

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 12 月 17 日 (17.12.2020)



(10) 国际公布号
WO 2020/249111 A1

(51) 国际专利分类号:

C12Q 1/68 (2018.01)

(21) 国际申请号:

PCT/CN2020/095927

(22) 国际申请日:

2020 年 6 月 12 日 (12.06.2020)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201910516755.X 2019年6月14日 (14.06.2019) CN

(71) 申请人: 山东大学 (SHANDONG UNIVERSITY)

[CN/CN]; 中国山东省青岛市即墨滨海路 72 号, Shandong 266237 (CN)。

(72) 发明人: 黄启来 (HUANG, Qilai); 中国山东省青岛市即墨滨海路 72 号, Shandong 266237 (CN)。

李博 (LI, Bo); 中国山东省青岛市即墨滨海路 72 号, Shandong 266237 (CN)。

(74) 代理人: 济南圣达知识产权代理有限公司 (JINAN SHENGDA INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国山东省济南市经十路 17703 号华特广场 B308 室, Shandong 250061 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家

保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

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

(84) 指定国 (除另有指明, 要求每一种可提供的地区

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

本国际公布:

- 包括国际检索报告 (条约第 21 条 (3))。
- 包括说明书序列表部分 (细则 5.2 (a))。

(54) Title: METHOD AND KIT FOR DETECTING GENOME EDITING AND APPLICATION THEREOF

(54) 发明名称: 一种基因组编辑检测方法、试剂盒及应用

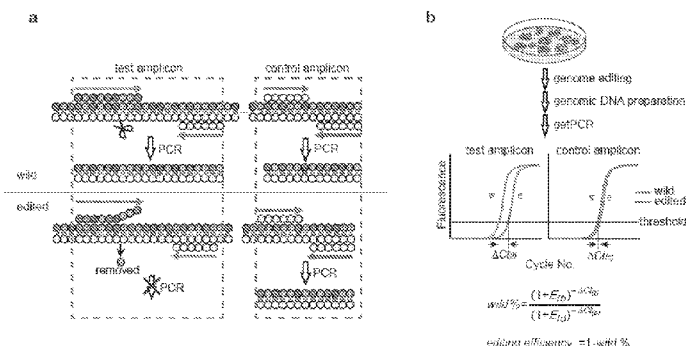


图 1

(57) Abstract: The present invention provides a getPCR method for determining genome editing efficiency, comprising quantifying wild-type DNA in a genome to be tested and calculating the percentage of the wild-type DNA to determine the genome editing efficiency.

(57) 摘要: 本发明提供了一种确认基因组编辑效率的 getPCR 方法, 其包括对待测基因组中的野生型 DNA 进行定量, 通过计算野生型 DNA 的百分比来确认基因组编辑效率。

WO 2020/249111 A1

一种基因组编辑检测方法、试剂盒及应用

技术领域

本公开属于基因编辑检测领域，具体涉及一种通过扩增定量基因组中野生型 DNA 比例间接确认基因组编辑概率的方法，及其在基因组编辑效率评价及单克隆筛选方面的应用。

背景技术

公开该背景技术部分的信息仅仅旨在增加对本公开的总体背景的理解，而不必然被视为承认或以任何形式暗示该信息构成已经成为本领域一般技术人员所公知的现有技术。

CRISPR / cas9 是一种目前主流应用的基因组编辑技术，其基因修饰效果与向导 RNA (sgRNA) 相关。在 CRISPR / cas9 系统中，Cas9 核酸酶通过 sgRNA 引导至含有原间隔序列相邻基序 (PAM) 的靶 DNA，然后在 PAM 序列上游 3bp 处切割靶 DNA 的两条链并产生双链断裂 (DSB)。一旦细胞感知到 DSB 的存在，就会通过两种不同的内在机制对断裂的基因组 DNA 进行修复，即同源重组 (HR) 或非同源末端连接 (NHEJ)。NHEJ 涉及断裂末端的直接连接，不需要同源模板并且以易错的方式修复 DNA 断裂，通常导致 DNA 断裂处不可预测的碱基插入或缺失，称之为 indel，可应用于基因的敲除，已广泛用于基因功能研究和在临床中去除致病基因。

在 CRISPR-Cas9 介导的基因组编辑中，预筛选出优秀的 sgRNA 对于获得良好的编辑效率和特异性具有重要意义，高效稳定的 sgRNA 可用于获得具有预期改变的单细胞克隆或子代。目前广泛使用的方法主要基于 DNA 测序或错配特异性核酸酶。对于 Sanger 测序方法，在分别读取每个 DNA 序列之前，需要涉及到靶区域 DNA 的 PCR 扩增和克隆。这种需要多个步骤的方法可以提供核酸酶诱导的每个突变事件的详细信息，但是非常耗时，昂贵且费力。第二代 DNA 测序 (NGS) 技术也用于分析由 sgRNA 引导 Cas9 核酸酶介导的 DNA 突变，因为它具有强大的并行分析能力。目前已经开发出了多种在线平台来分析 NGS 数据，包括 CRISPR-GA, BATCH-GE, CRISPResso, Cas-analyzer and CRISPRMatch 等。然而，发明人认为上述在线分析平台仍然需要多步骤的实验操作，时间及经济成本较高。基于错配特异性核酸酶的方法是目前最流行的方法，利用 T7 核酸内切酶 1 (T7E1) 或 Surveyor 核酸酶来切割含有序列差异的 DNA 链之间形成的含错配碱基的双链 DNA，而这种两条 DNA 链之间的差异就是由核酸酶切割引起的，从而可实现编辑效率的检测。这种方法的优点是只需要基本的实验室设备，但不适用于单核苷酸多态性区域的检测，并且往往会遗漏单核苷酸突变以及大片段的缺失的情况。此外，科学家还开发了许多其他的替代方案，但仅在某些方面有所改进，诸如 qEva-CRISPR21, 工程化核酸酶诱导的易位 (ENIT)、基于 Cas9 核酸酶的限制性片段长度多态性 (RFLP) 分析、通过扩增子分析进行 Indel 检测 (IDAA) 和基因编辑频率数字 PCR (GEF-dPCR)。发明人认为上述实验步骤较为繁琐，并且它们使用基因组靶 DNA 区域的 PCR 扩增产物而不是直接用基因组 DNA 本身来定量编辑效率。众所周知的，PCR 扩增过程中引入的序列和长度依赖性偏差将不可避免地影响检测的准确度。

发明内容

针对上述研究背景，发明人认为提供一种快速、简单并可靠的应用于基因组编辑效率定量和高通量基因分型、而且无需特定装置的方法具有重要意义。本公开提供了一种对基因组编辑效率进行检测的方法，以下文中称为 getPCR。所述 getPCR 利用 Taq 聚合酶选择性扩增特点对待测基因组 DNA 中野生型 DNA 进行扩增，通过定量扩增产物中野生型 DNA，确定野生型 DNA 比例，进而判断待测基因组中 indel 的发生频率，检测结果准确性更高，并且具有广泛的应用潜力。该方法应用于 Cas9 核酸内切酶诱导的 indel 检测具有良好的准确度，可适用于 Cas9 核酸酶技术相关的基因组编辑效率的检测，如用于 CRISPR/cas9 中 sgRNA 性能的评价、HDR 修复效率、碱基编辑器的评价；除此之外，还可以用于单细胞克隆基因型的

确认和筛选。

为了实现上述技术效果，本公开提供以下技术方案：

本公开第一方面，提供一种用于检测核酸酶切诱导的 indel 发生频率的方法，所述方法包括以下步骤：向待测基因组样品中加入引物及 Taq DNA 聚合酶，对基因组样品中的野生型 DNA 进行扩增，通过 PCR 定量野生型 DNA 比例，从而确认基因组中 indel 发生的频率；所述引物序列与野生型 DNA 序列相匹配，并且覆盖核酸酶切割位点。

优选的，所述核酸酶包括但不限于 Cas9 核酸酶、锌指核酸酶（ZFNs），转录激活因子样效应核酸酶（TALENs）和 CRISPR RNA 指导 FokI 核酸酶（RFNs），以及成对的 cas9 切口酶。进一步的，所述核酸酶为 Cas9 核酸酶。

锌指核酸酶（ZFNs）、转录激活因子样效应物核酸酶（TALENs）及 CRISPER-Cas9 系统为现代基因工程技术的常用手段，提供可靠、简便的方法对于上述基因修饰技术的效率进行评价，具有重要的意义。本领域内通常采用定量 indel 发生的频率来评估 CRISPR sgRNA 的效率，实时 PCR 技术是核酸定量中最有效的方法。然而，indel 发生的多样性和不可预测性使得不可能设计 indel 特异性引物，因此技术人员无法通过实时 PCR 直接量化 indel 频率。第一方面所述的检测方法，即 getPCR 技术，对基因组中的野生型 DNA 进行选择性的扩增，通过实时 PCR 的相对定量策略量化野生型 DNA 比例来绕过这一障碍。Taq 聚合酶能够特异性的扩增与引物完全匹配的模板，而不去扩增与引物发生错配的模板，并且 Taq 聚合酶对于引物与互补序列之间的碱基错配容忍程度较低。本公开方法利用 Taq 聚合酶的选择性扩增，可以准确的定量野生型 DNA，进而得到 indel 的发生概率。本公开以 Cas9 核酸酶为例，对于具有定向切割功能的核酸酶切位点进行引物设计并对引物参数进行优化，实现了良好的检测效果。证明本公开的研究思路及技术方案作为多种基因编辑技术的检测方法具有可行性，且有望具有良好的效果。

优选的，所述 PCR 定量为实时 PCR 或 ddPCR。

进一步优选的，所述扩增反应为实时 PCR，扩增反应退火温度为 $T_m \sim T_m + 4^\circ\text{C}$ 。

优选的，所述检测方法还包括以下步骤：在距离切割位点数百碱基对的位置处引入对照扩增，通过 $\Delta\Delta\text{Ct}$ 策略计算经过编辑后的基因组 DNA 样品中野生型 DNA 的百分比。

优选的，所述引物 3' 末端跨越 Cas9 核酸酶切割位点。

优选的，所述引物序列中包括值守碱基序列，所述值守碱基为核酸酶切位点与引物 3' 末端之间的序列，所述值守碱基长度为 1~8bp。

进一步优选的，所述引物为一条核苷酸序列，所述值守碱基的长度为 3~5bp。

进一步优选的，所述引物为正向和反向的一对序列组合，所述值守碱基的长度为 4bp。

进一步优选的，所述值守碱基 3' 末端碱基为腺嘌呤碱基或胞嘧啶或鸟嘌呤碱基；更优选的，为腺嘌呤碱基。

本公开第二方面，提供一种用于检测核酸酶切诱导的 indel 发生频率的试剂盒，所述试剂盒中包括引物、Taq DNA 聚合酶及 PCR 检测试剂；所述试剂盒的使用执行如第一方面所述检测方法。

本公开第三方面，提供第二方面所述试剂盒在评价基因组编辑效率、单细胞克隆筛选方面的应用。

优选的，所述基因组编辑包括 NHEJ 介导的 indel，HDR 介导的基因修饰及通过 BE4 产生的碱基编辑。

优选的，所述应用还包括筛选适配 CRISPR 的 gRNA。

本公开第四方面，提供一种对单细胞克隆进行基因分型的方法，所述方法包括以下步骤：以待测基因组中野生型 DNA 为模板，针对等位基因设计引物，提取待测单细胞克隆的基因组 DNA，通过第一方面所述的检测方法检测单细胞基因组 DNA 中等位基因是否发生了 indel 从而对单细胞基因实现分型。

本公开第五方面，提供一种 HDR 修复效率的检测方法，所述检测方法包括以下步骤：针对待测基因

组中 HDR 修复的基因组 DNA 设计引物, 提取待测细胞基因组 DNA, 采用第一方面所述方法检测 HDR 的发生概率; HDR 修复 DNA 所占百分比即 HDR 修复效率。

本公开第六方面, 提供一种碱基编辑器编辑效率的检测方法, 所述检测方法包括以下步骤, 以待测基因组 DNA 为模板, 针对碱基编辑后的靶序列设计引物, 采用第一方面所述的检测方法检测基因组中碱基编辑的发生概率, 即为编辑器的编辑效率。

本公开研究以 8 个 sgRNA 在 293T 细胞中的基因组编辑为例, getPCR 技术可以在所有基因组编辑案例中准确地定量基因组编辑效率, 包括 NHEJ 诱导的 indel, HDR 和碱基编辑。同时, 该方法在单细胞克隆基因型中表现出强大的能力, 因为它不仅可以表征所期望的基因组编辑是否发生, 而且还可以告知具体有几条等位基因携带了这种特定的编辑。

与现有技术相比, 本公开的有益效果是:

1. 随着 CRISPR 技术的快速发展和广泛应用, 提供一种简单、准确可靠的基因组编辑效率评价方法, 对于 gRNA 的筛选及实验方案的优化具有重要的意义。本公开提供的方法过程简单、定量结果可靠、节省时间和成本低、并且不涉及特定装置, 仅需要一个 qPCR 步骤。针对 CRISPR 靶标上精准测定 indel 频率, 检测的准确度与公认最准的 NGS 方法一致。
2. 基于 Cas 核酸酶切技术的基因编辑方法均可以使用本公开方法, 包括 NHEJ 诱导的 indel, HDR 和碱基编辑, 同时还可以应用于单细胞克隆的筛选。
3. 本公开提供的 getPCR 也可以很容易地扩展应用于由其它类型基因组切割核酸酶介导的基因组编辑实验, 去评估给定切割位置的编辑效率, 如锌指核酸酶 (ZFNs), 转录激活因子样效应核酸酶 (TALENs) 和 CRISPR RNA 指导 FokI 核酸酶 (RFNs), 以及成对的 cas9 切口酶等, 通过进一步的确定值守碱基的设计规则, 有望进一步推动该技术在基因组编辑技术在分子和细胞生物学研究中的广泛应用。

附图说明

构成本公开的一部分的说明书附图用来提供对本公开的进一步理解, 本公开的示意性实施例及其说明用于解释本公开, 并不构成对本公开的不当限定。

图 1 本公开 getPCR 原理及流程图;

(a) getPCR 鉴别 indel 和野生序列的原理 (b) getPCR 策略概述

图 2 实施例 1 中 getPCR 引物设计原理图;

(a) 26 个质粒在 HOXB13 基因 gRNA 靶标 4 处模拟 indel;

(b) 16 种具有不同值守碱基的 getPCR 值守碱基; 评估分别使用反向引物 (c) 和正向引物 (d) 以及正向与反向引物组合使用;

(e) 时能够区分 indels 与野生型序列的能力;

(f) 正反向引物组合使用时自扩增背景信号的研究;

(g) 引物 3' 末端的第一个碱基对扩增特异性的影响;

(h) 不同类型的碱基错配对扩增效率的影响;

(i) 3' 末端碱基类型在确定 getPCR 对错配的敏感性中的作用。平均值 $\pm$ s. e. m, n=3 个独立技术重复)

图 3 实施例 2 中运行 getPCR 的参数优化图;

(a-d) 在不同退火温度下, 使用四种值守碱基对 indel/野生型序列 DNA 模板进行扩增的扩增曲线。正向值守碱基包含 3 (a) 或 4 (b) 个正向值守碱基, 或 3 (c) 或 4 (d) 个反向的值守碱基。

(e-h) PCR 过程中不同的退火温度下, 不同 T_m 值的值守碱基的扩增效率以及选择性扩增能力的展示, 其中使用到了具有三个 (e) 或四个 (g) 观察值守的值守碱基和具有三个 (f) 或四个 (h) 值守碱基的反向值守碱基。PCR 效率表征为相对于 65°C 下的 C_t 值计算的 ΔC_t , 选择性表征为使用野生型模板和插入

缺失模板之间的 ΔC_t 。值守碱基序列显示在底部。小圆圈指在下降了 0.5 个循环时（如虚线所示）最佳扩增效率下的最佳选择性。

(i-1) 退火温度对 PCR 扩增效率的影响以及标准曲线的线性，以 R 平方值为特征。在检测中使用的四个值守碱基分别具有三个 (i) 或四个 (k) 正向值守碱基，或分别具有三个 (j) 或四个 (l) 反向值守碱基。（平均值 $\pm$ s. e. m, n=3 个独立技术重复）

图 4 实施例 4 中 getPCR 在模拟单细胞克隆的基因分型应用结果图；(a) Surveyor 检测电泳色谱图，检测的样品中含有给定百分比的插入缺失，用来模拟基因组编辑后的 DNA；

(b) Surveyor 检测得到的编辑频率结果的量化；

(c) 单独或组合使用正向和反向值守碱基运用 getPCR 方法检测 Indel 频率；

(d-f) 使用三种不同设计的值守碱基运用 getPCR 对模拟的单细胞克隆的基因分型。（平均值 $\pm$ s. e. m, n=3 个独立技术重复, *P<0.05, **P<0.01, ***P<0.001）

图 5 实施例 5 中 getPCR 确定单细胞克隆的编辑频率和基因型结果图；

在 gRNA 靶向 HOXB13, DYRK1A 和 EMX1 基因进行基因组编辑的 293T 细胞中进行 Indel 频率测定，并进行单细胞克隆的基因分型；

(a) 在对 8 种 gRNA 介导的基因组编辑组合，getPCR 对产生的 indel 频率进行定量，并与 NGS 和 Surveyor 方法相比较；

(b) getPCR 中使用的 gRNA 序列和值守碱基的图示；来自编辑的靶向 HOXB13 基因 (c, d), EMX1 基因 (e, f, i) 和 DYRK1A 基因 (g, h) 的 293T 细胞的单细胞克隆通过 getPCR 方法进行基因分型。箱形图分别展示了第一四分位数、中位数和第三四分位数，胡须表示 1.5IQR，异常值单独显示。在基因分型 (j-1) 中评估两种不同设计的值守碱基的相关性和组合效果。（平均值 $\pm$ s. e. m, n=3 个独立技术重复, *P<0.05, **P<0.01, ***P<0.001）

图 6 实施例 6 中 getPCR 应用于确定 HDR 频率和单细胞克隆的基因型结果图；

(a) getPCR 在 HDR 和碱基编辑方面的量化原理示意图；

(b) 靶向 EMX1 基因的 HDR 修复效率检测与靶向 HOXB13 基因的碱基编辑效率检测所用到的引物；

(c) 使用 getPCR 进行 HDR 效率定量并与 NGS 和 HindIII 酶切方法的相比；

(d-f) 单独或组合使用两种不同的值守碱基运用 getPCR 方法，对来自 HDR 实验的单细胞克隆的基因分型，箱形图分别展示了第一四分位数、中位数和第三四分位数，胡须表示 1.5IQR，异常值单独显示；

(g, h) 分别在靶向 EMX1 和 HOXB13 基因的碱基编辑实验中通过 getPCR 和 NGS 方法确定的每种基因型的频率，通过 getPCR 对来自 EMX1 基因碱基编辑的第 5 和第 6 位杂合的 10 个克隆进行详细的基因分型；

(i) 通过 getPCR 对来自 EMX1 基因碱基编辑的第 5 和第 6 位杂合的 10 个克隆进行详细的基因分型；

(j, k) 条形图和散点图在 EMX1 基因编辑实验中通过 getPCR 显示第 5 个核苷酸的单细胞克隆基因分型；

(l, m) 相应的第 6 个核苷酸的单细胞克隆基因分型；

(n, o) 条形图和散点图显示 HOXB13 基因的碱基编辑中的单细胞克隆基因分型。（平均值 $\pm$ s. e. m, n=3 个独立技术重复, *P<0.05, **P<0.01, ***P<0.001）

图 7 设计 getPCR 引物和运行 getPCR 的注意事项；

(a, b) 分别在正向和反向设计具有给定值守碱基但具有不同长度/ T_m 值的多个 getPCR 引物；

(c) 这些 getPCR 引物在野生型模板上的扩增效率；

(d) 条形图显示不同的值守碱基组合对模拟 indel 质粒 PCR 扩增的特异性，是图 2e 的替代展示图；

(e) 显示在不添加模板的情况下值守碱基组合的 PCR 自扩增信号的条形图，是图 2f 的替代展示；

(f, g) 相对于 3' 末端, 单碱基错配的位置对 PCR 扩增的影响, 分别展示正向和反向值守碱基的结果; (h, i) 比较 3' 末端碱基错配与 3' 末端碱基缺失对 PCR 扩增的阻碍能力, 分别展示正向和反向值守碱基; (j) 多种 qPCRSYBRGreenmix 产品在 getPCR 应用中的适用性比较。(平均值 $\pm$ s. e. m, n=3 个独立技术重复)

图 8 不同 DNA 聚合酶产物在错配识别中的表现;

(a, b) 电泳色谱显示不同 DNA 聚合酶的 PCR 扩增水平, PCR 中使用到的模板分别有不含错配碱基、含错配碱基, 引物分别为正向和反向值守碱基;

(c) 来自 a 和 b 的 PCR 产物的 Sanger 测序色谱;

(d, e) 条形图, 说明在多个 qPCR 产品的扩增中, 相对于 3' 末端, 对在不同位置的单碱基错配的敏感性, 分别使用了正向和反向值守碱基。(平均值 $\pm$ s. e. m, n=3 个独立技术重复)

图 9 使用模拟插入缺失的质粒进行编辑频率测定和单细胞克隆的基因分型;

(a-c) 使用正向和反向值守碱基组合的 getPCR 方法对模拟插入缺失的 DNA 进行频率定量;

(d-f) 通过组合两种不同设计的 getPCR 值守碱基对模拟单细胞克隆进行基因分型; 参考图 2a, 以获得模拟插入信息。(平均值 $\pm$ s. e. m, n=3 个独立技术重复)

图 10 针对 gRNA 靶向基因 HOXB13, DYRK1A 和 EMX1 进行基因组编辑以产生插入缺失突变的单细胞克隆进行基因分型;

(a, b) 分别用两种不同设计的值守碱基通过 getPCR 方法对单细胞克隆进行基因分型, 单细胞克隆来自 gRNA 靶向 DYRK1A 基因进行基因组编辑的 293T 细胞。箱形图分别展示了第一四分位数、中位数和第三四分位数, 胡须表示 1.5IQR, 异常值单独显示;

(c-g) 散点图显示两种不同设计的值守碱基在基因分型中的相关性和组合效应;

(h-l) 通过 Sanger 测序在单细胞克隆基因分型中确定的插入缺失突变, 分别针对 gRNAHOXB13 靶标 6, EMX1 靶标 5, DYRK1A 靶标 1 和 EMX1 靶标 1; (平均值 $\pm$ s. e. m, n=3 个独立技术重复, * $P<0.05$, ** $P<0.01$, *** $P<0.001$)

图 11 针对 gRNA 靶向 EMX1 基因完成碱基编辑后分离得到的单细胞克隆进行基因分型;

(a) 条形图, 显示在 EMX1 基因编辑实验中通过 getPCR 在第 5 个核苷酸处进行单细胞克隆基因分型, 即图 6j 用详细克隆编号注释;

(b) 条形图, 显示在 EMX1 基因编辑实验中通过 getPCR 在第 6 个核苷酸处进行单细胞克隆基因分型, 即图 6l 用详细克隆编号注释;

(c) 单细胞克隆基因分型的 Sanger 测序层析。(平均值 $\pm$ s. e. m, n=3 个独立技术重复)

图 12 由 HOXB13 基因上引入终止密码子的碱基编辑得到的单细胞克隆的基因分型;

(a) 对 HOXB13 基因碱基编辑实验中得到的单细胞克隆, 通过 getPCR 在第 8 个核苷酸处进行基因分型, 即图 6n 用详细克隆编号注释;

(b) 单细胞克隆基因分型的 Sanger 测序层析。(平均值 $\pm$ s. e. m, n=3 个独立技术重复)。

具体实施方式

应该指出, 以下详细说明都是例示性的, 旨在对本公开提供进一步的说明。除非另有指明, 本文使用的所有技术和科学术语具有与本公开所属技术领域的普通技术人员通常理解的含义。

需要注意的是, 这里所使用的术语仅是为了描述具体实施方式, 而非意图限制根据本公开的示例性实施方式。如在这里所使用的, 除非上下文另外明确指出, 否则单数形式也意图包括复数形式, 此外, 还应当理解的是, 当在本说明书中使用术语“包含”和/或“包括”时, 其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

正如背景技术所介绍的,现有技术中针对基因编辑方法效率进行检测的方法存在一定的缺陷:如 Sanger、NGS、基于错配特异性核酸酶的方法等具有操作复杂、成本高、检测准确度不足等。提供一种能够快速、简单并可靠的应用于基因组编辑效率定量和高通量基因分型,而且无需特定装置的方法具有重要的意义。为了实现该技术目的,本公开提供了一种getPCR检测方法,利用Taq聚合酶的特异性,以野生型DNA序列为模板,设计覆盖核酸酶切位点的引物序列,通过扩增定量基因组中野生型DNA的百分比间接确定基因组的编辑效率。经优化和验证,该方法的检测准确度高并且操作方便,具有广泛的应用价值。

为了使得本领域技术人员能够更加清楚地了解本公开的技术方案,以下将结合具体的实施例与对比例详细说明本公开的技术方案。

以下实施例中所使用的试剂和材料来源如下:

质粒与DNA片段 pcDNA3.1载体上含有HOXB13基因编码区的质粒由奥卢大学的卫功宏教授赠送。

模拟HOXB13gRNA靶标4的潜在不同插入缺失(图2a)的26个DNA变体和包含为了引入不同类型的引物-模板错配突变的其他15种变体都是通过定点诱变构建得到的。通过用PCR方法从pSpCas9(BB)载体(Addgene, # 42230)中删除cas9表达框构建得到了sgRNA表达质粒。将带有20-ntgRNA序列的退火寡核苷酸对连接到sgRNA表达质粒或pSpCas9(BB)载体的BbsI位点之间。基于pSpCas9(BB)通过定点诱变获得高保真CRISPR-Cas9核酸酶(R661A/Q695A/Q926A/D1135E)。

BE4-Gam质粒(Addgene, # 100806)用于碱基编辑实验。

含有EMX1-HindIII突变的99-nt长度的单链HDR模板是在英潍捷基公司(上海)合成得到,引入的HindIII位点序列与EMX1gRNA靶标5的PAM序列相邻。构建得到含有EMX1-HindIII突变的质粒,并作为100%的同源重组修复效率。所有使用的引物和寡核苷酸的序列均显示在表1中。

表1、用于质粒构建和转染的寡核苷酸序列

a、用于构建HOXB13变体的引物

Pladmid	Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')	Template
HB251d1L	AGGCGGGTACTACTCCTGC	CAAAGTAACCATAAGGCACGG	OE1
HB251d2L	AGGCGGGTACTACTCCTGC	AAAGTAACCATAAGGCACGGG	OE1
HB251d3L	AGGCGGGTACTACTCCTGC	AAGTAACCATAAGGCACGGGA	OE1
HB251d5L	AGGCGGGTACTACTCCTGC	GTAACCATAAGGCACGGGAGC	OE1
HB251d10L	AGGCGGGTACTACTCCTGC	CATAAGGCACGGGAGCTGG	OE1
HB251d15L	AGGCGGGTACTACTCCTGC	GGCACGGGAGCTGGGGACG	OE1
HB251d1R	GGCGGGTACTACTCCTGCC	CCAAAGTAACCATAAGGCACG	OE1
HB251d2R	GCGGGTACTACTCCTGCCG	CCAAAGTAACCATAAGGCACG	OE1
HB251d3R	CGGGTACTACTCCTGCCGA	CCAAAGTAACCATAAGGCACG	OE1
HB251d5R	GGTACTACTCCTGCCGAGTGT	CCAAAGTAACCATAAGGCACG	OE1
HB251d10R	TACTCCTGCCGAGTGTCCC	CCAAAGTAACCATAAGGCACG	OE1
HB251d15R	CTGCCGAGTGTCCCGAGC	CCAAAGTAACCATAAGGCACG	OE1
HB251d1L1R	GGCGGGTACTACTCCTGCC	CAAAGTAACCATAAGGCACGG	OE1
HB251d2L2R	GCGGGTACTACTCCTGCCG	AAAGTAACCATAAGGCACGGG	OE1
HB251d2L3R	CGGGTACTACTCCTGCCGA	AAAGTAACCATAAGGCACGGG	OE1
HB251d2L5R	GGTACTACTCCTGCCGAGTGT	AAAGTAACCATAAGGCACGGG	OE1
HB251d2L10R	TACTCCTGCCGAGTGTCCC	AAAGTAACCATAAGGCACGGG	OE1
HB251iA	AAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1

HB251iT	TAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
HB251iG	GAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
HB251iC	CAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
HB251iAA	AAAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
HB251iAT	ATAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
HB251iAG	AGAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
HB251iAC	ACAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
HB251iAAA	AAAAGGCGGGTACTACTCCTGC	CCAAAGTAACCATAAGGCACG	OE1
1MM-G	GGCGGGTACTACTCCTGCCG	CCCAAAGTAACCATAAGGCACG	OE1
2MM-G	GGCGGGTACTACTCCTGCCG	CTCAAAGTAACCATAAGGCACG	OE1
3MM-G	GGCGGGTACTACTCCTGCCG	CTTAAAGTAACCATAAGGCACG	OE1
1MM-C	CGCGGGTACTACTCCTGCCG	CCCAAAGTAACCATAAGGCACG	OE1
2MM-C	CGCGGGTACTACTCCTGCCG	CTCAAAGTAACCATAAGGCACG	OE1
3MM-C	CGCGGGTACTACTCCTGCCG	CTTAAAGTAACCATAAGGCACG	OE1
1MM-A	TGCGGGTACTACTCCTGCCG	CCCAAAGTAACCATAAGGCACG	OE1
2MM-A	TGCGGGTACTACTCCTGCCG	CTCAAAGTAACCATAAGGCACG	OE1
3MM-A	TGCGGGTACTACTCCTGCCG	CTTAAAGTAACCATAAGGCACG	OE1
1MM-T	AGCGGGTACTACTCCTGCCG	CCCAAAGTAACCATAAGGCACG	OE1
2MM-T	AGCGGGTACTACTCCTGCCG	CTCAAAGTAACCATAAGGCACG	OE1
3MM-T	AGCGGGTACTACTCCTGCCG	CTTAAAGTAACCATAAGGCACG	OE1
A252T	cttatggttacttttgTggcggtactactcc	GGAGTAGTACCCGCCACCAAAG TAACCATAAG	OE1
G253C	cttatggttacttttgAcgcggtactactcc	GGAGTAGTACCCGCGTCCAAAG TAACCATAAG	1MM-C
G253A	cttatggttacttttgAagcggtactactcc	GGAGTAGTACCCGCTTCCAAAG TAACCATAAG	1MM-A

b、用于构建空白sgRNA表达质粒的引物

Name	Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')	
sgRNA			
expression	TCCGGGAGCTGCATGTGTCA	GGGTACCTCTAGAGCCATTTG	addgene#42230
plasmid			

c、通过定点诱变构建HF-Cas9 (R661A, Q695A, Q926A, D1135E) 的引物

Cas9			
variation	Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')	
Cas9R661A	ACCGGCTGGGGCGCGCTGAGCCGGAAG CAACAGAACTTCATGGCGCTGATCCACG	CTTCCGGCTCAGCGCGCCCAGCCGGT GTCGTCGTGGATCAGCGCCATGAAGTTT	
Cas9Q695A	ACGAC	CTGTTGG	
Cas9Q926A	CTGGTGGAAACCCGGGCGATCACAAAGC	CACGTGCTTTGTGATCGCCCGGGTTTCC	

以模拟在一条等位基因上携带插入缺失的杂合单细胞克隆。getPCR引物的序列展示如表2所示。对于基因组DNA样品的indel的频率定量，使用2.5ng基因组DNA作为模板，并使用表3中所总结的引物进行扩增。

表2 基因组编辑效率测定

a. Surveyor DNA 扩增和 sanger 测序的引物

Gene	Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')
HOXB13	CCGGCAATTATGCCACCTTG	GGTGGGTTCTGTTCTCCCTG
DYRK1A	GGAGCTGGTCTGTTGGAGAA	TCCAATCCATAATCCCACGTT
EMX1	CCATCCCCTTCTGTGAATGT	GGAGATTGGAGACACGGAGA

b. getPCR 用于检测 HOXB13 gRNA 靶标 4 的插入缺失的引物

Primer ID	Sequence (5' - 3')	
universal-F	CCTGGGGTGCCCCAGGGGAC	Forward primer
R+1	GCAGGAGTAGTACCCGCCTC	watching primer-R
R+2	CAGGAGTAGTACCCGCCTCC	watching primer-R
R+3	CAGGAGTAGTACCCGCCTCCA	watching primer-R
R+4	CAGGAGTAGTACCCGCCTCCAA	watching primer-R
R+5	AGGAGTAGTACCCGCCTCCAAA	watching primer-R
R+6	GGAGTAGTACCCGCCTCCAAAG	watching primer-R
R+7	GGAGTAGTACCCGCCTCCAAAGT	watching primer-R
R+8	GGAGTAGTACCCGCCTCCAAAGTA	watching primer-R
universal-R	GGGGCGGCTGGGGTACTCTTC	Forward primer
F+1	CCCGTGCCTTATGGTTACTTTGGA	watching primer-F
F+2	CCGTGCCTTATGGTTACTTTGGAG	watching primer-F
F+3	CGTGCCTTATGGTTACTTTGGAGG	watching primer-F
F+4	GTGCCTTATGGTTACTTTGGAGGC	watching primer-F
F+5	GCCTTATGGTTACTTTGGAGGCG	watching primer-F
F+6	GCCTTATGGTTACTTTGGAGGCGG	watching primer-F
F+7	CTTATGGTTACTTTGGAGGCGGG	watching primer-F
F+8	CTTATGGTTACTTTGGAGGCGGGT	watching primer-F
universal-R	GGGGCGGCTGGGGTACTCTTC	Forward primer
F+3-1	TGCCTTATGGTTACTTTGGAGG	watching primer-F
F+3-2	GTGCCTTATGGTTACTTTGGAGG	watching primer-F
F+3-3	CGTGCCTTATGGTTACTTTGGAGG	watching primer-F
F+3-4	CCGTGCCTTATGGTTACTTTGGAGG	watching primer-F
F+3-5	CCCGTGCCTTATGGTTACTTTGGAGG	watching primer-F
F+4-1	GCCTTATGGTTACTTTGGAGGC	watching primer-F
F+4-2	TGCCTTATGGTTACTTTGGAGGC	watching primer-F
F+4-3	GTGCCTTATGGTTACTTTGGAGGC	watching primer-F
F+4-4	CGTGCCTTATGGTTACTTTGGAGGC	watching primer-F
F+4-5	CCGTGCCTTATGGTTACTTTGGAGGC	watching primer-F

universal-F	CCTGGGGTGCCCCAGGGGAC	Forward primer
R+4-1	GGAGTAGTACCCGCCTCCAA	watching primer-R
R+4-2	AGGAGTAGTACCCGCCTCCAA	watching primer-R
R+4-3	CAGGAGTAGTACCCGCCTCCAA	watching primer-R
R+4-4	GCAGGAGTAGTACCCGCCTCCAA	watching primer-R
R+4-5	GGCAGGAGTAGTACCCGCCTCCAA	watching primer-R
R+3-1	GGAGTAGTACCCGCCTCCA	watching primer-R
R+3-2	AGGAGTAGTACCCGCCTCCA	watching primer-R
R+3-3	CAGGAGTAGTACCCGCCTCCA	watching primer-R
R+3-4	GCAGGAGTAGTACCCGCCTCCA	watching primer-R
R+3-5	GGCAGGAGTAGTACCCGCCTCCA	watching primer-R
universal-R	CAGTGGGGCGGCTGGGGTA	Reverse primer
253-G	CCCGTGCCTTATGGTTACTTTGGAG	watching primer-F
253-C	CCCGTGCCTTATGGTTACTTTGGAC	watching primer-F
253-A	CCCGTGCCTTATGGTTACTTTGGAA	watching primer-F
253-T	CCCGTGCCTTATGGTTACTTTGGAT	watching primer-F

表 3.细胞基因组编辑效率

a. 用于 indel 效率定量的 getPCR 引物

Primer 1 sequence (5' - 3')			
Control Primers		Primer 2 sequence (5' - 3')	
HB13-104ctrl		GCGACATGACTCCCTGTTGCCTGTG	
GACCTGGTGGGTTCTGTTCTCCCTG			
Primer 1 sequence (5' - 3') (Watching primer)			
Target site	getPCR Primers	Primer 2 sequence (5' - 3')	
DYRK1A target 1	DY-Ro4	tggggcatctgtccagccatct	
DYRK1A target 1	DY-Fo3	ttgtaggaggagagacttcagcatgc	
DYRK1A target 1	DY-Fo3	tgetgctggccttcagatggc	
DYRK1A target 1	DY-Fo4	tatgataaggcagaaacctgttggtcac	
DYRK1A target 1	DY-Fo4	tgetgctggccttcagatggct	
DYRK1A target 1	DY-Fo5	tatgataaggcagaaacctgttggtcac	
DYRK1A target 1	DY-Fo5	ctgctggccttcagatggctg	
Primer 1 sequence (5' - 3')			
Control Primers		Primer 2 sequence (5' - 3')	
DY-99ctrl		GCGATGTTGTTTGCCGTAAACCTGGC	
GACTTCTCCAACAGACCAGCTCCTC			
Primer 1 sequence (5' - 3') (Watching primer)			
Target site	getPCR Primers	Primer 2 sequence (5' - 3')	
EMX1 target 1	EMX1T1-Ro4	CCATCCCCTTCTGTGAATGTtagaccc	
EMX1 target 1	EMX1T1-Ro4	ctgagctgagaGcctgatggga	
Primer 1 sequence (5' - 3')			
Control Primers		Primer 2 sequence (5' - 3')	
EMX1-113ctrl		CGATGTCACCTCCAATGACTAGGGTG	
CAGGGAGTGGCCAGAGTCCAGCT			
Primer 1 sequence (5' - 3') (Watching primer)			
Target site	getPCR Primers	Primer 2 sequence (5' - 3')	
EMX1 target 3	EMX1T3-Fo3	gtgggcaaccacaaaccacga	
EMX1 target 3	EMX1T3-Fo3	aggggcctggccagcagca	
EMX1 target 3	EMX1T3-Ro5	agcactctgcctcgtgggt	
EMX1 target 3	EMX1T3-Ro5	gatgtcacctccaatgactagggt	

EMX1 target 4	EMX1T4-Ro3	aggggcctggccagcagca	gtgggcaaccacaaacccacga
EMX1 target 4	EMX1T4-Fo4	ggcagagtgtgtgtgtgtgt	TCCCCAAAGCCTGGCCAGGGAGT
EMX1 target 5	EMX1T5-Fo4	gatgtcacctccaatgactagggt	agcactctgcctctgtgggt
EMX1 target 5	EMX1T5-Ro4	ggtttgtggttggccaccctagt	gcctgagtcaggagcagaagaagaa
EMX1 target 6	EMX1T6-Fo3	gcctgagtcaggagcagaagaagaa	ggtttgtggttggccaccctagt
EMX1 target 6	EMX1T6-Ro4	ttgatgtgatgggagcccttcttct	gaggccccagtggtgtgtct
Control Primers		Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')
EMX1-111ctrl		CCATCCCCTTCTGTGAATGTtagacc	TGAGCTGAGAGCCTGATGGGAAGAC

b. getPCR 引物用于碱基编辑效率定量

		Primer 1 sequence (5' - 3')	
Target site	getPCR Primers	(Watching primer)	Primer 2 sequence (5' - 3')
HOXB13 target 8	C8	cgccaggggtggtgcctG	ccgtgccttatggttactttggagg
HOXB13 target 8	T8	cgccaggggtggtgcctA	ccgtgccttatggttactttggagg
		Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')
Control Primers			
HB13-104ctrl		GCGACATGACTCCCTGTTGCCTGTG	GACCTGGTGGGTTCTGTTCTCCCTG
		Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')
Target site	getPCR Primers	(Watching primer)	Primer 2 sequence (5' - 3')
EMX1 target 6	5C	ggaggaggaagggcctgagtC	ggtttgtggttggccaccctagt
EMX1 target 6	5T	tggaggaggaagggcctgagtT	ggtttgtggttggccaccctagt
EMX1 target 6	6C	ggagcccttcttcttctgtctG	gaggccccagtggtgtgtct
EMX1 target 6	6T	ggagcccttcttcttctgtctA	gaggccccagtggtgtgtct
EMX1 target 6	CC	gaggaggaagggcctgagtCC	ggtttgtggttggccaccctagt
EMX1 target 6	TT	ggaggaggaagggcctgagtTT	ggtttgtggttggccaccctagt
EMX1 target 6	CT	gaggaggaagggcctgagtCT	ggtttgtggttggccaccctagt
EMX1 target 6	TC	gaggaggaagggcctgagtTC	ggtttgtggttggccaccctagt
		Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')
Control Primers			
EMX1-111ctrl		CCATCCCCTTCTGTGAATGTtagacc	TGAGCTGAGAGCCTGATGGGAAGAC

c. 用于 HDR 修复效率量化的 getPCR 引物

		Primer 1 sequence (5' - 3')	
Target site	getPCR Primers	(Watching primer)	Primer 2 sequence (5' - 3')
EMX1 target 5	EMX1T5HindIII-F	tcatgtcacctccaatgactAAGCTT	agcactctgcctctgtgggt
EMX1 target 5	EMX1T3o5-R	gggttgtggttggcccaAGCTTag	gcctgagtcaggagcagaagaagaa
Control Primers		Primer 1 sequence (5' - 3')	Primer 2 sequence (5' - 3')
EMX1-111ctrl		CCATCCCCTTCTGTGAATGTtagacc	TGAGCTGAGAGCCTGATGGGAAGAC

Surveyor核酸酶分析 使用已经报道过的Surveyor核酸酶测定法测定Indel频率，用到了Surveyor®突变检测试剂盒（IntegratedDNATechnologies, Cat # 706020）。过程简单陈述如下：根据产品手册，使用

TIANampGenomicDNAKit (TIANGEN, Cat # DP304-03) 提取基因组DNA; 然后使用高保真PrimeSTAR®MaxDNA聚合酶 (TaKaRa, Cat # R045B) 扩增得到DNA片段, 该片段的任一末端距离cas9的切割位点距离200-400bp, PCR所用的引物展示在表2a中。使用T100TM热循环仪 (Bio-Rad) 对270ng纯化的PCR产物进行退火, 以得到异源的双链体, 随后根据使用说明用SurveyorNuclease处理。用2%琼脂糖凝胶分离DNA片段, 并使用Quantum-ST5 (VILBERLOURMAT, France) 获得图像, 并用QuantumST5Xpress软件分析。

getPCR在HDR和BE4实验中的应用 如表3中所总结的, 在3' 末端设计具有修饰核苷酸的修饰特异性getPCR引物。在getPCR分析中, 使用2.5ng基因组DNA作为每个反应的模板。使用如图6a中所示的公式计算基因组修饰效率。

基于HindIII酶切的RFLP测定。在针对EMX1基因的HDR实验中, 在PAM序列附近引入一个HindIII位点, 这通过基于HindIII酶切的限制性片段长度多态性 (RFLP) 分析来实现HDR修复效率的定量。简言之, 使用PrimeSTARMaxDNA聚合酶扩增639bp片段, HindIII位点距离5' 末端有355bp的长度, PCR用到的引物与Surveyor测定法相同, 如表2a中所示。使用UniversalDNA纯化试剂盒 (TIANGEN, Cat # DP214) 将PCR产物纯化。取纯化的270ngPCR产物进行HindIII酶切实验, 并在2%琼脂糖凝胶上分离。使用Quantum-ST5 (VILBERLOURMAT, France) 获取图像并使用QuantumST5Xpress软件进行分析。

基于NGS的方法 将覆盖基因组编辑位点附近的DNA区域进行NGS扩增子文库的构建, 进行测序后通过计数NGS的读数来计算编辑效率。使用基因组DNA作为模板, 进行两轮PCR扩增以制备测序文库。在第一轮PCR中, 设计250-280bp的扩增子, 其中Cas9切割位点靠近中间部分, 两端引入Illumina测序引物的结合位点。在第二轮PCR中, 引入了接头序列用于测序过程中的簇生成, 同时还引入了索引序列。文库DNA进行纯化和定量后, 交付Genewiz公司在IlluminaHiSeqX-TEN平台上进行150bp双末端测序。对于NHEJ介导的插入缺失, 用野生型DNA的特征序列获得每个文库中的野生型读数计数, 并使用公式“编辑效率=1-野生型读数/总读数*100%”计算插入缺失编辑效率。关于碱基编辑和HDR实验中的编辑效率, 获得文库中预期DNA序列的读数计数, 并使用等式“效率=预期DNA序列的读数/总读数*100%”计算编辑效率。有关文库制备和计数方法的详细信息, 请参见表4。

表 4 通过 NGS 进行基因组编辑效率定量

a. 用于文库制备的引物

target site	Primer ID	First Round Primer 1 sequence (5' - 3')	First Round Primer 2 sequence (5' - 3')
EMX1-T1	EMX1-1lib	ACTCTTTCCCTACACGACGCT	GTGACTGGAGTTCAGAC
		CTTCCGATCTccatcccccttctgtgaatgtagacc	GTGTGCTCTTCCGATCT
			ccttctcctccagcttctgcccgt
EMX1-T2	EMX1-1lib	ACTCTTTCCACACGACGCT	GTGACTGGAGTTCAGA
		CTTCCGATCTccatcccccttctgtgaatgtagacc	CGTGTGCTCTTCCGATCT
			ccttctcctccagcttctgcccgt
EMX1-T3	EMX1-2lib	ACTCTTTCCCTACACGACGCT	GTGACTGGAGTTCAGAC
		CTTCCGATCTggttcagaaccggaggacaaagtac	GTGTGCTCTTCCGATCT
			gactccaggcctcccaaaagcctg
EMX1-T4	EMX1-2lib	ACTCTTTCCCTACACGACGCTC	GTGACTGGAGTTCAGAC
		TTCCGATCTggttcagaaccggaggacaaagtac	GTGTGCTCTTCCGATCT
			gactccaggcctcccaaaagcctg
EMX1-T5	EMX1-2lib	ACTCTTTCCCTACACGACGCTC	GTGACTGGAGTTCAGAC

		TTCCGATCTggttccagaaccggaggacaaagtac	GTGTGCTCTTCCGATCT gactccaggcctccccaagcctg GTGACTGGAGTTCAGAC
		ACTCTTTCCTACACGACGCTC	GTGTGCTCTTCCGATCT
EMX1-T6	EMX1-2lib	TTCCGATCTggttccagaaccggaggacaaagtac	gactccaggcctccccaagcctg GTGACTGGAGTTCAGAC
		ACTCTTTCCTACACGACGCTC	GTGTGCTCTTCCGATCT
EMX1-T6-BE4	EMX1-2lib	TTCCGATCTggttccagaaccggaggacaaagtac	gactccaggcctccccaagcctg GTGACTGGAGTTCAGAC
		ACTCTTTCCTACACGACGCTC	GTGTGCTCTTCCGATCT
EMX1-5 with HindIII	EMX1-3lib	TTCCGATCTgtggttccagaaccggaggacaaagta	ctccaggcctccccaagcctggc GTGACTGGAGTTCAGA CGTGTGCTCTTCCGATC Ttccatgaacttacctggttagtt
		ACTCTTTCCTACACGACGCTC	
DYRK1A	DYRK1A1lib	TTCCGATCTggatatgaatatttcctttaaacctcac	GTGACTGGAGTTCAGA CGTGTGCTCTTCCGATC
		ACTCTTTCCTACACGACGCT	
HOXB13-T6	HB13-251lib	CTTCCGATCTctgtcaactatgcccccttgatctg	Tggcaaaactcagtgggcggtgg GTGACTGGAGTTCAGA
		ACTCTTTCCTACACGACGCTC	CGTGTGCTCTTCCGATC
HOXB13-T8 (BE4)	HB13-251lib	TTCCGATCTctgtcaactatgcccccttgatctg	Tggcaaaactcagtgggcggtgg GTGACTGGAGTTCAGA CGTGTGCTCTTCCGATC
		ACTCTTTCCTACACGACGCTC	Tgttccagccaccagcgagagccc
HOXB13-T16	HB13-404lib	TTCCGATCTcccggagctcgctgaaaccctgtg	

1st round PCR, take 50ng gDNA as template, 28 cycles, 15μl system, set NTC control, annealed @60°C,
using PrimeSTAR® Max DNA Polymerase (TaKaRa)

Second Round	Primer ID	Primer sequence (5' - 3')
		AATGATACGGCGACCACCGAGATCTACACA
	libUni-F-universal	CACTCTTTCCTACACGACGCTCTTCCGA CAAGCAGAAGACGGCATACGAGATCGAGT
	libUni-D701-R	AATGTGACTGGAGTTCAGACGTGTGCTCTT CAAGCAGAAGACGGCATACGAGATTCTCC
	libUni-D702-R	GGAGTGACTGGAGTTCAGACGTGTGCTCTT CAAGCAGAAGACGGCATACGAGATAATGA
	libUni-D703-R	GCGGTGACTGGAGTTCAGACGTGTGCTCTT CAAGCAGAAGACGGCATACGAGATGGAAT
	libUni-D704-R	CTCGTGACTGGAGTTCAGACGTGTGCTCTT CAAGCAGAAGACGGCATACGAGATTTCTG
	libUni-D705-R	AATGTGACTGGAGTTCAGACGTGTGCTCTT

	CAAGCAGAAGACGGCATACGAGATACGAA
libUni-D706-R	TTCGTGACTGGAGTTCAGACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGCTT
libUni-D707-R	CAGGTGACTGGAGTTCAGACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGCGC
libUni-D708-R	ATTAGTGACTGGAGTTCAGACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCATA
libUni-D709-R	GCCGGTGACTGGAGTTCAGACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATTTCG
libUni-D710-R	CGGAGTGACTGGAGTTCAGACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGCGC
libUni-D711-R	GAGAGTGACTGGAGTTCAGACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCTAT
libUni-D712-R	CGCTGTGACTGGAGTTCAGACGTGTGCTCTT

2nd round PCR, take 1ng of purified DNA from 1st round PCR as template, 10cycles, 15μl system, annealed @65°C, using PrimeSTAR® Max DNA Polymerase (TaKaRa)

b. R 程序读取计数的特征序列

Target	Feature	characteristic sequence for counting
HOXB13 target 6	Wild Type	ttacttttgagg
HOXB13 target 16	Wild Type	tccgggaacct
DYRK1A target 1	Wild Type	tcagatggctgg
EMX1 target 1	Wild Type	cttcccatcagg
EMX1 target 3	Wild Type	aaccacagagg
EMX1 target 4	Wild Type	gcttgctgctgg
EMX1 target 5	Wild Type	tgactagggtgg
EMX1 target 6	Wild Type	gaagaagaagg
EMX1 target 5	HindIII	actAAGCTTggg
HOXB13 target 8	C8	gcggccagggtggctgctG
HOXB13 target 8	T8	gcggccagggtggctgctA
EMX1 target 6	T5	gaggaggaaggcctgagtT
EMX1 target 6	C5	gaggaggaaggcctgagtC
EMX1 target 6	T6	Tgagcagaagaagaagggt
EMX1 target 6	C6	Cgagcagaagaagaagggt
EMX1 target 6	C5C6	aggaggaagggcctgagtCC
EMX1 target 6	T5T6	aggaggaagggcctgagtTT
EMX1 target 6	C5T6	aggaggaagggcctgagtCT
EMX1 target 6	T5C6	aggaggaagggcctgagtTC

c. R 读取计数程序

library (ShortRead)

```

reads=readFastq("libraryName")
reads
total_counts=length(reads)
total_counts
sequences=sread(reads)
dict=DNAStrngSet(substr(sequences, 1, 150))
hits=vcountPattern("Wild Type characteristic sequence",dict,max.mismatch = 0,with.indels = FALSE)
wild_type_counts=sum(hits)
wild_type_counts

library(ShortRead)
reads=readFastq("libraryName")
reads
total_counts=length(reads)
total_counts
sequences=sread(reads)
dict=DNAStrngSet(substr(sequences, 1, 150))
hits=vcountPattern("expected_characteristic sequence",dict,max.mismatch = 0,with.indels = FALSE)
expected_sequence_counts=sum(hits)
expected_sequence_counts

```

单细胞克隆和基因分型转染后约48小时，通过有限稀释法分离单细胞并接种到96孔板中生长。当细胞长满96孔板时，将其进一步转移到24孔板中继续生长至愈合。然后根据制造商的说明用TIANamp基因组DNA试剂盒（TIANGEN，Cat#DP304-03）分离来自单细胞克隆的基因组DNA。通过getPCR检测确定每个克隆的基因型，并通过覆盖切割位点的扩增子的Sanger测序确认。使用到的引物如表2a中所示，用高保真PrimeSTARMaxDNA聚合酶（TaKaRa，Cat#R045B）进行PCR扩增，然后对PCR产物进行Sanger测序（TsingKeBiologicalTechnology或GeneWiz）。为了确定杂合细胞的每个等位基因的确切序列，使用TIDEWeb工具（<https://tide.nki.nl/>）直接分析Sanger测序ab1文件，或者将扩增子克隆到载体中之后，再对菌落进行Sanger测序。

不同DNA聚合酶对错配的敏感性 使用多种商业DNA聚合酶产物来比较引物错配对扩增的影响。它们是Taq master mix (Vazyme, Cat#P111, Lot#511151), Premix Taq™ (TaKaRa, Cat#RR901, Lot#A3001A), NOVA Taq-Plus PCR Forest Mix (Yugong Biolabs, Cat#EG15139, Lot#1393216101), DreamTaq Green PCR Master Mix (ThermoFisher, Cat#K1081, Lot#00291017), Platinum™ Green Hot Start PCR Master Mix (Invitrogen, Cat#13001012, Lot#00401653), PrimeSTAR® Max DNA Polymerase (TaKaRa, Cat#R045, Lot#A151995A), Phusion Hot Start II high-Fidelity PCR Master Mix (ThermoFisher, Cat#F-565, Lot#00633307) as well as Q5® Hot Start high-Fidelity DNA Polymerase (NEB, Cat#M0493)。在20 μl反应体系中，使用10ng质粒DNA作为模板，并按照产品手册所建议的程序进行热循环。然后将PCR产物直接进行2.0%琼脂糖凝胶电泳和Sanger测序。使用Quantum-ST5（VILBERLOURMAT, France）获得凝胶图像，并用QuantumST5Xpress软件分析。

不同qPCR SYBR Green产品在getPCR中的比较 为了测试getPCR的广泛可用性，将多种qPCRSYBRmix产品应用于getPCR，包括AceQ qPCR SYBR Green Master Mix (Vazyme, Cat#Q111-02), SYBR™ Select Master

Mix (Applied Biosystems™, Cat#4472908), Power SYBR Green PCR Master Mix (Applied Biosystems™, Cat#4367659), QuantiNova SYBR Green PCR Kit (QIAGEN, Cat#208054), FastStart Essential DNA Green Master (Roche, Cat#06402712001), NovoScript® SYBR One-Step qRT-PCR SuperMix (novoprotein, Cat#E092-01A), 2×T5 Fast qPCR Mix (TSINGKE, Cat#TSE202), UltraSYBR Mixture (CWBI0, Cat#CW0957), SYBR Premix Ex Taq (TaKaRa, Cat#RR420, A5405-1)。实时定量PCR在热循环仪Rotor-GeneQ (Qiagen, 德国) 或LightCycler®96qPCR仪 (德国罗氏应用科学公司) 上运行。根据制造商的说明书和设定的退火温度确定qPCR条件。

统计分析 基于Levene测试的结果应用学生t检验 (双尾) 以评估使用IBMSPSSStatistics版本的单细胞克隆基因分型的getPCR结果的统计学显著性。使用Pearson测试评估两种不同getPCR策略之间的相关性, 用到了IBMSPSSStatistics软件的第21个版本。

实施例1 getPCR中值守碱基的设计

为了使getPCR技术能更好的应用, 本实施例中针对值守碱基的设计规则进行研究。由于大多数indel出现在核酸酶切割位点附近, 且小于15bp的indel占主要部分, 此外, 要想使indel序列与野生型序列更好的区分开来, 本实施例中针对碱基数目较少的插入或缺失的情况进行考察。鉴于此, 发明人设计了设计构建了26个质粒, 分别带有1-15bp的插入缺失突变体, 以模拟在体内靶向HOXB13基因的核酸酶诱导的基因组编辑 (图2a)。

本实施例中设计了两个系列的事守碱基, 它们分别具有一至八个值守碱基 (图7a-c), 从中筛选出具有理想扩增效率的代表 (图2b) 以进一步检查它们的鉴别能力, 即区分indel和野生型DNA序列的能力。从理论上讲, 更多的事守碱基可以增加值守碱基的选择性。然而, 过多的值守碱基会使碱基错配从引物的3'末端向5'末端方向移动, 反而会降低Taq聚合酶的敏感性。单独使用单向的事守碱基时, 不论是对于反向 (图2c) 和正向 (图2d) 引物, 3至5个值守碱基就能表现出优越的区分野生型序列和indel序列的能力。当正向和反向值守碱基组合使用时, 总共4到6个值守碱基可成功区分插入缺失 (图2e, 图7d)。然而, 由于引物自扩增, 累计5到6个值守碱基的时候就会显示出更高的背景信号 (图2f, 图7e)。因此, 累计4个值守碱基的设计组合是getPCR引物的理想选择。

本实施例进一步对值守碱基的3'末端碱基类型在决定getPCR的鉴别能力中起重要作用。当与模板中非互补配对碱基形成错配时, 腺嘌呤碱基显示出最佳的特异性并且给出最低的非特异扩增信号。接下来是胞嘧啶和鸟嘌呤, 最后是胸腺嘧啶 (图2g)。当错配位于倒数第二位置时, 腺嘌呤碱基仍显示出最佳的特异性, 并且Taq聚合酶对腺嘌呤与非互补配对碱基之间的错配容忍度最低 (图2h)。此外, 3'末端碱基类型也决定了getPCR对上游错配的敏感性。在这里, 腺嘌呤碱基同样也是最佳选择, 它能够使getPCR扩增对其上游的倒数第二位的碱基错配更为敏感。值得注意的是, 如果在最后一个碱基附近发生多于一个错配, 则无论最后一个碱基是什么, 都将明显破坏PCR的扩增能力 (图2i)。此外, 不匹配碱基越接近3'端, getPCR变得越敏感 (图7f-g, 图8a-b)。

当探讨getPCR对错配敏感的潜在机制时, 本实施例中将3'末端带有错配碱基的引物与删除了该错配碱基的引物分别去进行PCR扩增并进行比较。有趣的是, 删除了错配碱基的引物在qPCR以及常规PCR分析中都能部分恢复扩增能力 (图7h-i, 图8a-b)。此外, 具有校对活性即3'至5'核酸外切酶活性的高保真DNA聚合酶如Phusion和Q5也可部分或完全恢复PCR扩增能力。PCR产物的Sanger测序结果显示引物3'末端的错配核苷酸在聚合过程中能够被3'至5'核酸外切酶活性除去。相反, 缺乏3'至5'核酸外切酶活性的Taq DNA聚合酶可以容忍并直接绕过错配 (图8c)。这表明, 错配一方面阻碍了引物与模板配对, 同时由错配引起的空间几何障碍也进一步阻碍了Taq聚合酶合成反应的起始。

实施例2 运行getPCR的参数

另一个需要确定的因素是getPCR运行的最佳参数，本实施例中针对getPCR反应时的退火温度进行研究。对于实施例1中设计的四组值守碱基，随着退火温度的升高，对比于含错配碱基indel模板，getPCR特异性扩增野生型模板DNA的能力明显增加（图3a-d）。然而，当退火温度升高到高于 T_m 值4℃以上时，PCR效率开始显著下降。由于PCR扩增通常优选最佳PCR效率，因此本实施例在最佳PCR效率下系统地评估每种值守碱基的选择性（图3e-h）。有趣的是，无论引物具有多少个值守碱基或者总碱基数，在退火温度比其 T_m 值高约4℃时通常能观察到最佳的选择性（图3e-h）。由于固定的值守碱基数目，通过在引物的5'末端添加更多碱基来增加引物 T_m 值不会显著改变区分插入缺失的能力。四种类型引物中的三种表现出稳定的鉴别插入缺失的能力（图3e-g）。只有一种类型的引物显示略微增加的能力，并在65.8℃左右的 T_m 值达到最佳值（图3h）。因此，在随后的实验中，值守碱基被设计为具有约65℃的 T_m 值并且在退火温度为69℃下进行getPCR。更重要的是，即使升高退火温度至超过 T_m 值可能会阻碍PCR效率，但对于这四种引物来讲，实时PCR定量的基础也就是Ct值和对数模板DNA量之间的线性相关性，却根本不会受到影响。（图3i-1）。DNA聚合酶在决定getPCR的辨别能力中起着重要作用。

本实施例中针对多种商业Taq酶进行测试，其性能表现不尽相同，但几乎全都表现出可观的足以区分插入缺失与野生型序列的能力（图7j）。然而，当评估对单碱基错配的敏感性时，9种SYBR Green qPCR商品中的7种都显示出较高的适用性能（图8d-e）。

实施例3 getPCR定量基因组编辑准确性的研究

图2a中所示的是用来模拟基因组编辑导致的插入缺失突变（indel）的质粒，首先用它们来评估getPCR定量基因组编辑效率的能力。本实施例中将二十六个indel质粒等份混合，然后再按特定比例与野生型质粒混合，以模拟0%，20%，40%，60%，80%和100%的插入缺失频率。通过getPCR以及经典的Surveyor方法对混合物进行插入缺失频率定量并进行比较。当indel频率不高于20%时，Surveyor方法的量化结果可以真实地反映预期值。然而，随着indel频率的进一步增加，观察值便逐渐偏离预期值（图4a-b）。相反，无论是值守碱基上携带3, 4或5个值守碱基，使用不同值守碱基的所有12种getPCR策略都可以准确地定量插入缺失频率（图4c，图9a-c）。

实施例4 getPCR在模拟单细胞克隆的基因分型中的应用

基因组编辑实验中的单细胞克隆筛选或子代基因分型是getPCR技术的另一个重要应用。单独使用图2a中所示的每个indel质粒，或将每个indel质粒与野生型质粒等比例混合，来模拟两个等位基因或一个等位基因发生编辑的单细胞克隆基因组DNA。所有三种getPCR策略不仅可以确定是否发生了插入缺失，而且还可以准确地确定是一个等位基因还是两个等位基因发生了插入缺失突变（图4d-f）。此外，当将任何两种getPCR策略进行组合分析时，它们的检测值还表现出极高的相关性，Pearson相关系数等于或高于0.995。有趣的是，两种getPCR策略的组合可以显著提高鉴定基因型的性能（图9d-f）。

实施例5 getPCR确定单细胞克隆的编辑频率和基因型

本实施例在Lenti-X 293T细胞中用Cas9和靶向HOXB13, DYRK1A或EMX1基因的8条不同的gRNA进行了基因组编辑，并应用getPCR检测编辑效率（图5b）。每种gRNA的编辑效率分别通过三种不同的方法确定，即getPCR、基于NGS的扩增子测序以及Surveyor分析。

对于设计的所有值守碱基，getPCR方法检测的编辑效率通常与NGS方法的结果相一致，而NGS是迄今为止被认为是最可靠的方法。相比之下，Surveyor方法测定出的编辑效率值与其它两种方法有明显的偏差，特别是当HOXB13基因上目标6和目标16的编辑效率较高时（图5a）。本实施例中对接受HOXB13基因的靶标6，EMX1基因的靶标1和靶标5以及DYRK1A基因的靶标1进行了基因组编辑的细胞，进行分离单细胞克隆并繁殖扩增。制备基因组DNA样品后通过getPCR进行基因分型，同时使用Sanger测序进行验证。总体而言，用这四种gRNA靶标进行基因组编辑实验的所有单细胞克隆都可以通过getPCR进行准确的基因分型。值得注意的

是，getPCR不仅可以检测出携带插入缺失的细胞克隆，而且同时可以成功鉴定出该细胞克隆是一条等位基因发生编辑还是两条等位基因均发生了编辑（图5c-i，图10a-b）。对于在HOXB13基因靶标6处进行的基因组编辑，使用分别含有3或4个值守碱基的两种不同设计的getPCR引物，从总共42个细胞克隆中准确鉴定出了24个双等位基因编辑的细胞和5个单等位基因编辑的细胞（图5c-d，图10h）。类似地，在EMX1基因靶标5处基因组编辑，分别使用带有4个值守碱基的正向和反向引物通过getPCR鉴定出8个双等位基因编辑的细胞和5个单等位基因编辑的细胞（图5e-f，图10i）。在DYRK1A基因的靶标1处，使用getPCR从总共53个单克隆细胞中筛选出11个双等位基因编辑的细胞，5个单等位基因编辑的细胞，用到了四条不同设计的值守碱基，其中三条是正向引物携带有3, 4或5个值守碱基，一条为反向引物携带有4个值守碱基（图5g-h，图10a-b, j）。对于EMX1基因的靶1处，使用携带4个值守碱基的引物的getPCR成功鉴定了来自45个克隆的1个双等位基因编辑的细胞克隆和9个单等位基因编辑的细胞克隆（图5i，图10k）。值得注意的是，任何两种不同设计的getPCR策略都显示出高度相关的检测值，并且当组合分析时可以帮助进行基因分型（图5j-l，图10c-g）。

实施例6 getPCR应用于确定HDR频率和单细胞克隆的基因型

本实施例中针对getPCR应用于HDR的基因组编辑修复效率的测定进行说明（图6a）。在Lenti-X 293T细胞中进行Cas9介导的基因组编辑实验，其中用到了EMX1基因的靶标5 gRNA和HDR模板，在PAM序列相邻的位置引入了HindIII位点序列（图6b）。应用getPCR方法以及基于NGS的扩增子测序和HindIII介导的限制性片段长度多态性（RFLP）分析来确定修复效率。结果表明，分别使用正向和反向的两条值守碱基的getPCR法都可以确定HDR频率，检测结果与基于RFLP和NGS方法的检测结果高度一致（图6c），根据评估，来自三个生物重复样品的HDR频率约为25%。此外，对在EMX1靶标5处进行Cas9介导的HDR修复的细胞，进行分离单细胞克隆并繁殖扩增，得到了50个单细胞克隆，分别使用两条值守碱基通过getPCR对它们进行基因分型，成功地挑选出6个纯合修复的细胞克隆和17个单等位基因修复的细胞克隆（图6d-e）。此外，这两个值守碱基的检测值具有高度一致性即强相关性（ $r = 0.982$, $P = 1.207 \times 10^{-36}$ ），并且二者的组合分析可以明显更好的实现基因分型，特别是对于杂合的细胞克隆（图6f）。

实施例7 getPCR确定碱基编辑器的编辑频率和单细胞克隆基因型

本实施例针对getPCR在碱基编辑器的编辑频率和单细胞克隆基因型检测方面的应用进行说明。本实施例在Lenti-X 293T细胞中用EMX1靶标6的gRNA或HOXB13靶标8的gRNA以及BE4碱基编辑器进行了基因组编辑，并应用getPCR检测编辑效率（图6b）。在碱基编辑频率的定量中，getPCR的检测结果与基于NGS的扩增子测序方法的结果高度一致（图6g-h）。对于EMX1靶标6，在gRNA靶向序列中第5位和第6位将约27%的“C”碱基转化为“T”。有趣的是，这两个位置的碱基编辑倾向于同时发生以产生T5T6基因型（图6g）。对于HOXB13靶标8的gRNA的碱基编辑，第8位的C到T的碱基变化频率约为15%，该变化能够通过引入一个前置的终止密码子‘TAG’来提前终止开放阅读框（图6h）。

在Lenti-X 293T细胞中用EMX1靶标6或HOXB13靶标8进行碱基编辑后，进一步分离单细胞克隆，并用getPCR方法进行基因分型。通过getPCR分析，对于EMX1靶标6的碱基编辑，确定46个克隆中的25个在第5位发生了C到T的转变（图6j-k），46个克隆中的22个克隆在第6位被证实携带C到T的转变（图6l-m）。getPCR检测结果显示丢失的碱基组成百分比显示，其中有三个克隆，E01，E29和E70可能在第5位含有除C和T以外的碱基，还有一个克隆E24可能在第6位携带这样的碱基。这些克隆的Sanger测序显示在E01和E29的第5位，E24的第6位发生了C到G的碱基编辑（图11a-c）。特别的，E70克隆在第5个核苷酸处不携带C到T之外的碱基转变，但在其一条等位基因上，在gRNA靶向序列的-8位核苷酸处具有A到T的突变（图11c）。该A到T的突变位于值守碱基的3'末端第14个核苷酸处，它阻止了引物退火到该等位基因上，并最终导致了丢失的getPCR信号。而Lenti-X 293T是一种基于HEK 293的细胞系，据报道其基因组接近三倍体，每个细胞有62-70

个染色体。于此一致地，在getPCR分析中，杂合克隆的每个等位基因的百分比通常为33%或66%左右，而不是50%（图6j，1）。

此外，在Sanger测序分析中这些三倍体特征得到了进一步验证，两个杂合等位基因峰图的高度通常具有两倍而不是对等的相互关系（图11c）。例如，在getPCR分析中，E11克隆的第5个核苷酸处的T和C碱基的百分比分别被确定为28.8%和62.9%，并且在Sanger测序中，C碱基的峰高几乎是T的两倍。然而，即使拿到了Sanger测序结果，仍有10个克隆的等位基因特异性基因型是未知的，仅仅知道它们在第5和第6核苷酸处均是杂合的（图11c）。本实施例设计了四个值守碱基以通过getPCR方法对这些克隆进行进一步基因分型（图6b），并且成功确定了这些克隆的确切等位基因特异性基因型（图6i）。克隆E02和E15定义为C5C6/C5C6/T5T6，并且E33，E39，E40以及E49被证明是C5C6/T5T6/T5T6。发现克隆E01和E29均为C5C6/T5T6/G5C6，并且E24，E34克隆最终分别确定为C5C6/T5C6/T5G6和C5C6/T5T6/T5C6。

对于在HOXB13靶标8处进行的碱基编辑以引入框内终止密码子，本实施例从49个细胞单克隆中确定了14个细胞克隆在sgRNA的第8位碱基处发生了C到T的转化，这将带来一个提前终止密码子（图6n-o）。值得注意的是，getPCR检测结果中碱基组成丢失的百分比表明，S37克隆可能在该位置携带除C和T碱基之外的额外碱基，Sanger测序显示在gRNA的第8位核苷酸处，三条等位基因之一的C碱基转变为了G碱基（图12a-b）。同样的，getPCR也可以确定杂合子克隆的精准基因型，这也被Sanger测序所证实。例如，在HOXB13 gRNA靶标8序列的第8个核苷酸处，对6个克隆S15，S47，S44，S18，S02和S35进行基因分型为C / C / T。

以上所述仅为本公开的优选实施例而已，并不用于限制本公开，对于本领域的技术人员来说，本公开可以有各种更改和变化。凡在本公开的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本公开的保护范围之内。

权利要求书

1. 一种用于检测核酸酶切诱导的 indel 发生频率的方法，其特征在于，所述方法包括以下步骤：向待测基因组样品中加入引物及 Taq DNA 聚合酶，对基因组样品中的野生型 DNA 进行扩增，通过 PCR 定量野生型 DNA 比例，从而确认基因组中 indel 发生的频率；所述引物序列与野生型 DNA 序列相匹配，并且覆盖核酸酶切割位点；优选的，所述 PCR 定量为实时 PCR 或 ddPCR；优选的，所述检测方法还包括以下步骤：在距离切割位点数百碱基对的位置处引入对照扩增，通过 $\Delta \Delta C_t$ 策略计算经过编辑后的基因组 DNA 样品中野生型 DNA 的百分比。
2. 如权利要求 1 所述的检测方法，其特征在于，所述核酸酶包括但不限于 Cas9 核酸酶、锌指核酸酶，转录激活因子样效应核酸酶和 CRISPR RNA 指导 FokI 核酸酶，以及成对的 cas9 切口酶；进一步的，所述核酸酶为 Cas9 核酸酶；所述引物 3' 末端跨越 Cas9 核酸酶切割位点。
3. 如权利要求 2 所述的检测方法，其特征在于，所述引物序列中包括值守碱基序列，所述值守碱基为核酸酶切位点与 3' 末端之间的序列，所述值守碱基长度为 1~8bp；优选的，所述引物为一条核苷酸序列，所述值守碱基的长度为 3~5bp；或所述引物为正向和反向的一对核苷酸序列，所述值守碱基的长度为 4bp。
4. 如权利要求 3 所述的检测方法，其特征在于，所述值守碱基 3' 末端碱基为腺嘌呤碱基或胞嘧啶或鸟嘌呤碱基；优选的，为腺嘌呤碱基。
5. 如权利要求 2 所述的检测方法，其特征在于，所述扩增反应的退火温度为 $T_m \sim T_m + 4^\circ\text{C}$ 。
6. 一种用于检测核酸酶切诱导的 indel 发生频率的试剂盒，所述试剂盒中包括引物、Taq DNA 聚合酶及 PCR 检测试剂。
7. 权利要求 6 所述的试剂盒在评价基因组编辑效率、单细胞克隆筛选方面的应用；优选的，所述基因组编辑包括 NHEJ 介导的插入缺失，HDR 介导的基因修饰及通过 BE4 产生的碱基编辑；优选的，所述应用还包括对筛选适配 CRISPR 的 gRNA。
8. 一种对单细胞克隆进行基因分型的方法，其特征在于，所述方法包括以下步骤：以待测基因组中野生型 DNA 为模板，针对等位基因设计引物，提取待测单细胞克隆的基因组 DNA，通过权利要求 1-5 任一项所述的检测方法检测单细胞基因组 DNA 中等位基因是否发生了 indel 从而对单细胞基因实现分型。
9. 一种 HDR 修复效率的检测方法，其特征在于，所述检测方法包括以下步骤：针对待测基因组中 HDR 修复的基因组 DNA 设计引物，提取待测细胞基因组 DNA，采用权利要求 1-5 任一项所述的检测方法检测 HDR 的发生概率；HDR 修复 DNA 所占百分比即 HDR 修复效率。
10. 一种碱基编辑器编辑效率的检测方法，其特征在于，所述检测方法包括以下步骤，以待测基因组 DNA 为模板，针对碱基编辑后的靶序列设计引物，采用权利要求 1-5 任一项所述的检测方法检测基因组中碱基编辑的发生概率，即为编辑器的编辑效率。

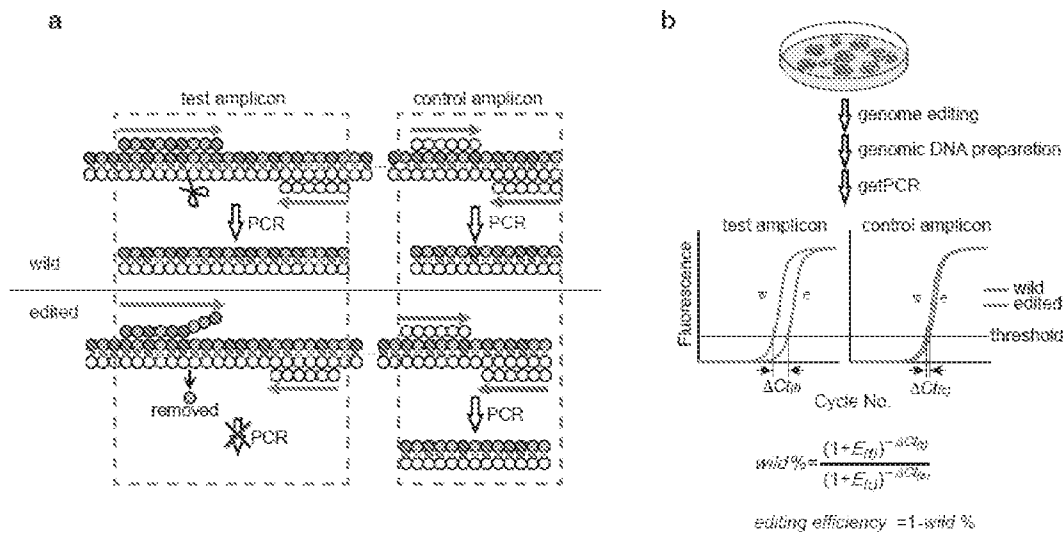
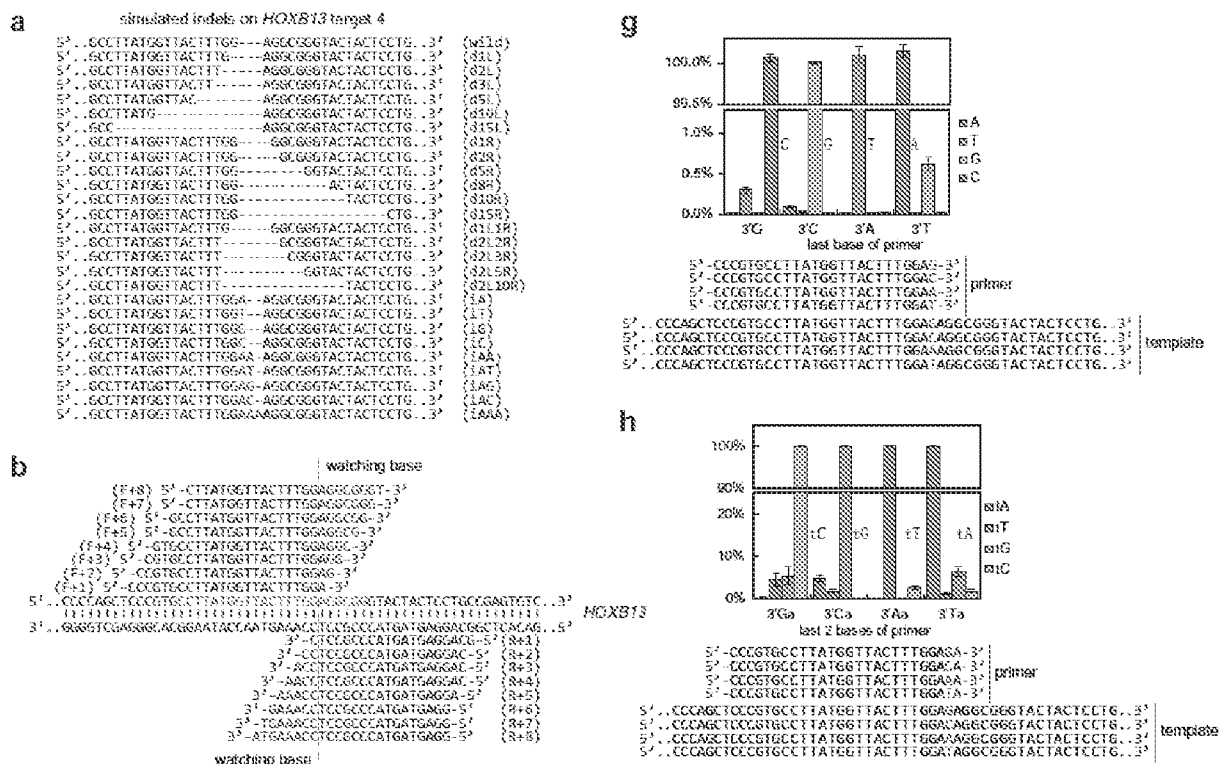


图 1



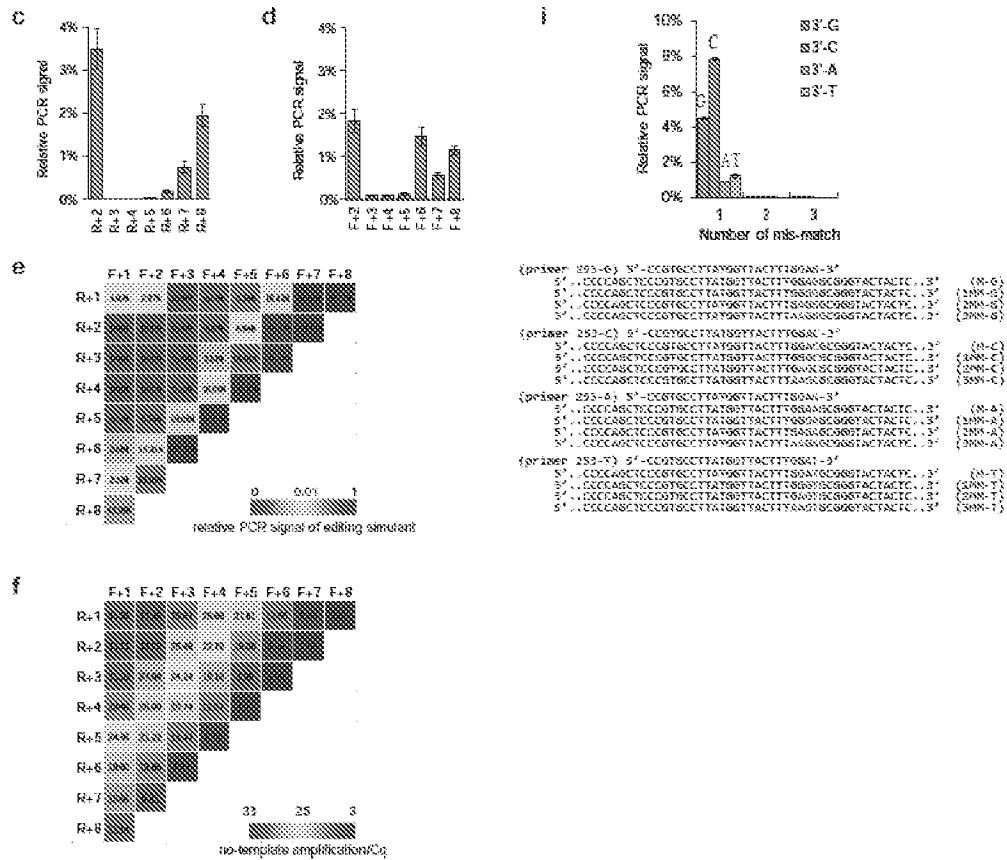
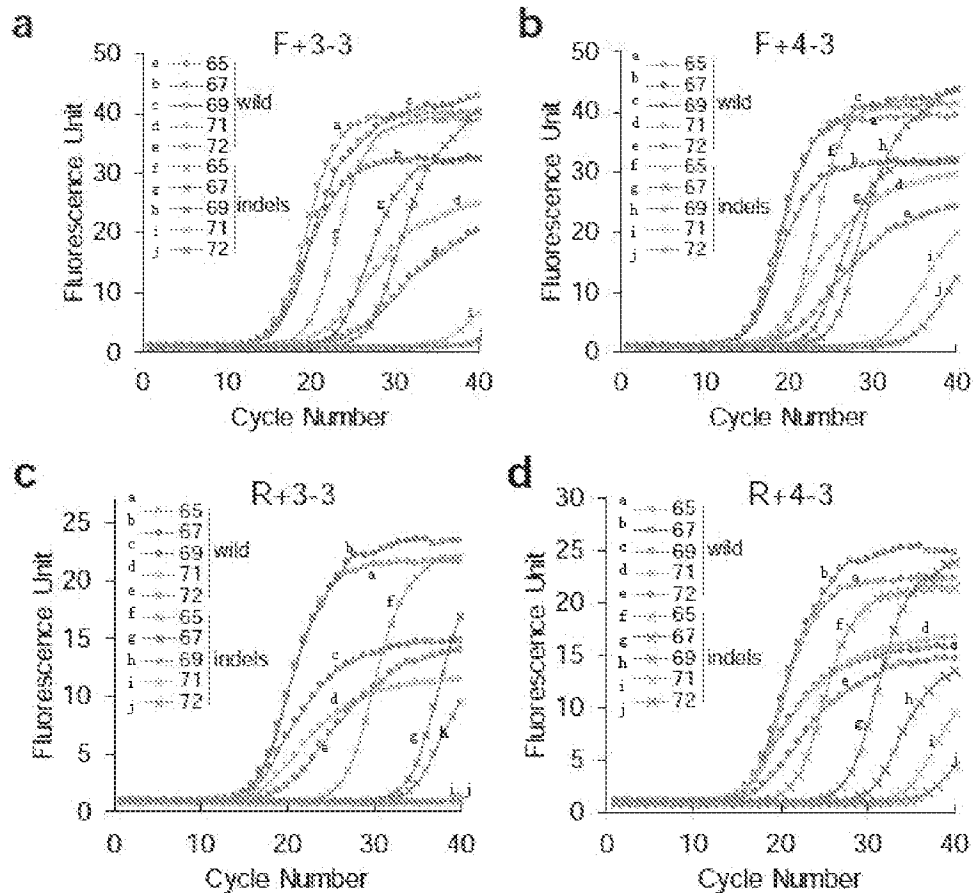


图 2



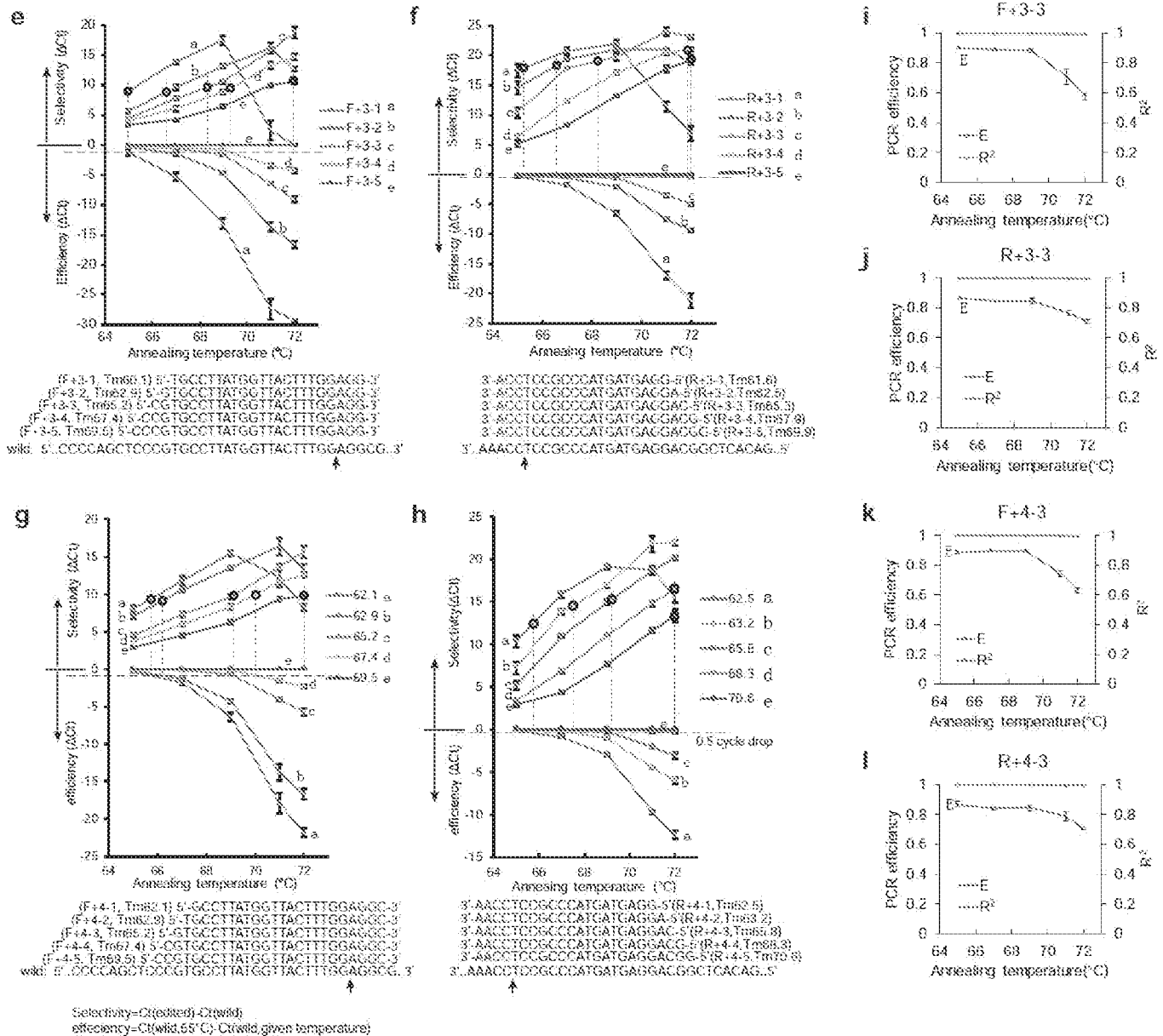
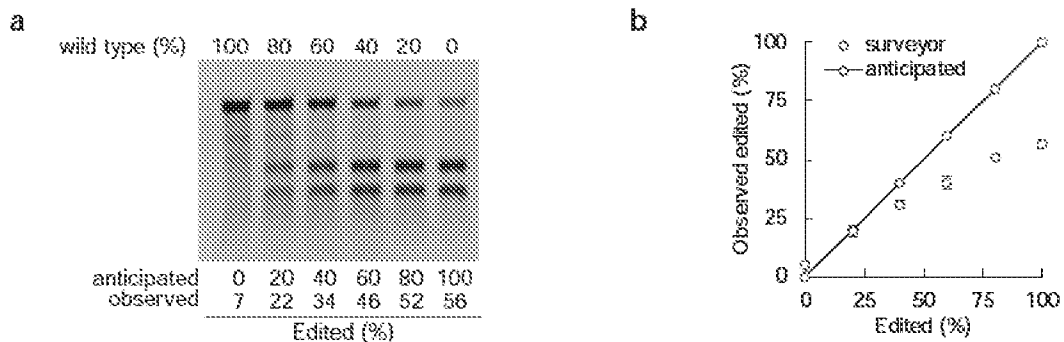


图 3



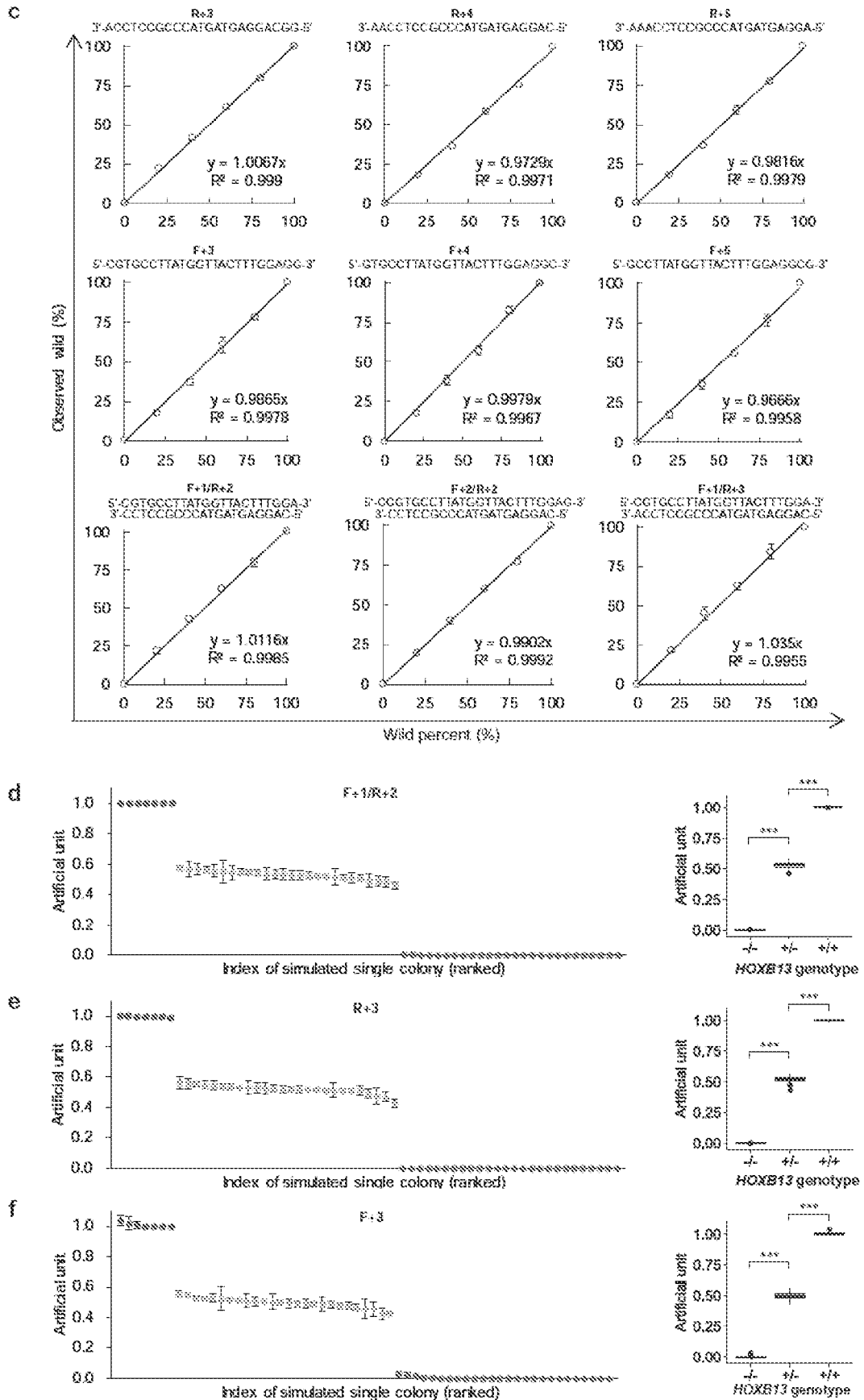
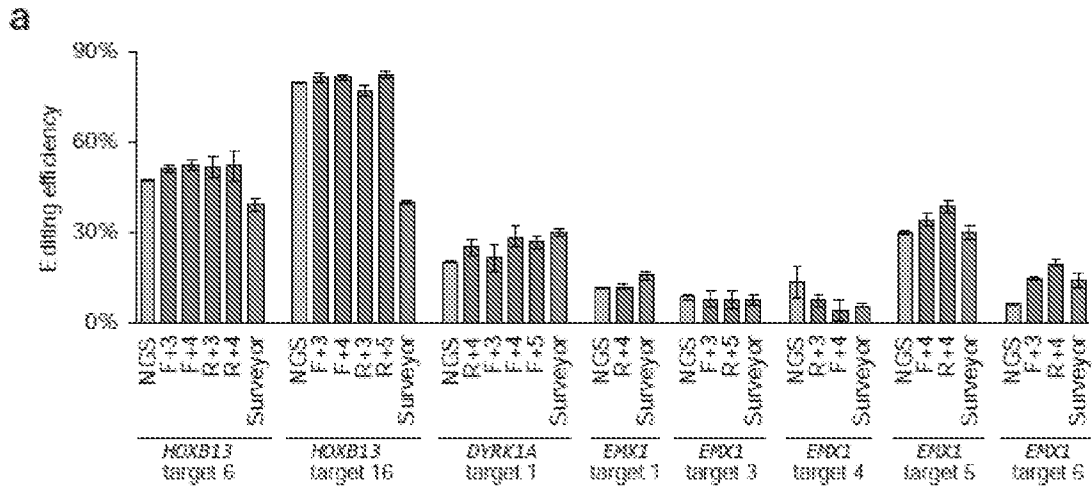


图 4



b

HOXB13T6F+3: 5'-TCCEGTGCTTATGCTTACTTTGG-3'
 HOXB13T6F+4: 5'-TCCEGTGCTTATGCTTACTTTGGA-3'
 5'...CTCCGTGCTTATGCTTACTTTGGAGCGGGTACTACTCCT...3' HOXB13 target 6
 3'-GAAACCTCGCCCATGATGAGG-5' HOXB13T6R+3
 3'-TGAAACCTCGCCCATGATGAGG-5' HOXB13T6R+4

HOXB13T16F+3: 5'-CTATCCGGGATATCCGGGAACC-3'
 HOXB13T16F+4: 5'-CTATCCGGGATATCCGGGAACCT-3'
 5'...TCTATCCGGGATATCCGGGAACCTTCAGGCTTATGSCAG...3' HOXB13 target 16
 3'-CCTTGGGATGCTGGGATACCGGT-5' HOXB13T16R+3
 3'-GCCCTTGGGATGCTGGGATACCGGT-5' HOXB13T16R+5

DYRK1AT1F+3: 5'-TGCTGCTGGCTTCAGATGGC-3'
 DYRK1AT1F+4: 5'-TGCTGCTGGCTTCAGATGGCT-3'
 DYRK1AT1F+5: 5'-CTGCTGGCTTCAGATGGCTG-3'
 5'...CATGCTGCTGGCTTCAGATGGCTGGACAGATGCCCAT...3' DYRK1A
 3'-TCTACCGACCTGCTACGGGGT-5' DYRK1AT1R+4

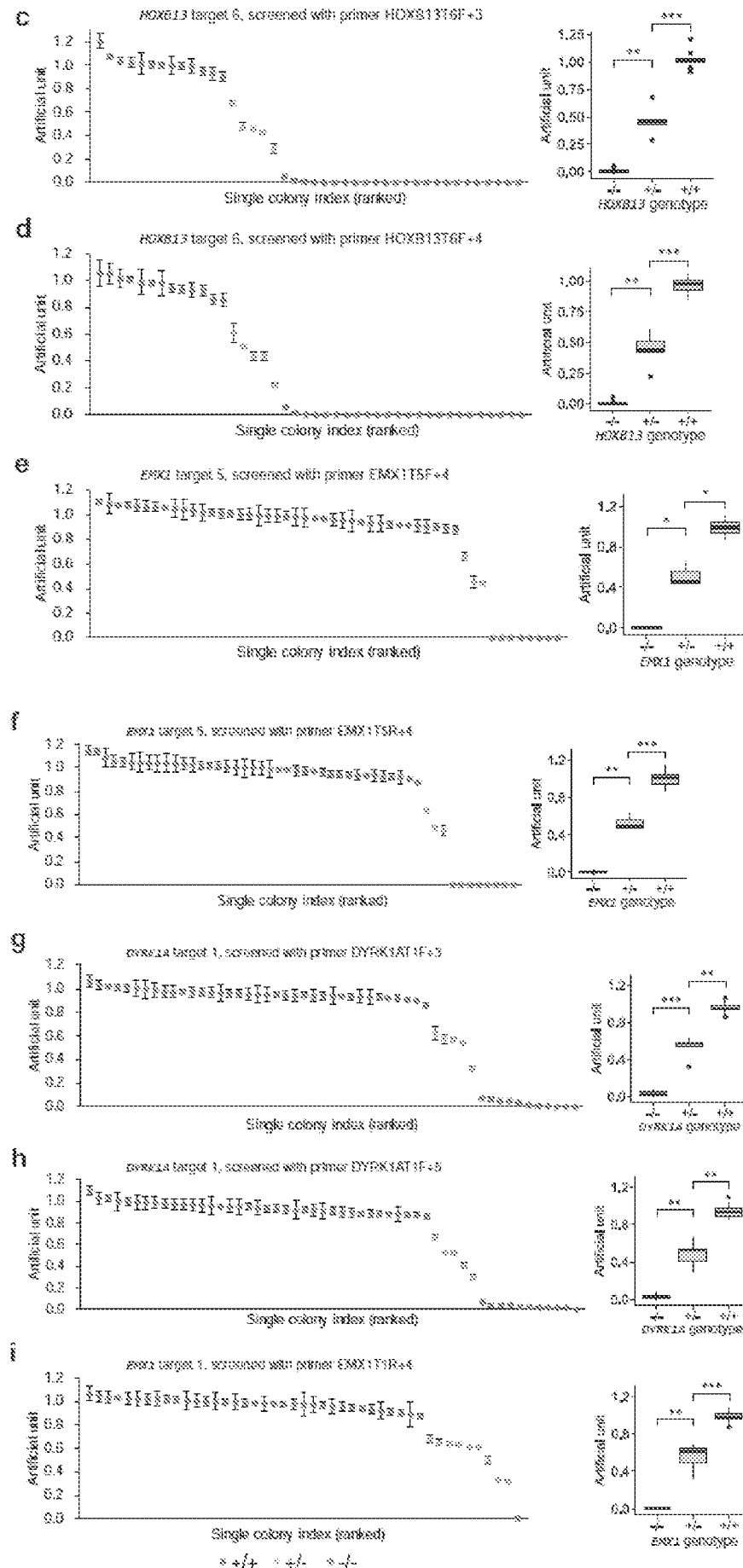
5'...CTATGTAGCTEAGTCTTCCATCAGGCTCTEAGCTCAGC...3' EMX1 target 1
 3'-AGGGTAGTCCGAGAGTCGAGTC-5' EMX1T1R+4

EMX1T3F+3: 5'-GTGGGCAACCAAAACCAACGA-3'
 5'...GGGTGGGCAACCAAAACCAACGAGAGTGGT...3' EMX1 target 3
 3'-TGGGTGGTCCCGTCTCAGGA-5' EMX1T3R+5

EMX1T4F+4: 5'-GGCAGAGTGGCTTCTGCTGCT-3'
 5'...GAGGGCAGAGTGGCTTCTGCTGCTGSCAGGCCCTG...3' EMX1 target 4
 3'-ACGACAGCCGGTCCGGGGA-5' EMX1T4R+3

EMX1T5F+4: 5'-GATGTACCTCCAATGACTAGGGT-3'
 5'...CGATGTACCTCCAATGACTAGGGTGGGCAACCAAAACCC...3' EMX1 target 5
 3'-TGATCCCAACCGTTGGTGTGTTGG-5' EMX1T5R+4

EMX1T6F+3: 5'-GCCTGAGTCCGAGCAGAAGAAGA-3'
 5'...GGCCTGAGTCCGAGCAGAAGAAGAAGGCTCCCATCAGATCAAC...3' EMX1 target 6
 3'-TCTTCTTCCCGAGGGTATGATGTT-5' EMX1T6R+4



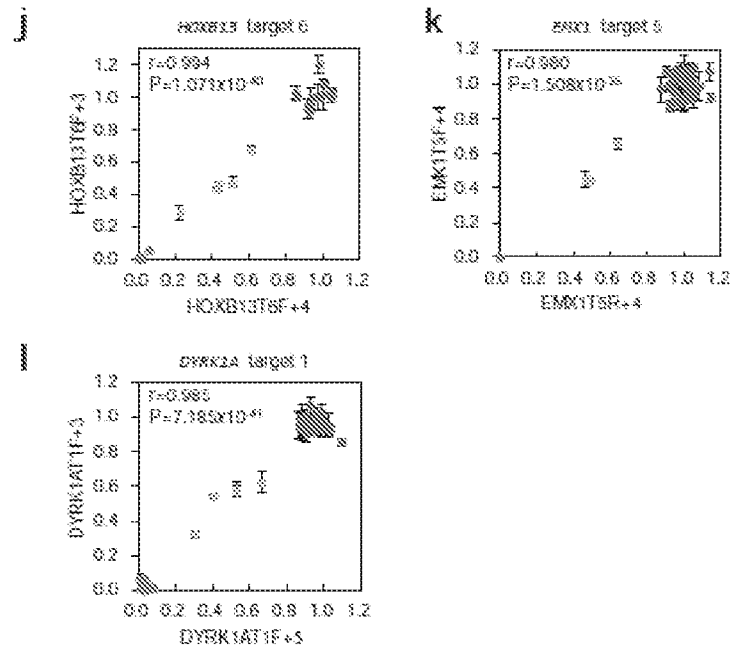
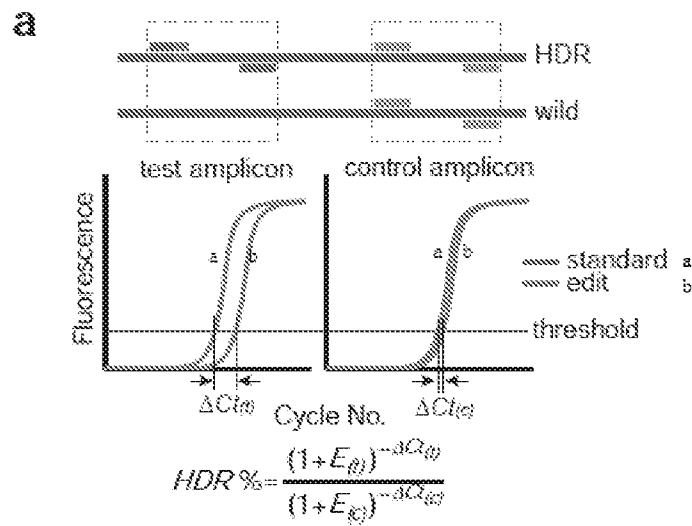


图 5



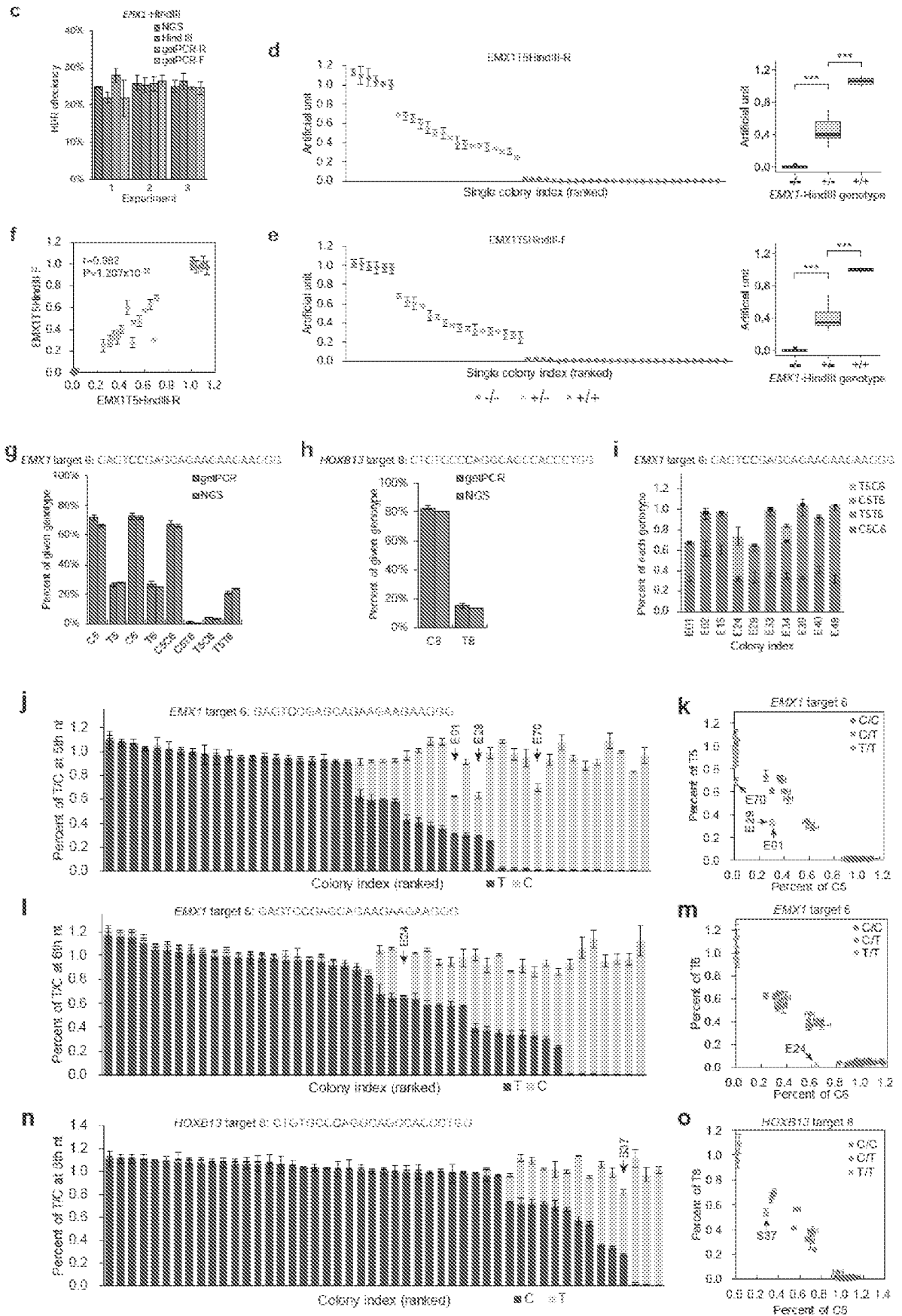


图 6

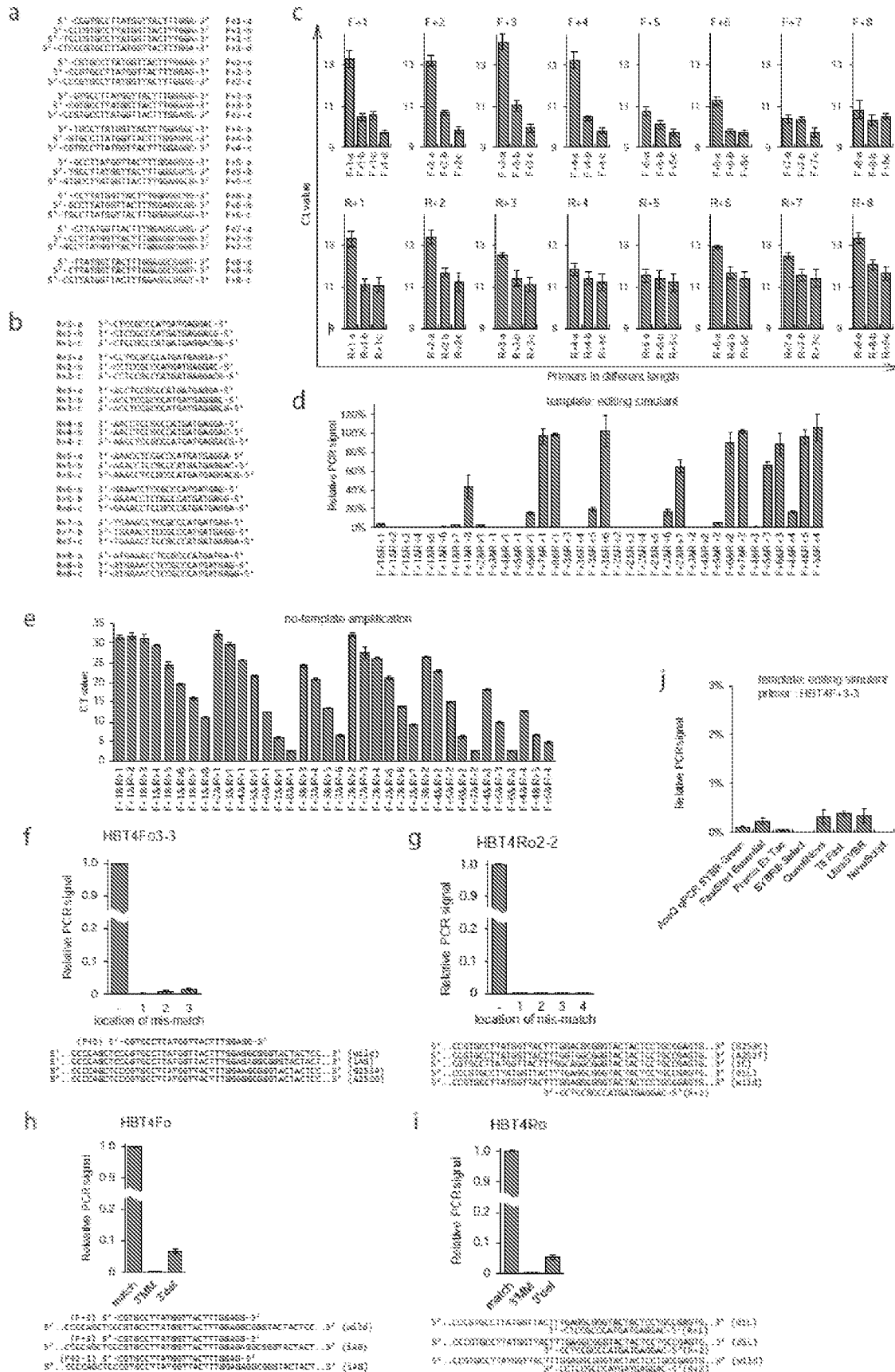
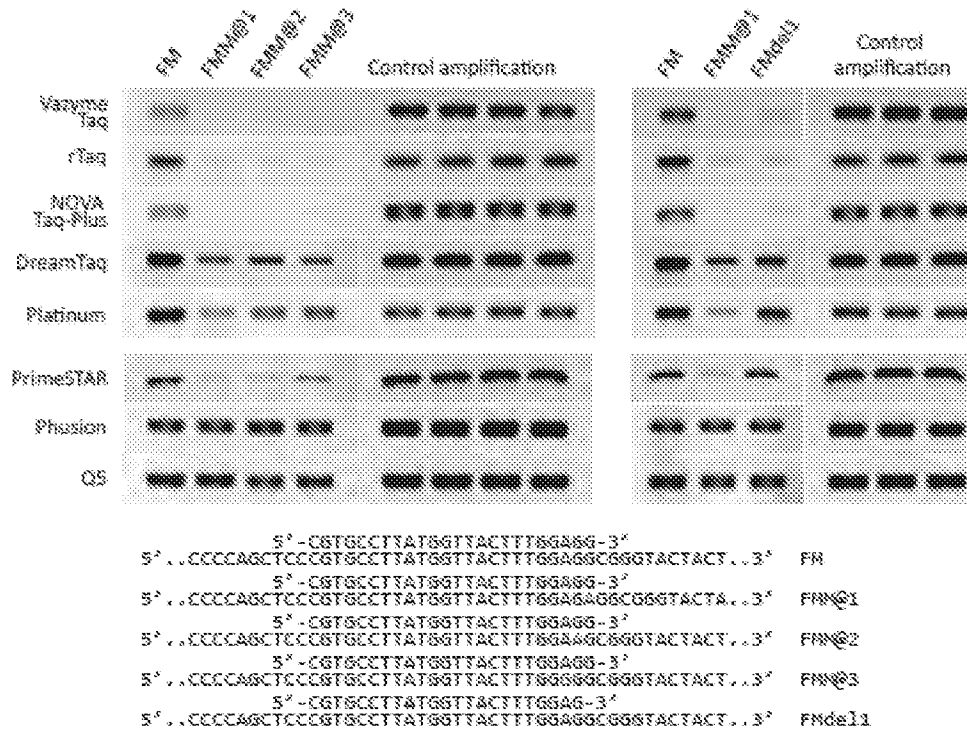
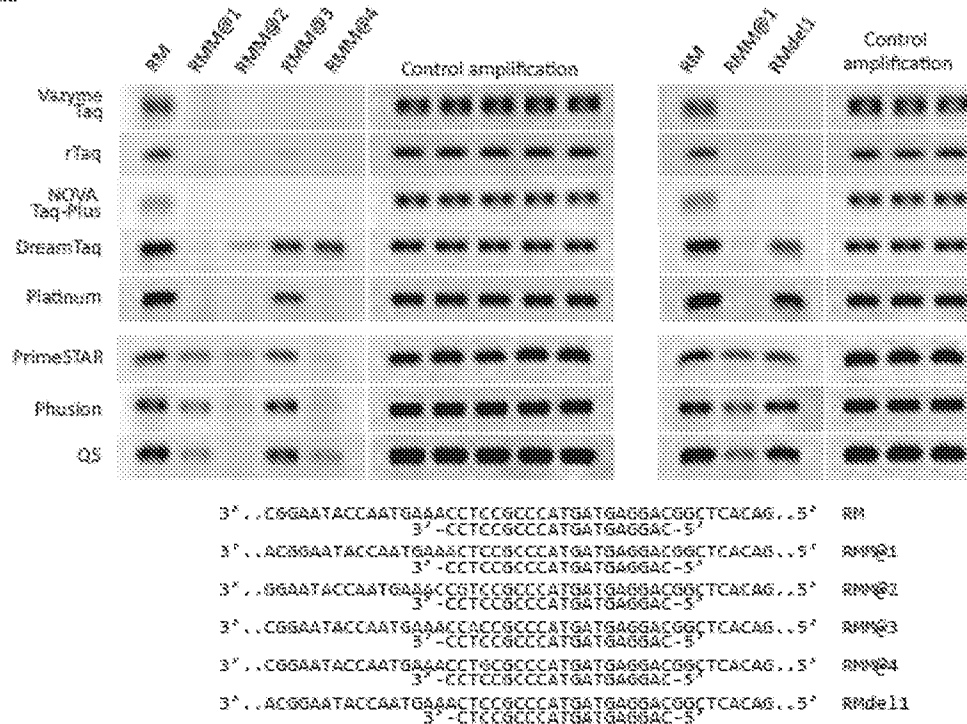


图 7

a**b**

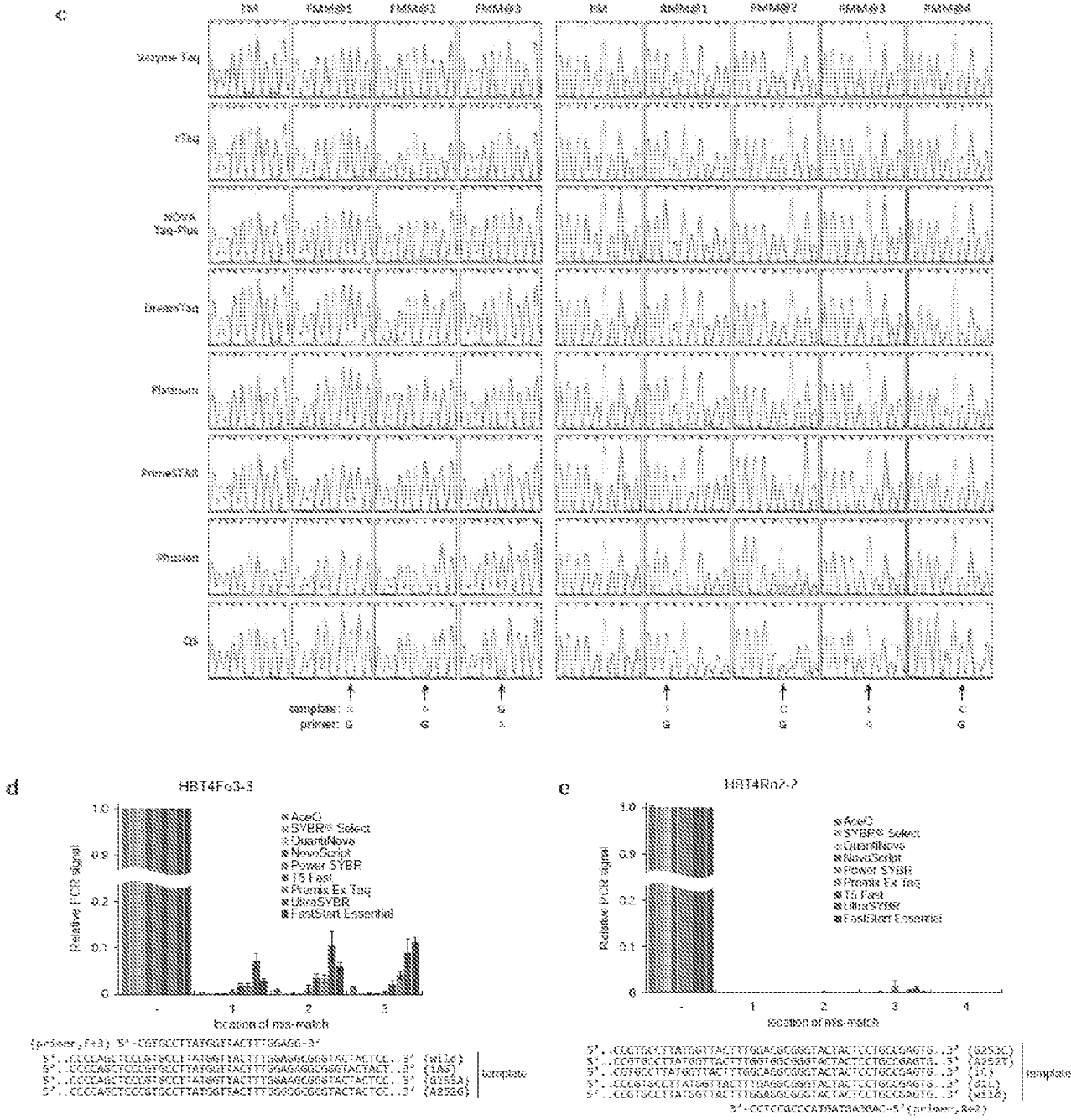


图 8

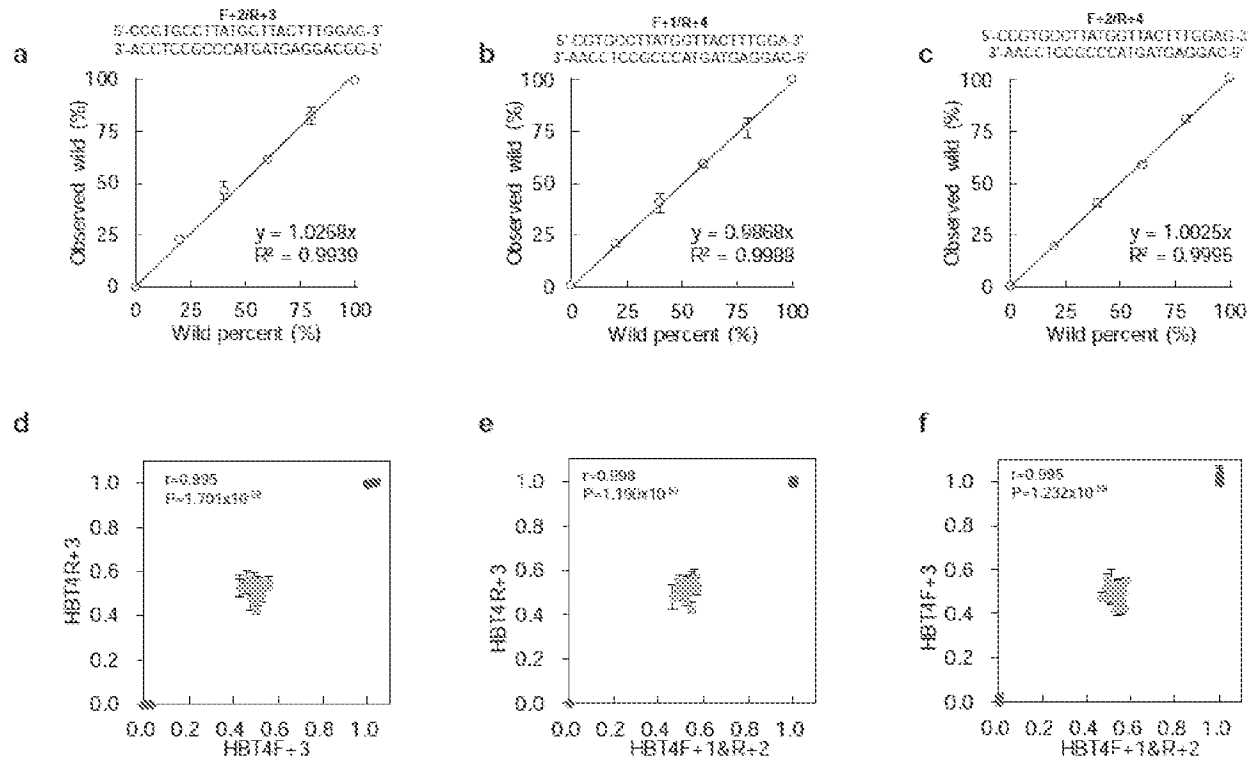
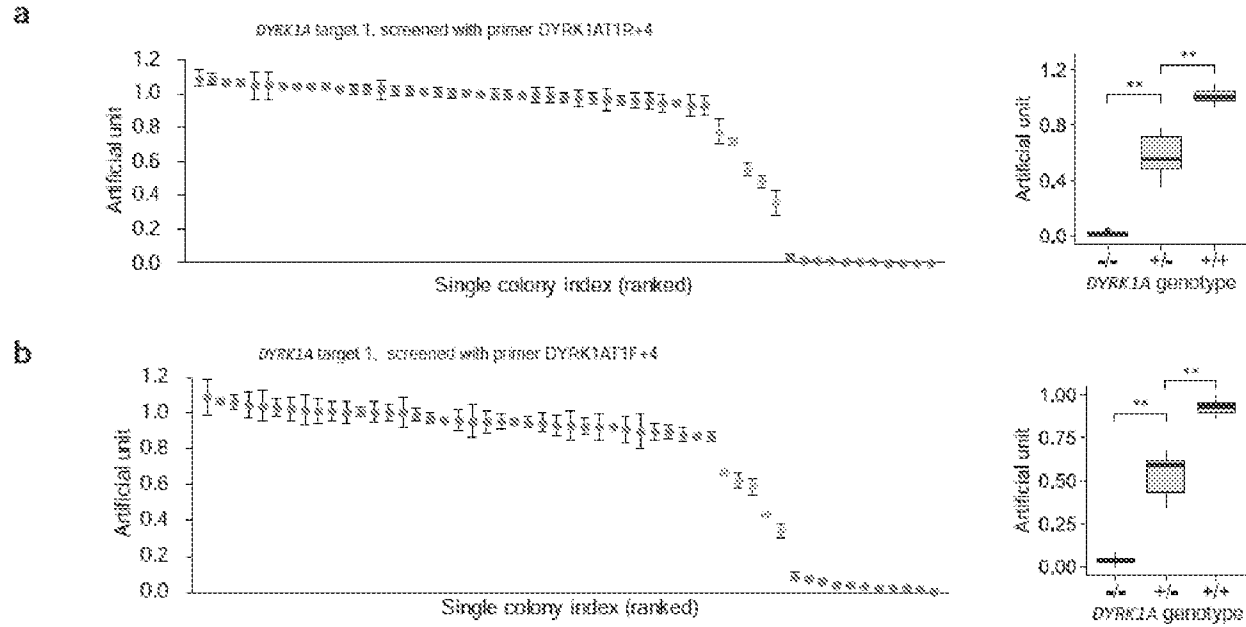


图 9



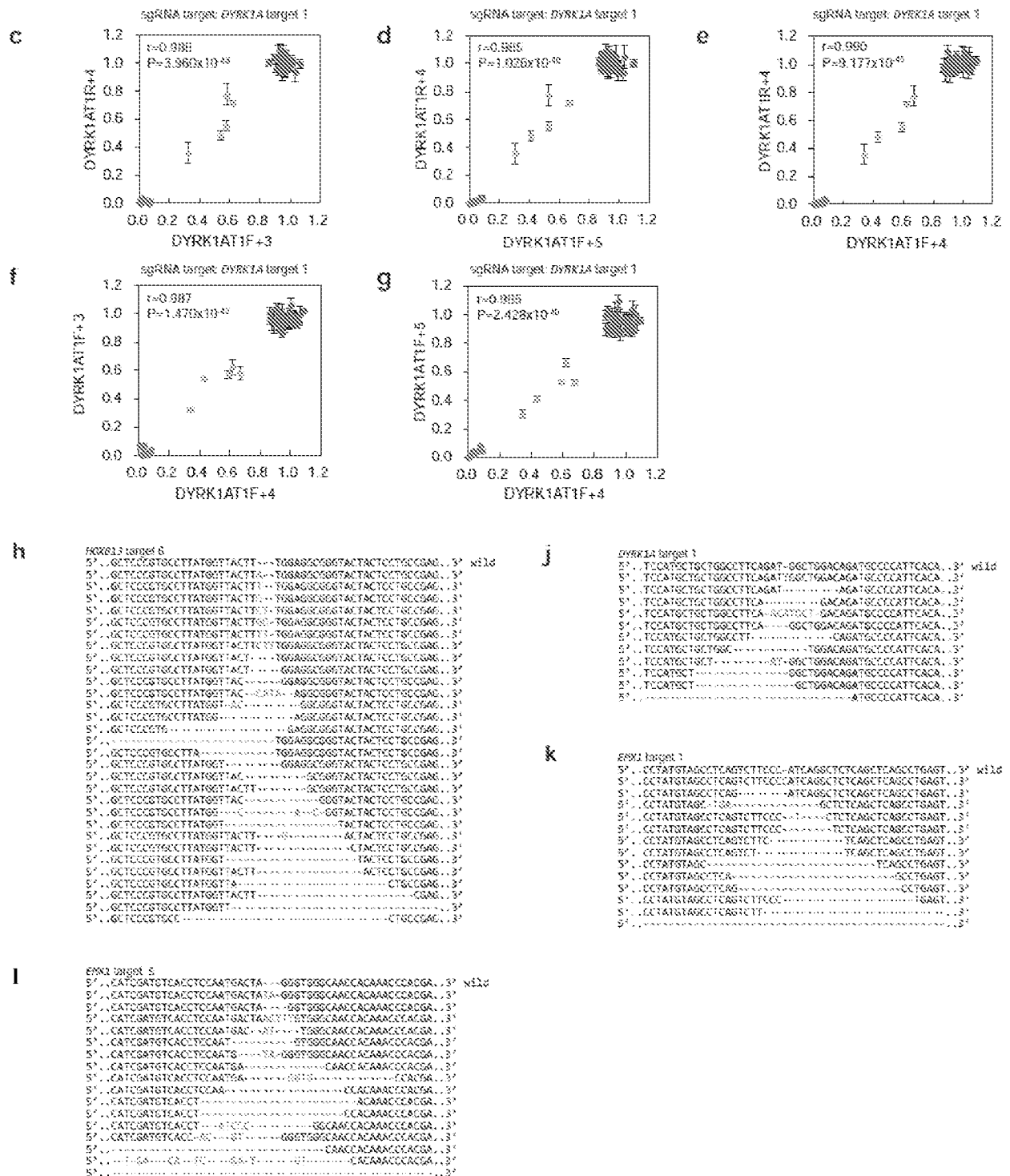
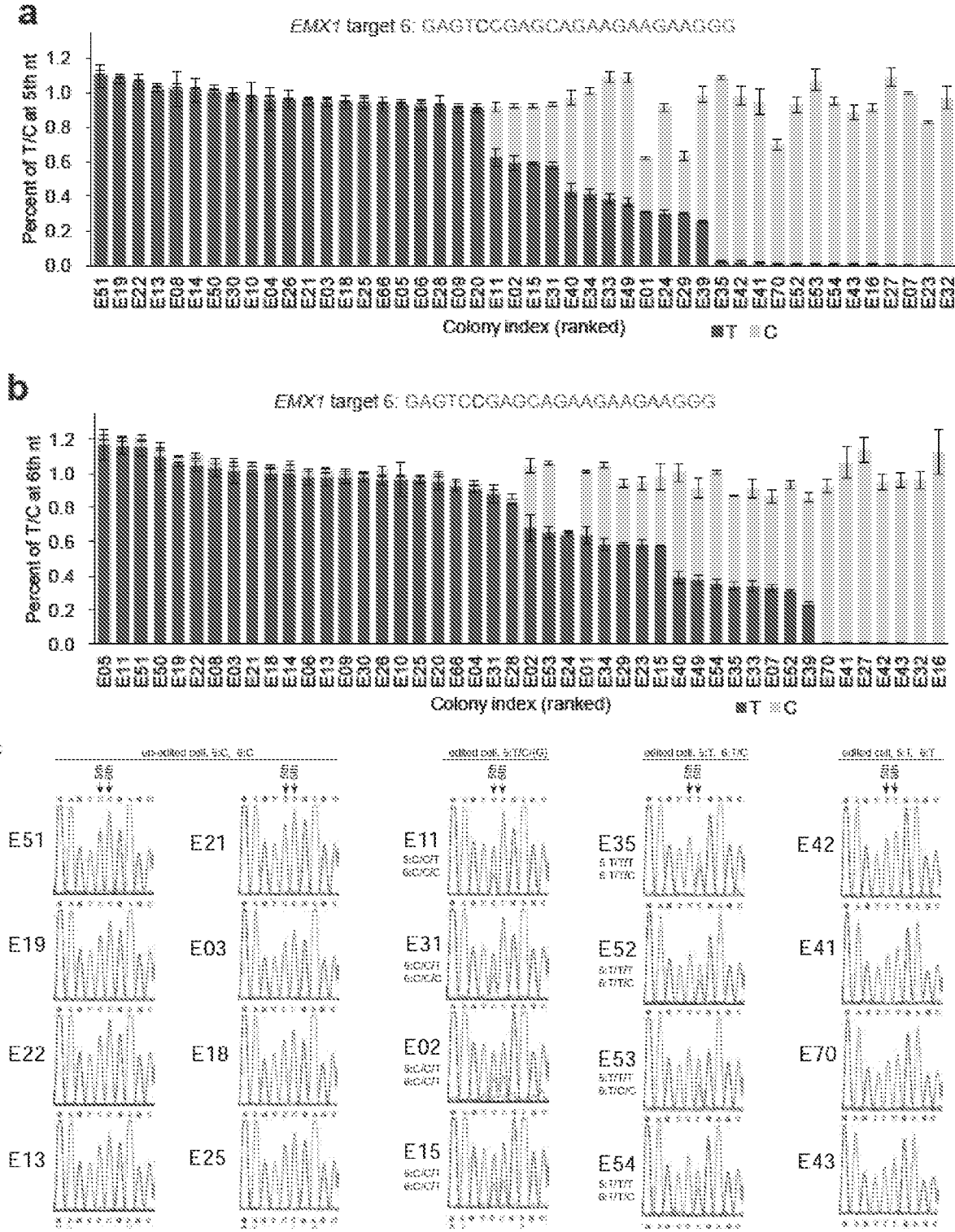


图 10



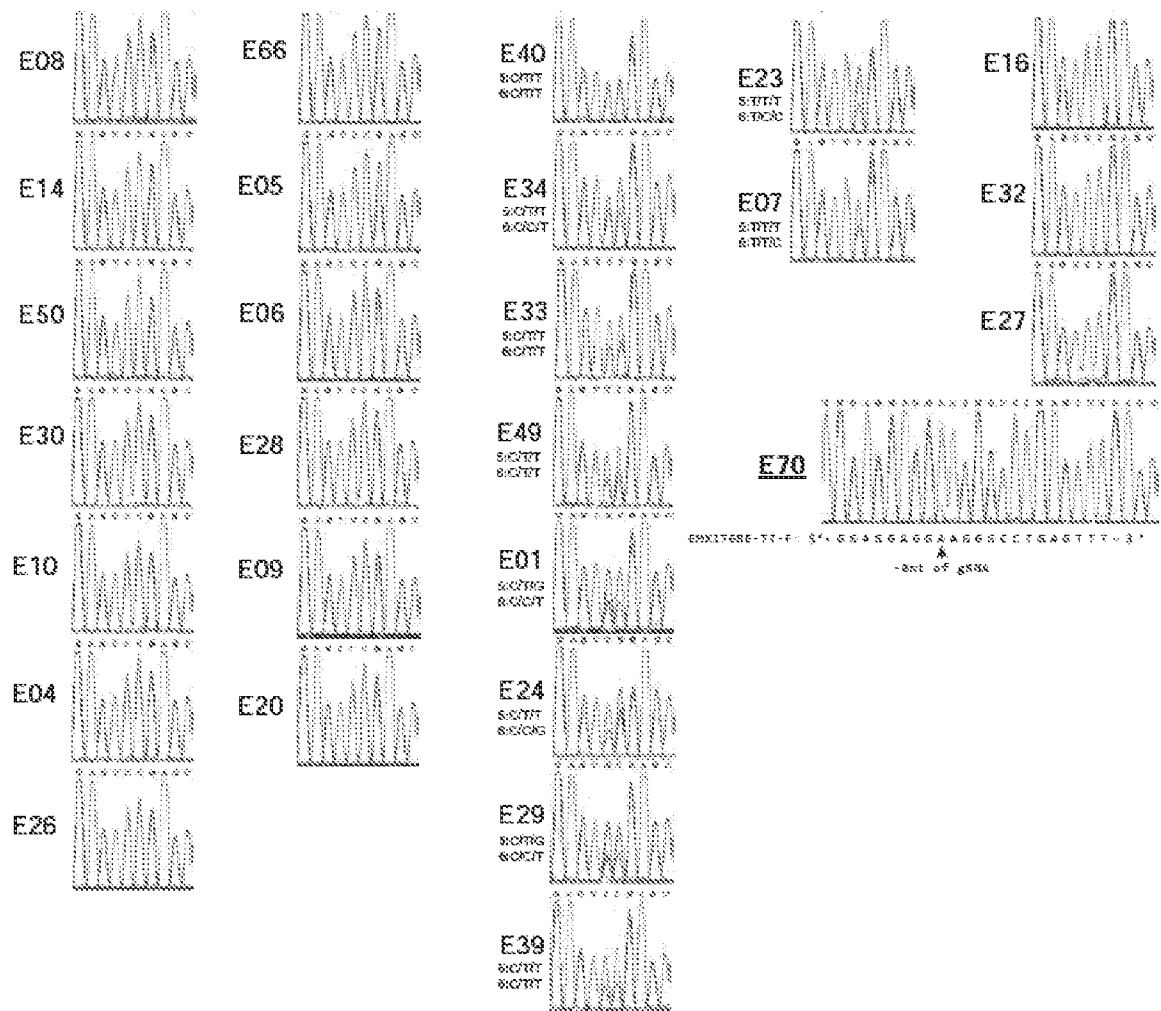
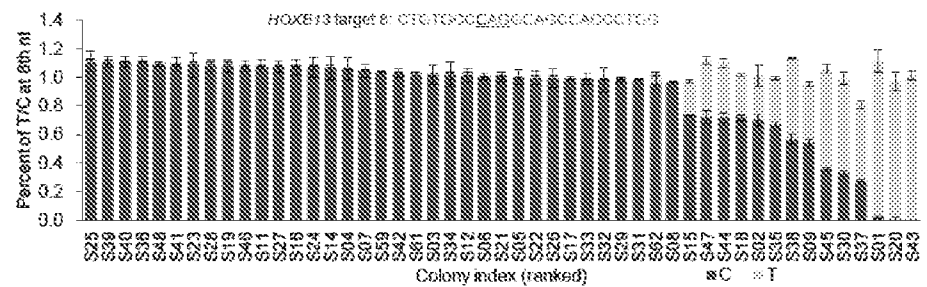


图 11

52



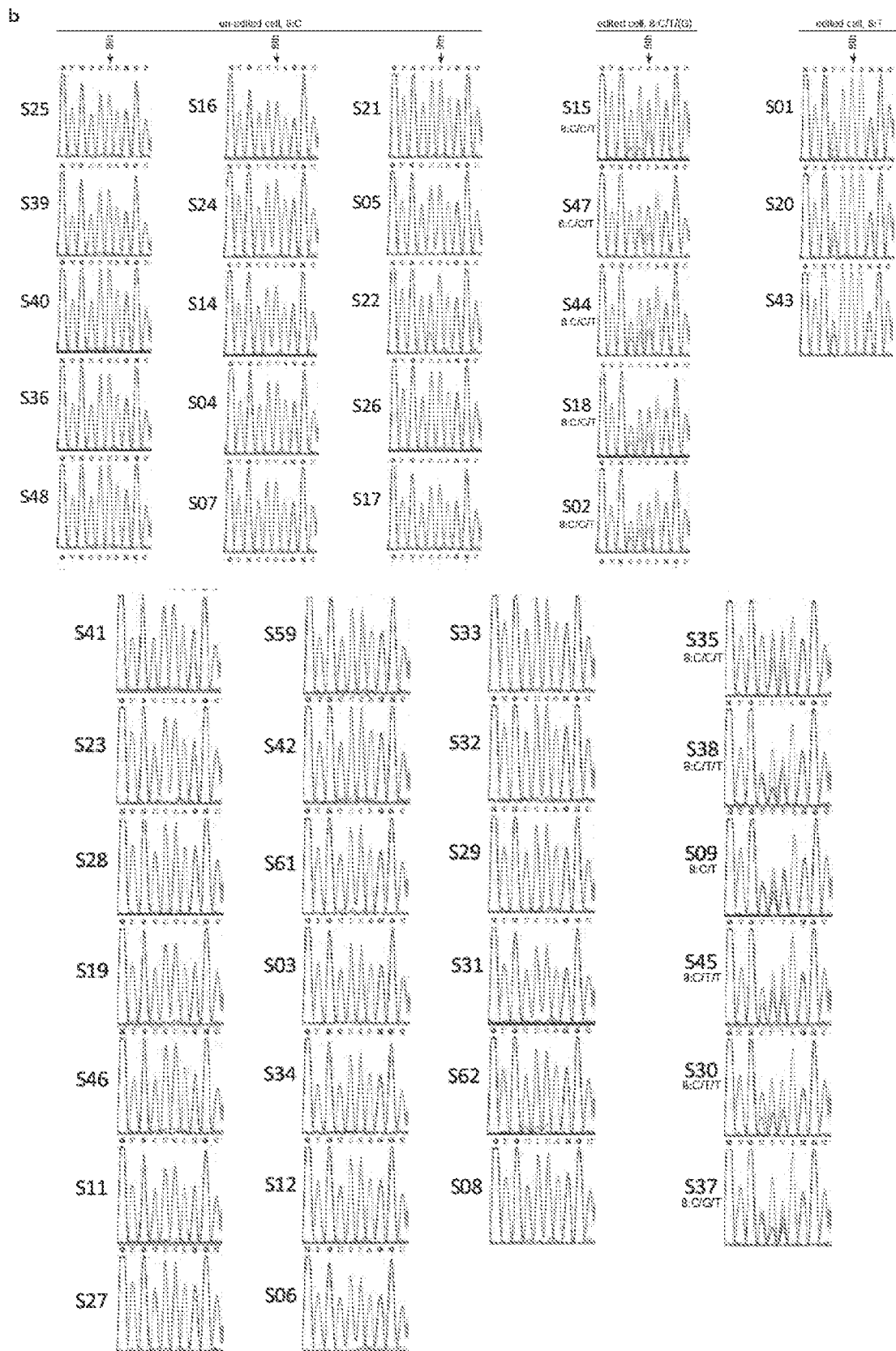


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/095927

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68(2018.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

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

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

DWPI, SIPOABS, CNABS, CNKI, ISI Web of knowledge, NCBI, Google Scholar: 插入缺失, 突变, 野生, 频率, 引物, 核酸酶, 定量, 基因编辑, indel, Mutation, wild, frequency, primer, nuclease, quantification, gene editing, Cas

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110607356 A (SHANDONG UNIVERSITY; SHANDONG DAOHUO GENE TECHNOLOGY CO., LTD.) 24 December 2019 (2019-12-24) claims	1-10
X	CN 103451311 A (AGCU SCIENTECH INCORPORATION) 18 December 2013 (2013-12-18) the claims	6
X	CN 101955996 A (NORTHWEST A&F UNIVERSITY) 26 January 2011 (2011-01-26) claims	1-10
A	CN 108728563 A (VEGETABLE RESEARCH INSTITUTE OF SHANDONG ACADEMY OF AGRICULTURAL SCIENCES) 02 November 2018 (2018-11-02) entire document	1-10
A	CN 103146823 A (NORTHWEST A&F UNIVERSITY) 12 June 2013 (2013-06-12) entire document	1-10

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

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

16 July 2020

Date of mailing of the international search report

11 September 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/095927**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed:
 - ☒ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/095927

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110607356	A	24 December 2019	None			
CN	103451311	A	18 December 2013	CN	103451311	B	20 April 2016
CN	101955996	A	26 January 2011	CN	101955996	B	23 May 2012
CN	108728563	A	02 November 2018	None			
CN	103146823	A	12 June 2013	None			

国际检索报告

国际申请号

PCT/CN2020/095927

A. 主题的分类 C12Q 1/68(2018.01) i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																				
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) C12Q 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) DWPI、SIP0ABS、CNABS、CNKI、ISI Web of knowledge、NCBI、Google Scholar: 插入缺失、突变、野生、频率、引物、核酸酶、定量、基因编辑、indel、Mutation、wild、frequency、primer、nuclease、quantification、gene editing、Cas																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110607356 A (山东大学 山东刀火基因科技有限公司) 2019年 12月 24日 (2019 - 12 - 24) 权利要求书</td> <td>1-10</td> </tr> <tr> <td>X</td> <td>CN 103451311 A (无锡中德美联生物技术有限公司) 2013年 12月 18日 (2013 - 12 - 18) 权利要求书</td> <td>6</td> </tr> <tr> <td>X</td> <td>CN 101955996 A (西北农林科技大学) 2011年 1月 26日 (2011 - 01 - 26) 权利要求书</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 108728563 A (山东省农业科学院蔬菜花卉研究所) 2018年 11月 2日 (2018 - 11 - 02) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 103146823 A (西北农林科技大学) 2013年 6月 12日 (2013 - 06 - 12) 全文</td> <td>1-10</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110607356 A (山东大学 山东刀火基因科技有限公司) 2019年 12月 24日 (2019 - 12 - 24) 权利要求书	1-10	X	CN 103451311 A (无锡中德美联生物技术有限公司) 2013年 12月 18日 (2013 - 12 - 18) 权利要求书	6	X	CN 101955996 A (西北农林科技大学) 2011年 1月 26日 (2011 - 01 - 26) 权利要求书	1-10	A	CN 108728563 A (山东省农业科学院蔬菜花卉研究所) 2018年 11月 2日 (2018 - 11 - 02) 全文	1-10	A	CN 103146823 A (西北农林科技大学) 2013年 6月 12日 (2013 - 06 - 12) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 110607356 A (山东大学 山东刀火基因科技有限公司) 2019年 12月 24日 (2019 - 12 - 24) 权利要求书	1-10																		
X	CN 103451311 A (无锡中德美联生物技术有限公司) 2013年 12月 18日 (2013 - 12 - 18) 权利要求书	6																		
X	CN 101955996 A (西北农林科技大学) 2011年 1月 26日 (2011 - 01 - 26) 权利要求书	1-10																		
A	CN 108728563 A (山东省农业科学院蔬菜花卉研究所) 2018年 11月 2日 (2018 - 11 - 02) 全文	1-10																		
A	CN 103146823 A (西北农林科技大学) 2013年 6月 12日 (2013 - 06 - 12) 全文	1-10																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
国际检索实际完成的日期	国际检索报告邮寄日期																			
2020年 7月 16日	2020年 9月 11日																			
ISA/CN的名称和邮寄地址	授权官员																			
中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	张起																			
传真号 (86-10)62019451	电话号码 62411036																			

第I栏

核苷酸和/或氨基酸序列 (续第1页第1. c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列, 国际检索是基于下列序列表进行的:

a. ☒ 作为国际申请的一部分提交的:

☒ 附件C/ST. 25文本文件形式

☐ 纸件或图形文件形式

b. ☐ 根据细则13之三. 1(a) 仅为国际检索目的以附件C/ST. 25文本文件形式与国际申请同时提交的:

c. ☐ 仅为国际检索目的在国际申请日之后提交的:

☐ 附件C/ST. 25文本文件形式 (细则13之三. 1(a))

☐ 纸件或图形文件形式 (细则13之三. 1(b) 和行政规程第713段)

2. ☐ 另外, 在提交/提供了多个版本或副本的序列表的情况下, 提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列表的信息相同或未超出申请时提交的申请中的信息范围 (如适用) 的所需声明。

3. 补充意见:

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/095927

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	110607356	A	2019年 12月 24日	无	
CN	103451311	A	2013年 12月 18日	CN 103451311	B 2016年 4月 20日
CN	101955996	A	2011年 1月 26日	CN 101955996	B 2012年 5月 23日
CN	108728563	A	2018年 11月 2日	无	
CN	103146823	A	2013年 6月 12日	无	