



(12) 发明专利申请

(10) 申请公布号 CN 115161301 A

(43) 申请公布日 2022. 10. 11

(21) 申请号 202210719339.1

(22) 申请日 2021.03.25

(62) 分案原申请数据

202110320668.4 2021.03.25

(71) 申请人 山东大学

地址 266237 山东省青岛市即墨滨海路72号

(72) 发明人 黄启来 刘晓丹 杜平 李博

(74) 专利代理机构 济南圣达知识产权代理有限公司 37221

专利代理师 宋海海

(51) Int. Cl.

C12N 9/12 (2006.01)

C12N 15/54 (2006.01)

C12Q 1/6827 (2018.01)

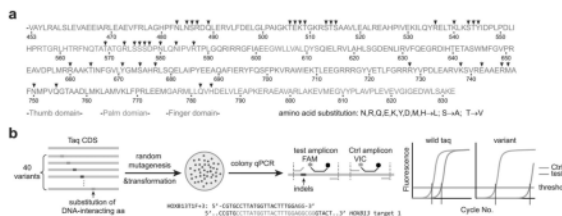
权利要求书1页 说明书18页  
序列表5页 附图9页

(54) 发明名称

一种高特异性Taq DNA聚合酶变体及其应用

(57) 摘要

本发明提供一种高特异性Taq DNA聚合酶变体及其应用,属于生物技术领域。本发明通过对野生型全长Taq DNA聚合酶进行了半理性的定向分子进化来提高其特异性。选取Taq酶上与引物/模板复合物有直接相互作用的全部极性氨基酸进行逐个突变,然后在这些变体及野生型序列的基础上进行广泛的随机诱变,生成Taq突变体文库,最终筛选出一系列具有高特异性的Taq突变体,在CRISPR/Cas9编辑效率评估和单细胞克隆基因分型中展现出了极大的优势,具有良好的实际应用之价值。



1. 一种Taq DNA聚合酶变体,其特征在于,所述Taq DNA聚合酶变体在SEQ ID NO.1所示的野生型Taq DNA聚合酶基础上进行突变,并且所述Taq DNA聚合酶变体其突变氨基酸具体为:R405Q,T569V。

2. 一种多核苷酸分子,其特征在于,所述多核苷酸分子编码权利要求1所述的Taq DNA聚合酶变体。

3. 一种重组表达载体,其特征在于,所述重组表达载体含有权利要求2所述的多核苷酸分子。

4. 一种宿主细胞,其特征在于,所述宿主细胞含有权利要求3所述的重组表达载体或染色体整合有权利要求2所述的多核苷酸分子,所述宿主细胞不包含动物细胞和植物细胞。

5. 如权利要求4所述宿主细胞,其特征在于,所述宿主细胞为原核细胞或真核细胞。

6. 一种制备权利要求1所述Taq DNA聚合酶变体的方法,其特征在于,包括步骤:培养权利要求4所述的宿主细胞,从而表达出所述的Taq DNA聚合酶变体;和分离所述的Taq DNA聚合酶变体。

7. 一种试剂盒,其特征在于,所述试剂盒包含权利要求1所述Taq DNA聚合酶变体。

8. 权利要求1所述Taq DNA聚合酶变体、权利要求2所述多核苷酸分子、权利要求3所述重组表达载体、权利要求4或5所述宿主细胞、权利要求7所述试剂盒在如下任意一种或多种中的应用:

- 1) 基因组编辑检测;
- 2) 基因突变检测。

9. 一种获得高特异性Taq DNA聚合酶变体的方法,其特征在于,所述方法包括:对野生型全长Taq DNA聚合酶进行了半理性的定向分子进化来提高其特异性;具体的,选取Taq酶上与引物/模板复合物有直接相互作用的全部极性氨基酸进行逐个突变,获得Taq变体,然后在这些变体及野生型序列的基础上进行广泛的随机诱变,生成Taq突变体文库;在qPCR筛选系统上,以基因组编辑indels质粒为模板,筛选具有高特异性的Taq突变体。

10. 如权利要求9所述的方法,其特征在于,所述方法包括:所述qPCR筛选系统使用26种在HOXB13基因上模拟indel的质粒作为模板;

优选的,所述高特异性Taq DNA聚合酶变体包含权利要求1所述的Taq DNA聚合酶变体。

## 一种高特异性Taq DNA聚合酶变体及其应用

### 技术领域

[0001] 本发明属于生物技术领域,具体涉及一种高特异性Taq DNA聚合酶变体及其应用。

### 背景技术

[0002] 公开该背景技术部分的信息仅仅旨在增加对本发明的总体背景的理解,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。

[0003] CRISPR/Cas9技术能够仅通过一小段引导RNA在特定位点进行便捷的基因组编辑,现已广泛应用于功能基因组学研究,并在涉及遗传变异的疾病治疗中具有巨大潜力。目的基因组修饰主要有三种类型,包括由于双链断裂导致的易错非同源末端连接(NHEJ)修复,该修复会引起indels随机突变;使用DNA模板进行同源介导的修复(HDR)或直接通过碱基编辑引起的精确碱基改变;以及通过招募转录因子或染色质修饰因子进行基因调控。对于基因组编辑应用,通常需要评估给定CRISPR靶标的编辑效率,并在某些情况下,对获得的单细胞克隆进行基因分型。目前已经开发了几种方法,包括GEF-dPCR, getPCR和(ACT-PCR),它们可在PCR扩增过程中将发生编辑修饰的DNA与野生型序列区分开。但由于Taq酶或TaqMan探针对DNA突变的鉴别能力有限,实验需要仔细优化才能得到较为准确的结果。使用修饰的荧光探针或使用比野生型Taq酶具有更好的错配选择能力的增强型DNA聚合酶变体,都可以提高PCR检测的准确性。DNA聚合酶变体能在不需要任何探针或引物修饰的情况下进行可靠的遗传变异检测,因此是提高基因变异检测准确度最经济有效的策略。

[0004] 聚合酶与引物/模板双链DNA在小沟处的相互作用对于复制起始复合物的组装至关重要,然而,这些相互作用力是高度冗余的,超过了有效DNA复制起始的最低需求,取代这些氨基酸以破坏相应的相互作用可以提高错配延伸中DNA聚合酶的选择性。基于此原理的DNA聚合酶理性进化主要集中在基序C中少数几个极性氨基酸和碱性氨基酸的替换上,比如,在12个氨基酸位点处进行功能突变并通过对分子改组产生的组合文库中进行筛选,鉴定出了选择性提高的Taq变体。但是,所有这些DNA聚合酶突变体的理性设计都是以提高3'末端单核苷酸错配延伸选择性为出发点的。然而,基因组编辑导致的插入缺失突变在很大程度上是复杂的以及不可预测的,这导致PCR检测引物和含有indel基因组DNA之间的错配类型也是极其多样化的。因此,基因组编辑研究非常需要一种新的DNA聚合酶变体,它具有更好的识别基因组修饰引起的引物-模板错配的能力,该Taq变体将使基因组编辑频率检测和单细胞克隆基因分型等实验更加准确和方便。

### 发明内容

[0005] 针对现有技术存在的问题,本发明提供一种高特异性Taq DNA聚合酶变体及其应用。对野生型全长Taq DNA聚合酶进行了半理性的定向分子进化来提高其特异性。选取Taq酶上与引物/模板复合物有直接相互作用的全部极性氨基酸进行逐个突变,获得40个Taq变体,然后在这些变体及野生型序列的基础上进行广泛的随机诱变,生成Taq突变体文库。在

我们的qPCR筛选系统上,以基因组编辑indels质粒为模板,筛选出一系列具有高特异性的Taq突变体,在CRISPR/Cas9编辑效率评估和单细胞克隆基因分型中展现出了极大的优势,因此具有良好的实际应用之价值。

[0006] 具体的,本发明涉及以下技术方案:

[0007] 本发明的第一个方面,提供一种Taq DNA聚合酶变体,所述Taq DNA聚合酶变体选自下组的一个或多个位点发生突变:S577A、W645R、I707V、R405Q、T569V、K354R、K531Q、L441M、S543A、R630W、F692Y、Y719F、M4I、D371E、V518D、A798V、G32D、D238V、W398C、N485L、I503F、R771K、E284K、I614L、T588S、L789F、G59W、V155F、K508Q、R229G、E255V、Q489L、E90K、E132Q、P369T、T513A、D151G、S515A、R741Q、A294S、A675V、E688D、V740A、G173D、L500I、R37Q、T140S、D365N、T140A、L538I、P10A、E303G、L484I、R492M、F272S、E794D、E170G、K508T、D578L、E818V、I799F、K206R、R229W、R249C、V390M、E404G、E267V、S577A、Q680H、R328M、R469C、E159D、D181H、P387L、A61T、D91N、K100E、K131N、A777V、P194H、P369T、T514V、Y719F、A118S、R435W、E708D、P6T、D177E、L252M、E465D、S699T、E135V、P316S、G422W、T385A、R137C、P685S、E818K、L828V、A414T、S515A、A600T、S36I、E171K、S576A、E57D、D222Y、H28L、E112D、L245P、R630L、L351F、L657P、P816S,其中,氨基酸残基编号采用SEQ ID NO.1(野生型Taq DNA聚合酶的氨基酸序列)所示的编号。

[0008] 所述Taq DNA聚合酶变体的氨基酸序列具有与SEQ ID NO.1相比至少80%的同源性;更优选地,具有至少90%的同源性;最优选地,具有至少95%的同源性;如具有至少96%、97%、98%、99%的同源性。

[0009] 所述Taq DNA聚合酶变体中的突变位点的数量为1-6个,进一步优选为1-4个,如1个、2个、3个或4个。

[0010] 所述Taq DNA聚合酶变体在SEQ ID NO.1所示的野生型Taq DNA聚合酶基础上进行突变,并且所述Taq DNA聚合酶变体选自下组中的突变体:

[0011]

突变体编号	突变氨基酸
Taq388	S577A,W645R,I707V
Taq92	R405Q,T569V
Taq99	K354R,K531Q
Taq393	L441M
Taq401	S543A,R630W,F692Y,Y719F
Taq506	M4I,D371E,V518D,A798V
Taq591	G32D,D238V,W398C,N485L,I503F,R771K
Taq664	E284K,I614L
Taq866	T588S,L789F
Taq9	G59W,V155F,K508Q
Taq1150	R229G,E255V,Q489L
Taq1140	E90K,E132Q,P369T,T513A
Taq761	D151G,S515A,R741Q
Taq812	A294S,A675V,E688D,V740A
Taq687	G173D,L500I
Taq808	R37Q,T140S,D365N
Taq1105	T140A,L538I
Taq1151	P10A,E303G,L484I,R492M
Taq1194	F272S,E794D

[0012]	Taq1108	E170G,K508T,D578L,E818V
	Taq1221	I799F,K206R,R229W
	Taq588	R249C,V390M,E404G
	Taq712	E267V,S577A,Q680H
	Taq1286	R328M,R469C,
	Taq1129	E159D,D181H,P387L
	Taq816	A61T,D91N,K100E,K131N,A777V
	Taq729	P194H,P369T,T514V,Y719F
	Taq1080	A118S,R435W,E708D
	Taq1312	P6T,D177E,L252M,E465D,S699T
	Taq1161	E135V,P316S,G422W
	Taq815	T385A
	Taq5	R137C, P685S, E818K, L828V
	Taq867	A414T,S515A,A600T
	Taq480	S36I,E171K,S576A
	Taq764	E57D,D222Y
	Taq926	H28L,E112D
	Taq903	L245P
	Taq1062	R630L
	Taq1201	L351F,L657P,P816S

[0013] 上述表格中的Taq DNA聚合酶变体是按照特异性从高到底排序,其中排在前十的变体为优秀变体,与野生型Taq相比,它们检测indels错配的Ct值至少多出7个循环,表示这些变体的选择性显著提高,其中突变体Taq388拥有最好的选择性,提高了约23个循环。同时,Taq388变异对源于indel和单核苷酸变异错配的PCR选择性均有极其显著的提高。在应用中,该Taq变体显著提高了getPCR方法对单细胞克隆基因分型的准确性,同时也使AS-qPCR SNP基因分型成为更可行的方法。

[0014] 本发明的第二方面,提供一种多核苷酸分子,所述多核苷酸分子编码上述第一方面所述的Taq DNA聚合酶变体。

[0015] 本发明的第三方面,提供了一种重组表达载体,所述重组表达载体含有本发明第二方面所述的多核苷酸分子。

[0016] 具体的,所述重组表达载体通过上述多核苷酸分子有效地连接到表达载体上获得,所述表达载体为病毒载体、质粒、噬菌体、噬菌粒、黏粒、F黏粒、噬菌体或人工染色体中的任意一种或多种;病毒载体可包括腺病毒载体、逆转录病毒载体或腺伴随病毒载体,人工染色体包括细菌人工染色体(BAC)、噬菌体P1衍生的载体(PAC)、酵母人工染色体(YAC)或哺乳动物人工染色体(MAC)。

[0017] 本发明的第四方面,提供了一种宿主细胞,所述宿主细胞含有本发明第三方面所述的载体或染色体整合有本发明第二方面所述的多核苷酸分子。

[0018] 所述宿主细胞可以为原核细胞或真核细胞。

[0019] 更具体的,所述宿主细胞是细菌细胞、真菌细胞或植物细胞中的任意一种或多种;

[0020] 其中所述细菌细胞为埃希氏菌属、农杆菌属、芽孢杆菌属、链霉菌属、假单胞菌属或葡萄球菌属内的任何种;

[0021] 更具体的,所述细菌细胞为大肠杆菌(如大肠杆菌DH5 $\alpha$ )、根癌农杆菌(如GV3101)、发根农杆菌、乳酸乳球菌、枯草芽孢杆菌、蜡状芽孢杆菌或荧光假单胞菌。

[0022] 所述真菌细胞包括酵母菌。

[0023] 转基因植物包括拟南芥植株、玉米植株、高粱植株、马铃薯植株、番茄植株、小麦植株、油菜植株、油菜籽植株、大豆植株、稻植株、大麦植株或烟草植株。

[0024] 本发明的第五方面,提供了一种制备本发明第一方面所述Taq DNA聚合酶变体的方法,包括步骤:培养本发明第四方面所述的宿主细胞,从而表达出所述的Taq DNA聚合酶变体;和分离所述的Taq DNA聚合酶变体。

[0025] 本发明的第六方面,提供了一种试剂盒,所述试剂盒包含本发明第一方面所述的Taq DNA聚合酶变体。

[0026] 本发明的第七个方面,提供上述第一方面所述Taq DNA聚合酶变体、第二方面所述多核苷酸分子、第三方面所述重组表达载体、第四方面所述宿主细胞、第六方面所述试剂盒在如下任意一种或多种中的应用:

[0027] 1) 基因组编辑检测(如基于CRISPR/Cas9的基因组编辑);

[0028] 2) 基因突变检测(如单细胞克隆基因分型,和SNP基因分型分析等)。

[0029] 以上一个或多个技术方案的有益技术效果:

[0030] 上述技术方案提供了一种高特异性Taq酶变体及其在基因组编辑和基因突变检测中的应用。本发明对野生型全长Taq DNA聚合酶进行了半理性的定向分子进化来提高其特异性。选取Taq酶上与引物/模板复合物有直接相互作用的全部极性氨基酸进行逐个突变,获得40个Taq变体,然后在这些变体及野生型序列的基础上进行广泛的随机诱变,生成Taq突变体文库。在我们的qPCR筛选系统上,以基因组编辑indels质粒为模板,筛选出一系列具有高特异性的Taq突变体。其中,特异性最好的一个变体Taq388在手掌区(S577A)和手指区(W645R和I707V)发生了三个氨基酸突变,在CRISPR/Cas9编辑效率评估和单细胞克隆基因分型中展现出了极大的优势。此外,该变体在检测自然发生的遗传变异如SNP也具有优良的性能,因此具有良好的实际应用之价值。

## 附图说明

[0031] 构成本发明的一部分的说明书附图用来提供对本发明的进一步理解,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。

[0032] 图1为本发明高特异性Taq定向进化策略的图解。

[0033] (a) 参与Taq-引物/模板相互作用的40个极性氨基酸的示意图。极性氨基酸在序列上用箭头表示。(b) Taq直接进化的原理和流程图。参与DNA相互作用的40个氨基酸被单独突变,然后用易错PCR进行随机突变,并在筛选系统上对Taq变体的活性和选择性进行了评估,

筛选系统使用了26个在H0XB13基因sgRNA靶1处含有indels的构建体,并给出了检测引物和退火区序列。与野生型Taq相比,高选择性Taq变异体具有更大的测试扩增Ct值。

[0034] 图2为本发明高选择性Taq变体的筛选

[0035] (a) 使用生长在含有IPTG的LB琼脂平板中的菌落,评估40个Taq变异体的酶活性以及在区分Indel造成的错配中的选择性。Ct值为45表示不再有聚合酶扩增活性。平均值±S.E.M,n=3个技术重复。(b) 在第一轮筛选中,对随机突变文库中的1316个转化子进行聚合酶活性和选择性评价。176个转化子保持了完整的聚合酶活性,并具有更高的特异性,并突出显示。(c) 对176个转化子进行进一步的活性和选择性评估,选出了39个证实其选择性提高的转化子,并突出显示。(d) 用纯化蛋白鉴定39个Taq变异体。用箭头表示特异性最好的三个突变体。

[0036] 图3为本发明Taq388对indel变异的选择性扩增能力分析。

[0037] (a) 基于TaqMan探针的qPCR系统中,Taq388在qPCR反应种对H0XB13基因上模拟indels突变混合物造成的引物-模板错配的选择性评价。(b) Taq388在SYBR Green qPCR体系中识别、选择上述indels能力评价。

[0038] 图4为本发明Taq388识别单核苷酸错配的能力。

[0039] (a) Taq变异体对位于引物3'末端最后一个核苷酸的引物-模板错配的敏感性评价,给出了引物和模板的序列。用匹配的模板计算相对PCR信号为100%。平均值±S.E.M,n=3个独立的技术重复。(b) 利用位于引物3'末端的倒数第二个核苷酸的引物-模板不匹配来评价Taq变异体的敏感性。平均值±S.E.M,n=3个独立的技术重复。(c-d) 在MCF7 (C/C) (c) 和T-47D (T/T) (d) 基因组DNA的等位基因特异性qPCR分析中,Taq388区分乳腺癌风险SNP rs4808611的不同等位基因的能力。

[0040] 图5为本发明Taq388在getPCR检测基因组编辑中的应用。

[0041] (a-b) 比较Taq388和野生型Taq在qPCR扩增种对H0XB13基因上26个不同的indels的识别能力,TaqMan探针法 (a) 或SYBR green法 (b) 检测携带每个Indel的质粒。(c) 比较Taq388和野生型Taq在H0XB13基因sgRNA靶2进行基因组编辑的Lenti-X 293T单细胞克隆的基因分型分析。所有20个克隆均含有先前确定的双等位基因插入缺失突变。(d) 在DYRK1A基因sgRNA靶1处进行基因组编辑的Lenti-X 293T单细胞克隆的基因分型中,比较Taq388和Taq的特异性。所有编辑后的克隆都是双等位基因indel变异,Sanger测序证实了这一点。检测引物中的观察碱基突出显示,PAM序列“NGG”显示为浅色。Ct值越大,酶的选择性越好。Ct值为45表示无扩增信号。(平均值±S.E.M,n=3个独立的技术重复)。

[0042] 图6为本发明Taq变体在SNP基因分型中的应用。

[0043] (a-e) 使用Taq388通过qPCR对30个基因组DNA样本上的5个SNP位点rs2236007 (a)、rs4808611 (b)、rs11055880 (c)、rs2290203 (d) 和rs2046210 (e) 进行的基因分型,并与野生型Taq进行比较。使用公式:等位基因1% =  $2^{-Ct(\text{allele1})} / (2^{-Ct(\text{allele1})} + 2^{-Ct(\text{allele2})})$  计算每个等位基因的百分比含量。轴上的点是纯合基因型,轴之间的点是杂合基因型。Taq388能成功区分每个基因型,但野生Taq因其特异性差而无法判定样品的基因型。(f-j) Taq388和野生型Taq对5个SNP的等位基因特异性qPCR分析的终点荧光散点图。原点附近的灰点是用于对照的无模板扩增样品。

[0044] 图7为本发明高特异度Taq的进化。



[0045] (a) Sanger测序确定的39个Taq变异体的氨基酸突变,阴影标示的克隆为10个选择性最好的变体。(b)对从大肠杆菌中表达和纯化的39个Taq突变体进行SDS-PAGE分析。(c)PCR扩增过程中野生型Taq和Taq388的突变频率,通过Sanger测序分析确定。将Taq388变异体扩增出的Taq编码序列克隆到质粒中,对每个Taq突变体的20个单细胞克隆进行测序,以鉴定突变。(d)利用Taq388和野生型Taq进行PCR扩增时产生的突变类型。

[0046] 图8为本发明Taq变体对错配的敏感性。

[0047] (a-c)在T-47D细胞(G/G)和VCaP细胞(A/A)基因组DNA的等位基因特异性qPCR分析中,Taq388区分乳腺癌风险SNP rs2236007的不同等位基因的能力。以及两种肿瘤细胞系中rs2236007位点基因型的Sanger测序分析。

[0048] (d)Taq388与图中标示的五种商用qPCR检测预混产品比较区分indels的能力;Taq388与图中标记的五种商品化qPCR主体混合物比较区分rs2236007的SNP等位基因的能力。

[0049] 图9为本发明Taq388与其他提高PCR选择性策略在SNP检测中的比较。

[0050] (a)用AS-qPCR检测SW620基因组DNA中TP53-G818A的遗传变异。Taq388与3'末端带有ddC的封闭型引物进行比较。(b)通过AS-qPCR检测MDA-MB-231基因组DNA中TP53-G839A的变异。Taq388与3'末端带有ddC的封闭型引物进行比较。(c)用AS-qPCR检测SW620基因组DNA中TP53-G818A变异。将Taq388与3'末端含LNA的引物进行比较。(d)用AS-qPCR检测MDA-MB-231基因组DNA中的TP53-G839A。Taq388与LNA引物进行比较。(e)通过qPCR从MDA-MB-231细胞中扩增出TP53-G839A。将Taq388与3'末端磷酸化的封闭型引物进行比较。

[0051] 图10为本发明在端点SNP基因分型中野生Taq的评估。

[0052] (a-e)七种DNA样品的Sanger测序色谱,在对这五种样品进行qPCR SNP基因分型时表现出广泛差异化的不同等位基因含量。Sanger测序结果与qPCR结果高度一致。

## 具体实施方式

[0053] 应该指出,以下详细说明都是例示性的,旨在对本申请提供进一步的说明。除非另有指明,本文使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解相同含义。

[0054] 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申请的示例性实施方式。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。下列具体实施方式中如果未注明具体条件的实验方法,通常按照本领域技术内的分子生物学的常规方法和条件,这种技术和条件在文献中有完整解释。参见例如Sambrook等人,《分子克隆:实验手册》中所述的技术和条件,或按照制造厂商所建议的条件。

[0055] 以下通过实施例对本发明做进一步解释说明,但不构成对本发明的限制。应理解这些实施例仅用于说明本发明而不用于限制本发明的范围。

### [0056] 实施例

#### [0057] 1. 实验材料和方法

##### [0058] 1.1 Taq聚合酶的定点和随机诱变

[0059] 细菌表达Taq聚合酶所用的质粒pAKTaq (Addgene#25712) 购自Addgene网站。通过在pAKTaq的基础上进行定点突变PCR, 逐个对参与Taq酶-DNA相互作用的40个极性氨基酸进行氨基酸替换(图1a)。在20 $\mu$ l定点突变PCR反应中含有4pmol定点突变引物和10 $\mu$ l 2x Prime STAR Max Premix (TaKaRa), PCR程序为98 $^{\circ}$ C预变性15秒, 然后98 $^{\circ}$ C变性10秒, 72 $^{\circ}$ C延伸2分钟, 循环25次, 最后在72 $^{\circ}$ C延伸5分钟。PCR产物中加入FastDigest DpnI (Thermo Fisher SCIENTIFIC) 在37 $^{\circ}$ C切割2小时后直接用于转化DH5 $\alpha$ 感受态细胞, 并将其涂布于含有氨苄青霉素的LB琼脂平板, 在37 $^{\circ}$ C培养箱中倒置培养过夜。第二天挑取单克隆接种到LB培养基中, 37 $^{\circ}$ C 250rpm震荡培养过夜, 从中提取质粒后用于Sanger测序。

[0060] 将Sanger测序确认的这40个突变体等比例混合, 并将其以1:1的比例与pAKTaq混合, 作为模板使用GeneMorph II Random Mutagenesis Kit (Agilent Technologies) 通过易错PCR方法进行随机诱变。在25 $\mu$ l易错PCR反应体系中含有2.5 $\mu$ l 10xMutazyme II reaction buffer, 0.5 $\mu$ l 40mM dNTP mix, 1pmol上游和下游引物, 0.5 $\mu$ l Mutazyme II DNA polymerase (2.5U/ $\mu$ l) 和15ng模板质粒。PCR程序为95 $^{\circ}$ C预变性2分钟, 然后95 $^{\circ}$ C变性30秒, 60 $^{\circ}$ C退火30秒, 72 $^{\circ}$ C延伸3分钟, 循环10次, 最后在72 $^{\circ}$ C延伸10分钟。PCR产物通过EcoRI/SalI双酶切克隆到原表达载体中。通过单克隆Sanger测序确定转化子发生突变的频率, 我们依据产品说明书对易错PCR的模板量和循环次数进行了调整, 直到获得达到我们的要求的突变频率。

[0061] 1.2 菌落qPCR筛选高特异性Taq变体

[0062] 用随机突变文库质粒转化大肠杆菌DH5 $\alpha$ 感受态细胞, 诱导Taq突变体在含有氨苄青霉素和IPTG的LB固体培养基中表达蛋白。为了确定不同Taq变异体的活性和特异性, 我们以26个基于pcDNA3.1载体的带有模拟CRISPR/Cas9基因编辑indel的HOXB13基因质粒作为PCR模板, 采用菌落实时定量PCR方法进行筛选。在单管qPCR反应中包含两个扩增子, 即检测扩增子和对照扩增子。检测扩增子的上游引物横跨模拟的基因组编辑位点, 用来考察Taq酶对于indel造成的引物-模板错配的选择性, 检测扩增子使用一条FAM标记的TaqMan探针。对照扩增子则匹配临近未发生突变的序列, 用来衡量Taq酶变体的聚合酶活性是否受到了影响, 对应一条VIC标记的TaqMan探针, 这里使用的引物都是根据getPCR策进行设计的, 值得注意的是, 质粒用Fast Digest NotI (Thermo Science<sup>TM</sup>, CAT#FD0593) 线性化, 以避免两个探针之间的荧光信号干扰。挑取生长在含有IPTG的LB琼脂平板上的表达Taq变体的单克隆菌落, 加入10 $\mu$ L的1XTaq enzyme screening buffer (50mM Tris-HCl [pH8.8], 16mM [NH<sub>4</sub>]<sub>2</sub>SO<sub>4</sub>, 0.1% [v/v] Tween<sup>®</sup>20, 2.5 $\mu$ M MgCl<sub>2</sub>, 每个dNTP 0.25mM) 混合均匀后, 加入7 $\mu$ L到20 $\mu$ L的qPCR体系中。每个引物和探针的工作浓度分别为0.2 $\mu$ M和0.1 $\mu$ M。定量PCR程序为: 95 $^{\circ}$ C预变性5分钟, 然后95 $^{\circ}$ C变性30秒, 68 $^{\circ}$ C退火30秒, 72 $^{\circ}$ C延伸10秒, 循环45次。检测扩增子Ct值增加而对照扩增子Ct值不变的Taq变体时希望得到的特异性增加的变体。

[0063] 1.3 Taq变体的纯化

[0064] 经过两轮菌落qPCR筛选, 最终获得了39个改进的变体, 通过Sanger测序分析确定每个变体的突变氨基酸, 并在大肠杆菌中进行表达和纯化。对于每个克隆, 将其对应的100 $\mu$ l的过夜培养物, 转接到含氨苄青霉素抗性的4ml LB液体培养基中, 并在37 $^{\circ}$ C和转速为250rpm活化约4h, 当OD<sub>600 nm</sub>达到0.8时, 加入终浓度为1mM的IPTG诱导蛋白表达, 在37 $^{\circ}$ C和转速为250rpm孵育12h。5000rpm离心3min收集菌体, 用400 $\mu$ l缓冲液 (50mM Tris-HCl

[pH7.9], 50mM蔗糖, 1mMEDTA [pH8.0]) 重悬菌体沉淀, 室温下5000rpm离心3min收集菌体。用200 $\mu$ L的预裂解液 (50mMTris-HCl [pH7.9], 50mM蔗糖, 1mMEDTA [pH8.0], 4mg/mL溶菌酶 [Amresco]), 室温孵育15min。然后, 将菌体悬液放入-80 $^{\circ}$ C的冰箱中冷冻30min, 然后将其置于室温下至完全解冻。在重复之前的冻融操作一次后, 立即将此溶液放在37 $^{\circ}$ C水浴中孵育15min。然后加入1 $\mu$ L的5mg/ml DNaseI、1 $\mu$ L的1M $\text{CaCl}_2$ 和2 $\mu$ L的1MMnCl<sub>2</sub>, 混合均匀。继续在37 $^{\circ}$ C孵育30min后, 加入200 $\mu$ L的裂解缓冲液 (10mMTris-HCl [pH7.9], 50mMKCl, 1mMEDTA [pH8.0], 0.5% [v/v] Tween<sup>®</sup>20, 0.5% [v/v] NP40) 并混合均匀, 然后将该裂解液在75 $^{\circ}$ C下孵育1h, 随后在4 $^{\circ}$ C下15000rpm离心10min, 收集上清溶液。在其中加入0.12g固体(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 于4 $^{\circ}$ C下旋转孵育30min。然后该溶液在4 $^{\circ}$ C下15000rpm离心20min收集沉淀, 将其重悬在300 $\mu$ L储存缓冲液中 (50mMTris-HCl [pH7.9], 50mMKCl, 0.1mMEDTA [pH8.0], 1xPI, 0.1% [v/v] Tween<sup>®</sup>20, 50% [v/v] glycerol), 并在-20 $^{\circ}$ C下储存。

[0065] 最后用SDS-PAGE电泳检测蛋白样品中Taq突变体的含量, 即将蛋白样品加入到由12%分离胶和5%浓缩胶组成的凝胶中, 跑完电泳并用eStainTML1蛋白染色剂 (GenScript) 染色, 并用Quantum-ST5 (VILBER LOURMAT, France) 进行凝胶成像分析。

[0066] 1.4 Taq388突变体的扩增保真度分析

[0067] 为了比较Taq388和野生型Taq的保真度, 我们以质粒pAKTaq中的Taq聚合酶编码序列为模板使用10X Taq enzyme screening buffer进行PCR扩增。PCR产物用FastDigest EcoRI (Thermo) 和FastDigest SalI (Thermo) 双酶切, 然后插入同样双酶切的载体pAKTaq中。将连接产物转化大肠杆菌DH5 $\alpha$ 感受态细胞, 挑选20个单细胞克隆进行Sanger测序, 计算每个克隆中扩增子序列的突变碱基数以得到突变频率。

[0068] 1.5 GetPCR分析条件

[0069] 在基于SYBR Green的getPCR方法中, 15 $\mu$ L的反应体系里含有7.5 $\mu$ L的2x Taqbuffer, 3pmol的每个引物, 0.005ng质粒DNA或3ng基因组作为模板, 1 $\mu$ L的Taq聚合酶。在qPCR仪Rotor-Gene Q 2plex, Qiagen上进行的分析, 程序为: 95 $^{\circ}$ C初始变性5min, 95 $^{\circ}$ C变性30s, 引物退火64~70 $^{\circ}$ C, 30s, 72 $^{\circ}$ C延伸10s, 在LightCycler<sup>®</sup>96热循环仪 (Roche Applied Science, Germany) 上进行的分析则使用以下条件: 95 $^{\circ}$ C初始变性5min。

[0070] 在使用TaqMan探针的getPCR方法中, 反应体系为20 $\mu$ L, 包括2 $\mu$ L的10x Taq enzyme screening Buffer, 0.1ng的质粒DNA或10ng的基因组作为模板, 4pmol的引物和2pmol的探针, 1 $\mu$ L的Taq聚合酶。在QPCR仪 (Rotor-Gene Q 2plex, Qiagen) 中采用以下程序进行实时PCR: 在95 $^{\circ}$ C下初始变性5min, 然后在95 $^{\circ}$ C变性30s, 引物退火在64~70 $^{\circ}$ C下退火30s, 在72 $^{\circ}$ C延伸10s, 当使用LightCycler<sup>®</sup>96热循环仪 (Roche Applied Science, Germany) 时, 使用了以下条件: 初始变性循环 (95 $^{\circ}$ C, 5min), 然后是45个PCR循环 (95 $^{\circ}$ C, 15s, 64-70 $^{\circ}$ C, 15s, 72 $^{\circ}$ C, 15s)。

[0071] 1.6 Taq388在indel检测中的选择性分析

[0072] 在SYBR Green和TaqMan探针法qPCR系统中检测Taq388对于indel导致的引物-模板错配的选择性。这里使用的PCR模板是在Taq变体筛选系统中用到的26个模拟indel的质粒。这26个质粒混合在一起时模拟基因组编辑产生的indels混合物, 而每个质粒单独作为模板则代表在基因组编辑实验中分离的带有纯合indel的单细胞克隆。对于TaqMan探针法qPCR检测, 在一个20 $\mu$ L的反应体系中使用1对检测引物及1个对应的TaqMan检测探针, 1对对

照引物及1个对照TaqMan探针。SYBR Green方法有所不同,它不使用TaqMan探针,需要在两个反应管中分别进行检测扩增和对照扩增。

[0073] 在基因组编辑的实际应用场景中检测Taq388的选择性时,使用了31个经过CRISPR/Cas9基因组编辑的lenti-X293T单克隆细胞基因组DNA,其中,20个单克隆细胞为HOXB13基因发生了双等位基因编辑,11个为DYRK1A基因发生双等位基因。未发生编辑的Lenti-X 293T细胞系基因组用作两个系列的内部参照物,结合SYBR Green或TaqMan探针的QPCR用LightCycler®96仪器(Roche)进行检测(图5c,d)。这里的PCR条件和程序参见getPCR分析条件部分。

[0074] 1.7 Taq388在SNP基因分型中的应用

[0075] 检测中使用了30份基因组DNA样本,其中10个来自乳腺癌细胞系(MCF7,T47D,MDA-MB-231,BT-474,BT-20,BT-549,SK-BR-3,ZR-75-1,MDA-MB-468,MDA-MB-453),5个来自前列腺癌细胞系(LNCaP,DU 145,PC3,22Rv1,VCaP)和4个来自其他类型的细胞系(HEK293T,Jurkat,HL-60,K562),11个是来自研究者自身经过隐去个人信息处理的基因组DNA。PCR反应中使用了针对5个SNP位点(rs2046210[C/T]、rs2290203[C/T]、rs11055880[C/T]、rs4808611[C/T]和rs2236007[GA/CT])设计的等位基因特异性引物。在qPCR进行SNP基因分型分析时,一方面我们依据qPCR得到的等位基因特异性的Ct值,来计算在样品中该位点处每个等位基因的百分比含量,据此确定其基因型,以rs4808611为例,从qPCR反应中获得C等位基因特异性引物以及T等位基因特异性引物的Ct值,然后使用公式分别计算两个等位基因的比例,C等位基因 $[C\% = 2^{-Ct(C)} / (2^{-Ct(C)} + 2^{-Ct(T)})]$ 和T等位基因 $[T\% = 2^{-Ct(T)} / (2^{-Ct(C)} + 2^{-Ct(T)})]$ 的比值;另一方面,我们可以将被测等位基因的荧光值直接绘制成散点图,直观地显示这些细胞系的基因型。这里的PCR条件和程序参见getPCR分析条件部分。作为对比,五种商品化产品也在rs2236007位点的基因分型中进行了应用,它们是2x Ultra SYBR Mix、THUNDERBIRD SYBR qPCR Mix、SYBR®Select Master Mix、Life Power和2x T5 Fast qPCR,每种商品的扩增条件参照各自的产品说明书进行。

[0076] 1.8封闭型引物或LNA引物的PCR

[0077] 3'端含有ddC或磷酸基团的封闭引物和LNA引物可用来提高等位基因扩增的选择性,我们针对SW620细胞基因组含有的纯合TP53-G818A位点和MDA-MB-231细胞基因组上含有的TP53-G839A位点,设计等位基因特异性引物、对照扩增引物以及阻断引物,对它们提高PCR选择性进行了评估。在15 $\mu$ l qPCR反应体系中含有1xTaqbuffer,3pmol上、下游引物,以及0.005ng带有变异位点的PCR产物作为模板,PCR扩增程序为95 $^{\circ}$ C预变性5分钟,然后45个循环的95 $^{\circ}$ C15s,68 $^{\circ}$ C15s,72 $^{\circ}$ C15s,最后跟着一个标准的融解曲线程序。

[0078] 2. 结果

[0079] 2.1高特异性Taq定向进化的合理设计

[0080] 尽管5'核酸外切酶缺失的大片段(KlenTaq)能提高保真度和热稳定性,但是为了使最终的DNA聚合酶变体同时适用于基于SYBR Green和基于TaqMan探针的qPCR分析,我们选择了全长的Thermus aquaticus(Taq)DNA聚合酶(SEQ ID NO.1)而不是KlenTaq作为分子进化的起始分子。科研人员认识到,替换掉与引物/模板复合物直接相互作用的氨基酸或影响结合口袋几何形状的氨基酸可以改变聚合酶的选择性。在以前的研究中,研究人员仅选择了一部分接触引物/模板的氨基酸进行突变。在本研究中,为了选出用于理性设计的候选

氨基酸,我们调查了DNA聚合酶开放形式和闭合形式的晶体结构,并选择了所有与引物/模板双链直接接触的40个极性氨基酸作为突变的靶标(图1a)。其中17个残基与引物链接触,24个残基与模板链接触,1个残基Arg573同时与两者接触。针对这些选出的氨基酸,我们首先进行了定点突变,用含有非极性侧链的亮氨酸、丙氨酸或缬氨酸取代了40个极性氨基酸残基,同时尽量保持它们的空间几何形状不变。具体地说,将氨基酸N、R、Q、E、K、Y、D、M和H替换为L,将S和T分别替换为A和V(见下表)。由于通常氨基酸的极性侧链是直接参与接触的基团,非极性氨基酸残基的取代将有效破坏相应的相互作用,从而使Taq聚合酶对引物/模板错配更加敏感,因此有望提高聚合酶在错配延伸方面的选择性。

[0081] 我们直接将生长在含有IPTG的LB琼脂平板上的转化子用于高通量筛选,无需复杂的蛋白质纯化操作。首先在基于TaqMan探针的菌落qPCR系统上评估了40个Taq变体的活性和选择性,该筛选系统使用26种在HOXB13基因上模拟indel的质粒作为模板。在该系统中,我们在一个反应管中设计了两个扩增子,其一是用来评估聚合酶选择性的检测扩增子,这里检测引物可退火到野生型DNA序列上,这是发生基因组编辑产生Indels的区域;另一个是用来评价聚合酶活性的对照扩增子,扩增引物退火到邻近区域(图1b)。26个indels会导致与检测引物的各种不匹配,与野生型Taq相比,检测扩增子Ct值的增加可以表明突变体选择性的增强。同时,如果对照扩增子Ct值保持不变,则表明被测试的Taq突变体活性不受突变的影响。

[0082] 我们发现其中9个变体严重丧失了聚合酶活性,包括R536L, Y545L, R573L, N580L, N583L, Y671L, N750L, Q754L和H784L。与野生型Taq相比,19个变体显示出更好的选择性,具有统计学意义,其中8个变体比野生型Taq多出5个循环,表明这几个变体具有更好的选择性(图2a)。但是,即使保留完整活性并具有最高选择性的变体T206V也只能提高13.9个循环,还存在很大的局限性。

[0083]

编号	突变氨基酸
1	Taq-N483L
2	Taq-N485L
3	Taq-S486A
4	Taq-R487L
5	Taq-Q489L
6	Taq-T506V
7	Taq-E507L
8	Taq-K508L
9	Taq-T509V
10	Taq-S513A
11	Taq-T514V
12	Taq-S515A
13	Taq-R536L
14	Taq-K540L
15	Taq-S543A
16	Taq-T544V
17	Taq-Y545L
18	Taq-T569V
19	Taq-R573L
20	Taq-S575A
21	Taq-S576A
22	Taq-S577A
23	Taq-D578L
24	Taq-N580L
25	Taq-N583L
26	Taq-R587L
27	Taq-R660L
28	Taq-Q782L
29	Taq-H784L
30	Taq-T664V
31	Taq-Y671L
32	Taq-S674A

[0084]

33	Taq-R677L
34	Taq-R728L
35	Taq-K738L
36	Taq-E742L
37	Taq-R746L
38	Taq-M747L
39	Taq-N750L
40	Taq-Q754L

[0085] 2.2高选择性Taq酶的广泛诱变分子进化

[0086] 进一步,我们在这40个变异体以及野生型Taq的基础上做了广泛的随机突变,以筛选特异性更好的Taq变体。用GeneMorph II随机突变试剂盒将野生型Taq表达载体与40个变异体混合后进行易错PCR,该试剂盒能以最小的突变偏倚引入合理水平的突变率。对于通过随机突变进行的定向蛋白质进化,通常每个构建体有2-7个核苷酸突变,对应1-3个氨基酸突变。通过调整输入模板量和循环次数,我们得到了Taq基因编码区上含平均5.3个突变的Taq突变体文库。然后将易错PCR产物克隆到原核表达质粒pAKTaq中,并将直接将生长在含有IPTG的LB琼脂平板上的单细胞菌落应用到qPCR筛选系统进行筛选。

[0087] 我们总共筛选了1316个克隆(图2b),其中1001个克隆(76.1%)的扩增曲线在x轴上向右偏移并超过5个循环表示它们失去了大部分或全部的聚合酶活性,有101个克隆(7.7%)不仅保持了完整的活性,还表现出极高的选择性,甚至对于检测indel错配的扩增反应完全没有扩增信号。为了进一步证实这些高选择性Taq变体的特异性,除了101个克隆外,我们扩大了范围,额外选择了75个克隆,这些克隆满足 $Ct(Contr1) < 14.5$ 和 $Ct(Test) > 30$ 的标准(图2c中的色点)。这次,我们在含有LB琼脂平板的IPTG上划线,收集直径大于2mm的菌落,并在qPCR筛选系统中进行了评估。我们发现只有62个菌落(35.2%)仍然符合 $Ct(Contr1) < 14.5$ 和 $Ct(Test) > 30$ 的高特异性标准,这可能反映之前菌落qPCR体系稳定性较差。此时,我们选择了39个符合较高标准( $Ct(Contr1) < 14.5$ 和 $Ct(Test) > 40$ )的克隆进行Sanger测序,并对这些Taq酶变体(见下表)在大肠杆菌中进行了蛋白表达和纯化,进一步用纯化的Taq聚合酶(图2c中的圆点)进一步验证。有趣的是,我们发现这39个变体中,只有13个变体的氨基酸替换涉及了Taq聚合酶和引物/模板复合物之间的直接接触(图7a)。

[0088]

突变体编号	突变氨基酸
Taq388	S577A,W645R,I707V
Taq92	R405Q,T569V
Taq99	K354R,K531Q
Taq393	L441M
Taq401	S543A,R630W,F692Y,Y719F
Taq506	M4I,D371E,V518D,A798V
Taq591	G32D,D238V,W398C,N485L,I503F,R771K

[0089]

Taq664	E284K,I614L
Taq866	T588S,L789F
Taq9	G59W,V155F,K508Q
Taq1150	R229G,E255V,Q489L
Taq1140	E90K,E132Q,P369T,T513A
Taq761	D151G,S515A,R741Q
Taq812	A294S,A675V,E688D,V740A
Taq687	G173D,L500I
Taq808	R37Q,T140S,D365N
Taq1105	T140A,L538I
Taq1151	P10A,E303G,L484I,R492M
Taq1194	F272S,E794D
Taq1108	E170G,K508T,D578L,E818V
Taq1221	I799F,K206R,R229W
Taq588	R249C,V390M,E404G
Taq712	E267V,S577A,Q680H
Taq1286	R328M,R469C,
Taq1129	E159D,D181H,P387L
Taq816	A61T,D91N,K100E,K131N,A777V
Taq729	P194H,P369T,T514V,Y719F
Taq1080	A118S,R435W,E708D
Taq1312	P6T,D177E,L252M,E465D,S699T
Taq1161	E135V,P316S,G422W
Taq815	T385A
Taq5	R137C, P685S, E818K, L828V
Taq867	A414T,S515A,A600T
Taq480	S36I,E171K,S576A
Taq764	E57D,D222Y
Taq926	H28L,E112D
Taq903	L245P
Taq1062	R630L
Taq1201	L351F,L657P,P816S

[0090] 2.3 Taq变体的纯化及其选择性的验证

[0091] 如上所述,我们在大肠杆菌中表达并纯化了这39个特异性提高的Taq变体。它们在



SDS-PAGE分析中表现出相似的纯度,表观分子量均为94kDa(图7b)。我们在qPCR筛选系统中评估了这些变体在indels检测系统中的聚合酶活性和选择性,最后鉴定出10个优秀变体,与野生型Taq相比,它们检测indels错配的Ct值至少多出7个循环,表示这些变体的选择性显著提高( $P < 0.05$ ) (图2d中的色点),其中突变体Taq388拥有最好的选择性,提高了约23个循环,在随后的实验中我们选择使用了该变体进行了系统的评估和应用。

[0092] 随后,我们通过Sanger测序来评价Taq388变异体在PCR扩增中的保真度。用Taq388扩增Taq编码序列,并将其克隆到原载体中,转化进大肠杆菌后,挑取单克隆进行Sanger测序分析由于PCR扩增产生的DNA突变。我们发现Taq388的保真度提高了4.7倍(图7c)。值得注意的是,野生型Taq发生了3种类型的突变,包括56.5%的转换、39.1%的颠换和4.4%的缺失,而Taq388只产生了转换类型的突变(图7d)。简而言之,我们获得了多个增强型Taq酶变异体,它在扩增indel引起的引物/模板错配时具有显著增强的选择性,并且在PCR扩增中保真度也提高了4.7倍。

[0093] 2.4增强型Taq鉴别错配的能力

[0094] 然后,我们系统地评估了Taq388变体对各种类型的引物/模板错配的区别能力。首先,在基于TaqMan探针的qPCR筛查系统上测试了其区分indels错配的能力。结果表明,Taq388比野生型Taq聚合酶的选择性提高了23个循环,这在筛选过程中已经得到了体现(图3a)。当在基于SYBR Green的qPCR系统中使用相同的引物和模板进行测试时,该变体区分Indels错配的能力也大大提高,但程度低于基于TaqMan探针的系统(图3b)。进一步,我们系统地研究了该变体在引物3'末端最后一个或倒数第二个位置上识别单核苷酸错配的能力。为了产生单核苷酸错配,我们构建了在HOXB13 c.251G位置含有三种类型的单核苷酸变异的质粒作为qPCR模板,这三种类型包括c.251G>A、c.251G>T、c.251G>C(图4a、b)。我们使用4种仅3'末端核苷酸有差异的引物进行基于SYBR green的qPCR分析,发现与野生型Taq相比,Taq388聚合酶变体在所有12种错配类型中都大大降低了来自错配模板的扩增信号(图4a)。类似地,使用具有不同的3'末端倒数第二位核苷酸的引物进行qPCR分析表明,Taq388变体也比野生型Taq在引物3'末端的倒数第二位错配时具有更高的选择性(图4b)

[0095] 接下来,我们在基因组DNA的实际应用场景中评估了Taq变体对单核苷酸错配的扩增选择性。我们使用3'末端靶向rs4808611位点的等位基因特异性引物,对SNP位点基因型分别为C/C和T/T的MCF7细胞(图4c)和T-47D细胞(图4d)的基因组DNA进行了qPCR分析。我们发现,对于这两个等位基因特异性引物,Taq388变异体均比野生型Taq具有更高的选择性。具体地说,对于T等位基因引物,来自C/C基因型的MCF7基因组DNA的Taq388变体的错配脱靶扩增强度减少了大约10个循环(图4c),而对于C等位基因引物,来自T/T基因型T-47D基因组DNA的扩增水平比Taq(图4d)减少了10个循环以上。此外,我们在另一个SNP位点rs2236007上观察到了类似的结果。具体地说,对于A等位基因特异性引物,用Taq388变体对G/G基因型T-47D基因组DNA的扩增水平减少了10.5个循环(图8a),而对于G等位基因引物,来自A/A基因型VCaP基因组DNA的扩增水平比Taq减少了多达7个循环(图8b)。

[0096] 此外,我们还将Taq388变体与5种商用的基于SYBR Green的qPCR预混液产品进行了比较。值得注意的是,Taq388聚合酶对Indel造成的引物/模板错配表现出了比所列的所有商业产品更高的选择性(图8c)。此外,该变体在使用G/G和A/A基因型的基因组DNA样本进行rs2236007位点的等位基因特异性PCR扩增中表现出比商品化产品更好的选择性(图8d)。

### [0097] 2.5 Taq388在基因组编辑单细胞克隆基因分型中的应用

[0098] 在功能基因组学研究中,我们通常需要在基因组编辑实验之后筛选大量子代个体或单细胞克隆,才能获得含有目标基因修饰的实验材料,而具有更高选择性的增强型Taq聚合酶能够大大提高基因分型的准确性。因此,我们将Taq388应用于单克隆的基因分型分析,模板为在筛选系统中作为模板使用的26个质粒。在基于TaqMan探针的qPCR分析中,使用野生型序列特异性测试引物,与野生型Taq聚合酶相比,Taq388区分插入/缺失的能力大大提高了,26个indel模板DNA平均提高了16.9个循环(图5a),其中23个indels模板甚至完全没有扩增信号。这表明Taq388在识别和区分indels导致的引物/模板错配方面拥有极其出色的能力。当在基于SYBR Green的qPCR分析中,Taq388在区分这26个indels与野生型的能力平均提高了10.7个循环,同样也显示出比野生Taq更强的扩增特异性(图5b)。尽管不如在基于TaqMan探针的qPCR分析中那样出色,但基于SYBR green的qPCR分析中野生型构建体和插入缺失构建体之间的最小Ct值差异仍超过9个循环,这足以用来准确鉴定插入缺失序列的单细胞克隆。

[0099] 接下来,我们在实际应用场景中用基因组DNA作为模板评估了Taq388在31个单细胞克隆的基因分型分析中的表现,这些克隆是在lenti-X 293T上针对HOXB13基因和DYRK1A基因进行了CRISPR/Cas9介导的基因组编辑<sup>7</sup>。Sanger测序表明,其中二十个克隆在HOXB13基因上产生了双等位基因的indel突变,十一个单细胞克隆在DYRK1A基因上产生了双等位基因indel突变。qPCR基因分型分析结果表明,不管是在HOXB13基因上发生的基因编辑,还是在DYRK1A基因上发生的基因编辑,Taq388都表现出了比Taq聚合酶更好的区分indel序列与野生型序列的能力(图5c,d)。对于在HOXB13 sgRNA靶标2上进行的基因组编辑,Taq388和Taq聚合酶区分indels与野生序列的能力的平均 $\Delta Ct$ 值分别为14.2和10.1个循环(图5c)。具体来说,在检测HT2-04克隆时,Taq聚合酶仅给出4个循环的 $\Delta Ct$ 值,但是Taq388在全部45个PCR循环运行结束都没有检测到有效扩增信号。关于在DYRK1A sgRNA靶标1上的基因组编辑,由Taq388和Taq聚合酶测定的indels突变引起的 $\Delta Ct$ 值分别为9.5和2.6个循环(图5d)。这表明,Taq388的应用可使基因组编辑检测更加准确可靠。

### [0100] 2.6 Taq388在SNP基因分型中的应用

[0101] 作为第三代分子标记,SNP位点具有许多优势,包括分布广泛和遗传稳定性高。它已广泛用于分子生物学,疾病预测和治疗等领域。但是,SNP检测很大程度上也受限于DNA聚合酶的特异性。因此,我们接下来测试了Taq388在SNP基因分型分析中的应用潜力,测试中使用了30个基因组DNA样品,其中有19个来自从ATCC购买的细胞系,11个样品来自发明人,经过随机打乱后排列和编号以隐藏个人信息。我们使用Taq388进行等位基因特异性SYBR Green qPCR扩增,针对五个SNP位点rs2236007,rs4808611,rs11055880,rs2290203和rs2046210进行基因分型分析,并通过Sanger测序确定了这30个样品的SNP基因型。

[0102] 我们采用了两种方法来确定样品的基因型。首先,我们通过图6附图说明中描述的方法,使用等位基因特异性Ct值来计算该等位基因的比例,并据此确定SNP基因型。理论上,对于等位基因1纯合的样本,计算得到的等位基因1和等位基因2的含量应该分别为100%和0%,杂合样本的两个等位基因的百分含量应介于这两个数值之间。对于SNP位点rs2236007,使用Taq388进行的qPCR分析表明,所有样本的SNP基因型都能准确鉴定。其中A/A样本和G/G样本位于相应的坐标轴上,G/A样本位于两者之间(图6a)。出乎意料的是,这10

个G/A样品分布在一个相当分散而不是聚焦在50%附近的区域。我们检查了相应样品的Sanger测序色谱图,发现这些样品的等位基因比例与Sanger测序峰图中的相对峰高高度相关(图10a)。例如,SK-BR-3细胞株具有最高的A等位基因比例,同时在Sanger测序中也表现为A峰远远高于G峰,这表明,我们用Taq388 qPCR基因分型计算出来的等位基因比例真实地反映了该样品的基因型。相反,在用野生Taq聚合酶进行的qPCR分析中,所有样本点都堆积在第一象限,无法确定每个样本的基因型(图6a)。使用Taq388聚合酶对剩下的四个SNP位点rs4808611(图6b)、rs11055880(图6c)、rs2290203(图6d)和rs2046210(图6e)进行的基因分型,成功对每一个样品的SNP基因型进行了确定。并且,杂合基因型样本的分散布局特性也与Sanger测序中相应的峰高有很好的相关性(图10b-e)。

[0103] 常用的终点法SNP基因分型技术使用TaqMan探针或等位基因特异性引物来区分不同的等位基因,现有情况下,为了准确进行SNP基因分型,PCR对等位基因之间的选择性仍急需进一步的提高。因此,接下来我们评估了Taq388在终点法基因分型方法中的应用,即在等位基因特异性PCR循环步骤结束后读取SYBR green荧光,据此来确定样本的基因型。对rs2236007位点的分析结果表明,与野生型Taq聚合酶相比,Taq388的qPCR扩增能完全区别基因型为G/G,G/A和A/A的三组样品(图6f),而野生型Taq qPCR扩增后三种基因型的样品完全堆积在一起,无法区分。同样,我们还使用Taq388聚合酶成功地对其他四个SNP位点rs4808611(图6g)、rs11055880(图6h)、rs2290203(图6i)和rs2046210(图6j)进行了基因分型。

[0104] 本发明中,对全长Taq聚合酶进行了半理性的定向进化,以提高其在PCR扩增中区分基因组编辑突变序列造成的引物-模板错配的能力。首先,我们对Taq聚合酶上与引物/模板双链结构有直接相互作用的40个极性氨基酸进行了逐个定点诱变。然后,在这些变体以及野生型Taq序列的基础上进行广泛的随机突变,生成全面的Taq突变体文库。以带有indel的HOXB13基因质粒为PCR扩增模板,在qPCR平台上经过三轮筛选和验证,筛选出了多个特异性显著提高的Taq变体,其中带有S577A、W645R和I707V替换的Taq388变体表现最好。Taq388变异对源于indel和单核苷酸变异错配的PCR选择性均有极其显著的提高。在应用中,该Taq变体显著提高了getPCR方法对单细胞克隆基因分型的准确性,同时也使AS-qPCR SNP基因分型成为更可行的方法。

[0105] 之前所有提高DNA聚合酶特异性的尝试都集中在辨别单核苷酸错配的能力上。本发明第一次针对基因组编辑indel导致的引物/模板错配,通过广泛的定向进化来获得性能更好的Taq聚合酶变体。此外,作为起始分子,我们使用了全长Taq聚合酶而不是其他研究中常用的Klenow片段,这使得Taq388变体不仅适用于基于SYBR Green的qPCR也适用于基于TaqMan探针的qPCR应用。

[0106] 并且,以前的研究大多是局限性的理性设计,侧重并局限于与引物/模板复合物相互作用的一部分极性氨基酸残基,以及它们之间进一步的简单组合应用。在这里,我们不仅包括了所有40个与引物/模板双链直接接触的极性氨基酸残基,并且在此基础上进一步进行了广泛的随机突变,以建立一个更全面的Taq突变体文库。值得注意的是,在最终的39个变体中,只有13个变体的氨基酸替换涉及了引物/模板接触的残基,并且,所有这些筛选出来的改良的变体都包含了未参与这种接触的氨基酸突变。此外,在我们最终获得的最好的10个变体中,多达5个Taq变体的氨基酸突变完全不涉及那些参与酶/引物/模板相互作用

的氨基酸。这表明,这些引物/模板非接触性氨基酸的替代也有助于提高DNA聚合酶的选择性,为DNA聚合酶的进化提供了新的方向。

[0107] 当应用于检测基因组编辑突变时,Taq388变体显示出极强的区分基因编辑序列和野生型序列的能力。这将使getPCR在基因组编辑实验中检测基因组编辑效率和单细胞克隆的基因分型变得更加准确和方便。当应用于检测那些自然发生的遗传变异时,Taq388在AS-qPCR分析中也显示出出色的SNP等位基因识别能力。我们得益于Taq388在PCR反应中的优秀的等位基因选择能力,实现了两种简单而有效的SNP基因分型方法,即利用等位基因特异性Ct值计算等位基因比例或绘制等位基因特异性PCR扩增的终点荧光散点图。对于这两种方法,三种基因型的样品都可以实现轻松、准确的鉴定。

[0108] 综上所述,通过半理性的定向进化,我们开发了多个Taq聚合酶变体,它们对来自基因组编辑indels的引物/模板错配具有显著提高的选择性,其中最好的突变体Taq388在基因组编辑测试和遗传变异检测方面显示出极大的潜力,该策略的成功为DNA聚合酶的进化提供了新的思路。

[0109] 最后应说明的是:以上所述仅为本发明的优选实施例而已,并不用于限制本发明,尽管参照前述实施例对本发明进行了详细的说明,对于本领域的技术人员来说,其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

## SEQUENCE LISTING

&lt;110&gt; 山东大学

&lt;120&gt; 一种高特异性Taq DNA聚合酶变体及其应用

&lt;130&gt;

&lt;160&gt; 2

&lt;170&gt; PatentIn version 3.3

&lt;210&gt; 1

&lt;211&gt; 833

&lt;212&gt; PRT

&lt;213&gt; 野生型Taq DNA聚合酶氨基酸序列

&lt;400&gt; 1

```

Met Asn Ser Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu
1           5           10           15
Leu Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys
           20           25           30
Gly Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe
           35           40           45
Ala Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile
           50           55           60
Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly
65           70           75           80
Gly Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
           85           90           95
Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu
           100          105          110
Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys
           115          120          125
Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys
           130          135          140
Asp Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu
145          150          155          160
Gly Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg
           165          170          175
Pro Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp
           180          185          190
Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu
           195          200          205
Leu Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg

```

210	215	220
Leu Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu		
225	230	235
Lys Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu		
245	250	255
Val Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala		
260	265	270
Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu		
275	280	285
Leu Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu		
290	295	300
Gly Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala		
305	310	315
Asp Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala		
325	330	335
Pro Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu		
340	345	350
Leu Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu		
355	360	365
Pro Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser		
370	375	380
Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr		
385	390	395
Glu Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn		
405	410	415
Leu Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg		
420	425	430
Glu Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr		
435	440	445
Gly Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val		
450	455	460
Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly		
465	470	475
His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe		
485	490	495
Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys		
500	505	510
Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro		
515	520	525

Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser  
 530 535 540  
 Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg  
 545 550 555 560  
 Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser  
 565 570 575  
 Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly  
 580 585 590  
 Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val  
 595 600 605  
 Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser  
 610 615 620  
 Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His  
 625 630 635 640  
 Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp  
 645 650 655  
 Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr  
 660 665 670  
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu  
 675 680 685  
 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val  
 690 695 700  
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr  
 705 710 715 720  
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala  
 725 730 735  
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met  
 740 745 750  
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys  
 755 760 765  
 Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val  
 770 775 780  
 His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val  
 785 790 795 800  
 Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val  
 805 810 815  
 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys  
 820 825 830  
 Glu

<210> 2

<211> 2502

<212> DNA

<213> 野生型Taq DNA聚合酶核苷酸序列

<400> 2

```

atgaattcgg ggatgctgcc cctctttgag cccaagggcc gggtcctcct ggtggacggc 60
caccacctgg cctaccgcac cttccacgcc ctgaagggcc tcaccaccag ccggggggag 120
ccggtgcagg cggctctacgg cttcgccaag agcctcctca aggccctcaa ggaggacggg 180
gacgcggtga tcgtggtcct tgacgccaag gccccctcct tccgccacga ggcctacggg 240
gggtacaagg cgggcccgggc ccccacgccg gaggactttc cccggcaact cgccctcatc 300
aaggagctgg tggacctcct ggggctggcg cgctcgagg tcccgggcta cgaggcggac 360
gacgtcctgg ccagcctggc caagaaggcg gaaaaggagg gctacgaggt ccgcctcctc 420
accgccgaca aagaccttta ccagctcctt tccgaccgca tccacgtcct ccaccccag 480
gggtacctca tcaccccggc ctggctttgg gaaaagtacg gcctgaggcc cgaccagtgg 540
gccgactacc gggcccctgac cggggacgag tccgacaacc ttcccggggt caagggcac 600
ggggagaaga cggcgaggaa gcttctggag gagtggggga gcctggaagc cctcctcaag 660
aacctggacc ggctgaagcc cgccatccgg gagaagatcc tggcccacat ggacgatctg 720
aagctctcct gggacctggc caaggtgctc accgacctgc ccctggaggt ggacttcgcc 780
aaaaggcggg agcccgaccg ggagaggctt agggcctttc tggagaggct tgagtttggc 840
agcctcctcc acgagttcgg ctttctggaa agccccaagg ccctggagga ggccccctgg 900
ccccgccggg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
gatcttctgg ccctggccgc cgccaggggg ggccgggtcc accgggcccc cgagccttat 1020
aaagccctca gggacctgaa ggaggcgcgg gggttctctg ccaaagacct gagcgttctg 1080
gccctgaggg aaggccttgg cctcccgcc ggcgacgacc ccatgctcct cgcctacctc 1140
ctggaccctt ccaacaccac ccccaggggg gtggcccggc gctacggcgg ggagtggacg 1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260
cttgaggggg aggagaggct ctttggctt taccgggagg tggagaggcc cttttccgct 1320
gtcctggccc acatggaggc cacgggggtg cgctggacg tggcctatct cagggccttg 1380
tccctggagg tggccgagga gatcggccgc ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaacte ccgggaccag ctggaaaggg tcctcttga cgagctaggg 1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgt ccaccagcgc cgccgtcctg 1560
gaggccctcc gcgaggccca cccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgaccttg ccggacctca tccaccccag gacgggccgc 1680
ctccacacc gttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcacccc cttgggcaga ggatccgccg ggccttcate 1800
gccgaggagg ggtggtatt ggtggcctg gactatagcc agatagagct caggggtgctg 1860
gcccacctct ccggcgacga gaacctgatc cgggtcttcc aggaggggcg ggacatccac 1920
acggagaccg ccagctggat gttcggcgtc ccccgggagg ccgtggacc cctgatgcgc 1980
cgggcccgcca agacatcaa cttcggggtc ctctacggca tgtcggccca ccgcctctcc 2040

```



caggagctag ccatccctta cgaggaggcc caggcettca ttgagcgcta ctttcagagc 2100  
ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160  
gtggagacc tcttcggccg ccgccgctac gtgccagacc tagaggccc ggtgaagagc 2220  
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccaggcac cgccgccgac 2280  
ctcatgaagc tggctatggt gaagctcttc cccaggctgg aggaaatggg ggcaggatg 2340  
ctccttcagg tccacgacga gctggtcctc gaggcccaa aagagaggc ggaggccgtg 2400  
gcccggctgg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460  
gaggtgggga taggggagga ctggctctcc gcccaaggagt ga 2502

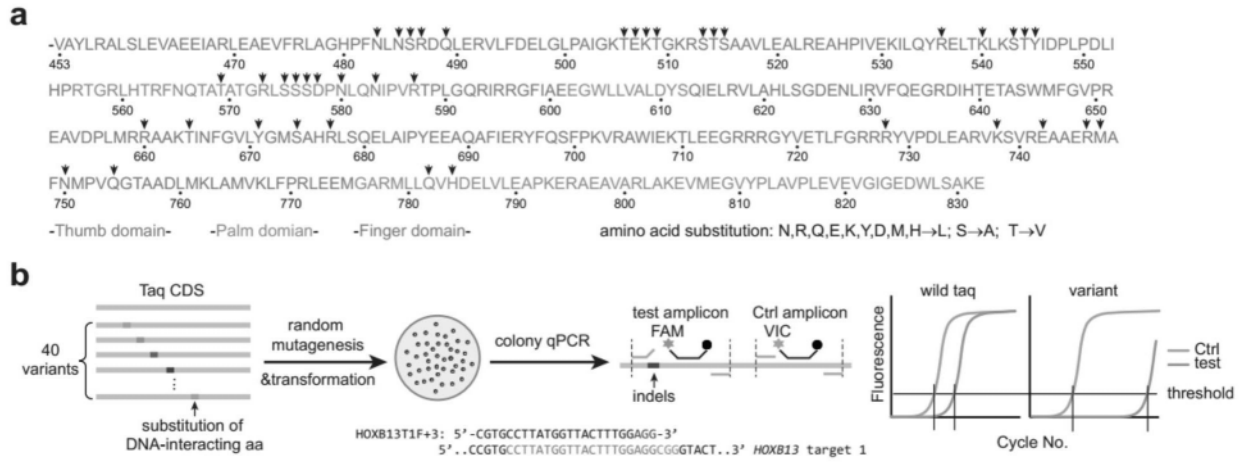


图1

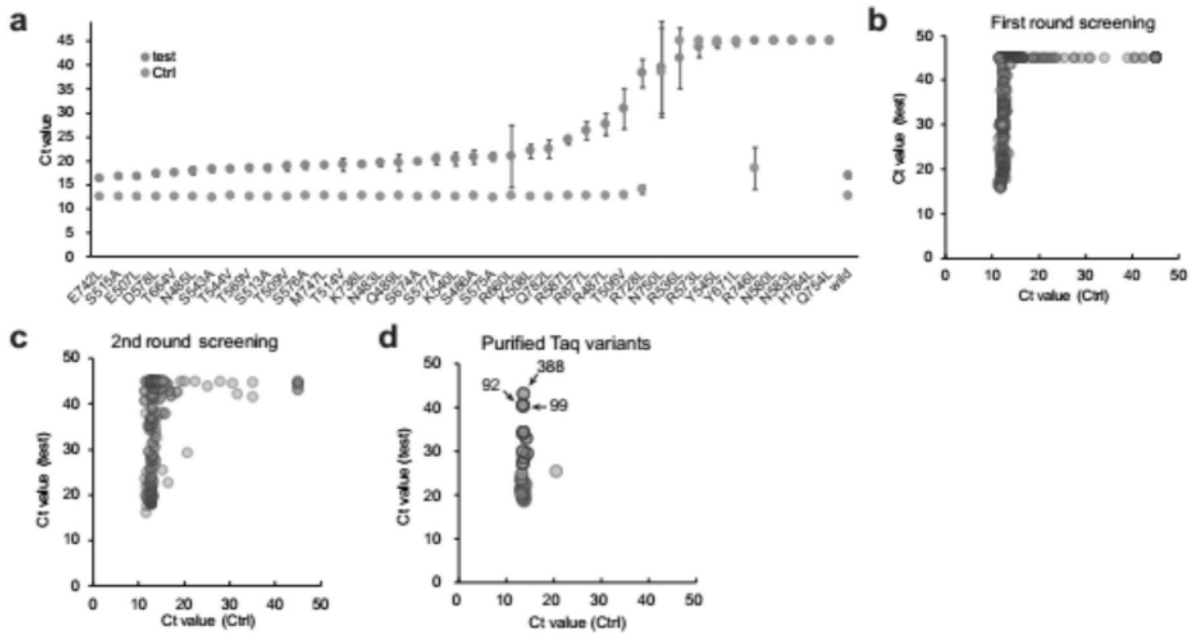


图2

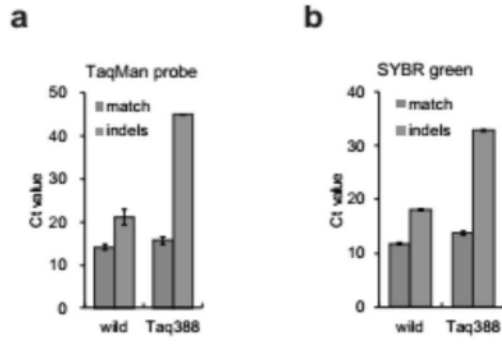


图3

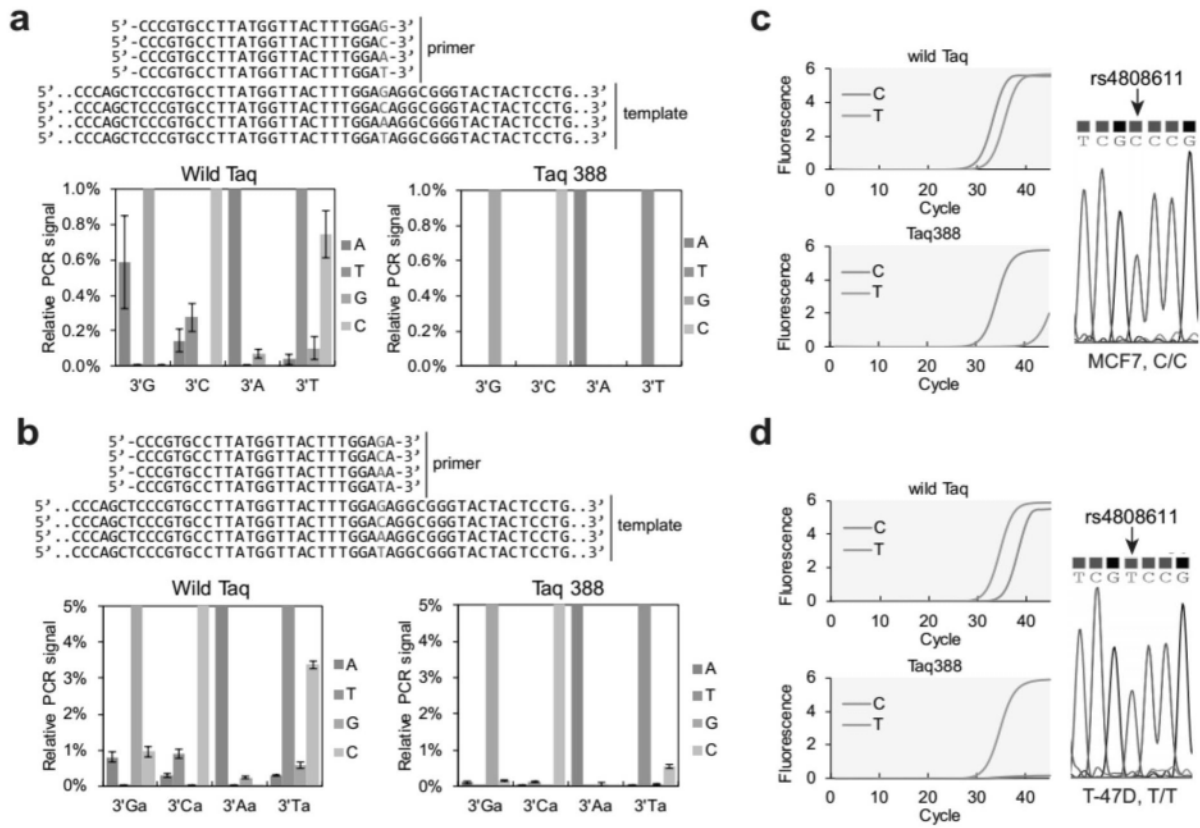


图4

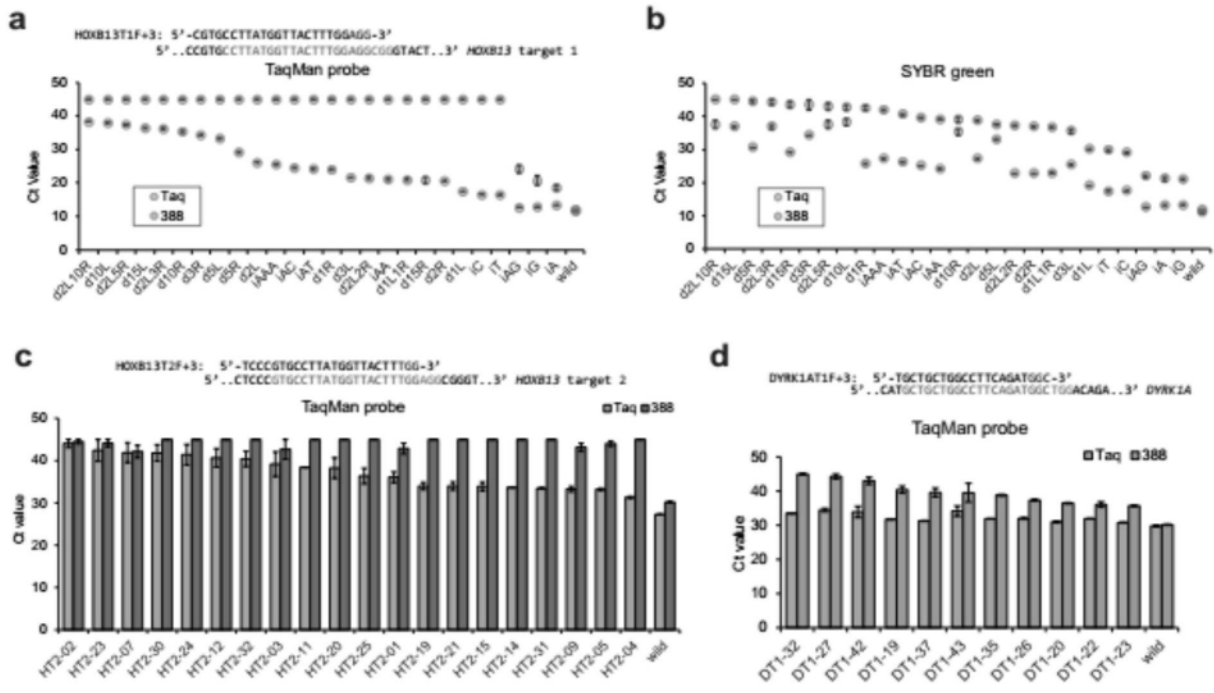


图5

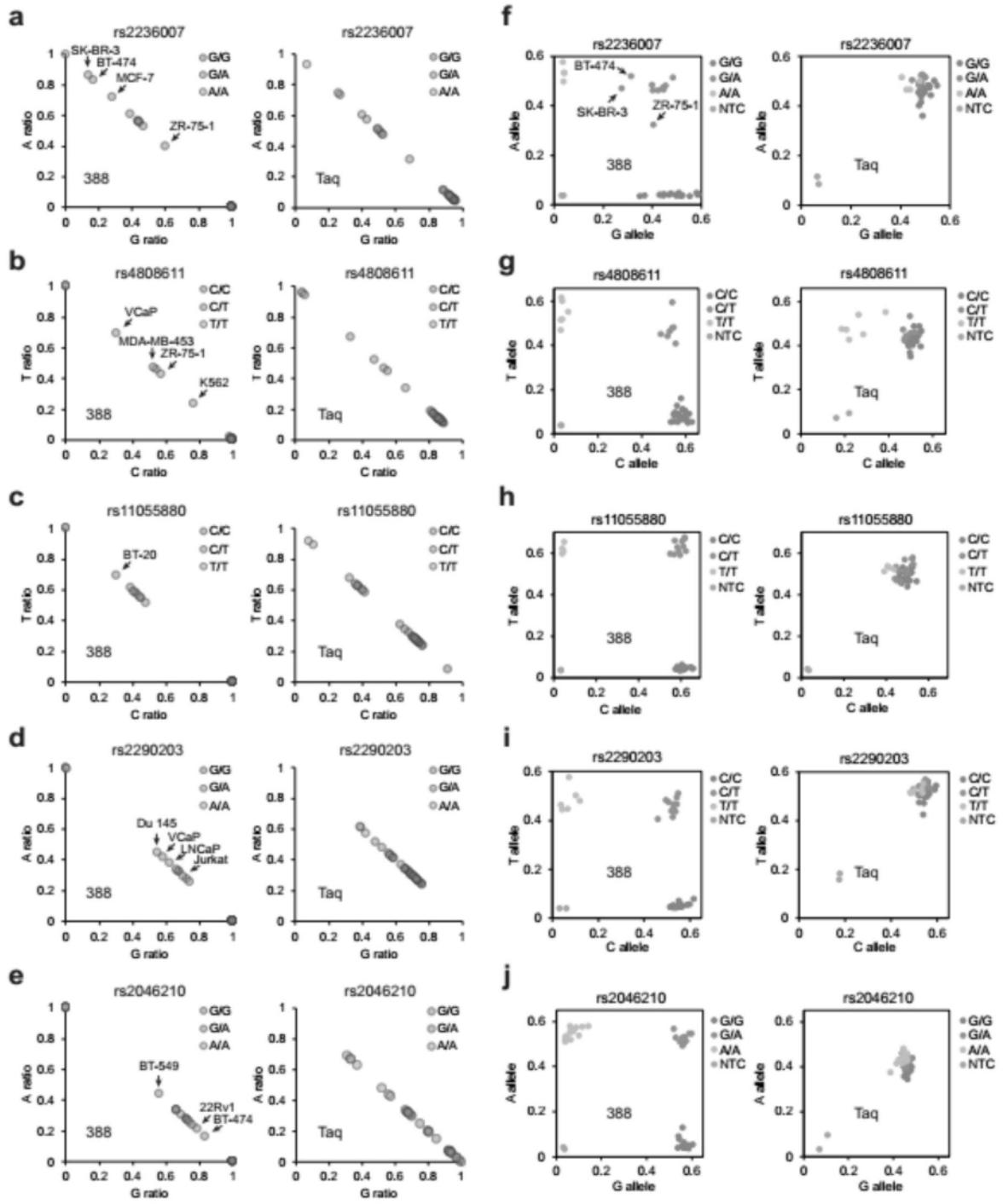


图6



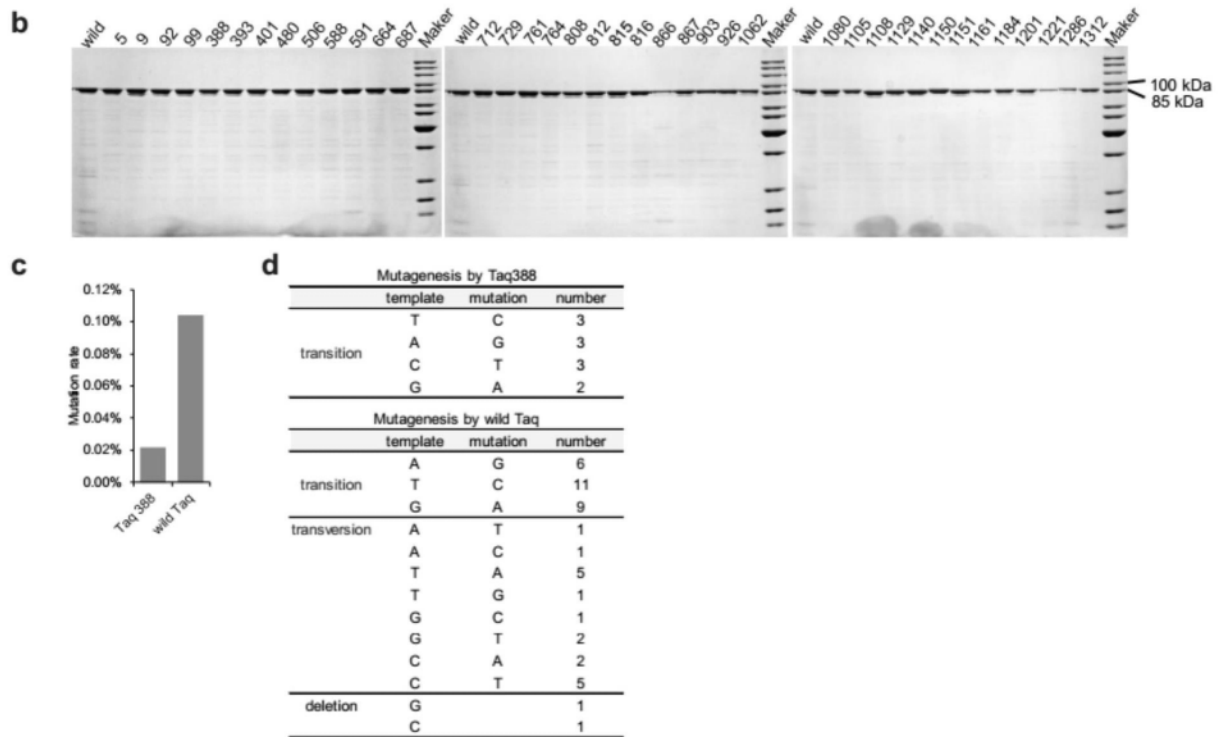


图7

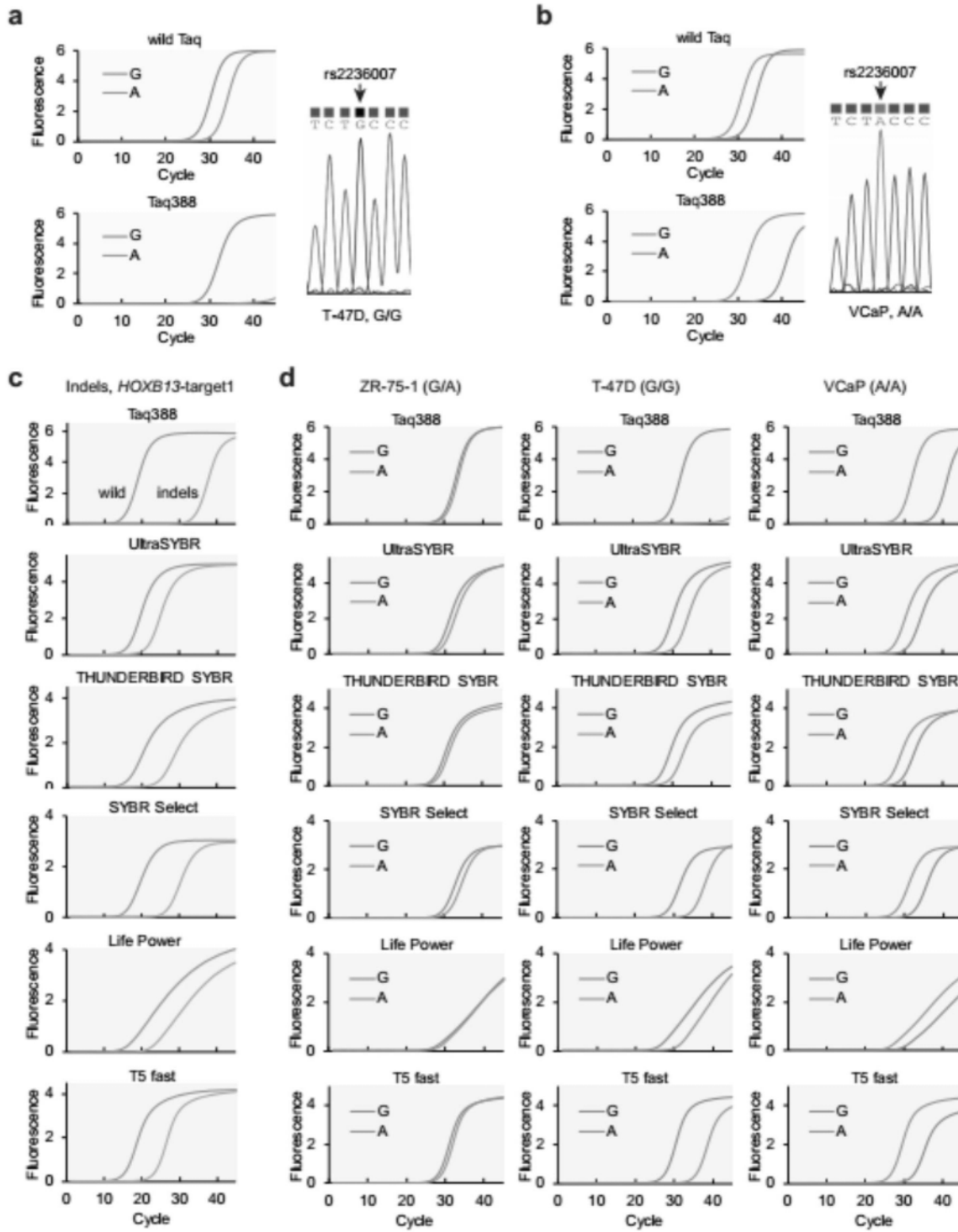


图8



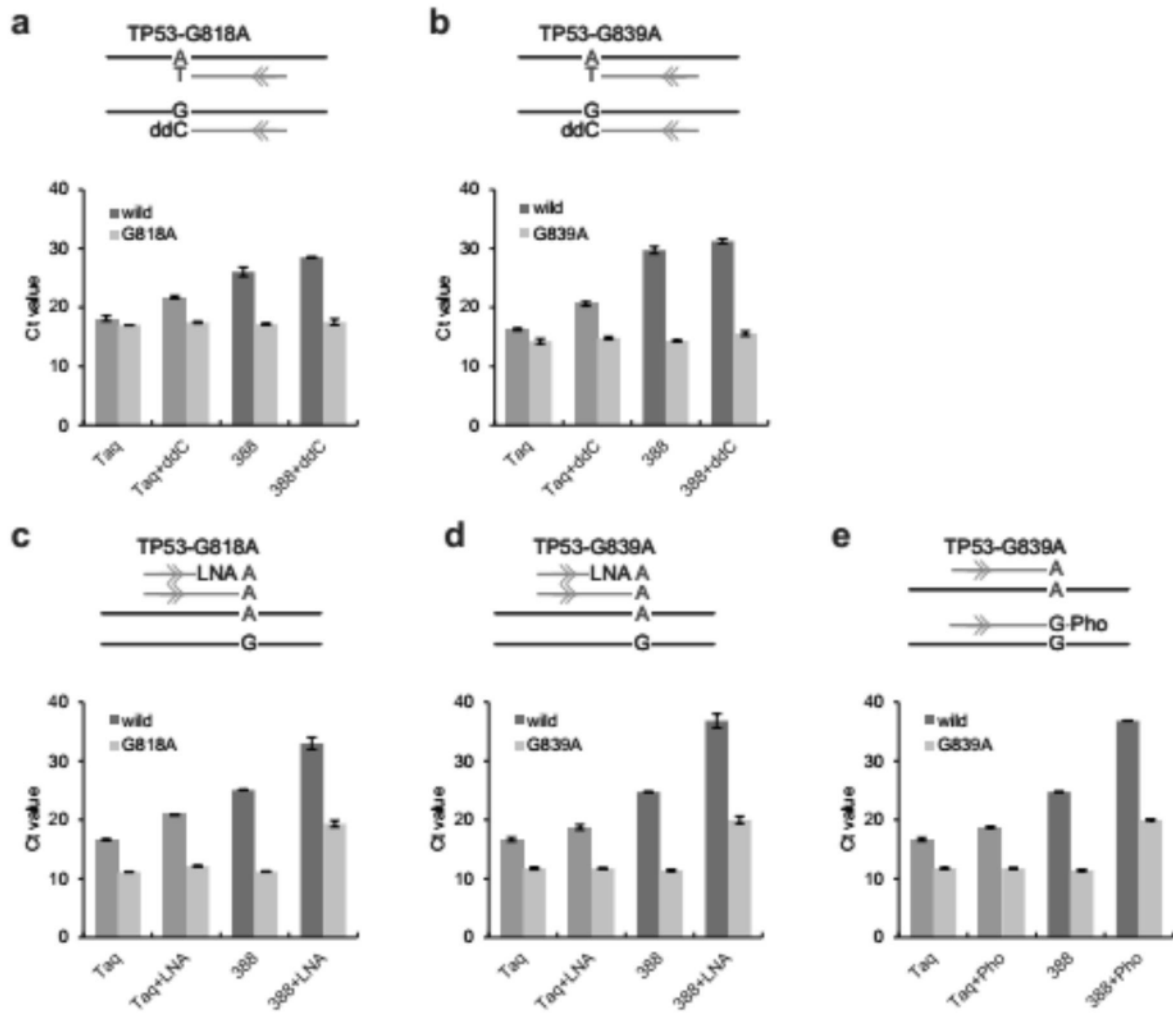


图9

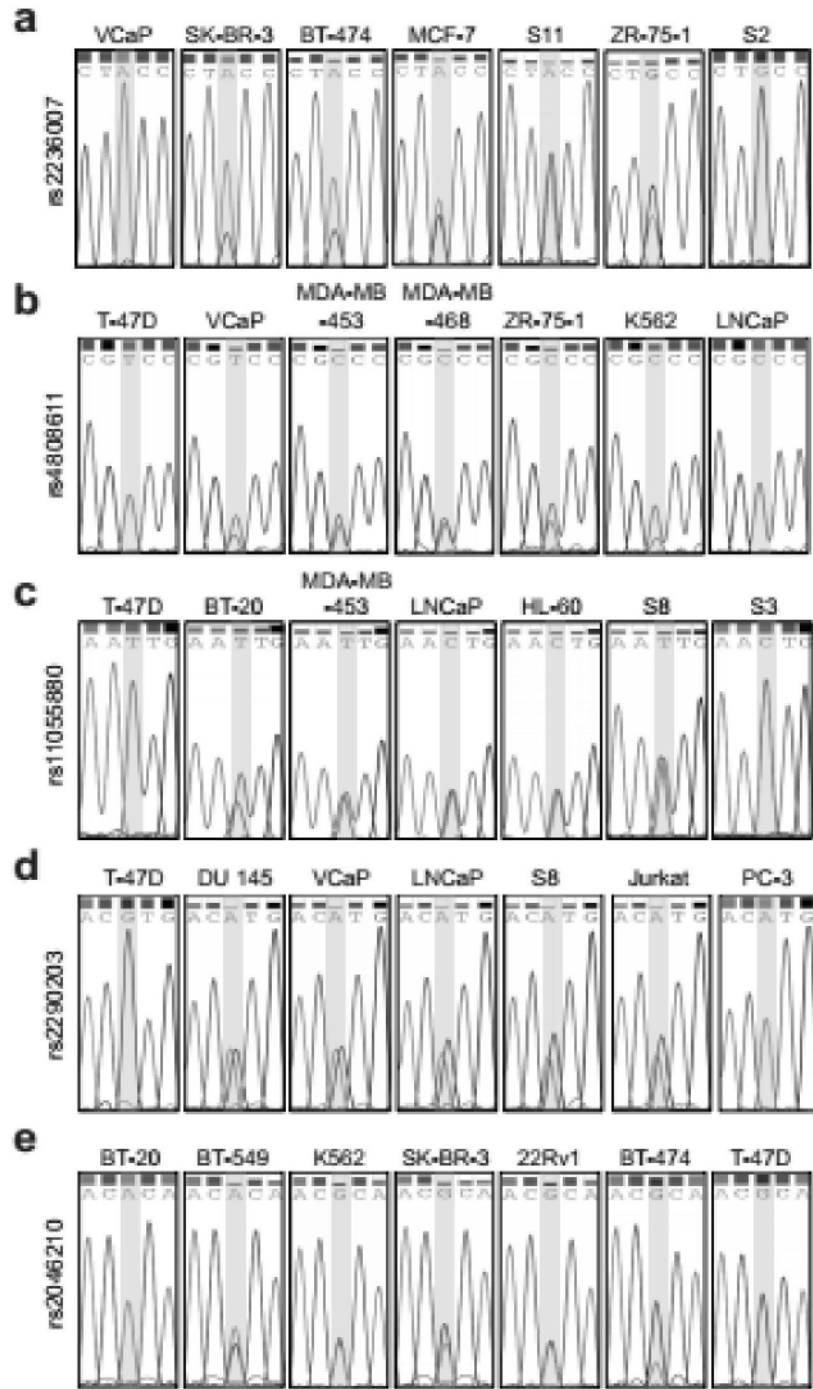


图10