基因组序列随机移动,同时保持A和B区域的数量和尺寸不变。 $n \sim 7,500$ 个染色体。

[0385] 图18D显示了从我们的成像数据得出的Chr2的p和q臂的基因组距离标准化邻近频率的Pearson相关矩阵,以及相应的A/B调用和G显带,如图18A所示。

[0386] 图18E显示了Chr2的个体拷贝的3D渲染,如图18B所示。比例尺:1微米。

[0387] 图18F显示了Chr2在单个细胞中的A/B分离评分的分布,如图18C所示。 $n \sim 3,100$ 个染色体。

[0388] 图25A-25G显示Chr21和Chr2的整体A/B隔室分析。图25A显示了基于Chr21的主成分分析的隔室调用。显示了对于来自源自成像(上图)和整体Hi-C(下图)实验的基因组距离标准化邻近频率的Pearson相关矩阵计算的第一主成分(PC1),PC1值 >0 对应于A隔室和PC1值 <0 对应于B隔室。

[0389] 图25B显示了Chr21的单个拷贝中A隔室基因座(A基因座)和B隔室基因座(B基因座)的3D渲染,以及A/B密度比。左图:Chr21的代表性拷贝的A和B基因座。来自源自成像的整体邻近频率图的A和B隔室调用显示在底部栏。右图:相同的染色体,但每个基因座都由其局部A/B密度比着色。

[0390] 图25C显示了跨所有成像的细胞平均的Chr21中每个成像基因座的平均A和B密度评分。下图表示来自源自成像的邻近频率图的每个基因座的A或B隔室调用。

[0391] 图25D显示了在细胞周期的G1和G2/S期细胞中Chr21的个体拷贝的A/B分离评分的分布的直方图。

[0392] 图25E显示了从Chr2的成像(左)和整体Hi-C接触矩阵(右)导出的邻近频率矩阵。Hi-C接触矩阵以50kb分箱,但仅显示来自成像的区段的接触(以250kb间隔选择)。

[0393] 图25F显示了与图25A相同的PC分析,但针对Chr2的p臂(上图)和q臂(下图)。

[0394] 图25G显示了与图25C相同的平均A和B密度分析,但针对Chr2。

[0395] 实施例10

[0396] 转录与局部A/B染色质含量的关系。为了研究染色质隔室化是否与单个染色体中的活性转录相关,设计了靶向Chr21中含有的86个基因的第一个内含子的寡核苷酸探针,并进行了连续轮次的杂交,以对这些基因的新生RNA转录物进行成像,然后进行染色质追踪。此外,为了更准确地检测基因的空间位置,对以每个靶基因的转录起始位点(TSS)为中心的5kb基因组基因座进行了成像。为了防止RNA探针与基因组DNA结合,反之亦然,在不加热变性双链基因组的情况下进行RNA探针杂交,然后使用核糖核酸酶处理消化RNA分子(单独成像染色质时也包括这一步骤),然后进行染色质追踪。使用该策略确认RNA和DNA信号之间的串扰可以忽略不计(图26A-26J)。

[0397] 通常,成像的基因的一个子集在任何个体细胞中都显示出转录活动(图18G)。研究了转录活动如何与局部染色质环境相关联。为了表征局部A/B染色质含量,计算每个基因的A和B基因座附近的局部密度,并将它们的比率(以下称为A/B密度比)用作活性染色质的局部富集的指标。发现对于所研究的约80%的基因,当基因活性转录时,其TSS处的局部A/B密度比高于基因未激发时(图18H)。作为推论,在基因的TSS具有较高的局部A/B密度比的细胞中,基因的激发率也往往较高(图18I)。这些结果表明,同一基因在于该基因附近具有A基因座的更高富集和/或B基因座的去富集的细胞中往往具有更高的转录活动。这种转录活动的

增加可能是由于转录机制的局部富集和/或沉默因子的去富集。可替代地,考虑到转录机制和辅助因子可以形成凝聚物,可能的是与转录机制相关的活性转录染色质具有与其他活性染色质相互作用的更强趋势。注意的是,这两种可能的机制并不相互排斥,而是可以协同工作以相互加强。

[0398] 图18G显示了Chr21的单个拷贝的3D渲染,与测量的基因的转录爆发一起显示。球代表该染色体中所有检测到的新生RNA爆发。比例尺:500nm。

[0399] 图18H显示了在活性激发状态和非激发状态之间成像的基因的转录起始位点(TSS)处的A/B密度比的变化(以对数差测量)。对于每个基因,中值A/B密度比是在其中基因正在激发的染色体中和在其中基因不激发的染色体中在其TSS处计算的。在Chr21上成像的84个基因的这些值的对数差根据其中值A/B密度比的变化幅度进行排序。79%的成像基因在活性激发时表现出A/B密度比与不激发相比的增加。

[0400] 图18I显示了当基因的TSS的局部环境从低(底部四分位数)变为高(顶部四分位数)A/B密度比时,成像基因的激发率的变化(以对数差异测量)。在Chr21上成像的84个基因的激发率对数差根据激发率的数值进行排序。79%的成像基因在其TSS位于顶部四分位数时与底部四分位数相比显示A/B密度比的更高的激发率。

[0401] 图26A-26J显示了RNA和DNA FISH探针串扰的测量。图26A显示了细胞核由DAPI标记的示例细胞(上图)和靶向基因(BRWD1)的新生RNA的FISH探针的荧光信号(下图)。

[0402] 图26B与图26A相同,但基因不同(SCAF4)。图26A和26B的染色遵循实施例19中“细胞培养物制备和初级/编码探针杂交”部分所述的RNA FISH方案中所述的方案。

[0403] 图26C和26D分别与图26A和26B相同,除了修改RNA FISH方案以包括额外的核糖核酸酶处理步骤以在添加FISH探针之前去除细胞RNA。图26C和图26D中的细胞在与图26A和26B相似的照明条件下成像,并且它们的荧光信号以与图26A和26B中相同的对比度显示。

[0404] 图26E显示了5个测量的基因中未处理和核糖核酸酶处理的细胞的信噪比>3的每个细胞的斑点数量。

[0405] 图26F显示了细胞核由DAPI标记的示例细胞(上图)和靶向基因组基因座(chr21:15.2Mb-15.25Mb)的探针的荧光信号(下图)。

[0406] 图26G与图26F相同,但用于不同的基因座(chr21:14.95Mb-15Mb)。图26F和图26G的染色遵循实施例19中“细胞培养物制备和初级/编码探针杂交”部分中描述的DNA FISH方案中所述的方案。

[0407] 图26H和26I分别与图26F和26G相同,除了修改DNA FISH方案以省略热变性步骤和因此去除了可接近的基因组DNA位点。图26H和26I中的细胞在与图26F和26G相似的照明条件下成像,并且它们的荧光信号以与图26H和26I中相同的对比度显示。

[0408] 图26J显示了对于用热变性步骤处理的细胞和省略了该步骤的细胞的信噪比>3的每个细胞的斑点数量。

[0409] 实施例11

[0410] 单个染色体中染色质结构域和隔室之间的关系。接下来,研究了单细胞染色质结构域之间的相互作用以及这些相互作用如何与隔室身份相关联。由于Chr2的大尺寸和其中的隔室分区的大数量,对Chr2的分析预期提供更多的见解,因此重点关注该染色体。

[0411] 尽管Ch2中的大多数单细胞结构域是完全由A或B基因座组成的“纯”结构域,但很

大一部分单细胞结构域跨越了整体A/B边界,并包含A和B基因座两者(图19A-19C)。这些“混合”结构域的存在表明,单个细胞中的结构域形成可能不与决定隔室身份的染色质特性密切相关,但表观遗传修饰中的细胞间变异性也可能导致一些单细胞中活性/非活性染色质边界的位移。

[0412] 接下来检查结构域如何彼此相互作用,重点关注结构域间相互作用如何依赖于结构域的A和B组成,以及结构域之间的基因组距离。结构域在短和长基因组分离中都发生接触,并且这种接触在个体染色体的空间距离图中表现为非对角框特征。这种接触模式在细胞间显著变化(图19D)。尽管存在这种异质性,结构域间相互作用似乎受到其基础染色质的A/B组成的调节(图19E):主要包含B基因座的结构域之间的接触频率平均高于主要包含A基因座的结构域之间的接触频率,这进而又高于由不同A/B身份的染色质主导的结构域之间的接触频率。该平均图片与最近基于通过Hi-C测量的A/B隔室化的染色质结构建模和通过成像测量的A和B基因座的全局排列提出的A和B染色质相互作用强度的层次结构一致。

[0413] 作为基因组距离的函数的结构域接触频率的进一步检查揭示了更复杂的情况。为简单起见,重点关注包含单个隔室身份的基因座的“纯”A和“纯”B结构域。正如预期的那样,对于所有组成的结构域对,接触频率随着基因组距离而降低(图19F)。然而,在较短的基因组距离(对于Chr2高达~75Mb)处,B结构域对(B-B)之间的接触频率高于A结构域对(A-A)之间的接触频率,而A-A结构域接触在较大基因组分离处胜过B-B结构域接触(图19F)。该结果与最近的整体Hi-C研究中报道的A-A和B-B染色质相互作用之间的基因组距离依赖性一致,并提供了关于单细胞结构域之间的优先相互作用可如何导致这些整体趋势的进一步见解。值得注意的是,在相对较大的基因组距离处,B-B结构域接触概率衰减到与A和B结构域之间(A-B)的接触概率相似的水平,而A-A结构域接触概率甚至在较大的基因组分离处仍然高于A-B结构域接触概率(图19F)。这导致在较大的基因组距离处A-A结构域相互作用明显占主导地位(图19G)。此外,接触结构域对也显示出不同程度的空间重叠,其中一些结构域对显示出相对表面的接触(图19F插图),而其他对显示出强烈的混合(图19H插图)。有趣的是,与A-A结构域对相比,B-B结构域对显示出明显更强的形成这种混合小球的趋势(图19H)。

[0414] 总体而言,这些结果表明优先A-A和B-B结构域相互作用导致染色质隔室的空间分离,并且A-A和B-B结构域相互作用的性质是不同的。这些相互作用的性质的差异可能源于参与A-A和B-B关联的不同分子因子。例如,异染色质因子(例如HP1)被认为参与B-B相互作用,而转录激活因子或共激活因子例如BRD4和介体可能参与活性染色质相互作用。这些不同的分子因子是否负责A-A和B-B结构域相互作用之间的基因组距离依赖性的观察到的差异以及它们混合的趋势有待进一步研究。

[0415] 图19A-19H显示了结构域间相互作用对其A/B组成和基因组距离的依赖性。图19A,左图是包含A和B基因座的“混合”染色质结构域的3D渲染,其两侧是单个细胞中Chr2拷贝中仅包含B基因座的“纯”结构域。比例尺:500nm。右图:左侧显示的同一区域的成对距离矩阵。矩阵底部和左侧的条显示基因座的A和B调用,轮廓突出显示染色质结构域的边界。A/B调用从Chr2的整体邻近频率图确定。

[0416] 图19B与图19A相同,但针对两个纯结构域,一个完全由A基因座组成,一个完全由B基因座组成,而不是混合的结构域。

[0417] 图19C是Chr2中单细胞染色质结构域中作为A基因座的基因座的分数的分布。

[0418] 图19D是Chr2的两个示例拷贝的单细胞空间距离矩阵。第一个和第三个图显示了两个完整染色体的矩阵,而第二个和最后一个图分别显示了在第一个和第三个图中以黄色突出显示的区域放大矩阵。侧边条显示来自整体邻近频率图的A/B隔室调用。

[0419] 图19E显示了Chr2中不同A/B组成的结构域的结构域接触概率。X轴和Y轴表示结构域内作为A基因座的基因座的分数(0%对应于纯B结构域,100%对应于纯A结构域)。如果两个结构域的绝缘评分 <2 ,则将它们定义为接触。对于绝缘评分的计算参见实施例19。

[0420] 图19F显示了Chr2中两个纯A结构域(A-A)、两个纯B结构域之间(B-B)以及一个纯A和一个纯B结构域(A-B)之间的结构域接触概率,绘制为两个相互作用结构域之间的基因组距离的函数。插图包含示例结构域对的3D渲染,显示绝缘评分 $=2$ 的远程相互作用。比例尺:500nm。

[0421] 图19G与图19E相同,但针对基因组距离大于80Mb的结构域对。

[0422] 图19H与图19F相同,但限于具有高度混合(由 <1 的低绝缘评分定义)的结构域对。插图包含示例结构域对的3D渲染,显示具有高度混合(绝缘评分 $=1$)的远程相互作用。比例尺:500nm。

[0423] 实施例12

[0424] 基因组规模的染色质成像。上述顺序成像方法允许获得个体染色体中染色质的高分辨率视图。这种直接顺序成像方法非常适合对与衍射极限分辨率相当或小于衍射极限分辨率的染色质结构进行成像。然而,在这种方法中,成像的基因组基因座数量仅随成像轮次数线性增加。对于基因组规模的染色质成像,由于许多基因组基因座可以同时解析并定位在细胞核中,因此推断成像的基因座的数量与成像的轮次的数量的更有效、非线性缩放是可能的。

[0425] 为了实现这一目标,设计了一种组合FISH方法,其灵感来自先前为转录组成像开发的多重抗错FISH方法,但通过考虑染色质的聚合性质(即基因组序列中的相邻基因座在空间上接近)和染色体的区域组织(即不同的染色体倾向于占据不同的空间区域)进行了特别设计用于染色质成像的重要修改。为了允许组合成像,每个基因组基因座被分配一个唯一的100位二进制条形码,Hamming权重为2,即每个条形码包含两个“1”位和98个“0”位(图20A)。这些条形码中的位值(“1”或“0”)决定了在连续几轮成像中每个基因座的信号的存在或不存在。为了避免在同一位中同时成像空间上接近的染色质区域,从这些100位Hamming权重2条形码中,进一步选择了一个子集来编码靶向的基因组基因座并优化条形码的分配,使得在相同的条形码位置中具有“1”位的基因座在基因组空间中被最大程度地分开(参见实施例19)。这种策略允许最大限度地减少由来自附近染色质基因座的重叠信号引起的检测错误。此外,由于绝大多数可能的100位二进制代码是无效的(即未分配给任何靶基因座),因此这种设计允许鉴定和丢弃检测错误,并进一步提高测量精度。

[0426] 使用编码探针的高度多样性文库将条形码物理印在靶向的基因组基因座上,每个编码探针包含用于结合靶基因座之一的靶区域和从100个预先设计的读出序列中选择的读出序列(图20A)。每个读出序列对应于100位中的一个,并且每个基因组基因座的编码探针组(每个基因座约400个探针)仅包含两个不同的读出序列,对应于在分配给该基因座的条形码中读取“1”的两个位。在编码探针结合后,通过荧光标记的读出探针的依次杂交检测印在染色质基因座上的条形码,每个探针与100个读出序列之一互补(图20A)。在一些情况下,

此处还使用了用于高分辨率全染色体追踪的衔接子探针策略。每个杂交轮次引入两个不同的衔接子/读出探针,并在双颜色通道中成像,使得在每个杂交轮次中读出2个位。这允许仅用50轮杂交对约1000个基因组基因座进行成像和鉴定(图20A-20C)。与具有相同数量颜色通道的相同数量基因座的顺序成像相比,这表示约10倍更少的杂交轮次,因此10倍更短的实验时间。由于二倍体细胞中的每个染色体都有两个同源物,因此使用聚类算法进一步分配了成像的基因座的同源物身份,利用染色体在每个细胞核中占据不同区域的趋势。

[0427] 在这项工作中,选择了1,041个基因组基因座进行成像,每个基因座尺寸为约30-kb,均匀覆盖IMR-90细胞中的22个常染色体和X染色体。另一个要求是每个染色体至少包含30个靶基因座,因此每个染色体同源物成像的基因座数量为30至80,取决于染色体的长度。对5个生物复制中约5,400个个体细胞中的这1,041个基因组基因座进行成像,每个基因座的检测效率为约80%,在每个细胞中检测到约1700个染色质基因座(图20D-20E)。在组合成像过程结束时,基因组基因座的一小部分通过顺序成像重新成像,一次一个基因座。通过组合成像确定的基因座位置和通过顺序成像确定的重新成像的位置之间的位移仅为~50nm(图27A),表明组合成像方法的高解码精度和成像过程中最小的样品降解/变形。

[0428] 为了获得染色质组织的群体平均视图,计算每个细胞中每对成像的染色质基因座之间的空间距离,然后确定所有成像的细胞中每对基因座之间的中位距离和邻近频率(图20F;图27B)。从成像数据确定的同一染色体内染色质基因座对之间的邻近频率显示出与通过整体Hi-C检测的接触频率高度相关,Pearson相关系数为0.89(图27C)。此外,成像结果显示独立生物学重复之间的高再现性(图27D)。

[0429] 通过探索个体细胞中的染色质组织,注意到染色体虽然具有占据每个细胞内的不同区域的倾向(图20F-20G),但也显示出彼此之间的大量重叠(图20G-20H)。这些结果与早期成像研究的观察结果一致并扩展了早期成像研究的观察结果。由于这一观察结果表明存在高度的跨染色体相互作用,因此进一步的分析重点关注探索它们。

[0430] 图20A-20H显示了通过大规模多重组合FISH成像的基因组规模染色质。图20A显示了成像方案。靶向的基因组基因座被分配了抗错条形码,例如Hamming权重为2的100位二进制条形码的子集(即100位中的两个读取“1”)。条形码用编码寡核苷酸探针印在基因组基因座上,所述寡核苷酸探针识别基因座并将两个不同的读出序列与每个基因座相关联,对应于分配给基因座的条形码中读取“1”的两个位。每个位都被唯一地分配一个读出序列。每个基因座由总共400个编码探针标记,但仅显示了4个。与读出序列互补的荧光读出探针被依次添加并成像,从而允许在每个基因座上读取“1”的位,从而确定该基因座的条形码身份。

[0431] 图20B显示了来自单个细胞的核中的多个成像轮次的代表性图像。显示了来自读出探针的染色质基因座的荧光信号和用作核标记的4',6-二脒基-2-苯基吡啶(DAPI)的信号。比例尺:5微米。

[0432] 图20C显示了在所有成像轮次中以一个染色质基因座为中心的小区域(B中的白框)的放大图像。基于发出信号的两个读出探针(1和13)确定基因座身份。比例尺:300nm。

[0433] 图20D是单个IMR-90细胞中所有检测到的染色质基因座(球体)的3D渲染,根据它们所属的染色体进行颜色编码(图像下方显示的染色体的索引)。基因组序列中的相邻基因座由柔性线连接。对约1000个基因组基因座进行成像。

[0434] 图20E显示了与图20D中相同的细胞的染色质基因座,但突出显示了所示染色体的

两个同源物。

[0435] 图20F显示了从约5,400个单细胞计算的中值距离矩阵。对于每对基因座,呈现了所有细胞中基因座之间观察到的3D空间距离的中值。

[0436] 图20G显示了示例图像,显示了单个细胞中多个染色体区域的位置。染色体如指示编码,阴影区域代表围绕所有成像的基因座的凸包。为清楚起见,每个染色体仅显示一个同源物。

[0437] 图20H显示了图20G中所示相同细胞的空间距离矩阵。显示了每对染色质基因座之间的空间距离。染色体顺序如矩阵下方所述,分别显示了每个染色体的两个同源物。

[0438] 图27A-27J显示了通过组合FISH进行的基因组规模成像:定位误差、再现性以及与Hi-C的比较。图27A显示了在组合成像运行中测量的基因组基因座的定位与在完成组合成像后使用依次杂交单独重新成像的相同基因座的定位之间的位移分布。Chr6中的10个基因组区域在约2000个细胞中重新成像。中位位移为~50nm。

[0439] 图27B显示了通过组合FISH成像的所有1,041个基因组基因座的邻近频率矩阵。一对基因座之间的邻近频率计算为其中基因座之间的测量距离小于500nm的截止距离的发生次数除以两个基因座之间的测量距离的总数。

[0440] 图27C显示了源自我们的成像数据的染色体内基因座对之间的邻近频率和源自整体Hi-C实验的接触数量的相关图,以500kb分箱并以靶基因座为中心。Pearson相关系数为0.91。IMR90细胞的可用Hi-C数据对于跨染色体接触来说是稀少的,因此阻碍了对我们的成像数据和Hi-C数据之间的跨染色体相互作用的可靠比较。

[0441] 图27D显示了在基因组规模成像实验的两个独立生物学重复中观察到的染色质基因座之间的成对距离的相关性。重复之间的Pearson相关系数为0.98。右上方的云表示跨染色体成对距离,左下方的云表示染色体内成对距离。

[0442] 实施例13

[0443] 跨染色体A-A相互作用的富集。接下来,研究了跨染色体相互作用如何依赖于染色质的A/B隔室身份。基于已发布的整体Hi-C数据,将每个成像的基因组基因座分类为A和B隔室。尽管基因组规模的成像数据也允许相当准确的A/B隔室调用,给出与基于Hi-C数据的调用的约80%的一致性(图27E),但由于整体Hi-C数据的更高基因组分辨率,Hi-C调用用于对成像的基因座的A/B隔室身份进行分类。38%的成像的基因座属于A隔室,而62%属于B隔室。为了检查活性和非活性染色质的跨染色体相互作用的程度是否不同,重新排列了跨染色体邻近频率矩阵中的基因组基因座,将所有A基因座彼此相邻放置,然后是所有B基因座。该矩阵显示,A隔室基因座与另一个A隔室基因座的跨染色体相互作用的平均趋势比与单独的B隔室基因座的相互作用更强(图21A-21B),这与之前对活性染色质之间的跨染色体相互作用的观察结果一致。相反,与针对A隔室基因座相比,B隔室基因座显示出针对彼此的相当或更低的跨染色体亲和力(图21A-21B)。换言之,跨染色体A-A相互作用表现出比A-B相互作用明显更强的趋势,A-B相互作用进而表现出比B-B相互作用略强的趋势。对于用于构建邻近频率图的大范围截止距离观察到类似趋势,前提是分析中包括足够多的细胞(图27F-27H)。值得注意的是,这与同一染色体内顺式相互作用的整体层次形成鲜明对比,其中B-B相互作用比A-A相互作用具有更强的形成趋势,A-A相互作用进而比A-B相互作用具有更强的形成趋势(图19E)。然而,有趣的是,跨染色体相互作用的这种观察到的趋势($A-A > A-B \approx$

B-B) 与高分辨率Chr2数据中在大基因组距离处针对顺式染色体相互作用观察到的趋势相似(图19F-19G)。对于基因组规模数据中的远程顺式染色体相互作用也观察到这种趋势,跨所有成像的染色体聚集(图21C)。

[0444] 接下来,检查了在单细胞水平上跨染色体相互作用和染色质隔室身份的关系。在个体细胞中,A和B基因座采用不同的空间分布,其中A基因座表现出比B基因座更集中地位于细胞核中的趋势,如预期的那样(图21D;图28)。为了进一步表征跨染色质相互作用,采用类似于前面介绍的基于密度的方法用于高分辨率全染色体追踪,除了此处仅考虑跨染色体相互作用。简而言之,对于每个染色体中的每个成像的基因座,计算来自同一细胞中所有其他染色体的A基因座和B基因座的局部密度,并确定这两个密度的比率(以下称为跨A/B密度比率)(图21D-21E)。该数量提供了对基因座附近跨染色体活性染色质的局部富集的测量。注意到大多数(62%)的成像的基因座属于B隔室,产生小于1的A/B比率总体偏差。为了控制这种偏差,比较了针对A基因座和B基因座观察到的跨A/B密度比的分布,其中分布在其中成像的基因座的A和B身份在成像的基因座之间随机打乱同时保持A和B基因座的数量不变的随机化对照中获得。值得注意的是,针对A基因座观察到的跨A/B密度比显著高于针对B基因座观察到的值,其进而高于源自随机化对照的值(图21E)。这些单细胞分析进一步支持了跨染色体相互作用优先富集活性染色质之间的相互作用的观点。

[0445] 进一步询问跨染色体A-A相互作用的富集是否需要转录。为了解决这个问题,通过 α -鹅膏蕈素处理抑制转录,并且发现这种处理不会导致跨染色体A-A相互作用的富集的显著降低(图29A-29C)。这一观察结果与先前的研究一致并扩展了先前的研究,其显示活化的 β -珠蛋白基因座与其他活性染色质基因座的远程和跨染色体相互作用不受转录抑制的抑制。

[0446] 图21A-21E显示跨染色体相互作用中活性-活性染色质相互作用的富集。图21A显示了标准化的跨染色体邻近频率矩阵。显示了每个跨染色体基因座对(不同染色体上的基因座对)之间的邻近频率,其中如果基因座对的距离小于500nm的截止距离,则认为它们是接近的。基因座被重新排序,使得A隔室基因座首先出现,然后是B隔室基因座,因此左上角的块代表A基因座对之间的相互作用,右下角代表B基因座对之间的相互作用。矩阵中的每个条目都通过源自同一对染色体的所有基因座对的中值邻近频率进行标准化,以解释染色体对之间相互作用的不同基础水平。

[0447] 图21B显示了A基因座对(A-A; $n=72,771$ 个基因座对)、B基因座对(B-B; $n=193,753$ 个基因座对)和由一个A和一个B基因座组成的对(A-B; $n=237,986$ 个基因座对)的跨染色体邻近频率的分布,源自图21A中所示的矩阵。分布在上图中以直方图表示,在下图中以箱线图表示,显示中位数(中心线)、第25-75个百分位数(框)和第5-95个百分位数(须线)。

[0448] 图21C显示了相同染色体内作为它们的基因组距离的函数的染色质基因座对之间的中值邻近频率,对跨所有染色体的由相同基因组距离隔开的基因座对进行平均。显示了A基因座对(A-A)、B基因座对(B-B)和A和B基因座的混合对(A-B)的中值接触频率。

[0449] 图21D显示了A隔室和B隔室基因座在两个单细胞中的分布。左图表示单个核中单个z平面内所有检测到的基因座的位置,具有A隔室基因座和B隔室基因座。在右图中,每个基因座的阴影表示跨染色体A和B基因座的局部密度比,即跨A/B密度比,与右侧显示的比例尺一致。

[0450] 图21E显示了成像的基因组基因座的局部跨A/B密度比的分布。对于每个基因座，确定了所有细胞的中值跨A/B密度比，并显示了A基因座的跨A/B密度比分布 ($n=382$ 个基因座)，以及显示了B基因座的跨A/B密度比分布 ($n=623$ 个基因座)。由于用于隔壁调用的Hi-C数据集和本研究中使用的不同形式的基因组组装，1041个成像的基因座中的36个未分配A/B身份。深灰色直方图表示随机对照，其中A和B隔壁身份被随机打乱，同时保持A基因座的总数和B基因座的总数不变。

[0451] 图27E是其A/B隔壁分配在对于每个人类常染色体的基因组规模成像数据和Hi-C数据之间一致的基因座的百分比的条形图。平均而言，每个染色体中约81%的基因座在我们的成像数据和Hi-C数据之间的A/B分配中表现出一致性。

[0452] 图27F显示了作为用于评估接近度的截止距离的函数的中值标准化跨染色体A-A、A-B和B-B邻近频率(如图21A和21B中定义)。标准化的近似值是从 $\sim 5,400$ 个IMR-90细胞计算的。还显示了用具有50nm标准偏差的3D高斯噪声项扰乱基因座位置后的中值标准化跨染色体A-A、A-B和B-B邻近频率，与估计的定位测量误差相当，如图27A所示。

[0453] 图27G与图27F相同，但是当将来自 α -鹅膏菌素处理的细胞的额外数据与未处理的细胞合并时(总共约9,500个细胞)。 α -鹅膏菌素处理的细胞显示跨染色体A-A相对于A-B和B-B邻近频率的与未处理细胞类似的富集(图29A、29B)。该合并结果表明，在图27F中在较低截止距离处观察到的A-A相互作用的较低富集可能是具有较少数量细胞的较差统计数据的结果。

[0454] 图27H显示了中值标准化跨染色体A-A、A-B和B-B邻近频率，作为分析中包括的细胞数量的函数。从成像的约5,400个未处理的IMR-90细胞中随机二次抽样细胞，并将邻近截止距离固定为500nm。

[0455] 图29A-29F显示转录抑制对跨染色体染色质相互作用和染色质基因座的核体缔合率的影响。图29A显示了标准化的跨染色体邻近频率矩阵，如图21A所示，但是针对用 α -鹅膏菌素处理以抑制转录的细胞。

[0456] 图29B显示了以箱线图显示的标准化跨染色体A-A、B-B和A-B邻近频率的分布，如图21B所示，但是针对用 α -鹅膏菌素处理的细胞。为了比较，此处再现了来自图21B的未处理细胞的标准化跨染色体A-A、B-B和A-B邻近频率。

[0457] 图29C显示了跨成像的A和B基因座的局部跨A/B密度比的分布，如图21E所示，但是针对用 α -鹅膏菌素处理的细胞。直方图表示其中A和B隔壁身份随机打乱同时保持A基因座的总数和B基因座的总数不变的随机化对照。

[0458] 实施例14

[0459] 染色质、新生RNA和核结构的多模态成像。为了将染色质的3D组织置于其功能活动和其他核结构的背景下，扩展组合成像方法以允许同时测量染色质组织以及成像的基因组基因座的转录活动，以及单细胞中的核标志物。具体而言，对上述1,041个基因组基因座以及从位于这些基因座的1,137个基因中的每一个转录的新生RNA进行成像，同时对重要的核结构(包括核斑点和核仁)进行成像(图22A)。

[0460] 为了允许在相同细胞内进行DNA、RNA和核结构成像，通过采用与上述对于染色质描述的相似的组合成像策略，对1,137个基因的内含子RNA进行多重成像(图22A)。考虑到并非所有基因都会在每个个体细胞中转录，因此转录灶的密度不应像染色质基因座的密度那

样高, RNA用54位Hamming权重2代码编码, 和选择1,137个编码基因的可能条形码, 其方式类似于如何选择用于染色质成像的条形码以最小化在同一位中成像空间邻近基因的机会。完成RNA成像后, RNA转录物被酶消化(也在单模态染色质成像实验中进行的步骤), 并如上所述进行多重DNA FISH以对1,041个基因组基因座进行成像(图22A)。基因组基因座和新生RNA转录物的解码在很大程度上是独立进行的, 另外还限制了转录物与含有它们的基因组基因座共定位(参见实施例19)。该程序进一步提高了RNA转录物的检测准确度, 并允许估计每个基因组基因座处的转录爆发的检测效率($\sim 90\%$) (参见实施例19)。最后, 使用针对这些结构的已知分子成分的免疫荧光对核斑点和核仁进行成像(图22A)。即使在DNA FISH后进行免疫荧光染色, 核斑点和核仁的荧光信号也显示出高信噪比(>25)。通过计算包含所有成像的基因组基因座的凸包表面来估计核纤层的位置。总之, 这些多模态测量允许3D基因组结构、转录活动和核组织的整体单细胞视图(图22B)。这些多模态成像实验是在约3700个个体细胞上在两个生物重复中进行的。来自这些多模态实验的染色质成像数据也包含在上文对于3D基因组组织分析所述的5个重复和约5,400个细胞中。

[0461] 从这些多模态实验的新生RNA转录物测量中, 对每个基因量化来自RNA内含子信号亮度的作为活性转录基因的细胞的分数的转录爆发频率(图22C)和中值爆发尺寸(图22D)。这些测量在重复实验中显示出高度相关性(图27I-27J)。爆发频率表现出双峰行为, 其中高爆发频率基因主要存在于A隔室中, 而低爆发频率基因存在于两个隔室中(图22C)。此外, 使用250nm的空间距离截止值估计特定染色质基因座是否与核体相关, 并观察到B基因座与核纤层的更高关联频率(图22E)和A基因座与核斑点的更高关联频率(图22F)。这些结果与先前观察到的非活性和活性染色质分别与核纤层和核斑点优先关联的观察结果一致。对于个体基因座, 它们的中位局部跨A/B密度比表现出与核纤层关联频率的负相关(图22G), 和与核斑点关联频率的正相关(图22H)。如前所述, 核仁另外显示出与着丝粒、某些染色体的端粒和含有核糖体编码基因的染色体的优先结合(图22I)。这些生物学结果为多模态测量提供了进一步的验证。

[0462] 值得注意的是, 对于大多数成像的基因座, 核纤层关联与较低的转录活动相关, 而核斑点关联与较高的转录活动相关(图22J)。这些结果与最近的单细胞测序研究一致, 该研究表明核纤层关联与单细胞中的基因表达负相关。此外, 观察到用转录抑制剂 α -鹅膏蕈素处理扰乱核斑点以及降低核斑点缔合率并增加成像的基因座的核纤层缔合率(图29D-29F)。总之, 这些结果扩展了先前关于转录激活或抑制后单个或几个基因组基因座的核重新定位的成像研究, 并提供了转录活动和与核结构的相互作用之间关系的基因组规模视图。

[0463] 图22A-22J显示了在核结构背景下染色质和转录活动的多模态基因组规模成像。图22A, 上图: 组合染色质(左图)、新生RNA转录物(中图)和核体(右图)成像以生成核结构和功能活动的背景下染色质组织的综合视图的多模态成像方案的图示。对约1000个基因组基因座、靶基因座中约1100个基因的新生RNA转录物和两种类型的核体(核斑点和核仁)进行成像。下图: 每种成像模式的代表性原始图像-跨多个成像轮次的染色质基因座(左)、跨多个成像轮次的新生RNA转录物(中)和核体(右: 核斑点, 使用抗SC35抗体成像; 和核仁, 使用抗纤维蛋白抗体成像)。比例尺: 5微米。

[0464] 图22B是单个细胞中染色质基因座、转录爆发和核体的3D渲染。左图: 所有检测到

的染色质基因座,由染色体编码(基于下面显示的染色体索引)。中间图:显示为彩色球体的所有检测到的内含子RNA,阴影表示成像的基因的身份,球体尺寸表示转录爆发尺寸。右图:检测到的核体的体积填充表示。核纤层被鉴定为围绕所有检测到的染色质基因座的凸包表面(阴影灰色区域)。

[0465] 图22C和22D显示了位于成像的A隔壁基因座($n=558$ 个基因)和B隔壁基因座($n=569$ 个基因)中的基因的转录爆发频率(图22C)和爆发尺寸(图22D)的分布。

[0466] 图22E和22F显示了A基因座($n=382$ 个基因座)和B基因座($n=623$ 个基因座)与核纤层(图22E)和核斑点(图22F)的关联率的分布。如果基因座与核外围或最近的核斑点的距离 $<250\text{nm}$,则认为染色质基因座与核纤层或核斑点相关联。

[0467] 图22G和22H显示了每个成像的基因组基因座的作为发现该基因座与核纤层(图22G, Pearson相关系数 $= -0.87$)和核斑点(图22H, Pearson相关系数 $= 0.66$)关联的函数的局部跨A/B密度比的散点图。所示的跨A/B密度比的值是所有成像的细胞的中值。

[0468] 图22I显示了所有成像的基因组基因座的与核仁的关联频率,按基因组位置排序。黑色垂直线表示着丝粒的位置,括号突出显示含有核糖体编码基因(rDNA)的染色体。

[0469] 图22J显示了转录与核结构关联的相关性。当比较其中基因座是核纤层关联的相对非核纤层关联的(左)和核斑点关联的相对与非核斑点关联的(右)的细胞群时,圆圈是个体基因组基因座的转录爆发频率的变化倍数。虚线突出显示没有变化,实线表示每种情况下的中值变化倍数。

[0470] 图27I和27J显示了每个基因的爆发频率(图27I)和爆发尺寸(图27J)的RNA成像重复之间的相关性。Pearson相关系数分别为0.94和0.81。

[0471] 图29D和29E显示了个体细胞核的代表性图像,其中显示了未处理细胞(图29D)和用 α -鹅膏蕈素处理的细胞(图29E)的成像的染色质基因座、核仁和核斑点。

[0472] 图29F显示了 α -鹅膏蕈素处理后每个基因座与核纤层(左)和核斑点(右)的关联率的变化倍数。每个基因组基因座的数据点以圆圈显示,实线是每种情况下所有基因座的中值变化倍数,虚线表示没有变化。值得注意的是,核体积和核斑点的尺寸和数量在用 α -鹅膏蕈素处理后也发生了变化,这些可能部分促成了核体关联的变化。

[0473] 实施例15

[0474] 检查各种核环境中的跨染色体相互作用。同一细胞中染色质组织和标志性核结构的同时成像进一步允许检查跨染色体A-A相互作用的观察到的富集如何依赖于核背景。因为核斑点是聚集活性转录基因座的最突出的核体之一,所以推测跨染色体A-A相互作用的观察到的富集是否仅仅是核斑点处这种局部聚集效应的结果。为了解决这个问题,分析限于不与核斑点关联的基因座,即对于每个基因座对,只考虑两个基因座都不与核斑点关联的那些细胞。有趣的是,在这种约束下,对于跨染色体A-A相互作用相对于A-B和B-B相互作用的富集仍然观察到相同的趋势(图30A-30B),表明与核斑点的关联不足以解释跨染色体A-A相互作用的观察到的富集。

[0475] 接下来,考虑了由于局部聚集效应导致的相对微不足道的情况:由于A隔壁基因座从核纤层中耗尽并且在细胞核的内部区域中更多地聚集(图28A-28B),因此推测跨染色体A-A相互作用的富集是否仅仅是由核内部的A隔壁染色质的局部富集引起的。为了测试这一点,对于每个基因座对,只考虑其中两个基因座都与核纤层关联的那些细胞。值得注意的

是,即使对于这些与核纤层关联的基因座,也观察到跨染色体A-A相互作用相对于A-B和B-B相互作用的富集(图30C-30D),尽管存在核纤层是富含非活性B隔室染色质的环境的事实。

[0476] 总体而言,这些结果表明了活性染色质之间观察到的跨染色体相互作用的重要分子机制。如前所述,对于大基因组距离处的顺式染色体相互作用也观察到A-A相对于A-B和B-B相互作用的富集(图19F-19G和图21C)。可能的是远程顺式A-A相互作用与跨染色体A-A相互作用具有共同的潜在机制。哪些分子因子导致这些活性染色质相互作用仍然是一个悬而未决的问题。

[0477] 图28A-28B显示A隔室和B隔室基因座在细胞核中显示出不同的空间分布。在图28A中,左图显示了在单细胞的单个z平面中显示A基因座和B基因座的示例图像。右图显示了这些单细胞中A基因座和B基因座的距离核外围的距离的分布。核外围被鉴定为围绕所有检测到的染色质基因座的凸包。

[0478] 图28B显示了A基因座($n=382$)和B基因座($n=623$)的距离核外围的距离的p平均分布。

[0479] 图30A-30D显示了不同核环境中跨染色体活性染色质相互作用的富集。图30A显示了标准化的跨染色体邻近频率矩阵,如图21A所示,但仅考虑不与核斑点关联的基因座。对于每个基因座对,仅考虑其中两个基因座都不与核斑点关联的细胞。

[0480] 图30B显示了A基因座对(A-A)、B基因座对(B-B)和由一个A和一个B基因座组成的对(A-B)的跨染色体邻近频率,如图21 5B所示,但仅考虑其中两个基因座都不与核斑点关联的细胞。

[0481] 图30C与图30A相同,但针对与核纤层关联的基因座对。对于每个基因座对,仅考虑其中两个基因座都与核纤层关联的细胞。

[0482] 图30D与图30B相同,但针对与核纤层关联的基因座对。对于每个基因座对,仅考虑其中两个基因座都与核纤层关联的细胞。

[0483] 实施例16

[0484] 跨染色体相互作用与转录活动之间的相关性。接下来,这些多模态单细胞测量用于表征个体染色质基因座的转录活动与其由跨染色体贡献定义的局部染色质环境之间的关系。为此,计算了跨A/B密度比,并确定了两个细胞群的该数量的中值(针对每个基因组基因座独立确定):(i)其中所考虑的基因座表现出转录活动的细胞,和(ii)其中基因座出现转录沉默的细胞(图23A)。值得注意的是,当基因座活性转录时,观察到更高跨A/B密度比的一致趋势:与沉默状态相比,86%的成像的基因座在活性转录状态下表现出更大的跨A/B密度比(图23B);同样,当具有较高的跨A/B密度比时,89%的基因座表现出较高的转录激发率(图23C)。在多个不同的核环境中观察到转录活动和局部A隔室染色质富集之间的这种正相关性,包括与核斑点关联的基因座、与核纤层关联的基因座和不与核斑点或核纤层关联的基因座(图23D),尽管与核斑点关联的基因座的相关性较弱。

[0485] 这些观察结果扩展了关于染色体内转录活动和顺式A/B密度比之间的关系的上述结果(图18G-18I),并且一起揭示了基因的转录活动和活性染色质在其局部环境中的富集之间普遍存在的正相关关系。

[0486] 图23A-23D显示转录活动与跨染色体活性染色质的局部富集之间的相关性。图23A显示了染色质基因座和转录活动的单细胞图像。左图:来自单个核的单个z平面中所有成像

的A和B基因座的位置。中间图：相同基因座的局部跨A/B密度比，基于比例尺编码。右图：与中间图相同，检测到的转录爆发叠加并显示为圆圈。比例尺：3微米。

[0487] 图23B显示了在活性激发和非激发状态之间每个成像的基因座的跨A/B密度比的变化(以对数差测量)。对于包含至少一个成像基因的每个基因组基因座，计算其中基因组基因座被活性转录的细胞(指定为转录)和其中基因组基因座未转录的细胞(指定为沉默)的跨A/B密度比。每个成像的基因座的这些值的中位数的对数差异根据变化的幅度进行排序。与不激发时相比，86%的成像的基因座在它们被活性激发时表现出A/B密度比的增加。

[0488] 图23C显示了其中在含有基因的基因座处的跨A/B密度比从低(底部四分位数)变化至高(顶部四分位数)的细胞之间成像的基因的激发率的变化(以对数差测量)。所有成像的基因的激发率对数差异根据激发率的大小进行排序。89%的成像基因在含有它们的基因座位于跨A/B密度比的顶部四分位数时与底部四分位数相比显示出更高的激发率。

[0489] 图23D显示了群体图，显示了成像的含基因的基因座的转录状态和沉默状态之间的局部跨A/B密度比的变化倍数，以它们的核体关联状态为条件。对于每个基因组基因座，计算在基因座的转录状态和沉默状态之间的跨A/B密度比中的变化倍数，从左到右分别考虑：所有细胞，仅其中基因座与核斑点关联的细胞，仅其中基因座与核纤层关联的细胞，以及仅其中基因座不与核斑点关联和不与核纤层关联的细胞(空心圆圈)。对于每个基因座和每个关联条件，确定每个状态(转录或沉默)中的中值跨A/B密度比，并显示了两种状态之间的变化倍数的 $\log_2$ 。虚线表示没有变化，实线表示每种情况下跨所有基因座的中值变化倍数。省略了一些异常值，以允许更清晰地可视化中值变化倍数(对于核斑点关联数据在呈现的标尺之上的5个基因座和之下的9个基因座，对于核纤层关联数据的在呈现的标尺之上的37个基因座和之下的17个基因座，对于非核纤层关联和非核斑点关联数据的在呈现的标尺之上的1个基因座和之下的2个基因座)。

[0490] 实施例17

[0491] 染色体范围和基因组规模的染色质成像。本文报道了大规模多重染色质成像，用于确定单个细胞中的多个基因组组织规模的染色质3D构象。展示了在数千个个体细胞中成像>1000个基因组基因座的能力。该方法通过将染色质追踪与新生转录和核结构成像相结合进一步允许将3D染色质组织置于其天然功能和结构环境中，并且展示了同时成像>1000个基因组基因座的能力、存在于这些基因座中的>1000个基因的转录活动以及标志性核结构，包括核斑点和核仁。

[0492] 具体来说，展示了用于高通量染色质追踪的两种互补策略。首先，扩展了先前报道的基于依次杂交的多重FISH技术的能力，并显示了使用数百轮杂交和多色成像对数百个基因组基因座的成像。通过提供整个染色体的构象的高分辨率视图，以及染色质结构域、隔室和转录与单个细胞中染色质组织的关系的系统表征，证明了这种方法的能力。其次，对于跨越比衍射极限分辨率大得多的空间因此允许在每个成像轮次中解析许多基因座的结构，开发了允许成像的基因座的数量随着成像轮次的数量的快得多的非线性增加的染色质成像的组合标记策略。后一种方法的效力通过对染色质组织和转录进行基因组规模成像来证明，使用仅数十轮杂交同时成像个体细胞中>1000个基因组基因座和>1000个基因的新生转录物。这些数据揭示了全基因组跨染色体相互作用，以及它们与核结构和转录的关系。将这种组合成像方法与已证明的执行数百轮杂交和成像的能力结合起来，应该可以同时成像>

10,000个基因组基因座,以提供单个细胞中染色质结构的高分辨率全基因组视图。

[0493] 实施例18

[0494] 此实施例中显示的高通量成像技术对于研究染色质组织具有几个优势。首先,与依赖邻近信息来推断染色质结构的基于高通量测序的方法相比,该方法提供了染色质组织的直接可视化,以及细胞核背景中个体成像基因座的空间位置和成像的基因座对之间的物理距离的直接测量。其次,该方法本质上是一种单细胞方法,并可以揭示个体细胞中详细的染色质结构。成像方法对个体染色质基因座的高(接近100%)检测效率允许成对染色质相互作用的高捕获率,其可以提供单个细胞中染色质结构的高分辨率视图。通过该方法测量的大量细胞允许对跨细胞的共同结构组织以及细胞间的变化进行稳健的统计分析。第三,染色质追踪技术可以容易地与其他成像方式相结合。这包括本研究中证明的多重转录成像和核结构成像,但也可以进一步扩展以包括其他模式,例如表观遗传修饰的成像或染色质可及性的程度。这种多模态成像可以为染色质结构、核组织和转录活动之间的关系提供关键见解。

[0495] 这里报道的高通量染色质成像方法有许多可能的应用。虽然在目前的工作中,跨染色体一致地靶向基因座以提供整体3D染色体和基因组组织的无偏视图,但这种方法也可用于靶向具有特定结构和功能特性的基因组基因座。一个有趣的方向是靶向含有特定基因或调控序列的基因座,或与特定核结构蛋白(例如CTCF或黏连蛋白)结合的基因座,以研究这些基因座之间的相互作用及其与转录的关系。作为一个更具体的例子,可以靶向大量潜在的启动子和增强子,并且可以研究它们的相互作用,同时对同一细胞中由启动子控制的基因的转录活动进行成像。这将允许推断哪些增强子控制哪些启动子,并揭示控制启动子和增强子的网络如何差异相互作用以调节转录的规则。在另一个方向上,据报道,许多转录因子和相关蛋白以及非编码RNA参与了在细胞核内组织染色质的物理凝聚物,这进而可能对基因表达调控很重要。与这些因子形成的结构以及转录输出同时成像染色质组织将提供有希望的途径以破译染色质结构、多组分组装、凝聚物形成和转录调控之间的关系。此外,不同的细胞类型表现出不同的基因表达谱,其可能部分受到3D基因组组织的调节。因此,成像染色质组织以及组织中个体细胞的基因表达谱有望为对细胞类型特异性基因表达模式很重要的染色质组织提供重要见解。

[0496] 实施例19

[0497] 此实施例说明了上述某些实施例中使用的某些实验模型和主题细节。

[0498] 分析中使用的细胞培养物和细胞系。细胞的制备与以前的研究类似。IMR-90细胞购自美国典型培养物保藏中心(ATCC,CCL-186),并根据推荐的方案进行培养。

[0499] 寡核苷酸探针设计:靶基因组区域的选择。对于通过依次杂交的高分辨率全染色体成像,首先将靶标染色体分成50kb区段。在筛选出重复元素和其中每50-kb区段可以设计<100个独特探针的区域后,针对Chr21总共保留了651个靶基因组基因座和针对Chr2总共保留了4,500个靶基因组基因座。然后为每个50-kb区段设计初级探针(约500个寡核苷酸探针),并保留每个区段的350个最中心定位的探针用于连续成像。对于Chr21成像,对所有651个基因组基因座进行成像。对于Chr2成像,靶标是250-kb基因组分辨率,因此仅为每五个50-kb区段中的一个设计初级探针。

[0500] 为了对Chr21上的新生RNA转录物进行成像,选择了对于其可以在来自Chr21上所

有蛋白质编码基因的第一个内含子上设计>50个初级探针的基因(参见下面的“初级/编码探针设计”部分)。总共选择了散布在Chr21中的86个基因。为了促进准确检测转录起始事件的空间位置,设计了靶向每个基因的转录起始位点(TSS)周围的5kb DNA区段的探针。

[0501] 对于通过组合成像策略进行的基因组规模染色质成像,基因组基因座被选择用于以下方式成像。对于每个人类染色体(Y染色体除外),每约3Mb间距选择一个30-kb区段。如果该间距导致给定染色体上的选定基因座少于30个,则该染色体的间距减小,直到所有染色体都至少选择了30个基因座。这导致共有1,041个靶基因组基因座用于成像,并且个体染色体中的基因座数量范围为30-80。然后为每个30-kb区段设计编码探针(约400个寡核苷酸探针)用于组合FISH成像。

[0502] 为了在基因组规模成像中对新生RNA转录物进行成像,选择了与1,041个靶向基因组基因座完全或部分重叠的所有包含内含子的基因。选择了针对所有这些RNA的内含子的编码探针,使得每个RNA有约20个编码探针,并且编码探针的靶向序列保持尽可能靠近转录起始位点。共有1,137个基因被靶向。

[0503] 通过组合FISH进行基因组规模成像的条形码设计。以下列方式选择用于对1,041个基因组基因座进行成像的二进制条形码。首先,生成所有可能的Hamming权重为2的100位二进制条形码(即每个条形码包含两个“1”位和98个“0”位),并从该列表中随机选择1,041个条形码。然后首先将选定的条形码任意分配到1041个基因组基因座。接下来,在使用和未使用的代码库之间以及来自不同染色体的基因座之间随机交换条形码,以最小化每个染色体在不同位上出现(即读取“1”)的基因座的数量差异。这导致每个染色体的每个位成像的基因座的大致相等的数量。为了优化条形码与每个染色体基因座的关联,允许同一染色体内的基因座交换条形码,并优化在相同代码位置具有读取“1”的条形码的基因座之间的最大的最小基因组距离。当比较具有相同的最小基因组距离的代码分配时,选择最小化基因组距离的变异系数的代码分配(使得基因组距离具有更大的平均值和更小的标准偏差)。

[0504] 类似地选择了用于对1,137个基因的新生RNA转录物进行成像的条形码,但使用54位Hamming距离2代码而不是100位Hamming距离2代码。

[0505] 初级/编码探针设计。用于染色质成像的初级/编码探针是从购自Twist Biosciences的寡核苷酸库合成的。该库中的每个寡核苷酸使用以下子序列(从5'到3'):用于PCR扩增和逆转录(RT)的20-核苷酸(nt)或19-nt正向引发区域,对应于在顺序成像情况下探针靶向的基因组基因座或在组合成像情况下探针靶向的基因组基因座将被成像的位之一的20-nt读出序列,42-nt或40-nt靶序列(分别用于顺序或组合成像),旨在唯一地结合单个靶向基因组基因座,上述20-nt读出序列的额外1-2个拷贝,以及用于PCR扩增的20-nt或19-nt反向引发序列。

[0506] 具有微小的修改的类似设计用于新生RNA成像。如前所述,从先前生成的针对PCR优化的随机20-nt序列列表中选择正向和反向引发序列。

[0507] 通过以下过程选择读出序列。首先,如前所述,创建了与人类基因组同源性最小的30-nt序列列表。然后,通过观察到的信噪比(SNR)对这些序列的一个子集进行排序,并选择前100个作为DNA读出探针。对于顺序成像,由于更大量的杂交轮次,需要显著更多的读出序列。因此,遵循之前概述的相同程序来选择约1,200个候选读出序列。然后过滤这些候选物

以确保40-60%的GC含量和57-67摄氏度的熔解温度。使用BLAST进一步过滤这些序列,使得没有读出序列具有HSP评分大于或等于17的命中。最后,通过反向互补这些序列的每一个的最后20-nt来选择读出序列。

[0508] 42-nt或40-nt靶序列的选择与之前描述的程序类似。简而言之,对每个感兴趣的基因组区域重复以下程序(参见上面的“靶基因组区域”部分)。首先,创建了与感兴趣的基因组区域互补的所有42-nt或40-nt序列的列表(从靶区域中的每个可能碱基开始)。然后,通过要求它们处于定义的熔解温度和GC含量范围内来过滤序列。然后使用与先前相同的程序,通过限制与人类基因组、人类转录组和包含重复序列的数据库的允许的同源程度来进一步过滤剩余的序列。用于通过依次杂交的全染色体成像的序列具有使用BLAST的额外过滤步骤,其中确保每个靶序列与预期的基因组基因座唯一匹配。最后,在最后的过滤步骤之后从剩余的序列中选择靶序列,使得任何一对靶序列之间不存在基因组重叠。

[0509] 对于通过依次杂交的全染色体成像,每个靶基因组基因座的所有42-nt靶序列都与与该基因座相关的独特读出序列相匹配。为了生成初级探针序列,将每个靶序列连接到分配的读出序列的两个相同拷贝,然后连接到正向和反向PCR引物。为了通过组合FISH生成用于基因组规模成像的全长编码探针,为每个靶基因组基因座选择的每个40-nt靶序列交替分配到跨越整个靶基因座的2个组。这些组中的每一个都与单个读出序列相关联,对应于其中基因座将被成像的两个位中的一个。然后,将每个靶序列连接到分配给其组的读出序列的两个相同拷贝,然后连接到正向和反向PCR引物。

[0510] 类似地设计用于RNA成像的探针,除了它们在每个探针上包含相同读出序列的3个拷贝,一个在5'端,两个在靶区域的3'端。用于RNA成像的读出序列与用于DNA成像的那些是正交的,并且是从测试的读出序列的相同排序的列表中选择。

[0511] 实验系统概述。用于执行这些实验的物理设置使用了几个组件。使用定制的荧光显微镜来获取图像,而定制的流体系统用于在显微镜台上自动执行缓冲液交换。定制软件用于同步和控制各种组件,并使许多实验步骤自动化。以下是这些元件中的每一个的详细描述。

[0512] 用于图像采集的显微镜设置。使用定制的显微镜系统进行图像采集。该系统围绕带有1.4NA的Nikon CFI Plan Apo Lambda 60x油浸物镜的Nikon Ti-U显微镜主体构建。照明基于以下两种选择之一:具有以下波长的固态单模激光器:405nm(Coherent,0bis 405nm LX 200mW),560nm(MPB Communications,2RU-VFL-P-2000-560-B1R)、647nm(MPB Communication,2RU-VFL-P-1500-647-B1R)和750nm(MPB Communication,2RU-VFL-P-500-750-B1R)。在这种情况下,560-nm、647-nm和750-nm激光器的输出通过声光可调谐过滤器(AOTF)控制,而405-nm激光器通过其激光控制箱直接控制。使用定制的二向色过滤器(Chroma,zy405/488/561/647/752RP-UF1)和发射过滤器(Chroma,ZET405/488/461/647-656/752m)来分离激发和发射照明。或者,具有以下波长的Lumencor CELESTA光引擎(基于光纤耦合固态激光器的照明系统):405nm、446nm、477nm、520nm、546nm、638nm和749nm。该系统与五带通二向色过滤器(IDEX,FF421/491/567/659/776-Di01-25x36)和五带通过滤器(IDEX,FF01-441/511/593/684/817-25)一起使用。在大多数实验中,使用折射光束整形器(Newport Optics,GBS-AR14)或振动光纤(Errol,custom Albedo unit)使照明变平。

[0513] 使用科学CMOS相机(具有针对单分子成像进行工厂校准的Hamamatsu FLASH4.0或

Hamamatsu C13440) 进行图像采集。使用XYZ载物台 (Ludl) 控制三个维度的样品位置。定制的自动对焦系统用于长时间保持恒定的焦平面。这是通过比较从玻璃-流体界面反射并在单独的CMOS相机 (Thorlabs, uc480) 上成像的两个IR激光器 (Thorlabs, LP980-SF15) 光束的相对位置来实现的。

[0514] 使用National Instruments数据采集卡 (NIPCIe-6353) 和定制软件 (参见下文“用于控制实验组件的软件”) 对不同的组件进行同步和控制。

[0515] 流体系统配置。流体系统使用了几个主要组件: 一个泵、一组串联的阀门、一个其中安装样品的流动室, 以及管道和连接器。蠕动泵 (Gilson, MINIPLUS3) 用于在系统中产生流动。该泵与一系列串联的8通阀 (Hamilton, MVP和HVXM 8-5) 相连。在这项研究中, 使用了3-5个以这种方式连接的阀门。每个阀门的最后一个连接被用作该系列中下一个阀门的输入 (最后一个除外), 而其余的连接含有缓冲液的管用于单轮杂交。阀门的一个固定子集用于成像、漂白和洗涤缓冲液 (参见“实验程序和方案”部分)。该阀门系统用于将各种缓冲液流入其中放置样品的流动室 (Bioprotechs, 060319-2)。腔室输出连接到废物收集容器, 形成开放式流动系统。使用弹性塑料管连接组件, 并使用压力粘合剂 (Blu-tack) 密封连接。使用定制软件控制系统 (参见下文“控制实验组件的软件”)。总体而言, 该系统允许进行20-36轮杂交 (取决于阀门的数量和为特殊缓冲液保留的点数)。在杂交轮次数量超过流动系统容量的实验中, 通过以下程序用新的缓冲液更换缓冲液: 阀门系统的输出直接连接到废物收集容器, 绕过样品容纳室。然后使用30%甲酰胺和双蒸水洗涤所有阀门。接下来, 引入了新的一组缓冲液, 并将腔室重新连接到流动系统。最后, 继续进行下一轮杂交实验。

[0516] 用于控制实验组件的软件。所有系统组件均使用定制软件进行控制。该软件包由协同工作的以下几个主要模块组成: “Hal”, 其是用于控制和同步所有照明和显微镜组件的软件包。值得注意的是, 在一些情况下, 有必要为组件编写驱动程序, 这些驱动程序不包含在此包中。Hal还用于定义成像参数, 例如照明强度、成像期间的阶段顺序和照明操作 (例如在z扫描期间)、曝光时间等。“Steve”, 其是用于拍摄镶嵌图像 (即由许多单独的视野组成的合成图像) 和选择区域进行实验成像的模块。“Kilroy”, 其是用于控制流体组件的软件, 并定义了预编程的操作序列以作为集合执行 (例如, 当执行新一轮杂交时发生的操作集合)。“Dave”, 其可以向Hal和Kilroy发出命令, 并用于通过预先定义完整的一组流体系统和显微镜操作、执行它们的顺序和间隔来自动执行数据收集。

[0517] 实验的一般流程是, 在实验开始之前, 用要使用的参数和规格加载Hal和Kilroy。加载样品并用成像缓冲液填充腔室后, 使用Steve拍摄DAPI通道的镶嵌图像, 并选择感兴趣的区域。然后生成一个文件以指定整个实验中的操作顺序, 并与所选感兴趣区域的坐标一起加载到Dave。实验的其余部分自动运行, 无需人工干预。如果实验中的轮次数量超过了流动系统的容量, 则自动序列指定达到系统的容量的动作。然后替换缓冲液 (参见上面的“流体系统配置”部分), 创建一个新的Dave文件, 并重复此操作, 直到完成所有成像轮次。

[0518] 初级/编码探针合成。从上述模板库中扩增初级/编码探针 (参见上文“初级/编码探针设计”)。这是使用先前描述的涉及以下步骤的扩增方案完成的: 首先, 使用有限循环PCR将初始寡核苷酸库扩增约20个循环。此步骤中使用的反向引物还通过引物延伸引入了T7启动子序列。然后, 所得产物通过柱纯化进行纯化, 并通过高产率的体外转录反应进一步扩增和转化为RNA。接下来, 通过逆转录反应将RNA产物转化回单链DNA。然后, 对前一步骤的

产物进行碱水解(以去除残留的RNA)并进行柱纯化(DNA Clean&Concentrator Kit,Zymo Research D4003和D4033)。最后,如有必要,将前一步骤的产物在真空中干燥并重新悬浮在水中以达到所需的初级探针浓度。所有引物均购自Integrated DNA Technologies(IDT)。

[0519] 细胞培养物制备和初级/编码探针杂交。细胞的制备与之前的研究类似。IMR-90细胞购自美国典型培养物保藏中心(ATCC,CCL-186),并根据推荐的方案进行培养。为避免染色质结构的潜在改变,本研究中的所有细胞均在培养开始后6周内以下述密度铺板。

[0520] 为了准备用于DNA成像,将细胞以每个盖玻片约500,000个细胞的密度铺在40毫米、圆形#1.5盖玻片(Bioptechs,0420-0323-2)上。在37℃和5%CO₂条件下,让细胞生长约2天直至汇合。在转录抑制实验中,在细胞固定前6小时,将细胞培养基更换为含有100微克/mL α -鹅膏蕈素(Sigma-Aldrich,A2263)的新鲜培养基。对于1,6-己二醇(Sigma-Aldrich,240117)的实验,我们在细胞铺板前用10微克/mL纤连蛋白(Sigma-Aldrich,F1141)涂覆盖玻片,并用含有2%w/v 1,6-己二醇的新鲜培养基替换培养基,持续45分钟。然后将培养物使用PBS中的4%多聚甲醛(PFA)在室温下固定10分钟,并在PBS中洗涤2-3次。然后分两步使细胞透化:首先,在室温下用PBS中的0.5%v/v Triton-X(Sigma-Aldrich,T8787)处理细胞10分钟。然后,在室温下用0.1M盐酸(HCl)处理细胞5分钟并在PBS中洗涤2-3次。在HCl处理后,用溶解在PBS中的0.1mg/mL核糖核酸酶A(ThermoFisher,EN0531)的溶液在37℃下处理细胞30-45分钟,以去除与RNA脱靶结合的潜在来源。在此处理之后,将细胞在预杂交缓冲液中孵育(使用2x盐水-柠檬酸钠缓冲液(SSC;Ambion,AM9763)和50%甲酰胺(Ambion,AM9342))约10分钟。接下来,将细胞盖玻片倒置并置于60毫米培养皿中的一滴50微升杂交缓冲液中(2xSSC、50%甲酰胺、10%硫酸葡聚糖(Sigma-Aldrich,D8906),含有添加或不添加10微克人类Cot-1 DNA(ThermoFisher,15279011)的约25微摩尔总浓度的初级/编码探针混合物)。将培养皿部分浸入约90℃的水浴中3分钟,然后在47℃的加湿室中孵育16-36小时。与初级/编码探针孵育后,将样品在2xSSC和40%甲酰胺中洗涤30分钟,然后在室温下用2xSSC中的4%PFA后固定10分钟。然后将样品与基准珠(ThermoFisher F8805或ThermoFisher F8792)在2xSSC中孵育2-3分钟,并在2xSSC中用1微摩尔4',6-二脒基-2-苯基吲哚(DAPI;ThermoFisher D1306)染色5-10分钟,然后存储在2xSSC中直到成像。

[0521] 对于包括RNA成像的实验,从固定细胞时开始使用的所有缓冲液都含有1:10-1:1,000稀释的核糖核酸酶抑制剂(NEB M0314或Fisher Scientific N2615)。RNA染色的处理与上述方案相同,直到用HCl处理。在此步骤之后,将细胞在预杂交缓冲液中孵育10分钟,然后将细胞盖玻片倒置并置于一滴杂交缓冲液上,该缓冲液含有以约1微摩尔总浓度的靶向RNA内含子的初级/编码探针,如前文针对DNA染色所述的。然而,在这种情况下,没有进行90℃的热变性,细胞立即在47℃下在加湿室中孵育16-36小时。在与初级/编码探针一起孵育后,将样品在甲酰胺溶液中洗涤,并如上文针对DNA所述用PFA后固定。然后将其与基准珠一起孵育并用1微摩尔DAPI染色,然后储存在2xSSC中直到成像。RNA成像后,将样品从显微镜中取出,用核糖核酸酶A处理细胞,然后DNA杂交以与上文对于没有RNA成像的DNA成像所述相同的方式进行。

[0522] 用于顺序或组合FISH成像的读出探针的依次杂交。方案的这一部分中的所有流体交换都是通过使用定制的流体系统实现的,盖玻片安装在FCS2流动室(Bioptechs,060319-2)中。该系统的设置在“流体系统配置”部分进行了详细描述。简而言之,流体系统使用3-5

个计算机控制的八通阀 (Hamilton, MVP和HVXM8-5) 和计算机控制的蠕动泵 (Gilson, MINIPLUS3)。放在一起, 这些组件允许控制流体流动的速率和在任何给定时间流动的流体的类型。杂交的每个轮次使用以下一般步骤: 首先, 杂交缓冲液与一组对每个轮次特异的寡核苷酸探针一起流入, 如下文所讨论的。然后, 在室温下孵育10分钟。接下来, 使洗涤缓冲液流过, 并孵育约200秒, 最后, 使成像缓冲液流过。

[0523] 如前所述制备成像缓冲液, 并使用60mM Tris pH 8.0、10%w/v葡萄糖、1%葡萄糖氧化酶除氧剂溶液 (含有~100mg/mL葡萄糖氧化酶 (Sigma-Aldrich, G2133) 和1:3稀释的过氧化氢酶 (Sigma-Aldrich, C3155))、0.5mg/mL 6-羟基-2,5,7,8-四甲基色满-2-羧酸 (Trolox; Sigma-Aldrich, 238813) 和50μM Trolox Quinone (通过Trolox溶液的UV辐照产生)。将Trolox溶解在甲醇中, 然后添加到溶液中。制备后, 成像缓冲液被一层约0.5厘米厚的矿物油覆盖, 以防止暴露于氧气中。

[0524] 杂交缓冲液和洗涤缓冲液分别由2xSSC中的35%和30%甲酰胺组成, 杂交缓冲液还含有0.01%v/v Triton-X。杂交缓冲液对于每个杂交轮次分开保存, 并包含两组 (用于通过组合FISH进行基因组规模染色质成像) 或三组 (用于通过依次杂交进行全染色体成像和通过组合FISH进行基因组规模染色质和RNA成像) 读出探针。通过以下方式引入荧光信号: 对于通过依次杂交进行的全染色体成像, 杂交缓冲液包含以30nM浓度添加的三种荧光读出探针 (Alexa750、Alexa647或Cy5和Cy3)。所有实验都涉及寡核苷酸衔接子和荧光读出探针的依次杂交, 如图24A中所述。在杂交的每个轮次期间, 首先流入一组衔接子探针 (每个浓度为100nM), 以检测三个不同颜色通道 (Alexa750、Alexa647或Cy5和Cy3) 中的三个靶向的基因组基因座。每个衔接子探针由与靶基因座之一特有的读出序列互补的区段和包含颜色通道特定的共同读出序列的区段组成。接下来, 三个不同的染料缀合读出探针 (每个与三个颜色通道特定的共同读出序列之一互补) 以每个探针30nM浓度流入。该程序允许在每个杂交轮次期间分别在三个颜色通道中对三个基因组基因座进行成像。如前所述, 荧光读出探针包含将荧光团连接到寡核苷酸的二硫键, 以允许在轮次之间有效去除信号。对于组合FISH的基因组规模染色质成像, 每个轮次的杂交缓冲液包括两种荧光读出探针, 一种用Cy5或Alexa647标记, 另一种用Alexa750标记。荧光读出探针使用: 1) 与在给定位点中成像的所有编码探针共有的读出序列互补的荧光标记的寡核苷酸, 以100nM浓度添加, 或2) 连接到附加的共有读出序列 (每个颜色通道中的所有衔接子共有) 的具有与读出序列互补的序列的衔接子寡核苷酸的组合, 如上所述, 以及与该共有读出序列互补的荧光标记的读出探针。对于一些实验, 衔接子和共有读出探针以1:1.5的比例预混合, 并添加到~100nM的最终浓度。对于其他实验, 衔接子和读出探针依次与样品杂交。对于RNA成像, 每个轮次的杂交缓冲液包含三种衔接子探针, 每个颜色通道一个, 如上所述。每个轮次都包括两个离散的杂交步骤-首先将衔接子流入、杂交, 然后洗涤多余的材料。然后依次流入与衔接子上的共同读出序列互补的分别用Cy3、Cy5 (或Alexa647) 和Alexa750标记的三种荧光读出探针。在荧光读出杂交后, 流入成像缓冲液并收集信号。

[0525] 在下一轮次的读出探针或衔接子探针杂交之前, 去除当前轮次中来自读出探针的荧光信号, 如下面的“杂交轮次之间的信号去除”部分所述。

[0526] 在第一轮杂交之前, 进行一轮成像以获取DAPI信号并鉴定核边界。对于通过依次杂交进行的全染色体成像, 在约220轮中对Chr21上的651个基因组基因座或在约320轮中对

Chr2上的935个基因组基因座进行成像,全部在3个颜色通道中。对于通过组合FISH进行的基因组规模染色质成像,整个组的1,041个基因组基因座在50轮杂交和每个轮次2个颜色通道中成像。在每个轮次中,基因组基因座通过在z维度上步进,以3D形式成像。Chr21上86个基因的新生RNA转录物在31轮中以3种颜色顺序成像,1,137个基因的RNA转录物的基因组规模成像在18轮中以3种颜色进行3D成像。特别是对于Chr21的连续成像,29轮中86个基因的TSS以3种颜色成像。额外的轮次用于重新标记基因组基因座的集合,并评估颜色通道之间的色差和渗色,以及样品和成像仪器的稳定性。对包含总共约1,000-3,000个细胞的约60个视野的成像,对于通过依次杂交对全染色体成像的顺序成像花费约12-18天,并且对于通过组合FISH的基因组规模染色质成像花费3天。

[0527] 3-5阀系统允许加载多达20-36种不同的杂交溶液。结果,在耗尽所有流体系统的通道后,样品室被绕过,并且用于杂交的所有通道都用30%的甲酰胺水溶液洗涤。接下来,腔室被重新连接,并进行下一组杂交和成像轮次。

[0528] 对于通过依次杂交的全染色体成像,在2xSSC中使用2%PFA定期(每约4天)进行温和的固定后步骤,持续5分钟,以保持样品的结构完整性。

[0529] 抗体标记和成像。在RNA或DNA成像后立即进行抗体成像。通过上述方案完成成像后,样品经历以下步骤:样品与封闭溶液(含0.1%v/v Tween-20(Sigma-Aldrich P9416)和1%w/v牛血清白蛋白(BSA;Jackson ImmunoResearch 001-000-162)的PBS))孵育30分钟。将样品与在封闭溶液中稀释的一抗孵育1小时。样品在含有0.1%Tween-20的PBS中洗涤3次,每次5分钟。对荧光标记的二抗重复步骤2和3。

[0530] 所有缓冲液交换均在显微镜上完成,使用下面的“流体系统设置”部分中描述的微流体系统。Cy5颜色通道用于成像,并使用光漂白消除连续抗体标记之间的信号。

[0531] 使用了一抗和二抗的以下组:为了对核斑点进行成像,使用了从原液1:200稀释的针对SC35的一抗(Abcam,ab11826)-一种通常用作核斑点的标记物的剪接因子,和从原液浓度1:1,000稀释的用Cy5染料标记的驴抗小鼠二抗(Jackson ImmunoResearch,715-175-150)。对于核仁成像,使用从原液1:200稀释的抗纤维蛋白抗体(Abcam,ab5821),以及从原液浓度1:1,000稀释的用Alexa 657染料标记的驴抗兔二抗(Jackson ImmunoResearch,711-605-152)。对于细胞周期状态测定,在RNA成像后立即使用从原液1:100稀释的抗双联蛋白抗体(Abcam,ab195047)以及从原液浓度1:1,000稀释的用Alexa 657染料标记的驴抗兔二抗(Jackson ImmunoResearch,711-605-152)进行免疫荧光染色。

[0532] 图像采集。对于每个实验,选择大约60个视野(FOV)用于成像,避开细胞稀疏的区域(我们通常每个FOV鉴定10-50个细胞)。每个相机FOV使用1,000x1,000像素,相机像素对应于成像平面中每个维度的153nm,或2048x2048像素,相机像素对应于成像平面中每个维度的108nm。

[0533] 在每个轮次的杂交后(参见上文“用于FISH成像的读出探针的依次杂交”),以3或4种颜色获取每个FOV的z堆栈图像:647nm和750nm照明(或560nm、647nm和750)用于获取FISH图像,使用560nm照明(或405nm照明)对基准珠进行成像。对于第一轮成像,使用405nm照明对DAPI信号进行成像,而对于抗体成像,在RNA或DNA成像后使用647nm激发通道。连续的z切片相隔85、100、150或200nm,覆盖所有成像的细胞的整个核体积。在每个z位置,在移动台之前在所有通道中获取图像,并以~10Hz的速率获取图像。

[0534] 杂交轮次之间的信号去除。在每个轮次的成像之前,来自前一轮的信号(或内源性背景,在第一轮的情况下)被熄灭。这是如前所述使用任选的光漂白步骤通过裂解连接荧光团和读出探针的二硫键来实现的。用于裂解的缓冲液含有50mM三(2-羧乙基)膦(TCEP; Sigma-Aldrich, C4706)以还原将荧光团连接到读出探针的二硫键,以及35%甲酰胺中的1mM无染料共有读出探针以阻止任何未占用的读出序列干扰下一轮杂交。在进行光漂白的实验中,伴随着将缓冲液更改为含或不含35%甲酰胺的2xSSC,并以560、647和750激光器的最大可用功率照明每个视野3-10秒。连续追踪实验中的光漂白步骤是在与寡核苷酸衔接子的杂交步骤中完成的,以最小化总实验时间。由于杂交和洗涤缓冲液中的高甲酰胺浓度,DAPI信号消失。

[0535] 在顺序DNA-FISH成像中重新标记基因组区域。在完成整个染色体的顺序DNA-FISH成像实验后,对区域的子集进行重新标记和重新成像。样品用2xSSC中的57%甲酰胺处理4分钟,重复此处理3次以剥离读出寡核苷酸探针(在第一轮成像后通过首先使用TCEP从寡核苷酸探针裂解染料然后通过光漂白来去除其荧光信号)。剥离读出探针后,在35%甲酰胺2xSSC中添加1mM无染料共有读出探针,以阻断未剥离的衔接子探针上的任何未占用的读出序列。接下来,通过遵循标准读出探针杂交方案(在“用于顺序或组合FISH成像的读出探针的依次杂交”部分中描述)来实现选定区域的重新标记。

[0536] 图像分析:分析管道概述。本研究中使用的图像分析管道是用Python实现的。整个管道使用以下步骤:鉴定和分割所有成像的细胞核,将3D高斯拟合到用于DNA或RNA成像以及基准珠的成像通道中所有检测到的荧光点。拒绝与已鉴定的细胞核不重叠的DNA和RNA点,使用基准珠校正样品漂移,校正不同颜色通道之间的色彩效应,并使用自定义算法和软件为DNA基因座和RNA分子分配身份(这些在下面针对DNA和RNA成像单独描述,用于全染色体和用于全基因组成像)。

[0537] 核分割。来自第一轮成像的DAPI图像用于鉴定个体细胞核的体积并允许进行细胞分割。这是通过卷积神经网络实现的,与之前发表的工作类似地构建和训练,其将DAPI图像在xy平面上的最大投影作为输入。

[0538] DNA和RNA成像的点拟合。以下分析管道应用于每个成像的FOV,以获得所有感兴趣的基因座的三维(3D)位置:在所有成像轮次中都拟合了基准点并将其用于图像对齐(参见下面的“漂移校正”部分)。在第一个成像轮次中(在第一轮杂交之前),DAPI信号用于鉴定个体细胞核的边界,以及用于RNA和DNA成像之间的图像配准。有关详细信息,参见“细胞核分割”和“DNA和RNA成像之间的图像配准”部分。将每个已鉴定的核内的衍射限制点拟合到3D高斯函数,以鉴定它们的质心和高于局部背景的亮度。为了使分析更易于管理,在通过组合FISH的基因组规模成像中,将保留用于解码的每个图像的拟合点数固定为125个或更少(比预期无噪声的不同基因座的数量大约3倍)。对于通过依次杂交进行的全染色体成像,每个图像的每个染色体的拟合点数固定为6个或更少。然后将来自步骤3的拟合点用于鉴定DNA基因座和转录灶并确定它们的位置,如下面的相应部分所述。

[0539] 漂移校正。以与上述相同的方式进行基准珠点拟合。然后在多个轮次的杂交之间比较一组基准珠位置,并应用刚性变换以最小化珠子的相对位置的平方差之和。

[0540] 校正色彩效果。通过独立标记每个成像通道中的同一组基因组基因座并分别比较不同颜色通道中相同基因座的信号来执行多色成像的渗色和色差。

[0541] DNA和RNA成像之间的图像配准。DAPI信号首先用于通过2D图像相关性在两组图像(即染色质和RNA)之间进行粗略的图像配准(每组中的所有图像都使用基准珠与DAPI图像对齐)。在执行了第一轮RNA解码后(参见下面的“在序列成像中从拟合的RNA点鉴定转录灶”和“在组合基因组规模成像中拟合的RNA点的解码算法”),通过假设当考虑到视野中所有成像的基因和细胞时新生RNA定位与含有其的DNA基因座之间的位移应该平均为零来计算更精细的比对。相应地,计算了额外的刚性变换,以最小化成像的新生RNA与其相应DNA基因座之间的平均位移,并将其用作最终比对。

[0542] 在序列成像中从拟合的DNA点鉴定染色质基因座。每个基因座的鉴定和3D定位是通过以下步骤实现的:为每个图像中所有拟合点的漂移和像差校正位置生成一个列表。因为点查找算法允许在对应于特定杂交轮次的特定颜色通道的每个连续图像中为每个染色体找到最多6个候选者(参见“DNA和RNA成像的点拟合”),执行了以下附加步骤以鉴定最有可能源自成像的染色质基因座的候选点。通过选择对应于每个杂交轮次的每个颜色通道中每个细胞中的每个染色体拷贝的最亮点来生成初始的暂定染色质迹线。对于每个拟合的点,无论是否选择它用于初始暂定染色质迹线,都计算了三个质量度量:局部背景之上的点亮度,点到局部质心的距离,其是从沿着暂定染色体迹线的上游的五个基因座和下游的五个基因座以及到整个暂定染色体迹线的质心的距离计算的。对于每个点,上述三个质量度量通过针对暂定染色质迹线中包含的点的质量度量值的分布(我们将其称为“有效分布”)计算每个候选点的组合Fisher p值组合成单个测量值。这可以被认为每个点的整体质量评分,并按以下方式每个点进行计算:对于每个度量,计算“有效分布”中具有较低质量度量的其他点的分数,将这三个分数相乘。然后使用期望最大化程序依次选择与每个靶染色质基因座相对应的具有最高质量评分的点,并基于得到的染色质迹线重新更新“有效分布”。重复这个优化过程直到收敛。收敛后,最后一组点(每个对应于一个染色质基因座)用于确定靶基因座的3D空间位置。最后,过滤来自步骤4的点,以去除质量评分低于设定截止值(因此具有低置信度)的点。为了设置组合质量评分的截止值,首先计算重新成像实验中包含的基因座的质量评分以确定位移误差(参见下面的“序列成像中重新成像的基因座的鉴定”部分)。然后,计算原始之间具有低位移误差($<500\text{nm}$)的点的质量评分的分布,并对具有高位移误差($>500\text{nm}$)的点重新成像。最后,设置质量评分阈值,使得最终染色质迹线(应用阈值后)中预期属于高位移错误类别的基因座分数 $<5\%$ 。步骤5后的剩余点用于确定染色质基因座的最终位置并追踪染色质结构。

[0543] 在序列成像中从拟合的RNA点鉴定转录灶。使用以下程序分析来自RNA成像轮次的信号:首先,首先使用基于DAPI的粗略对齐和距离保留的相应DNA基因座 1000nm 的距离内最亮的RNA点对每个细胞的拟合的RNA点的位置进行色差和漂移校正。然后,基于最初选择的RNA定位(来自步骤1和包含它们的DNA基因座的位置之间的位移对DNA和RNA成像之间的配准进行细化,如上文“DNA和RNA成像之间的图像配准”部分所述。最后,在步骤2的精细配准后,将所有候选RNA点的位置与包含该基因的 50kb DNA基因座的位置和相应的 5kb DNA转录起始位点进行比较。应用了距离 50-kb 基因座或 5-kb 转录起始位点 500nm 的更严格距离截止值,以及1的信噪比阈值。如果新生RNA定位通过这两个阈值,则它被认为是检测到的转录爆发。

[0544] 顺序成像中重新成像的基因座的鉴定和位移误差的估计。重新成像的基因座的鉴

定与“从顺序成像中拟合的DNA点鉴定染色质基因座”部分中的描述类似,除了重新成像的基因座用于替换原始成像轮次中的相应基因座子集。

[0545] 为了计算原始和重新成像的轮次之间的位移误差(图24B-24D),仅考虑基于以下观察结果的通过亮度阈值的基因座。值得注意的是,重新成像轮次中的荧光信号比原始成像轮次中的荧光信号要暗得多。这可能是由于原始读出探针的不完全去除和/或由于在甲酰胺处理期间部分去除了与基因组DNA结合的初级探针以剥离原始成像轮次中结合的荧光读出探针。较低的亮度降低了重新成像的基因座的定位精度,并导致人为地高估了初始成像轮次的定位误差。为了减轻这种影响,在估计定位误差时,只选择了与原始信号相比亮度大于20%的那些重新成像的基因座。

[0546] 作为附加说明,当检查距离两个相邻基因座 $>1000\text{nm}$ 的基因座时,其中一小部分($\sim 20\%$)表现出较大的重新成像位移误差并且具有较低的亮度。因此,远离它们的两个基因组邻居的基因座可能具有相对较低的置信度。

[0547] 组合基因组规模成像中拟合的DNA点的解码算法。每个基因座的鉴定和3D定位是通过以下步骤实现的:首先,为每个位图像中所有鉴定的点的漂移和像差校正位置生成一个列表(对应于特定成像轮次中的特定颜色通道)。对于每个位图像中的每个检测到的点,发现其他位图像中距其位置设定的截止距离(x 、 y 和 z 中约 150nm)内的所有点。所有这样的点对都被保留以供进一步分析,无论由点对产生的条形码(基于它们出现在其中的轮次和颜色通道)是否对应于有效的条形码(即分配给基因组基因座的条形码)。然后,对于每个点对,计算三个质量度量:两个点的3D定位之间的位移、两个点之间的亮度差异以及两个点的平均亮度。每个点的亮度通过相应位图像中所有点的中值亮度进行标准化。然后基于它们是否对应于有效的条形码(并因此可能对应于基因组基因座),将点对分为两组。在每组内,计算质量度量的分布。为方便起见,来自无效条形码的点对质量度量的分布被称为“无效分布”,来自所有有效条形码的被称为“有效分布”。对于每个点对,通过针对“有效分布”计算每个候选点对的组合Fisher p 值,将步骤3中的三个质量度量组合成一个单独的测量值。这可以被认为每个点对的整体质量评分,并按以下方式对每对进行计算:对于三个度量中的每一个,计算“有效分布”中具有较低质量度量的其他点对的分数并将这三个分数相乘。使用期望最大化程序依次选择与每个靶染色质基因座相对应的具有最高质量评分的两个点对并重新更新“有效分布”,并重复此优化程序直到收敛。收敛后,最后一组点对(每组对应于一个染色质基因座)用于确定基因座的3D空间位置。在步骤5之后,使用改进的K均值算法将属于同一染色体的染色质基因座分成两个同源物。与将点分成两组并最小化每组内的回转半径的标准K均值聚类算法相反,将组之间的点逐渐切换以首先最大化每个同源物中分配的点的分数,然后最小化每个同源物的回转半径。在分离两个同源物后,计算它们的质心和来自步骤2的每个点对距其亲本染色体的质心的距离。除了步骤3和重复的步骤3-6中考虑的3个度量之外,还添加了距离染色体中心的距离作为另一个质量度量。最后,对来自步骤7的点对进行过滤,以删除其质量评分与“无效分布”相似的对。步骤8后剩余的点对用于确定染色质基因座的最终位置并追踪染色质结构。

[0548] 组合基因组规模成像中拟合的RNA点的解码算法。使用以下程序对来自RNA成像轮次的信号进行解码:首先,为每个轮次的成像中所有已鉴定的点的漂移和像差校正位置生成一个列表。对于每个成像轮次中的每个检测到的点,发现了在距其位置设定的截止距离

内的其他轮次中的所有点,并且如果这些点对形成有效的条形码,则将它们保留为候选RNA爆发。然后,在初始图像配准(基于DAPI图像)和漂移和像差校正后,将这些候选RNA爆发中的每一个的位置与含有相关基因的DNA基因座的位置进行比较,如果它们在设定的阈值距离内,则保留它们。接下来,基于初始解码的RNA定位(来自步骤3)与包含它们的DNA基因座的位置之间的位移来精细化DNA和RNA成像之间的配准,如上文“DNA和RNA成像之间的图像配准”部分所述。最后,再次将所有候选RNA爆发的位置与包含它们解码的基因的DNA基因座的位置进行比较,这次使用经过精细化的图像配准。如果在这个阶段新生RNA定位在与含有其的DNA基因座的截止距离内,则它被认为是检测到的转录爆发。

[0549] 组合基因组规模成像中重新成像的基因座的鉴定。在组合基因组规模成像方法中,6号染色体上的靶基因组区域的一个子集被分配了探针,使得它们可以在组合成像之后使用连续的多色FISH单独重新成像。对于组合成像中这些基因座之一的每个解码实例,位移误差被估计为组合成像中确定的定位与连续重新成像轮次中最近点之间的距离。

[0550] 从免疫荧光成像中鉴定核体。通过对免疫荧光信号的强度应用阈值,从免疫荧光信号中提取核体(核斑点和核仁)的位置,从而产生鉴定高免疫荧光信号的像素化掩模。然后将其视为“包含”核体的一组像素化位置。

[0551] 从抗双联蛋白和DAPI图像确定细胞周期阶段。首先,经历有丝分裂的细胞通过目测消除并且不被考虑用于分析。接下来,使用双联蛋白免疫荧光信号和利用DAPI测量的核信号的组合,与之前的研究类似,将细胞分类为G1(低双联蛋白信号,低DAPI信号)或G2/S(高双联蛋白信号)。

[0552] 核纤层位置的估计。通过生成围绕给定细胞中所有已解码的染色质基因座的位置的最小3D凸包表面(使用Python的SciPy包)来估计核纤层的位置。

[0553] 空间距离。任何一对基因座之间的空间距离简单地计算为它们的拟合的3D高斯中心之间的欧几里得距离,乘以涉及相机像素和z步长与物理距离的适当比率。在距核体的距离的情况下,计算了距离所有已鉴定的核体“位置”的最小欧几里得距离或距离定义核纤层的凸包表面的最小距离。

[0554] 来自成像的邻近频率矩阵。为了计算任何给定基因座对之间的邻近频率,首先计数了小于设定截止距离的该基因座对之间的测量的距离的数量(本研究中为500nm,除非另有说明)。然后将该数字除以针对该对基因座测量的距离总数。截止距离是通过评估由一系列截止阈值产生的邻近频率矩阵与Hi-C接触矩阵之间的Pearson相关性,以及源自Chr21的成像和Hi-C数据的整体结构特征(例如TAD和隔室)的对齐来确定的。

[0555] 值得注意的是,对于200nm至800nm的截止距离,邻近频率图和Hi-C图之间的Pearson相关系数在0.82-0.88范围内基本保持不变,并在~400-500nm处达到最大值。此外,对于 ≤ 500 nm的截止阈值,源自成像数据的TAD边界与源自Hi-C数据的TAD边界以同样高的精度对齐,而对于大于600nm的截止距离,该对齐分解。在更大的规模,A/B隔室调用显示出400至600的截止距离的最高一致性水平。因此,400-500nm范围内的截止距离被认为对于TAD和A/B隔室的准确调用是最佳的,并选择500nm的截止距离用于所有分析。

[0556] 局部密度分析。为了计算每个解码的位置的A隔室和B隔室基因座的局部密度,计算了每个细胞的每对染色质基因座之间的空间距离。对于每个基因座,局部A/B密度比按以下方式计算:首先,设置了以每个A或B基因座为中心的高斯概率密度函数,标准偏差为

100nm (对于Chr21成像)、125nm (对于Chr2成像) 或500nm (对于基因组规模成像)。然后,该基因座处的总A密度然后计算为在全染色体成像中不包括其自身的所有A基因座的高斯概率密度函数值的总和。对于基因组规模的成像,基因座的总跨A密度是从所有跨染色体A基因座(即来自其他染色体的所有A基因座)中求和的。以类似的方式计算总B密度。最后,将A隔壁基因座的总密度除以B隔壁基因座的密度,以找到该基因座处的A/B密度比。跨A/B密度比以类似方式计算。

[0557] 来自成像数据的绝缘评分。之前已经为整体Hi-C定义了绝缘评分。类似的定义用于成像结果,并应用于计算单个细胞中个体染色体中相邻或非相邻结构域的绝缘评分。

[0558] 为了计算两个结构域之间的绝缘评分,通过考虑第一结构域内每对基因座之间的所有距离以及第二结构域内每对基因座之间的所有距离来计算结构域内距离分布。然后通过考虑位于不同结构域中的基因座对之间的所有距离来计算结构域间距离分布。然后将绝缘评分定义为所有结构域间距离的中值除以结构域内距离的中值。两个高度混合的结构域的绝缘评分接近1,而刚刚接触的结构域的绝缘评分为约2。

[0559] TAD调用的标准化绝缘评分。为了比较成像和Hi-C数据中的TAD调用,对上述绝缘评分定义进行了轻微修改,使得来自这些数据的绝缘评分适应相同的动态范围。对于来自成像数据的中值成对距离矩阵中的TAD调用,对于每个基因组基因座,选择具有固定窗口的上游和下游基因座(即,在所选基因座的每一侧的固定数量的基因座)。所选基因座的上游和下游的这两个染色质区域被视为两个结构域并计算绝缘评分,如上所述。然后将标准化的绝缘评分定义为区域间距离的中值与区域内距离的中值之间的差异,通过这两个中值的总和进行标准化。因此,标准化绝缘评分将始终介于0和1之间。根据标准化绝缘评分的这个定义,沿染色体应用滑动窗口来计算对应于基因组基因座的绝缘评分向量,并通过来自Scipy的标准峰值调用算法找到局部最大值,这些位置被视为TAD边界。类似地执行使用源自成像的邻近频率矩阵或Hi-C接触矩阵的TAD边界调用。

[0560] 染色体内的A/B分离评分。A/B分离评分量化了染色体内A和B基因座之间的空间分离水平。为了计算这个量,首先在每个染色体内将A密集体积在操作上定义为包含所有A基因座的3D空间,其中A密度评分在前2/3范围内。B密集体积以类似的方式在操作上定义。A和B密集体积的纯度度量被定义为这些体积内的所有基因座分别为A和B基因座的分数。最后,A/B分离评分被定义为A密集和B密集体积的纯度值的平均值。如果A和B基因座完全分离,则该分离评分将为1,并且具有完全混合的A和B基因座的染色体将具有大约0.5的分离评分。

[0561] 通过组合FISH估计RNA成像中的检测效率。转录爆发事件的检测效率的估计是通过以下方式进行的:首先,考虑所有含有其RNA内含子被成像的基因的靶向的基因组基因座。对于这些基因组基因座中的任何一个,如果基因被转录,它们相应的RNA信号应该出现在两个预定义的位中。知道这两个位中的每一个未被检测到的比率(p)允许推导出RNA的检测效率。鉴定了在其相应基因的两个预期位中的至少一个中与RNA信号共定位(约150nm内)的一组基因组基因座。然后,从步骤1中鉴定的全部染色质基因座集合中,与RNA信号共定位的基因座的分数(f)是从其基因的对应位中的仅一个(而不是两个位)确定的。根据测量的 f

(8.4%),其应该等于 $\frac{2p(1-p)}{1-p^2}$,估计 p (4.4%)。最后,使用等式 $\eta = (1-p)^2$ 计算了在两个位

中检测共定位信号的总体检测效率,发现为约92%。

[0562] Hi-C数据分析。IMR-90细胞的Hi-C数据是从吸管(straw)中获取和使用吸管加载的。为了鉴定个体染色体中的A/B隔室,遵循已建立的公开的方案。为了鉴定TAD,使用了“TAD调用的标准化绝缘评分”部分中描述的方法。为了将源自成像数据的邻近频率与Hi-C接触数进行比较,创建了以靶区域为中心的箱,并通过将更高分辨率Hi-C数据中的读取数量相加来获得这些箱的Hi-C数据。

[0563] CTCF和Rad21 ChIP-seq数据分析。CTCF和Rad21 ChIP-seq数据从ENCODE数据集下载,并通过UCSC Genome Browser Utilities转换为wig格式。靶向的基因组区段的读取计数被收集并通过相应的输入进行标准化。通过来自Scipy的标准峰值调用算法调用CTCF或Rad21 ChIP-seq信号富集在沿染色体的输入上的局部最大值。

[0564] 虽然已在本文中描述和说明了几个本公开内容的实施方案,但本领域普通技术人员将容易地设想用于进行所述功能和/或获得本文所述的结果和/或一个或多个有利方面的各种其它方法和/或结构,并且这样的变化和/或修改的每一种被认为在本公开内容的范围内。更通常地,本领域技术人员将容易地理解,本文所述的所有参数、尺度、材料和构型意欲为示例性的并且实际参数、尺度、材料和/或构型将取决于对其使用本公开内容的教导的具体一个或多个应用。本领域技术人员将认可或能够确定本文所述的特定的本公开内容的实施方案的许多等同物(通过使用不超过常规实验)。因此,应理解,前述实施方案仅通过举例的方式提出,并且在所附权利要求和其等同物的范围内,本公开内容可按与明确描述和要求保护的方式不同的方式来实施。本公开内容的涉及本文所述的各种个体特性、系统、物品、材料、试剂盒和/或方法。另外,如果这样的特性、系统、物品、材料、试剂盒和/或方法不是相互矛盾的,则两个或更多个这样的特性、系统、物品、材料、试剂盒和/或方法的任意组合包括在本公开内容的范围内。

[0565] 在本说明书和通过引用并入的文件包括冲突和/或不一致的公开内容的情况下,应以本说明书为准。如果通过引用并入的两个或更多个文件包含相互冲突和/或不一致的公开内容,则以较晚生效日期的文件为准。

[0566] 如本文中所定义和使用的,所有定义应当被理解为优先于词典定义、通过引用并入的文献中的定义和/或所定义的术语的一般含义。

[0567] 除非明确地指示与之相反,否则不定冠词“一种/一个(a)”和“一种/一个(an)”,如本文中在说明书和权利要求中所用,应当被理解为意指“至少一种/一个”。

[0568] 短语“和/或”,如在本文中在说明书和权利要求中所用,应当被理解为意指所连接的元素的“任一个或两者”,即元素在一些情况下结合地存在以及在其它情况下分离地存在。利用“和/或”列出的多个元素应当以相同的方式来解释,即所连接的元素的“一个或多个”。除通过“和/或”从句明确确定的元素外,还可任选地存在其它元素,无论与明确确定的那些元素相关还是无关。因此,作为非限定性实例,对“A和/或B”的提及,当与开放性措辞诸如“包含/包括”结合使用时,在一个实施方案中可仅指A(任选地包括除B外的元素);在另一个实施方案中可仅指B(任选地包括除A外的元素);在另一个实施方案中,可指A和B(任选地包括其它元素);依此类推。

[0569] 如本文中在说明书和权利要求中所用,“或”应当被理解为具有与如上定义的“和/或”相同的含义。例如,当在列表中分开各项时,“或”或“和/或”应当被理解为是包含性的,即包含至少一个,但也包括许多个元素或一系列元素中的多于一个,以及任选地,包括另外未

列出的项。只有明确地指明相反的术语,例如“……中的仅一个”或“……中的恰好一个”或当用于权利要求中时的“由……组成”将指包含多个元素或一系列元素中的正好一个元素。一般地,如本文中所用的术语“或”应当仅在冠有排他性的术语诸如“任一”、“……之一”、“……中的仅一个”或“……中的恰好一个”时被解释为表示排他性选择(即“一个或另一个但非两个”)。

[0570] 如本文中在说明书和权利要求中所用,关于一系列一个或多个元素的短语“至少一个”应当被理解为意指选自一系列元素中的任何一个或多个元素的至少一个元素,但不一定包括元素列表内明确列出的每一个元素的至少一个并且不排除元素列表中的元素的任何组合。该定义还允许可任选地存在除短语“至少一个”所指的元素列表内明确确定的元素外的元素,无论与那些明确确定的元素相关还是不相干。因此,作为非限定性实例,“A和B的至少一个”(或,等同地,“A或B的至少一个”或,等同地“A和/或B的至少一个”)可在一个实施方案中指至少一个(任选地包括不止一个)A而无B存在(且任选地包括除B外的元素);在另一个实施方案中指至少一个(任选地包括不止一个)B而无A存在(且任选地包括除A外的元素);在另一个实施方案中指至少一个(任选地包括不止一个)A和至少一个(任选地包括不止一个)B(且任选地包括其它元素);依此类推。

[0571] 当在本文中使用词语“约”来指代数字时,应理解本公开内容的又一实施方案包括未因词语“约”的存在而修改的数字。

[0572] 还应当理解,除非明确地指明与之相反,否则在包括不止一个步骤或行为的本文请求保护的任何方法中,所述方法的步骤或行为的顺序不必须地限定于其中叙述所述方法的步骤或行为的顺序。

[0573] 在权利要求中,以及在上述说明书中,所有过渡短语诸如“包含”、“包括”、“携带”、“具有”、“含有”、“拥有”、“涉及”、“持有”、“由……组成”等被理解为开放性的,即意指包括但不限于。只有过渡短语“由……组成”和“基本上由……组成”应当分别是封闭性的或半封闭性的过渡短语,如美国专利局专利审查程序手册第2111.03节中所示的。

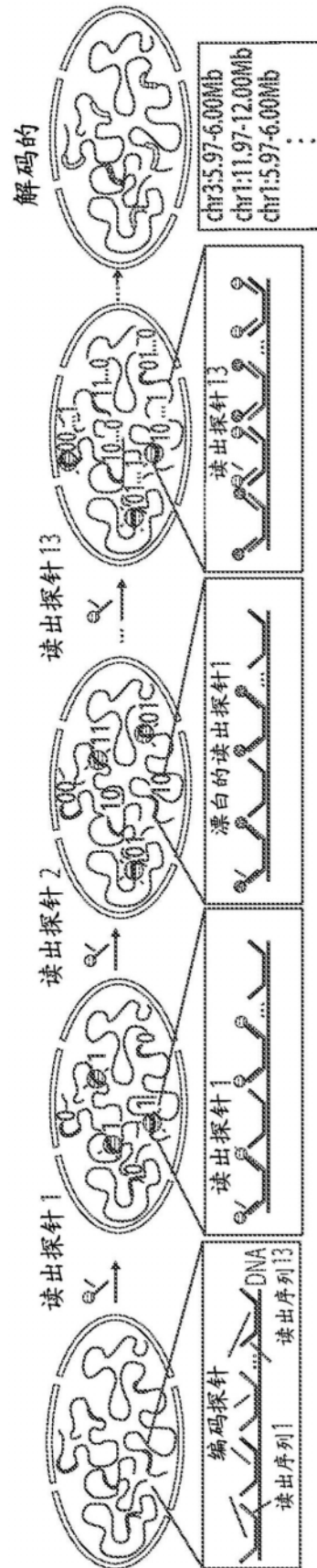


图1A

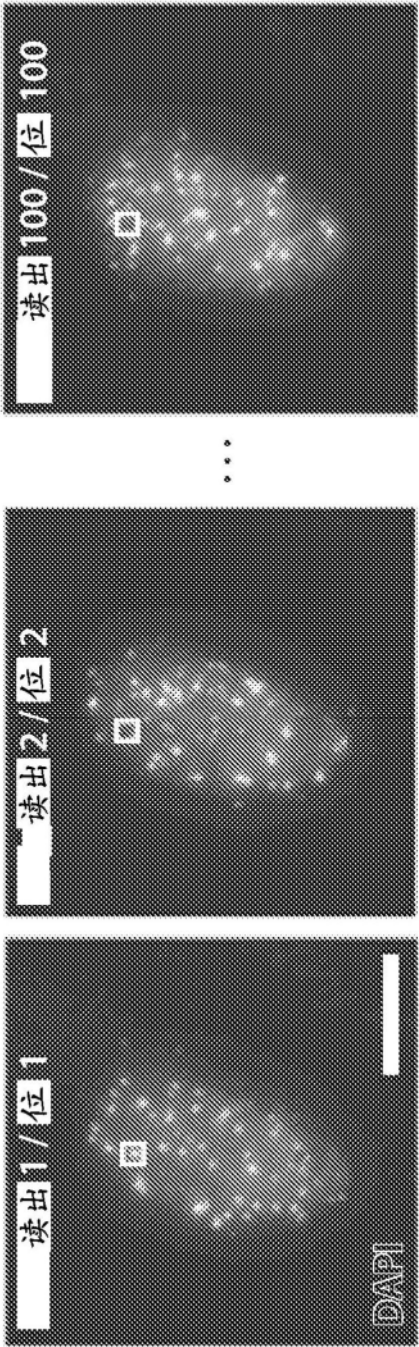


图1B

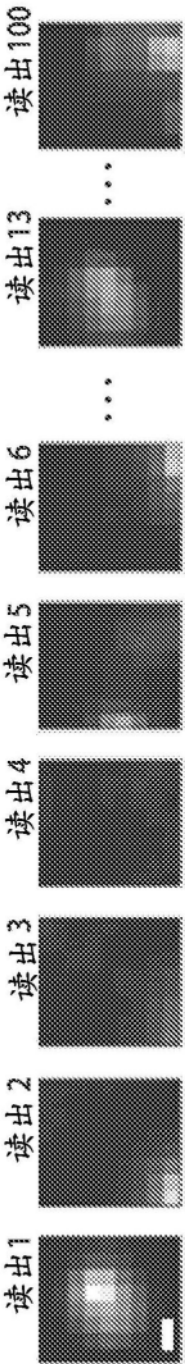


图1C

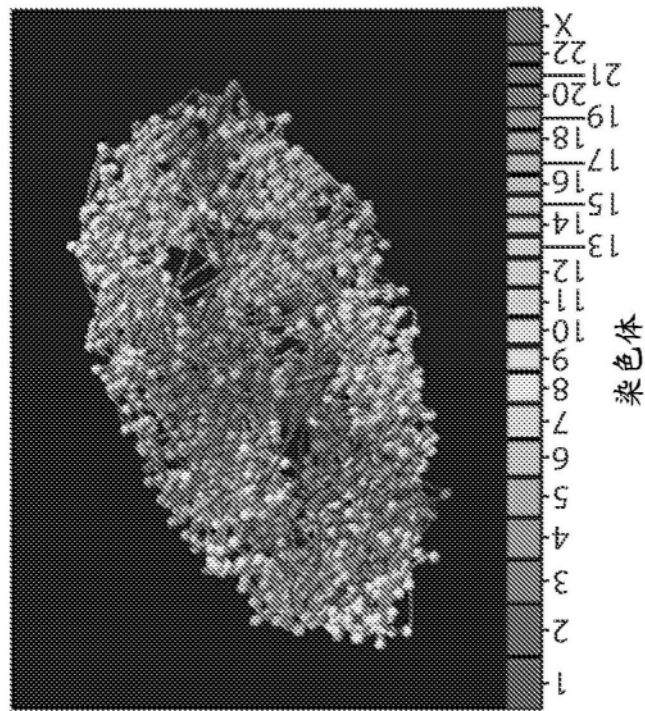


图1D

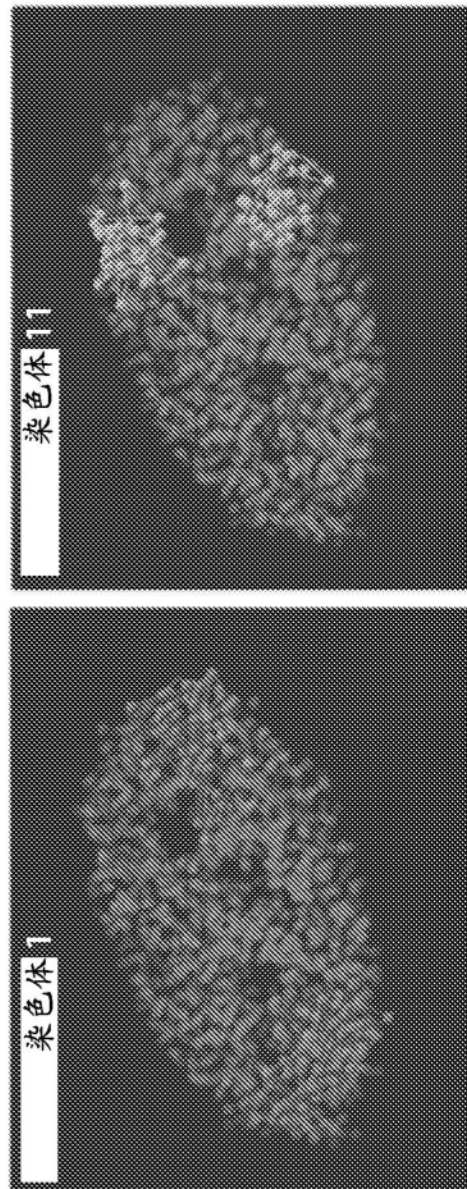


图1E

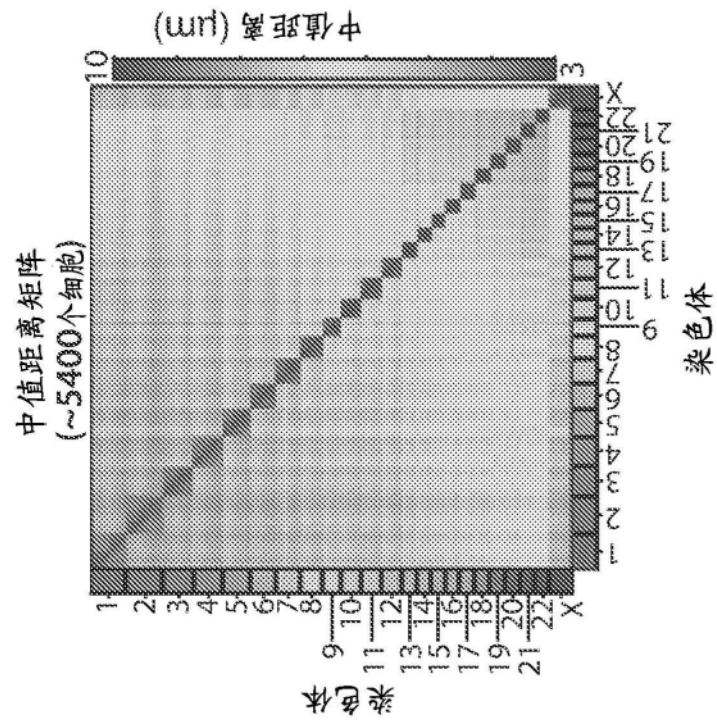


图1F

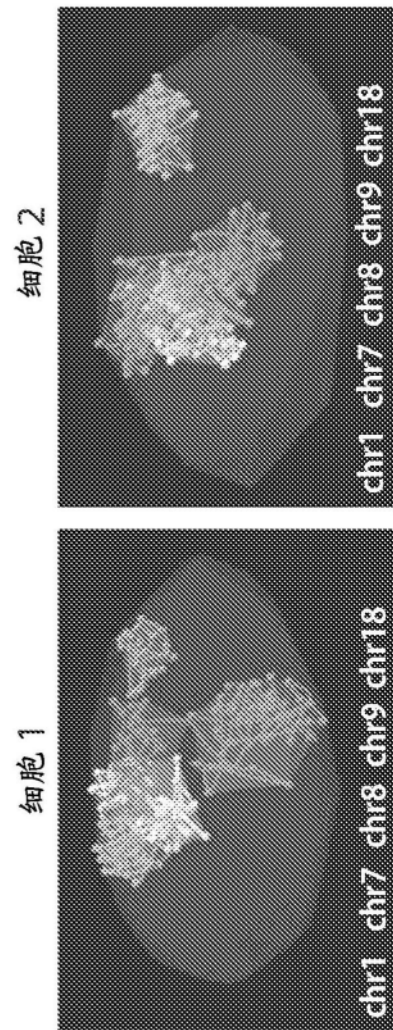


图1G

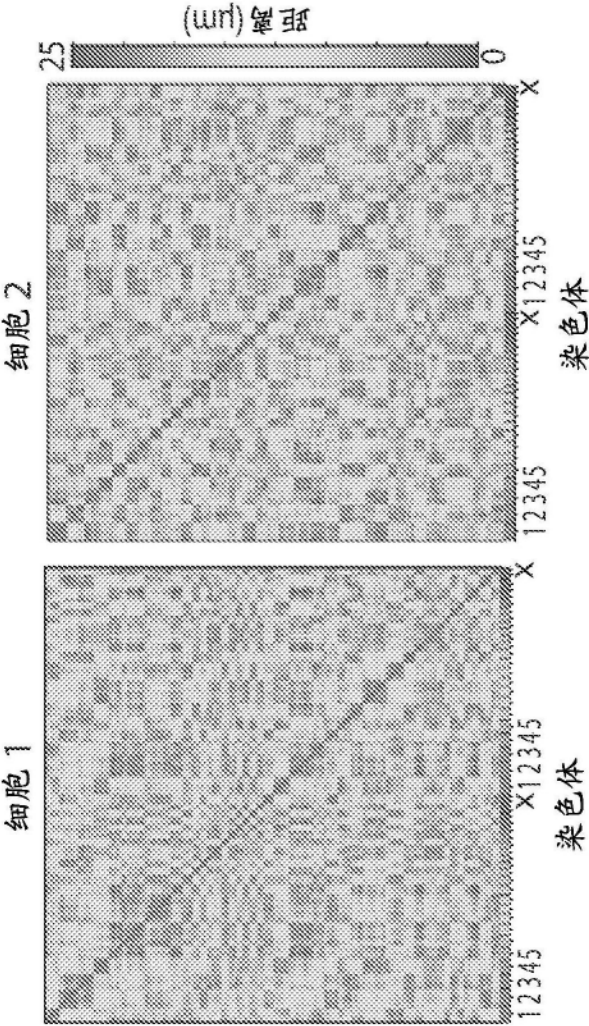


图1H

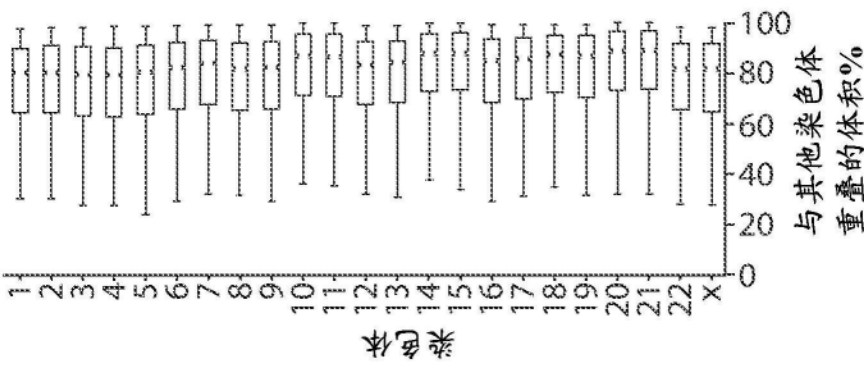


图1I

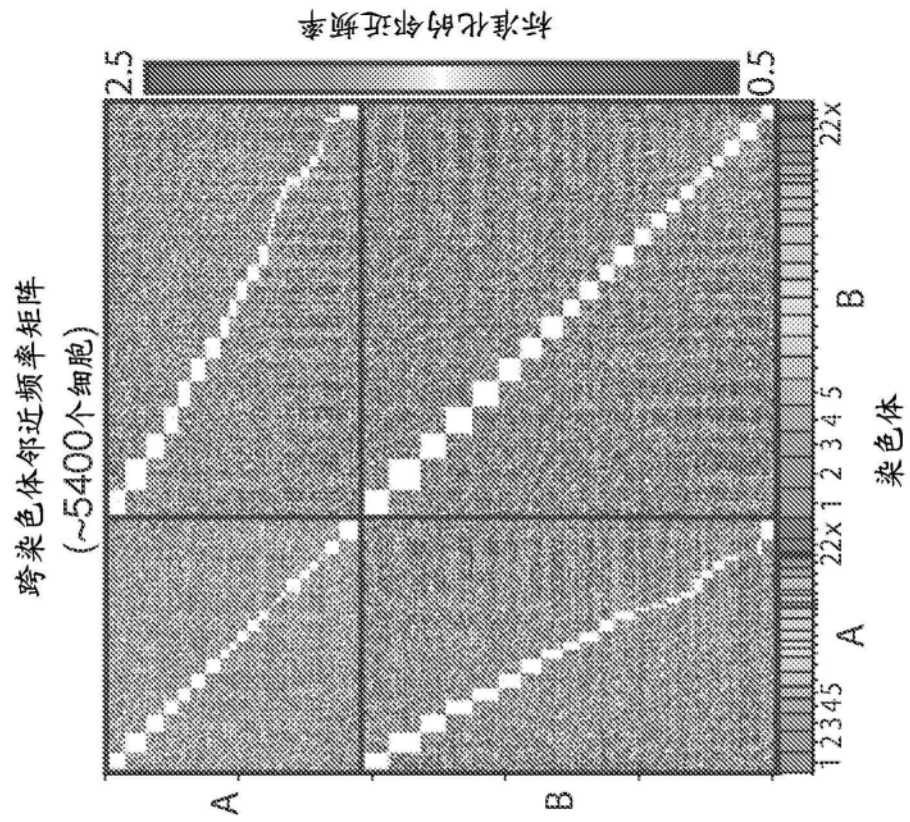


图2A

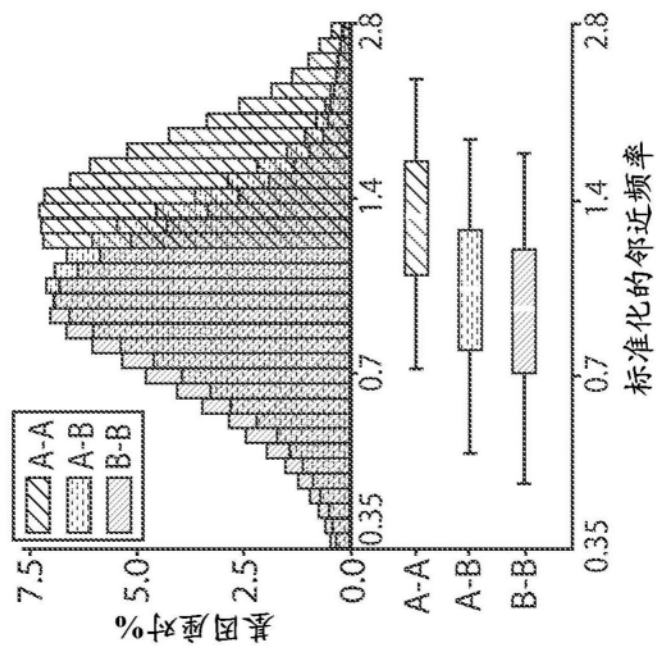


图2B

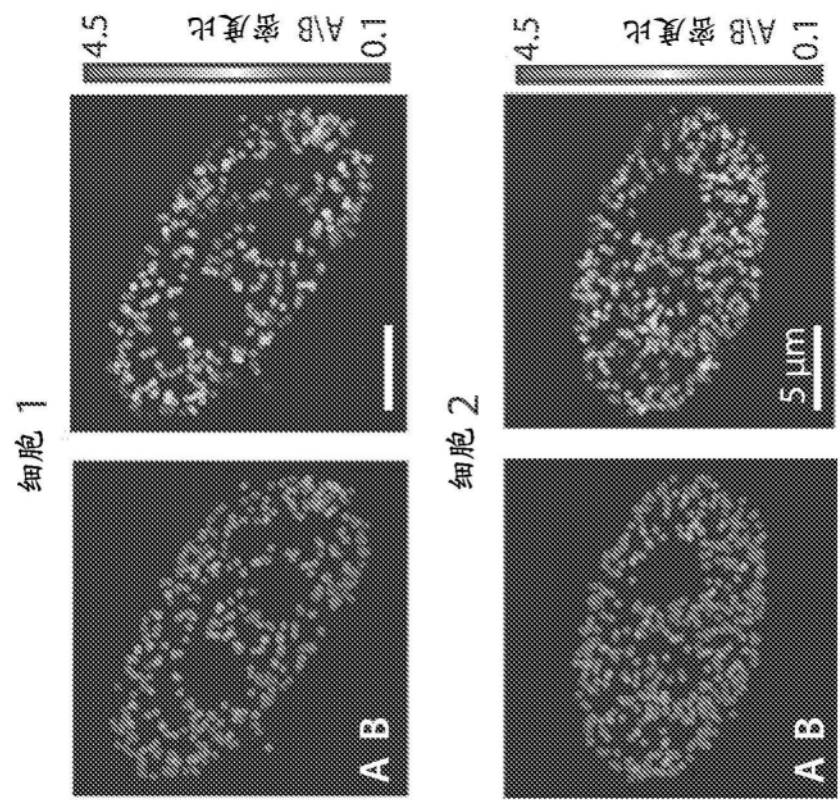


图2C

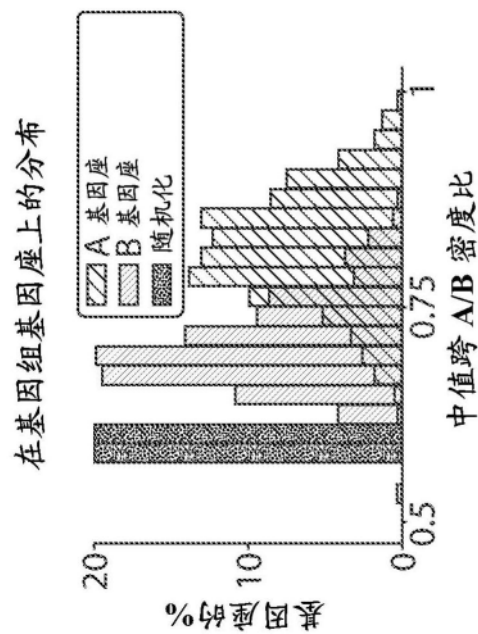


图2D

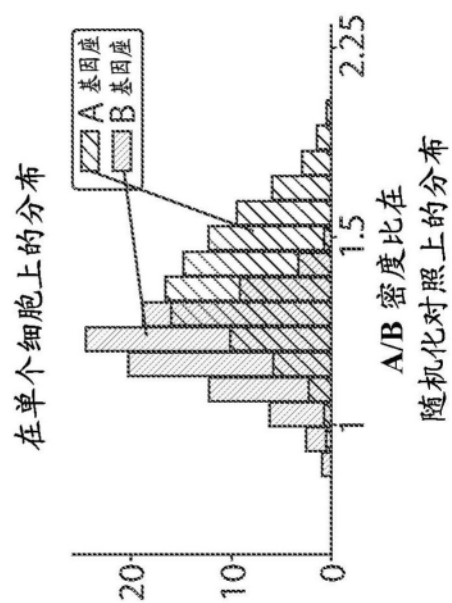


图2E

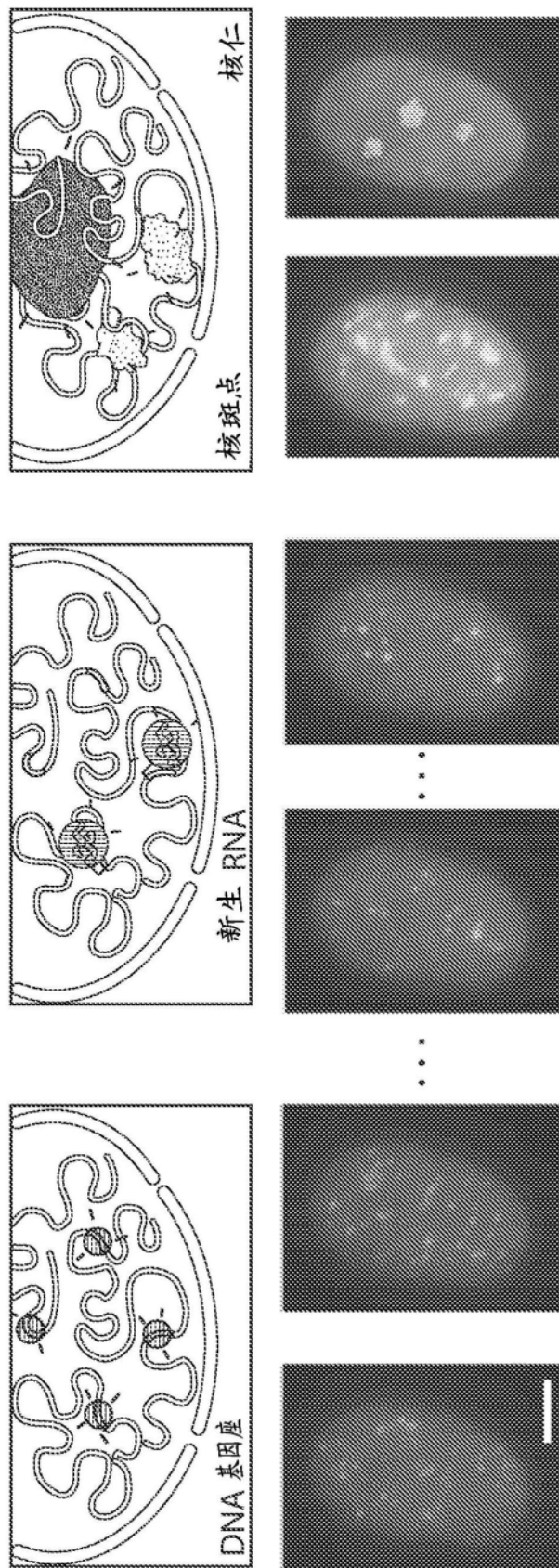


图3A

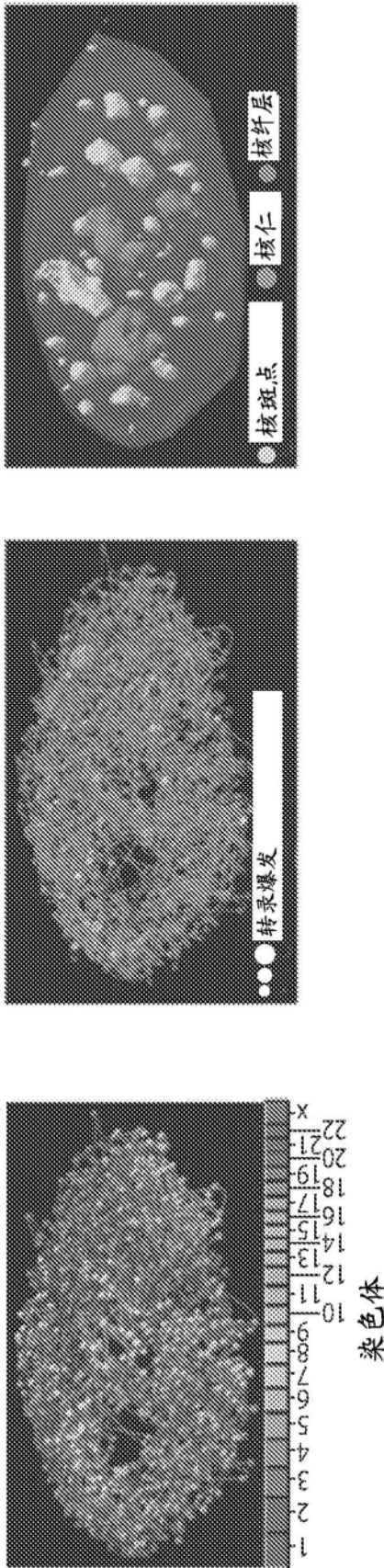


图3B

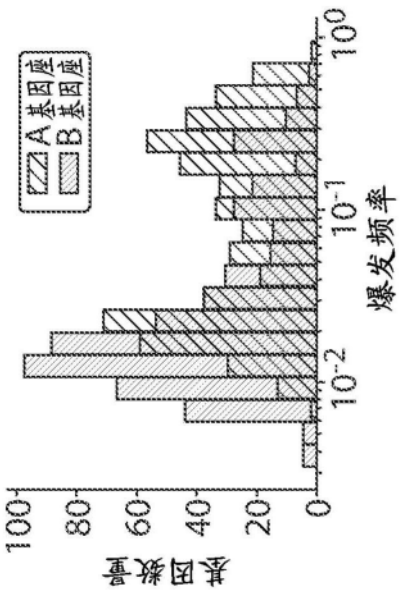


图3C

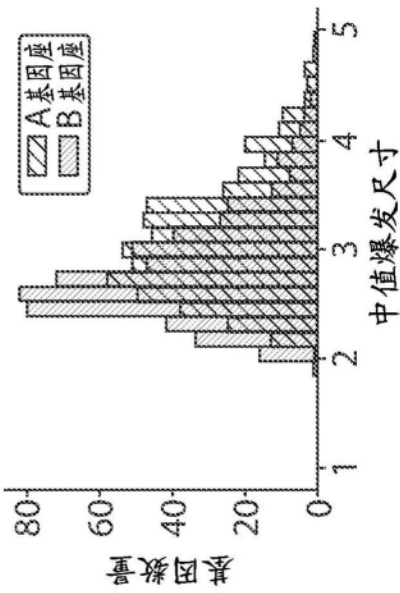


图3D

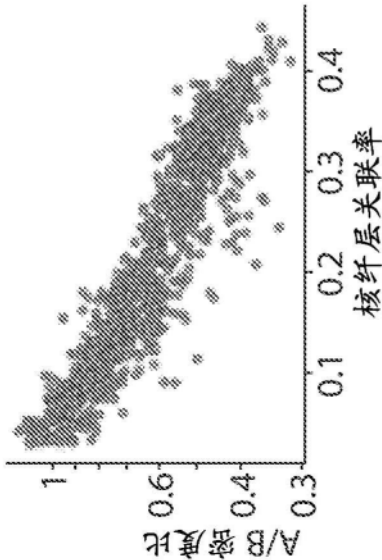


图3E

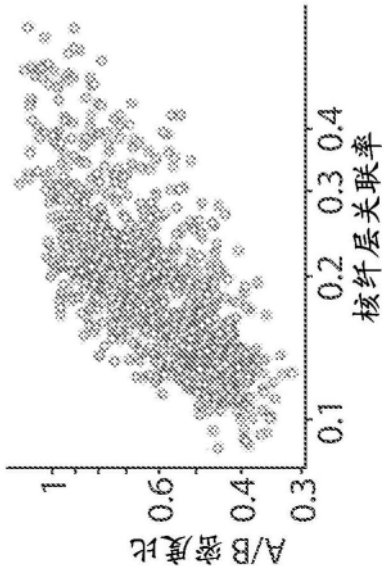


图3F

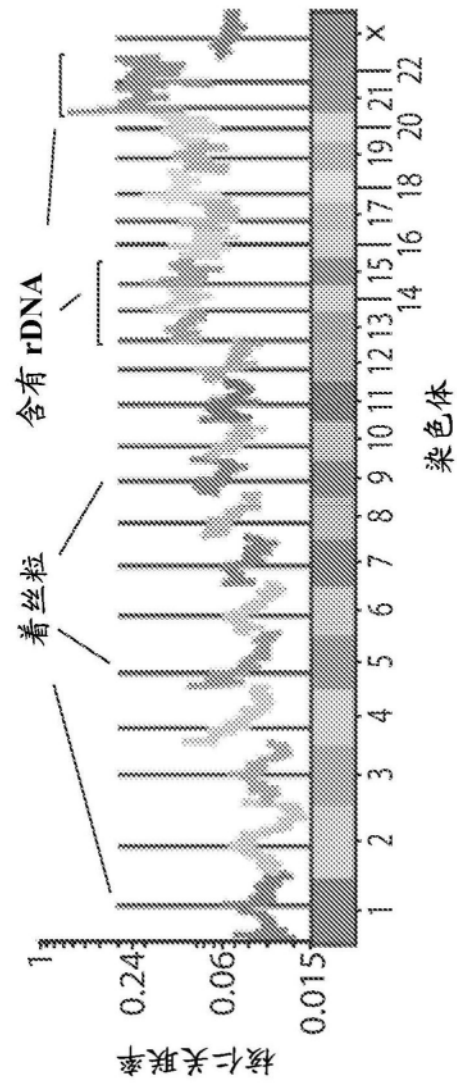


图3G

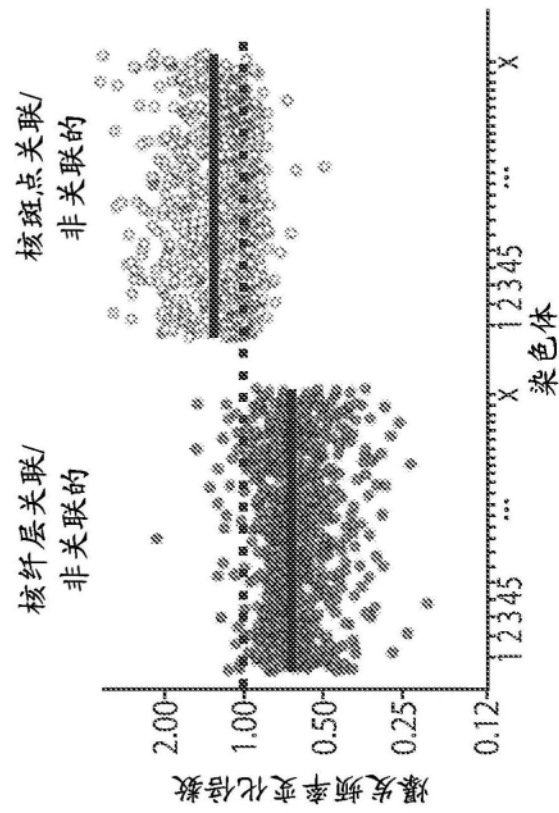


图3H

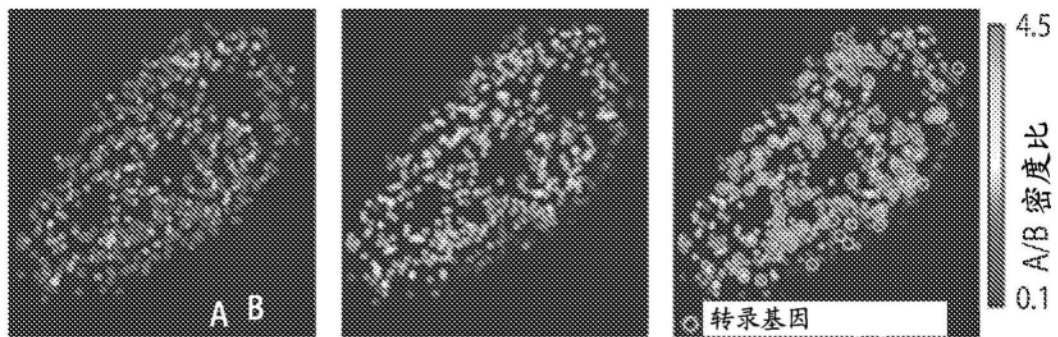


图4A

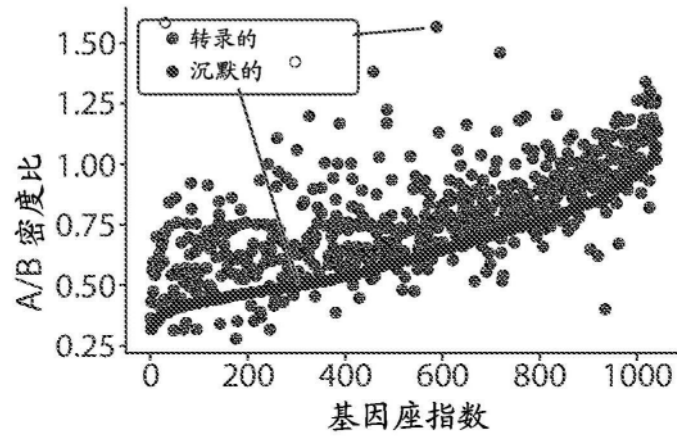


图4B

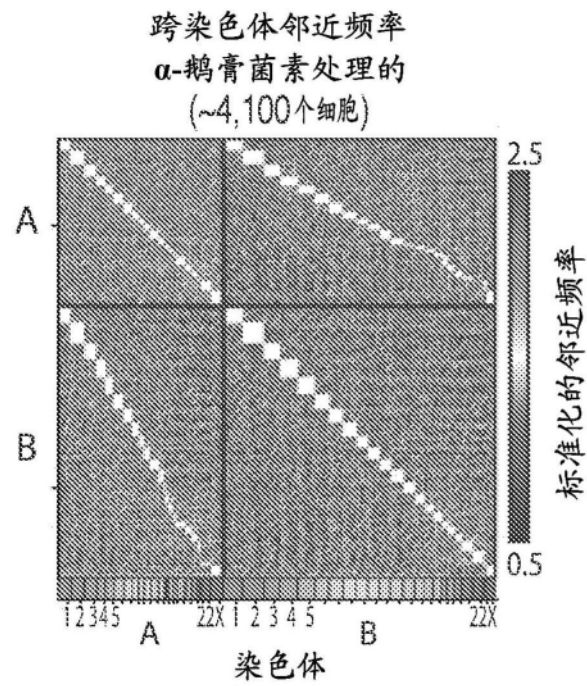


图4C

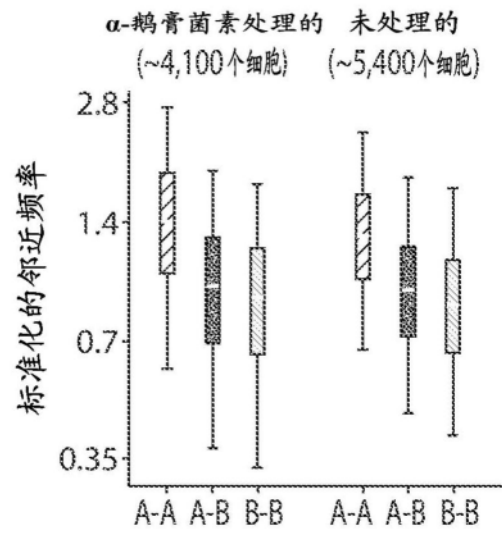


图4D

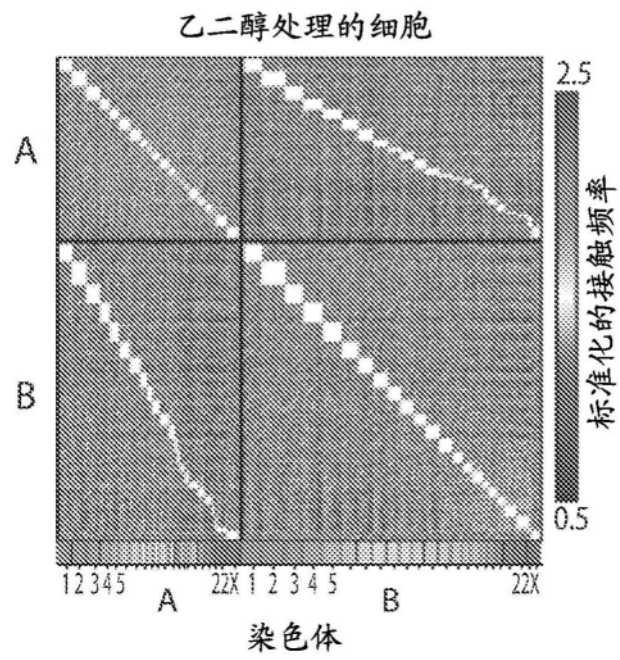


图4E

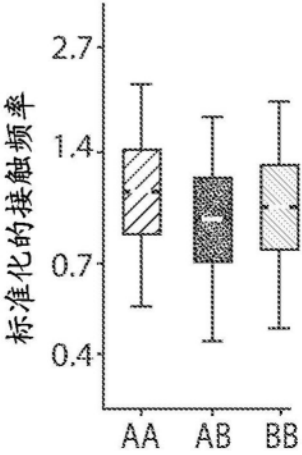


图4F

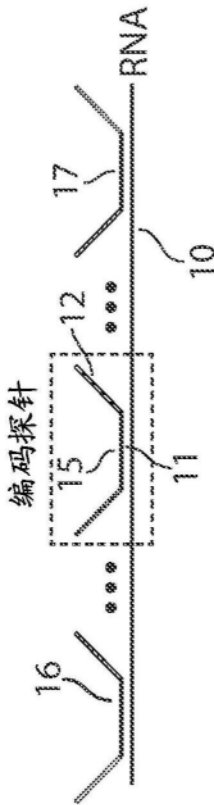


图5A

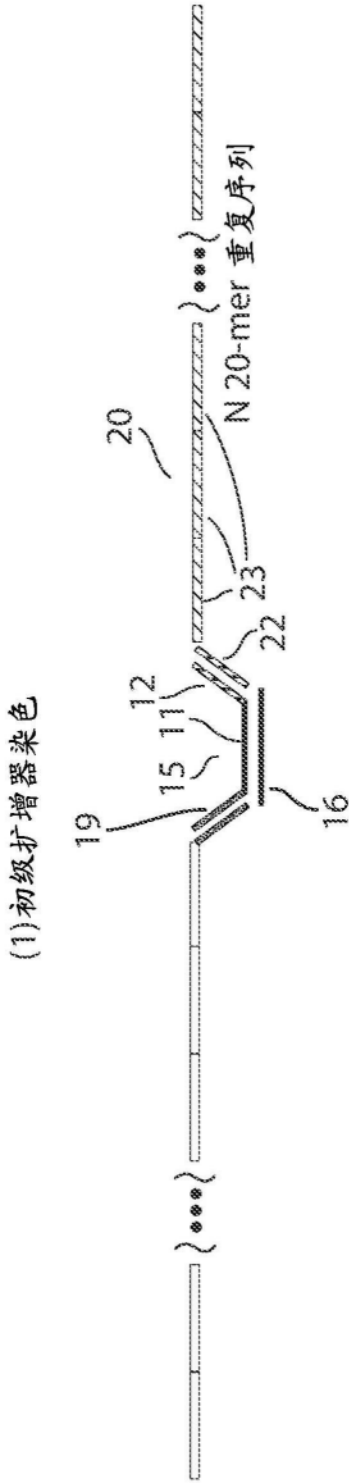


图5B

(2) 次级扩增器染色

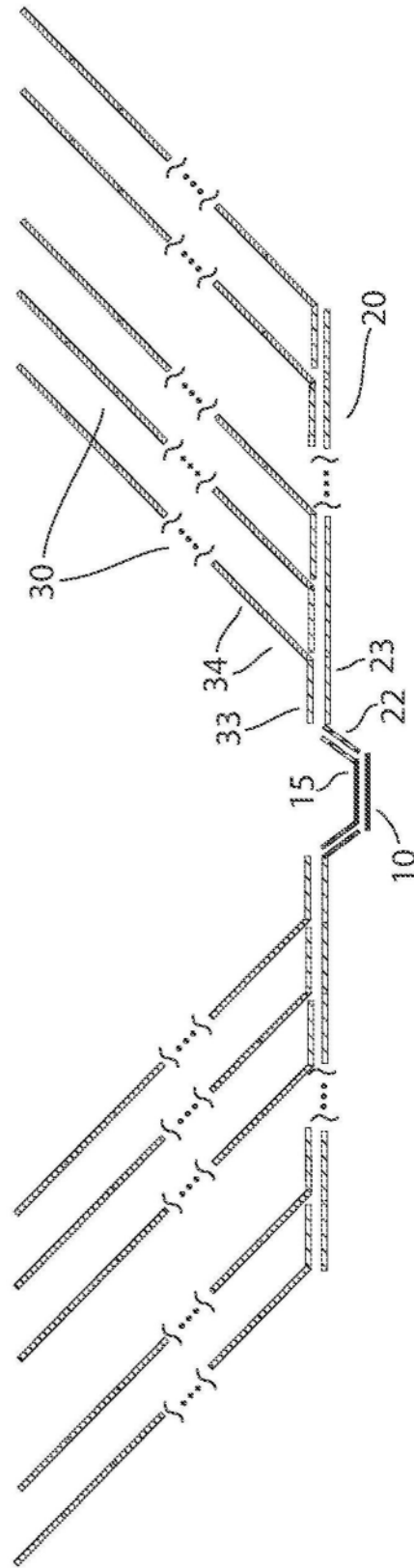


图5C

(3) NxM 扩增的 MERFISH 轮次 1

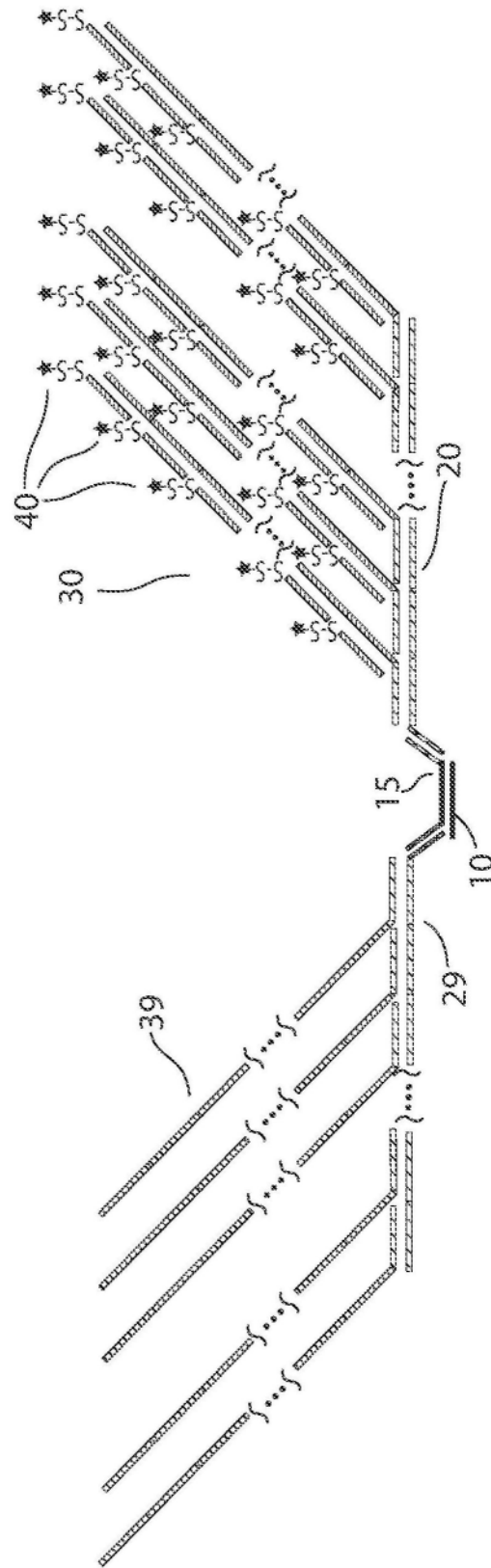


图5D

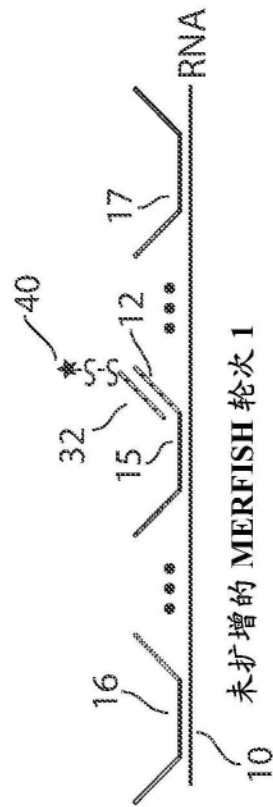


图5E

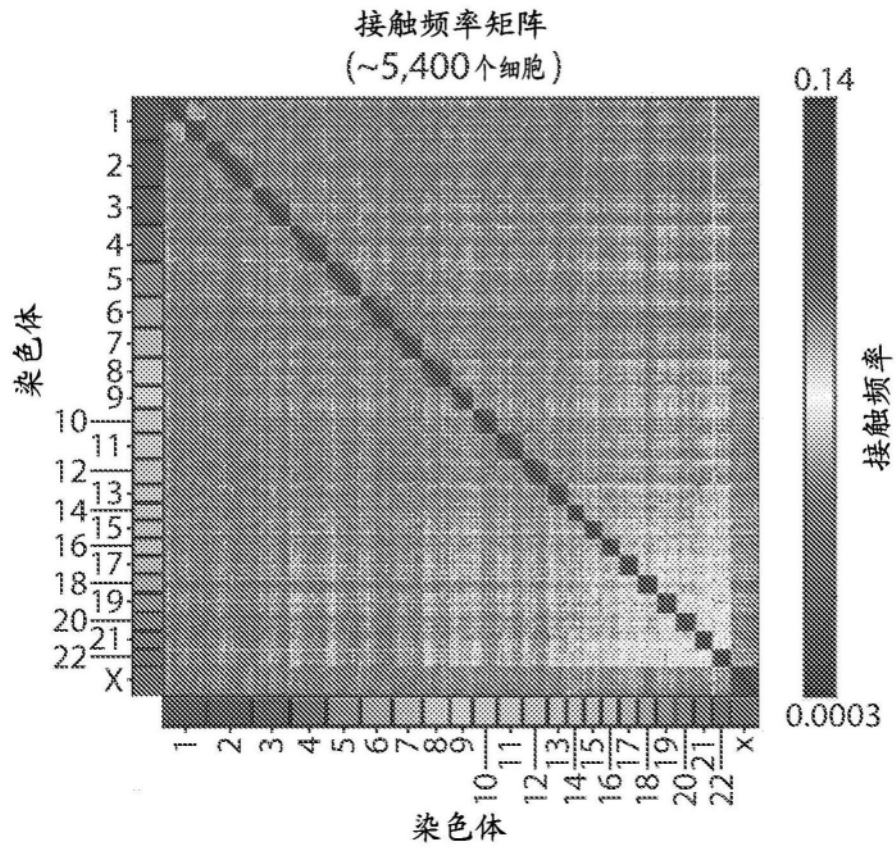


图6A

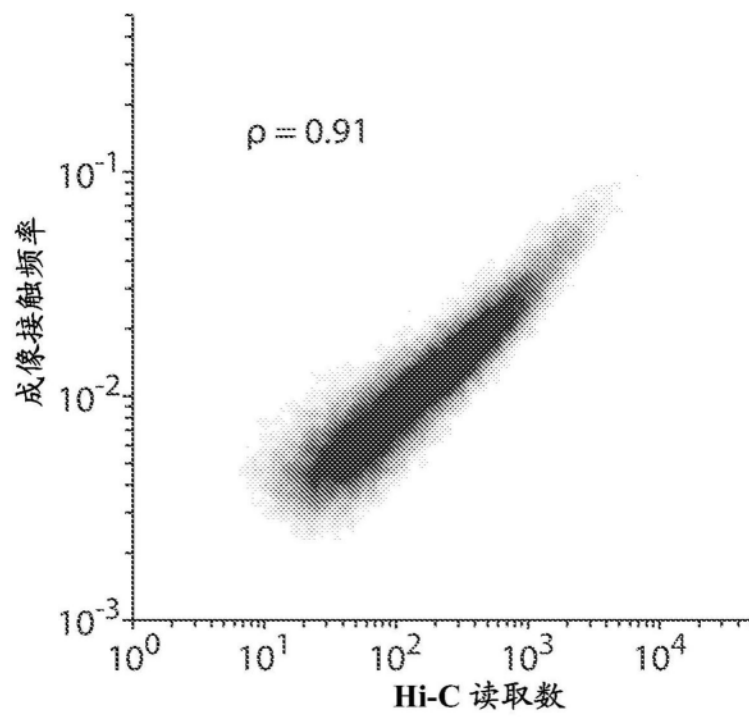


图6B

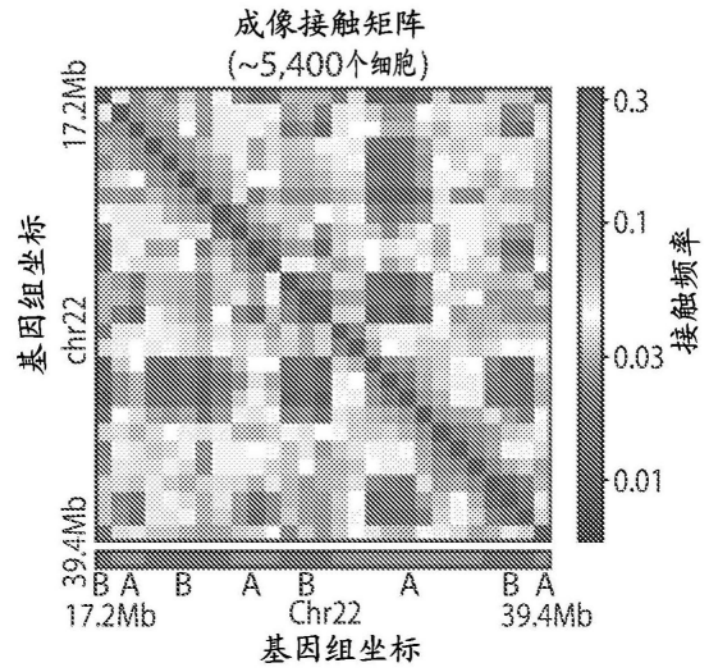


图7A

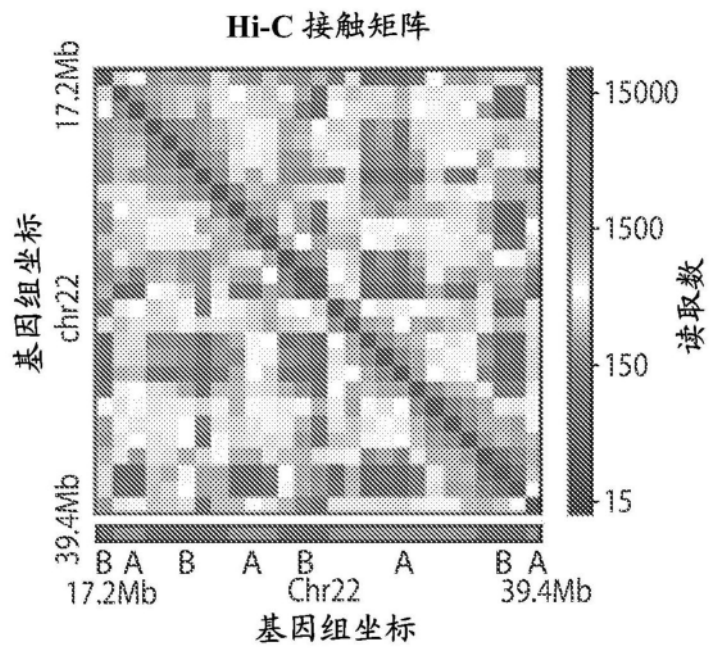


图7B

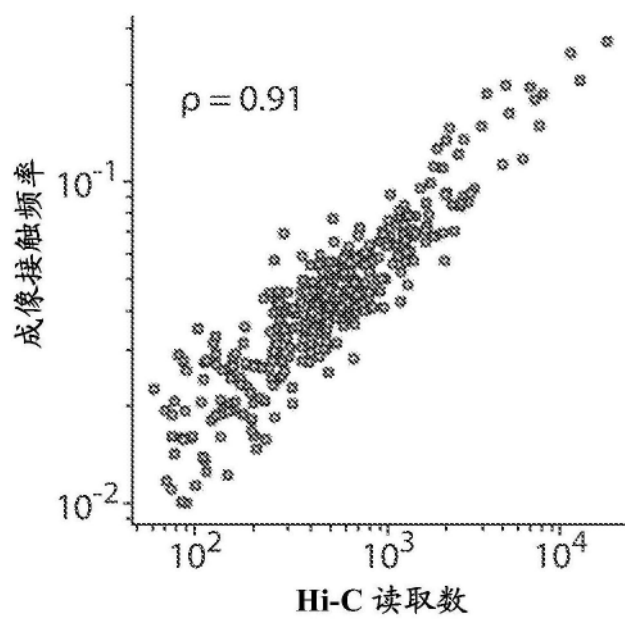


图7C

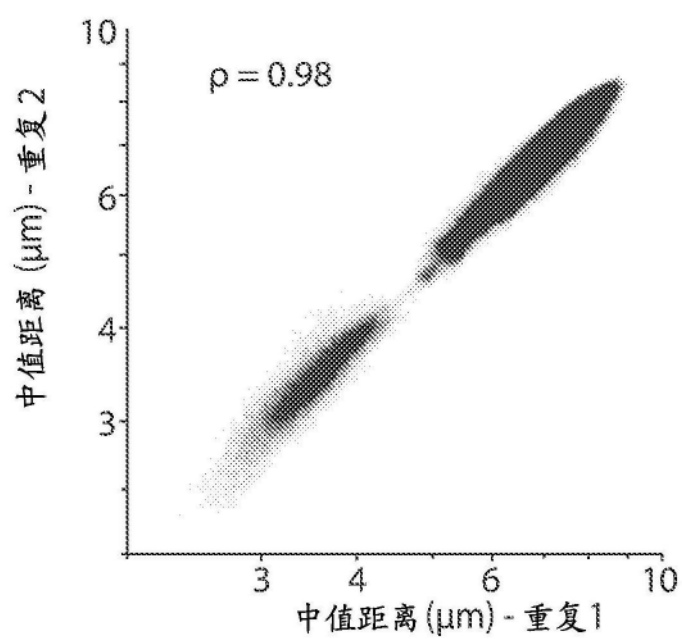


图8

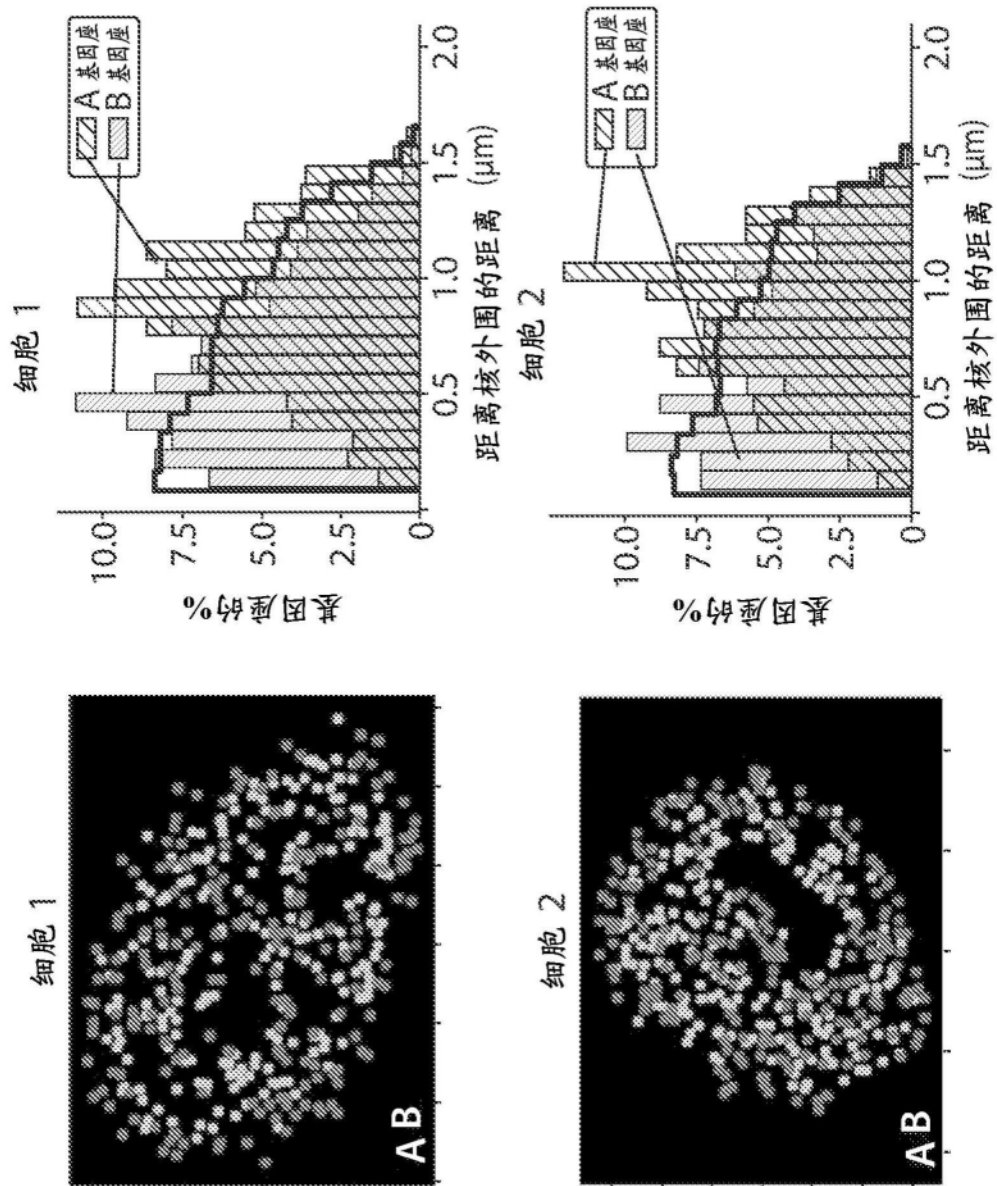


图9A

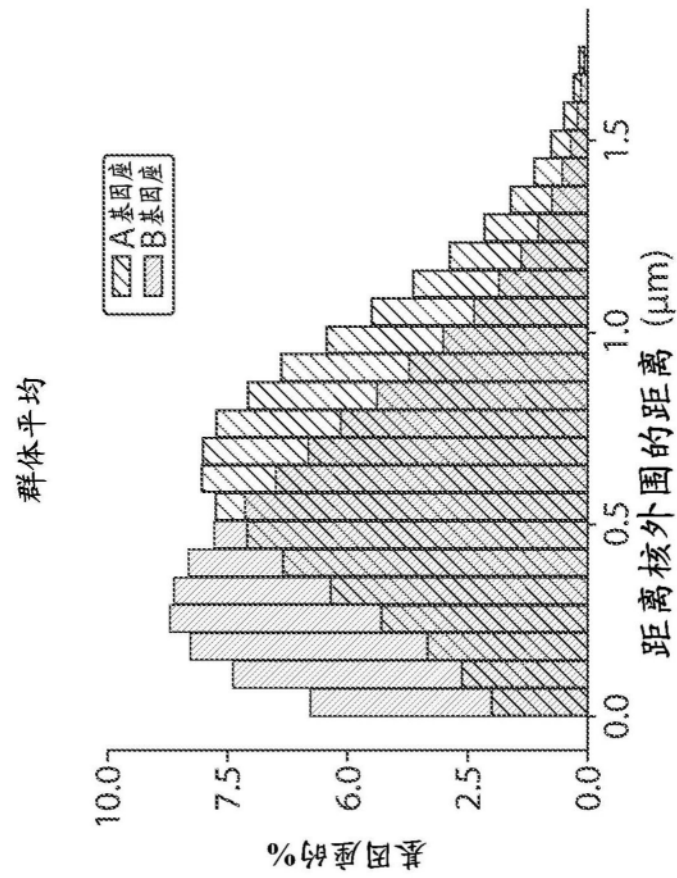


图9B

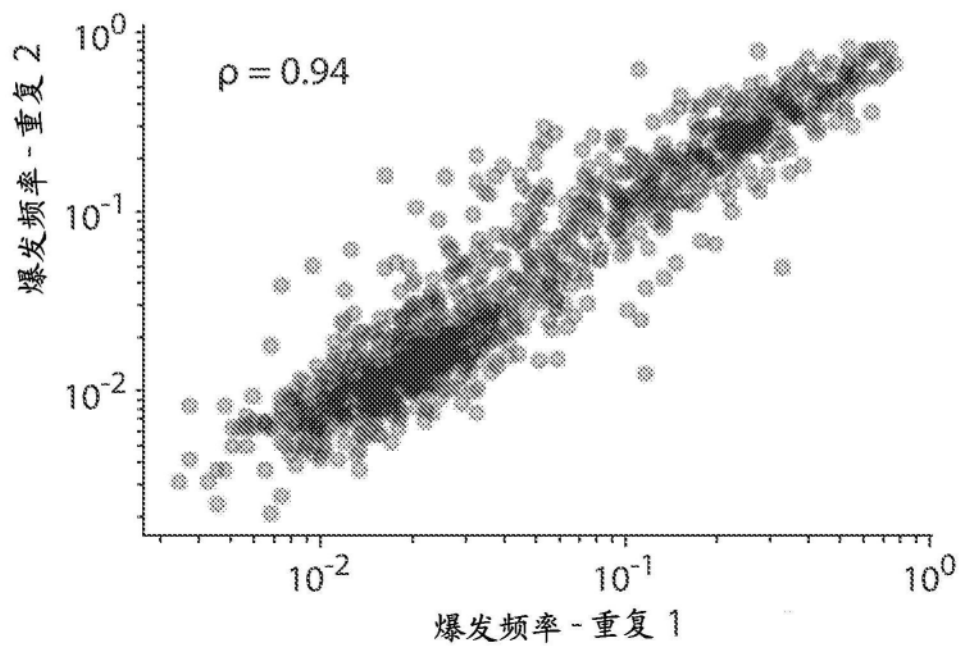


图10A

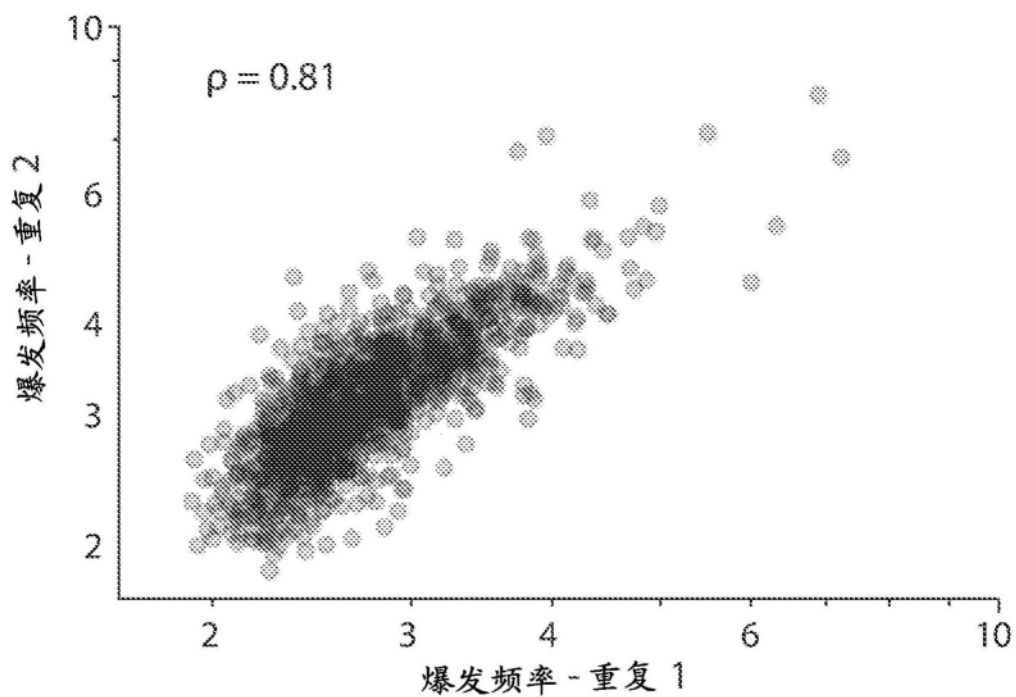


图10B

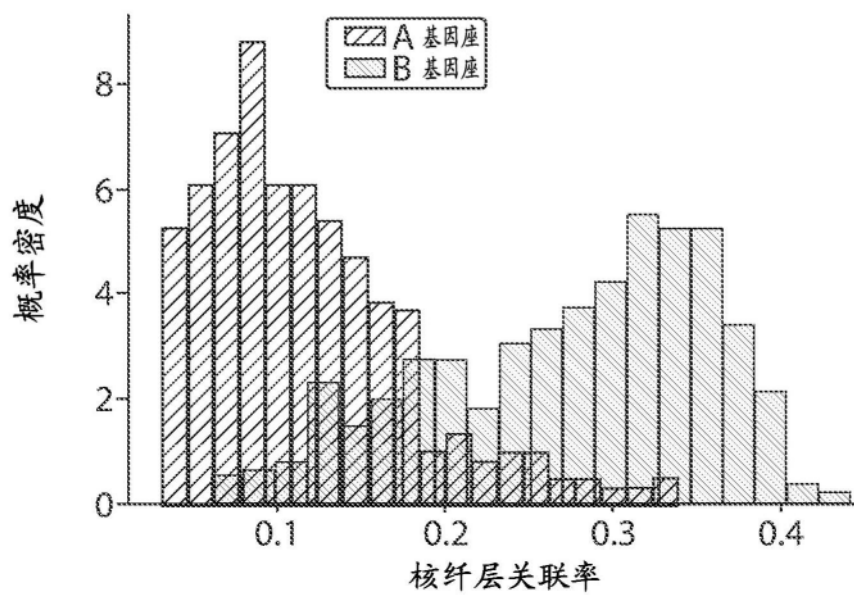


图11

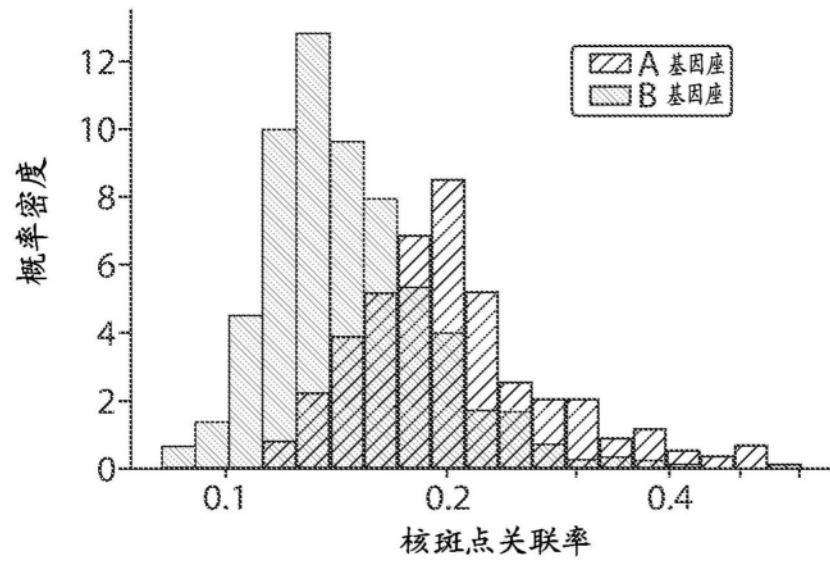


图12

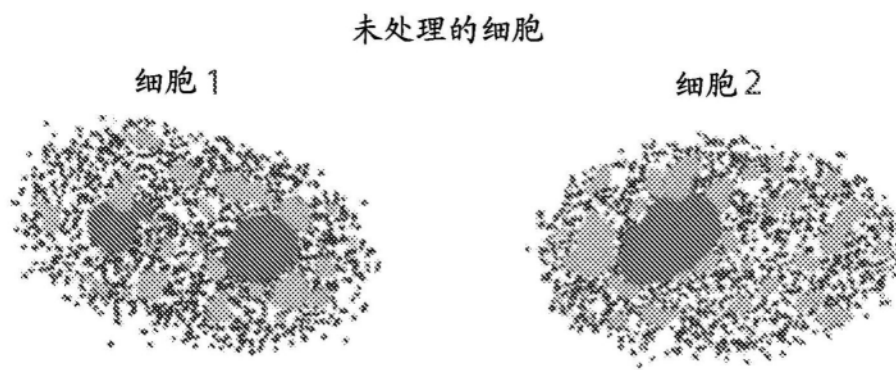


图13A

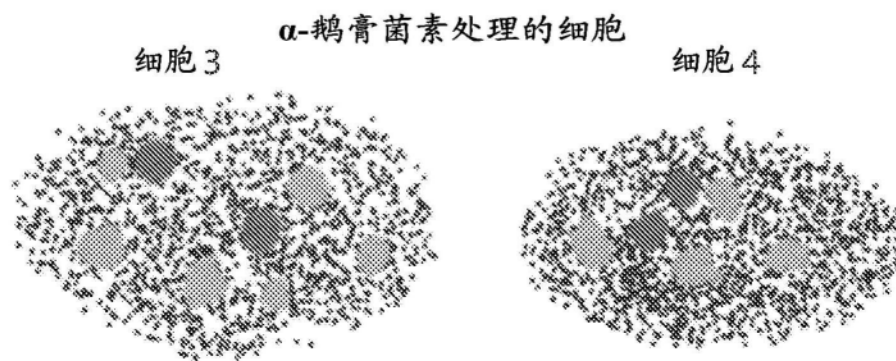


图13B

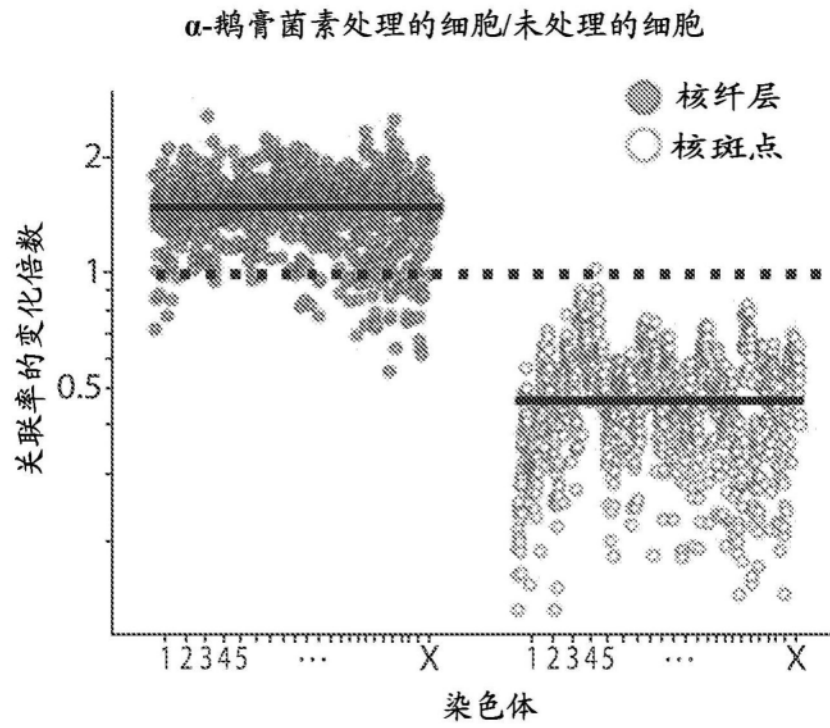


图13C

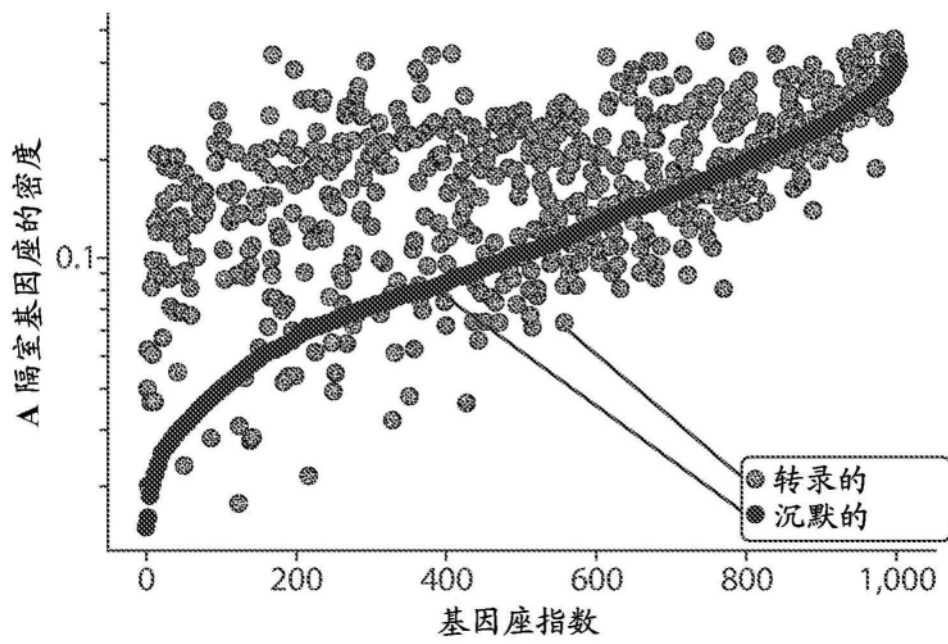


图14

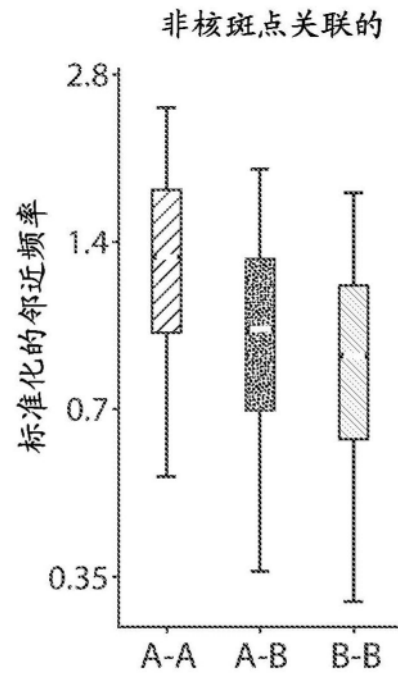


图15A

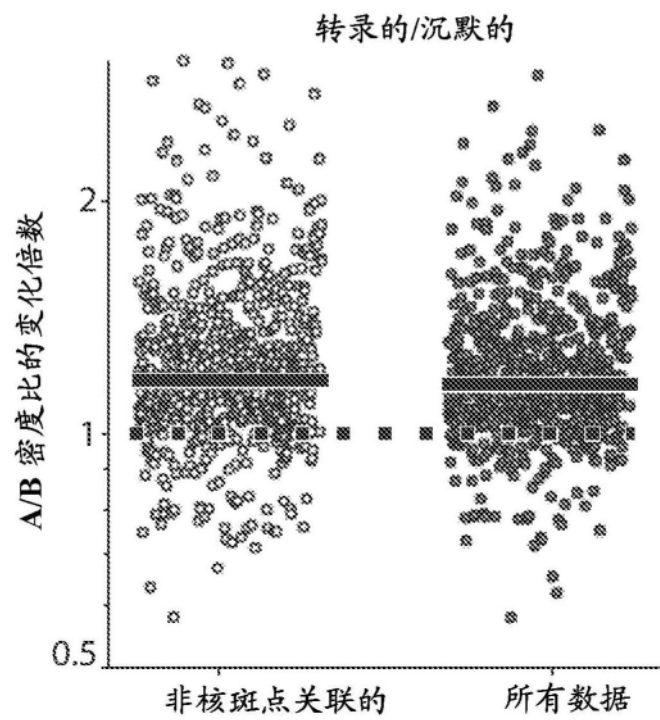


图15B

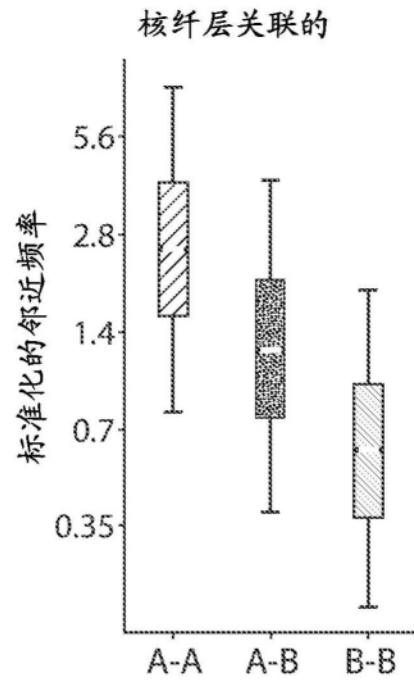


图16A

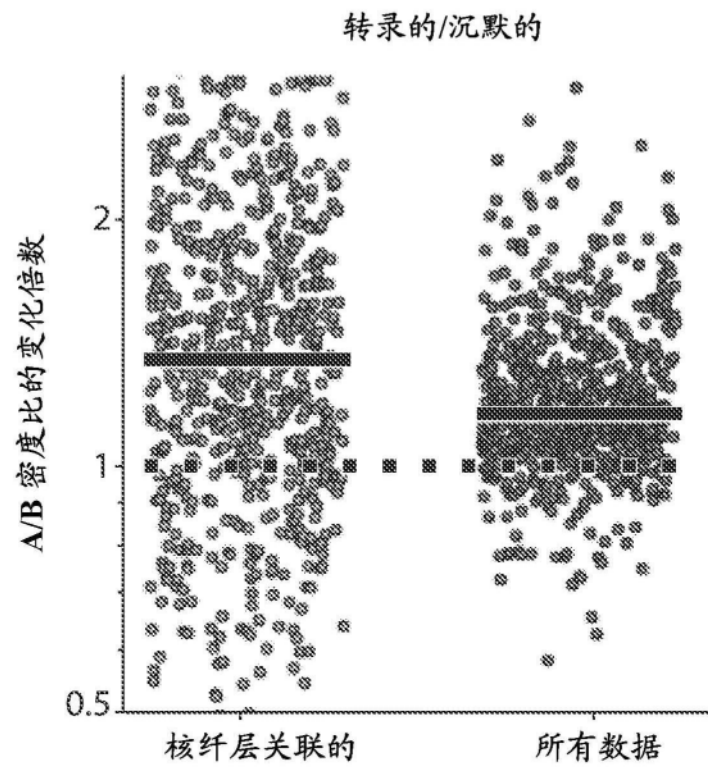


图16B

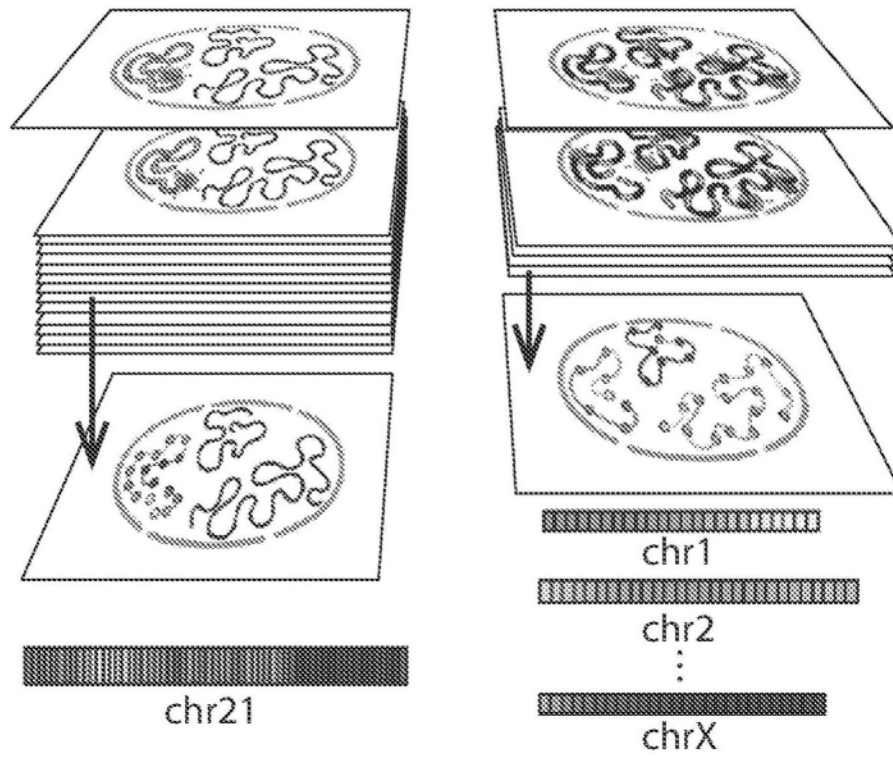


图17A

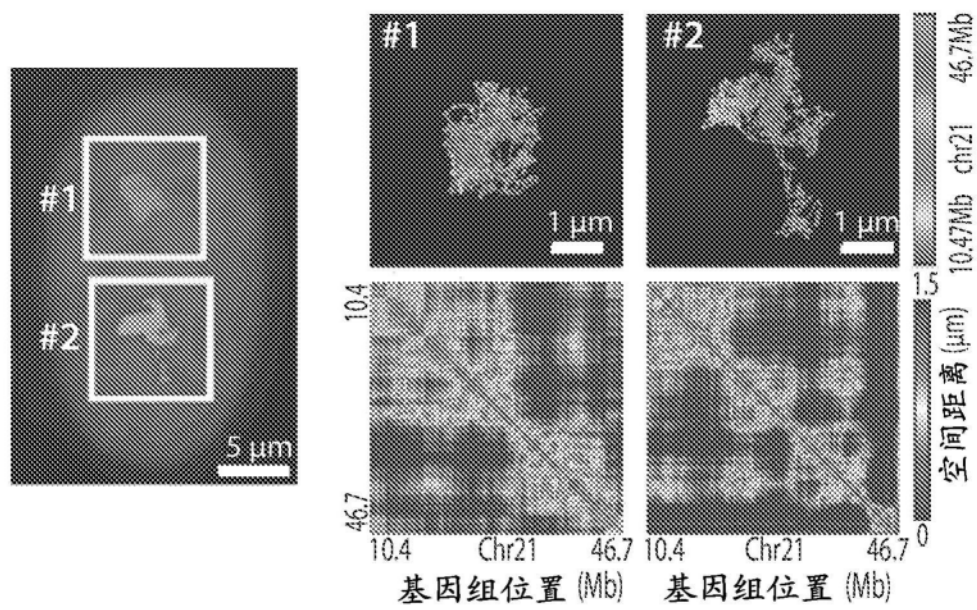


图17B

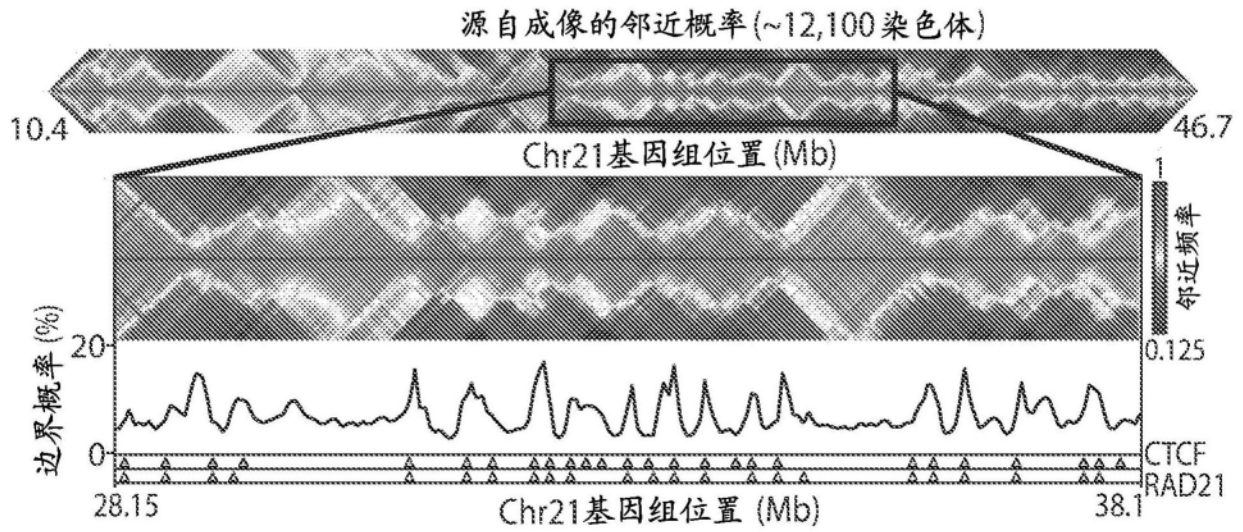


图17C

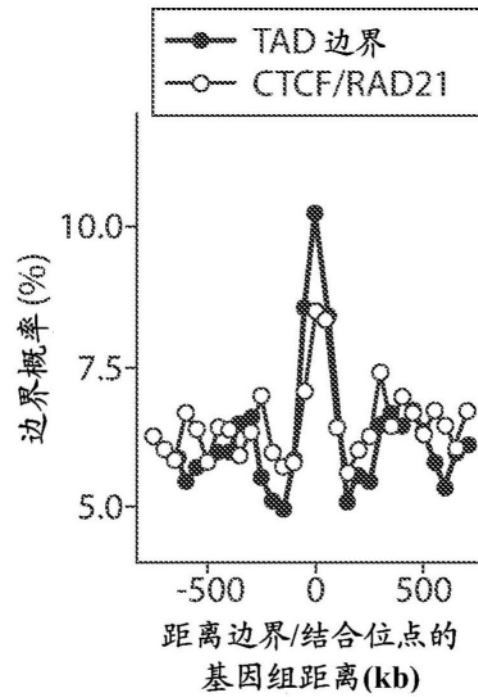


图17D

物理尺寸

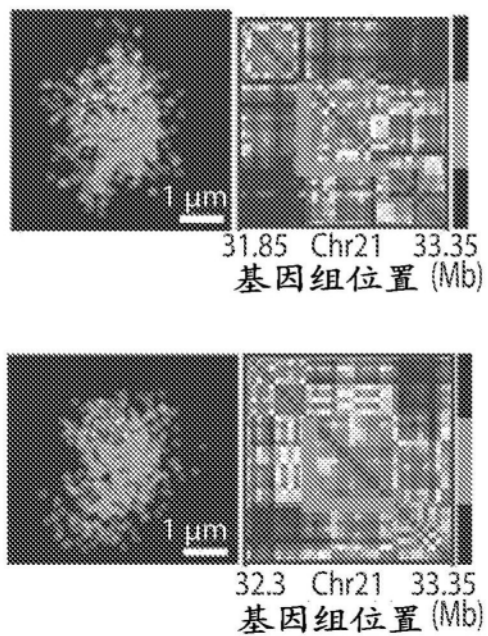


图17E

绝缘度

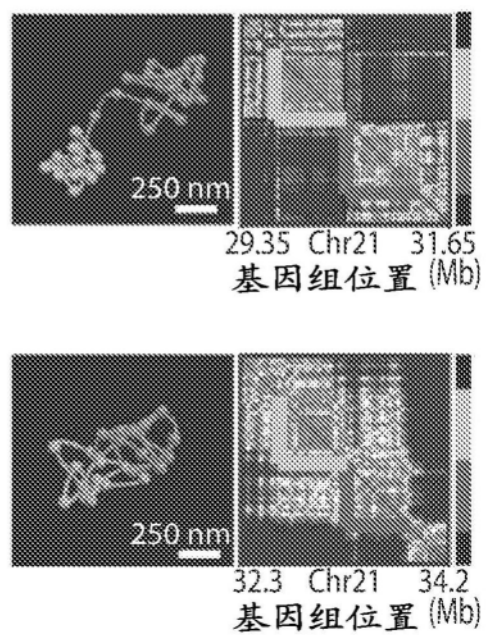


图17F

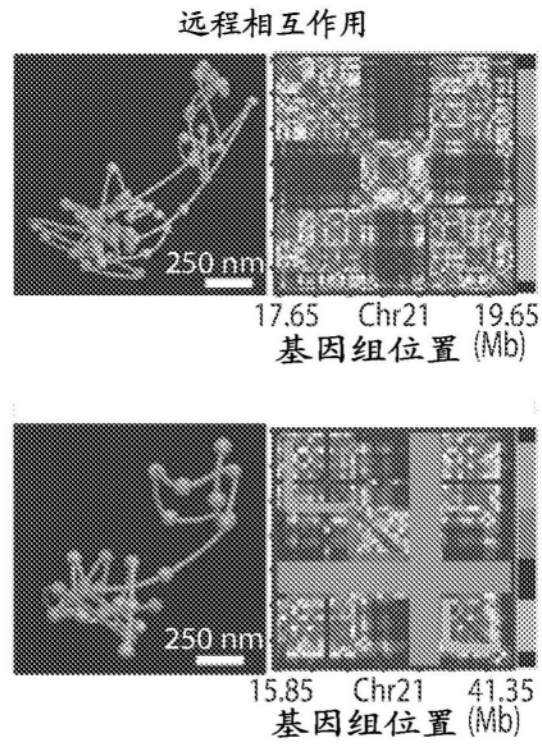


图17G

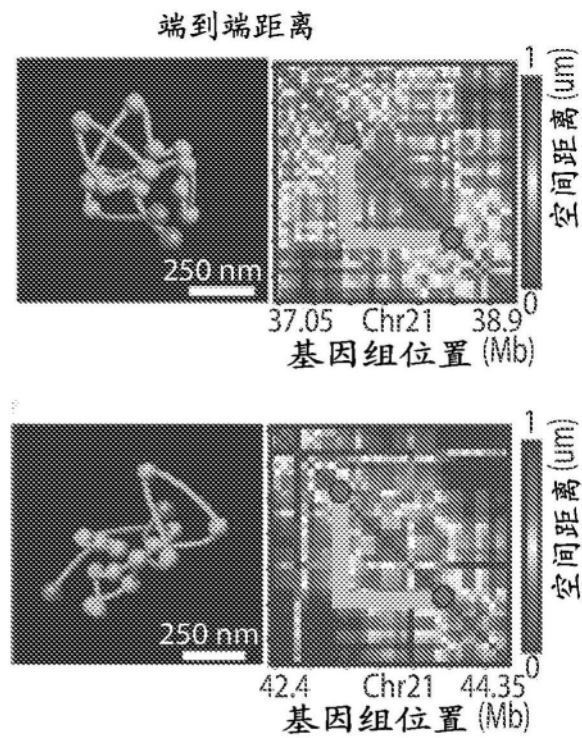


图17H

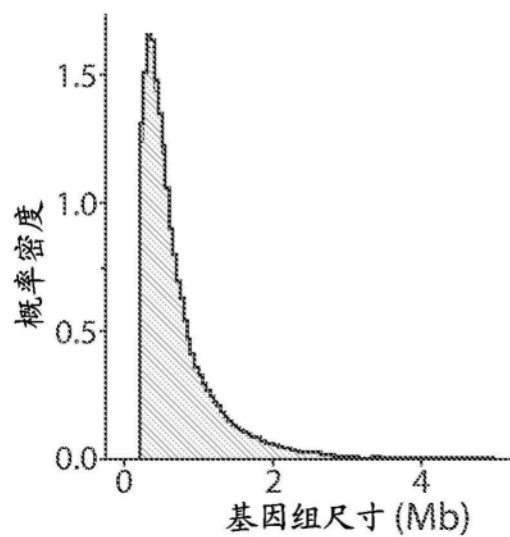


图17I

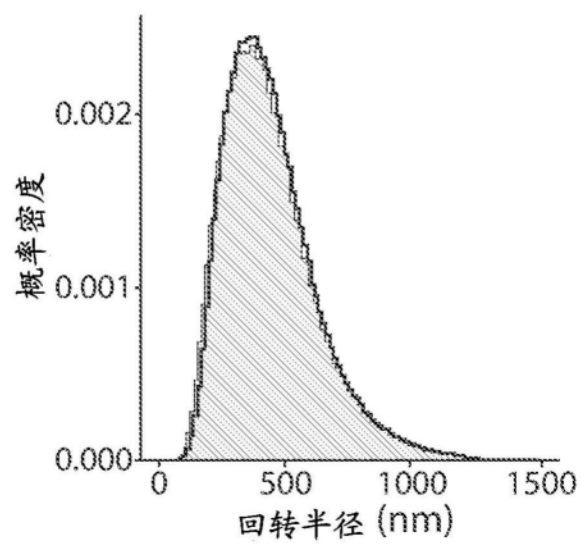


图17J

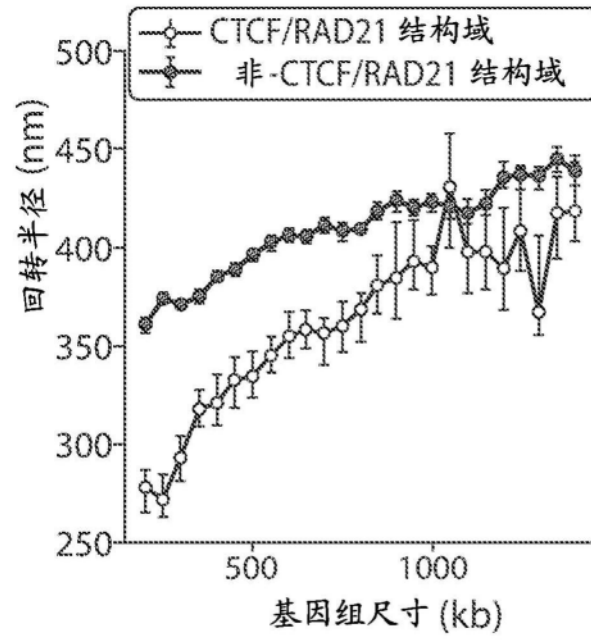


图17K

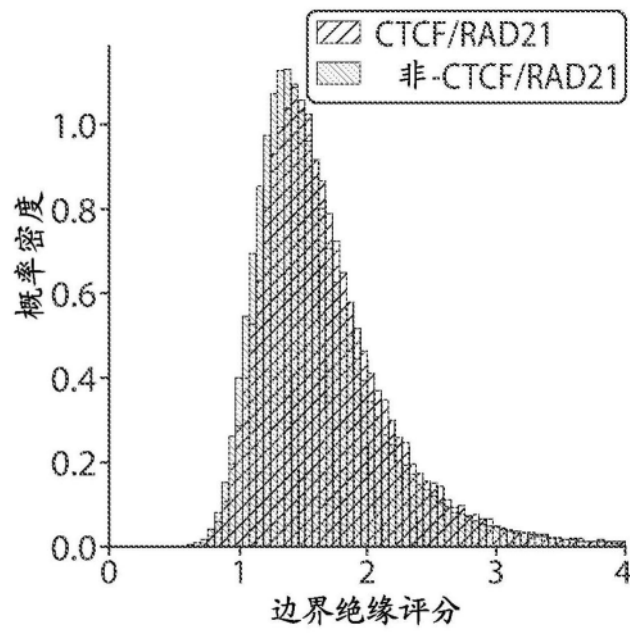


图17L

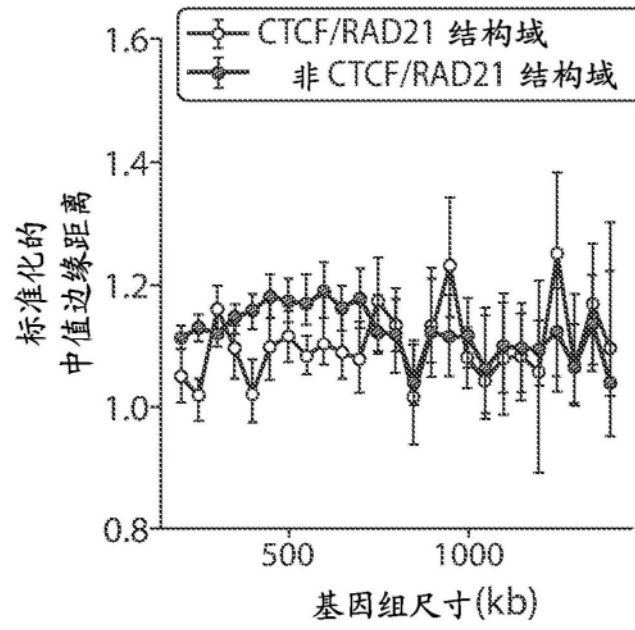


图17M

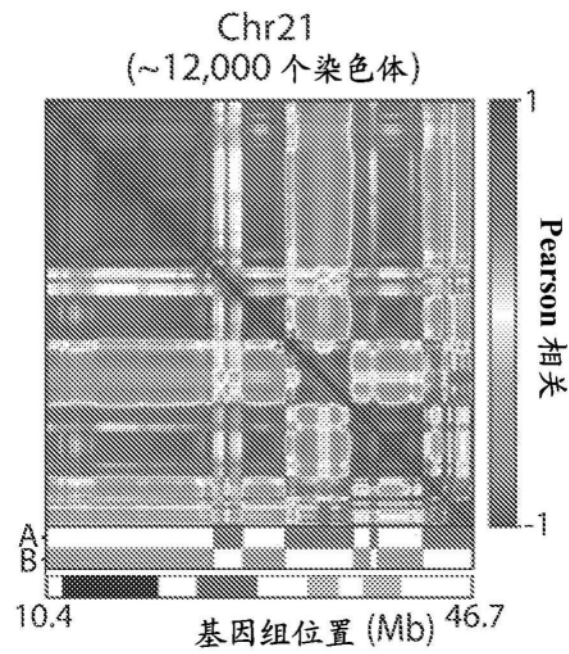


图18A

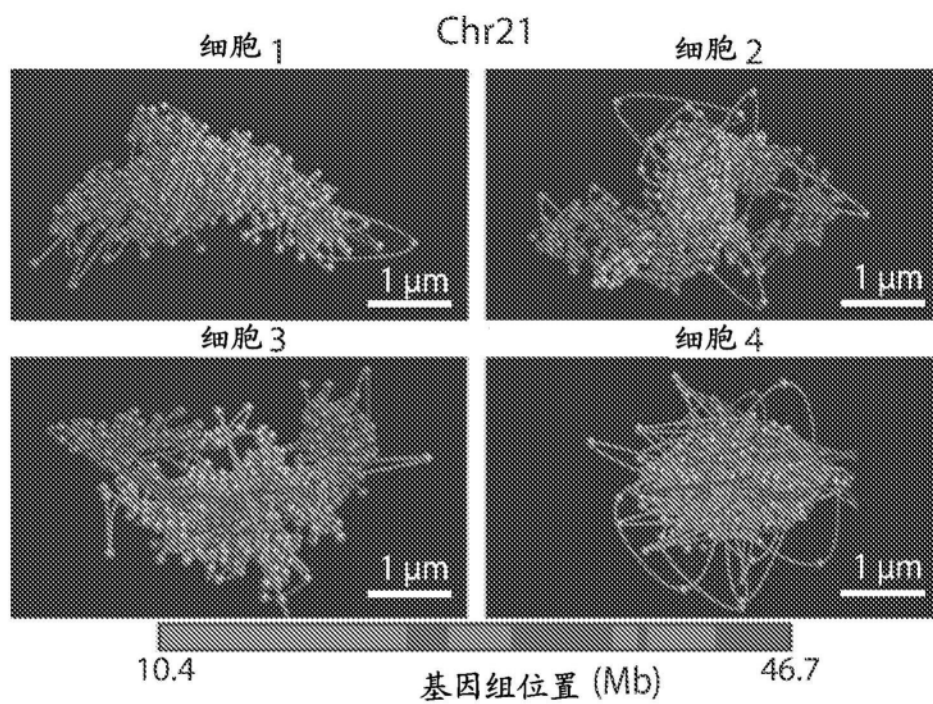


图18B

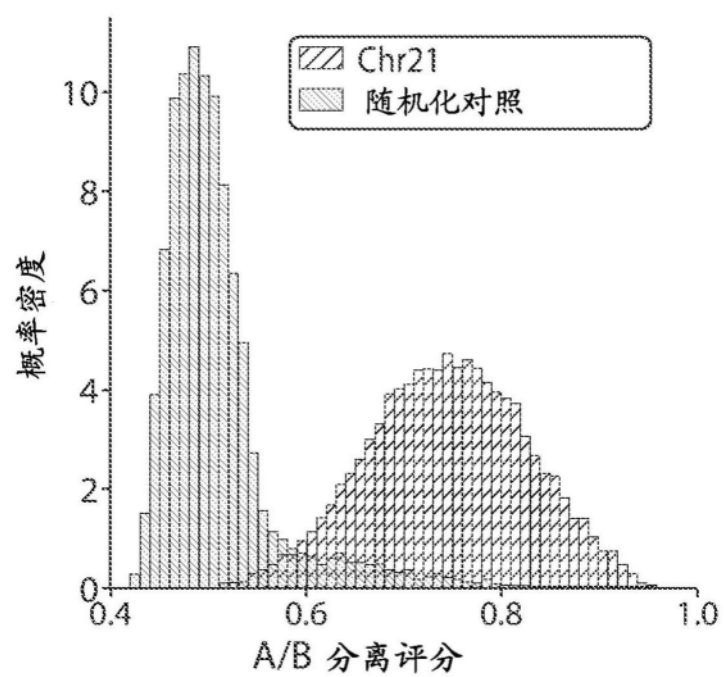


图18C

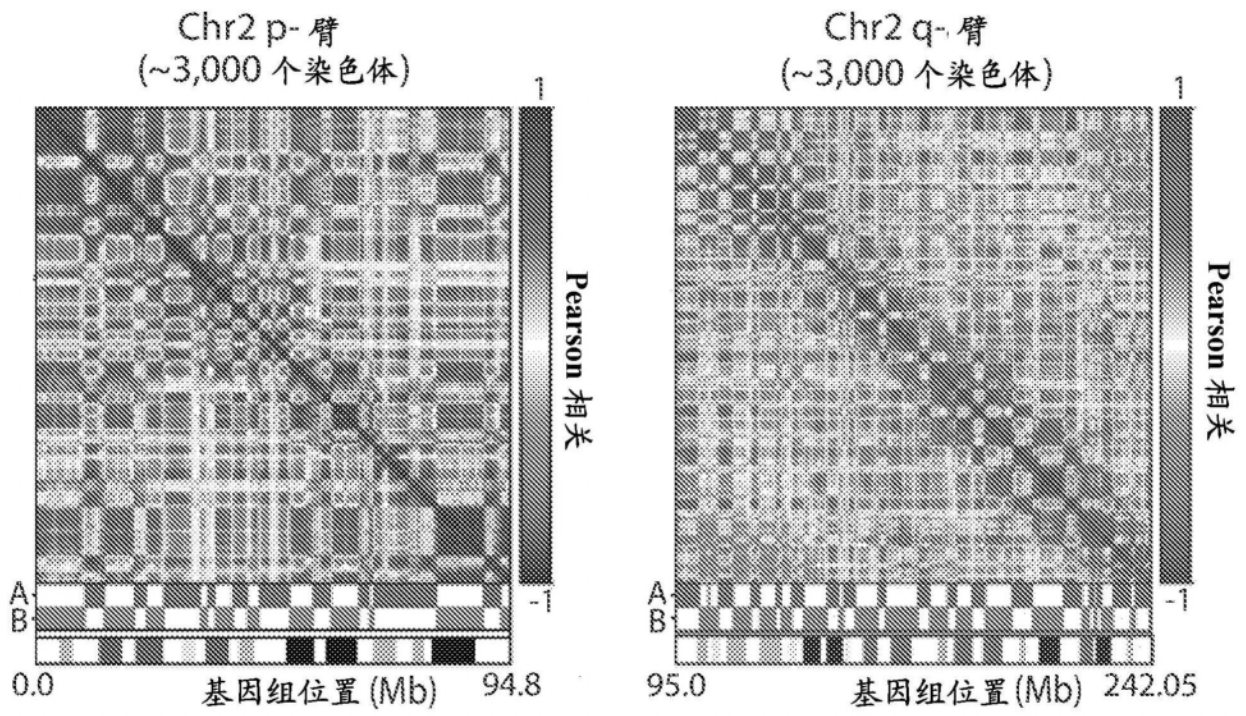


图18D

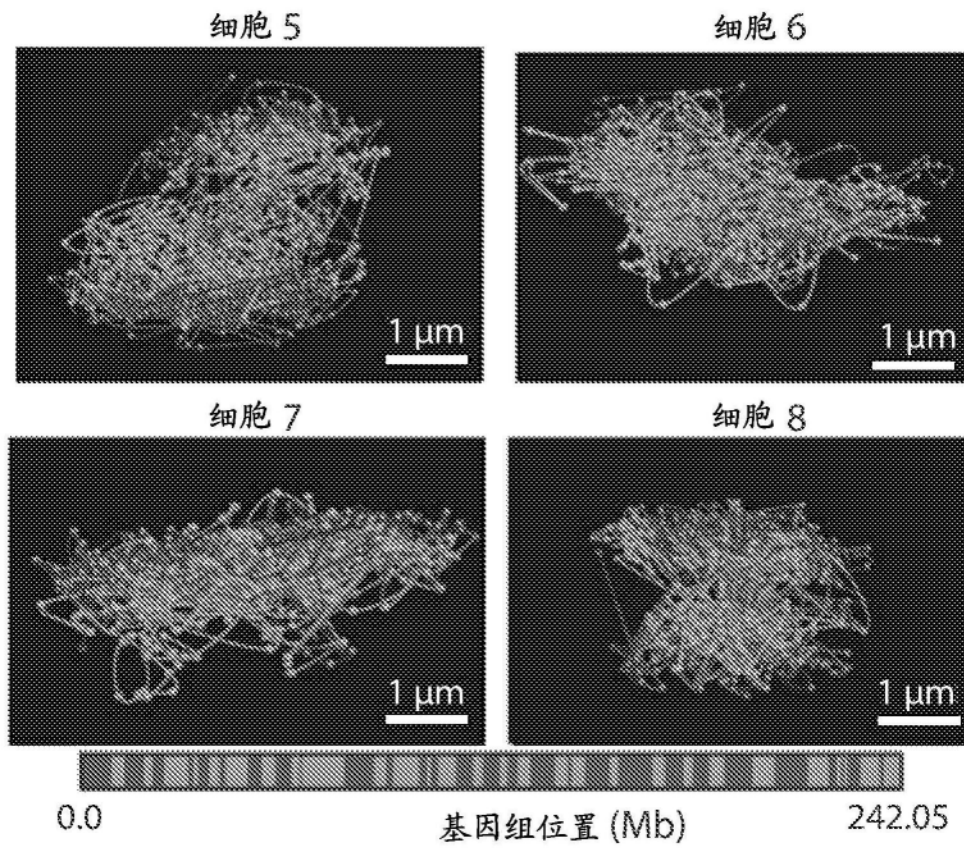


图18E

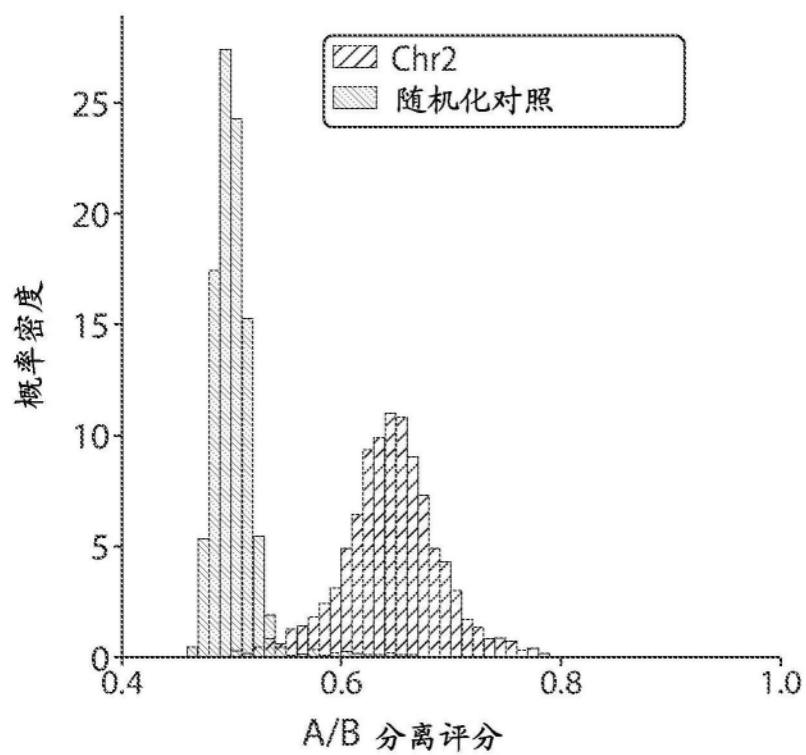


图18F

细胞 9

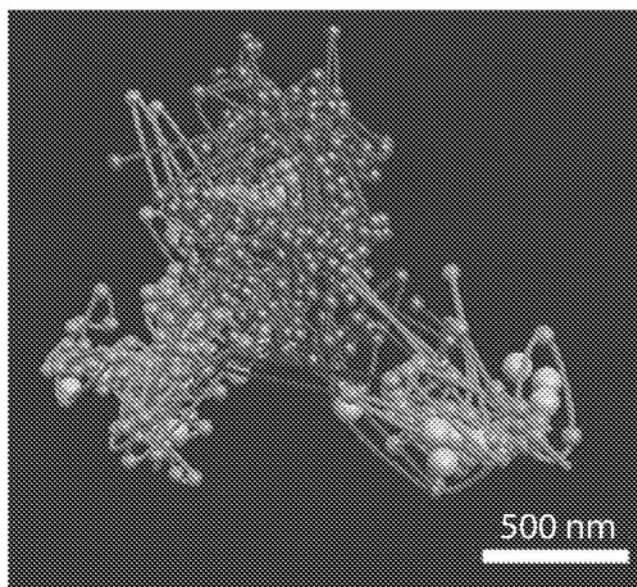


图18G

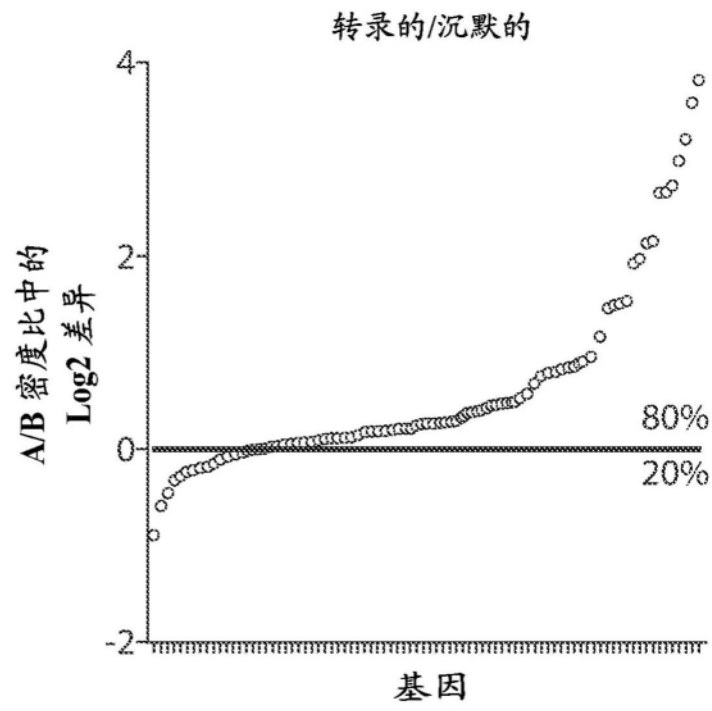


图18H

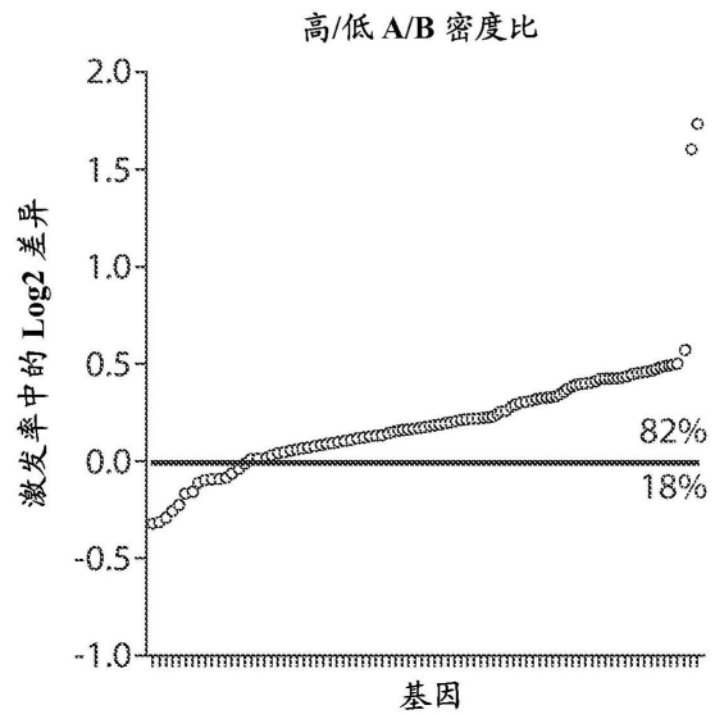


图18I

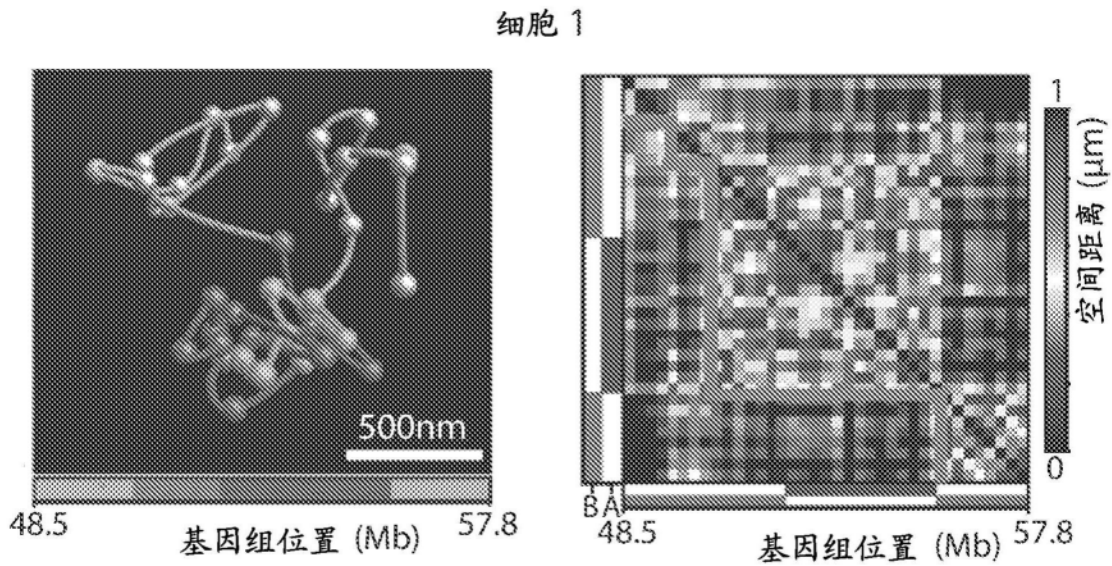


图19A

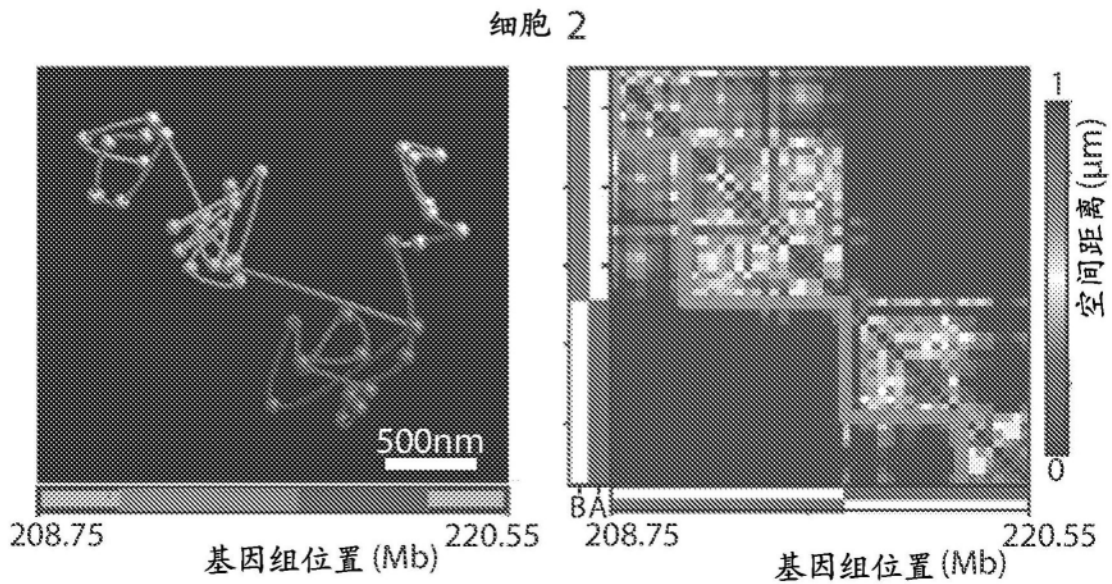


图19B

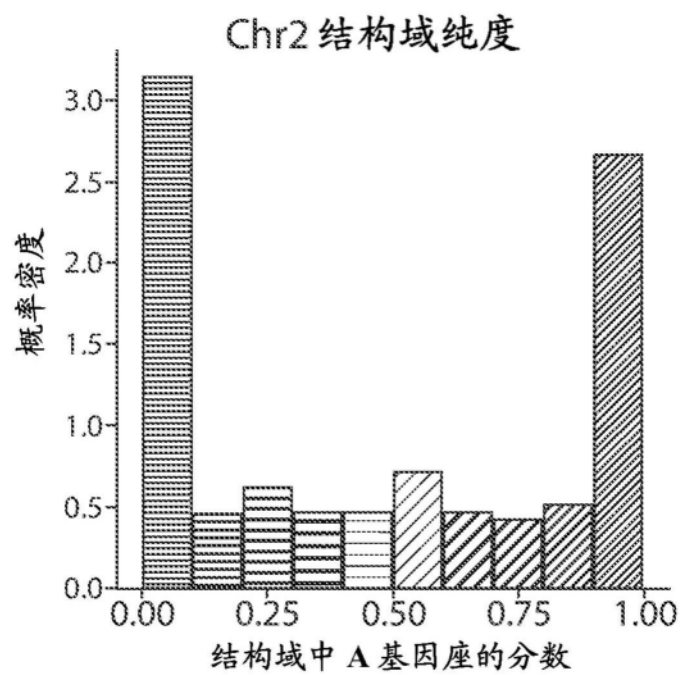


图19C

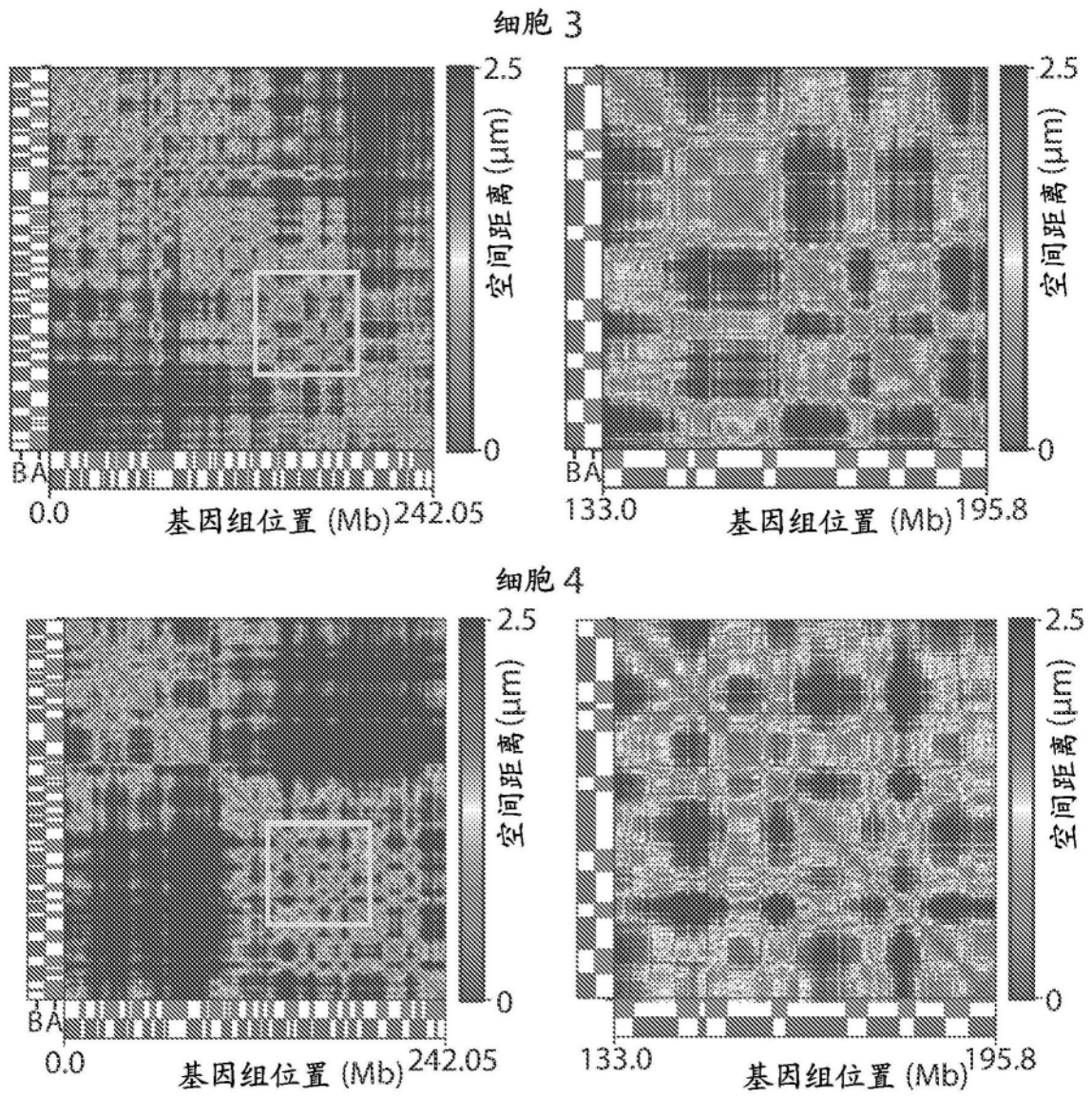


图19D

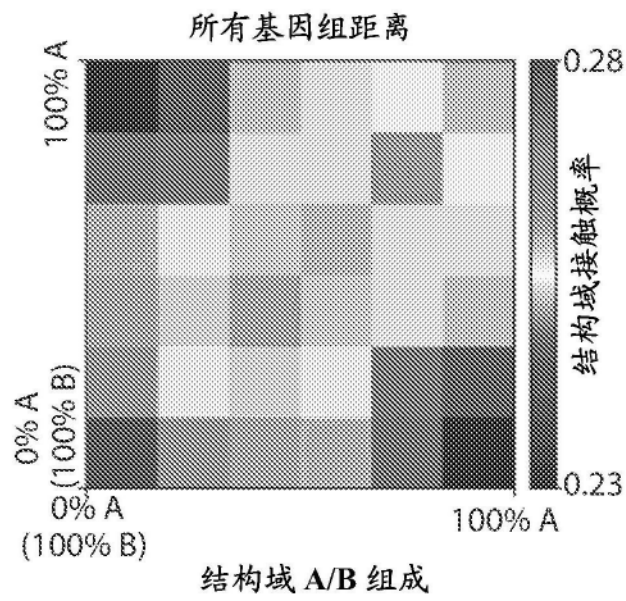


图19E

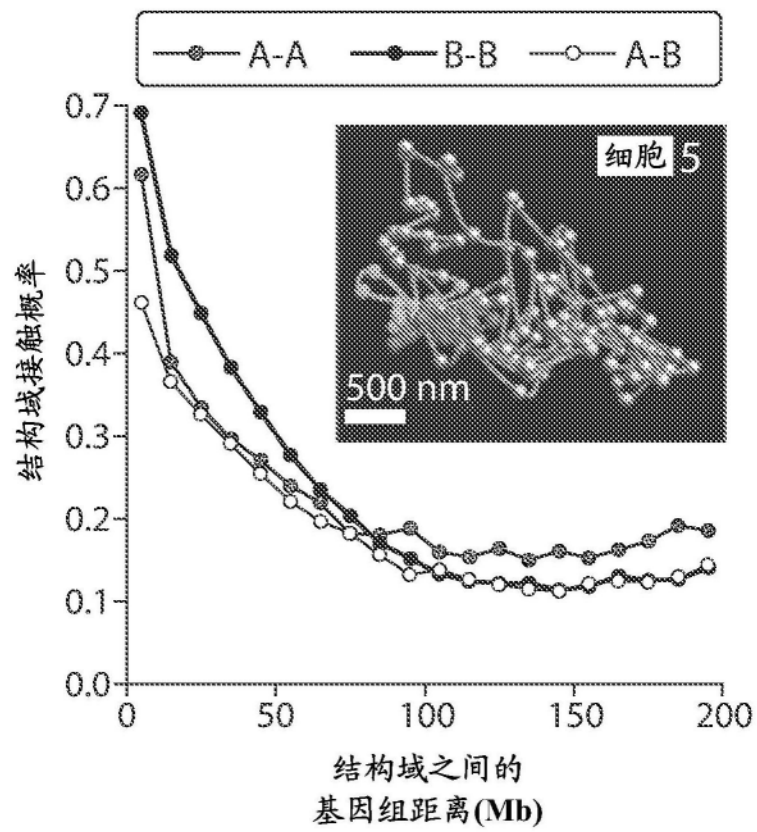


图19F

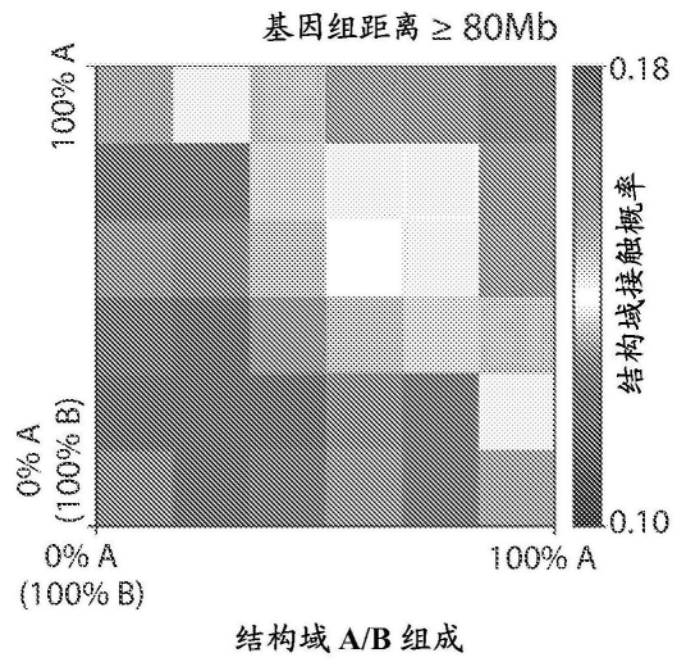


图19G

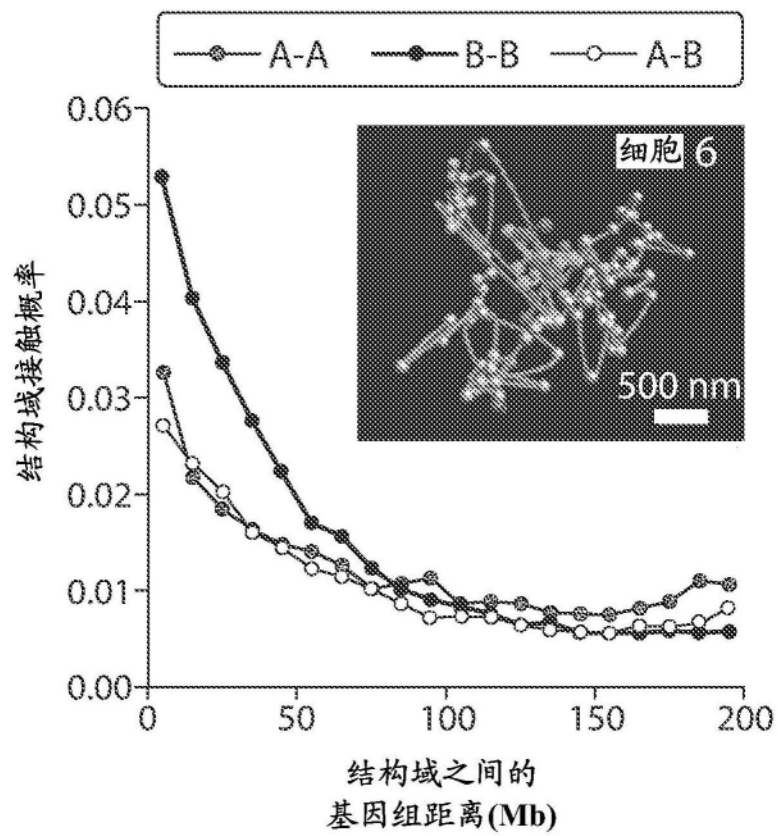


图19H

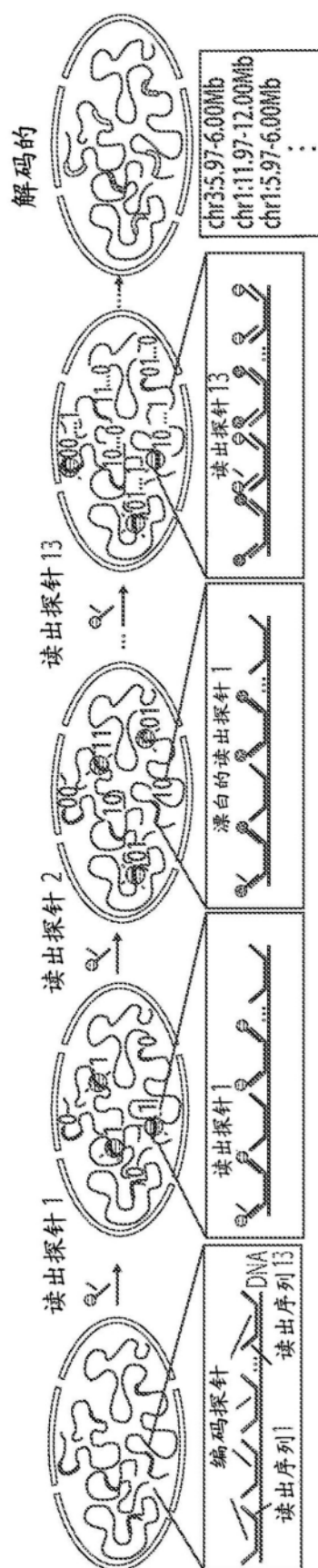


图20A

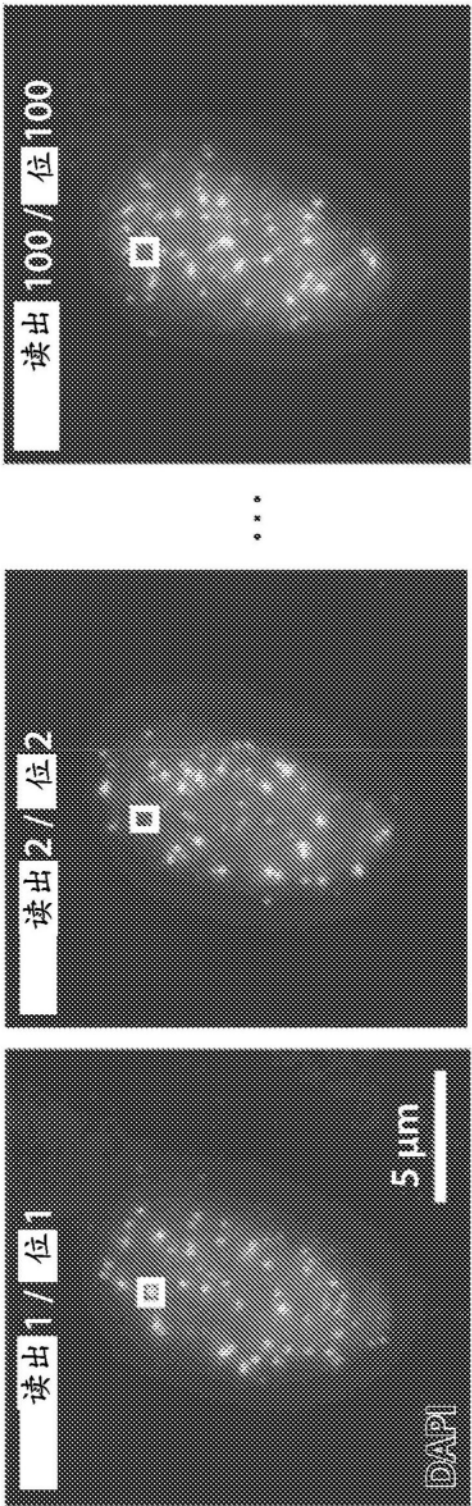


图20B

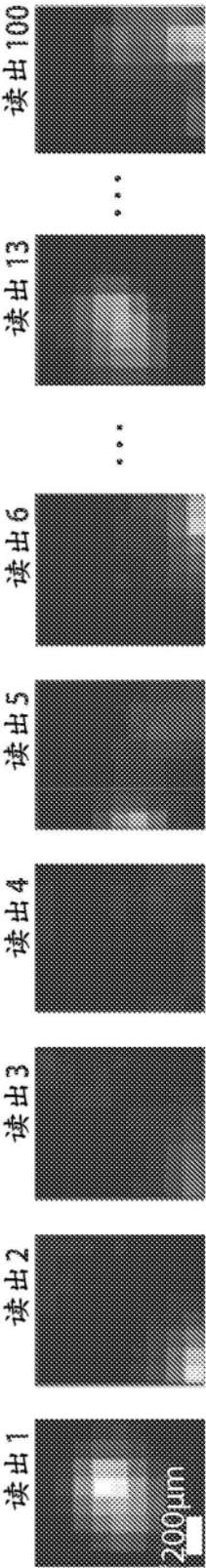


图20C

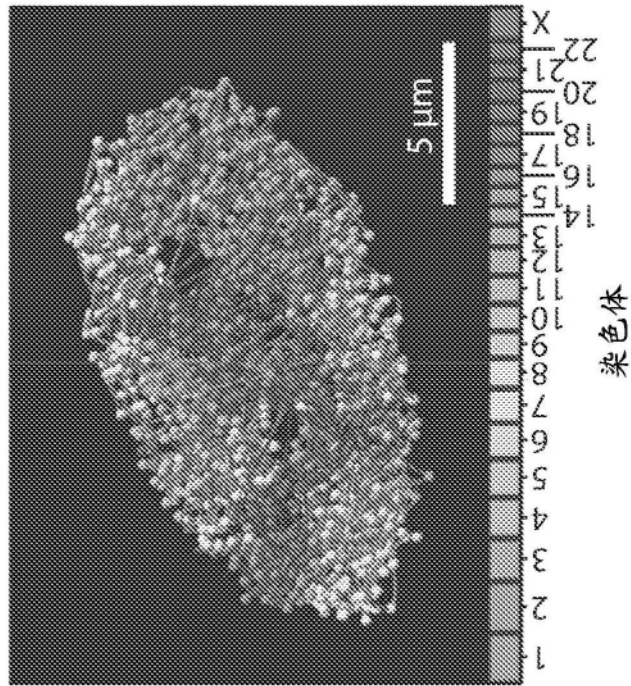


图20D

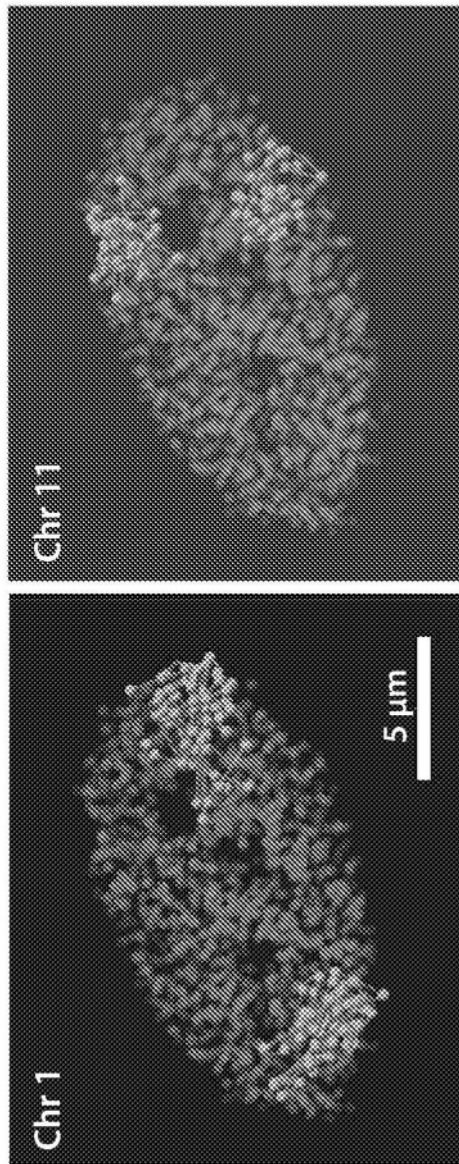


图20E

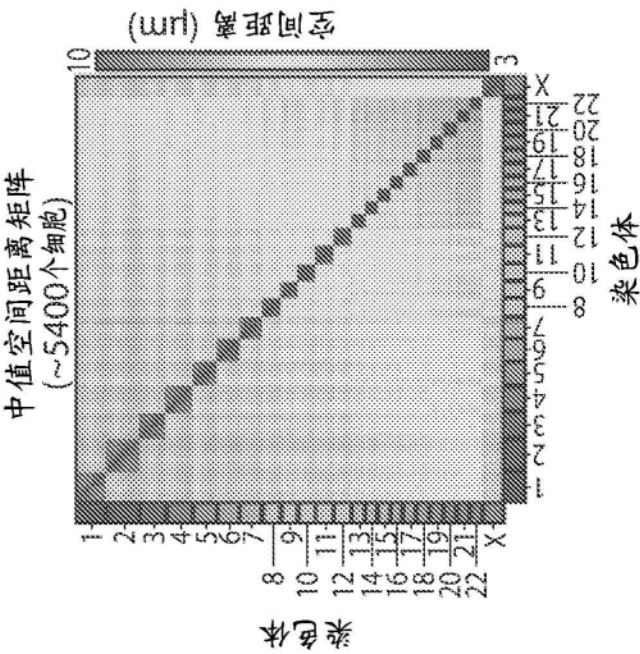


图20F

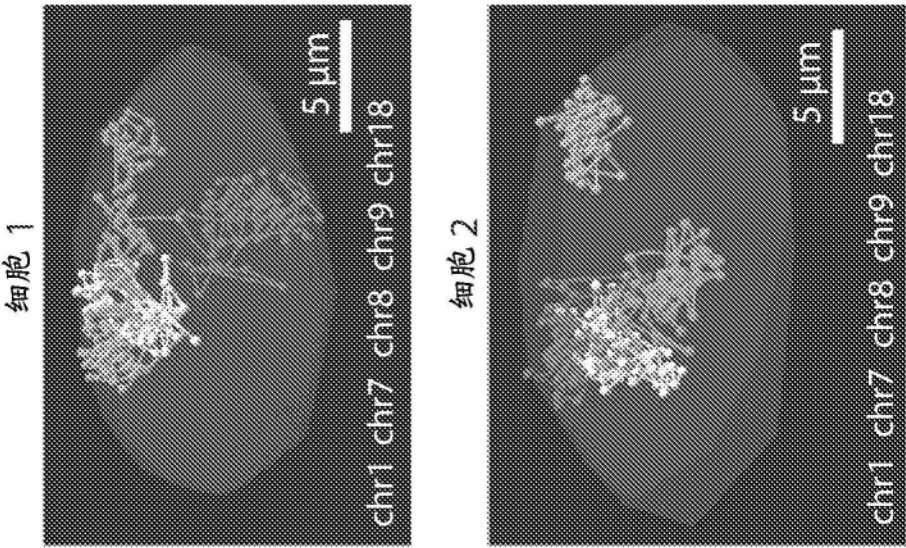


图20G

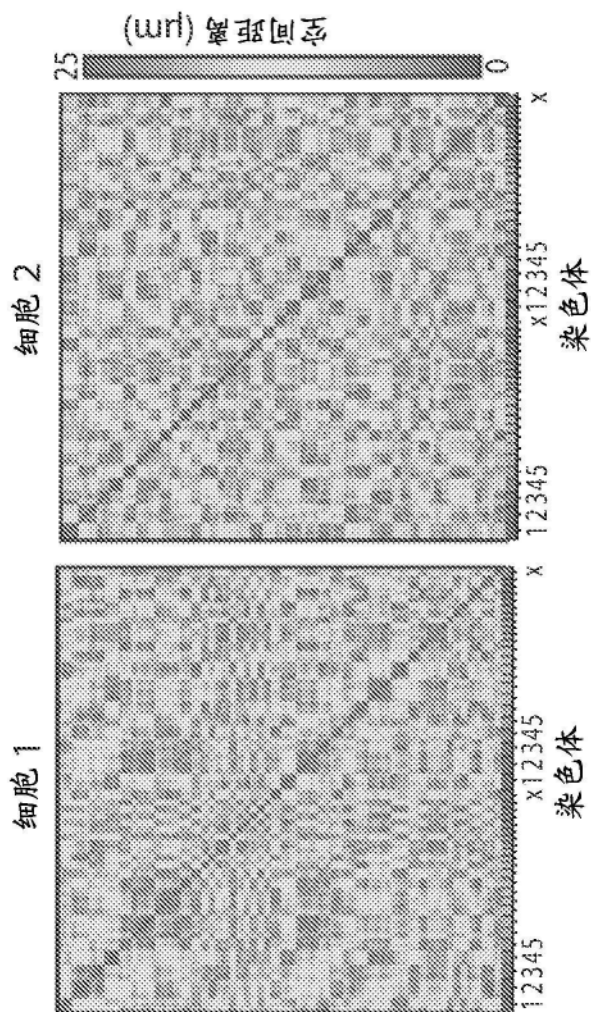


图20H

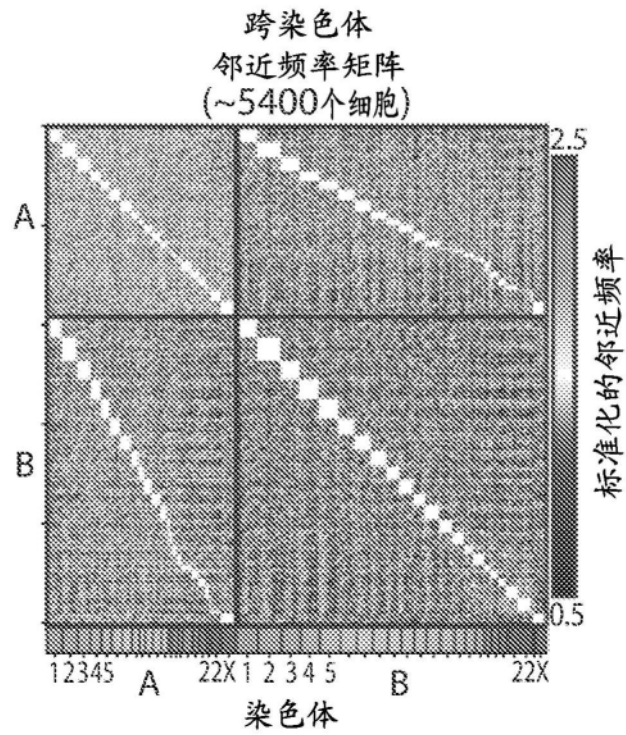


图21A

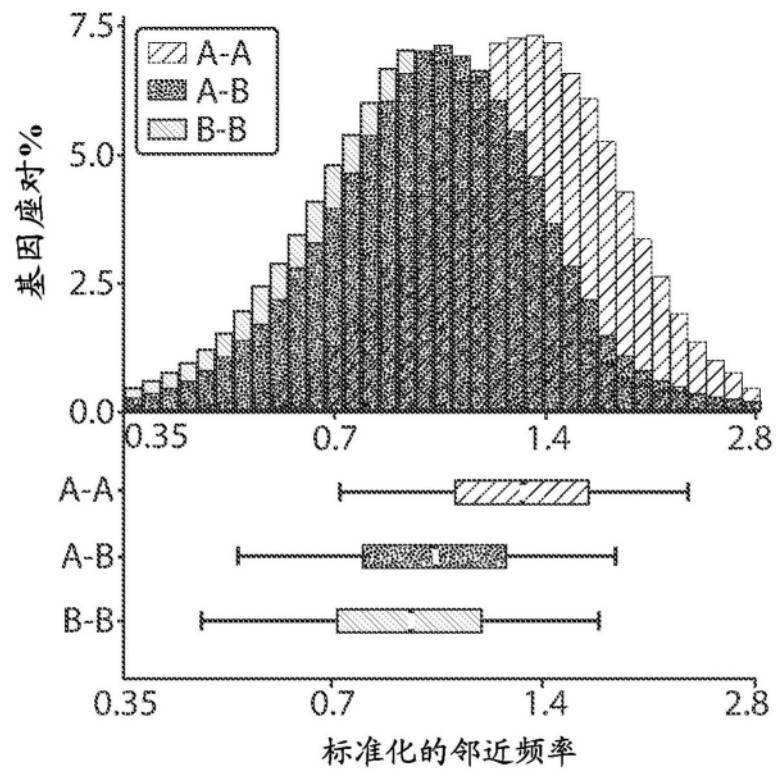


图21B

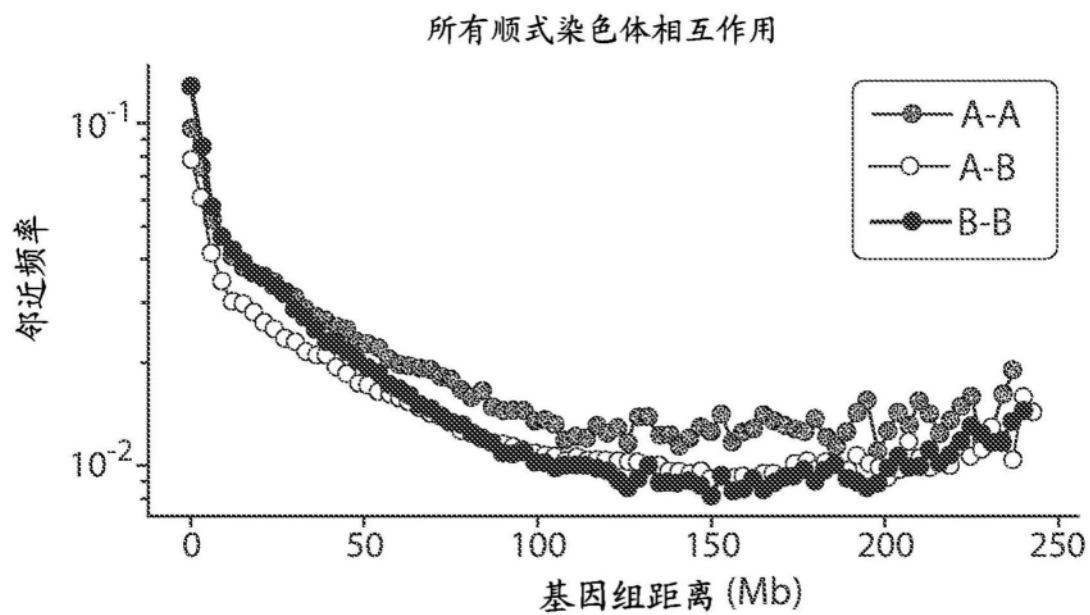


图21C

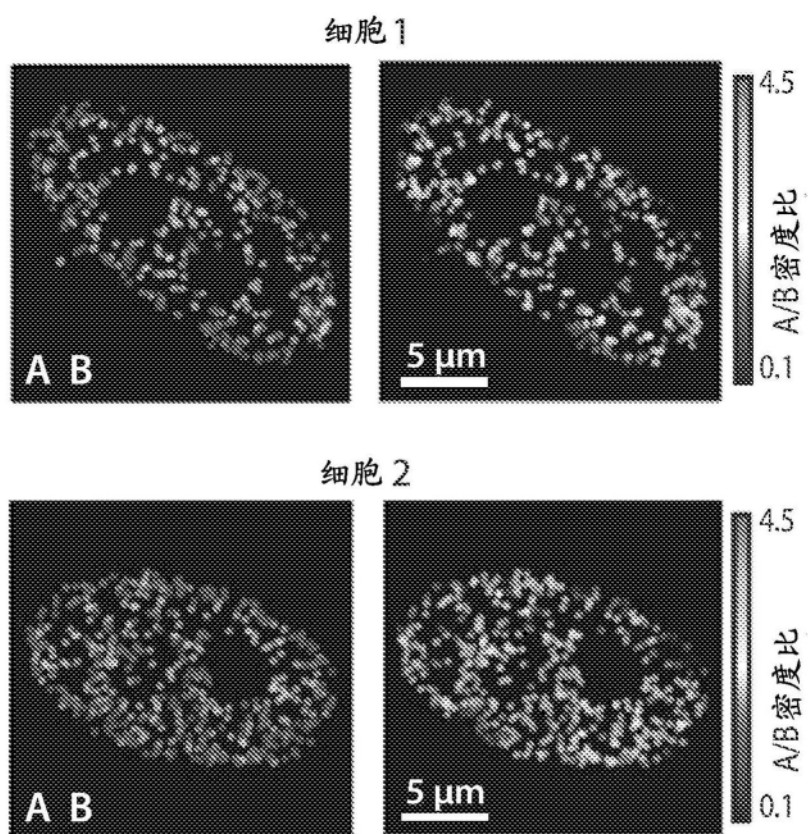


图21D

在基因组基因座上的分布

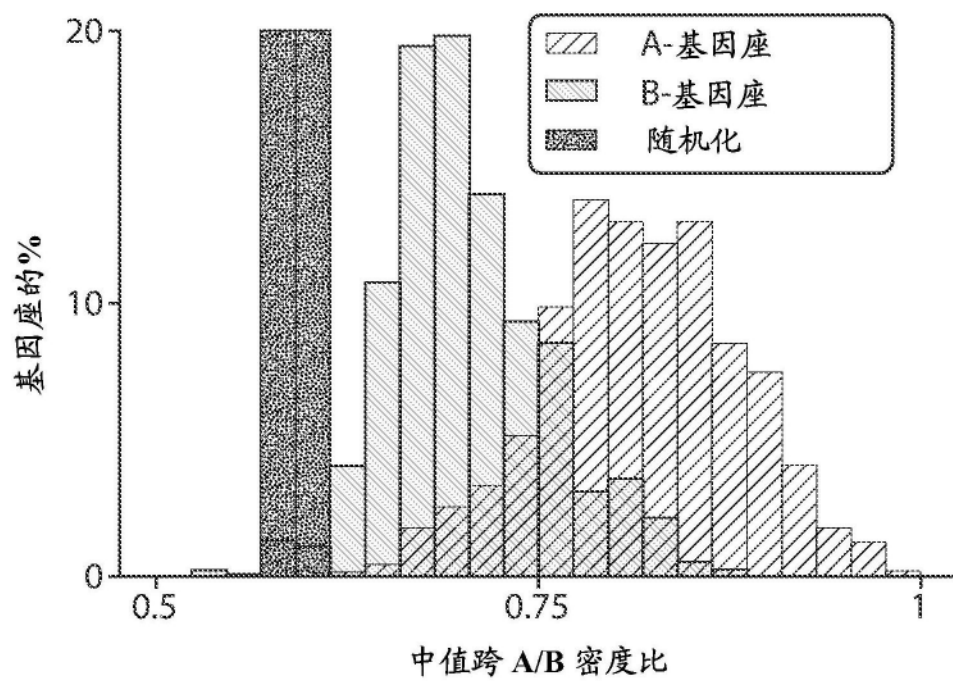


图21E

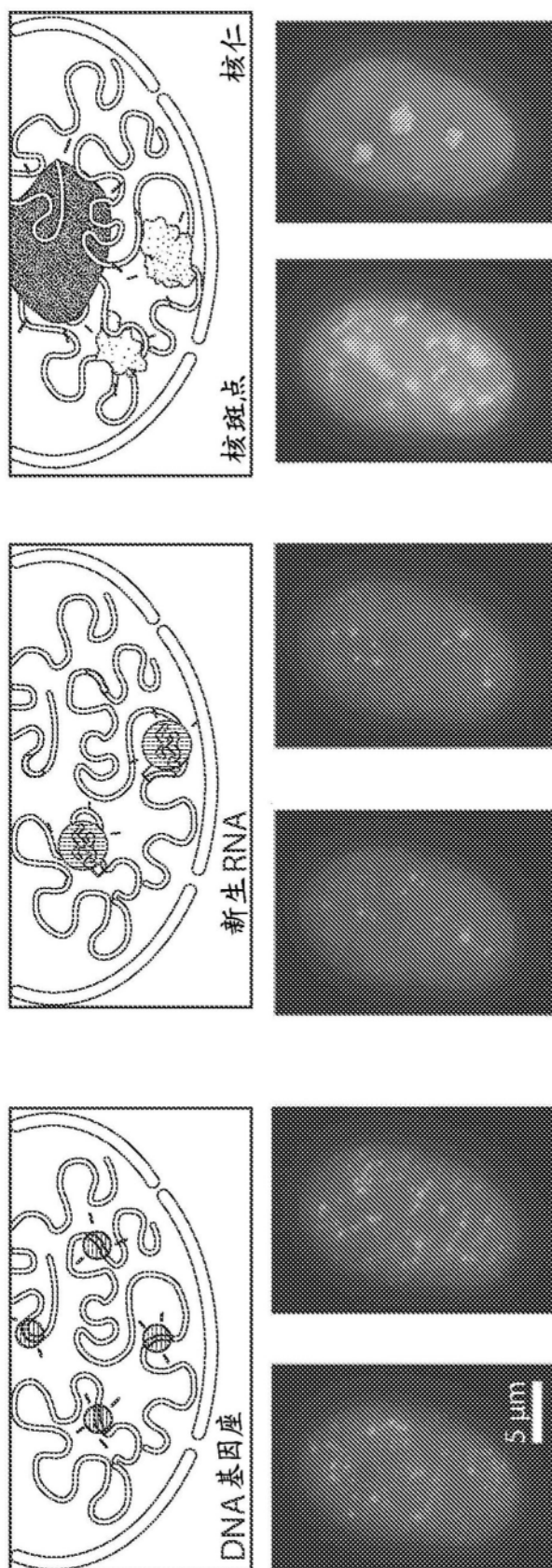


图22A

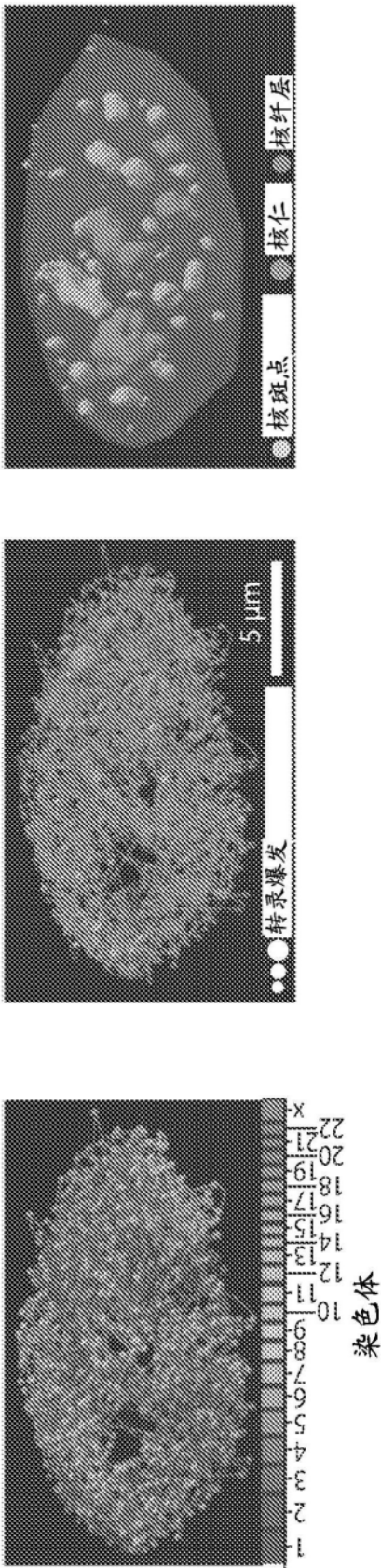


图22B

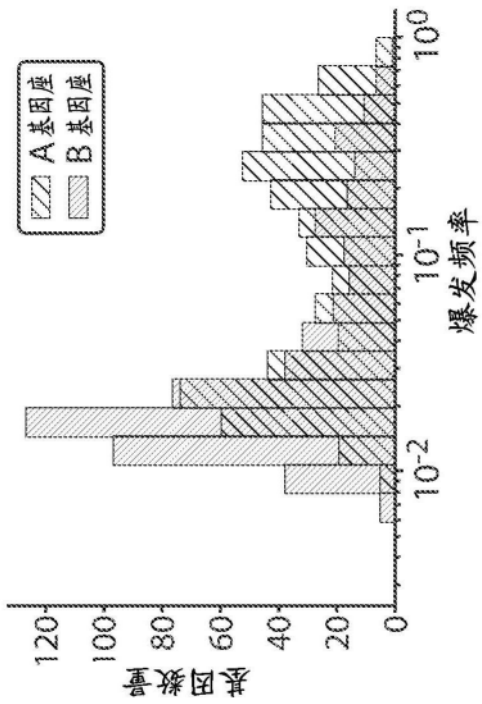


图22C

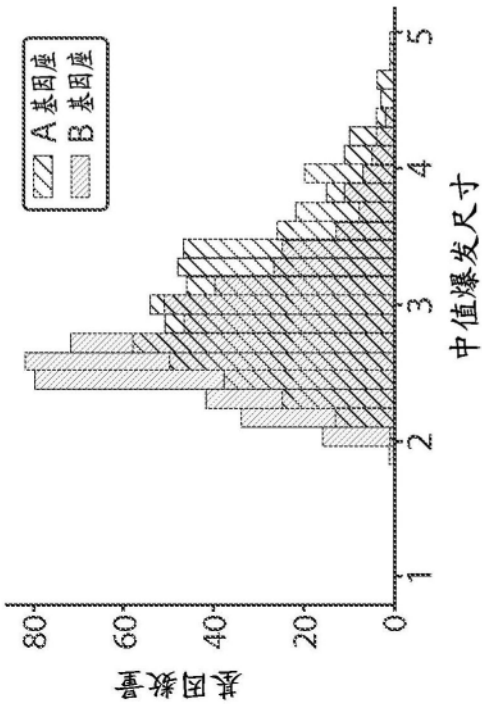


图22D

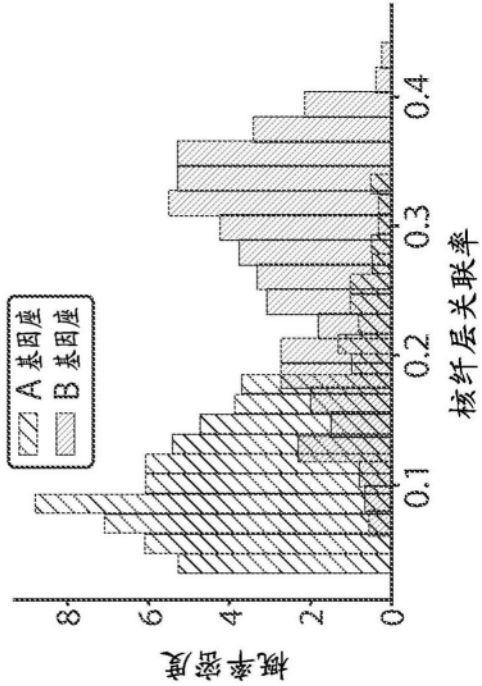


图22E

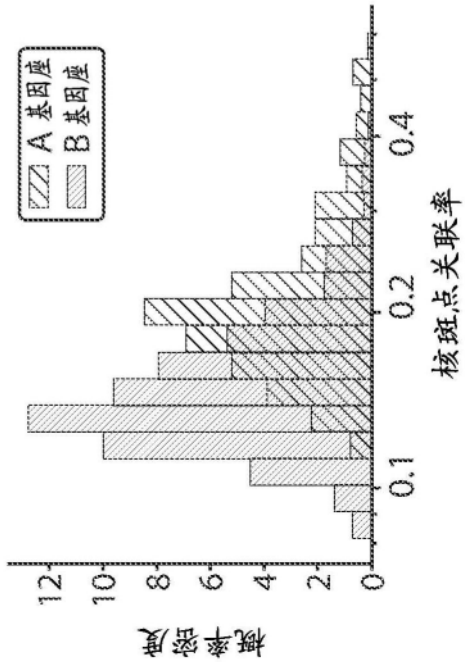


图22F

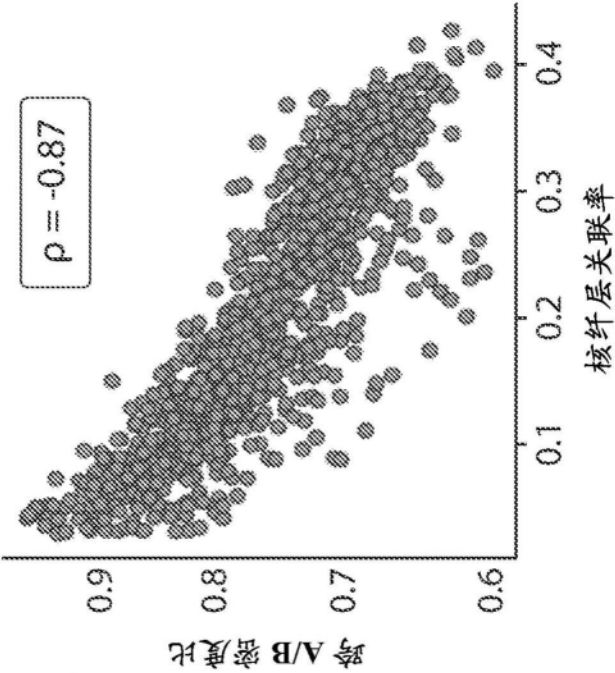


图22G

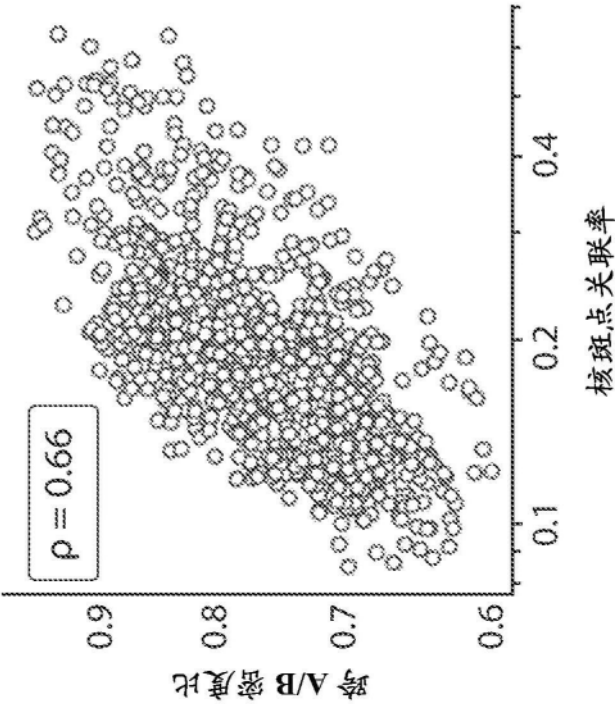


图22H

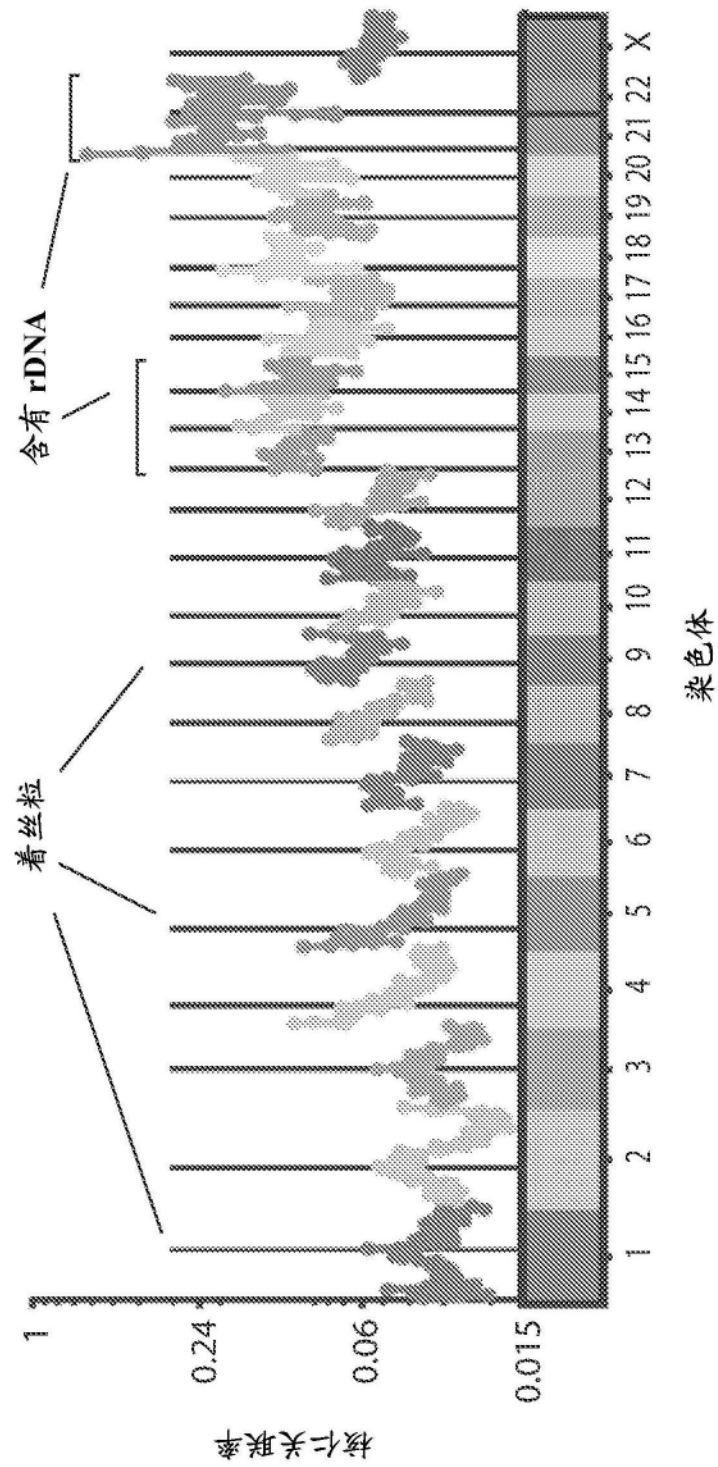


图22I

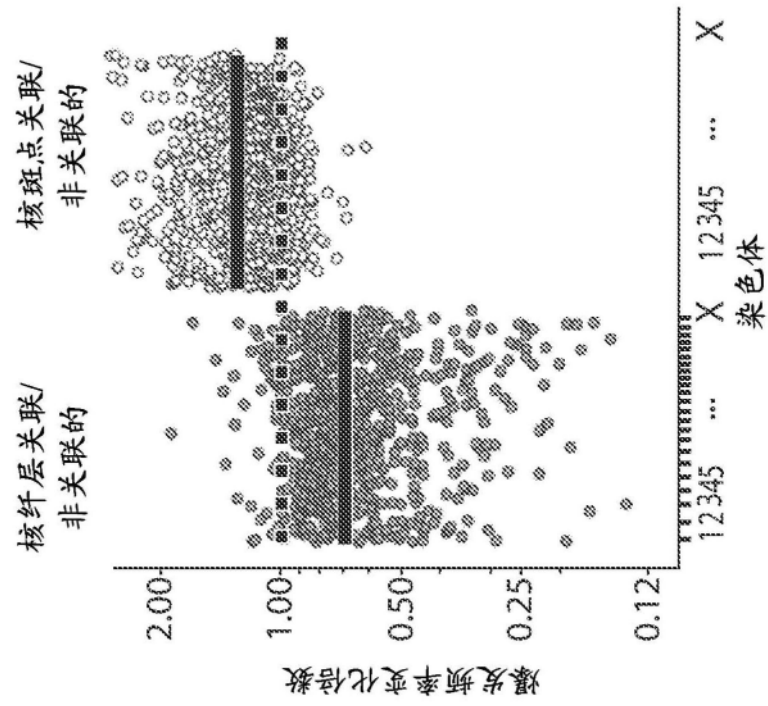


图22J

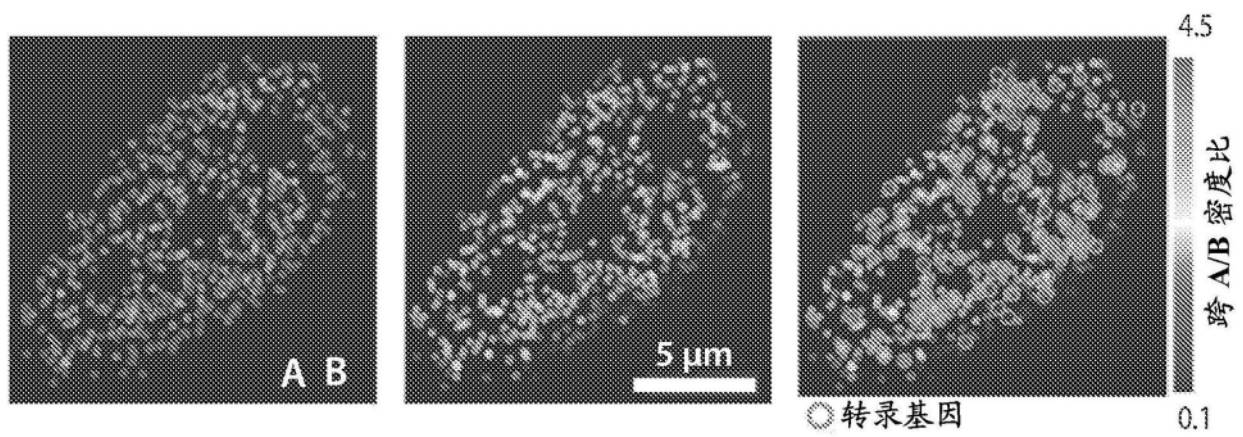


图23A

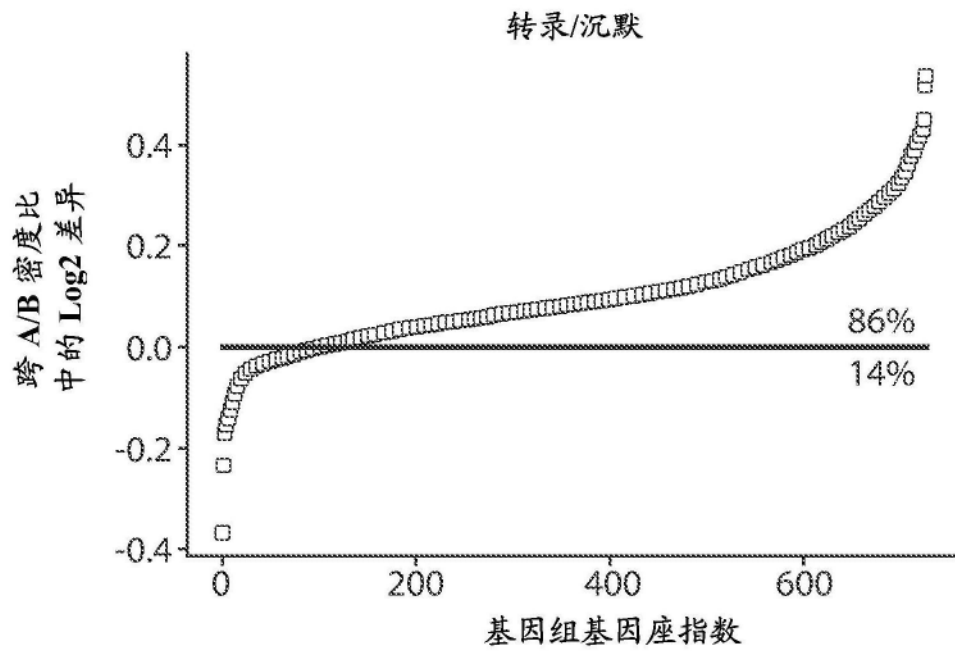


图23B

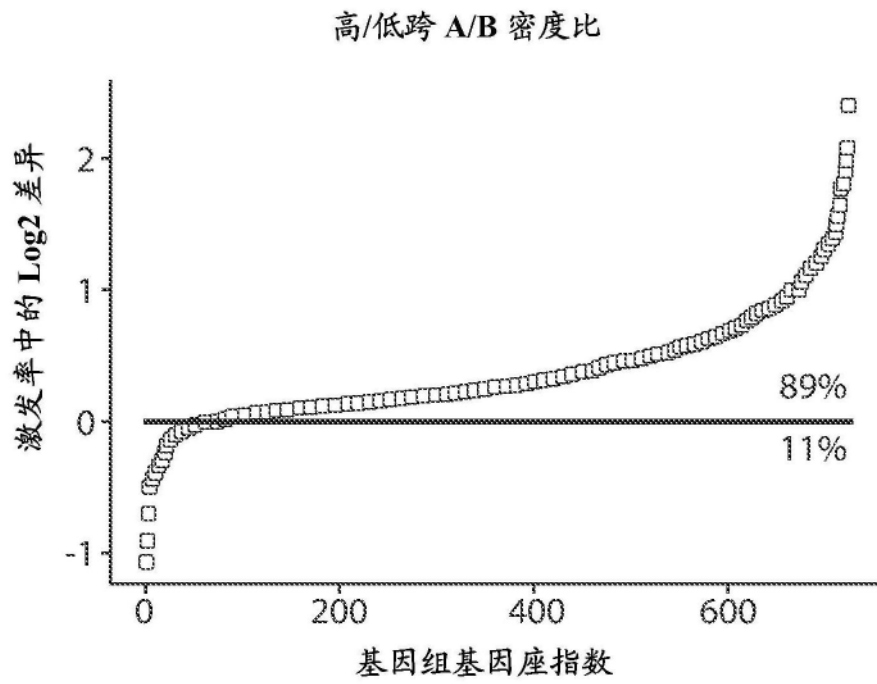


图23C

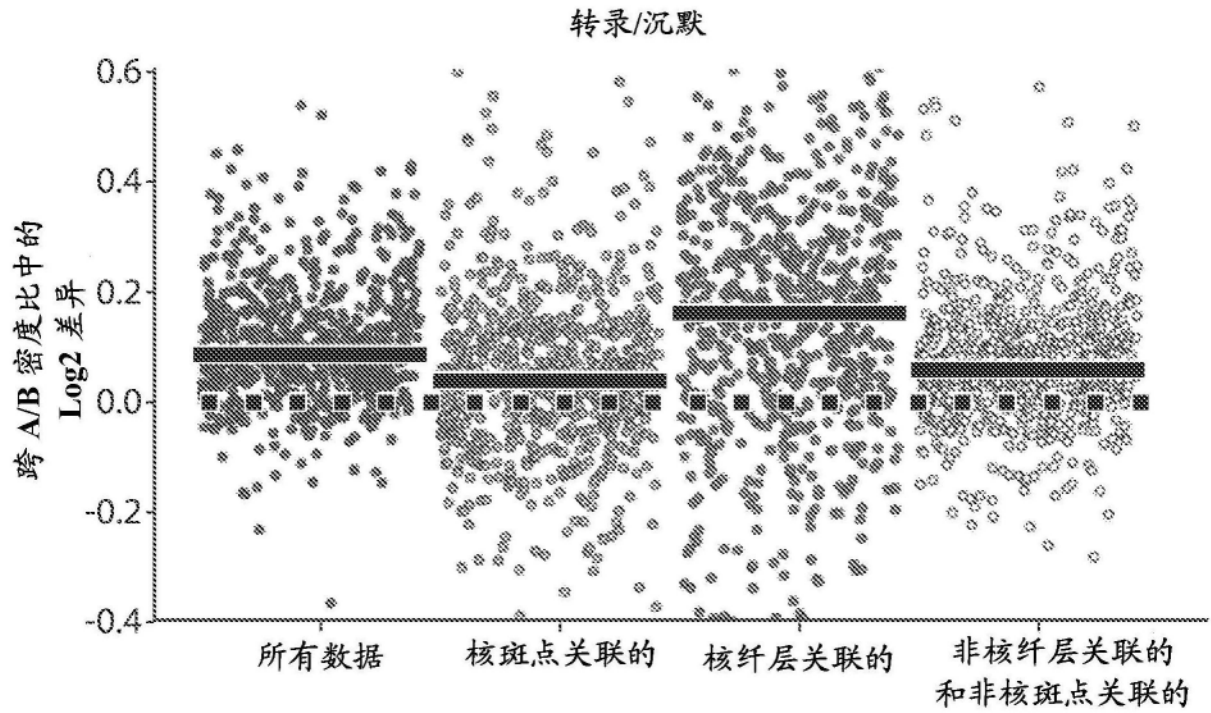


图23D

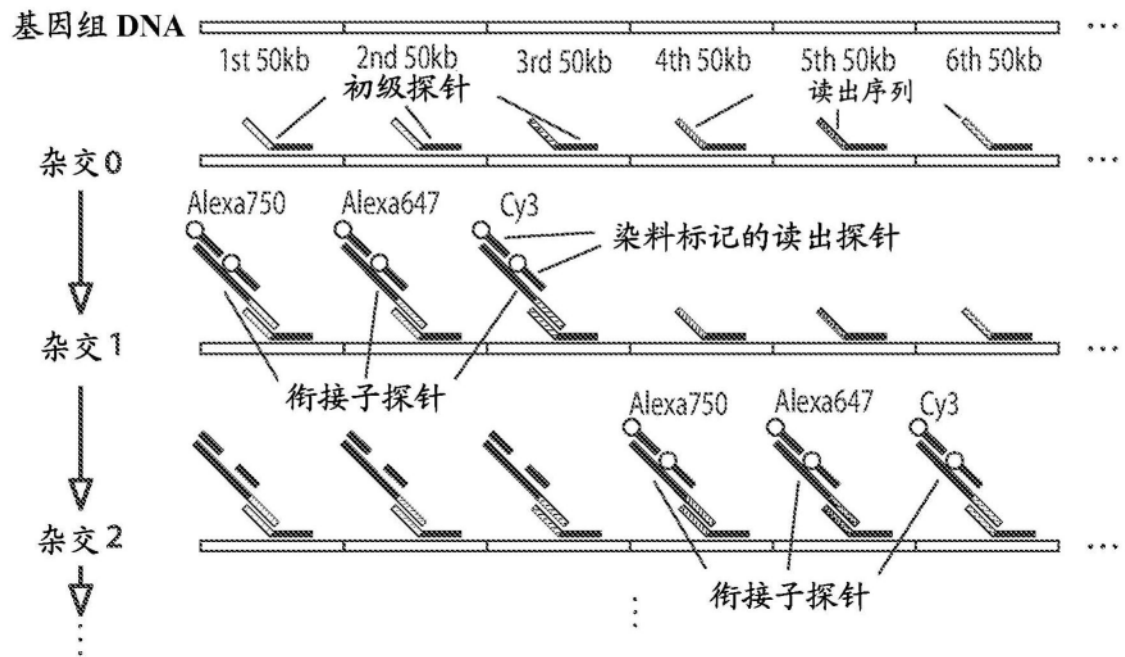


图24A

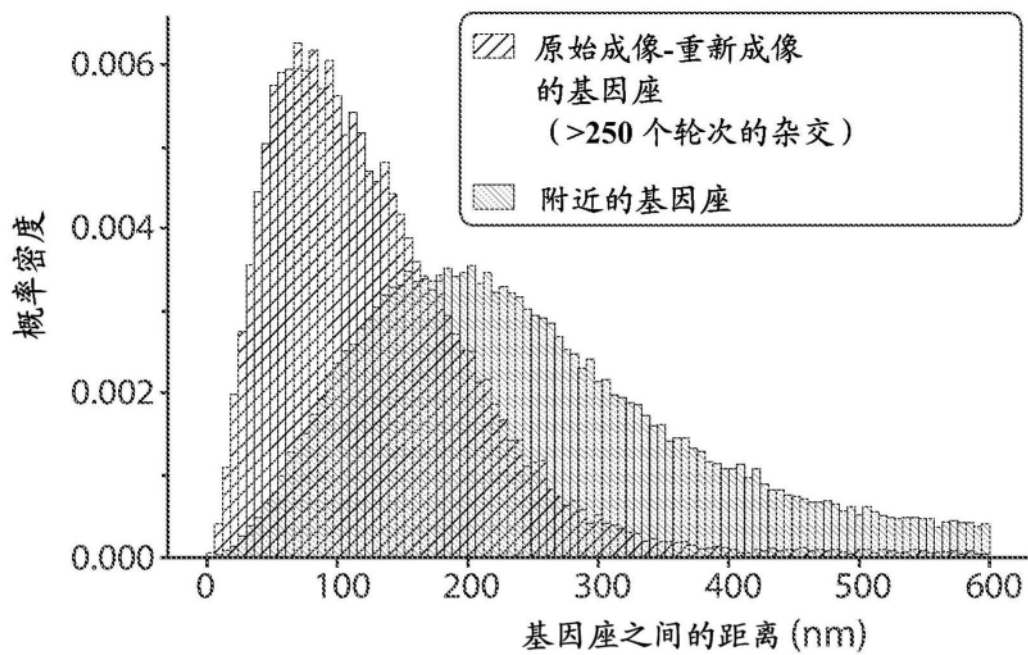


图24B

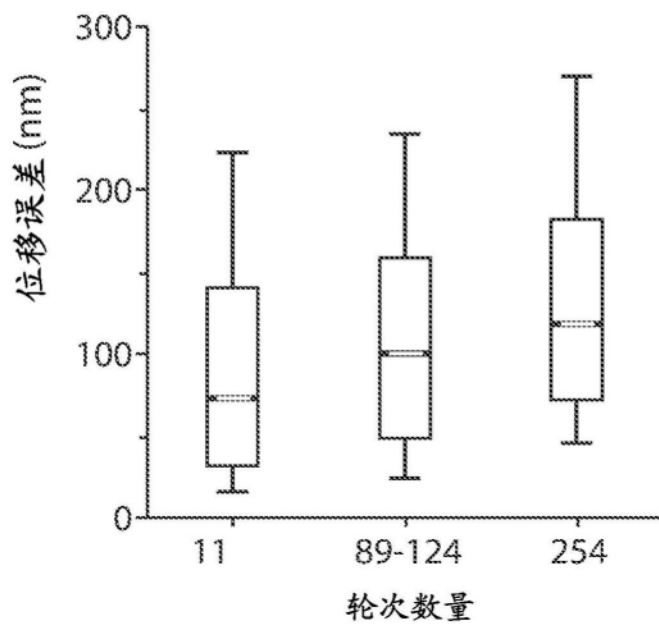


图24C

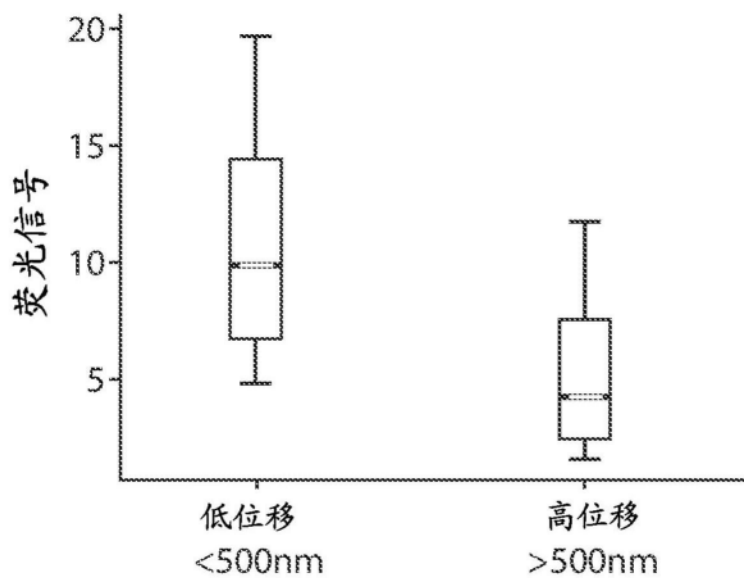


图24D

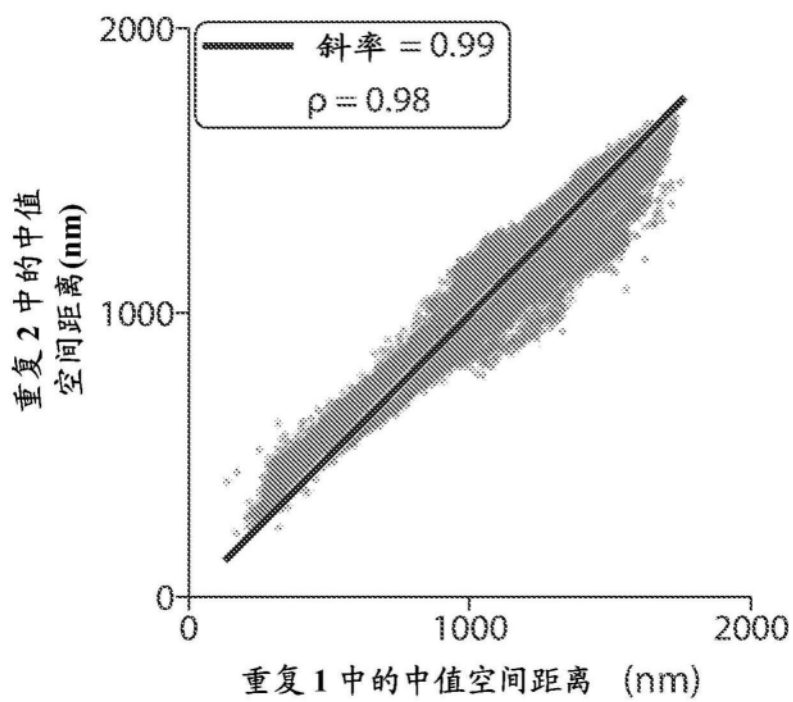


图24E

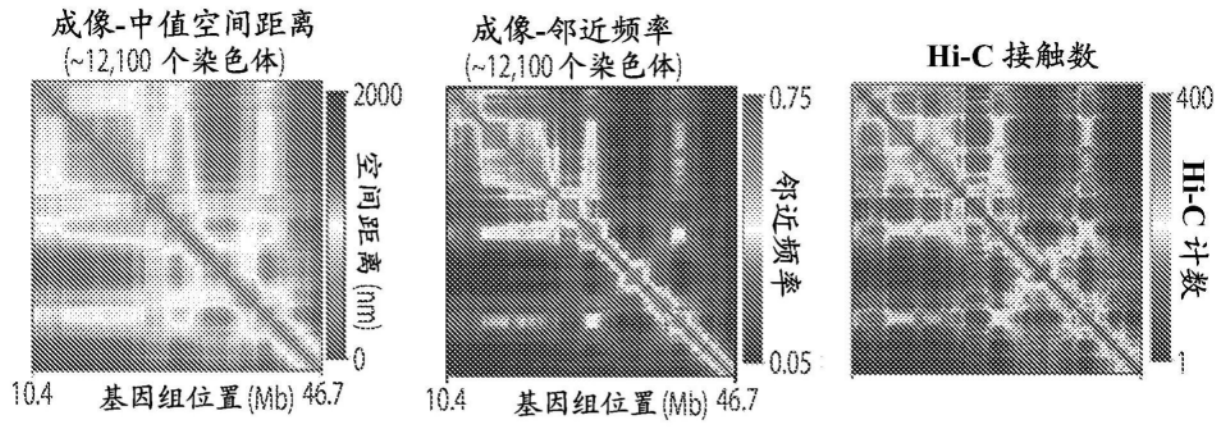


图24F

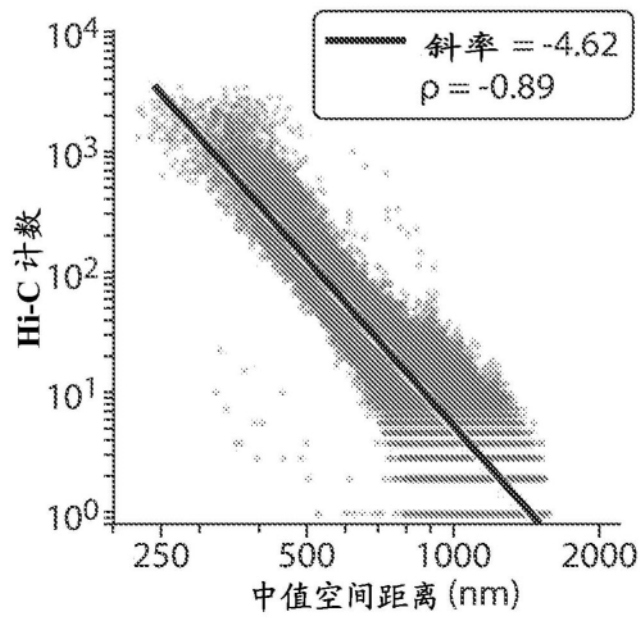


图24G

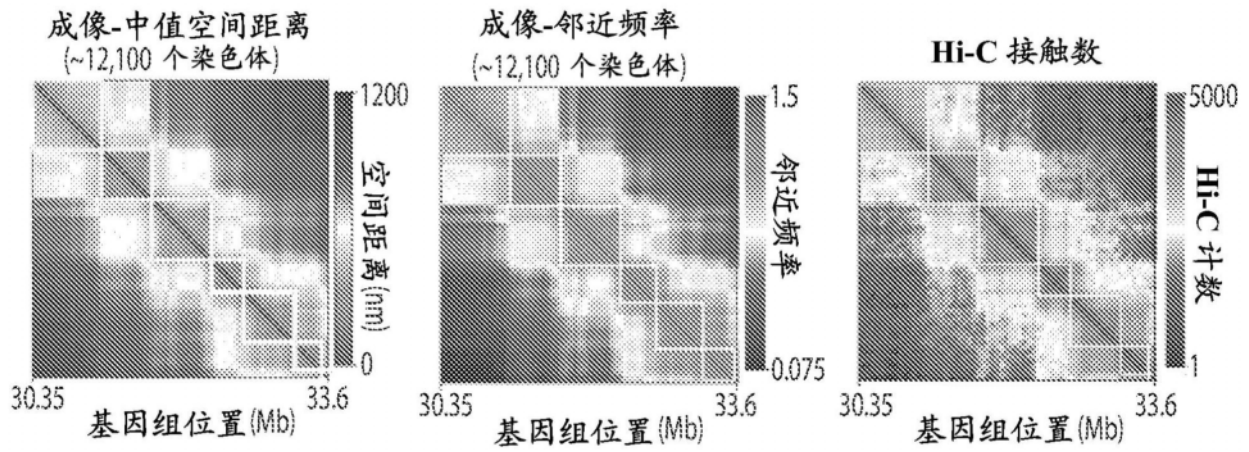


图24H

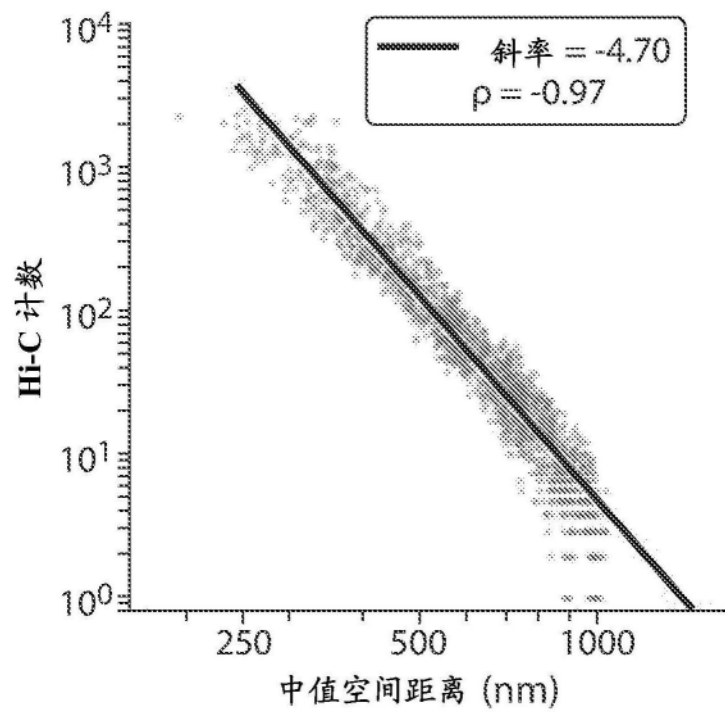


图24I

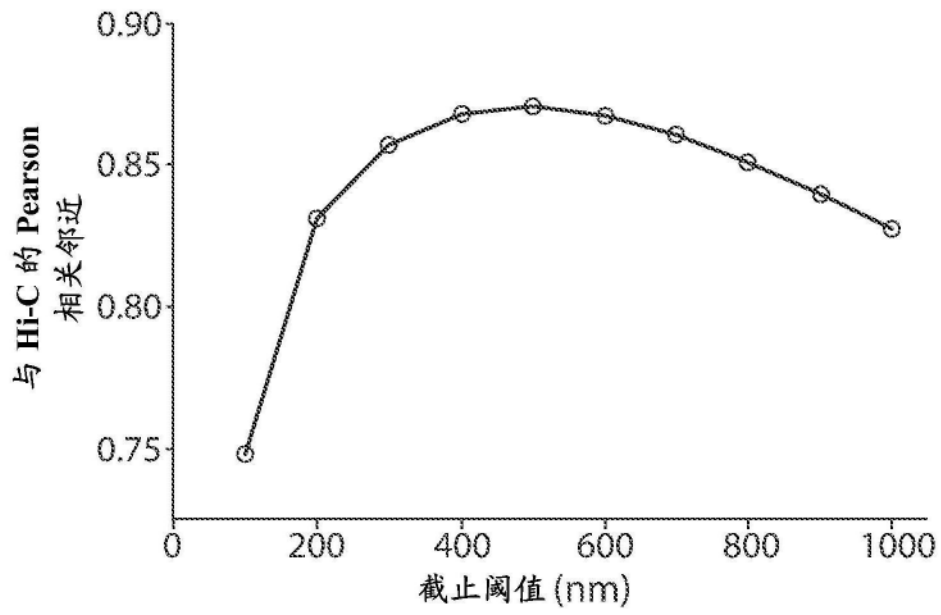


图24J

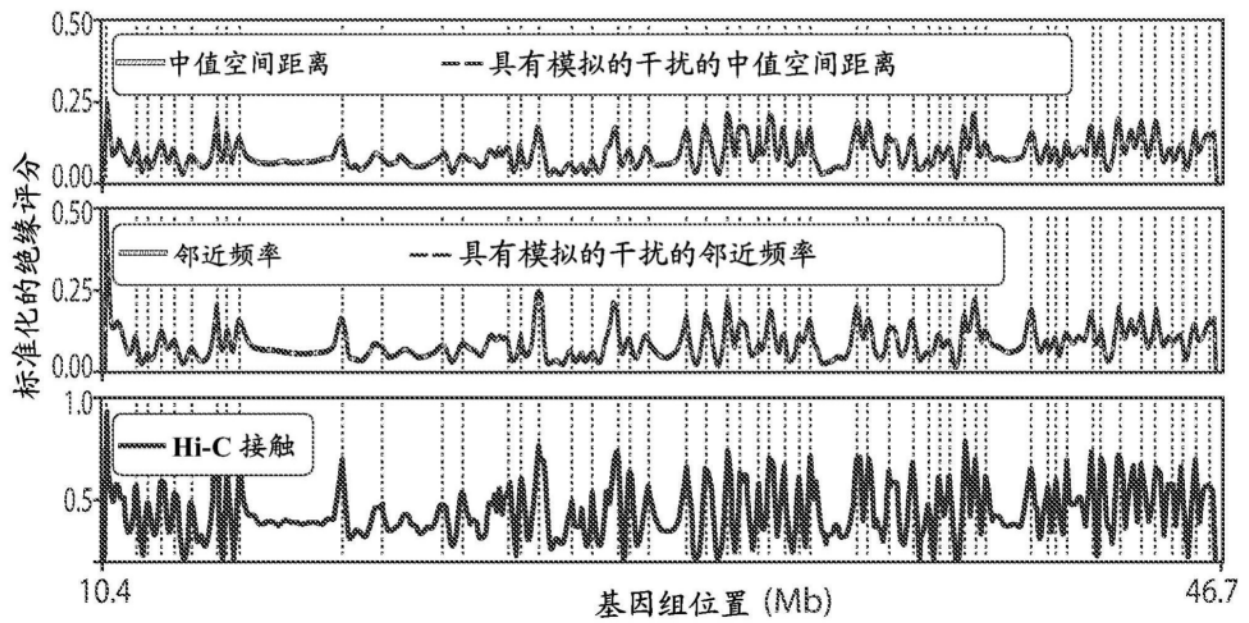


图24K

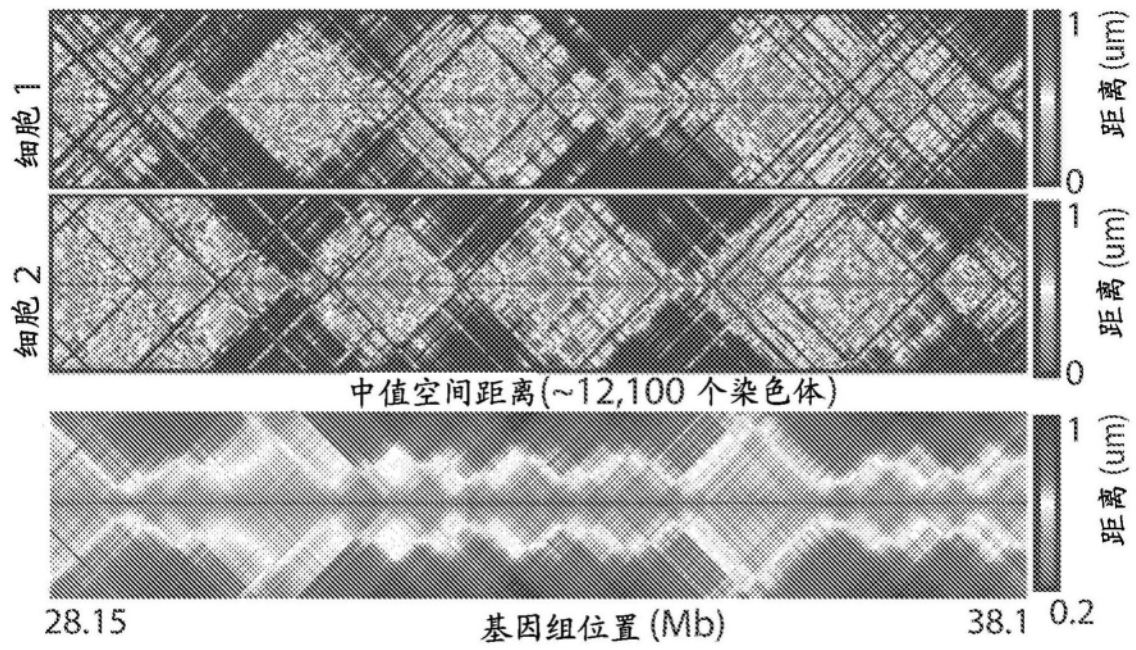


图24L

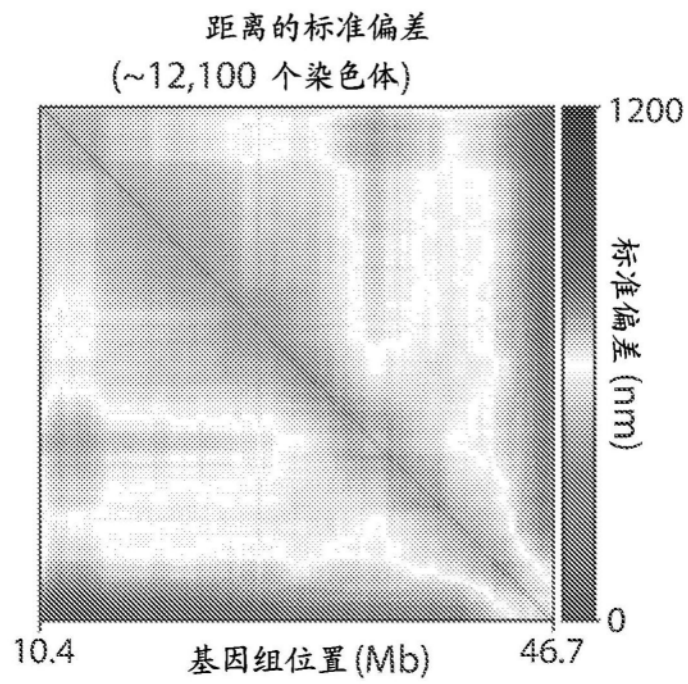


图24M

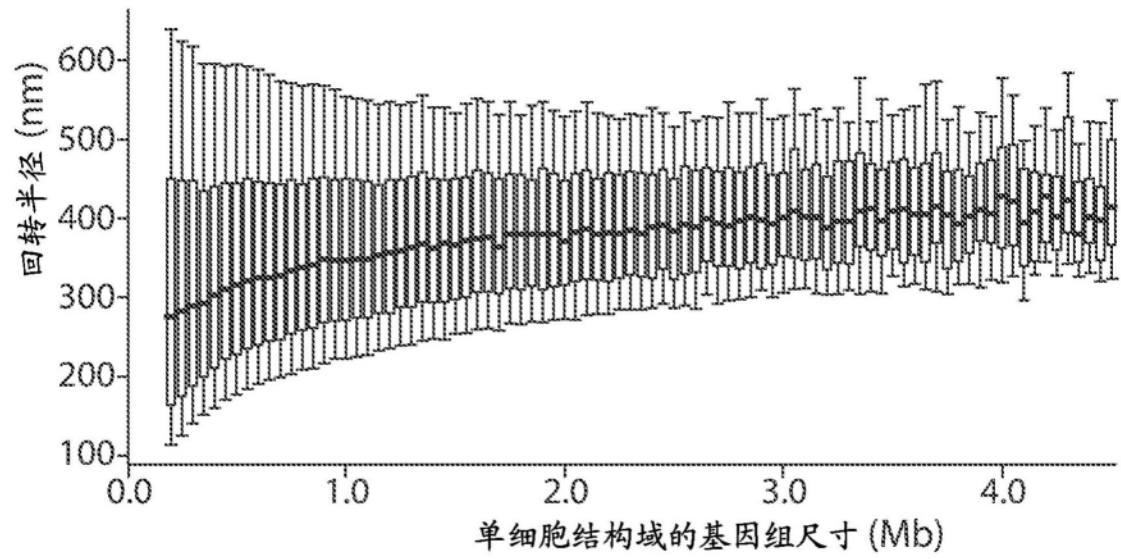


图24N

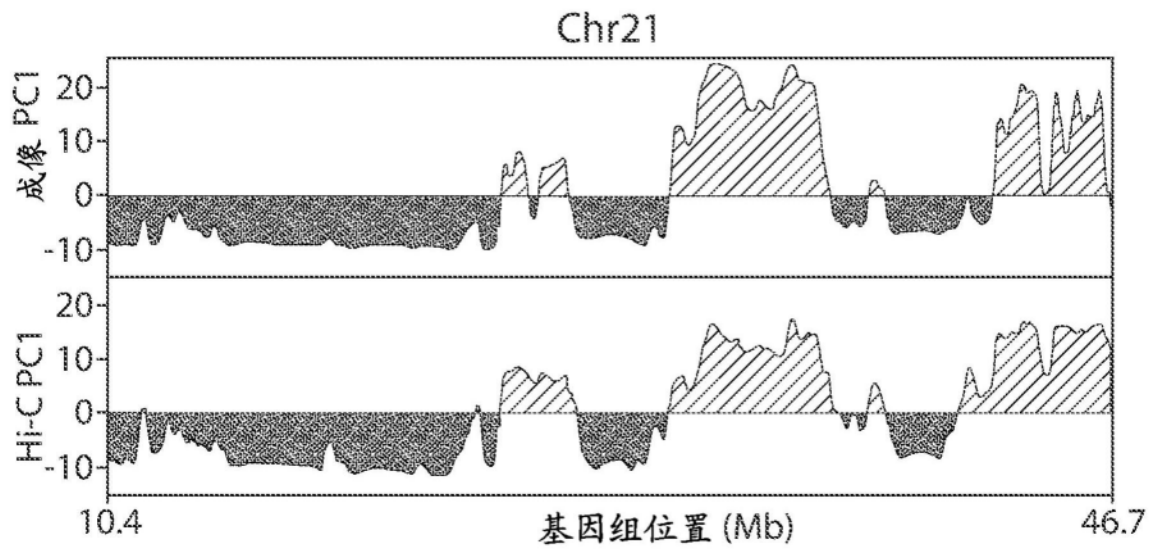


图25A

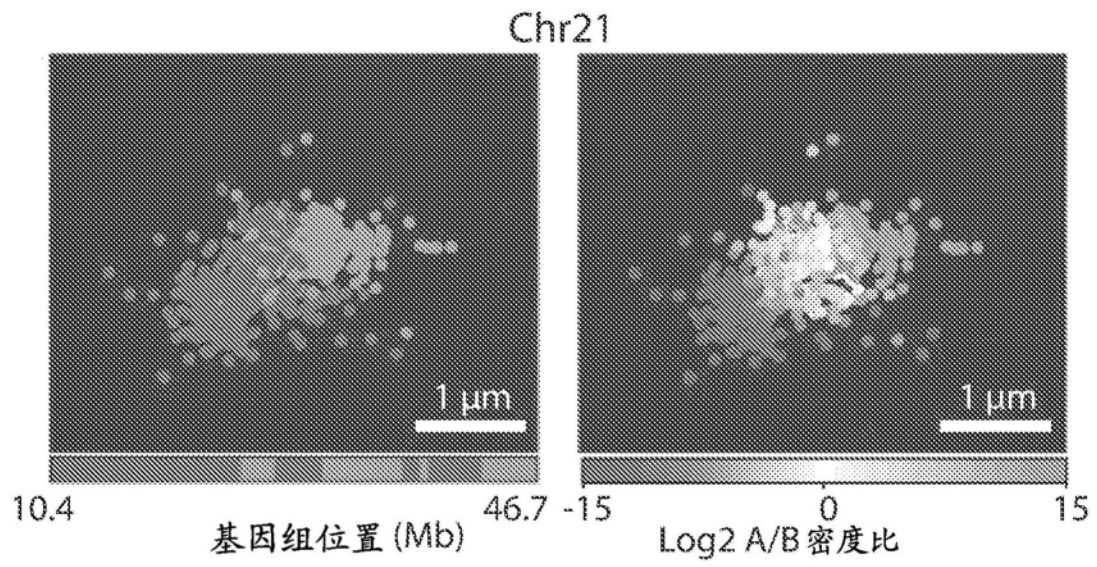


图25B

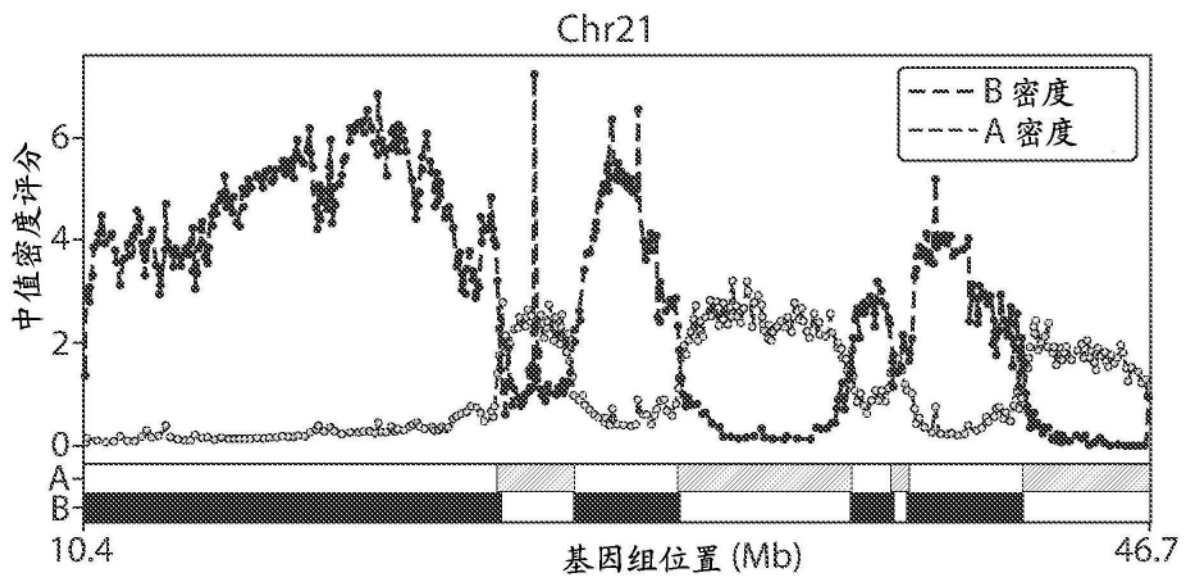
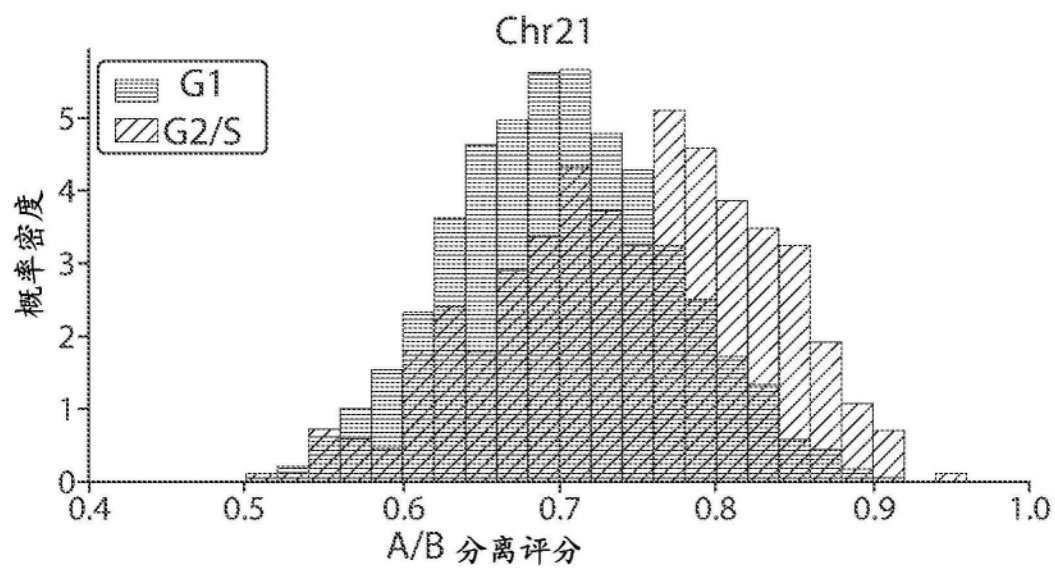


图25C



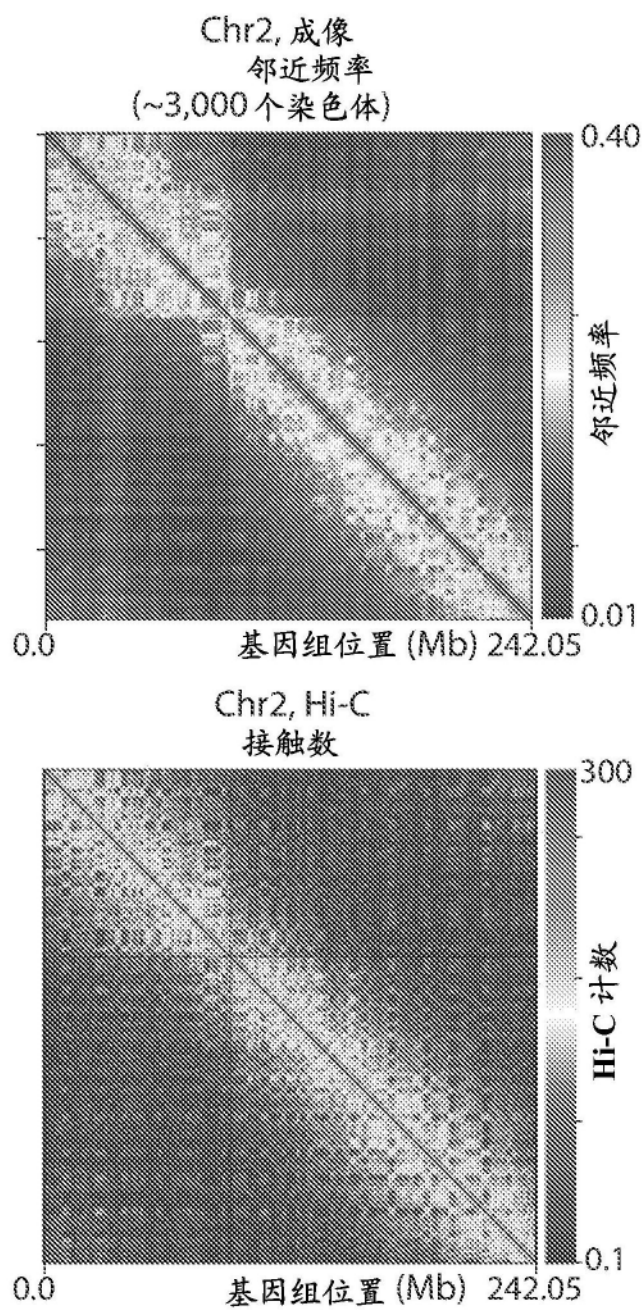


图25E

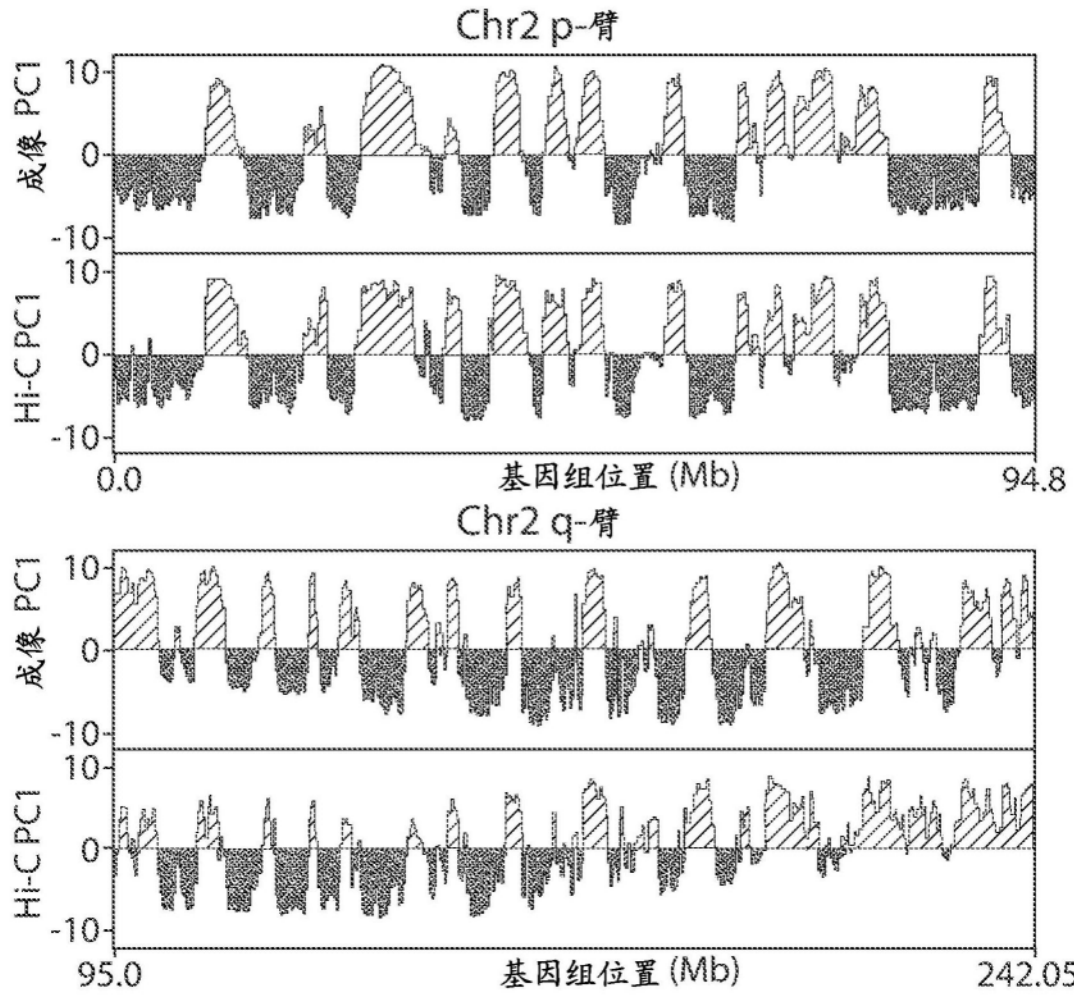


图25F

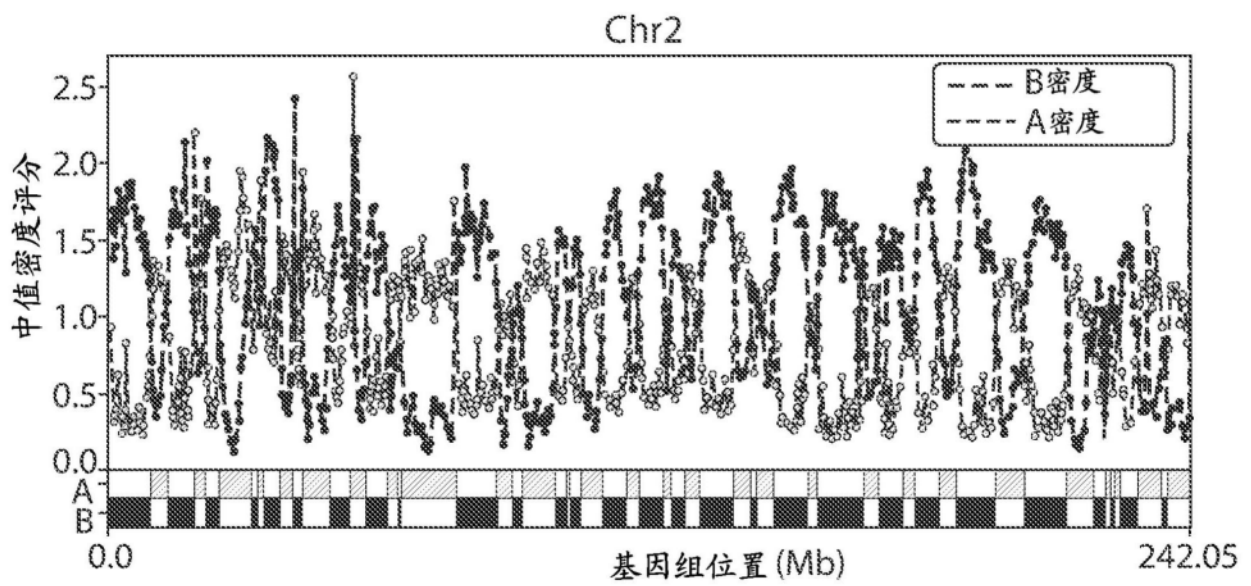


图25G

未处理的细胞中的
BRWD1 RNA 探针

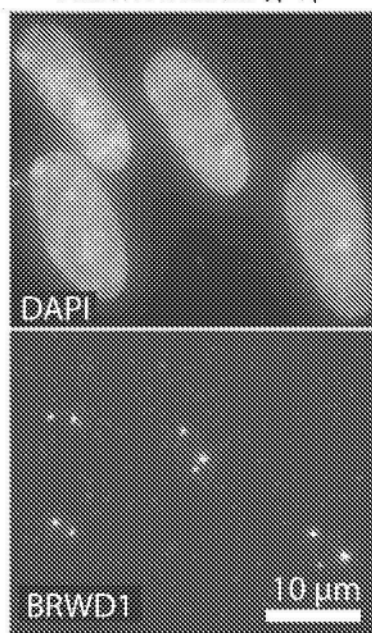


图26A

未处理的细胞中的
SCAF4 RNA 探针

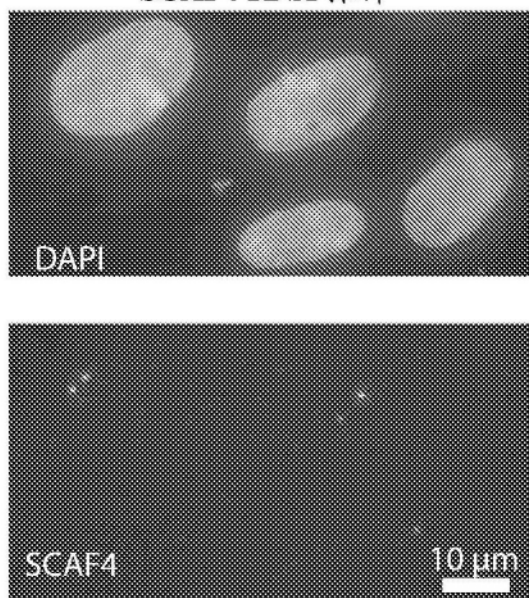


图26B

RNA 酶处理的细胞中的
BRWD1 RNA 探针

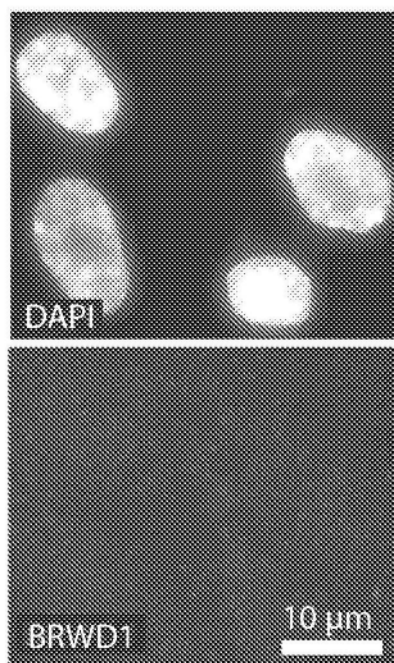


图26C

RNA 酶处理的细胞中的
SCAF4 RNA 探针

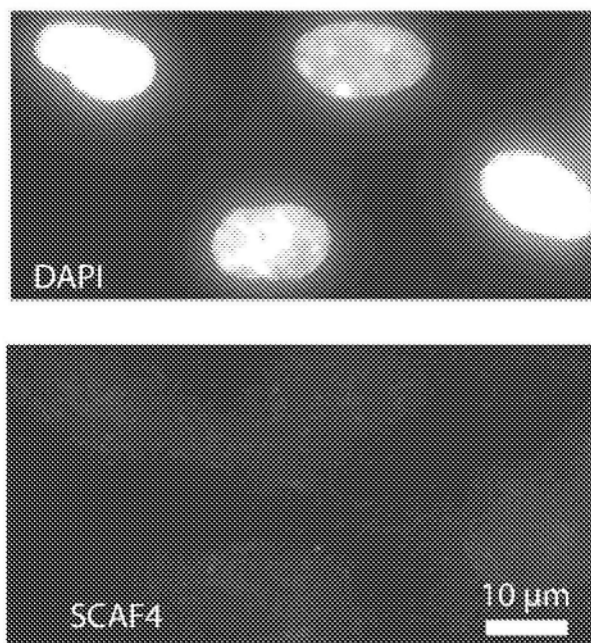


图26D

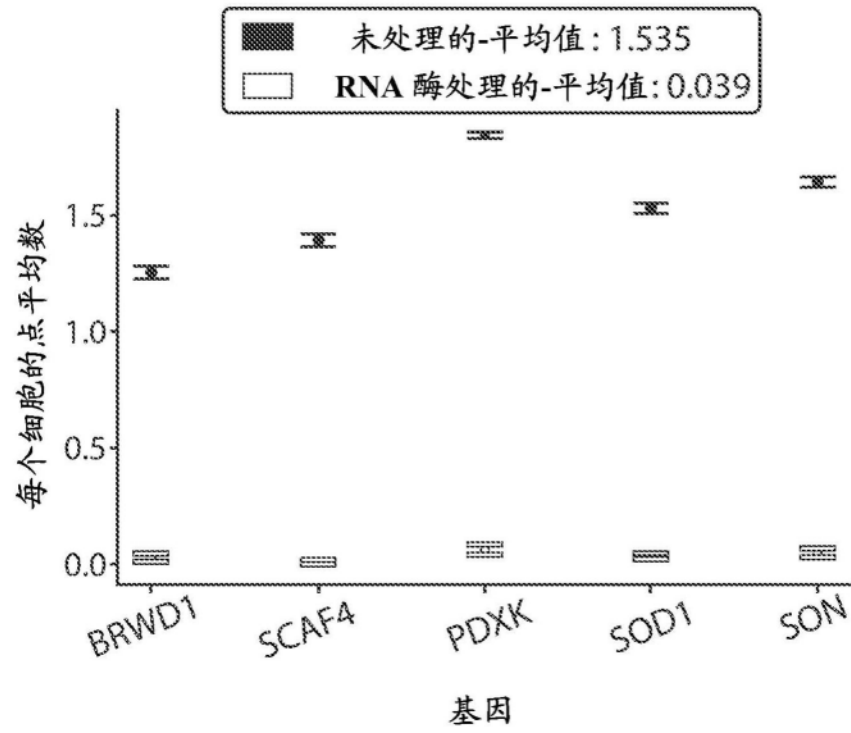


图26E

使用热变性的
基因座 1 DNA 探针

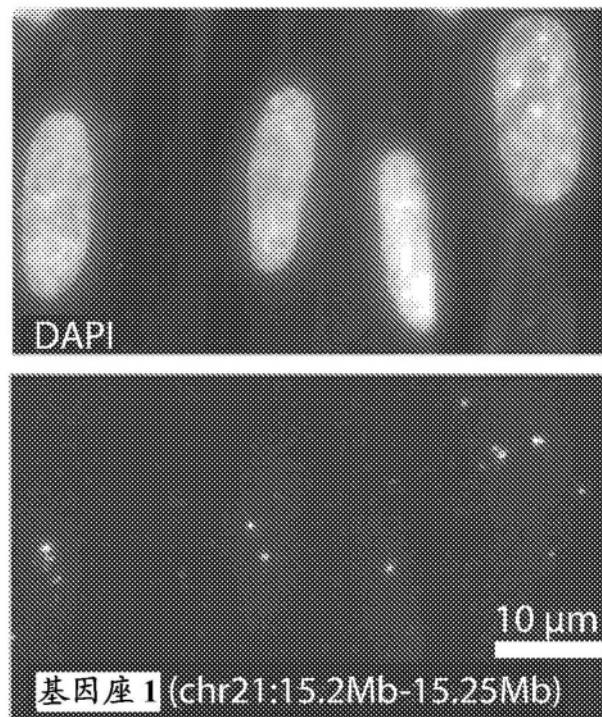


图26F

使用热变性的
基因座 2 DNA 探针

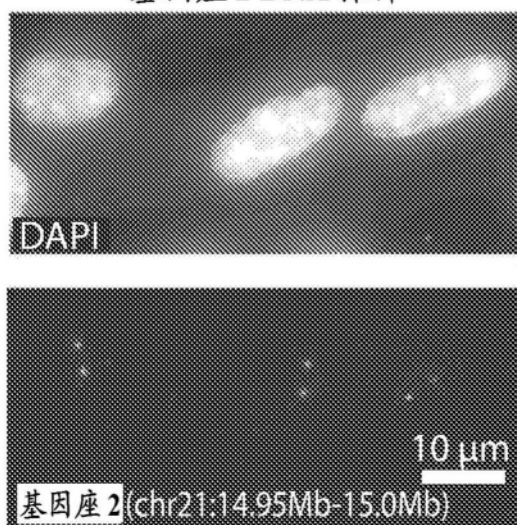


图26G

没有热变性的
基因座 1 DNA 探针

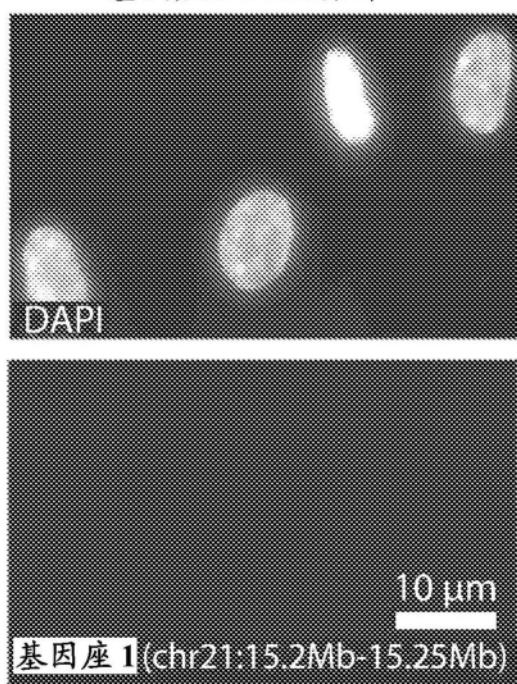


图26H

没有热变性的
基因座 2 DNA 探针

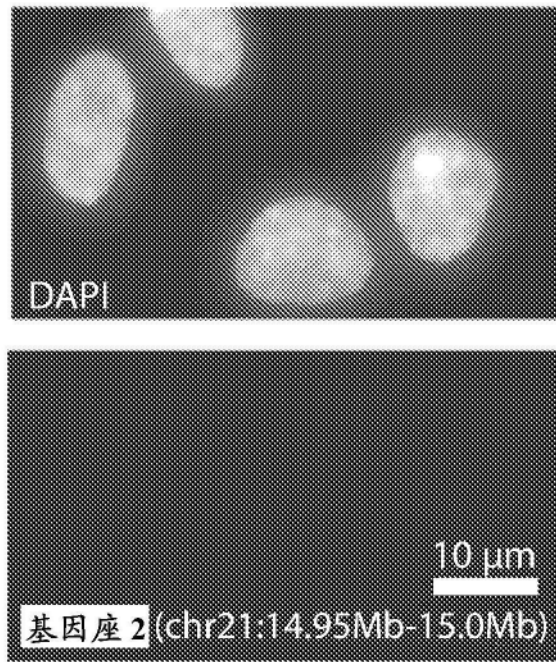


图26I

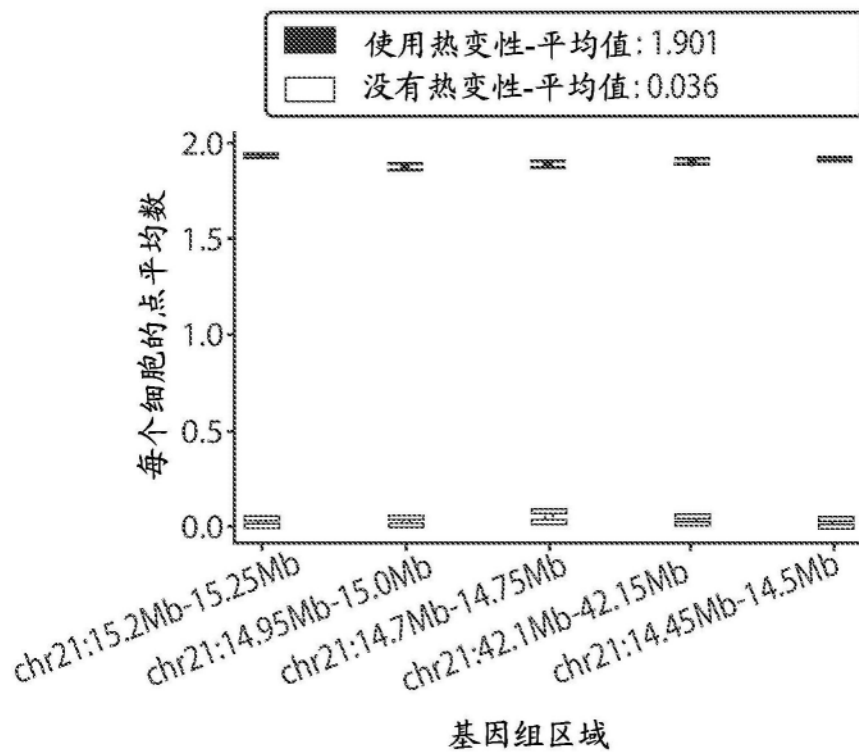


图26J

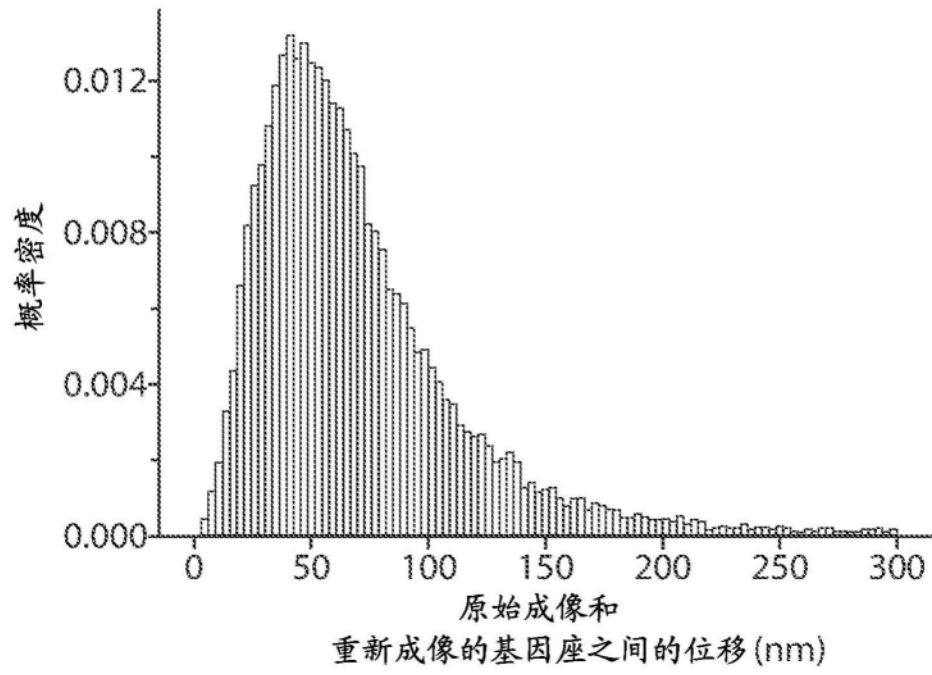


图27A

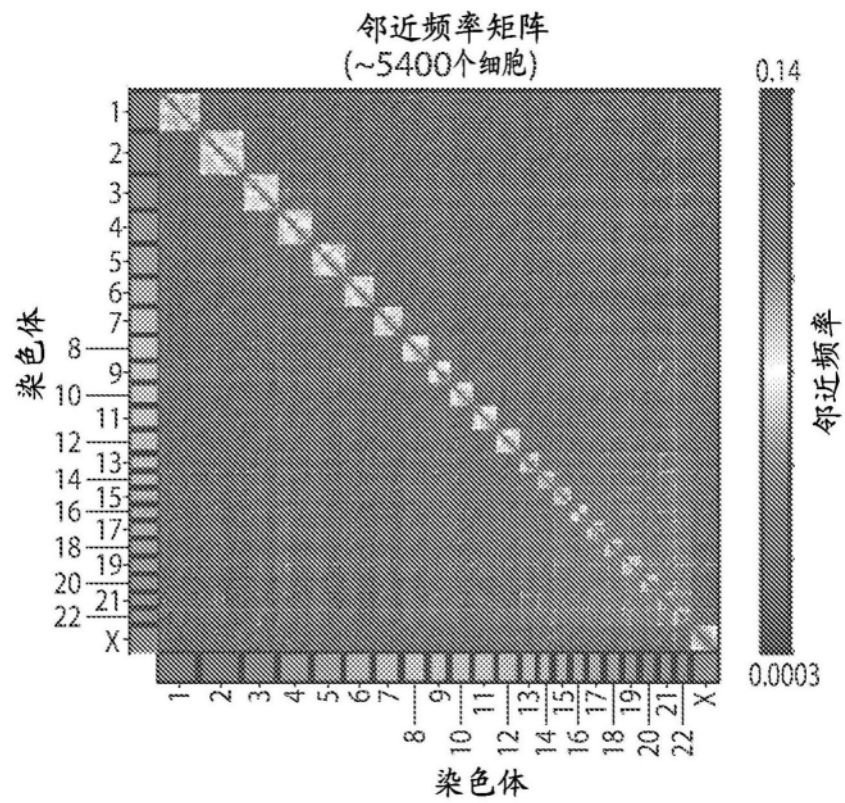


图27B

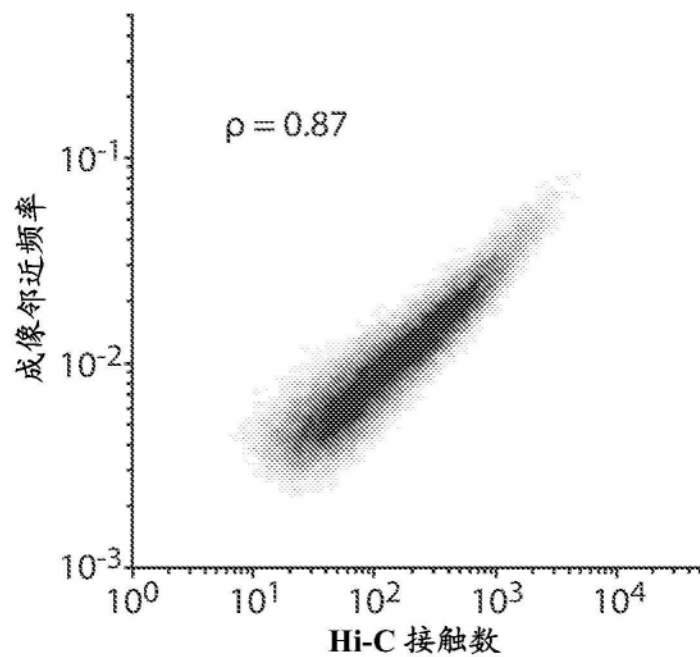


图27C

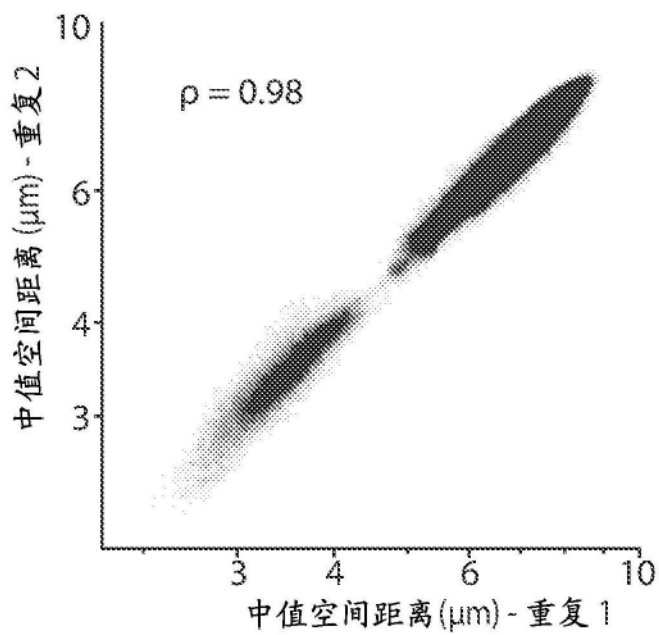


图27D

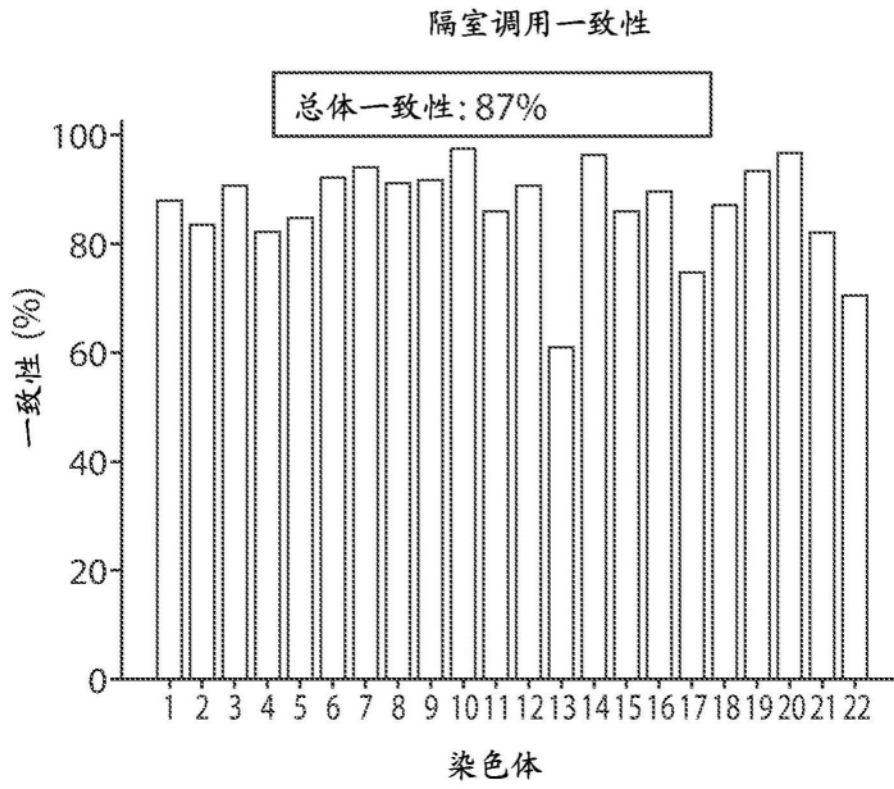


图27E

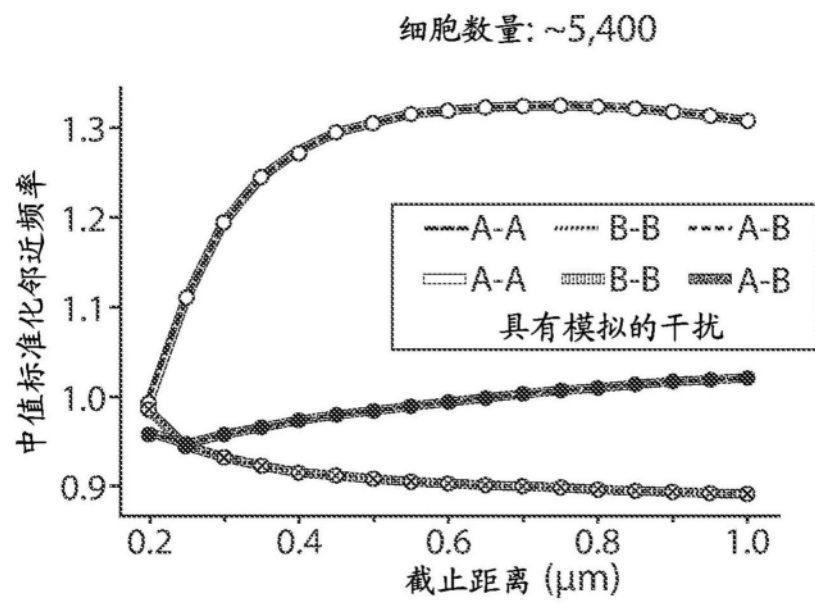


图27F

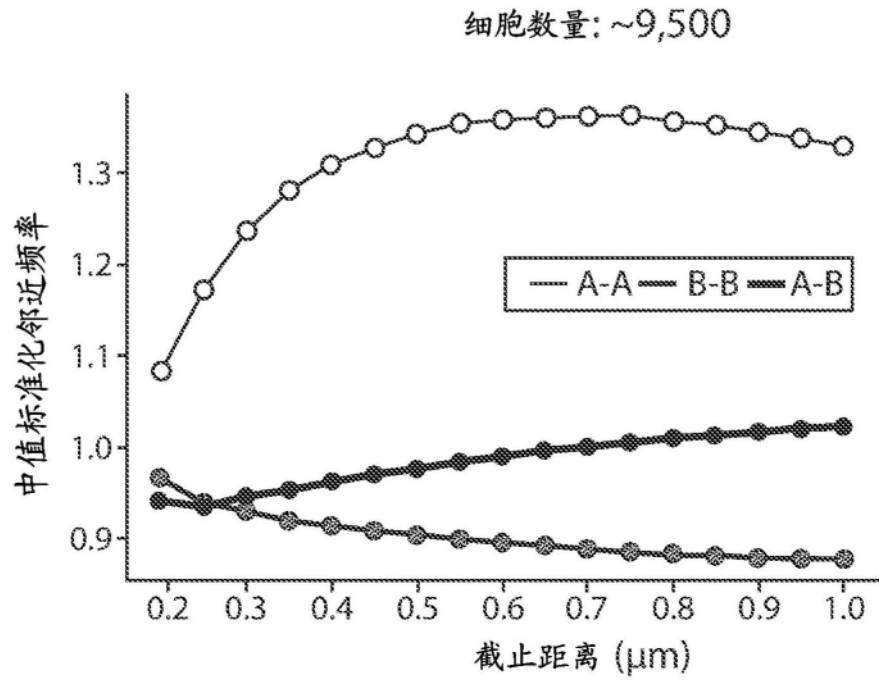


图27G

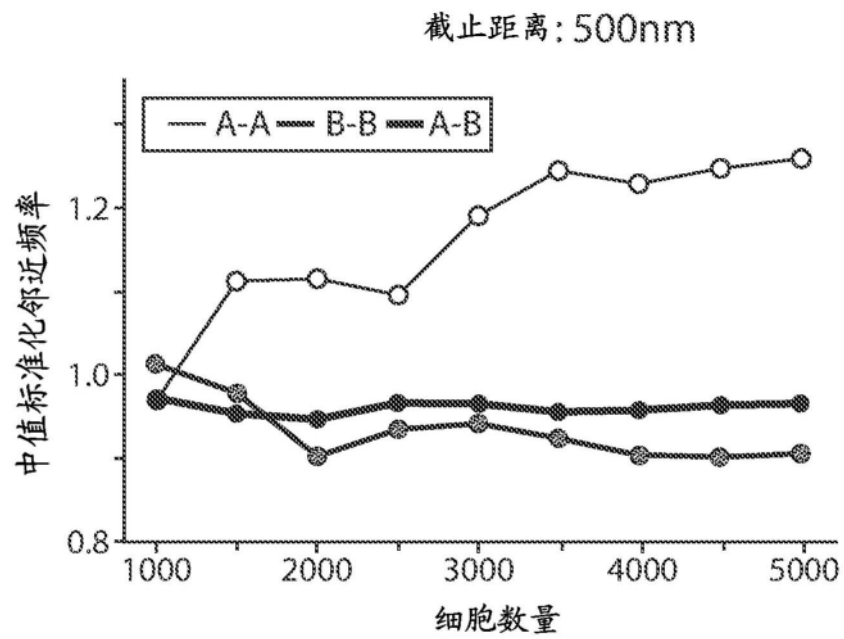


图27H

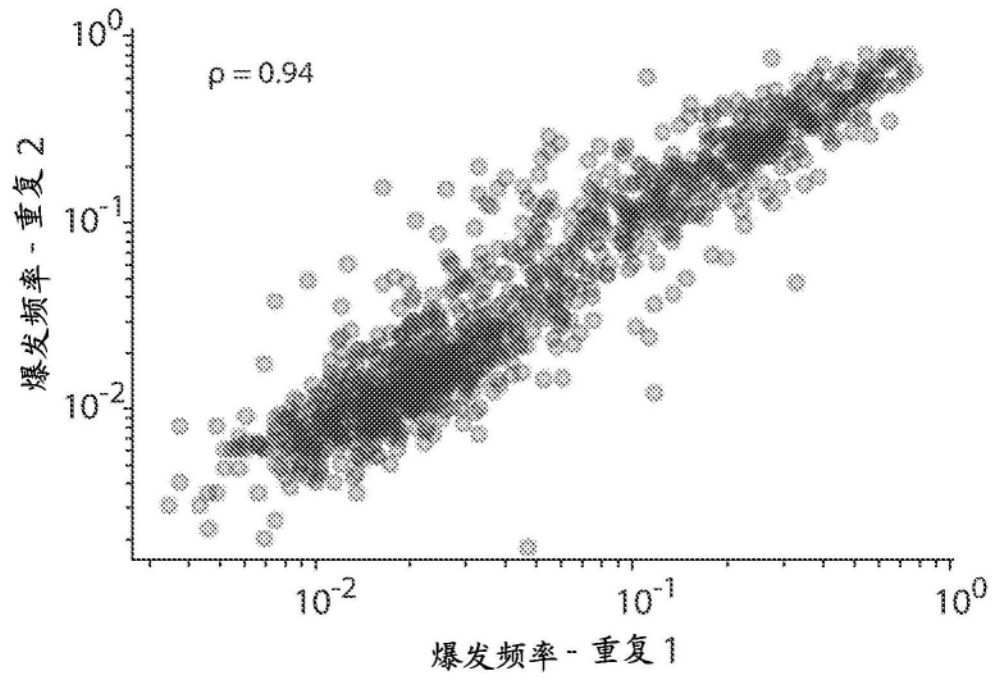


图27I

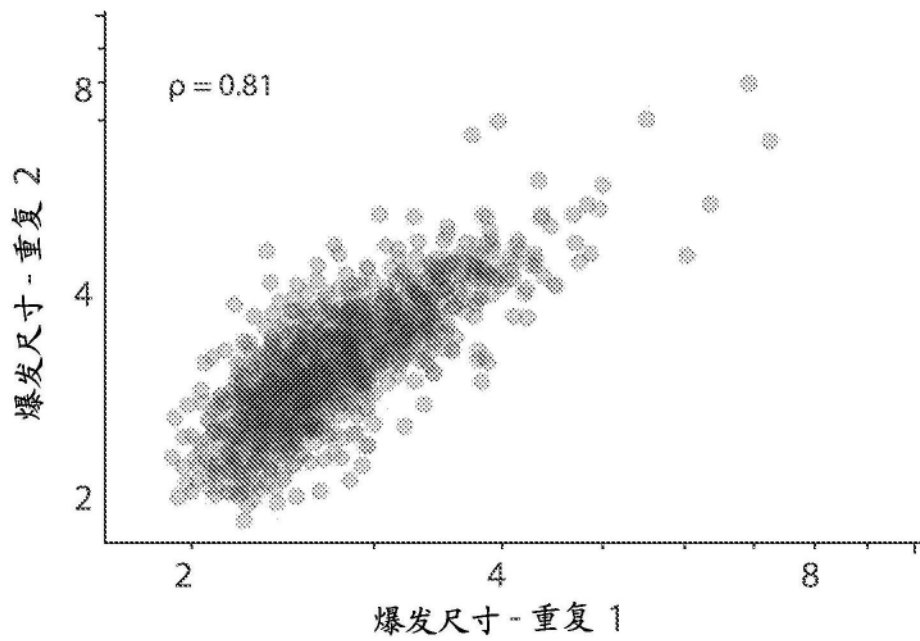
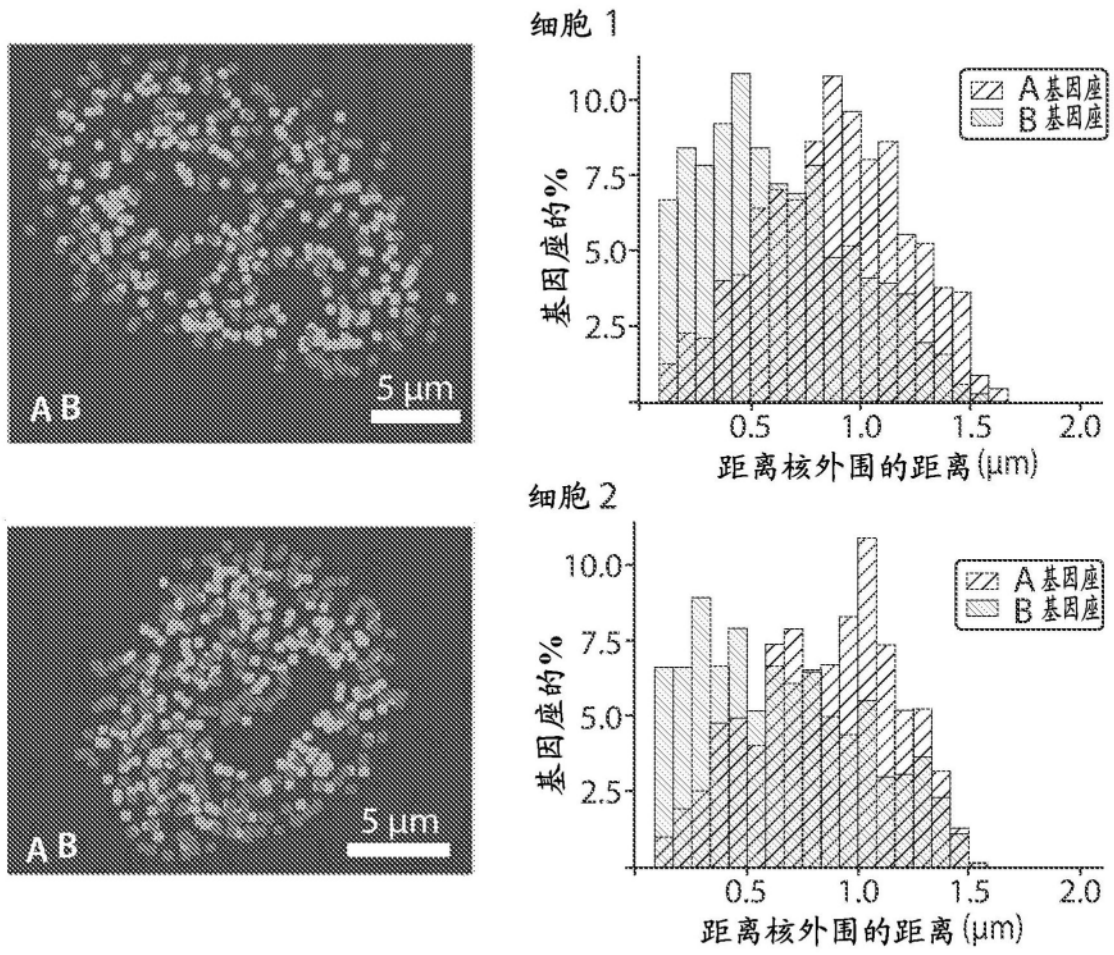


图27J



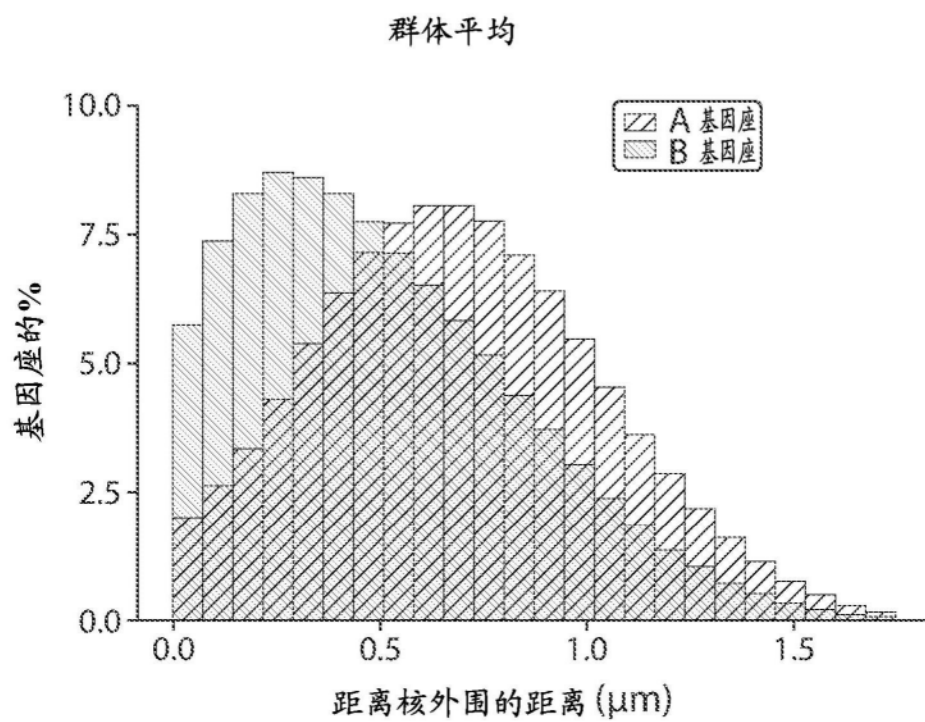


图28B

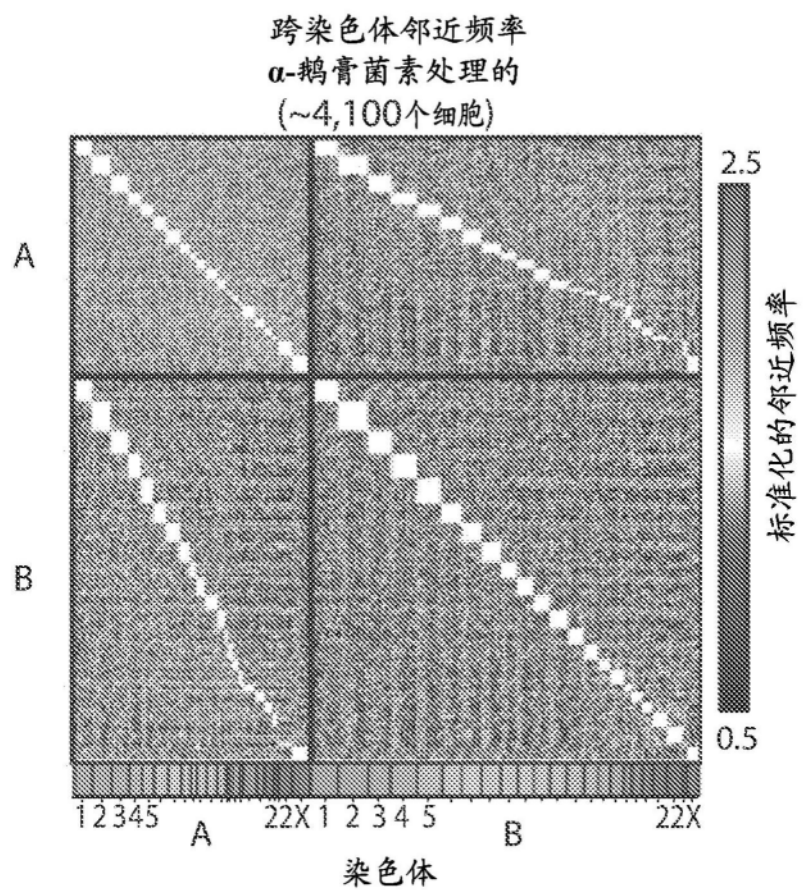


图29A

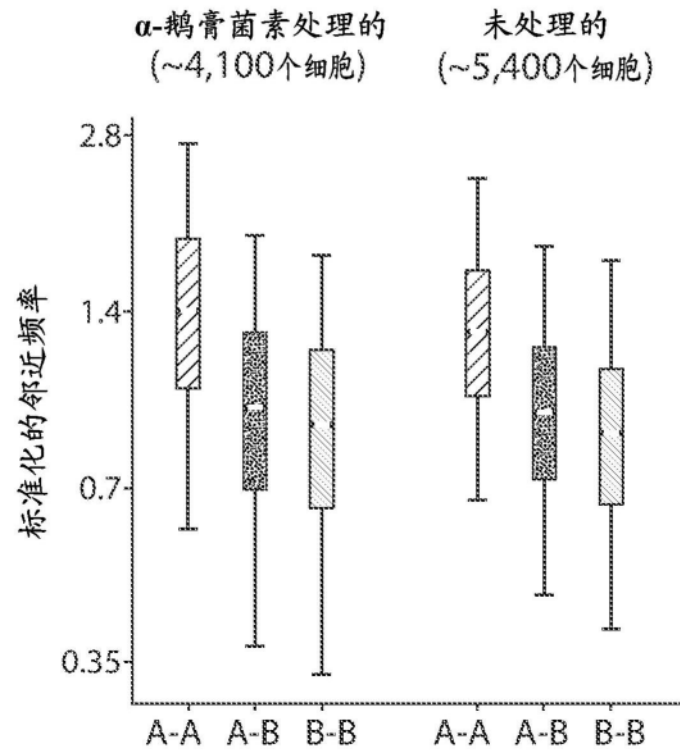


图29B

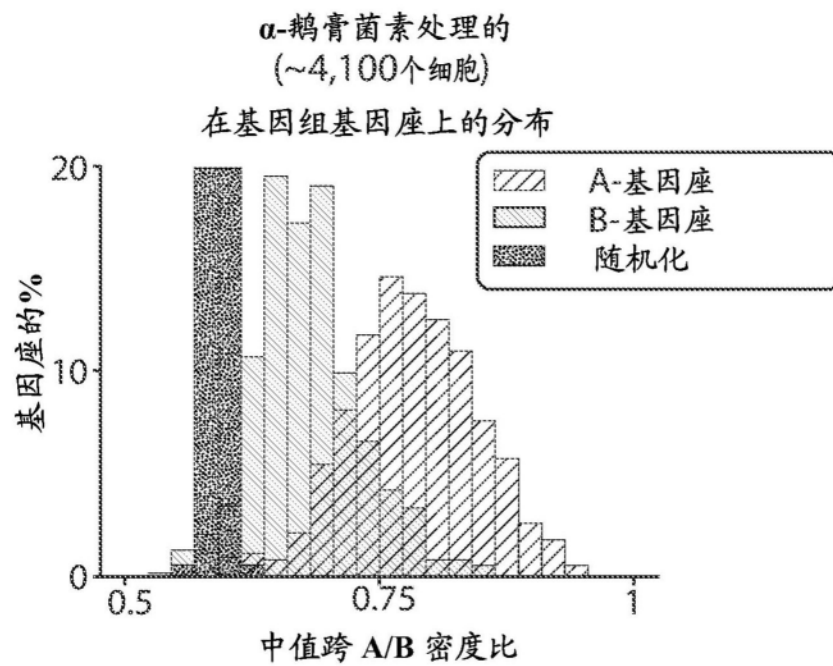


图29C

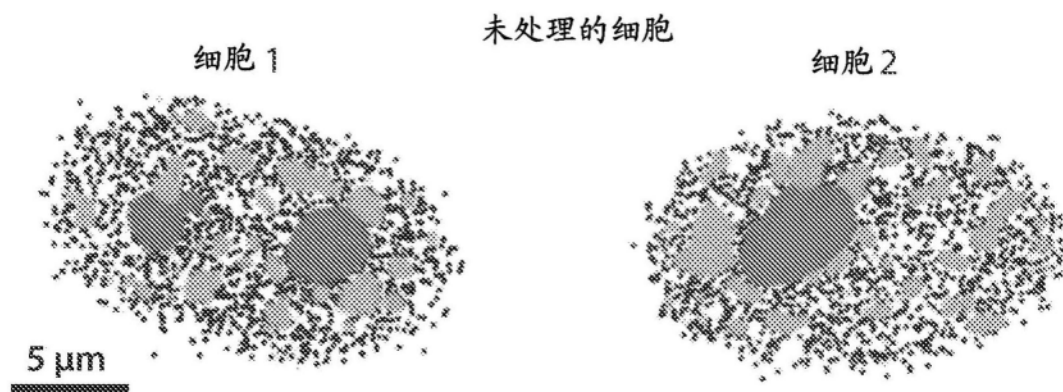


图29D

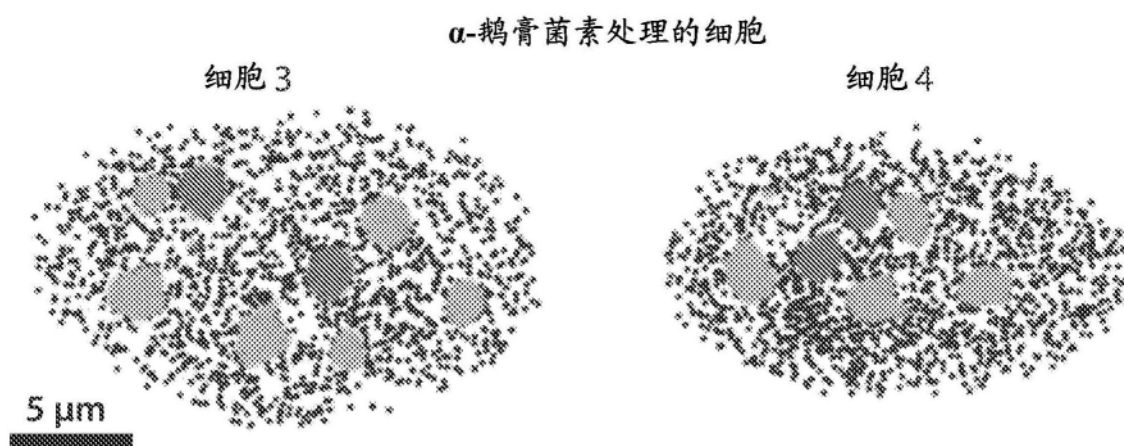


图29E

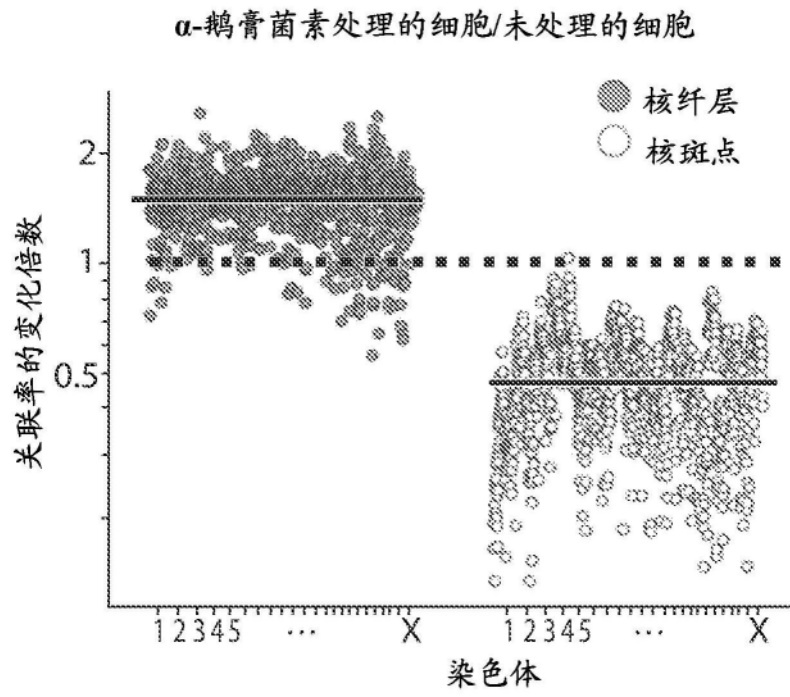


图29F

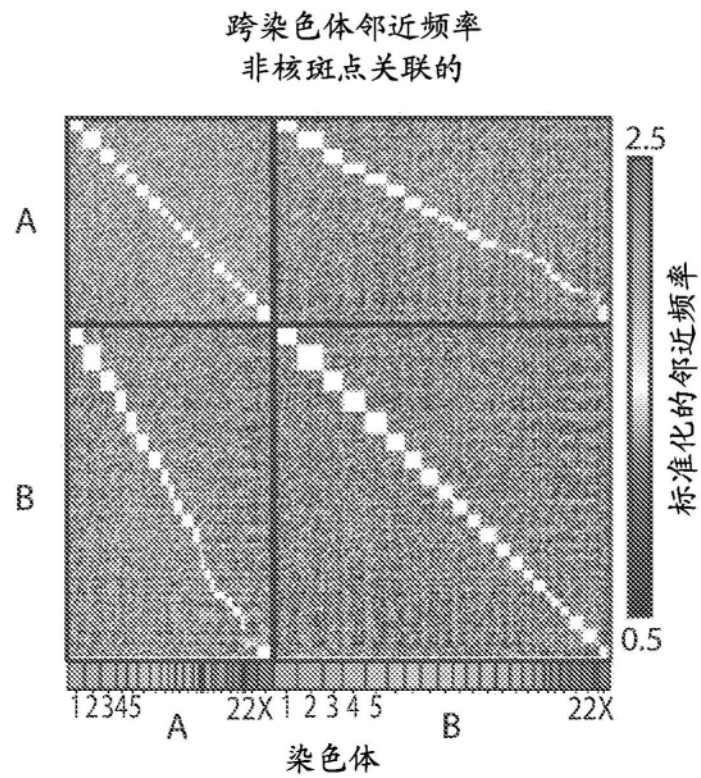


图30A

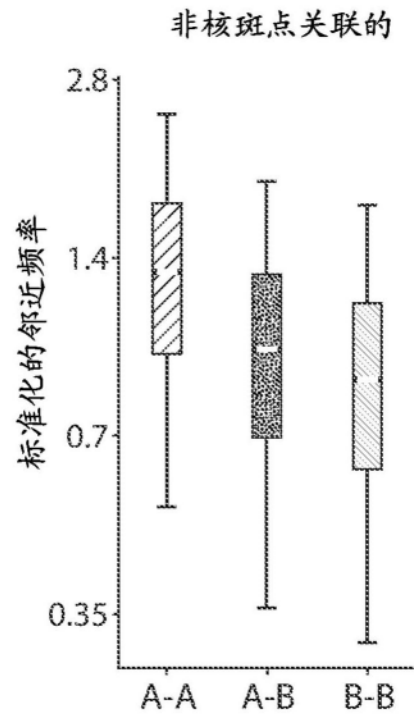


图30B

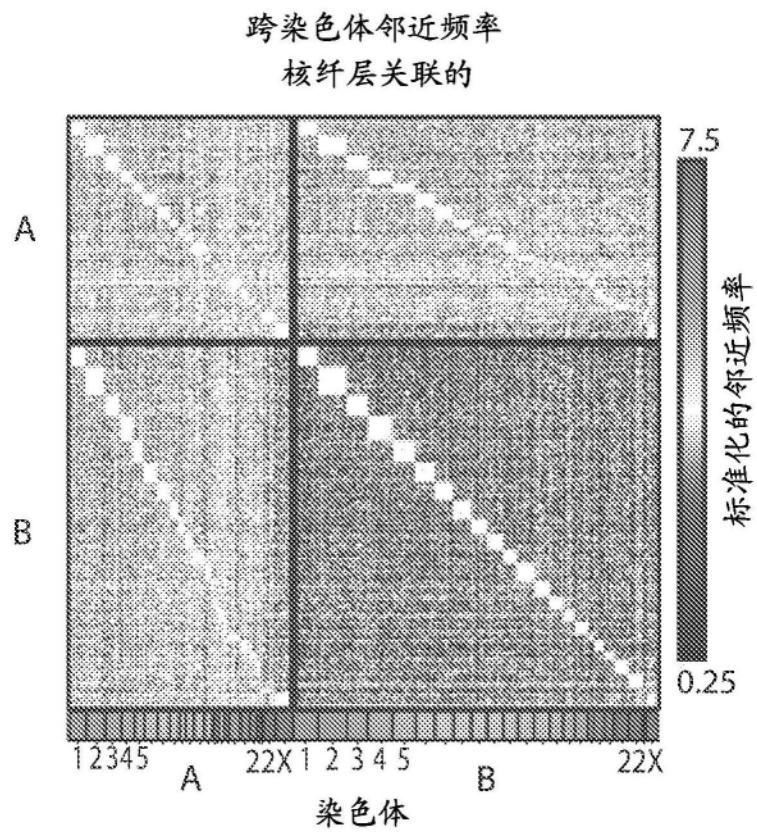


图30C

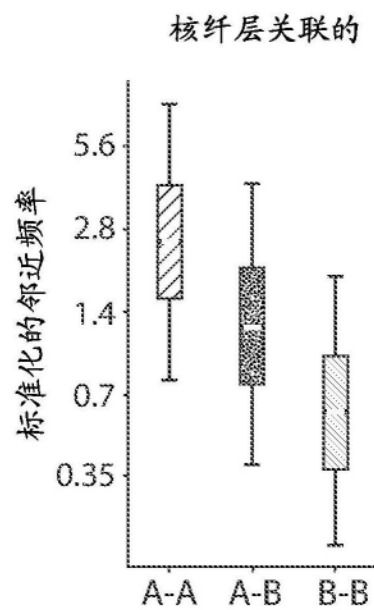


图30D