



(12) 发明专利

(10) 授权公告号 CN 113436683 B

(45) 授权公告日 2024. 08. 16

(21) 申请号 202010207885.8

C12N 15/11 (2006.01)

(22) 申请日 2020.03.23

C12N 15/13 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113436683 A

(56) 对比文件

CN 104204220 A, 2014.12.10

CN 105886616 A, 2016.08.24

(43) 申请公布日 2021.09.24

审查员 邹琴

(73) 专利权人 北京合生基因科技有限公司

地址 102206 北京市昌平区中关村生命科

学园创新大厦B203室

(72) 发明人 黄慧雅 曹玉冰 刘乙齐 郭亚琨

(74) 专利代理机构 北京清亦华知识产权代理事

务所(普通合伙) 11201

专利代理师 肖阳

(51) Int. Cl.

G16B 30/00 (2019.01)

G16B 40/00 (2019.01)

权利要求书1页 说明书15页

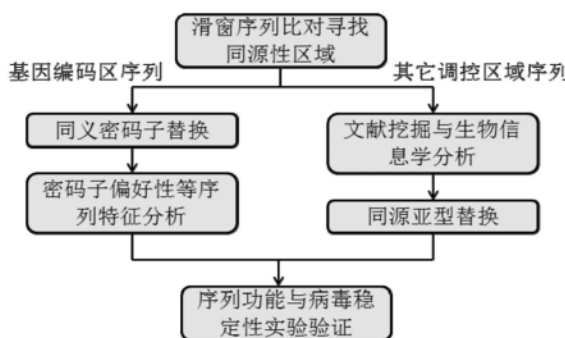
序列表13页 附图6页

(54) 发明名称

筛选候选插入片段的方法和系统

(57) 摘要

本发明提出了一种筛选候选插入片段的方法,所述插入片段用于改造待测样本核酸序列。该方法包括:基于用于改造待测样本核酸序列的候选基因,确定初始基因编码区序列;对所述初始基因编码区序列进行同义密码子替换,以便获得由多个候选编码区序列构成的候选编码区序列集合;将所述候选编码区序列与自身、宿主基因组和病毒基因组进行比对,并确定同源性高风险区域,以便对所述候选编码区序列进行同源性风险评分;和基于所述同源性风险评分,在所述候选编码区序列集合中确定优选编码区序列子集。



1. 一组UAS组合,其特征在于,包括一系列UAS序列,  
所述UAS序列具有SEQ ID NO:4~6以及17所示的核苷酸序列;或  
SEQ ID NO:1、23、31以及34所示的核苷酸序列;或  
SEQ ID NO:1、4、5、6以及14所示的核苷酸序列;或  
SEQ ID NO:17、23、25、31以及34所示的核苷酸序列;  
其中,所述UAS组合从5'端到3'端具有如下所示的UAS连接顺序:  
B1-A5-A6-A4-D5-A1-D2-C1、A5-A1-A6-A4-A14或B1-D5-D2-C1-C3。
2. 根据权利要求1所述的UAS组合,其特征在于,所述UAS组合具有SEQ ID NO:39、41、42  
或44所示的核苷酸序列。

## 筛选候选插入片段的方法和系统

### 技术领域

[0001] 本发明涉及生物信息领域,具体地,本发明涉及筛选候选插入片段的方法和系统。

### 背景技术

[0002] 病毒往往具有稳定的结构、简单的基因组、广谱的侵染能力与高效的包装能力而成为广泛使用的工程化DNA转运表达载体。而利用病毒自身的免疫特性,研究者使用灭活、减毒或工程改造的病毒作为有效的疫苗。更进一步,利用病毒在扩增过程中裂解宿主细胞的生物特性,研究者将病毒工程改造为具有复制包装能力并特异性实现肿瘤杀伤作用的溶瘤病毒。随着病毒相关研究的逐渐深入,目前已有多种工程病毒种类成为工程改造的对象,并有多种病毒产品应用于临床治疗。腺病毒与其它病毒类型相比,由于其相对稳定的结构、约9000bp的包装容量、较低的毒副作用而备受青睐,不仅可以作为DNA转运表达的非复制型载体,也可以作为溶瘤杀伤的复制型载体。

[0003] 虽然腺病毒作为工程载体具有上述优势,但即使是相对稳定的腺病毒亚型,如常用的5型腺病毒,在大规模药学生产过程中,随着培养代数增加,单核苷酸变异、小片段插入删除变异与大片段结构性变异的发生与积累也会逐渐增加。而另一方面,工程改造过程中引入的人工外源序列可能与自身及病毒基因组存在高度同源的区域,提升了由同源重组导致结构性变异的风险。这种同源性风险进一步影响了工程腺病毒基因组的稳定性。上述变异不仅影响了工程腺病毒药学生产的产量和纯度,也有可能影响工程腺病毒的功能与安全性。比如,非复制型工程腺病毒的药学生产过程中产生复制型腺病毒、受控复制型工程腺病毒的药学生产过程中产生不受控复制型腺病毒都是质量控制关注的重点。

### 发明内容

[0004] 本申请是基于发明人对以下事实和问题的发现和认识作出的:

[0005] 工程病毒的药学生产过程中,人工外源序列自身与病毒基因组可能存在高度同源的区域,提升了由同源重组导致结构性变异的风险。发明人通过对人工外源序列与病毒基因组序列特征进行信息学分析,预测分析存在同源性风险的区域,并结合同义密码子替换与同源亚型替换等方式降低序列的同源性风险,增加工程病毒基因组的稳定性。

[0006] 为此,在本发明的第一方面,本发明提出了一种筛选候选插入片段的方法,所述插入片段用于改造待测样本核酸序列。根据本发明的实施例,所述方法包括:(1)基于用于改造待测样本核酸序列的候选基因,确定初始基因编码区序列;(2)对所述初始基因编码区序列进行同义密码子替换,以便获得由多个候选编码区序列构成的候选编码区序列集合;(3)将所述候选编码区序列与宿主基因组和病毒基因组进行比对,并确定同源性高风险区域,以便对所述候选编码区序列进行同源性风险评分;(4)基于所述同源性风险评分,在所述候选编码区序列集合中确定优选编码区序列子集。

[0007] 根据本发明的实施例,上述方法还可以进一步包括如下附加技术特征至少之一:

[0008] 根据本发明的实施例,所述方法进一步包括:

[0009] 基于所述宿主中的密码子频率,对所述候选编码区序列进行密码子对偏好性评分;或基于所述宿主中的密码子频率,对所述候选编码区序列进行密码子偏好性评分;或基于所述候选编码区序列中的C碱基频率、G碱基频率和CpG序列频率,对所述候选编码区序列进行CpG评分;或基于所述候选编码区序列中的A碱基频率、T碱基频率和TpA序列频率,对所述候选编码区序列进行TpA评分;或基于所述候选编码区序列的编码mRNA序列,预测RNA二级结构的最小自由能;或基于所述候选编码区序列中的微卫星序列,对所述候选编码区序列进行微卫星不稳定性评分。

[0010] 根据本发明的实施例,步骤(4)进一步包括:基于所述同源性风险评分以及包括下列评分的至少之一,在所述候选编码区序列集合中确定优选编码区序列子集,

[0011] 所述密码子对偏好性评分、所述密码子偏好性评分、所述CpG评分、所述TpA评分、所述mRNA二级结构最小自由能和所述微卫星不稳定性评分。

[0012] 其中,同源性风险评分、密码子对偏好性评分、密码子偏好性评分、CpG评分、TpA评分、mRNA二级结构最小自由能与微卫星不稳定性评分优先级逐级降低。

[0013] 根据本发明实施例的上述方法可以获得基因编码区序列的候选插入片段(在本文实施例中称同源序列),该候选插入片段可以有效降低由同源重组导致的结构性变异的风险,提高待测样本核酸序列(如病毒基因组)的稳定性。

[0014] 根据本发明的实施例,采用滑窗序列比对法,所述滑窗的长度为12~17bp。

[0015] 根据本发明的实施例,在步骤(3)中,所述同源性风险评分是基于如下方式确定的:

[0016] 通过滑窗序列比对,统计错配数分别为0~4 bp、参考序列分别为自身或病毒基因组或宿主基因组序列时的最长一致序列长度与频数,按优先级由高至低依次比较错配数、参考序列与最长一致序列,按优先级由高至低比较错配数0~4 bp,按优先级由高至低比较参考序列为自身、病毒基因组与宿主基因组序列,按优先级由高至低比较最长一致序列长度与频数,比较不同候选编码区序列的同源性风险,优先级越高、数值越大时同源性风险越高;

[0017] 根据本发明的实施例,所述密码子对偏好性评分是基于下列公式确定的:

$$[0018] \quad \text{密码子对偏好性评分} = \sum \ln \frac{\text{密码子对频率} \cdot \text{各氨基酸频率乘积}}{\text{氨基酸对频率} \cdot \text{各密码子频率乘积}}$$

[0019] 根据本发明的实施例,所述密码子偏好性评分是基于下列公式确定的:

$$[0020] \quad \text{密码子偏好性评分} = \sum \ln \frac{\text{密码子频率}}{\text{氨基酸频率}}$$

[0021] 根据本发明的实施例,所述CpG评分是基于下列公式确定的:

$$[0022] \quad \text{CpG 评分} = \frac{\text{CpG 频率}}{\text{C 碱基频率} \cdot \text{G 碱基频率}}$$

[0023] 根据本发明的实施例,所述TpA评分是基于下列公式确定的:

$$[0024] \quad \text{TpA 评分} = \frac{\text{TpA 频率}}{\text{T 碱基频率} \cdot \text{A 碱基频率}}$$

[0025] 根据本发明的实施例,所述RNA二级结构最小自由能预测基于软件Mfold或

ViennaRNA。

[0026] 根据本发明的实施例,所述微卫星不稳定性评分是基于如下方式确定的:

[0027] 对于连续重复至少3次、微卫星序列长度小于同源性风险评分滑窗长度的微卫星区域,统计最长微卫星区域长度与频数,按优先级由高至低比较最长微卫星区域长度与频数,优先级越高、数值越大时微卫星不稳定性越高。

[0028] 根据本发明的实施例,在步骤(1)中,进一步确定初始非编码区序列,并且进一步包括:(5)获取所述初始非编码区序列的第一同源亚型序列;(6)基于所述初始非编码区序列和所述第一同源亚型序列,分别确定各自的保守性区域和非保守性区域;(7)通过对所述非保守性区域进行随机突变和截短处理的至少之一,获得由多个候选非编码区序列构成的候选非编码区序列集合;(8)对所述非编码区序列进行所述同源性风险评分以及任选地微卫星不稳定性评分,在所述候选非编码区序列集合中确定优选非编码区序列子集。

[0029] 在本发明的第二方面,本发明提出了一种筛选候选插入片段的方法,所述插入片段用于改造待测样本核酸序列。根据本发明的实施例,所述方法包括:

[0030] (a) 基于用于改造待测样本核酸序列的候选基因,确定初始非编码区序列;

[0031] (b) 获取所述初始非编码区序列的第一同源亚型序列;

[0032] (c) 基于所述初始非编码区序列和所述第一同源亚型序列,分别确定各自的保守性区域和非保守性区域;

[0033] (d) 通过对所述非保守性区域进行随机突变和截短处理的至少之一,获得由多个候选非编码区序列构成的候选非编码区序列集合;

[0034] (e) 对所述非编码区序列进行所述同源性风险评分以及任选地所述微卫星不稳定性评分;

[0035] (f) 基于所述同源性风险评分以及任选地所述微卫星不稳定性评分,在所述候选非编码区序列集合中确定优选非编码区序列子集。

[0036] 根据本发明实施例的上述方法可以获得基因非编码区序列的候选插入片段(在本文实施例中称同源序列),该候选插入片段可以有效降低由同源重组导致的结构性变异的风险,提高待测样本核酸序列(如病毒基因组)的稳定性。

[0037] 根据本发明的实施例,所述改造待测样本核酸序列包括病毒基因组序列。

[0038] 在本发明的第三方面,本发明提出了一种计算机可读存储介质,其上存储有计算机程序。根据本发明的实施例,该程序被处理器执行时实现前面所述的筛选候选插入片段的方法。

[0039] 在本发明的第四方面,本发明提出了一种电子设备。根据本发明的实施例,所述电子设备包括存储器、处理器;其中,所述处理器通过读取所述存储器中存储的可执行程序代码来运行与所述可执行程序代码对应的程序,以用于实现前面所述的筛选候选插入片段的方法。

[0040] 在本发明的第五方面,本发明提出了一种筛选候选插入片的系统,所述插入片用于改造待测样本核酸序列。根据本发明的实施例,所述系统包括:确定初始基因编码区序列装置,所述确定初始基因编码区序列装置基于用于改造待测样本核酸序列的候选基因,确定初始基因编码区序列;确定候选编码区序列集合装置,所述确定候选编码区序列集合装置与所述确定初始基因编码区序列装置相连,用于对所述初始基因编码区序列进行同

义密码子替换,以便获得由多个候选编码区序列构成的候选编码区序列集合;同源性风险评分装置,所述同源性风险评分装置与所述确定候选编码区序列集合装置相连,用于将所述候选编码区序列与宿主基因组和病毒基因组进行比对,并确定同源性高风险区域,以便对所述候选编码区序列进行同源性风险评分;以及任选地密码子对偏好性评分装置,所述密码子对偏好性评分装置与所述确定候选编码区序列集合装置相连,用于基于所述宿主中的密码子频率,对所述候选编码区序列进行密码子对偏好性评分;或密码子偏好性评分装置,所述密码子偏好性评分装置与所述确定候选编码区序列集合装置相连,用于基于所述宿主中的密码子频率,对所述候选编码区序列进行密码子偏好性评分;或CpG评分装置,所述CpG评分装置与所述确定候选编码区序列集合装置相连,用于基于所述候选编码区序列中的C碱基频率、G碱基频率和CpG序列频率,对所述候选编码区序列进行CpG评分;或TpA评分装置,所述TpA评分装置与所述确定候选编码区序列集合装置相连,用于基于所述候选编码区序列中的A碱基频率、T碱基频率和TpA序列频率,对所述候选编码区序列进行TpA评分;或mRNA二级结构最小自由能预测装置,所述mRNA二级结构最小自由能预测装置与所述候选编码区序列集合装置相连,用于基于所述候选编码区序列mRNA序列,对所述候选编码区序列进行mRNA二级结构最小自由能预测;或微卫星不稳定性评分装置,所述微卫星不稳定性评分装置与所述候选编码区序列集合装置相连,用于基于所述候选编码区序列中的微卫星序列,对所述候选编码区序列进行微卫星不稳定性评分;和确定优选编码区序列子集装置,所述确定优选编码区序列子集装置与所述同源性风险评分装置以及任选地密码子对偏好性评分装置、密码子偏好性评分装置、CpG评分装置、TpA评分装置、mRNA二级结构最小自由能预测装置和微卫星不稳定性评分装置相连,用于基于所述同源性风险评分以及任选地所述密码子对偏好性评分、所述密码子偏好性评分、所述CpG评分、所述TpA评分、所述mRNA二级结构最小自由能和所述微卫星不稳定性评分,在所述候选编码区序列集合中确定优选编码区序列子集。根据本发明实施例的上述系统适于执行上述筛选候选插入片段的方法,以降低序列的同源性风险,增加工程病毒基因组的稳定性。

[0041] 在本发明的第六方面,本发明提出了一种筛选候选插入片的系统,所述插入片用于改造待测样本核酸序列,所述系统包括:确定初始非编码区序列装置;第一同源亚型序列获取装置,所述第一同源亚型序列获取装置与所述确定初始非编码区序列装置相连,用于获取所述初始非编码区序列的第一同源亚型序列;确定保守性区域和非保守性区域装置,所述确定保守性区域和非保守性区域装置与所述确定初始非编码区序列装置和所述第一同源亚型序列获取装置相连,用于基于所述初始非编码区序列和所述第一同源亚型序列,分别确定各自的保守性区域和非保守性区域;确定候选非编码区序列集合装置,所述确定候选非编码区序列集合装置与所述确定保守性区域和非保守性区域装置相连,通过对所述非保守性区域进行随机突变和截短处理的至少之一,获得由多个候选非编码区序列构成的候选非编码区序列集合;确定优选非编码区序列子集装置,所述确定优选非编码区序列子集装置与所述确定候选非编码区序列集合装置相连,用于对所述非编码区序列进行所述同源性风险评分以及任选地所述微卫星不稳定性评分,在所述候选非编码区序列集合中确定优选非编码区序列子集。

[0042] 在本发明的第七方面,本发明提出了一种UAS序列。根据本发明的实施例,所述序列具有SEQ ID NO:1、4~6、14、17、23、25、31、34所示的核苷酸序列。发明人通过前面所述的

方法对UAS的同源序列进行筛选,发现,具有SEQ ID NO:1、4~6、14、17、23、25、31、34所示的核苷酸序列表达强度较强,可用于组合UAS的构建。

[0043] 在发明的第八方面,本发明提出了一组UAS组合。根据本发明的实施例,所述组合包括一系列UAS序列,所述UAS序列具有SEQ ID NO:4~6以及17所示的核苷酸序列;或SEQ ID NO:1、23、31以及34所示的核苷酸序列;或SEQ ID NO:1、4、5、6以及14所示的核苷酸序列;或SEQ ID NO:17、23、25、31以及34所示的核苷酸序列。发明人通过前面所述的方法对UAS同源序列进行筛选,发现,具有SEQ ID NO:1、4~6、14、17、23、25、31、34所示的核苷酸序列的UAS下游启动子表达强度较强,当采用上述的UAS序列组合构建互抑制开关型基因线路,所构建的四对组合启动子对与原启动子开关型基因线路相比,开关状态间的表达差异倍数均维持10倍以上,但源于启动子的病毒15代变异率从0.1降至未检出水平,因此,可以说,根据本发明实施例的上述UAS组合在维持下游启动子表达强度的同时有效降低了同源性风险。

[0044] 根据本发明的实施例,所述UAS组合从5'端到3'端具有如下所示的UAS连接顺序:

[0045] B1-A5-A6-A4 (SEQ ID NO:39)、D5-A1-D2-C1 (SEQ ID NO:41)、A5-A1-A6-A4-A14 (SEQ ID NO:42)或B1-D5-D2-C1-C3 (SEQ ID NO:44)。

[0046] 在发明的第九方面,本发明提出了一种Ni-scFv基因序列。根据本发明的实施例,所述序列具有SEQ ID NO:46或47所示的核苷酸序列。发明人通过前面所述的方法对Ni-scFv基因的同源序列进行筛选,发现,具有SEQ ID NO:46或47所示的核苷酸序列的Ni-scFv基因,最长无错配同源区域长度得到降低,15代变异率从0.01降至未检出水平,而有效含量与优化前相比无显著差异,因此,具有SEQ ID NO:46或47所示的核苷酸序列的Ni-scFv基因在维持目的基因功能的同时有效降低了同源性风险。

[0047] 在发明的第十方面,本发明提出了一种At-scFv基因序列。根据本发明的实施例,所述序列具有SEQ ID NO:49或50所示的核苷酸序列。发明人通过前面所述的方法对At-scFv基因的同源序列进行筛选,发现,具有SEQ ID NO:49或50所示的核苷酸序列的At-scFv基因,最长无错配同源区域长度得到降低,15代变异率从0.01降至未检出水平,而有效含量与优化前相比无显著差异,因此,具有SEQ ID NO:49或50所示的核苷酸序列的At-scFv基因在维持目的基因功能的同时有效降低了同源性风险。

## 附图说明

[0048] 图1是病毒同源性序列分析优化流程;

[0049] 图2是Gal4蛋白结合位点模式分布(引自10.1093/nar/gkl857);

[0050] 图3是4x UAS与5x UAS同源亚型启动子表达强度测试;

[0051] 图4是UAS同源亚型组合启动子表达强度测试;

[0052] 图5是UAS同源亚型组合启动子对构建互抑制开关型基因线路开关状态间表达差异测试;

[0053] 图6是UAS优化前后工程病毒功能与基因组稳定性测试;

[0054] 图7是目的基因序列优化前后工程病毒功能与基因组稳定性测试;

[0055] 图8是根据本发明实施例的筛选候选插入片段的系统的结构示意图;以及

[0056] 图9是根据本发明另一实施例的筛选候选插入片段的系统的结构示意图。

## 具体实施方式

[0057] 下面详细描述本发明的实施例,所述实施例的示例在附图中示出。下面通过参考附图描述的实施例是示例性的,旨在用于解释本发明,而不能理解为对本发明的限制。

[0058] 本发明提出了一种筛选候选插入片的系统,所述插入片用于改造待测样本核酸序列。根据本发明的实施例,参考图8,所述系统包括:确定初始基因编码区序列装置100,所述确定初始基因编码区序列装置100基于用于改造待测样本核酸序列的候选基因,确定初始基因编码区序列;确定候选编码区序列集合装置200,所述确定候选编码区序列集合装置200与所述确定初始基因编码区序列装置100相连,用于对所述初始基因编码区序列进行同义密码子替换,以便获得由多个候选编码区序列构成的候选编码区序列集合;同源性风险评分装置300,所述同源性风险评分装置300与所述确定候选编码区序列集合装置200相连,用于将所述候选编码区序列与宿主基因组和病毒基因组进行比对,并确定同源性高风险区域,以便对所述候选编码区序列进行同源性风险评分;以及任选地密码子对偏好性评分装置,所述密码子对偏好性评分装置与所述确定候选编码区序列集合装置相连,用于基于所述宿主中的密码子频率,对所述候选编码区序列进行密码子对偏好性评分;或密码子偏好性评分装置,所述密码子偏好性评分装置与所述确定候选编码区序列集合装置相连,用于基于所述宿主中的密码子频率,对所述候选编码区序列进行密码子偏好性评分;或CpG评分装置,所述CpG评分装置与所述确定候选编码区序列集合装置相连,用于基于所述候选编码区序列中的C碱基频率、G碱基频率和CpG序列频率,对所述候选编码区序列进行CpG评分;或TpA评分装置,所述TpA评分装置与所述确定候选编码区序列集合装置相连,用于基于所述候选编码区序列中的A碱基频率、T碱基频率和TpA序列频率,对所述候选编码区序列进行TpA评分;或mRNA二级结构最小自由能预测装置,所述mRNA二级结构最小自由能预测装置与所述候选编码区序列集合装置相连,用于基于所述候选编码区序列mRNA序列,对所述候选编码区序列进行mRNA二级结构最小自由能预测;或微卫星不稳定性评分装置,所述微卫星不稳定性评分装置与所述候选编码区序列集合装置相边,用于基于所述候选编码区序列中的微卫星序列,对所述候选编码区序列进行微卫星不稳定性评分;和确定优选编码区序列子集装置400,所述确定优选编码区序列子集装置与所述同源性风险评分装置300以及任选地密码子对偏好性评分装置、密码子偏好性评分装置、CpG评分装置、TpA评分装置、mRNA二级结构最小自由能预测装置和微卫星不稳定性评分装置相连,用于基于所述同源性风险评分以及任选地所述密码子对偏好性评分、所述密码子偏好性评分、所述CpG评分、所述TpA评分、所述mRNA二级结构最小自由能和所述微卫星不稳定性评分,在所述候选编码区序列集合中确定优选编码区序列子集。根据本发明实施例的上述系统适于执行上述筛选候选插入片的方法,以降低序列的同源性风险,增加工程病毒基因组的稳定性。

[0059] 在本发明的第六方面,本发明提出了一种筛选候选插入片的系统,所述插入片用于改造待测样本核酸序列,参考图9,所述系统包括:确定初始非编码区序列装置500;第一同源亚型序列获取装置600,所述第一同源亚型序列获取装置600与所述确定初始非编码区序列装置500相连,用于获取所述初始非编码区序列的第一同源亚型序列;确定保守性区域和非保守性区域装置700,所述确定保守性区域和非保守性区域装置700与所述确定初始非编码区序列装置500和所述第一同源亚型序列获取装置相连600,用于基于所述初始非编码区序列和所述第一同源亚型序列,分别确定各自的保守性区域和非保守性区域;确定



候选非编码区序列集合装置800,所述确定候选非编码区序列集合装置800与所述确定保守性区域和非保守性区域装置700相连,通过对所述非保守性区域进行随机突变和截短处理的至少之一,获得由多个候选非编码区序列构成的候选非编码区序列集合;确定优选非编码区序列子集装置900,所述确定优选非编码区序列子集装置900与所述确定候选非编码区序列集合装置800,用于对所述非编码区序列进行所述同源性风险评分以及任选地所述微卫星不稳定性评分,在所述候选非编码区序列集合中确定优选非编码区序列子集。

[0060] 本申请是基于发明人对以下事实和问题的发现和认识作出的:

[0061] 工程病毒的药学生产过程中,人工外源序列自身与病毒基因组可能存在高度同源的区域,提升了由同源重组导致结构性变异的风险。发明人通过对人工外源序列与病毒基因组序列特征进行信息学分析,预测分析存在同源性风险的区域,并结合同义密码子替换与同源亚型替换等方式降低序列的同源性风险,增加工程病毒基因组的稳定性。

[0062] 实施例1

[0063] 具体流程如下:

[0064] 通过文献调研、一代测序等方式获得工程病毒(如腺病毒)的参考基因组序列与待优化的人工外源序列。通过数据整理等方式获得人基因组密码子频率与密码子对频率。

[0065] 1)使用短序列比对软件(如Bowtie)对人工外源序列与自身、人基因组和工程腺病毒基因组序列进行滑窗序列比对,滑窗长度为12~17bp,错配数为0~4bp,输出最长一致序列长度与频数,并进行评分;按优先级由高至低依次比较错配数、参考序列与最长一致序列,按优先级由高至低比较错配数0~4bp,按优先级由高至低比较参考序列为自身、病毒基因组与宿主基因组序列,按优先级由高至低比较最长一致序列长度与频数,比较不同候选编码区序列的同源性风险,优先级越高、数值越大时同源性风险越高;根据腺病毒实验经验,最长一致序列长度15bp以上、错配0bp的区域为同源性风险高的区域。

[0066] 2)对于位于基因编码区的序列,进行服从人基因组密码子频率分布的同义密码子随机替换,过滤其中包含影响序列克隆与载体构建的酶切位点序列、同源克隆序列与条形码序列的随机序列,获得编码氨基酸不变的5000条备选优化序列。

[0067] 3)分别对备选序列进行步骤1)的同源性风险评分。分别对备选序列进行密码子对频率统计,并结合人基因组密码子对偏好性进行密码子对偏好性评分和密码子偏好性评分,评分方式为

$$\text{密码子对偏好性评分} = \sum \ln \frac{\text{密码子对频率} * \text{各氨基酸频率乘积}}{\text{氨基酸对频率} * \text{各密码子频率乘积}}$$

[0068]

$$\text{密码子偏好性评分} = \sum \ln \frac{\text{密码子频率}}{\text{氨基酸频率}}$$

[0069] 分别对备选序列进行CpG与TpA序列特征频率统计并进行评分,评分方式为

$$\text{CpG 评分} = \frac{\text{CpG 频率}}{\text{C 碱基频率} * \text{G 碱基频率}}, \quad \text{TpA 评分} = \frac{\text{TpA 频率}}{\text{T 碱基频率} * \text{A 碱基频率}}$$

[0071] 使用软件Mfold或ViennaRNA分别对备选序列进行mRNA二级结构最小自由能预测。

[0072] 分别对备选序列进行微卫星序列特征统计并进行评分,对于连续重复至少3次、微

卫星序列长度小于同源性风险评分滑窗长度的微卫星区域,统计最长微卫星区域长度与频数,按优先级由高至低比较最长微卫星区域长度与频数,优先级越高、数值越大时微卫星不稳定性越高。

[0073] 基于上述评分对备选序列进行排序,筛选各项评分均为前50的备选序列,从中选择备选序列进行生物学功能与工程腺病毒基因组稳定性实验验证。

[0074] 4) 对于位于其它调控区域的序列,通过文献挖掘获得生物学功能相近的同源亚型序列;通过序列相似性搜索软件(如Blast)进行分析,获得序列保守性区域;通过对非保守性区域进行随机突变、截短等方式获得其它备选同源亚型序列,并过滤其中包含影响序列克隆与载体构建的酶切位点序列、同源克隆序列与条形码序列的随机序列。

[0075] 5) 分别对备选序列进行步骤1)的同源性风险评分和微卫星不稳定性评分并排序,从中选择备选序列进行生物学功能与工程腺病毒基因组稳定性实验验证。

[0076] 6) 最终筛选出序列功能相近而病毒保持传代稳定性同源性优化序列。

[0077] 根据本发明实施例的流程示意图如图1所示。

[0078] 下面将进一步详细描述本发明的实施例,所述实施例的示例在附图中示出。下面通过参考附图描述的实施例是示例性的,旨在用于解释本发明,而不能理解为对本发明的限制。

[0079] Gal4/UAS是存在于酵母中的基因表达调控系统。转录调控因子Gal4的结合域可以与UAS序列(启动子上游的转录因子结合位点)结合,而激活域可以激活下游启动子。多个串联的UAS可以有效增强Gal4的结合能力以增强对于下游启动子的激活强度。在病毒载体中,UAS的同源序列可能导致重组变异而失去功能。

[0080] 由此,发明人对UAS进行了同源序列优化。

[0081] 实施例2UAS同源序列优化

[0082] 实验一

[0083] 检测4x UAS与5x UAS不同同源亚型下游启动子表达强度

[0084] 通过生物信息学分析,可得Gal4蛋白结合位点的序列模式如图2所示。分别参考不同的野生型Gal4蛋白结合位点,对非保守区域进行同源亚型替换,并将过同源性风险评分,获得一系列UAS同源亚型序列如SEQ ID NO:1~37所示。

[0085] >A1

[0086] CGGAGTACTGTCCTCCG (SEQ ID NO:1)。

[0087] >A2

[0088] CGGAGGACTGTCCTCCG (SEQ ID NO:2)。

[0089] >A3

[0090] CGGAGAACTGTTCTCCG (SEQ ID NO:3)。

[0091] >A4

[0092] CGGAGCACTGTGCTCCG (SEQ ID NO:4)。

[0093] >A5

[0094] CGGACTACTGTAGTCCG (SEQ ID NO:5)。

[0095] >A6

[0096] CGGACGACTGTCGTCCG (SEQ ID NO:6)。

- [0097] >A7  
[0098] CGGACAACACTGTTGTCCG (SEQ ID NO:7)。  
[0099] >A8  
[0100] CGGACCACTGTGGTCCG (SEQ ID NO:8)。  
[0101] >A9  
[0102] CGGAATACTGTATTCCG (SEQ ID NO:9)。  
[0103] >A10  
[0104] CGGAAGACTGTCTTCCG (SEQ ID NO:10)。  
[0105] >A11  
[0106] CGGAAAACACTGTTTTCCG (SEQ ID NO:11)。  
[0107] >A12  
[0108] CGGAACACTGTGTTCCG (SEQ ID NO:12)。  
[0109] >A13  
[0110] CGGATTACTGTAATCCG (SEQ ID NO:13)。  
[0111] >A14  
[0112] CGGATGACTGTCATCCG (SEQ ID NO:14)。  
[0113] >A15  
[0114] CGGATAACTGTTATCCG (SEQ ID NO:15)。  
[0115] >A16  
[0116] CGGATCACTGTGATCCG (SEQ ID NO:16)。  
[0117] >B1  
[0118] CGGGGTGCCGCCCCCG (SEQ ID NO:17)。  
[0119] >B2  
[0120] CGGGATGCCGCATCCCG (SEQ ID NO:18)。  
[0121] >B3  
[0122] CGGGAAGCCGCTTCCCG (SEQ ID NO:19)。  
[0123] >B4  
[0124] CGGGTTGCCGCAACCCG (SEQ ID NO:20)。  
[0125] >B5  
[0126] CGGGTAGCCGCTACCCG (SEQ ID NO:21)。  
[0127] >B6  
[0128] CGGGCCGCCGCGTCCCG (SEQ ID NO:22)。  
[0129] >C1  
[0130] CGGAGTGCCGCACTCCG (SEQ ID NO:23)。  
[0131] >C2  
[0132] CGGGTACTGTACCCCG (SEQ ID NO:24)。  
[0133] >C3  
[0134] CGGAGTGCTGTACTCCG (SEQ ID NO:25)。  
[0135] >C4

- [0136] CGGAGTACCGTACTCCG (SEQ ID NO:26)。  
 [0137] >C5  
 [0138] CGGGGTGCTGCACCCCG (SEQ ID NO:27)。  
 [0139] >C6  
 [0140] CGGGGTACCGTACCCCG (SEQ ID NO:28)。  
 [0141] >C7  
 [0142] CGGGGTGCCGCACCCCG (SEQ ID NO:29)。  
 [0143] >D1  
 [0144] CGGTCCACTGTGTGCCG (SEQ ID NO:30)。  
 [0145] >D2  
 [0146] CGGGTGACAGCCCTCCG (SEQ ID NO:31)。  
 [0147] >D3  
 [0148] CGGCCATATGTCTTCCG (SEQ ID NO:32)。  
 [0149] >D4  
 [0150] CGGCGGTCTTTCGTCCG (SEQ ID NO:33)。  
 [0151] >D5  
 [0152] CGGGTGACCGCCCTCCG (SEQ ID NO:34)。  
 [0153] >D6  
 [0154] CGGTCCACAGTGTGCCG (SEQ ID NO:35)。  
 [0155] >D7  
 [0156] CGGCCATATCGCTTCCG (SEQ ID NO:36)。  
 [0157] >D8  
 [0158] CGGGTGACAGCCCTCCG (SEQ ID NO:37)。

[0159] 使用UAS同源亚型A1、C3、D1~D5构建4x启动子,分别转染靶细胞测试原启动子与各亚型启动报告基因的表达强度,并以原启动子表达强度归一化为相对表达强度,比较各亚型与原启动子间的表达差异。

[0160] 使用UAS同源亚型A2~A16、B1~B5、C1~C6、D1~D2构建5x启动子,分别转染靶细胞测试原启动子与各亚型启动报告基因的表达强度,并以原启动子表达强度归一化为相对表达强度,比较各亚型与原启动子间的表达差异。

[0161] 检测结果如图3。各同源亚型中,A1、A4、A5、A6、A14、B1、C1、C3、D2、D5相对表达强度较强,可用于构建组合UAS启动子。

[0162] 实验二

[0163] 检测4x UAS与5x UAS组合不同同源亚型下游启动子表达强度

[0164] 根据实验一结果与序列间同源性,构建4x UAS与5x UAS组合启动子如表2,并添加两种不同的抑制位点,分别转染靶细胞测试原启动子与各亚型组合启动子报告基因的表达强度,并以原启动子表达强度归一化为相对表达强度,比较各亚型组合启动子与原启动子间的表达差异。

[0165] 表2:

[0166]	组合启动子	UAS同源亚型
--------	-------	---------

组合A	A5-A1-A6-A4
组合B	B1-A5-A6-A4
组合C	B1-D5-D2-C1
组合D	D5-A1-D2-C1
组合E	A5-A1-A6-A4-A14
组合F	B1-A5-A6-A4-A14
组合G	B1-D5-D2-C1-C3
组合H	D5-A1-D2-C1-C3

[0167] >A5-A1-A6-A4

[0168] CGGACTACTGTAGTCCGCGGAGTACTGTCCTCCGCGGACGACTGTCGTCCGCGGAGCACTGTGCTCCG  
(SEQ ID NO:38)。

[0169] >B1-A5-A6-A4

[0170] CGGGGTGCCGCCCCCGCGGACTACTGTAGTCCGCGGACGACTGTCGTCCGCGGAGCACTGTGCTCCG  
(SEQ ID NO:39)。

[0171] >B1-D5-D2-C1

[0172] CGGGGTGCCGCCCCCGCGGGTGACCGCCCTCCGCGGGTGACAGCCCTCCGCGGAGTGCCGCACTCCG  
(SEQ ID NO:40)。

[0173] >D5-A1-D2-C1

[0174] CGGGTGACCGCCCTCCGCGGAGTACTGTCCTCCGCGGGTGACAGCCCTCCGCGGAGTGCCGCACTCCG  
(SEQ ID NO:41)。

[0175] >A5-A1-A6-A4-A14

[0176] CGGACTACTGTAGTCCGCGGAGTACTGTCCTCCGCGGACGACTGTCGTCCGCGGAGCACTGTGCTCCG  
CGGATGACTGTCATCCG (SEQ ID NO:42)。

[0177] >B1-A5-A6-A4-A14

[0178] CGGGGTGCCGCCCCCGCGGACTACTGTAGTCCGCGGACGACTGTCGTCCGCGGAGCACTGTGCTCCG  
CGGATGACTGTCATCCG (SEQ ID NO:43)。

[0179] >B1-D5-D2-C1-C3

[0180] CGGGGTGCCGCCCCCGCGGGTGACCGCCCTCCGCGGGTGACAGCCCTCCGCGGAGTGCCGCACTCCG  
CGGAGTGCTGTACTCCG (SEQ ID NO:44)。

[0181] >D5-A1-D2-C1-C3

[0182] CGGGTGACCGCCCTCCGCGGAGTACTGTCCTCCGCGGGTGACAGCCCTCCGC GGAGTGCCGCACTCC  
GCGGAGTGCTGTACTCCG (SEQ ID NO:45)。

[0183] 使用添加了不同抑制位点的组合启动子对启动相应抑制子表达,构建互抑制开关型基因线路(构建方法参考专利PCT/CN2017/096043),分别转染靶细胞测试原启动子对与各亚型组合启动子对开关型基因线路中报告基因开关状态的表达差异倍数,比较各亚型组合启动子对与原启动子对开关型基因线路开关状态间的表达差异倍数。

[0184] 检测结果如图4~5。各同源亚型组合启动子中,组合B、组合D、组合E、组合G表达强度较强,用于构建互抑制开关型基因线路。所构建的四对组合启动子对与原启动子开关型基因线路相比,开关状态间的表达差异倍数均维持10倍以上。

[0185] 实验三

[0186] 检测启动子优化前后工程病毒功能与稳定性差异

[0187] 装载原启动子开关型基因线路的工程病毒传代15代,以测序深度1G的深度测序检测变异情况,变异率检测极限为0.001;装载组合B-组合D启动子对开关型基因线路的工程病毒传代15代,以测序深度1G的深度测序检测变异情况,变异率检测极限为0.001;等量的两种工程病毒侵染靶细胞后,测定启动子启动的目的基因产物功能性有效含量,并以启动子优化前水平归一化为相对表达强度;比较序列同源性优化前后启动子表达强度与病毒基因组稳定性。

[0188] 检测结果如图6。通过启动子序列优化,源于启动子的病毒15代变异率从0.1降至未检出水平,而表达的目的基因有效含量与优化前相比无显著差异。可见通过同源性序列优化,在维持启动子表达强度的同时有效降低了同源性风险。

[0189] 程序性死亡受体1 (PD-1) 与其配体 (PD-L1) 是重要的免疫检查点分子,T细胞表面的PD-1分子与配体PD-L1 (常见于肿瘤表面) 结合后会抑制T细胞活化与增殖并诱导T细胞凋亡,是肿瘤微环境免疫抑制性的重要机制。使用单链抗体 (scFv) 结合PD-1或PD-L1以阻止二者结合可以有效增强肿瘤微环境中的免疫作用,可单独作为小分子靶向药物或装载于其它靶向药物载体。在病毒载体中,scFv中的同源序列可能导致重组变异而失去功能。在本申请中,靶向PD-1的单链抗体被称为Ni-scFv,靶向PD-L1的被称为At-scFv。

[0190] 由此,发明人对Ni-scFv基因和At-scFv基因进行了同源序列优化。

[0191] 实施例3目的基因同源序列优化

[0192] 分别对靶向PD-L1的At-scFv与靶向PD-1的Ni-scFv基因进行同源序列优化。

[0193] 实验一

[0194] Ni-scFv基因同源序列优化

[0195] 通过同义密码子替换,并依据实施例1的方式获得的Ni-scFv基因同源亚型序列如SEQ ID NO:46~48所示

[0196] >Ni-scFv-a

[0197] ATGTCCGTTCCAACCCAGGTTCTCGGCCTCCTACTGCTGTGGCTCACCGATGCCAGATGCCAGGTGCA  
ATTGGTAGAGTCCGGGGCGGGGTGGTGCAGCCAGGGCGATCTTTGCGGTTAGACTGCAAGGCATCCGGAATAACC  
TTTTCTAATAGCGGGATGCATTGGGTGAGACAAGCACCCGAAAAGGCCTTGAATGGGTAGCAGTCATATGGTATG  
ACGGTAGCAAACGCTACTATGCAGACTCGGTCAAGGGTCGATTCACTATTAGTCGCGATAACAGCAAGAATACCTT  
GTTCTGAGATGAACTCACTACGGGCTGAAGACACAGCTGTCTATTACTGTGCAACCAACGACGATTATTGGGGA  
CAGGGGACCTTGGTCACCGTCAGTTCCGGGGGAGGGGGATCCGGAGGCGGCGGCTCAGGAGGAGGTGGCTCCGAGA  
TCGTAAGTACTCAGAGCCCGCAACTTTGTCTCTGTGCGCTGGCGAGCGGGCCACACTGTCCTGTAGGGCCTCACA  
GAGCGTGAGTTCATATCTAGCTTGGTATCAACAGAAGCCGGGGCAGGCCCTAGATTGCTAATCTACGATGCAAGT  
AACAGAGCAACTGGCATCCCCGCGAGATTTAGCGGATCCGGTTCGGAAACCGACTTTACTACTACTATCTCCTCCC  
TAGAACCAGAAGATTTGCGTCTACTATTGCCAACAGTCGTCCAACCTGGCCTCGCACATTTGGGCAGGAACCAA  
GGTTGAAATTTAAA (SEQ ID NO:46)。

[0198] >Ni-scFv-b

[0199] ATGTCCGTTCCACCCAGGTGCTCGGCCTCCTATTACTCTGGCTAACTGATGCCCGGTGTCAGGTGCA  
GCTCGTGGAGAGTGGAGGGGGTGTGGTTCAGCCGGGTCGGTCACTGCGGCTGGACTGTAAAGCTAGCGGTATCACG

TTCAGTAACTCAGGTATGCACTGGGTACGGCAGGCCCCCCGGCAAAGGCTTGGAGTGGGTTGCTGTGATATGGTACG  
ATGTTCTAAAAGGTATTATGCTGATTCCGTGAAGGGCCGGTTTACCATATCACGCGACAACCTCCAAGAACACGCT  
CTTCCTCCAGATGAATTCCTCCGAGCGGAAGACACCGCGTTTATTATTGCGCCACCAATGATGACTACTGGGGC  
CAGGGCACCTTGGTTACCGTGTCTAGCGGAGGGGGGGGCTCGGGGGCGGGCGGTAGCGGTGGAGGTGGGTTCCGAGA  
TCGTCTCACGCAATCCCCGCCACCCTTAGTCTCAGCCCTGGGGAGCGGGCAACCCTTAGCTGCCGGGCGTCCCA  
GTCAGTCAGTTCTACCTTGCTGTACCAGCAAAGCCCCGGCCAGGCACCTCGCCTCCTTATTTATGATGCATCG  
AACCGAGCAACCGGAATTCCTGCGCGTTTTCAGTGGTTCTGGTAGCGGGACCGACTTCACACTTACAATATCTAGCC  
TAGAACCAGAAGACTTCGCTGTCTACTACTGCCAACAGAGCTCGAACTGGCCTAGAACATTCGGGCAGGGCACCAA  
GGTAGAAATCAAA (SEQ ID NO:47)。

[0200] >Ni-scFv-c

[0201] ATGAGCGTGCACCACCCAGGTGCTGGGCCTGCTGCTGCTGTGGCTGACCGACGCCAGGTGCCAGGTGCA  
GCTGGTGGAGAGCGGCGGGCGGCGTGGTGCAGCCCAGGAGCCTGAGGCTGGACTGCAAGGCCAGCGGCATCAC  
TTCAGCAACAGCGGCATGCACTGGGTGAGGCAGGCCCCCCGGCAAAGGGCCTGGAGTGGGTGGCCGTGATCTGGTACG  
ACGGCAGCAAGAGGTACTACGCCAGCGTGAAGGGCAGTTTACCATCAGCAGGGACAACAGCAAGAACCCT  
GTTCTTGCAGATGAACAGCCTGAGGGCCGAGGACACCGCGTGTACTACTGCGCCACCAACGACTACTGGGGC  
CAGGGCACCTGCTGACCGTGAGCAGCGGGCGGCGGCAGCGGCGGGCGGCAGCGGCGGGCGGCGGCAGCGGCGGGCGG  
TCGTGCTGACCCAGAGCCCCGCCACCCTGAGCCTGAGCCCCGGCGAGAGGGCCACCCTGAGCTGCAGGGCCAGCCA  
GAGCGTGAGCAGCTACCTGGCCTGGTACCAGCAGAAGCCCCGGCCAGGCCCCCAGGCTGCTGATCTACGACGCCAGC  
AACAGGGCCACCGGCATCCCCGCCAGTTTCAGCGGCAGCGGCAGCGGCACCGACTTCACCCTGACCATCAGCAGCC  
TGGAGCCCAGGACTTCGCCGTGTACTACTGCCAGCAGAGCAGCAACTGGCCCAGGACCTTCGCCAGGGCACCAA  
GGTGGAGATCAAG (SEQ ID NO:48)。

[0202] 装载目的基因的工程病毒传代15代,以测序深度1G的深度测序检测变异情况,变异率检测极限为0.001;装载同源性优化后的目的基因的工程病毒传代15代,以测序深度1G的深度测序检测变异情况,变异率检测极限为0.001;等量的两种工程病毒侵染靶细胞后,测定目的基因产物功能性有效含量,并以目的基因优化前水平归一化为相对有效含量;比较序列同源性优化前后目的基因功能与病毒基因组稳定性。

[0203] 检测结果见表3与图7。目的基因序列优化前,与自身序列最长无错配同源区域为16bp,为高同源性风险区域,传代15代后可检出变异率达0.01;与病毒骨架序列最长无错配同源区域为14bp,未达到高同源性风险,传代15代后未检出变异;通过目的基因序列优化,获得的最优Ni-scFv基因同源亚型序列Ni-scFv-a与Ni-scFv-b,最长无错配同源区域长度得到降低,15代变异率从0.01降至未检出水平,而有效含量与优化前相比无显著差异。可见通过同源性序列优化,在维持目的基因功能的同时有效降低了同源性风险。

[0204] 表3:

[0205]	优化前		优化后	
	与自身序列	与病毒骨架序列	与自身序列	与病毒骨架序列
最长无错配同源区域	16 bp	14 bp	<12 bp	12 bp
传代15代检出变异比例	0.01	未检出	未检出	未检出

[0206] 实验二

[0207] 通过同义密码子替换,并依据实施例1的方法获得的At-scFv基因同源亚型序列如SEQ ID NO:49~51所示。

[0208] >At-scFv-a

[0209] ATGTCCGTCCCTACTCAGGTACTAGGCCTCCTCTTGCTATGGCTGACCGACGCTAGATGCGAAGTGCA  
ACTGGTCGAATCCGGCGGTGGCTTGGTCCAGCCAGGCGGATCATTACGCTGTCTTGTGCAGCATCAGGCTTCACC  
TTTAGTGACAGTTGGATCCATTGGGTCCGGCAAGCCCCAGGCAAAGGGCTGGAATGGGTGCGCTGGATTAGCCCAT  
ATGGCGGCAGCACCTATTACGCCGACAGCGTCAAGGGGCGCTTTACCATTAGTGCTGACACAAGTAAGAATACCGC  
TTATCTGCAGATGAATAGCCTGCGGGCCGAAGACACGGCTGTTTACTACTGTGCACGACGCCACTGGCCGGGCGGT  
TTTACTATTGGGGACAAGGAACTCTCGTGACAGTCTCGTCCGGAGGCGGCGGTTTCAGGTGGCGGCGGCTCAGGAG  
GGGGGGTTCTGATATACAGATGACACAATCCCCTTCTTCTCTGAGCGCAAGTGTGGGCGATCGGGTAACCATCAC  
CTGTCCGGCTTCCCAGGACGTGAGTACAGCCGTGGCTTGGTATCAACAGAAGCCAGGCAAGGCCCCGAAGCTGCTA  
ATCTACAGCGCTAGTTTCTGTACTCAGGGGTCCGAGCCGCTTTAGCGGAAGTGGATCTGGTACTGACTTCACAC  
TTACTATCAGTTCTCTACAGCCGGAAGACTTCGCTACATATTACTGCCAGCAGTATCTCTATCACCCCGCTACTTT  
TGGACAAGGAACAAAAGTTGAGATCAAG (SEQ ID NO:49)。

[0210] >At-scFv-b

[0211] ATGTCAGTCCCCACCCAGGTCCTTGGACTACTCCTTCTATGGTTGACAGATGCCCGCTGCGAAGTCCA  
GTTGGTGAATCCGGGGGCGCCTTGTACAGCCCGGGGGAGCCTGAGACTCAGCTGTGCCGCTTCAGGATTTACC  
TTCTCTGATTCTGGATACATTGGGTACGTACGGCCCTGGGAAGGGATTGGAGTGGGTGGCTGGATCAGTCCAT  
ACGGTGGCTCTACGTACTATGCGGACAGCGTCAAAGGGCGCTTTACTATTAGTGACATACATCGAAGAATACAGC  
CTACCTGCAGATGAATTCATTGAGGGCAGAGGACACTGCCGTCTATTACTGTGCAAGAAGGCACTGGCCCGGCGGC  
TTCGACTATTGGGGACAGGGCACCCCTGGTCACAGTATCTTCAGGCGGAGGAGGTTCCGGGGGCGGCGGCTCCGGCG  
GTGGAGGCTCAGATATCCAGATGACACAGAGCCCCAGCTCTTTATCAGCTTCAGTGGGCGATCGGGTCACCATCAC  
TTGTCTGCCTCTCAGGATGTGTCTACCGCCGTGGCCTGGTATCAACAGAAACCCGGCAAGGCCCCAAAATTACTG  
ATATATAGTGCAAGCTTCCTGTACTCAGGAGTCCCTTACGCTTCTCCGGCTCGGGCAGTGGGACCGACTTTACTC  
TGACGATATCCAGTCTGCAGCCTGAAGATTTGCTACCTACTACTGTCAACAGTACTTGTACCACCCCGCAACATT  
CGGACAGGGGACCAAAGTAGAAATTA (SEQ ID NO:50)。

[0212] >At-scFv-c

[0213] ATGAGCGTGCCACCCAGGTGCTGGGCCTGCTGCTGCTGTGGCTGACCGACGCCAGGTGCGAGGTGCA  
GCTGGTGGAGAGCGGCGGCGCCTGGTGCAGCCCGGCGGCAGCCTGAGGCTGAGCTGCGCCGCCAGCGGCTTCACC  
TTCAGCGACAGCTGGATCCACTGGGTGAGGCAGGCCCCGGAAGGGCCTGGAGTGGGTGGCTGGATCAGCCCCT  
ACGGCGGCAGCACCTACTACGCCGACAGCGTGAAGGGCAGGTTTACCATCAGCGCCGACACCAGCAAGAACACCGC  
CTACCTGCAGATGAACAGCCTGAGGGCCGAGGACACCGCCGTGTACTACTGCGCCAGGAGGCACTGGCCCGGCGGC  
TTCGACTACTGGGGCCAGGGCACCCCTGGTGACCGTGAGCAGCGGCGGCGGCAGCGGCGGCGGCAGCGGCG  
GCGGCGGCAGCGACATCCAGATGACCCAGAGCCCCAGCAGCCTGAGCGCCAGCGTGGGCGACAGGGTGACCATCAC  
CTGCAGGGCCAGCCAGGACGTGAGCACCGCCGTGGCCTGGTACCAGCAGAAGCCCGGCAAGGCCCCCAAGCTGCTG  
ATCTACAGCGCCAGCTTCTGTACAGCGGCGTGGCCAGCAGGTTTACGCGGCAGCGGCAGCGGCACCGACTTCACCC  
TGACCATCAGCAGCCTGCAGCCGAGGACTTCGCCACCTACTACTGCCAGCAGTACCTGTACCACCCCGCCACCTT  
CGGCCAGGGCACCAAGGTGGAGATCAAG (SEQ ID NO:51)。



[0214] 装载目的基因的工程病毒传代15代,以测序深度1G的深度测序检测变异情况,变异率检测极限为0.001;装载同源性优化后的目的基因的工程病毒传代15代,以测序深度1G的深度测序检测变异情况,变异率检测极限为0.001;等量的两种工程病毒侵染靶细胞后,测定目的基因产物功能性有效含量,并以目的基因优化前水平归一化为相对有效含量;比较序列同源性优化前后目的基因功能与病毒基因组稳定性。

[0215] 通过目的基因序列优化,获得的最优At-scFv基因同源亚型序列At-scFv-a与At-scFv-b,最长无错配同源区域长度得到降低,15代变异率降低,而有效含量与优化前相比无显著差异。可见通过同源性序列优化,在维持目的基因功能的同时有效降低了同源性风险。

[0216] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0217] 尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

- [0001] SEQUENCE LISTING
- [0002] <110> 北京合生基因科技有限公司
- [0003] <120> 筛选候选插入片段的方法和系统
- [0004] <130> PIDC3193676
- [0005] <160> 51
- [0006] <170> PatentIn version 3.5
- [0007] <210> 1
- [0008] <211> 17
- [0009] <212> DNA
- [0010] <213> Artificial Sequence
- [0011] <220>
- [0012] <223> UAS同源亚型序列
- [0013] <400> 1
- [0014] cggagtactg tcctccg 17
- [0015] <210> 2
- [0016] <211> 17
- [0017] <212> DNA
- [0018] <213> Artificial Sequence
- [0019] <220>
- [0020] <223> UAS同源亚型序列
- [0021] <400> 2
- [0022] cggaggactg tcctccg 17
- [0023] <210> 3
- [0024] <211> 17
- [0025] <212> DNA
- [0026] <213> Artificial Sequence
- [0027] <220>
- [0028] <223> UAS同源亚型序列
- [0029] <400> 3
- [0030] cggagaactg ttctccg 17
- [0031] <210> 4
- [0032] <211> 17
- [0033] <212> DNA
- [0034] <213> Artificial Sequence
- [0035] <220>
- [0036] <223> UAS同源亚型序列
- [0037] <400> 4
- [0038] cggagcactg tgctccg 17

[0039] <210> 5  
[0040] <211> 17  
[0041] <212> DNA  
[0042] <213> Artificial Sequence  
[0043] <220>  
[0044] <223> UAS同源亚型序列  
[0045] <400> 5  
[0046] cggactactg tagtccg 17  
[0047] <210> 6  
[0048] <211> 17  
[0049] <212> DNA  
[0050] <213> Artificial Sequence  
[0051] <220>  
[0052] <223> UAS同源亚型序列  
[0053] <400> 6  
[0054] cggacgactg tcgtccg 17  
[0055] <210> 7  
[0056] <211> 17  
[0057] <212> DNA  
[0058] <213> Artificial Sequence  
[0059] <220>  
[0060] <223> UAS同源亚型序列  
[0061] <400> 7  
[0062] cggacaactg ttgtccg 17  
[0063] <210> 8  
[0064] <211> 17  
[0065] <212> DNA  
[0066] <213> Artificial Sequence  
[0067] <220>  
[0068] <223> UAS同源亚型序列  
[0069] <400> 8  
[0070] cggaccactg tgggccg 17  
[0071] <210> 9  
[0072] <211> 17  
[0073] <212> DNA  
[0074] <213> Artificial Sequence  
[0075] <220>  
[0076] <223> UAS同源亚型序列  
[0077] <400> 9

[0078] cggaatactg tattccg 17  
[0079] <210> 10  
[0080] <211> 17  
[0081] <212> DNA  
[0082] <213> Artificial Sequence  
[0083] <220>  
[0084] <223> UAS同源亚型序列  
[0085] <400> 10  
[0086] cggaagactg tcttccg 17  
[0087] <210> 11  
[0088] <211> 17  
[0089] <212> DNA  
[0090] <213> Artificial Sequence  
[0091] <220>  
[0092] <223> UAS同源亚型序列  
[0093] <400> 11  
[0094] cggaaaactg ttttccg 17  
[0095] <210> 12  
[0096] <211> 17  
[0097] <212> DNA  
[0098] <213> Artificial Sequence  
[0099] <220>  
[0100] <223> UAS同源亚型序列  
[0101] <400> 12  
[0102] cggaacactg tgttccg 17  
[0103] <210> 13  
[0104] <211> 17  
[0105] <212> DNA  
[0106] <213> Artificial Sequence  
[0107] <220>  
[0108] <223> UAS同源亚型序列  
[0109] <400> 13  
[0110] cggattactg taatccg 17  
[0111] <210> 14  
[0112] <211> 17  
[0113] <212> DNA  
[0114] <213> Artificial Sequence  
[0115] <220>  
[0116] <223> UAS同源亚型序列

[0117] <400> 14  
[0118] cggatgactg tcatccg 17  
[0119] <210> 15  
[0120] <211> 17  
[0121] <212> DNA  
[0122] <213> Artificial Sequence  
[0123] <220>  
[0124] <223> UAS同源亚型序列  
[0125] <400> 15  
[0126] cggataactg ttatccg 17  
[0127] <210> 16  
[0128] <211> 17  
[0129] <212> DNA  
[0130] <213> Artificial Sequence  
[0131] <220>  
[0132] <223> UAS同源亚型序列  
[0133] <400> 16  
[0134] cggatcactg tgatccg 17  
[0135] <210> 17  
[0136] <211> 17  
[0137] <212> DNA  
[0138] <213> Artificial Sequence  
[0139] <220>  
[0140] <223> UAS同源亚型序列  
[0141] <400> 17  
[0142] cggggtgccg cccccg 17  
[0143] <210> 18  
[0144] <211> 17  
[0145] <212> DNA  
[0146] <213> Artificial Sequence  
[0147] <220>  
[0148] <223> UAS同源亚型序列  
[0149] <400> 18  
[0150] cgggatgccg catccg 17  
[0151] <210> 19  
[0152] <211> 17  
[0153] <212> DNA  
[0154] <213> Artificial Sequence  
[0155] <220>

[0156] <223> UAS同源亚型序列  
[0157] <400> 19  
[0158] cgggaagccg cttcccg 17  
[0159] <210> 20  
[0160] <211> 17  
[0161] <212> DNA  
[0162] <213> Artificial Sequence  
[0163] <220>  
[0164] <223> UAS同源亚型序列  
[0165] <400> 20  
[0166] cggggtgccg caaccg 17  
[0167] <210> 21  
[0168] <211> 17  
[0169] <212> DNA  
[0170] <213> Artificial Sequence  
[0171] <220>  
[0172] <223> UAS同源亚型序列  
[0173] <400> 21  
[0174] cgggtagccg ctaccg 17  
[0175] <210> 22  
[0176] <211> 17  
[0177] <212> DNA  
[0178] <213> Artificial Sequence  
[0179] <220>  
[0180] <223> UAS同源亚型序列  
[0181] <400> 22  
[0182] cgggccgccg cgtcccg 17  
[0183] <210> 23  
[0184] <211> 17  
[0185] <212> DNA  
[0186] <213> Artificial Sequence  
[0187] <220>  
[0188] <223> UAS同源亚型序列  
[0189] <400> 23  
[0190] cggagtgccg cactccg 17  
[0191] <210> 24  
[0192] <211> 17  
[0193] <212> DNA  
[0194] <213> Artificial Sequence

[0195] <220>  
[0196] <223> UAS同源亚型序列  
[0197] <400> 24  
[0198] cggggtactg tacccccg 17  
[0199] <210> 25  
[0200] <211> 17  
[0201] <212> DNA  
[0202] <213> Artificial Sequence  
[0203] <220>  
[0204] <223> UAS同源亚型序列  
[0205] <400> 25  
[0206] cggagtgctg tactccg 17  
[0207] <210> 26  
[0208] <211> 17  
[0209] <212> DNA  
[0210] <213> Artificial Sequence  
[0211] <220>  
[0212] <223> UAS同源亚型序列  
[0213] <400> 26  
[0214] cggagtaccg tactccg 17  
[0215] <210> 27  
[0216] <211> 17  
[0217] <212> DNA  
[0218] <213> Artificial Sequence  
[0219] <220>  
[0220] <223> UAS同源亚型序列  
[0221] <400> 27  
[0222] cggggtgctg cacccccg 17  
[0223] <210> 28  
[0224] <211> 17  
[0225] <212> DNA  
[0226] <213> Artificial Sequence  
[0227] <220>  
[0228] <223> UAS同源亚型序列  
[0229] <400> 28  
[0230] cggggtaccg tacccccg 17  
[0231] <210> 29  
[0232] <211> 17  
[0233] <212> DNA

- [0234] <213> Artificial Sequence  
[0235] <220>  
[0236] <223> UAS同源亚型序列  
[0237] <400> 29  
[0238] cggggtgccg caccccg 17  
[0239] <210> 30  
[0240] <211> 17  
[0241] <212> DNA  
[0242] <213> Artificial Sequence  
[0243] <220>  
[0244] <223> UAS同源亚型序列  
[0245] <400> 30  
[0246] cggtcactg tgtgccg 17  
[0247] <210> 31  
[0248] <211> 17  
[0249] <212> DNA  
[0250] <213> Artificial Sequence  
[0251] <220>  
[0252] <223> UAS同源亚型序列  
[0253] <400> 31  
[0254] cgggtgacag ccctccg 17  
[0255] <210> 32  
[0256] <211> 17  
[0257] <212> DNA  
[0258] <213> Artificial Sequence  
[0259] <220>  
[0260] <223> UAS同源亚型序列  
[0261] <400> 32  
[0262] cggccatag tcttccg 17  
[0263] <210> 33  
[0264] <211> 17  
[0265] <212> DNA  
[0266] <213> Artificial Sequence  
[0267] <220>  
[0268] <223> UAS同源亚型序列  
[0269] <400> 33  
[0270] cggcgtctt tcgtccg 17  
[0271] <210> 34  
[0272] <211> 17



- [0273] <212> DNA  
[0274] <213> Artificial Sequence  
[0275] <220>  
[0276] <223> UAS同源亚型序列  
[0277] <400> 34  
[0278] cgggtgaccg ccctccg 17  
[0279] <210> 35  
[0280] <211> 17  
[0281] <212> DNA  
[0282] <213> Artificial Sequence  
[0283] <220>  
[0284] <223> UAS同源亚型序列  
[0285] <400> 35  
[0286] cgggtccacag tgtgccg 17  
[0287] <210> 36  
[0288] <211> 17  
[0289] <212> DNA  
[0290] <213> Artificial Sequence  
[0291] <220>  
[0292] <223> UAS同源亚型序列  
[0293] <400> 36  
[0294] cggccatata gcttccg 17  
[0295] <210> 37  
[0296] <211> 17  
[0297] <212> DNA  
[0298] <213> Artificial Sequence  
[0299] <220>  
[0300] <223> UAS同源亚型序列  
[0301] <400> 37  
[0302] cgggtgacag ccctccg 17  
[0303] <210> 38  
[0304] <211> 68  
[0305] <212> DNA  
[0306] <213> Artificial Sequence  
[0307] <220>  
[0308] <223> 组合启动子A5-A1-A6-A4序列  
[0309] <400> 38  
[0310] cggactactg tagtccgcgg agtactgtcc tccgcggacg actgtcgtcc gggagcact 60  
[0311] gtgctccg 68

- [0312] <210> 39  
[0313] <211> 68  
[0314] <212> DNA  
[0315] <213> Artificial Sequence  
[0316] <220>  
[0317] <223> 组合启动子B1-A5-A6-A4序列  
[0318] <400> 39  
[0319] cggggtgccg cccccgcgg actactgtag tccgcggacg actgtcgtcc gcgagcaact 60  
[0320] gtgctccg 68  
[0321] <210> 40  
[0322] <211> 68  
[0323] <212> DNA  
[0324] <213> Artificial Sequence  
[0325] <220>  
[0326] <223> 组合启动子B1-D5-D2-C1序列  
[0327] <400> 40  
[0328] cggggtgccg cccccgcgg gtgaccgccc tccgcgggtg acagccctcc gcgagtgcc 60  
[0329] gcactccg 68  
[0330] <210> 41  
[0331] <211> 68  
[0332] <212> DNA  
[0333] <213> Artificial Sequence  
[0334] <220>  
[0335] <223> 组合启动子D5-A1-D2-C1序列  
[0336] <400> 41  
[0337] cgggtgaccg ccctccgcgg agtactgtcc tccgcgggtg acagccctcc gcgagtgcc 60  
[0338] gcactccg 68  
[0339] <210> 42  
[0340] <211> 85  
[0341] <212> DNA  
[0342] <213> Artificial Sequence  
[0343] <220>  
[0344] <223> 组合启动子A5-A1-A6-A4-A14序列  
[0345] <400> 42  
[0346] cggactactg tagtccgcgg agtactgtcc tccgcggacg actgtcgtcc gcgagcaact 60  
[0347] gtgctccgcg gatgactgtc atccg 85  
[0348] <210> 43  
[0349] <211> 85  
[0350] <212> DNA

- [0351] <213> Artificial Sequence  
[0352] <220>  
[0353] <223> 组合启动子B1-A5-A6-A4-A14序列  
[0354] <400> 43  
[0355] cggggtgccg cccccgcgg actactgtag tccgcggacg actgtcgtcc gcgagcact 60  
[0356] gtgctccgcg gatgactgtc atccg 85  
[0357] <210> 44  
[0358] <211> 85  
[0359] <212> DNA  
[0360] <213> Artificial Sequence  
[0361] <220>  
[0362] <223> 组合启动子B1-D5-D2-C1-C3序列  
[0363] <400> 44  
[0364] cggggtgccg cccccgcgg gtgaccgccc tccgcgggtg acagccctcc gcgagtgcc 60  
[0365] gcaactccgcg gactgctgta ctccg 85  
[0366] <210> 45  
[0367] <211> 85  
[0368] <212> DNA  
[0369] <213> Artificial Sequence  
[0370] <220>  
[0371] <223> 组合启动子 D5-A1-D2-C1-C3序列  
[0372] <400> 45  
[0373] cgggtgaccg ccctccgcg agtactgtcc tccgcgggtg acagccctcc gcgagtgcc 60  
[0374] gcaactccgcg gactgctgta ctccg 85  
[0375] <210> 46  
[0376] <211> 765  
[0377] <212> DNA  
[0378] <213> Artificial Sequence  
[0379] <220>  
[0380] <223> Ni-scFv基因同源亚型序列  
[0381] <400> 46  
[0382] atgtccgttc caaccaggt tctcggcctc ctactgctgt ggctcaccga tgccagatgc 60  
[0383] caggtgcaat tggtagagtc cgggggcggg gtggtgcagc cagggcgatc tttgcggtta 120  
[0384] gactgcaagg catccggaat aaccttttct aatagcggga tgcattgggt gagacaagca 180  
[0385] cccgaaaag gccttgaatg ggtagcagtc atatggtatg acggtagcaa acgctactat 240  
[0386] gcagactcgg tcaagggtcg attcactatt agtcgcgata acagcaagaa taccttgttc 300  
[0387] ctgcagatga actcactacg ggctgaagac acagctgtct attactgtgc aaccaacgac 360  
[0388] gattattggg gacaggggac cttggtcacc gtcagttccg ggggaggggg atccggagge 420  
[0389] ggcggctcag gaggaggtgg ctccgagatc gtactgactc agagcccggc aactttgtct 480

[0390] ctgtcgctg gcgagcgggc cacactgtcc tgtagggcct cacagagcgt gagttcatat 540  
 [0391] ctagcttgggt atcaacagaa gccggggcag gcccctagat tgctaatcta cgatgcaagt 600  
 [0392] aacagagcaa ctggcatccc cgcgagattt agcggatccg gttccggaac cgactttaca 660  
 [0393] ctactatct cctccctaga accagaagat tttgcagtct actattgcca acagtcgtcc 720  
 [0394] aactggcctc gcacatttgg gcaggggaacc aaggttgaaa ttaa 765  
 [0395] <210> 47  
 [0396] <211> 765  
 [0397] <212> DNA  
 [0398] <213> Artificial Sequence  
 [0399] <220>  
 [0400] <223> Ni-scFv基因同源亚型序列  
 [0401] <400> 47  
 [0402] atgtccgttc ccaccaggt gctcggcctc ctattactct ggctaactga tgcccgggtg 60  
 [0403] caggtgcagc tcgtggagag tggaggggggt gtggttcagc cgggtcggtc actgcggctg 120  
 [0404] gactgtaaag ctagecgtat cacgttcagt aactcaggta tgcactgggt acggcaggcc 180  
 [0405] cccggcaaag gcttggagtg gtttgctgtg atatggtacg atggttctaa aaggtattat 240  
 [0406] gctgattccg tgaagggccg gtttaccata tcacgcgaca actccaagaa cacgctcttc 300  
 [0407] ctccagatga attcactccg agcgggaagac accgcggttt attattgcgc caccaatgat 360  
 [0408] gactactggg gccagggcac cttggttacc gtgtctagcg gagggggggg ctcggggggc 420  
 [0409] ggcggtagcg gtggaggtgg gtccgagatc gtcctcacgc aatccccgc cacccttagt 480  
 [0410] ctacgccctg gggagcgggc aacccttagc tgccggcgt cccagtcagt cagttcctac 540  
 [0411] cttgcctggt accagcaaaa gcccgccag gcacctgcc tccttattta tgatgcatcg 600  
 [0412] aaccgagcaa ccggaattcc tgcgcggttc agtggttctg gtagcgggac cgacttcaca 660  
 [0413] cttacaatat ctagecctaga accagaagac ttcgctgtct actactgcca acagagctcg 720  
 [0414] aactggccta gaacattcgg gcagggcacc aaggtagaaa tcaa 765  
 [0415] <210> 48  
 [0416] <211> 765  
 [0417] <212> DNA  
 [0418] <213> Artificial Sequence  
 [0419] <220>  
 [0420] <223> Ni-scFv基因同源亚型序列  
 [0421] <400> 48  
 [0422] atgagcgtgc ccaccaggt gctgggcctg ctgctgctgt ggctgaccga cgccaggtgc 60  
 [0423] caggtgcagc tgggtggagag cggcggcggc gtgggtcagc ccggcaggag cctgaggctg 120  
 [0424] gactgcaagg ccagcggcat caccttcagc aacagcggca tgcactgggt gaggcaggcc 180  
 [0425] cccggcaagg gcctggagtg gttggccgtg atctggtacg acggcagcaa gaggtactac 240  
 [0426] gccgacagcg tgaagggcag gttcaccatc agcagggaca acagcaagaa caccctgttc 300  
 [0427] ctgcagatga acagcctgag ggccgaggac accgccgtgt actactgcgc caccaacgac 360  
 [0428] gactactggg gccagggcac cctggtgacc gtgagcagcg gcggcggcgg cagcggcggc 420

[0429]	ggcggcagcg gcggcggcgg cagcgagatc gtgctgacce agagccccgc caccctgagc	480
[0430]	ctgagccccg gcgagagggc caccctgagc tgcagggcca gccagagcgt gagcagctac	540
[0431]	ctggcctggt accagcagaa gcccgccag gccccaggc tgctgatcta cgacgccagc	600
[0432]	aacagggcca ccggcatccc cgccaggttc agcggcagcg gcagcggcac cgacttcacc	660
[0433]	ctgaccatca gcagcctgga gcccgaggac ttcgccgtgt actactgcca gcagagcagc	720
[0434]	aactggccca ggaccttcgg ccagggcacc aaggtggaga tcaag	765
[0435]	<210> 49	
[0436]	<211> 780	
[0437]	<212> DNA	
[0438]	<213> Artificial Sequence	
[0439]	<220>	
[0440]	<223> At-scFv基因同源亚型序列	
[0441]	<400> 49	
[0442]	atgtccgtcc ctactcaggt actaggcctc ctcttgctat ggctgaccga cgctagatgc	60
[0443]	gaagtgaac tggtcgaatc cggcgggtggc ttggtccagc caggcggatc attacgcctg	120
[0444]	tcttgtgcag catcaggctt cacctttagt gacagttgga tccattgggt ccggcaagcc	180
[0445]	ccaggcaaag ggctggaatg ggtcgcctgg attagcccat atggcggcag cacctattac	240
[0446]	gccgacagcg tcaagggcg ctttaccatt agtgctgaca caagtaagaa taccgcttat	300
[0447]	ctgcagatga atagcctgcg ggccgaagac acggctgttt actactgtgc acgacgccac	360
[0448]	tggccggcg gttttgacta ttggggacaa ggaactctcg tgacagtctc gtcgggaggc	420
[0449]	ggcggttcag gtggcggcgg ctcaggaggg gggggttctg atatacagat gacacaatcc	480
[0450]	ccttcttctc tgagcgcaag tgtgggcgat cgggtaacca tcacctgtcg ggcttcccag	540
[0451]	gacgtgagta cagccgtggc ttggtatcaa cagaagccag gcaaggcccc gaagctgcta	600
[0452]	atctacagcg ctagtcttct gtactcaggg gtgccgagcc gctttagcgg aagtggatct	660
[0453]	ggtactgact tcacacttac tatcagttct ctacagccgg aagacttcgc tacatattac	720
[0454]	tgccagcagt atctctatca ccccgctact tttggacaag gaacaaaagt tgagatcaag	780
[0455]	<210> 50	
[0456]	<211> 780	
[0457]	<212> DNA	
[0458]	<213> Artificial Sequence	
[0459]	<220>	
[0460]	<223> At-scFv基因同源亚型序列	
[0461]	<400> 50	
[0462]	atgtcagtc ccaccaggt ccttggaacta ctcttctat gtttgacaga tgcccgtgc	60
[0463]	gaagtccagt tgggtggaatc cgggggcggc cttgtacagc ccggggggag cctgagactc	120
[0464]	agctgtgccg cttcaggatt taccttctct gattcctgga tacattgggt acgtcagcc	180
[0465]	cctgggaagg gattggagtg ggtggcctgg atcagtccat acggtggctc tacgtactat	240
[0466]	gcggacagcg tcaaagggcg ctttactatt agtcagata catcgaagaa tacagcctac	300
[0467]	ctgcagatga attcattgag ggcagaggac actgccgtct attactgtgc aagaaggcac	360

[0468]	tggcccggcg gcttcgacta ttggggacag ggcaccctgg tcacagtatc ttcaggcggga	420
[0469]	ggaggttccg ggggcggcgg ctccggcggg ggaggctcag atatccagat gacacagagc	480
[0470]	cccagctctt tatcagcttc agtgggcgat cgggtcacca tcaactgtcg tgcctctcag	540
[0471]	gatgtgtcta ccgccgtggc ctggtatcaa cagaaaccgg gcaaggcccc aaaattactg	600
[0472]	atatatagtg caagcttcct gtactcagga gtcccttcac gcttctccgg ctccggcagt	660
[0473]	gggaccgact ttactctgac gatatccagt ctgcagcctg aagatttcgc tacctactac	720
[0474]	tgtcaacagt acttgtacca ccccgcaaca ttcggacagg ggaccaaaagt agaaattaaa	780
[0475]	<210>	51
[0476]	<211>	780
[0477]	<212>	DNA
[0478]	<213>	Artificial Sequence
[0479]	<220>	
[0480]	<223>	At-scFv基因同源亚型序列
[0481]	<400>	51
[0482]	atgagcgtgc ccaccaggt gctgggcctg ctgctgctgt ggctgaccga cgccaggtgc	60
[0483]	gaggtgcagc tgggtggagag cggcggcggc ctggtgcagc ccggcggcag cctgaggctg	120
[0484]	agctgcgccg ccagcggctt caccttcagc gacagctgga tccactgggt gaggcaggcc	180
[0485]	cccggcaagg gcctggagtg ggtggcctgg atcagcccct acggcggcag cacctactac	240
[0486]	gccgacagcg tgaagggcag gttcaccatc agcgcgcaca ccagcaagaa caccgcctac	300
[0487]	ctgcagatga acagcctgag ggccgaggac accgccgtgt actactgcgc caggaggcac	360
[0488]	tggcccggcg gcttcgacta ctggggccag ggcaccctgg tgaccgtgag cagcggcggc	420
[0489]	ggcggcagcg gcggcggcgg cagcggcggc ggccggcagc acatccagat gaccagagc	480
[0490]	cccagcagcc tgagcggcag cgtgggcgac aggggtgacca tcacctgcag ggccagccag	540
[0491]	gacgtgagca ccgccgtggc ctggtaccag cagaagcccg gcaaggcccc caagctgctg	600
[0492]	atctacagcg ccagcttcct gtacagcggc gtgccagca ggttcagcgg cagcggcagc	660
[0493]	ggcaccgact tcaccctgac catcagcagc ctgcagcccg aggacttcgc cacctactac	720
[0494]	tgccagcagt acctgtacca ccccgccacc ttcggccagg gcaccaaggt ggagatcaag	780

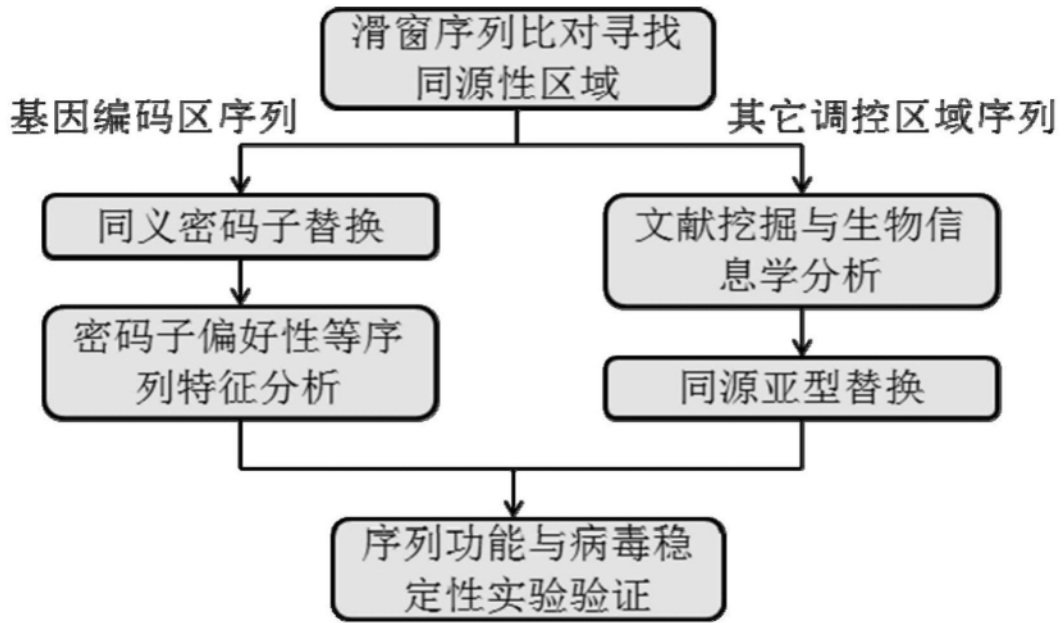


图1

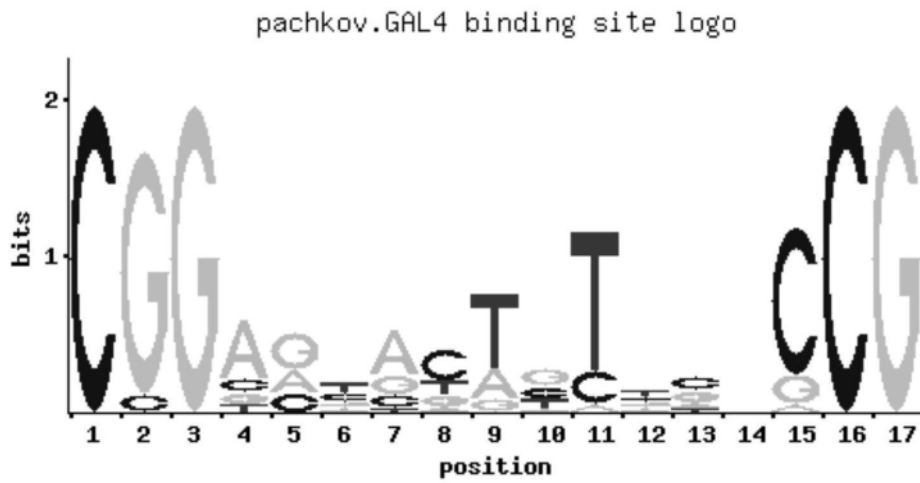


图2

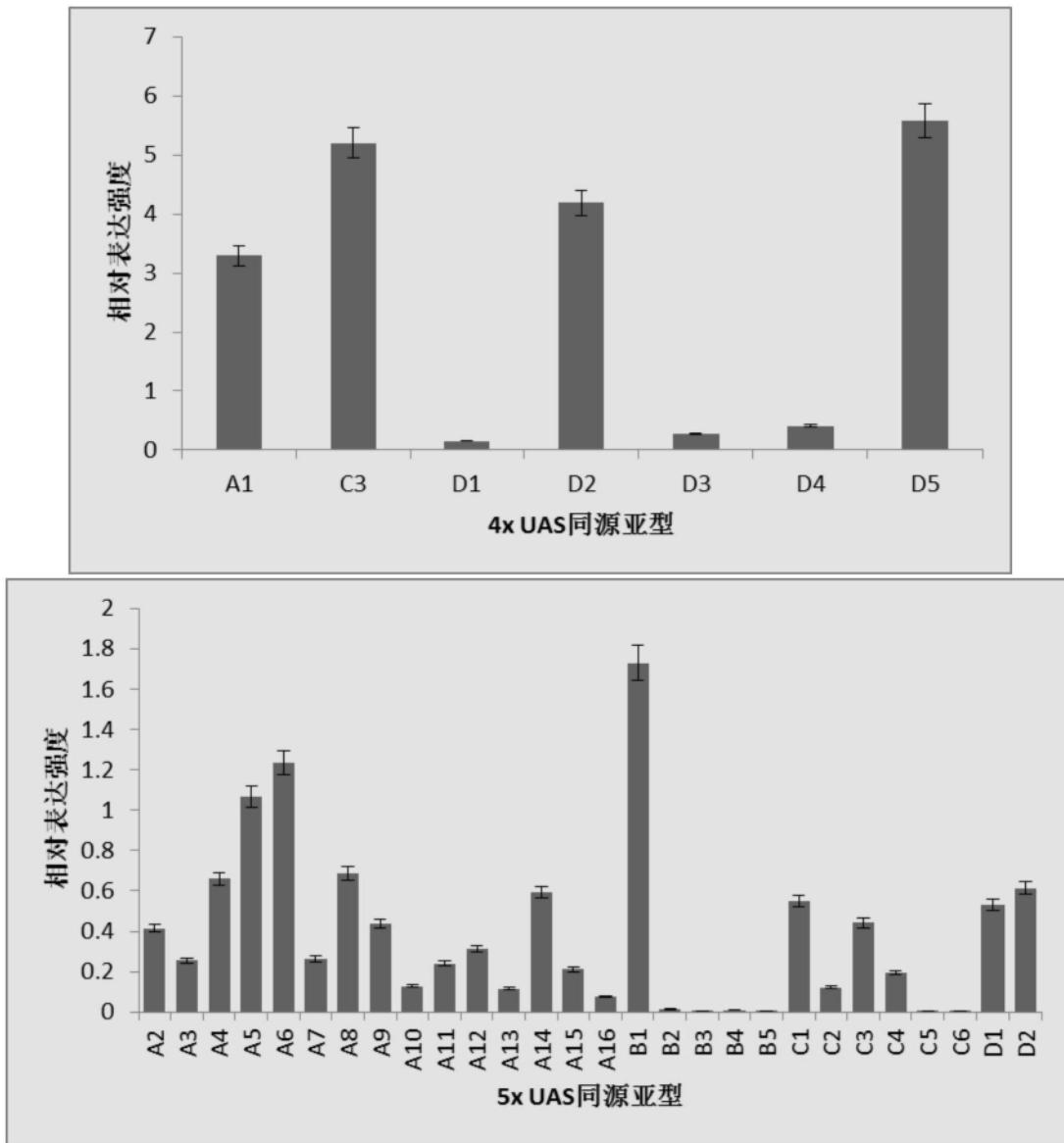


图3



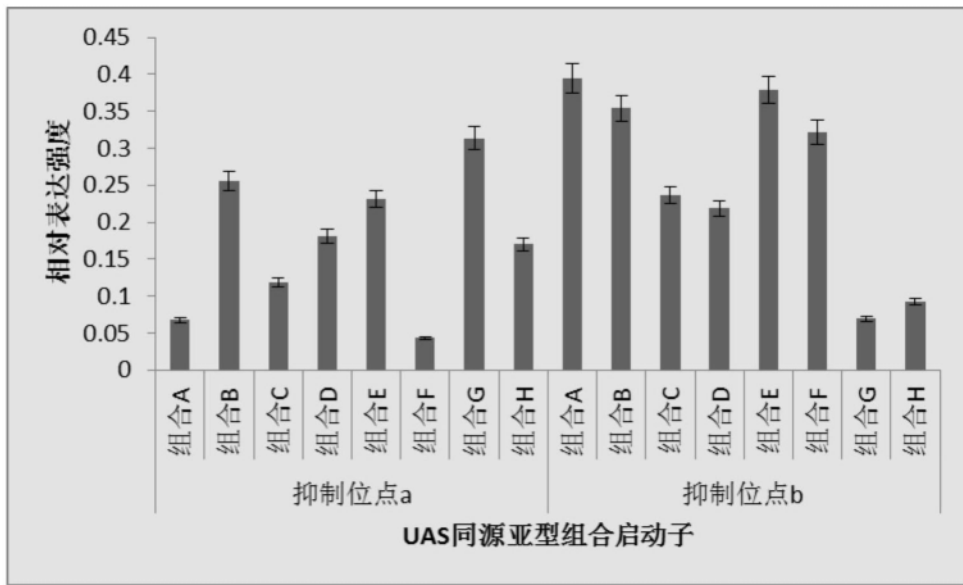


图4

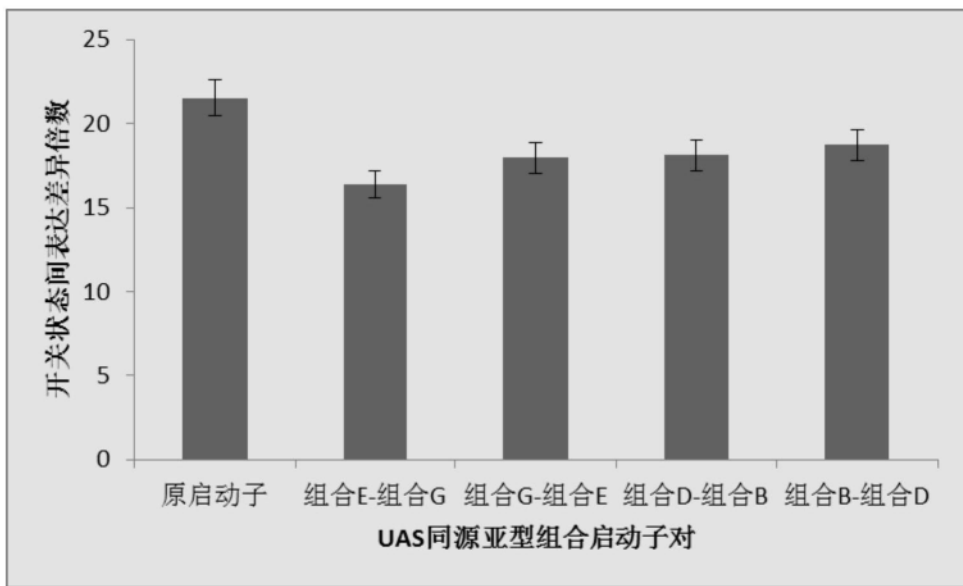


图5

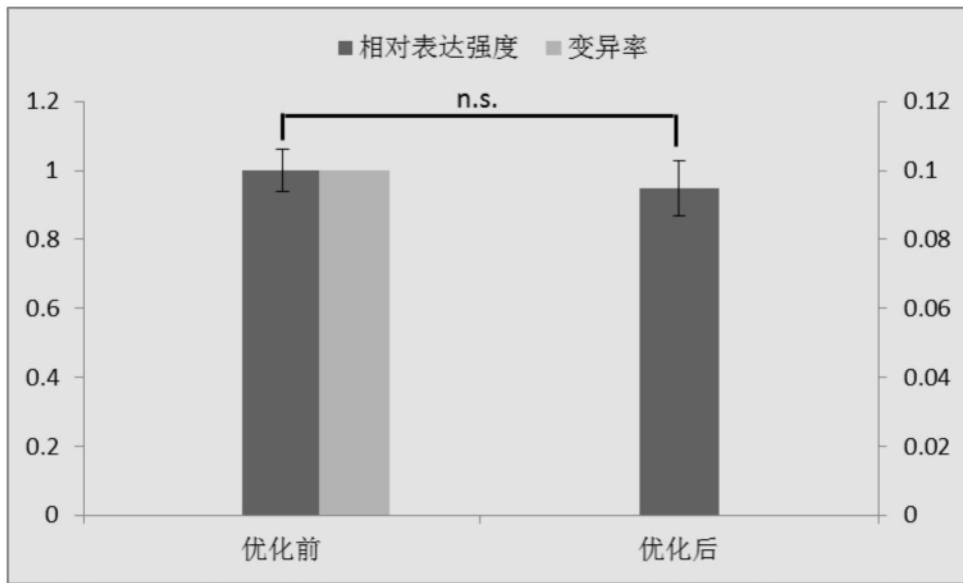


图6

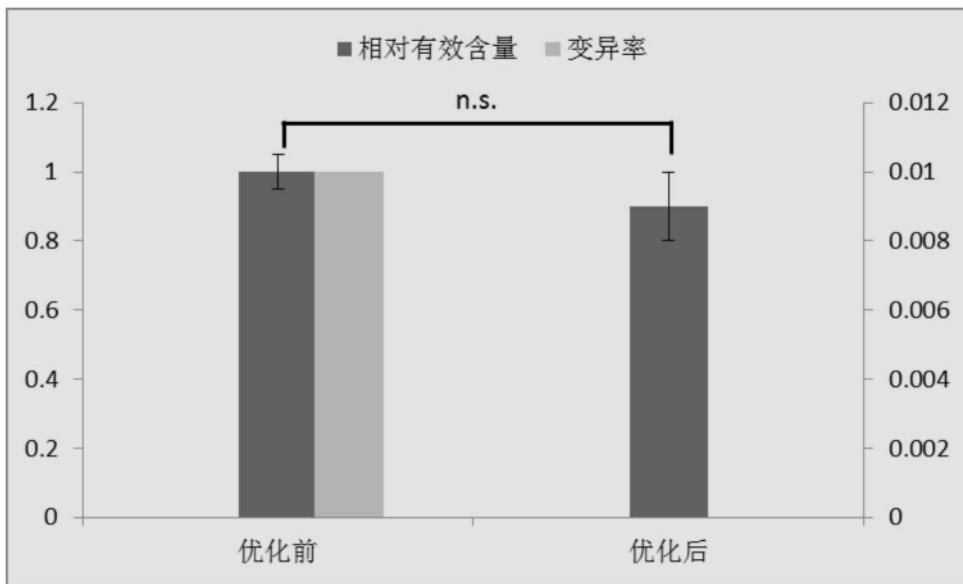


图7

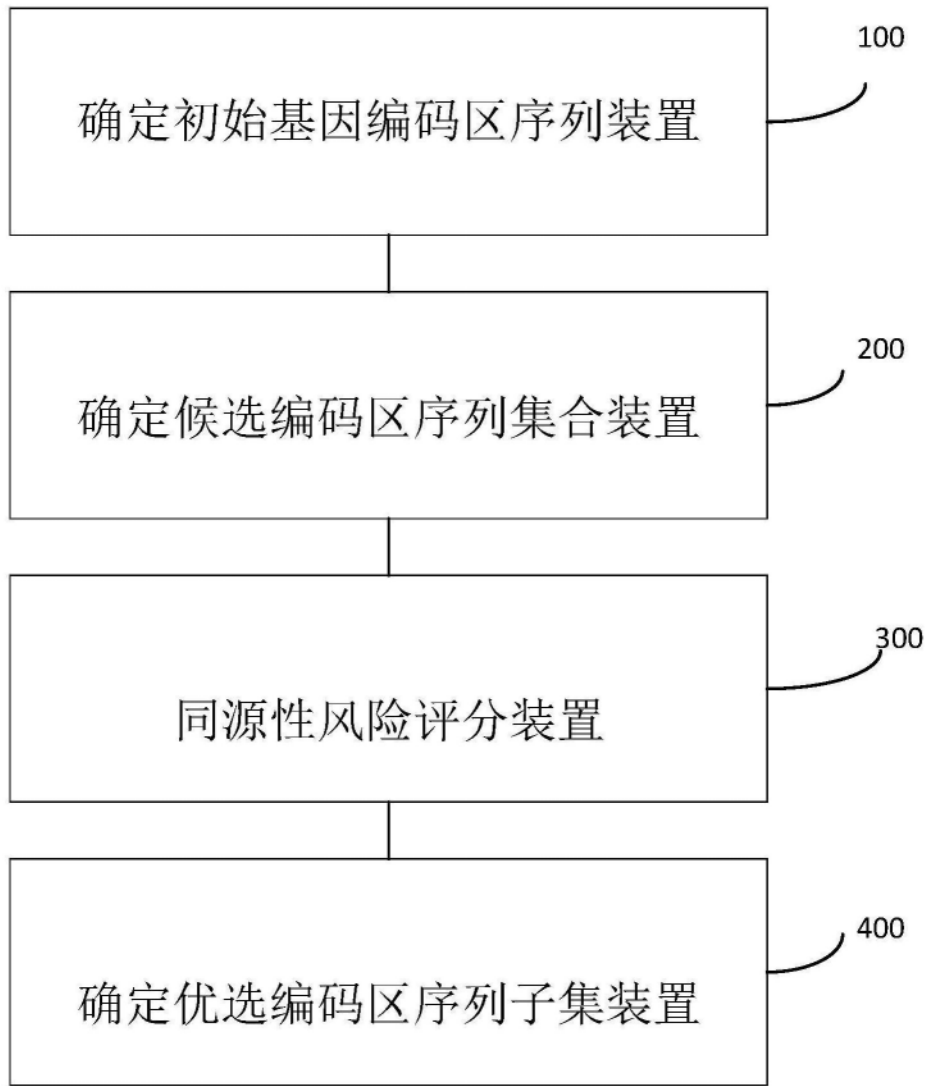


图8

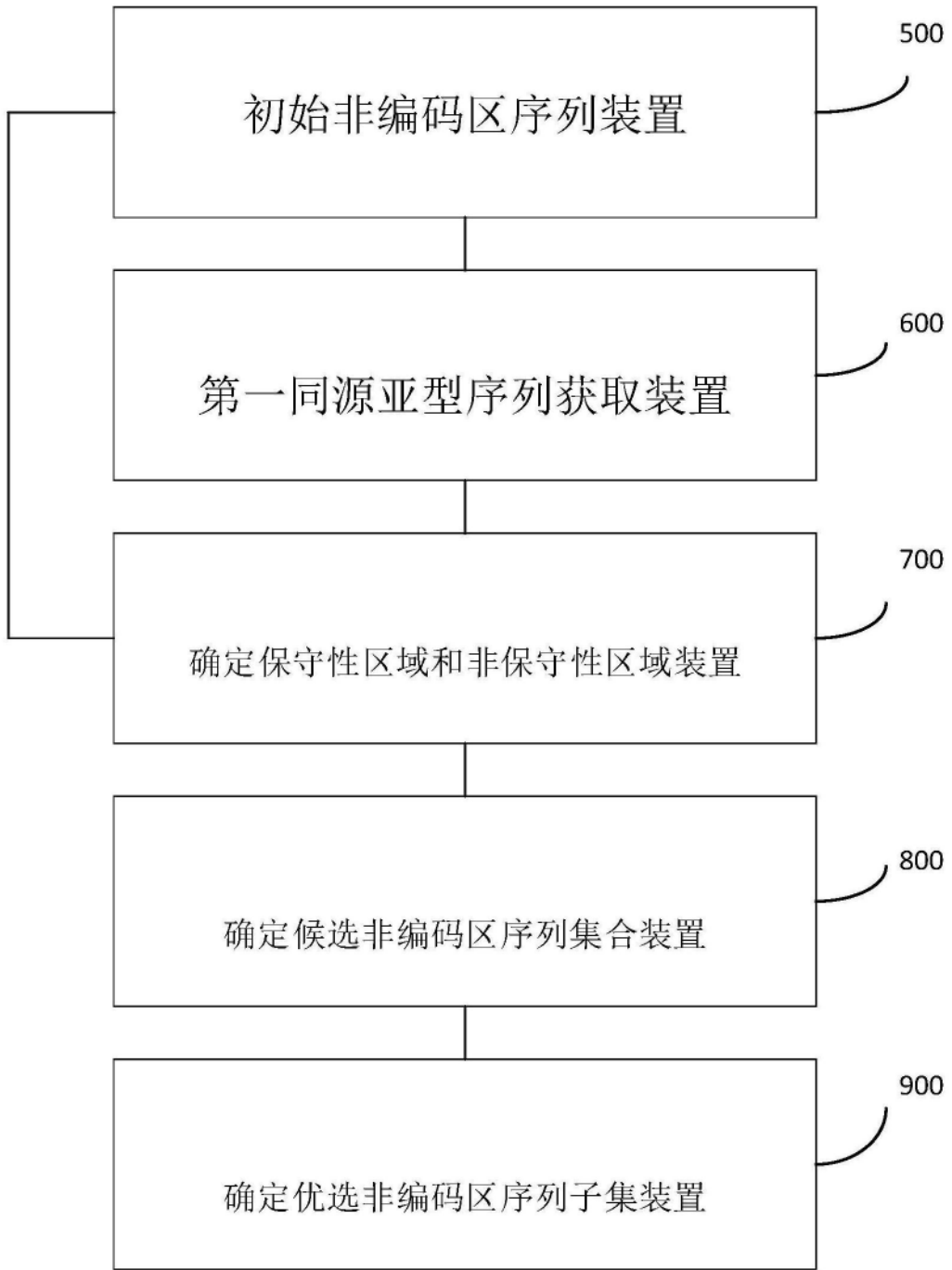


图9