

说明书

发明名称: 靶向乙型肝炎病毒基因的表观编辑工具

技术领域

[0001] 本申请涉及生物医药领域,具体地涉及一种靶标特异性的表观遗传编辑工具及其用途。

背景技术

[0002] 乙型肝炎病毒(Hepatitis B virus, HBV)是一种嗜肝DNA病毒,其引起的慢性乙肝感染可导致持久的肝脏炎症,从而显著增加病人肝硬化和肝癌的几率,据世界卫生组织(WTO)预估,目前在全球范围内乙肝病毒引起的慢性乙型肝炎病例超过2.5亿例,其中每年造成超过80万病人死亡,为世界公共卫生健康造成极大负担。中国作为人口大国,乙肝携带者的数量大约占全球的三分之一。自上世纪八十年代以来,随着乙肝疫苗接种率的提高,我国的新发乙肝病例得到了有效的控制,但是目前国内的乙肝感染者依然存量巨大。

[0003] 目前针对HBV的治疗手段比如核苷(酸)类似物和IFN可以有效地抑制病毒的复制,但是很难实现功能性治愈,即降低表面抗原(HBsAg)至不可检测的水平,并且需要长期服药,一旦治疗停止,HBV的复制将会反弹,造成这一现象的主要原因是HBV感染肝细胞之后会形成共价闭环环状DNA(cccDNA),并且HBV基因组可以整合进入宿主肝细胞基因组,cccDNA以及整合DNA可以长期稳定地作为病毒复制和蛋白表达的模板,目前的常规治疗并不会消除二者,因此,消除或者沉默cccDNA以及整合DNA可能有效地实现乙肝感染的治疗。由于cccDNA在宿主肝细胞核中可以与组蛋白组装形成类似于染色体的结构,因此其转录调控也可受到表观修饰的影响。HBV基因组中存在三个CpG岛,以下分别称为CG I、CG II和CG III, CpG岛是实现表观调控的关键区域。HBV基因组的转录是由四个启动子(Xp、Cp、Sp1、Sp2)和两个增强子(Enh I和Enh II)调控的,其中Xp、Cp、Enh I和Enh II位于CG II区域内,Sp1和Sp2位于CG I和CG III附近,因此,针对CpG岛进行表观编辑可能影响多个调控元件,从而影响病毒基因组的转录。

- [0004] 在HBV基因组的特定调控区域或特定位点引入表观修饰，可改变染色质结构，从而将靶标基因调整为转录抑制状态，实现靶标基因的沉默调控，在此过程中不会切割DNA，避免了产生基因组双链断裂的可能性，从根本机理上杜绝了激活不可预测的DNA修复机制并可能产生免疫原性截短或突变蛋白的风险。但目前针对特异性靶向HBV基因组的表观编辑工具的研究大多还处于早期阶段，能够持续治愈或清除HBV病毒的表观编辑技术开发尚面临许多未知性和局限性的挑战。
- [0005] 发明内容
- [0006] 本申请以开发安全有效的沉默调控HBV靶标基因的疗法为目的，提供了一种通过表观编辑工具在HBV基因组特定调控区域引入抑制性的表观修饰的方法，从而改变乙肝病毒cccDNA以及整合基因组DNA转录活性，实现降低HBV的复制和蛋白的表达，达到治疗乙肝的目的。
- [0007] 一方面，本申请提供了一种组合物，所述组合物包含：(1)融合分子或编码所述融合分子的核酸序列，其中所述融合分子包含至少一种DNA结合蛋白和至少一种基因表达调节剂，以及(2)至少一种单一引导RNA(sgRNA)，或编码所述sgRNA的核酸序列，其中所述sgRNA与乙型肝炎病毒(HBV)基因附近和/或HBV基因调控元件内的靶DNA序列互补，所述HBV基因因为包含SEQ ID NO:1所示的核苷酸序列的B型HBV基因、包含SEQ ID NO:2所示的核苷酸序列的C型HBV基因或者包含SEQ ID NO:3所示的核苷酸序列的D型HBV基因，所述HBV基因调控元件包含转录起始位点、核心启动子、启动子、增强子、沉默子、绝缘子元件、边界元件和/或基因座控制区，且所述靶DNA序列位于所述HBV基因的第1056位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。
- [0008] 在某些实施方式中，所述靶DNA序列位于所述HBV基因的第1056位至第1900位核苷酸之间、第1972位至第2082位核苷酸之间、第2134位至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。

- [0009] 在某些实施方式中，所述靶DNA序列位于所述HBV基因的第1060位至第1079位核苷酸之间、第1149位至第1612位核苷酸之间、第1693位至第1852位核苷酸之间、和/或第2863位至第2882位核苷酸之间。
- [0010] 在某些实施方式中，所述靶DNA序列位于下列所述B型HBV基因的一个或多个区域：第1149至第1190位核苷酸之间、第1210位至第1310位核苷酸之间、第1350位至第1400位核苷酸之间、第1420位至第1450位核苷酸之间和第1470位至第1592位核苷酸之间。
- [0011] 在某些实施方式中，所述靶DNA序列位于下列所述C型HBV基因的一个或多个区域：第1150位至第1180位核苷酸之间、第1200位至第1310位核苷酸之间、第1350位至第1390位核苷酸之间、第1420位至第1460位核苷酸之间和第1480位至第1593位核苷酸之间。
- [0012] 在某些实施方式中，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1056位至第1079位核苷酸之间、第1101位至第1900位核苷酸之间、第1972位至第2082位核苷酸之间、第2134位至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2882位核苷酸之间、第2911位至第2930位核苷酸之间、和第3048位至第3067位核苷酸之间。
- [0013] 在某些实施方式中，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1060位至第1079位核苷酸之间、第1160位至第1310位核苷酸之间、第1253位至第1284位核苷酸之间、第1370位至第1470位核苷酸之间、第1490位至第1612位核苷酸之间、第1693位至第1852位核苷酸之间、和第2863位至第2882位核苷酸之间。
- [0014] 例如，所述B型HBV(血清型adw)全基因组的序列如下所示：

CTCCACCACTTTCCACCAAACCTCTTCAAGATCCCAGAGTCAGGGCCCTGTACTTTCCTGC
TGGTGGCTCCAGTTCAGGAACAGTGAGCCCTGCTCAGAATACTGTCTCTGCCATATCGTC
AATCTTATCGAAGACTGGGGACCCTGTACCGAACATGGAGAACATCGCATCAGGACTCC
TAGGACCCCTGCTCGTGTACAGGCGGGGTTTTTCTTGTTGACAAAAATCCTCACAATAC
CACAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTCTAGGGGGAACACCCGTGTGT
CTTGGCCAAAATTCGCAGTCCCAAATCTCCAGTCACTCACCAACCTGTTGTCTCTCAATT
TGTCTGGTTATCGCTGGATGTGTCTGCGGCGTTTTATCATCTTCTCTGCATCCTGCTGC
TATGCCTCATCTTCTTGTTGGTTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCTCT
AATTCCAGGATCATCAACAACCAGCACCGGACCATGCAAAACCTGCACAACCTCTGCTC
AAGGAACCTCTATGTTTCCCTCATGTTGCTGTACAAAACCTACGGACGGAAACTGCACC
TGTATTCCCATCCCATCATCTTGGGCTTTCGCAAAATACCTATGGGAGTGGGCCTCAGTCC
GTTTCTCTTGGCTCAGTTTACTAGTGCCATTTGTTTCAGTGGTTTCGTAGGGCTTTCCTCCAC
TGTCTGGCTTTCAGTTATATGGATGATGTGGTTTTGGGGGCAAGTCTGTACAACATCTT
GAGTCCCTTTATGCCGCTGTTACCAATTTTCTTTTGTCTTTGGGTATACATTAAACCCCTC
ACAAAACAAAAGATGGGGATATCCCTTAACTTCATGGGATATGTAATTGGGAGTTGGG
GCACATTGCCACAGGAACATATTGTACAAAAAATCAAAATGTGTTTTAGGAAACTTCCTG
TAAACAGGCCTATTGATTGGAAAGTATGTCAACGAATTGTGGGTCTTTTGGGGTTTGCCG
CCCCTTTCACGCAATGTGGATATCCTGCTTAAATGCCTTTATATGCATGTATACAAGCGAA
ACAGGCTTTTACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTATCTGAACCT
TTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTGTTGCTGACGCAACCCCA

CTGGTTGGGGCTTGGCCATAGGCCATCAGCGCATGCGTGGAACCTTTGTGTCTCCTCTAC
CGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGCAGCAGGTCTGGGGCAAAC
CTCATCGGGACTGACAATTCGTGCTGCTCTCCCGCAAGTATACATCATTTCCATGGCTG
CTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTACGTCCCGTCGGC
GCTGAATCCCGCGGACGACCCCTCTCGGGGCCGCTTGGGACTCTACCGCCCGCTTCTCC
GCCTATTGTACCGGCCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTCTGTG
CCTTCTCATCTGCCGGACCGTGTGCACCTTCGCTTCACCTCTGCACGTCGCATGGAGACC
ACCGTGAACGCCCACAGGAACCTGCCCAAGGTCTTGCATAAGAGGACTCTTGGACTTTC
AGCAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTGTGTTTACTGAGTGGG
AGGAGTTGGGGGAGGAGGTTAGGTTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAA
TTGGTGTGTTTACCAGCACCATGCAACTTTTTACCTCTGCCTAATCATCTCATGTTTCATG
TCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATTGACC
CGTATAAAGAATTTGGAGCTTCTGTGGAGTTACTCTCTTTTTTGCCCTTCTGACTTCTTTCC
TTCTATTTCGAGATCTCCTCGACACCGCCTCTGCTCTGTATCGGGAGGCCTTAGAGTCTCC
GGAACATTGTTACCTCACCATACGGCACTCAGGCAAGCTATCCTGTGTTGGGGTGAGT
TGATGAATCTAGCTACCTGGGTGGGAAGTAATTTGGAAGATCCAGCATCCAGGGAATTA
GTAGTCAGCTATGTCAACGTTAACATGGGCCTAAAAATCAGACAACTCTTGTGGTTTCAC
ATTTCTGTCTTACTTTTGGGAGAGAACTGTCTTGAATATTTGGTGTCTTTTGGAGTGT
GGATTTCGCACTCCTCCTGCATATAGACCACCAAATGCCCCCTATCTTATCAACTCTTCCGGA
AACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGCA
GACGAAGGTCTCAATCGCCGCGTCGCAGAAGATATCAATCTCGGGAATCTCAATGTTAG
TATTCCTTGGACACATAAGGTGGGAACTTTACGGGGCTTTATTCTTCTACGGTACCTTG
CTTTAATCCTAAATGGCAAACCTCTTTTCTGACATTCATTTGCAGGAGGACATTGTT
GATAGATGTAAGCAATTTGTGGGGCCCCCTTACAGTAAATGAAAACAGGAGACTGAAATT
AATTATGCCTGCTAGGTTTTATCCCAATGTTACTAAATATTTGCCCTTAGATAAAGGGATC
AAACCGTATTATCCAGAGTATGTAGTTAATCACTTCCAGACGCGACATTATTTACACA
CTCTTTGGAAGGCGGGGATCTTATATAAAAGAGAGTCCACACGTAGCGCCTCATTTTGCG
GGTCACCATAATTCTTGGGAACAAGATCTACAGCATGGGAGGTTGGTCTTCCAAACCTCG
AAAAGGCATGGGGACAAATCTTCTGTCCCCAATCCCCTGGGATTCTTCCCCGATCATCA
GTTGGACCTGCATTCAAAGCCAACCTCAGAAAGTCCAGATTGGGACCTCAACCCGCAC

AAGGACAACCTGGCCGGACGCCAACAAAGGTGGGAGTGGGAGCATTGCGGCCAGGGTTC
ACCCCTCCCCATGGGGGACTGTTGGGGTGGAGCCCTCAGGCTCAGGGCCTACTCACAAC
TGTGCCAGCAGCTCCTCCTCCTGCCTCCACCAATCGGCAGTCAGGAAGGCAGCCTACTC
CCTTATCTCCACCTCTAAGGGACACTCATCCTCAGGCCATGCAGTGGAA (SEQ ID NO: 1)。

[0015] 例如，所述C型HBV(血清型adr)全基因组的序列如下所示：

CTCCACAACATTCCACCAAGCTCTGCTAGATCCCAGAGTGAGGGGCCATATATTTCTCTGC
TGGTGGCTCCAGTTCCGGAACAGTAAACCTGTTCCGACTACTGCCTCACCCATATCGTC
AATCTTCTCGAGGACTGGGGACCCTGCACCGAACATGGAGAACAACATCAGGATTCC
TAGGACCCCTGCTCGTGTACAGGCGGGGTTTTCTTGTGACAAGAATCCTCACAATAC
CACAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTCTAGGGGGAGCACCCACGTGT
CCTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCTTGTCTCCAATT
TGTCTTGGCTATCGCTGGATGTGTCTGCGGCGTTTATCATAATCCTCTTCATCCTGCTGCT
ATGCCTCATCTTCTTGTGGTTCTTCTGGACTACCAAGGTATGTTGCCCCGTTTGTCTCTA
CTTCCAGGAACATCAACCACCAGCACGGGGCCATGCAAGACCTGCACGATTCTGCTCA
AGGAACCTCTATGTTTCCCTCTTGTGCTGTACAAAACCTTCGGACGGAACTGCACTT
GTATTCCCATCCCATCATCCTGGGCTTTCGCAAGATTCTATGGGAGTGGGCCTCAGTCC
GTTTCTCTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTTCGTAGGGCTTTCCCCCA
CTGTTTGGCTTTTCAGTATATGGATGATGTGGTATTTGGGGGCCAAGTCTGTACAACATCTT
GAGTCCCTTTTTACCTCTATTACCAATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAA
TAAACCAAACGTTGGGGCTACTCCCTTAACCTCATGGGATATGTAATTGGAAGTTGGGG
TACTTTACCGCAGGAACATATTGTACTAAAACCTCAAGCAATGTTTTCGAAAATTACCTGT
AAATAGACCTATTGATTGGAAAGTCTGTCAAAGAATTGTGGGTCTTTTGGGCTTTGCTGC
CCCTTTTACACAATGTGGCTATCCTGCCTTGATGCCTTTATATGCATGTATACAATCTAAGC
AGGCTTTCACCTTCTCGCCAACCTACAAGGCCTTCTGTGTAAACAATATCTGAACCTTT
ACCCCGTTGCCCGGCAACGGTCAGGTCTCTGCCAAGTGTGTGCTGACGCAACCCCCACT
GGATGGGGCTTGGCCATAGGCCATCGGCGCATGCGTGGAACCTTTGTGGCTCCTCTGCC
GATCCATACTGCGGAACCTCCTAGCAGCTTGTTCGCTCGCAGCCGGTCTGGAGCGAAAC
TTATCGGCACCGACAACCTCTGTTGTCTCTCTCGGAAATACACCTCCTTTCCATGGCTGC
TAGGGTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTCTACGTCCCGTCGGCG
CTGAATCCCGCGGACGACCCGTCTCGGGGCCGTTTGGGGCTCTACCGTCCCCCTTCTTCAT

CTGCCGTTCCGGCCGACCACGGGGCGCACCTCTCTTTACGCGGTCTCCCCGTCTGTGCC
TTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTCTGCACGTGCGATGGAGACCAC
CGTGAACGCCCACCAGGTCTTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCTCAG
CAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTGTGTTAAAGACTGGGAG
GAGTTGGGGGAGGAGATTAGGTTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAATT
GGTCTGTTACCAGCACCATGCAACTTTTTACCTCTGCCTAATCATCTCATGTTTCATGTC
CTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCG
TATAAAGAATTTGGAGCATCTGTGGAGTTACTCTCTTTTTTGCTTCTGACTTCTTTTCCTT
CTATTGAGATCTCCTCGACACCGCTTCTGCTCTGTATCGGGAGGCCTTAGAGTCTCCGG
AACATTGTTACCTCACCATAACAGCACTCAGGCAAGCTATTCTGTGTTGGGGTGAGTTG
ATGAATCTGGCCACCTGGGTGGGAAGTAATTTGGAAGACCCAGCATCCAGGGAATTAGT
AGTCAGCTATGTCAATGTTAATATGGGCCTAAAAATCAGACAACCTATTGTGGTTTCACATT
TCCTGTCTTACTTTTGGAAGAGAACTGTTCTTGAGTATTTGGTATCTTTTGAGTGTGG
ATTGCACTCCTCCCGCTTACAGACCACCAAATGCCCCCTATCTTATCAACACTTCCGGAA
ACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGCAG
ACGAAGATCTCAATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAATCTCAATGTTAGTA
TCCCTTGGA CTCATAAGGTGGGAACTTTACTGGGCTTTATTCTTCTACTGTACCTGTCTT
TAATCCCGAGTGGCAAACCTCCCTCCTTTTCCTCACATTCAATTACAGGAGGACATTATTAAT
AGATGTCAACAATATGTGGGCCCTCTTACAGTTAATGAAAAAGGAGATTAAAATTAATT
ATGCCTGCTAGGTTCTATCCTAACCTTACCAAATATTTGCCCTTAGACAAAGGCATTAAAC
CGTATTATCCTGAACATGCAGTTAATCATTACTTCAAACTAGGCATTATTTACATACTCTG
TGGAAGGCTGGCATTCTATATAAGAGAGAACTACACGCAGCGCCTCATTTTGTGGGTC
ACCATATTCTTGGAACAAGAGCTACAGCATGGGAGGTTGGTCTTCCAAACCTCGACAA
GGCATGGGGACGAATCTTTCTGTTCCCAATCCTCTGGGATTCTTTCCCGATCACCAGTTG
GACCCTGCGTTCGGAGCCAACTCAAACAATCCAGATTGGGACTTCAACCCCAACAAGG
ATCACTGGCCAGAGGCAAATCAGGTAGGAGCGGGAGCATTCTGGGCCAGGGTTACCCCC
ACCACACGGCGGTCTTTTGGGGTGGAGCCCTCAGGCTCAGGGCATATTGACAACAGTGC
CAGCAGCACCTCCTCCTGCCTCCACCAATCGGCAGTCAGGAAGACAGCCTACTCCCATC
TCTCCACCTCTAAGAGACAGTCATCCTCAGGCCATGCAGTGGA (SEQ ID NO: 2)。

[0016] 例如，所述D型HBV(血清型ayw)全基因组的序列如下所示：

TTCCACAACCTTTCACCAAACCTCTGCAAGATCCCAGAGTGAGAGGCCTGTATTCCCTG
CTGGTGGCTCCAGTTCAGGAGCAGTAAACCCTGTCCGACTACTGCCTCTCCCTTATCGT
CAATCTTCTCGAGGATTGGGGACCCTGCGCTGAACATGGAGAACATCACATCAGGATTC
CTAGGACCCCTTCTCGTGTTACAGGCGGGGTTTTCTTGTGACAAGAATCCTCACATA
CCGCAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGAACCTACCGTGTG
TCTTGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAACCTCCTGTCCTCAA
CTTGTCCTGGTATCGCTGGATGTGTCTGCGGCGTTTTATCATCTTCTCTCACTCTGCT
GCTATGCCCTCATCTTCTTGTGGTTCTTCTGGACTATCAAGGTATGTTGCCCGTTTGTCT
CTAATTCAGGATCCTCAACCACCAGCACGGGACCATGCCGAACCTGCATGACTACTGC
TCAAGGAACCTCTATGTATCCCTCCTGTTGCTGTACCAAACCTTCGGACGGAAATTGCAC
CTGTATTCCCATCCCATCATCTGGGCTTTCGGAAAATTCCTATGGGAGTGGGCCTCAGC
CCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTTGTTTCAGTGGTTCGTAGGGCTTTCCCC
ACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCTGTACAGCATCT
TGAGTCCCTTTTTACCGCTGTTACCAATTTTCTTTTGTCTTTGGGTATACATTAAACCCTA
ACAAAACAAAGAGATGGGGTTACTCTCTGAATTTTATGGGTATGTCAATTGGAAGTTATG
GGTCTTGCCACAAGAACACATCATAAAAAATCAAAGAATGTTTTAGAAAACCTCCT
ATTAACAGGCCTATTGATTGGAAAGTATGTCAACGAATTGTGGGTCTTTTGGGTTTTGCT
GCCCCATTACACAATGTGGTTATCCTGCGTTAATGCCCTTGATGCATGTATTCAATCTAA
GCAGGCTTTCACCTTCTCGCCAACTTACAAGGCCTTCTGTGTAAACAATACCTGAACCT
TTACCCCGTTGCCCCGCAACGGCCAGGTCTGTGCCAAGTGTGCTGACGCAACCCCCA
CTGGCTGGGGCTTGGTCATGGGCCATCAGCGCGTGCGTGGAACCTTTTCGGCTCCTCTG
CCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGCAGCAGGTCTGGAGCAAA
CATTATCGGGACTGATAACTCTGTGTCTCTCCCGCAAATATACATCGTATCCATGGCTG
CTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTACGTCCCGTCGGC
GCTGAATCCTGCGGACGACCCTTCTCGGGGTCGCTTGGGACTCTCTCGTCCCCCTTCTCC
GTCTGCCGTTCCGACCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTCTGTG
CCTTCTCATCTGCCGACCGTGTGCACTTCGCTTACCTCTGCACGTCGCATGGAGACC
ACCGTGAACGCCACCGAATGTTGCCAAAGGTCTTACATAAGAGGACTCTTGGACTCTC
TGCAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTTTGTTTAAAGACTGGG
AGGAGTTGGGGGAGGAGATTAGATTAAAGGTCTTTGTACTAGGAGGCTGTAGGCATAAA

TTGGTCTGCGCACCAGCACCATGCAACTTTTTACCTCTGCCTAATCATCTCTTGTTCATG
TCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATCGACC
CTTATAAAGAATTTGGAGCTACTGTGGAGTTACTCTCGTTTTTGCCTTCTGACTTCTTTCC
TTCAGTACGAGATCTTCTAGATAACCGCCTCAGCTCTGTATCGGGAAGCCTTAGAGTCTCC
TGAGCATTGTTACCTCACCATACTGCACTCAGGCAAGCAATTCTTTGCTGGGGGGAAC
TAATGACTCTAGCTACCTGGGTGGGTGTTAATTGGAAGATCCAGCATCTAGAGACCTAG
TAGTCAGTTATGTCAACACTAATATGGGCCTAAAGTTCAGGCAACTCTTGTGGTTTCACA
TTTCTTGTCTCACTTTTGGGAAGAGAAACCGTTATAGAGTATTGGTGTCTTCGGAGTGT
GGATTGCGACTCCTCCAGCTTATAGACCACCAAATGCCCTATCTATCAACACTTCCGG
AACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGC
AGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAACCTCAATGTTA
GTATTCCTTGGACTCATAAGGTGGGGAACCTTACTGGTCTTTATTCTTCTACTGTACCTGT
CTTTAATCCTCATTGGAAAACACCATCTTTTCTAATATACATTTACACCAAGACATTATC
AAAAAATGTGAACAGTTTGTAGGCCCACTTACAGTTAATGAGAAAAGAAGATTGCAATT
GATTATGCCTGCTAGGTTTTATCCAAAGGTTACCAAATATTTACCATTGGATAAGGGTATT
AAACCTTATTATCCAGAACATCTAGTTAATCATTACTTCCAACTAGACACTATTTACACA
CTCTATGGAAGGCGGGTATATTATATAAGAGAGAAACAACACATAGCGCCTCATTTTGTG
GGTCACCATATTCTTGGGAACAAGATCTACAGCATGGGGCAGAATCTTCCACCAGCAAT
CCTCTGGGATTCTTTCCCGACCACCAGTTGGATCCAGCCTTCAGAGCAAACACAGCAAA
TCCAGATTGGGACTTCAATCCCAACAAGGACACCTGGCCAGACGCCAACAAGGTAGGA
GCTGGAGCATTCGGGCTGGGTTTCACCCACCGCACGGAGGCCTTTTGGGGTGGAGCC
CTCAGGCTCAGGGCATACTACAACTTTGCCAGCAAATCCGCCTCCTGCCTCCACCAAT
CGCCAGACAGGAAGGCAGCCTACCCCGCTGTCTCCACCTTTGAGAAACACTCATCTCA
GGCCATGCAGTGGA (SEQ ID NO: 3)。

- [0017] 在某些实施方式中，所述sgRNA包含SEQ ID NOs：4-1165中任一项所示的核苷酸序列。
- [0018] 在某些实施方式中，所述sgRNA包含SEQ ID NOs：4-1165中任一项所示的核苷酸序列的部分序列，所述部分序列的长度为15-20个碱基对。
- [0019] 在某些实施方式中，所述至少一种DNA结合蛋白是CRISPR酶、锌指核酸酶(ZNF)、转录激活子样效应因子核酸酶(TALEN)、归巢核酸内切酶、dCas9-FokI核酸酶或MegaTal核酸酶。
- [0020] 在某些实施方式中，所述CRISPR酶为2类Cas蛋白和/或其突变体。

- [0021] 在某些实施方式中，所述CRISPR酶为下列所述的一种或多种Cas蛋白：2类II-A型Cas蛋白、2类II-B型Cas蛋白、2类II-C型Cas蛋白、2类V-A型Cas蛋白、2类V-B型Cas蛋白、2类V-C型Cas蛋白、2类V-U型Cas蛋白、和以上所述的突变体。
- [0022] 在某些实施方式中，所述CRISPR酶为Cas9蛋白和/或其突变体。
- [0023] 在某些实施方式中，所述至少一种DNA结合蛋白是dCas9。
- [0024] 在某些实施方式中，所述dCas9包括金黄色葡萄球菌(*Staphylococcus aureus*)dCas9、化脓性链球菌(*Streptococcus pyogenes*)dCas9、空肠弯曲菌(*Campylobacter jejuni*)dCas9、白喉棒状杆菌(*Corynebacterium diphtheria*)dCas9、凸腹真杆菌(*Eubacterium ventriosum*)dCas9、巴氏链球菌(*Streptococcus pasteurianus*)dCas9、香肠乳杆菌(*Lactobacillus farciminis*)dCas9、球形球毛菌(*Sphaerochaeta globus*)dCas9、固氮螺菌属(*Azospirillum*，例如，菌株B510)dCas9、嗜重氮葡萄糖醋杆菌(*Gluconacetobacter diazotrophicus*)dCas9、灰色奈瑟球菌(*Neisseria cinerea*)dCas9、肠道罗斯拜瑞氏菌(*Roseburia intestinalis*)dCas9、食清洁剂细小棒菌(*Parvibaculum lavamentivorans*)dCas9、卤水硝酸盐裂解菌(*Nitratifactor salsuginis*，例如，菌株DSM 16511)dCas9、海鸥弯曲菌(*Campylobacter lari*，例如，菌株CF89-12)dCas9、嗜热链球菌(*Streptococcus thermophilus*，例如，菌株LMD-9)dCas9。
- [0025] 在某些实施方式中，所述dCas9包含SEQ ID NO：1170-1187所示的氨基酸序列。
- [0026] 在某些实施方式中，所述至少一种基因表达调节剂提供所述HBV基因附近和/或HBV基因调控元件内至少一个核苷酸的修饰。
- [0027] 在某些实施方式中，所述至少一种基因表达调节剂包含选自DNA甲基转移酶、DNA羟甲基化酶、DNA去甲基化酶、组蛋白甲基转移酶、组蛋白去甲基化酶、组蛋白乙酰转移酶、组蛋白去乙酰化酶、磷酸酶、激酶、转录活化物、转录阻遏物、以上所述的部分、以及以上所述的任意组合中的一种或多种。
- [0028] 在某些实施方式中，所述至少一个核苷酸的所述修饰是DNA甲基化。

- [0029] 在某些实施方式中，所述至少一种基因表达调节剂包含选自DNA甲基转移酶(DNMT)、基于锌指蛋白的转录因子、以上所述的部分、以及以上所述的任意组合中的一种或多种。
- [0030] 在某些实施方式中，所述至少一种基因表达调节剂包含DNA甲基转移酶(DNMT)或其部分，以及基于锌指蛋白的转录因子或其部分。
- [0031] 在某些实施方式中，所述DNA甲基转移酶是DNMT3A、DNMT3B、DNMT3L、DNMT1或DNMT2。
- [0032] 在某些实施方式中，所述DNMT3A包含SEQ ID NO:1166所示的氨基酸序列，以及所述DNMT3L包含SEQ ID NO:1167或1195所示的氨基酸序列。
- [0033] 在某些实施方式中，所述基于锌指蛋白的转录因子是Krüppel相关抑制因子(KRAB)或衍生自ZIM3的KRAB结构域(ZIM3 KRAB)。
- [0034] 在某些实施方式中，所述基于锌指蛋白的转录因子包含SEQ ID NO:1168或1196所示的氨基酸序列。
- [0035] 在某些实施方式中，所述DNA甲基转移酶选自DNMT3A和DNMT3L及其组合，且所述基于锌指蛋白的转录因子是KRAB或ZIM3 KRAB。
- [0036] 在某些实施方式中，所述融合分子包含与所述至少一种DNA结合蛋白的C端、N端或两端融合的所述至少一种基因表达调节剂。
- [0037] 在某些实施方式中，所述至少一种基因表达调节剂与所述至少一种DNA结合蛋白直接融合。
- [0038] 在某些实施方式中，所述至少一种基因表达调节剂通过非调节剂、第二调节剂或接头与所述至少一种DNA结合蛋白间接融合。
- [0039] 在某些实施方式中，所述融合分子包含结构域DNMT3A-DNMT3L-dCas9-KRAB或结构域DNMT3A-DNMT3L-ZIM3 KRAB-dCas9，其中，-表示所述融合分子各部分结构域直接和/或间接地连接，且所述各部分结构域按照从N端到C端的顺序。
- [0040] 在某些实施方式中，所述融合分子包含SEQ ID NO:1169或1194所示的氨基酸序列。
- [0041] 在某些实施方式中，所述融合分子还包含至少一个核定位序列(NLS)。

- [0042] 在某些实施方式中，所述至少一个核定位序列与所述至少一种DNA结合蛋白的C端、N端或两端直接或间接融合。
- [0043] 在某些实施方式中，编码所述融合分子的所述核酸序列是脱氧核糖核酸(DNA)或信使核糖核酸(mRNA)。
- [0044] 在某些实施方式中，所述融合分子包装在脂质体或脂质纳米颗粒中。
- [0045] 在某些实施方式中，所述融合分子和所述sgRNA包装在脂质体或脂质纳米颗粒中。
- [0046] 在某些实施方式中，所述融合分子和所述sgRNA包装在相同的脂质体或脂质纳米颗粒中，或者包装在不同的脂质体或脂质纳米颗粒中。
- [0047] 在某些实施方式中，所述脂质体或所述脂质纳米颗粒包含可电离脂质(20%–70%，摩尔比)、聚乙二醇化脂质(0%–30%，摩尔比)、支撑脂质(30%–50%，摩尔比)和胆固醇(10%–50%，摩尔比)。
- [0048] 在某些实施方式中，所述可电离脂质选自pH响应性可电离脂质、热响应性可电离脂质和光响应性可电离脂质。
- [0049] 在某些实施方式中，所述融合分子包装在AAV载体中。
- [0050] 在某些实施方式中，所述融合分子和所述sgRNA包装在AAV载体中。
- [0051] 在某些实施方式中，所述融合分子和所述sgRNA包装在相同的AAV载体或不同的AAV载体中。
- [0052] 在某些实施方式中，所述组合物是包含药学上可接受的载体的药物组合物。
- [0053] 另一方面，本申请提供一种单一引导RNA(sgRNA)，所述sgRNA包含与靶DNA序列互补的序列，所述靶DNA序列位于乙型肝炎病毒(HBV)基因附近和/或HBV基因调控元件内；所述HBV基因为包含SEQ ID NO:1所示的核苷酸序列的B型HBV基因、包含SEQ ID NO:2所示的核苷酸序列的C型HBV基因或者包含SEQ ID NO:3所示的核苷酸序列的D型HBV基因，所述HBV基因调控元件包含转录起始位点、核心启动子、启动子、增强子、沉默子、绝缘子元件、边界元件和/或基因座控制区，且所述靶DNA序列位于所述HBV基因的第1056位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。

- [0054] 在某些实施方式中，所述靶DNA序列位于所述HBV基因的第1056位至第1900位核苷酸之间、第1972位至第2082位核苷酸之间、第2134位核苷酸至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。
- [0055] 在某些实施方式中，所述靶DNA序列位于所述HBV基因的第1060位至第1079位核苷酸之间、第1149位至第1612位核苷酸之间、第1693位至1852位核苷酸之间、和/或第2863位至第2882位核苷酸之间。
- [0056] 在某些实施方式中，所述靶DNA序列位于下列所述B型HBV基因的一个或多个区域：第1149至第1190位核苷酸之间、第1210位至第1310位核苷酸之间、第1350位至第1400位核苷酸之间、第1420位至第1450位核苷酸之间和第1470位至第1592位核苷酸之间。
- [0057] 在某些实施方式中，所述靶DNA序列位于下列所述C型HBV基因的一个或多个区域：第1150位至第1180位核苷酸之间、第1200位至第1310位核苷酸之间、第1350位至第1390位核苷酸之间、第1420位至第1460位核苷酸之间和第1480位至第1593位核苷酸之间。
- [0058] 在某些实施方式中，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1056位至第1079位核苷酸之间、第1101位至第1900位核苷酸之间、第1972位至第2082位核苷酸之间、第2134位至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2882位核苷酸之间、第2911位至第2930位核苷酸之间、和第3048位至第3067位核苷酸之间。
- [0059] 在某些实施方式中，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1060位至第1079位核苷酸之间、第1160位至第1310位核苷酸之间、第1253位至第1284位核苷酸之间、第1370位至第1470位核苷酸之间、第1490位至第1612位核苷酸之间、第1693位至第1852位核苷酸之间、和第2863位至第2882位核苷酸之间。。
- [0060] 在某些实施方式中，所述sgRNA包含SEQ ID NOs:4-1165中任一项所示的核苷酸序列。

- [0061] 在某些实施方式中，所述sgRNA包含所述SEQ ID NOs:4-1165中任一项所示的核苷酸序列的部分序列，所述部分序列的长度为15-20个碱基对。
- [0062] 另一方面，本申请提供编码本申请所述的sgRNA的核酸分子。
- [0063] 另一方面，本申请提供减少或消除细胞中表达乙型肝炎病毒(HBV)基因产物的方法，所述方法包括将本申请所述的组合物引入所述细胞中，从而减少或消除所述细胞中表达所述HBV基因产物的步骤。
- [0064] 另一方面，本申请提供体内减少或消除受试者中表达乙型肝炎病毒(HBV)基因产物的方法，所述方法包括将本申请所述的组合物引入所述受试者的细胞中，从而减少或消除所述受试者中表达所述HBV基因产物的步骤。
- [0065] 另一方面，本申请提供用于治疗受试者乙型肝炎病毒(HBV)感染相关疾病或减轻受试者HBV感染相关疾病症状的方法，所述方法包括将有效量的本申请所述的组合物引入到所述受试者的细胞的步骤。
- [0066] 在某些实施方式中，所述受试者为哺乳动物，诸如人、猴、小鼠、大鼠、兔、猪、马、猫和狗。
- [0067] 在某些实施方式中，所述方法包括向所述受试者施用一次或一次以上的所述组合物。
- [0068] 在某些实施方式中，所述方法包括向所述受试者施用至少两次所述组合物。
- [0069] 在某些实施方式中，所述施用一次以上或施用至少两次所述组合物的间隔为1天、2天、3天、4天、5天、6天、7天、8天、9天、10天、11天、12天、13天、14天或15天。例如，所述方法包括每天一次、每2天一次、每3天一次、每4天一次、每5天一次、每6天一次、每7天一次、每8天一次、每9天一次、每10天一次、每11天一次、每12天一次、每13天一次、每14天一次、或每15天一次地向所述受试者施用所述组合物，直至所述受试者达到临床治愈。
- [0070] 在某些实施方式中，所述HBV感染相关疾病包括由HBV感染引起的肝炎、肝硬化、肝纤维化和肝细胞癌。
- [0071] 在某些实施方式中，所述组合物用于治疗受试者乙型肝炎病毒(HBV)感染相关疾病或减轻受试者HBV感染相关疾病症状。

[0072] 另一方面，本申请提供一种试剂盒，所述试剂盒包含本申请所述的组合物以及用于放置所述组合物的容器和/或用法说明书。

[0073] 本领域技术人员能够从下文的详细描述中容易地洞察到本申请的其它方面和优势。下文的详细描述中仅显示和描述了本申请的示例性实施方式。如本领域技术人员将认识到的，本申请的内容使得本领域技术人员能够对所公开的具体实施方式进行修改而不脱离本申请所涉及发明的精神和范围。相应地，本申请的附图和说明书中的描述仅仅是示例性的，而非为限制性的。

附图说明

[0074] 本申请所涉及的发明的具体特征如所附权利要求书所显示。通过参考下文中详细描述 of 示例性实施方式和附图能够更好地理解本申请所涉及发明的特点和优势。对附图简要说明如下：

[0075] 图1A显示的是本申请所述的HBV病毒蛋白表达的报告系统(reporter)示意图。

[0076] 图1B显示的利用图1A所示的reporter细胞系进行sgRNA功能验证的筛选方案示意图。

[0077] 图2A–2F显示的是利用图1A所示的HBx reporter细胞系对B型HBV基因分别进行表观编辑14天、21天和28天后的荧光强度变化。图2G显示的是sgRNA在B型HBV基因组中的靶向位置与敲降能力的关系(纵坐标为转染后28天不同sgRNA敲降细胞的比例，可代表sgRNA的编辑能力；横坐标为各个sgRNA的靶序列起始位点在B型HBV基因组中的位置)。

[0078] 图3A–3B显示的是利用图1A所示的HBx reporter细胞系对C型HBV基因进行表观编辑14天后的荧光强度变化。图3C显示的是sgRNA在C型HBV基因组中的靶向位置与敲降能力的关系(纵坐标为转染后14天不同sgRNA敲降细胞的比例，可代表sgRNA的编辑能力；横坐标为各个sgRNA的靶序列起始位点在C型HBV基因组中的位置)。

[0079] 图4A–4F显示的是利用图1A所示的HBx reporter细胞系对D型HBV基因进行表观编辑16天、21天和28天后的荧光强度变化。图4G显示的是sgRNA在D型HBV基因组中的靶向位置与敲降能力的关系(纵坐标为转染后28天不同sgRNA敲降细胞

的比例，可代表sgRNA的编辑能力；横坐标为各个sgRNA的靶序列起始位点在D型HBV基因组中的位置)。

[0080] 图5A显示的是本申请所述的方法在原代肝细胞(PHH)HBV感染模型中进行表观编辑功能验证的方案示意图。

[0081] 图5B显示的是本申请所述的药物组合物(分别包含不同sgRNA)在分别递送至HBV感染的原代肝细胞2天、4天、6天、8天和10天后的抗原水平结果。

[0082] 图6显示的是reporter细胞系筛选sgRNA文库的流程示意图。

[0083] 图7显示的是D型HBV基因HBx、HBcAg和HBsAg三种reporter细胞系筛选sgRNA文库的结果(纵坐标为转染14天后，不同sgRNA在GFP阴性细胞群中与在原始文库中的NGS测序读数的变化倍数，可代表sgRNA的编辑能力；横坐标为各个sgRNA的靶序列起始位点在D型HBV基因组中的位置)。

[0084] 图8显示的是文库筛选出的候选sgRNA在原代肝细胞(PHH)中效果验证，图示结果为HBV感染后14天HBV各标志物的抑制率水平以及PHH细胞活力。

[0085] 图9A-9C显示的是EPIREG在转基因HBV小鼠中对HBV标记物的敲低效果。图示结果分别为各个对照/给药组HBV标记物的总体变化曲线、各个给药组每只小鼠的HBV标记物的变化曲线、以及给药后产生抗体的三只小鼠的HBsAb变化曲线，各曲线所示的HBV标记物含量均以10为底取对数用于数据分析。

[0086] 图10显示的是EPIREG在AAV-HBV小鼠中对HBV标记物的敲低效果。图示的HBV标记物含量均以10为底取对数用于数据分析。

具体实施方式

[0087] 以下由特定的具体实施例说明本申请发明的实施方式，熟悉此技术的人士可由本说明书所公开的内容容易地了解本申请发明的其他优点及效果。

[0088] 术语定义

[0089] 在申请中，术语“融合分子”通常是指至少两个部分组成的分子(bipartite molecule)，例如，在本申请中，其包含本申请所述的至少一种DNA结合蛋白和本申请所述的至少一种基因表达调节剂偶联，从而形成单个实体。例如，所述至少一种基因表达调节剂可以在N端、C端或末端氨基酸外的任意氨基酸处与所述至少一种DNA结合蛋白融合，其他分子或部分则也可以与已经包含在该融

合分子中的部分融合。组成该融合分子的部分可以由接头分开，或者可以直接偶联。在某些实施方式中，该融合分子是一种融合蛋白，所述融合蛋白可以通过直接或间接共价或非共价连接两个或更多个基因而产生的嵌合蛋白质，所述基因最初编码单独的蛋白质。在某些实施方式中，融合基因的翻译产生具有源自每种原始蛋白质的功能特性的单一多肽。本领域的技术人员充分了解用于确定本申请的融合分子中的部分的测定最佳顺序和/或组合。

[0090] 在本申请中，术语“多核苷酸”、“核苷酸”、“核苷酸序列”、“核酸”和“寡核苷酸”可互换地使用。它们通常是指具有任何长度的核苷酸的聚合形式，是脱氧核糖核苷酸或核糖核苷酸、或其类似物。多核苷酸可具有任何三维结构，并且可以执行已知或未知的任何功能。以下是多核苷酸的非限制性实例：基因或基因片段的编码区或非编码区、根据连接分析定义的多个座位(一个座位)、外显子、内含子、信使RNA(mRNA)、转运RNA、核糖体RNA、短发夹RNA(shRNA)、micro-RNA(miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离的DNA、任何序列的分离的RNA、核酸探针、和引物。多核苷酸可以包含一个或多个经修饰的核苷酸，如甲基化的核苷酸和核苷酸类似物。如果存在，可以在聚合物组装之前或之后进行核苷酸结构的修饰。核苷酸的序列可以被非核苷酸组分中断。多核苷酸可以在聚合后，如通过与标记的组分缀合来进一步修饰。

[0091] 在本申请中，术语“DNA结合蛋白”通常是指由一个或多个DNA结合结构域(不同功能的结构域)组成的较大蛋白质，所述DNA结合结构域是独立折叠的蛋白质结构域，其含有识别双链或单链DNA的至少一个基序。例如，所述DNA结合结构域可识别特异性DNA序列(识别或调节序列)或具有对DNA的一般亲和性。在某些情形下，DNA结合蛋白的其他结构域通常调节DNA结合结构域的活动性；DNA结合功能可以是结构性的或者包括转录调节，有时这两种作用是重叠的。在根据本申请所提供的方法和组合物的某些实施方案中，DNA结合蛋白可包含(DNA)核酸酶，诸如能够以序列特异性方式靶向DNA或者能够被指导或指示以序列特异性方式靶向DNA的核酸酶，诸如CRISPR-Cas系统、锌指核酸酶(ZFN)、转录激活子样效应因子核酸酶(TALEN)或大范围核酸酶。在一些实施方

案中，DNA结合蛋白是源自CRISPR-Cas系统的DNA核酸酶。例如，所述源自CRISPR-Cas系统的DNA核酸酶是Cas蛋白。

[0092] 在本申请中，术语“Cas蛋白”可与“CRISPR蛋白”、“CRISPR酶”、“CRISPR-Cas蛋白”、“CRISPR-Cas酶”、“Cas”、“CRISPR效应子”或“Cas效应子蛋白”互换地使用，是CRISPR-Cas系统的组分之一。Cas蛋白(例如，工程化的Cas蛋白)可具有与野生型对应Cas蛋白基本相同(例如，介于80%与100%之间，介于90%与100%之间，介于95%与100%之间，介于98%与100%之间，介于99%与100%之间，介于99.9%与100%之间，或约100%)的核酸酶活性。在某些情况下，工程化的Cas蛋白具有比野生型对应Cas蛋白高(例如，高至少5%、至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%或至少90%)的核酸酶活性。可选地或另外地，Cas蛋白(例如，工程化的Cas蛋白)可具有比野生型对应Cas蛋白高至少5%、至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%或至少90%的特异性。在特定的实例中，Cas蛋白(例如，工程化的Cas蛋白)具有比野生型对应Cas蛋白高至少30%的特异性。如本文中所用，术语Cas的“特异性”可对应于中靶多核苷酸切割事件相对于所有多核苷酸切割事件(包括中靶和脱靶事件)的数量或百分比的数量或百分比。Cas蛋白的活性和特异性与以下文献中描述的那些一致：Hsu PD等人，DNA targeting specificity of RNA-guided Cas9 nucleases, Nat Biotechnol. 2013年9月；31(9):827-832和Slaymaker IM等人，Rationally engineered Cas9 nucleases with improved specificity, Science. 2016年1月1日；351(6268):84-88，所述文献也描述了用于检测Cas蛋白的活性和特异性的方法的实例，并通过引用以其整体并入本文。

[0093] 可对编码Cas的核酸分子进行密码子优化。经密码子优化的序列的实例在这种情况下是针对在真核生物(例如，人)中表达而优化(即针对在人中表达而优化)的序列，或针对另一真核生物如本文所论述的动物或哺乳动物而优化的序列；参见，例如，WO 2014/093622(PCT/US2013/074667)中的经SaCas9人密码子优化的序列。虽然这是优选的，但是应当理解，其它实例也是可能的，并且针对除人外的宿主物种的密码子优化或针对特定器官的密码子优化是已知的。在一些

实施方案中，对编码Cas的酶编码序列进行密码子优化以用于在特定细胞诸如真核细胞中表达。真核细胞可以是特定生物体(诸如哺乳动物，包括但不限于人或非人真核生物或本文所述的动物或哺乳动物，例如小鼠、大鼠、兔、狗、家畜或非人哺乳动物或灵长类动物)的细胞或源自特定生物体的细胞。在一些实施方案中，可以排除用于改变人类种系遗传特征的方法和/或用于改变动物遗传特征的方法(所述方法可能导致他们遭受痛苦，而对人或动物没有任何实质性的医学益处)，以及由此类方法产生的动物。一般而言，密码子优化是指通过用在该宿主细胞的基因中更频繁或最频繁使用的密码子替换天然序列的至少一个密码子(例如，约或超过约1个、2个、3个、4个、5个、10个、15个、20个、25个、50个或更多个密码子)，同时保持天然氨基酸序列，来修饰核酸序列以增强在目标宿主细胞中的表达的过程。不同物种对特定氨基酸的特定密码子表现出特定的偏性。密码子偏性(生物体之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关，而这又被认为取决于，除其它以外，被翻译的密码子的性质和特定转运RNA(tRNA)分子的可用性。细胞中所选tRNA占优势通常反映了肽合成中最常用的密码子。因此，可基于密码子优化，针对给定生物中的最佳基因表达来定制基因。密码子使用表可在例如于www.kazusa.or.jp/codon/可获得的“密码子使用数据库”中容易地获得，并且这些表可以以多种方式进行修改。参见Nakamura,Y等人“Codon usage tabulated from the international DNA sequence databases:status for the year 2000” Nucl.Acids Res.28:292(2000)。用于密码子优化特定序列以在特定宿主细胞中表达的计算机算法也是可用的，诸如Gene Forge(Appagen; Jacobus, PA)，也是可用的。在一些实施方案中，编码Cas的序列中的一个或多个密码子(例如1个、2个、3个、4个、5个、10个、15个、20个、25个、50个或更多个，或所有密码子)对应于特定氨基酸最常用的密码子。

[0094] 在一些实施方案中，Cas蛋白可具有核酸切割活性。Cas蛋白可具有RNA结合和DNA切割功能。在一些实施方案中，Cas可指导靶序列的位置处或其附近，诸如靶序列内和/或靶序列的互补序列内或与靶序列相关的序列处，例如距靶序列的第一个或最后一个核苷酸约1个、2个、3个、4个、5个、6个、7个、8个、9

个、10个、15个、20个、25个、50个、100个、200个、500个或更多个碱基对内的一条或两条核酸链的切割。在一些实施方案中，Cas蛋白可指导靶序列内和/或靶序列的互补序列内或与靶序列相关的序列处和/或距靶序列的第一个或最后一个核苷酸约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、25个、50个、100个、200个、500个或更多个碱基对内的一条或两条链的不止一次切割(例如一次、两次、三次、四次、五次或更多次切割)。在一些实施方案中，切割可以是平端的，即产生平端。在一些实施方案中，裂解可以是交错的，即产生粘性末端。

- [0095] 在一些实施方案中，载体编码靶向核酸的Cas蛋白，可相对于相应的野生型酶对所述Cas蛋白进行突变，使得突变的靶向核酸的Cas蛋白缺乏切割含有靶序列的靶多核苷酸的一条或两条链的能力，例如，HNH结构域中的改变或突变产生基本上缺乏所有DNA切割活性的突变的Cas，例如，突变的酶的DNA切割活性约为不超过该酶的非突变形式的核酸切割活性的25%、10%、5%、1%、0.1%、0.01%或更低；一个实例是当与非突变形式相比，突变形式的核酸切割活性为零或可忽略时。
- [0096] 在一些实施方案中，Cas蛋白可形成诱导型系统的组分。该系统的可诱导性质将允许使用一种形式的能量对基因编辑或基因表达进行时空控制。能量的形式可包括但不限于电磁辐射、声能、化学能和热能。诱导型系统的实例包括四环素诱导型启动子(Tet-On或Tet-Off)、小分子双杂交转录激活系统(FKBP、ABA等)或光诱导型系统(光敏色素、LOV结构域或隐花色素)。在一个实施方案中，CRISPR效应蛋白可以是光诱导型转录效应子(LITE)的一部分，以序列特异性方式指导转录活性的变化。光的组分可包括CRISPR效应蛋白、光响应性细胞色素异二聚体(例如来自拟南芥)和转录激活/阻遏结构域。在US 61/736465和US 61/721,283以及WO 2014018423 A2(其据此通过引用以其整体并入)中提供了诱导型DNA结合蛋白及其使用方法的其它实例。
- [0097] 在一些实施方案中，突变的Cas可具有一个或多个导致脱靶效应降低的突变，例如，用于实现对靶基因座的修饰但降低或消除脱靶活性的改进的CRISPR酶(诸如当与引导RNA复合时)，以及用于增强CRISPR酶活性的改进的CRISPR酶

(诸如当与引导RNA复合时)。应当理解，如下文所述的突变的酶可用于本文如其它地方所述的根据本申请的任何方法中。如本文其它地方所述的任何方法、产品、组合物和用途同样适用于如下文进一步详述的突变的CRISPR酶。

- [0098] 可用于各种组合中以增强或降低中靶活性对比脱靶活性的活性和/或特异性，或者增强或降低中靶结合对比脱靶结合的结合和/或特异性的方法和突变，可用于补偿或增强为促进其它作用而进行的突变或修饰。为促进其它作用而进行的此类突变或修饰包括对Cas的突变或修饰和/或对引导RNA进行的突变或修饰。本申请的方法和突变用于调节Cas核酸酶活性和/或与化学修饰的引导RNA的结合。
- [0099] 在某些实施方案中，本申请的Cas蛋白的催化活性被改变或修饰。应当理解，如果催化活性不同于相应的野生型Cas蛋白(例如，未突变的Cas蛋白)的催化活性，则突变的Cas具有改变的或经修饰的催化活性。催化活性可通过本领域已知的方法测定。举例来说，而非限制，通过测定插入/缺失(indel)百分比(例如在给定时间后，或在给定剂量下)，可以在体外或体内测定催化活性。在某些实施方案中，催化活性得以增强。在某些实施方案中，催化活性增强了至少5%、至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少100%。在某些实施方案中，催化活性得以降低。在某些实施方案中，催化活性降低至少5%，优选至少10%，更优选至少20%，诸如至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%，或(基本上)100%。本文中的一个或多个突变可使催化活性失活，这可显著降低所有催化活性，将活性降低至低于可检测水平，或降低至不可测量的催化活性。
- [0100] 工程化的Cas蛋白的一种或多种特征可以不同于相应的野生型Cas蛋白。此类特征的实例包括催化活性、gRNA结合、Cas蛋白的特异性(例如，编辑确定靶标的特异性)、Cas蛋白的稳定性、脱靶结合、靶结合、蛋白酶活性、切口酶活性、PFS识别。在一些实例中，工程化的Cas蛋白可包含相应的野生型Cas蛋白的一个或多个突变。在一些实施方案中，与相应的野生型Cas蛋白相比，工程化的Cas蛋白的催化活性增强。在一些实施方案中，与相应的野生型Cas蛋白相比，工程化的Cas蛋白的催化活性降低。在一些实施方案中，与相应的野生型Cas蛋白相

比，工程化的Cas蛋白的gRNA结合增加。在一些实施方案中，与相应的野生型Cas蛋白相比，工程化的Cas蛋白的gRNA结合减少。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的特异性增强。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的特异性降低。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的稳定性增强。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的稳定性降低。在一些实施方案中，工程化的Cas蛋白还包含一个或多个使催化活性失活的突变。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的脱靶结合增加。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的脱靶结合减少。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的靶结合增加。在一些实施方案中，与相应的野生型Cas蛋白相比，Cas蛋白的靶结合降低。在一些实施方案中，与相应的野生型Cas蛋白相比，工程化的Cas蛋白具有更高的蛋白酶活性或多核苷酸结合能力。在一些实施方案中，与相应的野生型Cas蛋白相比，PFS识别被改变。

[0101] Cas蛋白的实例包括I类(例如，I型、III型和IV型)和2类(例如，II型、V型和VI型)Cas蛋白，例如，Cas9、Cas12(例如，Cas12a、Cas12b、Cas12c、Cas12d)、Cas13(例如，Cas13a、Cas13b、Cas13c、Cas13d)、CasX、CasY、Cas14、其变体(例如，突变形式、截短形式)、其同源物及其直系同源物。术语“直系同源物”和“同源物”在本领域是公知的。通过进一步的指导，本文所用蛋白质的“同源物”是与作为其同源物的蛋白质执行相同或相似功能的相同物种的蛋白质。同源蛋白质可以是但不必是结构相关的，或者只是部分结构相关的。本文所用蛋白质的“直系同源物”是与作为其直系同源物的蛋白质执行相同或相似功能的不同物种的蛋白质。直系同源蛋白质可以是但不必是结构相关的，或者只是部分结构相关的。

[0102] 在一些实施方案中，Cas蛋白是2类Cas蛋白，即2类CRISPR-Cas系统的Cas蛋白。2类CRISPR-Cas系统可具有亚型，例如II-A型、II-B型、II-C型、V-A型、V-B型、V-C型或V-U型。在一些实施方案中，Cas蛋白是Cas9、Cas12a、Cas12b、Cas12c或Cas12d。在一些实施方案中，Cas9可以是SpCas9、SaCas9、StCas9和其它Cas9直系同源物。Cas12可以是Cas12a、Cas1

2b和Cas12c, 包括FnCas12a或其同源物或直系同源物。CRISPR-Cas系统的定义和示例性成员包括以下文献中描述的那些: Kira S.Makarova和Eugene V.Koonin, Annotation and Classification of CRISPR-Cas systems, *Methods Mol Biol.* 2015; 1311:47-75和Sergey Shmakov等人, Diversity and evolution of class 2 CRISPR-Cas systems, *Nat Rev Microbiol.* 2017年3月; 15(3):169-182。

- [0103] 在一些实例中, Cas蛋白包含至少一个RuvC结构域和至少一个HNH结构域。Cas蛋白还可包含连接RuvC结构域与HNH结构域的第一和第二接头结构域。在Nishimasu,H.等人“Crystal structure of Cas9 in complex with guide RNA and target RNA” *Cell* 156(2014年2月27日, 2014):935-949和Ribeiro,L.等人(2018)“Protein engineering strategies to expand CRISPR-Cas9 applications” *International Journal of Genomics Volume 2018, Article ID 1652567*(doi:10.1155/2018/1652567)的研究中描述了Cas9中连接HNH与RuvC结构域的第一接头(L1)和第二接头(L2)。Ribeiro的图1显示了Cas9的总体组织、结构和功能, 该文献特别地通过引用并入本文。具体而言, 图1A显示了SpCas9的结构域组织的示意图, 表明了包括如本文所述的接头L1(跨越氨基酸765-780)和L2(跨越氨基酸906-918)的HNH和RuvC结构域的遗传结构。类似地, 当提及第一和第二接头结构域时, 可以利用金黄色葡萄球菌Cas9(SaCas9)的结构域组织。一方面, 接头1结构域区域跨越残基481-519, 并将RuvC-II结构域与SaCas9中的HNH结构域相连接。在一些实施方案中, 接头2区域跨越残基629-649, 并连接SaCas9的RuvC-III结构域与HNH结构域。因此, 可在Cas9直系同源物中突变第一和/或第二接头结构域, 并且可参考对应于野生型SaCas9的氨基酸的氨基酸残基。参见, Nishimasu, *Cell.* 2015年8月27日; 162(5):1113-1126; doi:10.1016/j.cell.2015.08.007, 通过引用并入本文。特别地, 图1, Nishimasu的S1-S3详细描述了Cas9蛋白的结构域组织, 其教导通过引用特别并入本文。第一和第二接头可包含约10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42

个、43个、44个、45个或更多个氨基酸。第一和第二接头可对应于野生型接头。在一些方面，第一和第二接头可在第一和/或第二接头中包含一个或多个突变。一方面，第一和/或第二接头包含一个或多个提高Cas9蛋白特异性的突变。在一些实施方案中，连接Cas9的HNH与RuvC结构域的接头L1和L2含有野生型氨基酸序列。在一些实施方案中，连接HNH与RuvC结构域的接头在一个或多个氨基酸中含有突变。在实施方案中，第一接头(L1)包含对应于SpCas9的氨基酸T769I的突变，和/或第二接头(L2)包含对应于SpCas9的氨基酸G915M的突变。在实施方案中，一个或多个接头突变，例如T769I和G915M，赋予Cas9蛋白提高的特异性。在一个实施方案中，如本文中所述，可将第一和第二接头中的一个或多个突变与Cas9蛋白的其它部分中的一个或多个突变组合，以进一步提高特异性和/或保持与野生型Cas9蛋白基本等同的活性。在一个实施方案中，可利用本文详述的方法鉴定接头中的突变和/或Cas蛋白中的额外突变，所述方法增强/改善了对野生型Cas9的特异性并基本上保留了其野生型活性。

[0104] 在一些实施方案中，Cas蛋白可以是2类II型CRISPR-Cas系统的Cas蛋白(II型Cas蛋白)。在一些实施方案中，Cas蛋白可以是2类II型Cas蛋白，例如Cas9。在一些实施方案中，基于CRISPR/Cas9的系统可包括Cas9蛋白或其片段、Cas9融合蛋白、编码Cas9蛋白或其片段的核酸、或编码Cas9融合蛋白的核酸。

“Cas9(CRISPR相关蛋白9)”是指与NCBI保藏号NP_269215具有至少约85%氨基酸同一性并具有RNA结合活性、DNA结合活性和/或DNA切割活性(例如，核酸内切酶或切口酶活性)的多肽或其片段。Cas9的功能可通过多种测定法中的任一种来定义，所述测定法包括但不限于基于荧光偏振的核酸结合测定法、基于荧光偏振的链侵入测定法、转录测定法、EGFP破坏测定法、DNA切割测定法和/或Surveyor测定法。“Cas9核酸分子”是指编码Cas 9多肽或其片段的多核苷酸。在基因组序列号NC_002737下提供了示例性Cas9核酸分子序列。在一些实施方案中，本文公开了Cas9的抑制剂，例如化脓性链球菌(SpCas9)或金黄色葡萄球菌(SaCas9)中天然存在的Cas9或其变体。Cas9利用前间区序列邻近基序(Protospacer Adjacent Motif)(PAM)序列和引导RNA(gRNA)对靶DNA的碱基配对来识别外源DNA。Cas9在任何基因组座位上诱导靶向链断裂的相对容易性使得

在多种细胞类型和生物体中进行高效的基因组编辑成为可能。Cas9衍生物也可用作转录激活物/阻遏物。

- [0105] 在一些情况下，CRISPR-Cas蛋白是Cas9或其变体。在一些实例中，Cas9可以是包括任何天然存在的细菌Cas9在内的野生型Cas9，可以是经密码子优化的或经修饰的形式，包括任何嵌合体、突变体、同源物或直系同源物。在本申请的另一方面，Cas9酶可包含一个或多个突变，并且可用作与功能域融合或不与其融合的通用DNA结合蛋白。突变可以是人工引入的突变或功能获得或功能丧失的突变。本申请的其它方面涉及与结构域融合的突变的Cas9酶，所述结构域包括但不限于核酸酶、转录激活物、转录阻遏物、重组酶、转座酶、组蛋白重塑物、去甲基化酶、DNA甲基转移酶、隐色素、光诱导型/可控结构域或化学诱导型/可控结构域。在一些情况下，Cas9酶可来自或衍生自SpCas9(化脓性链球菌Cas9)、saCas9(金黄色葡萄球菌Cas9)、或StCas9(嗜热链球菌的野生型Cas9)。如本文中所用，关于酶的术语“衍生”意指衍生的酶在很大程度上基于与野生型酶具有高度序列同源性的意义，但已经以本领域已知的或本文所述的一些方式对其进行了突变(修饰)。在实例中，突变可包括蛋白质的第一接头结构域、第二接头结构域和/或其它部分中的一个或多个突变。相对于野生型酶，高度序列同源性可包括至少80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高。
- [0106] 在特定的实施方案中，CRISPR酶可以是来自或源自包括以下的属的生物体的Cas9蛋白：链球菌属(*Streptococcus*)、弯曲杆菌属(*Campylobacter*)、硝化细菌属(*Nitratifactor*)、葡萄球菌属(*Staphylococcus*)、细小棒菌属(*Parvibaculum*)、罗氏菌属(*Roseburia*)、奈瑟菌属(*Neisseria*)、葡糖醋杆菌属(*Gluconacetobacter*)、固氮螺菌属(*Azospirillum*)、鳞球菌属(*Sphaerochaeta*)、乳杆菌属(*Lactobacillus*)、肠道真杆菌属(*Eubacterium*)、棒状杆菌属(*Corynebacter*)、肉食杆菌属(*Carnobacterium*)、红杆菌属(*Rhodobacter*)、李斯特菌属(*Listeria*)、Paludibacter、梭菌属(*Clostridium*)、毛螺菌科(*Lachnospiraceae*)、Clostridiaridium、纤毛菌属(*Leptotrichia*)、弗朗西斯菌属(*Francisella*)、军团菌属(*Legionella*)、脂环酸芽孢杆菌属(*Alicyclobacillus*)、Methanomethyophilus、卟啉单胞菌属(*Porphyromonas*)、普雷沃氏菌属

(Prevotella)、拟杆菌门(Bacteroidetes)、创伤球菌属(Helcococcus)、钩端螺旋体属(Letospira)、脱硫弧菌属(Desulfovibrio)、脱硫嗜盐碱菌属(Desulfonatronum)、丰佑菌科(Opitutaceae)、肿块芽孢杆菌属(Tuberibacillus)、芽孢杆菌属(Bacillus)、短小芽孢杆菌属(Brevibacillus)、甲基杆菌属(Methylobacterium)或氨基酸球菌属(Acidaminococcus)、链球菌属、弯曲杆菌属、硝化细菌属、葡萄球菌属、细小棒菌属(Parvibaculum)、罗氏菌属、奈瑟菌属、葡萄糖杆菌属、固氮螺菌属、鳞球菌属、乳杆菌属、肠道真杆菌属、棒状杆菌属、萨特氏菌属(Sutterella)、军团菌属、(Treponema)、产线菌属(Filifactor)、肠道真杆菌属、链球菌属、乳杆菌属、支原体属(Mycoplasma)、拟杆菌属(Bacteroides)、Flaviivola、黄杆菌属(Flavobacterium)、鳞球菌属、固氮螺菌属、葡萄糖杆菌属、奈瑟菌属、罗氏菌属、细小棒菌属、葡萄球菌属、硝化细菌属、支原体属或弯曲杆菌属。

- [0107] 在一些实施方案中，CRISPR酶可以是来自或源自包括以下生物体的Ca_s9蛋白：变形链球菌(*S.mutans*)、无乳链球菌(*S.agalactiae*)、似马链球菌(*S.equisimilis*)、血链球菌(*S.sanguinis*)、肺炎链球菌(*S.pneumonia*)、空肠弯曲菌、大肠杆菌(*C.coli*)；卤水硝酸盐裂解菌、*N. tergarctus*；耳葡萄球菌(*S.auricularis*)、肉葡萄球菌(*S.carnosus*)；脑膜炎奈瑟菌、淋病奈瑟菌(*N. gonorrhoeae*)、单核增生李斯特菌(*L.monocytogenes*)、伊氏李斯特菌(*L.ivanovii*)；肉毒梭菌(*C.botulinum*)、艰难梭菌(*C.difficile*)、破伤风梭菌(*C.tetani*)或索氏梭菌(*C.sordellii*)、土拉热弗朗西丝菌1(*Francisella tularensis* 1)、土拉热弗朗西斯菌诺维亚种(*Francisella tularensis* subsp.*novicida*)、阿尔伯普雷沃菌(*Prevotella albensis*)、毛螺旋菌MC20171(*Lachnospiraceae* bacterium MC2017 1)、蛋白溶解丁酸弧菌(*Butyrivibrio proteoclasticus*)、异域菌门菌GW2011 GWA2_33_10(*Peregrinibacteria* bacterium GW2011 GWA2_33_10)、*Parcubacteria* bacterium GW2011GWC2_44_17、史密斯氏菌属的种SCADC(*Smithella* sp.*SCADC*)、酸氨球菌属的种BV3L6(*Acidaminococcus* sp.*BV3L6*)、毛螺旋菌MA2020(*Lachnospiraceae* bacterium MA2020)、候选白蚁甲烷支原体(*Candidatus Methanoplasma termitum*)、挑剔真杆菌(*Eubacterium eligens*)、博沃库利莫拉菌237(*Moraxella bovoculi* 237)、稻田钩端螺旋体(*Leptospira inadai*)、毛螺科菌ND2006(*Lachnospiraceae* bacterium ND2006)、白

菜卟啉单孢菌3(*Porphyromonas crevioricanis* 3)、解糖脲普雷沃菌(*Prevotella disiens*)和猕猴卟啉单孢菌(*Porphyromonas macacae*)。在一些实施方案中, Cas9蛋白来自于来自或源自化脓性链球菌、金黄色葡萄球菌或嗜热链球菌Cas9的生物体。

[0108] 在更优选实施方案中, Cas9蛋白源自选自化脓性链球菌、金黄色葡萄球菌或嗜热链球菌Cas9的细菌物种。在某些实施方案中, Cas9源自细菌物种, 所述细菌物种选自土拉弗朗西斯菌1、阿尔伯普雷沃菌、毛螺旋菌MC20171、蛋白溶解丁酸弧菌、异域菌门菌GW2011 GWA233JO、旁氏杆菌GW2011 GWC2_44_17、史密斯氏菌属种SCADC、酸氨球菌属种BV3L6BV3L6、毛螺旋菌MA2020、候选白蚁甲烷支原体、挑剔真杆菌、博沃库利莫拉菌237237、稻田钩端螺旋体、螺杆菌ND2006、白菜卟啉单孢菌3、解糖脲普雷沃菌和猕猴卟啉单孢菌。在某些实施方案中, Cas9蛋白源自选自酸氨球菌属的种BV3L6、毛螺旋菌MA2020的细菌物种。在某些实施方案中, 效应蛋白源自土拉弗朗西斯菌1的亚种, 包括但不限于土拉弗朗西斯菌诺维亚种。

[0109] Cas9蛋白包括但不限于化脓性链球菌M1血清型(UniProt ID:Q99ZW2)、金黄色葡萄球菌Cas9(UniProt ID:J7RUA5)、凸腹真杆菌(*Eubacterium ventriosum*)Cas9(UniProt ID:A5Z395)、固氮螺菌属(菌株B510)Cas9(UniProt ID:D3NT09)、嗜重氮葡萄糖醋杆菌(菌株ATCC 49037)Cas9(UniProt ID:A9HKP2)、灰色奈瑟球菌Cas9(UniProt ID:D0W2Z9)、肠道罗斯拜瑞氏菌Cas9(UniProt ID:C7G697)、食清洁剂细小棒菌(菌株DS-1)Cas9(UniProt ID:A7HP89)、卤水硝酸盐裂解菌(菌株DSM 16511)Cas9(UniProt ID:E6WZS9)、海鸥弯曲菌Cas9(UniProt ID:G1UFN3)。

[0110] 源自化脓性链球菌的Cas9或任何密切相关的Cas9的酶促作用在靶位点序列上产生双链断裂, 所述靶位点序列与引导序列的20个核苷酸杂交, 并且在靶序列的20个核苷酸之后具有前间区序列邻近基序(PAM)序列(实例包括NGG/NRG, 其可以如本文所述进行确定)。通过Cas9进行位点特异性DNA识别和切割的CRISPR活性由引导序列、与引导序列部分杂交的tracr序列和PAM序列定义。CRISPR系统的更多方面描述于Karginov和Hannon, *The CRISPR system: small RNA-guided defense in bacteria and archaea*, *Molecular Cell* 2010年1月15日; 37(1):7中。来自化脓性链球菌SF370的II型CRISPR基因座, 其包含一簇四个基因Cas9、Cas1、Cas2

和Csn1，以及两个非编码RNA元件tracrRNA和重复序列(正向重复序列)的特征阵列，所述重复序列被非重复序列的短区段(间隔区，各约30bp)隔开。在该系统中，靶向DNA双链断裂(DSB)在四个连续步骤中产生。首先，从CRISPR基因座转录两个非编码RNA(前crRNA阵列和tracrRNA)。第二，tracrRNA与前crRNA的正向重复序列杂交，然后被加工成含有单个间隔区序列的成熟crRNA。第三，成熟的crRNA:tracrRNA复合物通过crRNA的间隔区与前间隔序列DNA之间的异源双链体形成，将Cas9导向由前间隔序列和相应的PAM组成的DNA靶标。最后，Cas9介导PAM上游靶DNA的切割，以在原间隔区内产生DSB。在某些实施方案中，Cas9可以是组成型存在或诱导型存在或条件性存在或施用或递送的。Cas9优化可用于增强功能或开发新功能。可产生嵌合Cas9蛋白，Cas9可用作通用DNA结合蛋白。为Cas9提供的结构信息可用于进一步工程化和优化CRISPR-Cas系统，并且这也可推断出其它CRISPR酶系统中的结构-功能关系，特别是其它II型CRISPR酶或Cas9直系同源物中的结构-功能关系。此外，Cas9蛋白包含可易于鉴定的C端区域，该区域与转座子ORF-B同源，并包括活性RuvC样核酸酶(一个富含精氨酸的区域)。

- [0111] 在本申请中，术语“基因表达调节剂”通常可选自基因表达的阻遏物(例如，KRAB)、基因表达的激活物或表观遗传修饰的调节剂(例如，DNMT3A、DNMT3L、DNMT3A-DNMT3L融合肽)或其任意组合。本领域已知不同的基因表达调节剂，参见例如Thakore等人，Nat Methods.2016；13:127-37，通过引用以其整体并入本文。
- [0112] 在一些实施方案中，基因表达调节剂包括基因表达的阻遏物。阻遏物可以是任何已知的基因表达的阻遏物，例如，可选自Kr ü ppe1相关盒(KRAB)结构域、mSin3相互作用结构域(SID)、MAX-相互作用蛋白1(MXI1)、色影结构域(chromo shadow domain)、耳阻遏结构域(SRDX)、真核释放因子1(ERF1)、真核释放因子3(ERF3)、四环素阻遏物、lad阻遏物、Catharanthus roseus G-box结合因子1和2、格鲁乔果蝇属(Drosophila Groucho)、三联基序28(TRIM28)、核受体辅阻遏物1、核受体辅阻遏物2或以上所述的片段或融合物。例如，Kr ü ppe1相关盒(KRAB)结构域是一种类型的转录阻遏结构域，存在于许多基于锌指蛋

白的转录因子的N端部分。KRAB结构域在通过DNA结合结构域与靶DNA系连时起转录阻遏物的作用。KRAB结构域富含带电荷的氨基酸，可分为亚结构域A和B。KRAB A和B亚结构域可被可变间隔区段分开，许多KRAB蛋白仅含有A亚结构域。KRAB A亚结构域中的45个氨基酸的序列已被证明对转录抑制是重要的。B亚结构域本身并不抑制转录，但会加强KRAB A亚结构域施加的抑制作用。KRAB结构域招募辅阻遏物KAP1(KRAB相关蛋白-1，也称为转录中介因子1 β (transcription intermediary factor 1beta)、KRAB-A相互作用蛋白和三联基序蛋白28)和异染色质蛋白1(HP1)，以及其它染色质调节蛋白，通过异染色质形成导致转录抑制。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与KRAB结构域或其片段融合的dCas9。在一些实施方案中，KRAB结构域或其片段与dCas9的N端融合。在一个实施方案中，KRAB结构域或其片段与dCas9的C端融合。在一个实施方案中，KRAB结构域或其片段与dCas9分子的N端和C端两端融合。在一些实施方案中，融合分子包含KRAB结构域，所述结构域包含SEQ ID NO:1168所示的序列、与SEQ ID NO:1168基本上相同的(例如，至少80%、85%、90%、92%、95%、97%、98%、99%或更高的同一性)的序列或相对于SEQ ID NO:1160具有1个、2个、3个、4个、5个或更多个改变(例如，氨基酸取代、插入或缺失)的序列或其任意片段。在一些实施方案中，所述基于锌指蛋白的转录因子是在许多Krüppel型C2H2锌指蛋白中发现的KRAB结构域，例如，其为衍生自ZIM3的KRAB结构域(ZIM3KRAB)。例如，所述融合分子包含ZIM3 KRAB结构域，所述结构域包含SEQ ID NO:1196所示的序列、与SEQ ID NO:1196基本上相同的(例如，至少80%、85%、90%、92%、95%、97%、98%、99%或更高的同一性)的序列或相对于SEQ ID NO:1196具有1个、2个、3个、4个、5个或更多个改变(例如，氨基酸取代、插入或缺失)的序列或其任意片段。其他KRAB结构域的活性片段可以通过本领域内任何合适的比对方法进行鉴定。

- [0113] 在一些实施方案中，基因表达调节剂包括基因表达的激活物。激活物可以是任何已知的基因表达激活物，例如，VP16激活域、VP64激活域、p65激活域、爱泼斯坦-巴尔(Epstein-Barr)病毒R反式激活因子Rta分子或以上所述的片段。

可用于dCas9的激活是本领域已知的。参见，例如，Chavez等人，Nat Methods. (2016)13:563–67，通过引用以其整体并入本文。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与VP64、p65、Rta或其任意组合融合的dCas9。三联激活物VP64–p65–Rta(也称为VPR)(其中三个转录激活结构域使用短氨基酸接头融合)在与dCas9融合时，可以有效地上调靶基因表达。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与VPR融合的dCas9。

[0114] 在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与基因表达调节剂融合的dCas9，且所述基因表达调节剂包括表观遗传修饰调节剂。在一些实施方案中，融合分子通过表观遗传修饰，例如通过组蛋白乙酰化或甲基化、或DNA甲基化，在靶基因的调控元件(例如启动子、增强子或转录起始位点)处调节靶基因的表达。所述调节剂可以是任何已知的表观遗传修饰调节剂，例如，组蛋白乙酰转移酶(例如，p300催化结构域)、组蛋白脱乙酰酶、组蛋白甲基转移酶(例如，SUV39H1或G9a(EHMT2))、组蛋白去甲基化酶(例如，LSD1)、DNA甲基转移酶(例如，DNMT3A或DNMT3A–DNMT3L)、DNA去甲基化酶(例如，TET1催化结构域或TDG)或以上所述的片段。

[0115] 在某些实施方式中，所述表观遗传修饰调节剂可具有组蛋白修饰活性；组蛋白修饰活性可包括但不限于组蛋白脱乙酰酶、组蛋白乙酰转移酶、组蛋白去甲基化酶或组蛋白甲基转移酶活性。例如，所述表观遗传修饰调节剂可具有组蛋白乙酰转移酶活性，所述组蛋白乙酰转移酶可以是p300或CREB结合蛋白(CBP)蛋白或其片段。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与乙酰转移酶p300或其片段(例如，p300的催化核心)融合的dCas9。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与CREB结合蛋白(CBP)蛋白或其片段融合的dCas9。再例如，所述表观遗传修饰调节剂可具有组蛋白去甲基化酶活性。在一些实施方案中，表观遗传修饰调节剂可包括从核酸或蛋白质(例如，组蛋白)中去除甲基(CH₃–)基团的酶。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与Lys特异性组蛋白去甲基化酶1(LSD1)或其片段融合的dCas9。再例如，所述表观遗传修饰调节剂可具有组蛋白甲基转移酶活性。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包

含与SUV39H1或其片段融合的dCas9。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与G9a(EHMT2)或其片段融合的dCas9。

[0116] 在某些实施方式中，所述表观遗传修饰调节剂可具有DNA去甲基化酶活性。在一些实施方案中，表观遗传修饰调节剂可将甲基转化为羟甲基胞嘧啶，作为去甲基化DNA的机制。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与10-11易位甲基胞嘧啶双加氧酶1(TET1)或其片段融合的dCas9。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与胸腺嘧啶DNA糖基化酶(TDG)或其片段融合的dCas9。再例如，所述表观遗传修饰调节剂可具有DNA甲基化酶活性。在一些实施方案中，所述表观遗传修饰调节剂可具有涉及将甲基转移到DNA、RNA、蛋白质、小分子、胞嘧啶或腺嘌呤上的甲基化酶活性。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与DNMT3A或其片段融合的dCas9。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与DNMT3L或其片段融合的dCas9。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与DNMT3L和DNMT3L或其片段融合的dCas9。在一些实施方案中，本文提供的方法和组合物包括融合分子，其包含与DNMT3A-DNMT3L融合肽融合的dCas9。在一些实施方案中，所述具有DNA甲基化酶活性的表观遗传修饰调节剂为衍生自人的DNMT3L，其包含SEQ ID NO:1195所示的氨基酸序列。

[0117] 在本申请中，术语“引导RNA(gRNA)”与“引导分子”、“引导序列”和“单一引导RNA(sgRNA)”可互换地使用，在CRIPR-Cas系统的情形下其通常包括与靶DNA序列具有足够互补性以与靶DNA序列杂交并指导核酸靶向复合物(例如，本申请所述的组合物)与靶DNA序列特异性结合的任何多核苷酸序列。引导RNA可与靶DNA序列形成双链体。在一些实施方式中，引导RNA能够与CRISPR-Cas蛋白形成复合物，并包含与靶DNA序列具有足够互补性以与靶DNA序列杂交，并指导复合物与靶DNA序列的序列特异性结合的引导序列。CRISPR-Cas蛋白的引导分子或引导RNA可包含tracr-mate序列(在内源CRISPR系统的情况下包括“正向重复序列”)和引导序列(在内源CRISPR系统的情况下也称为“间隔区”)。在一些实施方案中，本文所述的CRISPR-Cas系统或复合

物不包含tracr序列和/或不依赖于tracr序列的存在。在某些实施方案中，引导分子可包含与引导序列或间隔区序列融合或连接的正向重复序列或基本上由其组成或由其组成。一般而言，CRISPR-Cas系统的特征在于促进CRISPR复合物在靶DNA序列位点形成的元件，其中靶DNA序列与引导序列之间的杂交促进CRISPR复合物的形成。

[0118] 在某些实施方案中，引导分子的引导序列或间隔区的长度为15至50个核苷酸。在某些实施方案中，引导RNA的间隔区长度至少为15个核苷酸。在某些实施方案中，间隔区长度为15至17个核苷酸、长度为17至20个核苷酸、长度为20至24个核苷酸、长度为23至25个核苷酸、长度为24至27个核苷酸、长度为27至30个核苷酸、长度为30至35个核苷酸或长度超过35个核苷酸。在一些实施方案中，引导序列的长度为15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个、60个、61个、62个、63个、64个、65个、66个、67个、68个、69个、70个、71个、72个、73个、74个、75个、76个、77个、78个、79个、80个、81个、82个、83个、84个、85个、86个、87个、88个、89个、90个、91个、92个、93个、94个、95个、96个、97个、98个、99个或100个核苷酸。

[0119] 在一些实施方案中，选择引导分子(正向重复序列和/或间隔区)的序列以降低引导分子内的二级结构程度。在一些实施方案中，当最佳折叠时，核酸靶向引导RNA的约或少于约75%、50%、40%、30%、25%、20%、15%、10%、5%、1%或更少的核苷酸参与自互补碱基配对。最佳折叠可通过任何合适的多核苷酸折叠算法来确定。一些程序基于计算最小吉布斯自由能。一个这样的算法的实例是mFold，如由Zuker和Stiegler(Nucleic Acids Res 9(1981),133-148)所描述的。另一个示例性折叠算法是由维也纳大学理论化学研究所开发的在线网络服务器RNAfold，其使用质心结构预测算法

(参见, 例如A.R.Gruber等人, 2008, Cell 106(1):23-24以及PA Carr和GM Church, 2009, Nature Biotechnology 27(12):1151-62)。

- [0120] 如上所述, CRISPR/Cas9系统利用gRNA提供基于CRISPR/Cas9的系统的靶向。gRNA是两个非编码RNA: crRNA和tracrRNA的融合物。通过交换编码20bp前间隔序列的序列, sgRNA可以靶向任何所需的DNA序列, 所述前间隔序列通过与所需DNA靶标的互补碱基配对赋予靶向特异性。gRNA模拟参与II型效应系统的天然存在的crRNA:tracrRNA双链体。该双链体(其可包括例如42个核苷酸的crRNA和75个核苷酸的tracrRNA)用作Cas9切割靶核酸的向导。
- [0121] 另一方面, 本申请提供的sgRNA可进一步包含支架(scaffold)序列, 所述支架序列具有: SEQ ID NO:1190所示的核苷酸序列; 或者与SEQ ID NO:1190所示的核苷酸序列具有至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%、至少99.9%或者至少100%的序列同一性且保留其生物学活性的核苷酸序列; 或者基于SEQ ID NO:1190所述的核苷酸序列改造得到的且保留其生物学活性的核苷酸序列。例如, 所述改造可以为碱基磷酸化、碱基硫化、碱基甲基化、碱基羟基化、序列的缩短和序列的加长中的一种或者多种。进一步地, 所述序列的缩短和所述序列加长可以包括相对于基础序列存在1个、2个、3个、4个、5个、6个、7个、8个、9个或者10个碱基的缺失或者添加。
- [0122] 仅作为示例, 所述支架序列可以为:
- NNNNNNNNNNNNNNNNNNNGUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAA
GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (SEQ ID NO:
1190)。
- [0123] 在一些实施方案中, 本申请提供的sgRNA可以在所述支架序列的5'端进一步包括所述CRISPR间隔区序列, 所述CRISPR间隔区序列为长度为20、21、22、23、24、25、26、27、28、29、30个核苷酸且能够与靶序列互补配对的序列。在一些优选的实施方案中, 所述CRISPR间隔区序列为长度为20个或21个核苷酸且能够与靶序列互补配对的序列。在一些实施方案中, 所述sgRNA

在所述间隔区序列的3'端可进一步包括终止子。例如,所述终止子可以为多个如至少6个(例如7个或者8个)U构成的终止子。

[0124] 在本申请中,术语“靶DNA”和“靶序列”或“前间隔序列”可互换地使用,其通常是指存在于靶核酸中的核苷酸序列,其包含与本申请的寡核苷酸(例如,引导RNA)互补的核碱基序列。在某些情形中,靶序列由靶核酸上与本申请寡核苷酸的连续核苷酸序列互补的区域组成。在某些情形中,靶序列比单个寡核苷酸的互补序列更长,并且可以例如表示可以被本申请的几种寡核苷酸靶向的靶核酸的可选区域。在某些情形中,“靶序列”可意指靶基因的部分,例如靶基因一个或多个外显子序列,内含子序列,或靶基因的调节序列,或靶基因的外显子和内含子序列、内含子和调节序列、外显子和调节序列、或外显子、内含子和调节序列的组合。在本申请的CRISPR复合物或系统形成的背景下,“靶DNA”是指引导RNA序列被设计为对其具有互补性的序列,其中在靶DNA与引导RNA序列之间的杂交促进CRISPR复合物或系统的形成。在一些实施例中,靶DNA位于细胞的细胞核或细胞质中。基于CRISPR/Cas9的系统可包括至少一个gRNA,其中gRNA靶向不同的DNA序列。靶DNA序列可以是重叠的。靶序列或前间隔序列之后是位于前间隔序列3'末端的PAM序列。不同的II型系统有不同的PAM要求。例如,II型化脓性链球菌系统使用“NGG”序列,其中“N”可以是任意核苷酸。

[0125] 在一些实施方案中,向细胞施用的gRNA的数量可以是至少1种gRNA、至少2种不同的gRNA、至少3种不同的gRNA、至少4种不同的gRNA、至少5种不同的gRNA、至少6种不同的gRNA、至少7种不同的gRNA、至少8种不同的gRNA、至少9种不同的gRNA、至少10种不同的gRNA、至少11种不同的gRNA、至少12种不同的gRNA、至少13种不同的gRNA、至少14种不同的gRNA、至少15种不同的gRNA、至少16种不同的gRNA、至少17种不同的gRNA、至少18种不同的gRNA、至少19种不同的gRNA、至少20种不同的gRNA、至少25种不同的gRNA、至少30种不同的gRNA、至少35种不同的gRNA、至少40种不同的gRNA、至少45种不同的gRNA或至少50种不同的gRNA。

[0126] 在一些实施方案中，向细胞施用的gRNA的数量可以是至少1种gRNA至至少50种不同的gRNA、至少1种gRNA至至少45种不同的gRNA、至少1种gRNA至至少40种不同的gRNA、至少1种gRNA至至少35种不同的gRNA、至少1种gRNA至至少30种不同的gRNA、至少1种gRNA至至少25种不同的gRNA、至少1种gRNA至至少20种不同的gRNA、至少1种gRNA至至少16种不同的gRNA、至少1种gRNA至至少12种不同的gRNA、至少1种gRNA至至少8种不同的gRNA、至少1种gRNA至至少4种不同的gRNA、至少4种不同的gRNA至至少50种不同的gRNA、至少4种不同的gRNA至至少45种不同的gRNA、至少4种不同的gRNA至至少40种不同的gRNA、至少4种不同的gRNA至至少35种不同的gRNA、至少4种不同的gRNA至至少30种不同的gRNA、至少4种不同的gRNA至至少25种不同的gRNA、至少4种不同的gRNA至至少20种不同的gRNA、至少4种不同的gRNA至至少16种不同的gRNA、至少4种不同的gRNA至至少12种不同的gRNA、至少4种不同的gRNA至至少8种不同的gRNA、至少8种不同的gRNA至至少50种不同的gRNA、至少8种不同的gRNA至至少45种不同的gRNA、至少8种不同的gRNA至至少40种不同的gRNA、至少8种不同的gRNA至至少35种不同的gRNA、8种不同的gRNA至至少30种不同的gRNA、至少8种不同的gRNA至至少25种不同的gRNA、8种不同的gRNA至至少20种不同的gRNA、至少8种不同的gRNA至至少16种不同的gRNA或8种不同的gRNA至至少12种不同的gRNA。在一些实施方案中，通过选择gRNA来增加或减少靶基因的转录。

[0127] 如本文中所用，术语“调控元件”是指能够控制核酸序列表达的遗传元件。例如，剪接信号、启动子序列、多腺苷酸化信号、转录终止序列、上游调节结构域、复制起点、内部核糖体进入位点(“IRES”)、增强子等，它们共同提供了编码序列在受体细胞中的复制、转录和翻译。并非所有这些控制序列都需要存在。真核生物中的转录控制信号通常包含“启动子”和“增强子”元件。启动子和增强子由DNA序列的短阵列组成，启动子是促进可操作连接的编码区转录起始的调控元件，增强子则通过增加位于同一DNA分子上最接近的启动子的活性从而增加遗传转录速度的调控元件，所述的这些序列特异性地参与转录的细胞蛋白质相互作用(Maniatis等人，Science 236:1237(1987)，通过

引用以其整体并入本文)。启动子和增强子元件已可以从各种真核生物来源中分离出来,包括酵母、昆虫和哺乳动物细胞以及病毒中的基因(类似的控制序列,即启动子,也在原核生物中发现)。特定启动子和增强子的选择取决于受体细胞类型。一些真核启动子和增强子具有广泛的宿主范围,而其它启动子和增强子在有限的细胞类型的亚组内具有功能(关于综述,参见,例如,Voss等人,Trends Biochem.Sci.,11:287(1986);和Maniatis等人(同上),通过引用以其整体并入本文)。例如,SV40早期基因增强子在来自许多哺乳动物物种的多种细胞类型中非常活跃,并已被用于在多种哺乳动物细胞中表达蛋白质(Dijkema等人,EMBO J.4:761(1985),通过引用以其整体并入本文)。源自人延伸因子1- α 基因的启动子和增强子元件(Uetsuki等人,J.Biol.Chem.,264:5791(1989); Kim等人, Gene 91:217(1990);和Mizushima and Nagata,Nucl.Acids.Res.,18:5322(1990))、劳斯肉瘤病毒的长末端重复序列(Gorman等人, Proc.Natl.Acad.Sci.U.S.A.79:6777(1982))和人巨细胞病毒(Boshart等人, Cell 41:521(1985))也可用于在不同的哺乳动物细胞类型中表达蛋白质,所述参考文献通过引用以其整体并入本文。启动子和增强子可以单独或一起天然存在。例如,逆转录病毒长末端重复序列包含启动子和增强子元件。一般来说,启动子和增强子的作用独立于被转录或翻译的基因。因此,所用的增强子和启动子相对于它们可操作连接的基因可以是“内源的”、“外源的”或“异源的”。“内源性”增强子/启动子是基因组中给定的基因天然连接的增强子/启动子。“外源”或“异源”增强子或启动子是通过遗传操作(即,分子生物学技术)与基因并置的增强子或启动子,所述并置使得该基因的转录由连接的增强子/启动子指导。表达载体上“剪接信号”的存在通常导致重组转录物的高水平表达。

- [0128] 在某些实施方式中,“剪接信号”介导从初级RNA转录物上去除内含子,由剪接供体和受体位点组成(Sambrook等人, Molecular Cloning:A Laboratory Manual, 第2版, Cold Spring Harbor Laboratory Press,New York(1989),第16.7-16.8页,通过引用以其整体并入本文)。常用的剪接供体和受体位点是来自SV40的16S RNA的剪接点。

- [0129] 在某些实施方式中，“转录终止信号”通常存在于多腺苷酸化信号的下游，长度为数百个核苷酸。例如，术语“poly A信号”或“poly A序列”表示指导新生RNA转录物的终止和多聚腺苷酸化的DNA序列。重组转录物的高效多腺苷酸化往往是必要的，因为缺少poly A信号的转录物是不稳定的，并且会被快速降解。表达载体中使用的poly A信号可以是“异源的”或“内源的”。内源性poly A信号是天然存在于基因组中给定基因的编码区的3'末端的信号。异源poly A信号是从一个基因中分离出来并与另一个基因的3'末端可操作地连接的信号。常用的异源poly A信号是SV40 poly A信号。SV40 poly A信号包含在237bp的BamHI/BclI限制性片段上，并指导终止和多聚腺苷酸化(Sambrook等人，同上，16.6–16.7，通过引用以其整体并入)。
- [0130] 在本申请中，术语“失活Cas9蛋白”可以称为“dCas9”蛋白。已知用于生成具有失活DNA切割结构域的Cas9蛋白(或其片段)的方法参见例如，Jinek等，Science.337: 816–821(2012)；Qi等，“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”，Cell.28, 152(5): 1173–83(2013)，其各自的全部内容通过引用纳入本文)。例如，已知Cas9的DNA切割结构域包括两个亚结构域，HNH核酸酶亚结构域和RuvC1亚结构域。HNH亚结构域切割与gRNA互补的链，而RuvC1亚结构域切割非互补链。这些亚结构域中的突变可使Cas9的核酸酶活性沉默。例如，突变D10A和H840A使酿脓链球菌Cas9的核酸酶活性完全失活(Jinek等，Science. 337: 816–821(2012)；Qi等，Cell.28; 152(5): 1173–83(2013))。合适的CRISPR失活或切口DNA结合结构域包括但不限于，核酸酶失活的变体Cas9结构域，其包括D10A，D10A/D839A/H840A和D10A/D839A/H840A/N863A突变结构域，如WO2015089406A1中所述，其通过引用纳入本文。在某些情形下，来自化脓性链球菌的无核酸内切酶活性的dCas9已被gRNA靶向细菌、酵母和人细胞中的基因，以通过空间位阻沉默基因表达。如本文中所用，“dCas”可指dCas蛋白或其片段。如本文中所用，“dCas9”可指dCas9蛋白或其片段。如本文中所用，术语“iCas”和“dCas”可互换使用，指无催化活性的CRISPR相关蛋白。在一个实施方案中，dCas蛋白在DNA切割结构域中包含一个或多个突

变。在一个实施方案中，dCas蛋白在RuvC或结构域中包含一个或多个突变。在一个实施方案中，dCas分子在RuvC和HNH结构域中都包含一个或多个突变。在一个实施方案中，dCas蛋白是野生型Cas蛋白的片段。在一个实施方案中，dCas蛋白包含来自野生型Cas蛋白的功能结构域，其中该功能结构域选自Reel结构域、桥螺旋结构域或PAM相互作用结构域。在一个实施方案中，与相应的野生型Cas蛋白的核酸酶活性相比，dCas的核酸酶活性降低了至少40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或99%。

[0131] 合适的dCas可衍生自野生型Cas蛋白。Cas蛋白可来自I型、II型或III型CRISPR-Cas系统。在一个实施方案中，合适的dCas可衍生自Cas1、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9或Cas10。在一个实施方案中，dCas衍生自Cas9蛋白。例如，可通过在Cas9蛋白的DNA切割结构域(例如核酸酶结构域，例如RuvC和/或HNH结构域)中引入点突变(例如，取代、缺失或添加)来获得dCas9。参见，例如，Jinek等人，Science(2012)337:816-21，通过引用以其整体并入本文。例如，在RuvC和HNH结构域中引入两个点突变降低了Cas9核酸酶活性，同时保留了Cas9 sgRNA和DNA结合活性。在一个实施方案中，RuvC和HNH活性部位内的两个点突变是化脓性链球菌Cas9的D10A和H840A突变。或者，可以删除化脓性链球菌Cas9的D10和H840，以消除Cas9核酸酶活性，同时保留其sgRNA和DNA结合活性。在一个实施方案中，RuvC和HNH活性部位内的两个点突变是化脓性链球菌Cas9的D10A和N580A突变。

[0132] 在各种实施方案中，本申请涉及dCas蛋白或其任何变体或突变体。dCas9的所有变体和突变体可用于本文公开的方法、组合物、融合分子或试剂盒，所述变体和突变体包括但不限于衍生自SpCas9(从化脓链球菌分离的Cas9)、SaCas9(从金黄色葡萄球菌分离的Cas9)、StCas9(从嗜热链球菌分离的Cas9)、NmCas9(从脑膜炎奈瑟菌(*Neisseria meningitidis*)分离的Cas9)、FnCas9(从诺维弗朗西斯菌(*Francisella novicida*)分离的Cas9)、CjCas9(从空肠弯曲杆菌分离的Cas9)、ScCas9(从犬链球菌(*Streptococcus canis*)分离的Cas9)以及上面列出的Cas9的任何变体和突变体形式，诸如高保真Cas9(Kleinstiver等人，Nature.2016年1月28日)和增强的SpCas9(Slaymaker等人，Sciences.2016年1月1日)的那些变体和突变体。例如，本

申请SEQ ID NOs:1170–1187所示的dCas9序列仅提供几个示例性选项，并不是排他性的。在一个实施方案中，dCas9蛋白是在D10和/或H840(如SEQ ID NO:1170所示的)处包含突变的化脓性链球菌dCas9蛋白。在一个实施方案中，dCas9蛋白是包含D10A和/或H840A突变(如SEQ ID NO:1170所示的)的化脓性链球菌dCas9蛋白。在一个实施方案中，dCas9蛋白是金黄色葡萄球菌dCas9蛋白，其包含SEQ ID NO:1171–1173中任一项所示的氨基酸序列、与SEQ ID NO:1171–1173中任一项基本相同的(例如，至少80%,85%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%或更高的序列同一性)序列、或相对于SEQ ID NO:1171–1173中任一项具有1个、2个、3个、4个、5个或更多个改变(例如，氨基酸取代、插入或缺失)的序列或其任意片段。

[0133] 类似的突变也可以应用于任何其它天然存在的Cas9(例如，来自其它物种的Cas9)或工程化的Cas9。在某些实施方案中，dCas9包括化脓性链球菌(*Streptococcus pyogenes*)dCas9、金黄色葡萄球菌(*Staphylococcus aureus*)dCas9、空肠弯曲菌(*Campylobacter jejuni*)dCas9、白喉棒状杆菌(*Corynebacterium diphtheria*)dCas9、凸腹真杆菌(*Eubacterium ventriosum*)dCas9、巴氏链球菌(*Streptococcus pasteurianus*)dCas9、香肠乳杆菌(*Lactobacillus farciminis*)dCas9、球形球毛菌(*Sphaerochaeta globus*)dCas9、固氮螺菌属(*Azospirillum*，例如，菌株B510)dCas9、嗜重氮葡萄糖醋杆菌(*Gluconacetobacter diazotrophicus*)dCas9、灰色奈瑟球菌(*Neisseria cinerea*)dCas9、肠道罗斯拜瑞氏菌(*Roseburia intestinalis*)dCas9、食清洁剂细小棒菌(*Parvibaculum lavamentivorans*)dCas9、卤水硝酸盐裂解菌(*Nitratifactor salsuginis*，例如，菌株DSM 16511)dCas9、海鸥弯曲菌(*Campylobacter lari*，例如，菌株CF89–12)dCas9、嗜热链球菌(*Streptococcus thermophilus*，例如，菌株LMD–9)dCas9或以上所述的片段。在某些实施方案中，本申请还提供了包含编码以下蛋白分子的核苷酸的载体：化脓性链球菌dCas9、金黄色葡萄球菌dCas9、空肠弯曲菌dCas9、白喉棒状杆菌dCas9、凸腹真杆菌dCas9、巴氏链球菌dCas9、香肠乳杆菌dCas9、球形球毛菌dCas9、固氮螺菌属(菌株B510)dCas9、嗜重氮葡萄糖醋杆菌dCas9、灰色奈瑟球菌dCas9、肠道罗斯拜瑞氏菌dCas9、食清洁剂细小棒菌dCas9、卤水硝酸盐裂解菌(菌株DSM 16511)dCas9、

海鸥弯曲菌(菌株CF89-12)dCas9、嗜热链球菌(菌株LMD-9)dCas9或以上所述的片段。

- [0134] 在本申请中，术语“核苷酸的修饰”可意指本发明中所述的核酸通过本领域成熟的方法，例如“Current protocols innucleic acid chemistry” Beaucage, S.L.等人, (Edrs.), John Wiley&Sons, Inc., NewYork, NY, USA中所描述的方法(其在此通过引用并入本文)进行合成或修饰。该修饰可包括但不限于：末端修饰，如5’-末端修饰(例如，磷酸化，缀合，反式键(inverted linkage))或3’-末端修饰(例如，缀合，DNA核苷酸，反式键等)；碱基修饰，如用稳定化的碱基、去稳定化的碱基或者与扩展的配对组库碱基配对的碱基进行置换，除去碱基(无碱基核苷酸)，或者缀合的碱基；糖修饰(例如，在2’-位或4’-位进行糖修饰)或置换糖；或者骨架修饰，包括对磷酸二酯键进行修饰或置换。
- [0135] 在本申请中，术语“DNA甲基化”与“核酸甲基化”可互换地使用，其通常是指本申请中基因片段、核苷酸或其碱基具有的甲基化状态，常出现在已经用核酸转染的细胞内部的过程，所述细胞已转染了包含与启动子有效连接的编码多肽的结构基因的核酸，在该过程中启动子核酸的胞嘧啶转化成5-甲基胞嘧啶。其中至少一个胞嘧啶转化成5-甲基胞嘧啶的启动子核酸称作“甲基化”核酸或DNA。本申请中基因所在的DNA片段可以在一条链或多条链上具有甲基化，还可以在一个位点或多个位点上具有甲基化。
- [0136] 在本申请中，术语“其部分”通常是指指定整体的部分或片段。例如，当在本申请中、相对于指定的多肽序列使用时，术语“其部分”是指短于指定的多肽的全长序列的一段连续长度的指定多肽的序列。指定多肽的一部分可以通过它的第一位置和它的最后位置得到限定，其中所述第一和最后位置各自对应于指定多肽的序列中的位置，其中对应于第一位置的序列位置位于对应于最后位置的序列位置的N端，并且由此该部分的序列是指定多肽中的连续的氨基酸序列，其在对应于第一位置的序列位置开始并且在对应于最后位置的序列位置结束。也可以通过参照指定多肽序列中的位置和相对于参照位置的残基长度来限定一部分，由此该部分的序列是指定多肽中的连续的氨基酸序列，其具有限定的长度并且根据限定的位置定位在指定多肽中。

- [0137] 在本申请中，术语“直接或间接融合”通常是指相对的“直接融合”或“间接融合”。术语“直接融合”通常是指直接相连或直接结合。例如，所述直接相连可以为相连的物质(例如氨基酸序列区段)之间没有间隔成分(例如氨基酸残基或其衍生物)而直接相连接的情况；例如氨基酸序列区段X与另一氨基酸序列区段Y通过氨基酸序列区段X的C端氨基酸与氨基酸序列区段Y的N端氨基酸形成的酰胺键直接连接。“间接融合”通常是指相连的物质(例如氨基酸序列区段)之间有间隔成分(例如氨基酸残基或其衍生物)而间接相连接的情况。
- [0138] 在本申请中，用来包装本申请所述的组合物、融合分子和/或引导分子(sgRNA)的载体可包含脂质颗粒，例如，所述脂质颗粒可以是脂质纳米颗粒(LNP)和脂质体。
- [0139] 例如，如本文中所用，术语“脂质纳米颗粒(LNP)”或“一个LNP”或“多个LNP”通常是指包含通过分子间力彼此物理结合(例如，共价或非共价)的多个(即多于一个)脂质分子的颗粒。LNP可以是例如微球(包括单层和多层囊泡，例如脂质体)、乳液中的分散相、胶团或悬浮液中的内相。LNP可将核酸包封在阳离子脂质颗粒(例如，脂质体)内，并且可被相对容易地递送至细胞。在一些实例中，脂质纳米颗粒不含任何病毒组分，这有助于最小化安全性和免疫原性问题。所述脂质颗粒可用于体外、离体和体内递送。所述脂质颗粒还可用于各种规模的细胞群。本申请的LNP可通过本领域已知的各种方法，例如通过混合有机相与水相来容易地制备。两相的混合可通过微流体装置和撞击流反应器来实现。有机相和水相混合越充分，获得的LNP的包埋率和粒径分布就越好。优选地，LNP的粒径可通过改变有机相与水相的混合速度来调节。混合速度越快，制备的LNP的粒径将越小。包埋效率可通过调节LNP系统的N/P(可电离脂质/核酸)比值来优化。在一些优选实施方案中，N/P比为1:1–9:1。在一些实例中，LNP可用于递送DNA分子(例如，包含DNA结合蛋白和/或sgRNA的编码序列的分子)和/或RNA分子(例如，Cas、sgRNA的mRNA)。在某些情况下，LNP可用于递送Cas/gRNA的RNP复合物。在一些实施方案中，LNP用于递送mRNA和gRNA(例如，包含DNMT3A–DNMT3L(3A–3L)–dCas9–KRAB或DNMT3A–DNMT3L–ZIM3 KRAB–dCas9，和至少一种靶向HBV基因的sgRNA的mRNA融合分子)。

[0140] LNP的组分可包含阳离子脂质1,2-苯二甲酸-3-二甲基铵-丙烷(DLinDAP)、1,2-二亚油基氧基-3-N,N-二甲基氨基丙烷(DLinDMA)、1,2-二亚油基氧基酮-N,N-二甲基-3-氨基丙烷(DLinK-DMA)、1,2-二亚油基-4-(2-二甲基氨基乙基)-。在一些实施方案中，LNP可包含可电离脂质。在一些实施方案中，可电离脂质包括但不限于pH响应性可电离脂质、热响应性可电离脂质和光响应性可电离脂质。在一些实施方案中，可电离脂质包括在某些条件(诸如但不限于pH、温度或光)下电离的阳离子脂质和阴离子脂质。在一些实施方案中，LNP的可电离脂质的摩尔比为20%至约70%(例如，约20%至约70%、约20%至约65%、约20%至约60%、约20%至约55%、约20%至约50%、约20%至约45%、约20%至约40%、约20%至约35%、约20%至约30%、约20%至约25%、约30%至约70%、约30%至约65%、约30%至约60%、约30%至约55%、约30%至约50%、约30%至约45%、约30%至约40%、约30%至约35%、约40%至约70%、约40%至约65%、约40%至约60%、约40%至约55%、约40%至约50%、约40%至约45%、约50%至约70%、约50%至约65%、约50%至约60%、约50%至约55%、约60%至约70%或约60%至约65%)。在一些实施方案中，LNP可包含聚乙二醇化脂质。在一些实施方案中，LNP的聚乙二醇化脂质的摩尔比为0%至约30%(例如，约0%至约30%、约0%至约25%、约0%至约20%、约0%至约15%、约0%至约10%、约10%至约30%、约10%至约25%、约10%至约20%、约10%至约15%、约20%至约30%或约20%至约25%)。在一些实施方案中，LNP可包含支撑脂质。在一些实施方案中，LNP的支撑脂质的摩尔比为30%至约50%(例如，约30%至约50%、约30%至约45%、约30%至约40%、约30%至约35%、约40%至约50%或约40%至约45%)。在一些实施方案中，LNP可包含胆固醇。在一些实施方案中，LNP的胆固醇的摩尔比为10%至约50%(例如，约10%至约50%、约10%至约45%、约10%至约40%、约10%至约35%、约10%至约30%、约10%至约25%、约10%至约20%、约10%至约15%、约20%至约50%、约20%至约45%、约20%至约40%、约20%至约35%、约20%至约30%、约20%至约25%、约30%至约50%、约30%至约45%、约30%至约40%、约30%至约35%、约40%至约50%或约40%至约45%)。在一些实施方案

中，LNP可包含可电离脂质(20%–70%，摩尔比)、聚乙二醇化脂质(0%–30%，摩尔比)、支撑脂质(30%–50%，摩尔比)和胆固醇(10%–50%，摩尔比)的混合物。

[0141] 例如，如本文中所用，术语“脂质体”通常是指通过一个或多个双层的膜与外部介质隔离的具有内部空间的囊泡。在一些实施方案中，所述双层的膜可以通过两性分子形成，如包含空间隔离的亲水性和疏水性结构域的合成或天然来源的脂质；在另一些实施方案中，所述双层的膜可以通过两亲性聚合物和表面活性剂形成。在一些实施方案中，所述脂质体是球形囊泡结构，其由围绕内部水性区室的单层或多层脂质双分子层、和相对不可渗透的外部亲脂性磷脂双分子层组成。在一些实施方案中，脂质体是生物相容的、无毒的，可以递送亲水性和亲脂性药物分子，保护它们的运载物不被血浆酶降解，并且将它们的负载运输穿过生物膜和血脑屏障(BBB)。脂质体可由几种不同类型的脂质例如磷脂制成。脂质体可包含天然磷脂和脂质(诸如1,2-二硬脂酰基-sn-甘油-3-磷脂酰胆碱(DSPC)、鞘磷脂、卵磷脂酰胆碱、单唾液酸神经节苷脂或其任意组合。为了改变脂质体的结构和性质，可向脂质体中加入几种其它添加剂。例如，脂质体还可包含胆固醇、鞘磷脂和/或1,2-二油酰基-sn-甘油-3-磷酸乙醇胺(DOPE)，例如，以增加稳定性和/或防止脂质体内部运载物的泄漏。

[0142] 在本申请中，术语“腺病毒相关病毒(AAV)载体”通常是指具有功能性或部分功能性ITR序列和转基因的载体。如本文中所用，术语“ITR”是指反向末端重复序列。ITR序列可源自腺相关病毒血清型，包括但不限于AAV-1、AAV-2、AAV-3、AAV-4、AAV-5、AAV-6、AAV-7、AAV-8、AAV-9、AAV-10、AAV-11、AAV-12和AAV-13，以及任何AAV变体或混合物。然而，ITR不必是野生型核苷酸序列，并且可被改变(例如，通过核苷酸的插入、缺失或取代)，只要序列保留提供功能性拯救、复制和包装的功能。AAV载体可具有一个或多个全部或部分删除的AAV野生型基因，优选为rep和/或cap基因，但保留功能性侧翼ITR序列。功能性ITR序列起到例如拯救、复制和包装AAV病毒粒子或颗粒的作用。因此，“AAV载体”在本申请中被定义为至少包括将转

基因插入受试者细胞所需的那些序列。任选包括病毒复制和包装(例如, 功能性 ITR)所必需的那些顺式序列。

[0143] 在本申请中, 术语“药学上可接受的载体”通常是指给予治疗剂, 例如抗体或多肽、基因和其它治疗剂的载体。该术语指本身不诱导对接受组合物的个体有害的抗体产生并且可以给予而不产生过度毒性的任何药物载体。合适的载体可以是大的、代谢缓慢的大分子, 例如蛋白质、多糖、聚乳酸、聚乙醇酸、多聚氨基酸、氨基酸共聚物、脂质聚集物和灭活的病毒颗粒。本领域技术人员熟知这些载体。治疗组合物中药学上可接受的载体可包括液体, 例如水、盐水、甘油和乙醇。这些载体中也可存在辅助物质, 例如润湿剂或乳化剂、pH缓冲物质等。

[0144] 在本申请中, 术语“编码……的序列”或“编码……的核酸”通常是指包含编码蛋白质的核苷酸序列的核酸(RNA或DNA分子)。编码序列还可包括与调控元件可操作地连接的起始和终止信号, 所述调控元件包含能够在对其施用了核酸的个体或哺乳动物的细胞中指导表达的启动子和多腺苷酸化信号。可对编码序列进行密码子优化。在一些实施方案中, 所述编码核酸可以为mRNA; 可使用一种或多种修饰技术用于产生更稳定的mRNA。已知的mRNA修饰技术大致可分为三类: 用人工合成的非天然核糖核酸代替天然核糖核酸合成mRNA; 添加5' caps、3' poly(A) “尾” 和UTR(未翻译区)序列; 采用特殊的新型配方技术, 有效保护mRNA。其中, 优选的mRNA修饰技术可以通过人工合成非天然核糖核酸取代天然核糖核酸合成mRNA。真核mRNA上的化学修饰大致可以分为三类: 甲基化、伪尿苷(Ψ)和次黄嘌呤。例如, 所述化学修饰可选自: 假尿苷、N1-甲基假尿苷、N1-乙基假尿苷、2-硫代尿苷、4'-硫代尿苷、5-甲基胞嘧啶、2-硫代-1-甲基-1-脱氮-假尿苷、2-硫代-1-甲基-假尿苷、2-硫代-5-氮杂-尿苷、2-硫代-二氢假尿苷、2-硫代-二氢尿苷、2-硫代-假尿苷、4-甲氧基-2-硫代-假尿苷、4-甲氧基-假尿苷、4-硫代-1-甲基-假尿苷、4-硫代-假尿苷、5-氮杂-尿苷、二氢假尿苷、5-甲基尿苷、5-甲氧基尿苷和2'-O-甲基尿苷。

- [0145] 在本申请中，术语“互补”通常是指核酸分子的核苷酸或核苷酸类似物之间的沃森—克里克(例如，A-T/U和C-G)或Hoogsteen碱基配对。“互补性”是指两个核酸序列之间共有的性质，使得当它们彼此反向平行排列时，每个位置处的核苷酸碱基将是互补的。
- [0146] 在本申请中，术语“受试者”和“患者”可互换地使用，其通常是指人和非人动物。本申请所用的术语“非人动物”包括所有脊椎动物，例如哺乳动物和非哺乳动物，诸如非人灵长类动物、羊、狗、猫、马、牛、鸡、两栖动物、爬行动物等。
- [0147] 在本申请中，术语“有效量”和“治疗有效量”或“治疗有效剂量”可互换地使用，其通常是指能够产生足够量的所需蛋白质以所需方式调节蛋白质活性，从而为临床干预提供了缓解工具的融合分子(蛋白)、多肽、核酸、脂质纳米颗粒、脂质体、一种或多种AAV颗粒或一种或多种病毒体的量或剂量。在某些实施方式中，治疗有效量或剂量的如本文所述的转染的融合蛋白、多肽、核酸、一种或多种AAV颗粒或一种或多种病毒体足以赋予对融合蛋白/基因治疗构建体靶向的基因的抑制。
- [0148] 如本文中所用，术语“治疗”，例如疾病，意指当施用例如本文所述的融合分子或编码该融合分子的核酸和/或gRNA或编码该gRNA的核酸时，与从未施用该融合分子或编码该融合分子的核酸和/或该gRNA或编码该gRNA的核酸时相比，患有疾病、有患疾病的风险和/或经历疾病症状的受试者(例如，人)在一个实施方案中将经历症状更轻和/或将更快恢复。
- [0149] 不欲被任何理论所限，下文中的实施例仅仅是为了阐释本申请的组合物、使用方法和用途等，而不适用于限制本申请发明的范围。
- [0150] 实施例
- [0151] 实施例1
- [0152] 靶向乙型肝炎病毒(HBV)基因的sgRNA文库构建及其功能筛选
- [0153] (1)构建B型、C型和D型HBV基因的sgRNA文库
- [0154] 本实施例使用的是乙型肝炎病毒(HBV)的基因组序列，其包括编码HBV聚合酶Pol、表面抗原HBsAg、HBV X蛋白和核心区抗原前体HBcAg的多核苷酸的全基

基因组，这些基因组分为B型HBV(血清型adw亚型)、C型HBV(血清型adr亚型)、D型HBV(血清型ayw亚型)，如下表所示。

向导 RNA（与靶序列互补的核苷酸序列 + 3'端 PAM 序列）	基因型
AGACCACCGTGAACGCCCACAGG	B 型
G TTCCTGTGGGCGTTCACGGTGG	B 型
CAGGTT CCTGTGGGCGTTCACGG	B 型
GGTCTCCATGCGACGTGCAGAGG	B 型
CGCCCACAGGAACCTGCCCCAAGG	B 型
GACCTTGGGCAGGTTCTGTGGG	B 型
CTTCACCTCTGCACGTCGCATGG	B 型
AGACCTTGGGCAGGTTCTGTGG	B 型
TCTTATGCAAGACCTTGGGCAGG	B 型
TGCCCCAAGGTCTTG CATAAGAGG	B 型
GTCTCTTATGCAAGACCTTGGG	B 型
AGTCCTCTTATGCAAGACCTTGG	B 型
GAGGTGAAGCGAAGTGCACACGG	B 型
GTCTTG CATAAGAGGACTCTTGG	B 型
GAAGCGAAGTG CACACGGTCCGG	B 型
ACACGGTCCGGCAGATGAGAAGG	B 型
GTCTGTGCCTTCTCATCTGCCGG	B 型
GGCAGATGAGAAGGCACAGACGG	B 型
GCAGATGAGAAGGCACAGACGGG	B 型
CAGATGAGAAGGCACAGACGGGG	B 型
AATGTCAACGACCGACCTTGAGG	B 型
CGGGGAGTCCGCGTAAAGAGAGG	B 型
TTTGAAGTATGCCTCAAGGTCGG	B 型
AGTCTTTGAAGTATGCCTCAAGG	B 型
GGGGCGCACCTCTCTTTACGCGG	B 型
GTAAAGAGAGGTGCGCCCCGTGG	B 型
AGAGAGGTGCGCCCCGTGGTCGG	B 型
AGGTGCGCCCCGTGGTCGGCCGG	B 型
AAGACTGTGTGTTTACTGAGTGG	B 型

AGACTGTGTGTTTACTGAGTGGG	B 型
TATTGTACCGGCCGACCACGGGG	B 型
CTGTGTGTTTACTGAGTGGGAGG	B 型
CTATTGTACCGGCCGACCACGGG	B 型
CCTATTGTACCGGCCGACCACGG	B 型
CCGTGGTCGGCCGGTACAATAGG	B 型
TGGTCGGCCGGTACAATAGGCGG	B 型
GTTTACTGAGTGGGAGGAGTTGG	B 型
TTTACTGAGTGGGAGGAGTTGGG	B 型
TTACTGAGTGGGAGGAGTTGGGG	B 型
TACTGAGTGGGAGGAGTTGGGGG	B 型
CGCTTCTCCGCCTATTGTACCGG	B 型
TGAGTGGGAGGAGTTGGGGGAGG	B 型
CGGTACAATAGGCGGAGAAGCGG	B 型
GGTACAATAGGCGGAGAAGCGGG	B 型
GTGGGAGGAGTTGGGGGAGGAGG	B 型
ACAATAGGCGGAGAAGCGGGCGG	B 型
AGGAGTTGGGGGAGGAGGTTAGG	B 型
GGGGGAGGAGGTTAGGTTAAAGG	B 型
GCGGGCGGTAGAGTCCCAAGCGG	B 型
AGGTTAAAGGTCTTTGTACTAGG	B 型
GAGTCCCAAGCGGCCCCGAGAGG	B 型
TTAAAGGTCTTTGTACTAGGAGG	B 型
AGTCCCAAGCGGCCCCGAGAGGG	B 型
GTCCCAAGCGGCCCCGAGAGGGG	B 型
GACCCCTCTCGGGGCCGCTTGGG	B 型
CGACCCCTCTCGGGGCCGCTTGG	B 型
TCTTTGTACTAGGAGGCTGTAGG	B 型
GCCCCGAGAGGGGTCGTCCGCGG	B 型
CCCGCGGACGACCCCTCTCGGGG	B 型
CCCCGAGAGGGGTCGTCCGCGGG	B 型
TCCCGCGGACGACCCCTCTCGGG	B 型
ATCCCGCGGACGACCCCTCTCGG	B 型
AGGAGGCTGTAGGCATAAATTGG	B 型
TCCGCGGGATTGAGCGCCGACGG	B 型
CCGCGGGATTGAGCGCCGACGGG	B 型
CCCGTCGGCGCTGAATCCCGCGG	B 型
CGCCGACGGGACGTAAACAAAGG	B 型
GTCCTTTGTTTACGTCCCGTCGG	B 型
GTGAAAAAGTTGCATGGTGCTGG	B 型
GCAGAGGTGAAAAAGTTGCATGG	B 型
AAACAAAGGACGTCCCGCGCAGG	B 型
CGTCCCGCGCAGGATCCAGTTGG	B 型
CTGCCAACTGGATCCTGCGCGGG	B 型

GCTGCCAACTGGATCCTGCGCGG	B 型
AACATGAGATGATTAGGCAGAGG	B 型
GACATGAACATGAGATGATTAGG	B 型
TGCTAGGCTGTGCTGCCAACTGG	B 型
GCAGCACAGCCTAGCAGCCATGG	B 型
ACATCATTTCCATGGCTGCTAGG	B 型
AGCTTGGAGGCTTGAACAGTAGG	B 型
GCAAGTATACATCATTTCCATGG	B 型
ATGGAAATGATGTATACTTGCGG	B 型
CAAGCCTCCAAGCTGTGCCTTGG	B 型
TGGAAATGATGTATACTTGCGGG	B 型
AAGCCTCCAAGCTGTGCCTTGGG	B 型
CCACCCAAGGCACAGCTTGGAGG	B 型
CCTCCAAGCTGTGCCTTGGGTGG	B 型
AAGCCACCCAAGGCACAGCTTGG	B 型
AGCTGTGCCTTGGGTGGCTTTGG	B 型
GCTGTGCCTTGGGTGGCTTTGGG	B 型
CTGTGCCTTGGGTGGCTTTGGGG	B 型
CCTTGGGTGGCTTTGGGGCATGG	B 型
CCATGCCCCAAAGCCACCCAAGG	B 型
CAGAATTGTCAGTCCCGATGAGG	B 型
ATTGACCCGTATAAAGAATTTGG	B 型
GGTCTGGGGCAAACCTCATCGGG	B 型
AGGTCTGGGGCAAACCTCATCGG	B 型
AAGCTCCAAATTCTTTATACGGG	B 型
GAAGCTCCAAATTCTTTATACGG	B 型
TAAAGAATTTGGAGCTTCTGTGG	B 型
TTTTGCTCGCAGCAGGTCTGGGG	B 型
GTTTTGCTCGCAGCAGGTCTGGG	B 型
TGTTTTGCTCGCAGCAGGTCTGG	B 型
CCTGCTGCGAGCAAAACAAGCGG	B 型
CCGCTTGTTTTGCTCGCAGCAGG	B 型
TGCGAGCAAAACAAGCGGCTAGG	B 型
CGGCTAGGAGTTCCGCAGTATGG	B 型
AGGAGTTCCGCAGTATGGATCGG	B 型
ATAGAAGGAAAGAAGTCAGAAGG	B 型
TCCGCAGTATGGATCGGTAGAGG	B 型
TCCTCTACCGATCCATACTGCGG	B 型
GATCGGTAGAGGAGACACAAAGG	B 型
TCGAGGAGATCTCGAATAGAAGG	B 型
ACAGAGCAGAGGCGGTGTCGAGG	B 型
AGGTTCACGCATGCGCTGATGG	B 型
ACACCGCCTCTGCTCTGTATCGG	B 型
CACCGCCTCTGCTCTGTATCGGG	B 型

CTCCCGATACAGAGCAGAGGCGG	B 型
ATAGGCCATCAGCGCATGCGTGG	B 型
CGCCTCTGCTCTGTATCGGGAGG	B 型
ACGCATGCGCTGATGGCCTATGG	B 型
GGCCTCCCGATACAGAGCAGAGG	B 型
TCGGGAGGCCTTAGAGTCTCCGG	B 型
ACTGGTTGGGGCTTGGCCATAGG	B 型
TATGGCCAAGCCCCAACCAGTGG	B 型
ATGGCCAAGCCCCAACCAGTGGG	B 型
TGGCCAAGCCCCAACCAGTGGGG	B 型
GGCCAAGCCCCAACCAGTGGGGG	B 型
ACAATGTTCCGGAGACTCTAAGG	B 型
AACCCCCACTGGTTGGGGCTTGG	B 型
GACGCAACCCCCACTGGTTGGGG	B 型
TGACGCAACCCCCACTGGTTGGG	B 型
CTGACGCAACCCCCACTGGTTGG	B 型
TGGTGAGGTGAACAATGTTCCGG	B 型
TTTGCTGACGCAACCCCCACTGG	B 型
ACATTGTTACCTCACCATACGG	B 型
GGGTTGCGTCAGCAAACACTTGG	B 型
CACCTCACCATACGGCACTCAGG	B 型
TGCCCTGAGTGCCGTATGGTGAGG	B 型
GCAAACACTTGGCACAGACCAGG	B 型
TAGCTTGCCCTGAGTGCCGTATGG	B 型
TCAGGCAAGCTATCCTGTGTTGG	B 型
CAGGCAAGCTATCCTGTGTTGGG	B 型
AGGCAAGCTATCCTGTGTTGGGG	B 型
ACCAGGCCGTTGCCGAGCAACGG	B 型
CCCGTTGCTCGGCAACGGCCTGG	B 型
CCAGGCCGTTGCCGAGCAACGGG	B 型
CAGGCCGTTGCCGAGCAACGGGG	B 型
TTTACCCCGTTGCTCGGCAACGG	B 型
TCATCAACTCACCCCAACACAGG	B 型
GTTGCCGAGCAACGGGGTAAAGG	B 型
TGAACCTTTACCCCGTTGCTCGG	B 型
AGTTGATGAATCTAGCTACCTGG	B 型
GTTGATGAATCTAGCTACCTGGG	B 型
GATGAATCTAGCTACCTGGGTGG	B 型
ATGAATCTAGCTACCTGGGTGGG	B 型
CAGATACTGTTTACTTAGAAAGG	B 型
TACCTGGGTGGGAAGTAATTTGG	B 型
TTCCAAATTACTTCCCACCCAGG	B 型
CTTAGAAAGGCCTTGTAAGTTGG	B 型
ATTTGGAAGATCCAGCATCCAGG	B 型

TTTGGAAGATCCAGCATCCAGGG	B 型
TACTTTCTCGCCAACTTACAAGG	B 型
ACTACTAATTCCCTGGATGCTGG	B 型
ATAGCTGACTACTAATTCCTGG	B 型
TGCATGTATACAAGCGAAACAGG	B 型
CAGCTATGTCAACGTTAACATGG	B 型
AGCTATGTCAACGTTAACATGGG	B 型
GCTTGTATACATGCATATAAAGG	B 型
GCATATAAAGGCATTAAAGCAGG	B 型
ACAAGAGTTGTCTGATTTTTAGG	B 型
TAAAAATCAGACAACTCTTGTGG	B 型
GGATATCCACATTGCGTGAAAGG	B 型
GATATCCACATTGCGTGAAAGGG	B 型
ATATCCACATTGCGTGAAAGGGG	B 型
TCCACATTGCGTGAAAGGGGCGG	B 型
GCCGCCCTTTCACGCAATGTGG	B 型
CACATTTCTGTCTTACTTTTGG	B 型
ACATTTCCTGTCTTACTTTTGGG	B 型
TTCTCTCCCAAAGTAAGACAGG	B 型
AACGAATTGTGGGTCTTTTGGGG	B 型
CAACGAATTGTGGGTCTTTTGGG	B 型
TCAACGAATTGTGGGTCTTTTGG	B 型
AAAGTATGTCAACGAATTGTGGG	B 型
AGAAACTGTCCTTGAATATTTGG	B 型
GAAAGTATGTCAACGAATTGTGG	B 型
CAAAAGACACCAAATATTCAAGG	B 型
CTTGAATATTTGGTGTCTTTTGG	B 型
TGACATACTTCCAATCAATAGG	B 型
ATTTGGTGTCTTTTGGAGTGTGG	B 型
CTGTAAACAGGCCTATTGATTGG	B 型
CAATCAATAGGCCTGTTTACAGG	B 型
TTAGGAACTTCCTGTAAACAGG	B 型
TTTGGTGGTCTATATGCAGGAGG	B 型
AAAAAATCAAAATGTGTTTTAGG	B 型
GCATTTGGTGGTCTATATGCAGG	B 型
GATAAGATAGGGGCATTGGTGG	B 型
GTTGATAAGATAGGGGCATTGG	B 型
TTTTGTACAATATGTTCCCTGTGG	B 型
CGGAAGAGTTGATAAGATAGGGG	B 型
CCCTATCTTATCAACTCTCCGG	B 型
CCGGAAGAGTTGATAAGATAGGG	B 型
TCCGGAAGAGTTGATAAGATAGG	B 型
GAGTTGGGGCACATTGCCACAGG	B 型
CGTCTAACAACAGTAGTTTCCGG	B 型

ACTACTGTTGTTAGACGACGAGG	B 型
GGATATGTAATTGGGAGTTGGGG	B 型
GGGATATGTAATTGGGAGTTGGG	B 型
TGGGATATGTAATTGGGAGTTGG	B 型
CTGTTGTTAGACGACGAGGCAGG	B 型
ACTTCATGGGATATGTAATTGGG	B 型
AACTTCATGGGATATGTAATTGG	B 型
TTACATATCCCATGAAGTTAAGG	B 型
TACATATCCCATGAAGTTAAGGG	B 型
GGATATTCCTTAACTTCATGGG	B 型
GGGATATTCCTTAACTTCATGG	B 型
CGAGGGAGTTCTTCTTCTAGGGG	B 型
GCGAGGGAGTTCTTCTTCTAGGG	B 型
GGCGAGGGAGTTCTTCTTCTAGG	B 型
CCCCATCTTTTTGTTTTGTGAGG	B 型
CTCCCTCGCCTCGCAGACGAAGG	B 型
CCTCACAAAACAAAAAGATGGGG	B 型
CCCTCACAAAACAAAAAGATGGG	B 型
CCCATCTTTTTGTTTTGTGAGGG	B 型
GACCTTCGTCTGCGAGGCGAGGG	B 型
ACCCTCACAAAACAAAAAGATGG	B 型
AGACCTTCGTCTGCGAGGCGAGG	B 型
GATTGAGACCTTCGTCTGCGAGG	B 型
GATTGATATCTTCTGCGACGCGG	B 型
ACCCAAAGACAAAAGAAAATTGG	B 型
GTCGCAGAAGATATCAATCTCGG	B 型
ACCAATTTTCTTTTGTCTTTGGG	B 型
TCGCAGAAGATATCAATCTCGGG	B 型
TACCAATTTTCTTTTGTCTTTGG	B 型
CAAAAGAAAATTGGTAACAGCGG	B 型
AATTGGTAACAGCGGCATAAAGG	B 型
ATTGGTAACAGCGGCATAAAGGG	B 型
ATCTCAATGTTAGTATTCCTTGG	B 型
TAGTATTCCTTGGACACATAAGG	B 型
TATTCCTTGGACACATAAGGTGG	B 型
ATTCCTTGGACACATAAGGTGGG	B 型
CTCAAGATGTTGTACAGACTTGG	B 型
TTTCCCACCTTATGTGTCCAAGG	B 型
ACATAAGGTGGGAACTTTACGG	B 型
CATAAGGTGGGAACTTTACGGG	B 型
ATAAGGTGGGAACTTTACGGGG	B 型
TATGGATGATGTGGTTTTGGGGG	B 型
ATATGGATGATGTGGTTTTGGGG	B 型
TATATGGATGATGTGGTTTTGGG	B 型

TTATATGGATGATGTGGTTTTGG	B 型
TACGGGGCTTTATTCTTCTACGG	B 型
TTTCAGTTATATGGATGATGTGG	B 型
CTGTCTGGCTTTCAGTTATATGG	B 型
TATAACTGAAAGCCAGACAGTGG	B 型
ATAACTGAAAGCCAGACAGTGGG	B 型
TAAGTAAAGCCAGACAGTGGGG	B 型
AACTGAAAGCCAGACAGTGGGGG	B 型
TACCTTGCTTTAATCCTAAATGG	B 型
TGCCATTTAGGATTAAAGCAAGG	B 型
TAGGGCTTTCCCCCACTGTCTGG	B 型
AAAGAAGGAGTTTGCCATTTAGG	B 型
CATTTGTTTCAGTGGTTCGTAGGG	B 型
CCTACGAACCACTGAACAAATGG	B 型
CCATTTGTTTCAGTGGTTCGTAGG	B 型
AAATGAATGTCAGGAAAAGAAGG	B 型
TACTAGTGCCATTTGTTTCAGTGG	B 型
TTTTCTGACATTCATTTGCAGG	B 型
TCCTGACATTCATTTGCAGGAGG	B 型
TCCTCCTGCAAATGAATGTCAGG	B 型
GTAAACTGAGCCAAGAGAAACGG	B 型
GAGCCAAGAGAAACGGACTGAGG	B 型
GGGCCTCAGTCCGTTTCTCTTGG	B 型
TGATAGATGTAAGCAATTTGTGG	B 型
GATAGATGTAAGCAATTTGTGGG	B 型
ATAGATGTAAGCAATTTGTGGGG	B 型
GGACTGAGGCCCCACTCCCATAGG	B 型
CGCAAAATACCTATGGGAGTGGG	B 型
TCGCAAAATACCTATGGGAGTGG	B 型
GGCTTTCGCAAAATACCTATGGG	B 型
GGGCTTTCGCAAAATACCTATGG	B 型
CTGTTTTCATTTACTGTAAGGGG	B 型
CCCTTACAGTAAATGAAAACAGG	B 型
CCTGTTTTCATTTACTGTAAGGG	B 型
TCCTGTTTTCATTTACTGTAAGG	B 型
TTTGCGAAAGCCCAAGATGATGG	B 型
TTGCGAAAGCCCAAGATGATGGG	B 型
GAAAGCCCAAGATGATGGGATGG	B 型
AAAGCCCAAGATGATGGGATGGG	B 型
TATTCCTATCCCATCATCTTGGG	B 型
GTATTCCTATCCCATCATCTTGG	B 型
AGATGATGGGATGGGAATACAGG	B 型
TGAAATTAATTATGCCTGCTAGG	B 型
CAGGTGCAGTTTCCGTCCGTAGG	B 型

ACATTGGGATAAAAACCTAGCAGG	B 型
TGCTGTACAAAACCTACGGACGG	B 型
ATGTTGCTGTACAAAACCTACGG	B 型
GGCAAATATTTAGTAACATTGGG	B 型
GGGCAAATATTTAGTAACATTGG	B 型
AGGTTTTGTACAGCAACATGAGG	B 型
GGTTTTGTACAGCAACATGAGGG	B 型
AAATATTTGCCCTTAGATAAAGG	B 型
AATATTTGCCCTTAGATAAAGGG	B 型
GCAACATGAGGGAAACATAGAGG	B 型
GGTTTGATCCCTTTATCTAAGGG	B 型
CGGTTTGATCCCTTTATCTAAGG	B 型
AACATAGAGGTTTCCTTGAGCAGG	B 型
TCCTTGAGCAGGAGTTGTGCAGG	B 型
ACCTGCACAACCTCCTGCTCAAGG	B 型
ACTACATACTCTGGATAATACGG	B 型
GGAGTTGTGCAGGTTTTGCATGG	B 型
TAATGATTAACCTACATACTCTGG	B 型
TGTGCAGGTTTTGCATGGTCCGG	B 型
GGTTTTGCATGGTCCGGTGCTGG	B 型
GGATCATCAACAACCAGCACCGG	B 型
GTGCTGGTTGTTGATGATCCTGG	B 型
TGTGTAAATAATGTCGCGTCTGG	B 型
GACATTATTTACACACTCTTTGG	B 型
GTTGATGATCCTGGAATTAGAGG	B 型
TTATTTACACACTCTTTGGAAGG	B 型
TTTACACACTCTTTGGAAGGCGG	B 型
TTACACACTCTTTGGAAGGCGGG	B 型
TACACACTCTTTGGAAGGCGGGG	B 型
TCCTGGAATTAGAGGACAAACGG	B 型
CCCGTTTGTCTCTAATTCCAGG	B 型
CCTGGAATTAGAGGACAAACGGG	B 型
TTGGTTCTTCTGGACTATCAAGG	B 型
CATCTTCTTGTGGTTCTTCTGG	B 型
GCAAAATGAGGCGCTACGTGTGG	B 型
ACACGTAGCGCCTCATTTTGCGG	B 型
GAAGAACCAACAAGAAGATGAGG	B 型
CACGTAGCGCCTCATTTTGCGGG	B 型
GCTATGCCTCATCTTCTTGTGG	B 型
ATATGGTGACCCGCAAAATGAGG	B 型
AGAAGATGAGGCATAGCAGCAGG	B 型
TTTGCGGGTCACCATATTCTTGG	B 型
TTGCGGGTCACCATATTCTTGGG	B 型
GGCATAGCAGCAGGATGCAGAGG	B 型

AGATCTTGTCCCAAGAATATGG	B 型
TGGGAACAAGATCTACAGCATGG	B 型
GGGAACAAGATCTACAGCATGGG	B 型
AACAAGATCTACAGCATGGGAGG	B 型
AGATCTACAGCATGGGAGGTTGG	B 型
GTTATCGCTGGATGTGTCTGCGG	B 型
AGACACATCCAGCGATAACCAGG	B 型
TGGTCTTCCAAACCTCGAAAAGG	B 型
TTCCAAACCTCGAAAAGGCATGG	B 型
CAATTTGTCCTGGTTATCGCTGG	B 型
TCCAAACCTCGAAAAGGCATGGG	B 型
CAGCGATAACCAGGACAAATTGG	B 型
CCCCATGCCTTTTCGAGGTTTGG	B 型
CCAAACCTCGAAAAGGCATGGGG	B 型
CGATAACCAGGACAAATTGGAGG	B 型
TTTGTCCCCATGCCTTTTCGAGG	B 型
TGTTGTCTCCAATTTGTCCTGG	B 型
AGGACAAATTGGAGGACAACAGG	B 型
CAAATTGGAGGACAACAGGTTGG	B 型
ACAACAGGTTGGTGAGTGAAGTGG	B 型
TCTTTCTGTCCCAATCCCCCTGG	B 型
CTTTCTGTCCCAATCCCCCTGGG	B 型
TTGGTGAGTGACTGGAGATTTGG	B 型
TGGTGAGTGACTGGAGATTTGGG	B 型
GGAAGAATCCCAGGGGATTGGGG	B 型
GGGAAGAATCCCAGGGGATTGGG	B 型
GGGGAAGAATCCCAGGGGATTGG	B 型
TGATCGGGGAAGAATCCCAGGGG	B 型
ATGATCGGGGAAGAATCCCAGGG	B 型
GATGATCGGGGAAGAATCCCAGG	B 型
AGATTTGGGACTGCGAATTTTGG	B 型
ATTCTTCCCGATCATCAGTTGG	B 型
CAGGGTCCAACCTGATGATCGGGG	B 型
GCAGGGTCCAACCTGATGATCGGG	B 型
TGCAGGGTCCAACCTGATGATCGG	B 型
CGAATTTTGGCCAAGACACACGG	B 型
GAATTTTGGCCAAGACACACGGG	B 型
GGGGGAACACCCGTGTGTCTTGG	B 型
CTGAGTTGGCTTTGAATGCAGGG	B 型
TCTGAGTTGGCTTTGAATGCAGG	B 型
CCAATCTGGACTTTCTGAGTTGG	B 型
ACTTCTCTCAATTTTCTAGGGGG	B 型
CCAATCAGAAAGTCCAGATTGG	B 型
CAACTCAGAAAGTCCAGATTGGG	B 型

GACTTCTCTCAATTTTCTAGGGG	B 型
GGACTTCTCTCAATTTTCTAGGG	B 型
TGGACTTCTCTCAATTTTCTAGG	B 型
TGCGGGTTGAGGTCCCAATCTGG	B 型
TTGGGACCTCAACCCGCACAAGG	B 型
CACAGAGTCTAGACTCGTGGTGG	B 型
CACCACGAGTCTAGACTCTGTGG	B 型
AGTTGTCCTTGTGCGGGTTGAGG	B 型
TACCACAGAGTCTAGACTCGTGG	B 型
TCAACCCGCACAAGGACAAGTGG	B 型
CCGGCCAGTTGTCCTTGTGCGGG	B 型
CCCGCACAAGGACAAGTGGCCGG	B 型
TCCGGCCAGTTGTCCTTGTGCGG	B 型
CTAGACTCTGTGGTATTGTGAGG	B 型
CAACTGGCCGGACGCCAACAAGG	B 型
CTGGCCGGACGCCAACAAGGTGG	B 型
TGGCCGGACGCCAACAAGGTGGG	B 型
ACTCCCACCTTGTGGCGTCCGG	B 型
GGACGCCAACAAGGTGGGAGTGG	B 型
GACGCCAACAAGGTGGGAGTGGG	B 型
TGCTCCCACCTCCCACCTTGTGG	B 型
AAGGTGGGAGTGGGAGCATTCGG	B 型
AGGTGGGAGTGGGAGCATTCGGG	B 型
GGAGTGGGAGCATTCGGGCCAGG	B 型
GAGTGGGAGCATTCGGGCCAGGG	B 型
ACCCCGCCTGTAACACGAGCAGG	B 型
CCCCGCCTGTAACACGAGCAGGG	B 型
CCCTGCTCGTGTTACAGGCGGGG	B 型
CCCGCCTGTAACACGAGCAGGGG	B 型
CCCCTGCTCGTGTTACAGGCGGG	B 型
ACCCCTGCTCGTGTTACAGGCGG	B 型
AGGACCCCTGCTCGTGTTACAGG	B 型
GTAACACGAGCAGGGGTCCTAGG	B 型
CCATGGGGAGGGGTGAACCCTGG	B 型
CCAGGGTTCACCCCTCCCCATGG	B 型
CAGGGTTCACCCCTCCCCATGGG	B 型
AGGGTTCACCCCTCCCCATGGGG	B 型
GGGTTCACCCCTCCCCATGGGGG	B 型
CAACAGTCCCCCATGGGGAGGGG	B 型
CCCTCCCCATGGGGGACTGTTGG	B 型
CCAACAGTCCCCCATGGGGAGGG	B 型
CCCAACAGTCCCCCATGGGGAGG	B 型
CCTCCCCATGGGGGACTGTTGGG	B 型
CTCCCCATGGGGGACTGTTGGGG	B 型

CACCCCAACAGTCCCCCATGGGG	B 型
CCCATGGGGGACTGTTGGGGTGG	B 型
AACATCGCATCAGGACTCCTAGG	B 型
CCACCCCAACAGTCCCCCATGGG	B 型
TCCACCCCAACAGTCCCCCATGG	B 型
AACATGGAGAACATCGCATCAGG	B 型
ACTGTTGGGGTGGAGCCCTCAGG	B 型
GATGCGATGTTCTCCATGTTCTGG	B 型
GGGGTGGAGCCCTCAGGCTCAGG	B 型
GGGTGGAGCCCTCAGGCTCAGGG	B 型
ATGTTCTCCATGTTCTCGGTACAGG	B 型
TGTTCTCCATGTTCTCGGTACAGGG	B 型
TGAGTAGGCCCTGAGCCTGAGGG	B 型
TGGGGACCCTGTACCGAACATGG	B 型
GTGAGTAGGCCCTGAGCCTGAGG	B 型
CTGCTGGCACAGTTGTGAGTAGG	B 型
GTCAATCTTATCGAAGACTGGGG	B 型
CGTCAATCTTATCGAAGACTGGG	B 型
TCGTCAATCTTATCGAAGACTGG	B 型
CTTCGATAAGATTGACGATATGG	B 型
GAGGCAGGAGGAGGAGCTGCTGG	B 型
CGATTGGTGGAGGCAGGAGGAGG	B 型
CTCCTCCTGCCTCCACCAATCGG	B 型
TGCCGATTGGTGGAGGCAGGAGG	B 型
GACTGCCGATTGGTGGAGGCAGG	B 型
GCAGAGACAGTATTCTGAGCAGG	B 型
CAGAGACAGTATTCTGAGCAGGG	B 型
GCCTCCACCAATCGGCAGTCAGG	B 型
TCCTGACTGCCGATTGGTGGAGG	B 型
CCTTCCTGACTGCCGATTGGTGG	B 型
CCACCAATCGGCAGTCAGGAAGG	B 型
CTGCCTTCCTGACTGCCGATTGG	B 型
AGGGCTCACTGTTCTGAACTGG	B 型
AGAGGTGGAGATAAGGGAGTAGG	B 型
CCTGAACTGGAGCCACCAGCAGG	B 型
CCTGCTGGTGGCTCCAGTTCAGG	B 型
CTCCCTTATCTCCACCTCTAAGG	B 型
TCCCTTATCTCCACCTCTAAGGG	B 型
TCCCTTAGAGGTGGAGATAAGGG	B 型
GTCCCTTAGAGGTGGAGATAAGG	B 型
AGCCACCAGCAGGAAAGTACAGG	B 型
GCCACCAGCAGGAAAGTACAGGG	B 型
GCCCTGTACTTTCCTGCTGGTGG	B 型
GGATGAGTGTCCCTTAGAGGTGG	B 型

AGGGCCCTGTACTTTCCTGCTGG	B 型
TGAGGATGAGTGTCCCTTAGAGG	B 型
TCTAAGGGACACTCATCCTCAGG	B 型
AAAGTACAGGGCCCTGACTCTGG	B 型
AAGTACAGGGCCCTGACTCTGGG	B 型
CTCATCCTCAGGCCATGCAGTGG	B 型
TCTTCAAGATCCCAGAGTCAGGG	B 型
CTCTTCAAGATCCCAGAGTCAGG	B 型
CTCTGGGATCTTGAAGAGTTTGG	B 型
TGGGATCTTGAAGAGTTTGGTGG	B 型
TTGAAGAGTTTGGTGGAAAGTGG	B 型
AAGAGTTTGGTGGAAAGTGGTGG	B 型
GAGTTCCACTGCATGGCCTGAGG	B 型
AGTGGTGGAGTTCCACTGCATGG	B 型
CAGAGCTTGGTGGAAATGTTGTGG	C 型
TGGGATCTAGCAGAGCTTGGTGG	C 型
CTCTGGGATCTAGCAGAGCTTGG	C 型
CTCTGCTAGATCCCAGAGTGAGG	C 型
TCTGCTAGATCCCAGAGTGAGGG	C 型
CTGCTAGATCCCAGAGTGAGGGG	C 型
AAATATAGGCCCCCTCACTCTGGG	C 型
AAAATATAGGCCCCCTCACTCTGG	C 型
AGGGGCCTATATTTTCCTGCTGG	C 型
GGCCTATATTTTCCTGCTGGTGG	C 型
AGCCACCAGCAGGAAAATATAGG	C 型
CCTGCTGGTGGCTCCAGTTCCGG	C 型
CCGGAACCTGGAGCCACCAGCAGG	C 型
AGGGTTTACTGTTCCGGAACCTGG	C 型
CGGAACAGGGTTTACTGTTCCGG	C 型
GTGAGGCAGTAGTCGGAACAGGG	C 型
GGTGAGGCAGTAGTCGGAACAGG	C 型
GATATGGGTGAGGCAGTAGTCGG	C 型
GAAGATTGACGATATGGGTGAGG	C 型
CTCGAGAAGATTGACGATATGGG	C 型
CCTCGAGAAGATTGACGATATGG	C 型
CCATATCGTCAATCTTCTCGAGG	C 型
TCGTCAATCTTCTCGAGGACTGG	C 型
CGTCAATCTTCTCGAGGACTGGG	C 型
GTCAATCTTCTCGAGGACTGGGG	C 型
TGGGGACCCTGCACCGAACATGG	C 型
TGTTCTCCATGTTTCGGTGCAGGG	C 型
GTGTTCTCCATGTTTCGGTGCAGG	C 型
GATGTTGTGTTCTCCATGTTTCGG	C 型
AACATGGAGAACAACATCAGG	C 型

AACACAACATCAGGATTCCTAGG	C 型
GTAACACGAGCAGGGTCCTAGG	C 型
AGGACCCCTGCTCGTGTTACAGG	C 型
ACCCCTGCTCGTGTTACAGGCGG	C 型
CCCCTGCTCGTGTTACAGGCGGG	C 型
CCCGCCTGTAACACGAGCAGGGG	C 型
CCCTGCTCGTGTTACAGGCGGGG	C 型
CCCCGCCTGTAACACGAGCAGGG	C 型
ACCCCGCCTGTAACACGAGCAGG	C 型
CTAGACTCTGTGGTATTGTGAGG	C 型
TACCACAGAGTCTAGACTCGTGG	C 型
CACCACGAGTCTAGACTCTGTGG	C 型
CACAGAGTCTAGACTCGTGGTGG	C 型
TGGACTTCTCTCAATTTTCTAGG	C 型
GGACTTCTCTCAATTTTCTAGGG	C 型
GACTTCTCTCAATTTTCTAGGGG	C 型
ACTTCTCTCAATTTTCTAGGGGG	C 型
GGGGGAGCACCCACGTGTCCTGG	C 型
GAATTTTGCCAGGACACGTGGG	C 型
CGAATTTTGCCAGGACACGTGG	C 型
GGGGACTGCGAATTTTGCCAGG	C 型
AGGTTGGGGACTGCGAATTTTGG	C 型
TGGTGAGTGATTGGAGGTTGGGG	C 型
TTGGTGAGTGATTGGAGGTTGGG	C 型
GTTGGTGAGTGATTGGAGGTTGG	C 型
AGAGGTTGGTGAGTGATTGGAGG	C 型
ACAAGAGGTTGGTGAGTGATTGG	C 型
CAAATTGGAGGACAAGAGGTTGG	C 型
AGGACAAATTGGAGGACAAGAGG	C 型
TCTTGTCTCCAATTTGTCTCTGG	C 型
CGATAGCCAGGACAAATTGGAGG	C 型
CAGCGATAGCCAGGACAAATTGG	C 型
CAATTTGTCTGGCTATCGCTGG	C 型
AGACACATCCAGCGATAGCCAGG	C 型
GCTATCGCTGGATGTGTCTGCGG	C 型
GGCATAGCAGCAGGATGAAGAGG	C 型
AGAAGATGAGGCATAGCAGCAGG	C 型
GCTATGCCTCATCTTCTTGTGG	C 型
GAAGAACCAACAAGAAGATGAGG	C 型
CATCTTCTTGTGGTTCTTCTGG	C 型
TTGGTTCTTCTGGACTACCAAGG	C 型
GACAAACGGGCAACATACCTTGG	C 型
CCTGGAAGTAGAGGACAAACGGG	C 型
CCCGTTTGTCTCTACTTCCAGG	C 型

TCCTGGAAGTAGAGGACAAACGG	C 型
GTTGATGTTCTGGAAGTAGAGG	C 型
GTGCTGGTGGTTGATGTTCTGG	C 型
AGGAACATCAACCACCAGCACGG	C 型
GGAACATCAACCACCAGCACGGG	C 型
GAACATCAACCACCAGCACGGGG	C 型
CTTGCATGGCCCCGTGCTGGTGG	C 型
GGTCTTGCATGGCCCCGTGCTGG	C 型
GGAATCGTGCAGGTCTTGCATGG	C 型
ACCTGCACGATTCTGCTCAAGG	C 型
TCCTTGAGCAGGAATCGTGCAGG	C 型
AACATAGAGGTTCTTGAGCAGG	C 型
GCAACAAGAGGGAAACATAGAGG	C 型
GGTTTTGTACAGCAACAAGAGGG	C 型
AGGTTTTGTACAGCAACAAGAGG	C 型
TTGTTGCTGTACAAAACCTTCGG	C 型
TGCTGTACAAAACCTTCGGACGG	C 型
CAAGTGCAGTTTCCGTCCGAAGG	C 型
GTATTCCCATCCCATCATCCTGG	C 型
TATTCCCATCCCATCATCCTGGG	C 型
AAAGCCCAGGATGATGGGATGGG	C 型
GAAAGCCCAGGATGATGGGATGG	C 型
TTGCGAAAGCCCAGGATGATGGG	C 型
CTTGCGAAAGCCCAGGATGATGG	C 型
TAGGAATCTTGCGAAAGCCCAGG	C 型
GGGCTTTCGCAAGATTCCTATGG	C 型
GGCTTTCGCAAGATTCCTATGGG	C 型
TCGCAAGATTCCTATGGGAGTGG	C 型
CGCAAGATTCCTATGGGAGTGGG	C 型
GGACTGAGGCCCACTCCCATAGG	C 型
GGGCCTCAGTCCGTTTCTCCTGG	C 型
GAGCCAGGAGAAACGGACTGAGG	C 型
GTAAACTGAGCCAGGAGAAACGG	C 型
TGGCACTAGTAAACTGAGCCAGG	C 型
TACTAGTGCCATTTGTTTCAGTGG	C 型
CCTACGAACCACTGAACAAATGG	C 型
CCATTTGTTTCAGTGGTTCGTAGG	C 型
CATTTGTTTCAGTGGTTCGTAGGG	C 型
TAGGGCTTTCCTCCCACTGTTTGG	C 型
AACTGAAAGCCAAACAGTGGGGG	C 型
TAACTGAAAGCCAAACAGTGGGG	C 型
ATAACTGAAAGCCAAACAGTGGG	C 型
TATAACTGAAAGCCAAACAGTGG	C 型
CTGTTTGGCTTTCAGTTATATGG	C 型

TTTCAGTTATATGGATGATGTGG	C 型
TTATATGGATGATGTGGTATTGG	C 型
TATATGGATGATGTGGTATTGGG	C 型
ATATGGATGATGTGGTATTGGGG	C 型
TATGGATGATGTGGTATTGGGGG	C 型
CTCAAGATGTTGTACAGACTTGG	C 型
ATTGGTAATAGAGGTAAAAAGGG	C 型
AATTGGTAATAGAGGTAAAAAGG	C 型
CAAAAGAAAATTGGTAATAGAGG	C 型
TACCAATTTTCTTTTGTCTTTGG	C 型
ACCAATTTTCTTTTGTCTTTGGG	C 型
ACCCAAAGACAAAAGAAAATTGG	C 型
ACCCTAATAAAACCAAACGTTGG	C 型
CCCAACGTTTGGTTTTATTAGGG	C 型
CCCTAATAAAACCAAACGTTGGG	C 型
CCTAATAAAACCAAACGTTGGGG	C 型
CCCCAACGTTTGGTTTTATTAGG	C 型
AAGGGAGTAGCCCCAACGTTTGG	C 型
GGGCTACTCCCTTAACTTCATGG	C 型
GGCTACTCCCTTAACTTCATGGG	C 型
TACATATCCCATGAAGTTAAGGG	C 型
TTACATATCCCATGAAGTTAAGG	C 型
AACTTCATGGGATATGTAATTGG	C 型
TGGGATATGTAATTGGAAGTTGG	C 型
GGGATATGTAATTGGAAGTTGGG	C 型
GGATATGTAATTGGAAGTTGGGG	C 型
AAGTTGGGGTACTTTACCGCAGG	C 型
TTTAGTACAATATGTTCCCTGCGG	C 型
CAATCAATAGGTCTATTTACAGG	C 型
CTGTAAATAGACCTATTGATTGG	C 型
TGACAGACTTTCCAATCAATAGG	C 型
GAAAGTCTGTCAAAGAATTGTGG	C 型
AAAGTCTGTCAAAGAATTGTGGG	C 型
TCAAAGAATTGTGGGTCTTTTGG	C 型
CAAAGAATTGTGGGTCTTTTGGG	C 型
GCTGCCCCCTTTACACAATGTGG	C 型
ATAGCCACATTGTGTAAAAGGGG	C 型
GATAGCCACATTGTGTAAAAGGG	C 型
GGATAGCCACATTGTGTAAAAGG	C 型
GCATATAAAGGCATCAAGGCAGG	C 型
ACATGCATATAAAGGCATCAAGG	C 型
GATTGTATACATGCATATAAAGG	C 型
TGCATGTATACAATCTAAGCAGG	C 型
CACTTTCTCGCCAACTTACAAGG	C 型

CACAGAAAGGCCTTGTAAGTTGG	C 型
CAGATATTGTTTACACAGAAAGG	C 型
TGAACCTTTACCCCGTTGCCCGG	C 型
GTTGCCGGGCAACGGGGTAAAGG	C 型
TTTACCCCGTTGCCCGGCAACGG	C 型
CTGACCGTTGCCGGGCAACGGGG	C 型
CCCGTTGCCCGGCAACGGTCAGG	C 型
CCTGACCGTTGCCGGGCAACGGG	C 型
ACCTGACCGTTGCCGGGCAACGG	C 型
GCAGAGACCTGACCGTTGCCGGG	C 型
GGCAGAGACCTGACCGTTGCCGG	C 型
GGGTTGCGTCAGCAAAACACTTGG	C 型
TTTGCTGACGCAACCCCCACTGG	C 型
CTGACGCAACCCCCACTGGATGG	C 型
TGACGCAACCCCCACTGGATGGG	C 型
GACGCAACCCCCACTGGATGGGG	C 型
AACCCCCACTGGATGGGGCTTGG	C 型
GGCCAAGCCCCATCCAGTGGGGG	C 型
TGGCCAAGCCCCATCCAGTGGGG	C 型
ATGGCCAAGCCCCATCCAGTGGG	C 型
TATGGCCAAGCCCCATCCAGTGG	C 型
ACTGGATGGGGCTTGGCCATAGG	C 型
GGGGCTTGGCCATAGGCCATCGG	C 型
ACGCATGCGCCGATGGCCTATGG	C 型
ATAGGCCATCGGCGCATGCGTGG	C 型
AGGTTCACGCATGCGCCGATGG	C 型
GCGCATGCGTGGAACCTTTGTGG	C 型
GATCGGCAGAGGAGCCACAAAGG	C 型
TCCTCTGCCGATCCATACTGCGG	C 型
TCCGCAGTATGGATCGGCAGAGG	C 型
AGGAGTTCCGCAGTATGGATCGG	C 型
CTGCTAGGAGTTCCGCAGTATGG	C 型
TGCGAGCGAAACAAGCTGCTAGG	C 型
CAGCTTGTTTCGCTCGCAGCCGG	C 型
TGTTTCGCTCGCAGCCGGTCTGG	C 型
CGATAAGTTTCGCTCCAGACCGG	C 型
CGGTCTGGAGCGAAACTTATCGG	C 型
AGAGAGGACAACAGAGTTGTCTGG	C 型
ACAACTCTGTTGTCTCTCTCGG	C 型
AGGAGGTGTATTTCCGAGAGAGG	C 型
GGAAATACACCTCCTTTCCATGG	C 型
ACCTCCTTTCCATGGCTGCTAGG	C 型
CCTCCTTTCCATGGCTGCTAGGG	C 型
CCCTAGCAGCCATGGAAAGGAGG	C 型

ACACCCTAGCAGCCATGGAAAGG	C 型
GCAGCACACCCTAGCAGCCATGG	C 型
TGCTAGGGTGTGCTGCCAACTGG	C 型
GCTGCCAACTGGATCCTGCGCGG	C 型
CTGCCAACTGGATCCTGCGCGGG	C 型
CGTCCCGCGCAGGATCCAGTTGG	C 型
AGACAAAGGACGTCCCGCGCAGG	C 型
GTCCTTTGTCTACGTCCCGTCGG	C 型
CGCCGACGGGACGTAGACAAAGG	C 型
CCGCGGGATTACAGCGCCGACGGG	C 型
CCCGTCGGCGCTGAATCCCGCGG	C 型
TCCGCGGGATTACAGCGCCGACGG	C 型
ATCCCGCGGACGACCCGTCTCGG	C 型
TCCCGCGGACGACCCGTCTCGGG	C 型
CCCCGAGACGGGTCGTCCGCGGG	C 型
CCCGCGGACGACCCGTCTCGGGG	C 型
GCCCCGAGACGGGTCGTCCGCGG	C 型
CGACCCGTCTCGGGGCCGTTTGG	C 型
GACCCGTCTCGGGGCCGTTTGGG	C 型
ACCCGTCTCGGGGCCGTTTGGGG	C 型
GCCCCAAACGGCCCCGAGACGGG	C 型
AGCCCCAAACGGCCCCGAGACGG	C 型
GGGGACGGTAGAGCCCCAAACGG	C 型
ACGGCAGATGAAGAAGGGGACGG	C 型
CGGAACGGCAGATGAAGAAGGGG	C 型
CCGGAACGGCAGATGAAGAAGGG	C 型
CCCTTCTTCATCTGCCGTTCGGG	C 型
GCCGGAACGGCAGATGAAGAAGG	C 型
TCTGCCGTTCGGGCCGACCACGG	C 型
CTGCCGTTCGGGCCGACCACGGG	C 型
TGCCGTTCGGGCCGACCACGGGG	C 型
CGCCCCGTGGTCGGCCGGAACGG	C 型
AGGTGCGCCCCGTGGTCGGCCGG	C 型
AGAGAGGTGCGCCCCGTGGTCGG	C 型
GTAAAGAGAGGTGCGCCCCGTGG	C 型
GGGGCGCACCTCTCTTTACGCGG	C 型
CGGGGAGACCGCGTAAAGAGAGG	C 型
CAGATGAGAAGGCACAGACGGGG	C 型
GCAGATGAGAAGGCACAGACGGG	C 型
GGCAGATGAGAAGGCACAGACGG	C 型
GTCTGTGCCTTCTCATCTGCCGG	C 型
ACACGGTCCGGCAGATGAGAAGG	C 型
GAAGCGAAGTGACACGGTCCGG	C 型
GAGGTGAAGCGAAGTGACACCGG	C 型

CTTCACCTCTGCACGTCGCATGG	C 型
GGTCTCCATGCGACGTGCAGAGG	C 型
GACCACCGTGAACGCCACCAGG	C 型
GACCTGGTGGGCGTTCACGGTGG	C 型
CAAGACCTGGTGGGCGTTCACGG	C 型
CGCCCACCAGGTCTTGCCCAAGG	C 型
GACCTTGGGCAAGACCTGGTGGG	C 型
AGACCTTGGGCAAGACCTGGTGG	C 型
GTAAGACCTTGGGCAAGACCTGG	C 型
TGCCCAAGGTCTTACATAAGAGG	C 型
GTCTCTTATGTAAGACCTTGGG	C 型
AGTCCTCTTATGTAAGACCTTGG	C 型
GTCTTACATAAGAGGACTCTTGG	C 型
AATGTCAACGACCGACCTTGAGG	C 型
TTTGAAGTATGCCTCAAGGTCGG	C 型
AGTCTTTGAAGTATGCCTCAAGG	C 型
AAGACTGTGTGTTTAAAGACTGG	C 型
AGACTGTGTGTTTAAAGACTGGG	C 型
CTGTGTGTTTAAAGACTGGGAGG	C 型
GTTTAAAGACTGGGAGGAGTTGG	C 型
TTTAAAGACTGGGAGGAGTTGGG	C 型
TTAAAGACTGGGAGGAGTTGGGG	C 型
TAAAGACTGGGAGGAGTTGGGGG	C 型
AGACTGGGAGGAGTTGGGGGAGG	C 型
AGGAGTTGGGGGAGGAGATTAGG	C 型
GGGGGAGGAGATTAGGTAAAGG	C 型
AGGTAAAGGTCTTTGTACTAGG	C 型
TTAAAGGTCTTTGTACTAGGAGG	C 型
TCTTTGTACTAGGAGGCTGTAGG	C 型
AGGAGGCTGTAGGCATAAATTGG	C 型
GTGAAAAAGTTGCATGGTGCTGG	C 型
GCAGAGGTGAAAAAGTTGCATGG	C 型
AACATGAGATGATTAGGCAGAGG	C 型
GACATGAACATGAGATGATTAGG	C 型
AGCTTGGAGGCTTGAACAGTAGG	C 型
CAAGCCTCCAAGCTGTGCCTTGG	C 型
AAGCCTCCAAGCTGTGCCTTGGG	C 型
CCTCCAAGCTGTGCCTTGGGTGG	C 型
CCACCCAAGGCACAGCTTGGAGG	C 型
AAGCCACCAAGGCACAGCTTGG	C 型
AGCTGTGCCTTGGGTGGCTTTGG	C 型
GCTGTGCCTTGGGTGGCTTTGGG	C 型
CTGTGCCTTGGGTGGCTTTGGGG	C 型
CCATGCCCCAAAGCCACCCAAGG	C 型

CCTTGGGTGGCTTTGGGGCATGG	C 型
ATTGACCCGTATAAAGAATTTGG	C 型
ATGCTCCAAATTCITTATACGGG	C 型
GATGCTCCAAATTCITTATACGG	C 型
TAAAGAATTTGGAGCATCTGTGG	C 型
ATAGAAGGAAAAGAAGTCAGAAGG	C 型
TCGAGGAGATCTCGAATAGAAGG	C 型
ACAGAGCAGAAGCGGTGTCGAGG	C 型
ACACCGCTTCTGCTCTGTATCGG	C 型
CACCGCTTCTGCTCTGTATCGGG	C 型
CTCCCGATACAGAGCAGAAGCGG	C 型
CGCTTCTGCTCTGTATCGGGAGG	C 型
TCGGGAGGCCTTAGAGTCTCCGG	C 型
ACAATGTTCCGGAGACTCTAAGG	C 型
TGGTGAGGTGAACAATGTTCGGG	C 型
CACCTCACCATACAGCACTCAGG	C 型
TGCCTGAGTGCTGTATGGTGAGG	C 型
TAGCTTGCCTGAGTGCTGTATGG	C 型
TCAGGCAAGCTATTCTGTGTTGG	C 型
CAGGCAAGCTATTCTGTGTTGGG	C 型
AGGCAAGCTATTCTGTGTTGGGG	C 型
TTGGGGTGAGTTGATGAATCTGG	C 型
AGTTGATGAATCTGGCCACCTGG	C 型
GTIGATGAATCTGGCCACCTGGG	C 型
GATGAATCTGGCCACCTGGGTGG	C 型
ATGAATCTGGCCACCTGGGTGGG	C 型
CAAATTACTTCCCACCCAGGTGG	C 型
CACCTGGGTGGGAAGTAATTTGG	C 型
TTCCAAATTACTTCCCACCCAGG	C 型
ATTTGGAAGACCCAGCATCCAGG	C 型
TTTGGAAGACCCAGCATCCAGGG	C 型
CTACTAATCCCTGGATGCTGGG	C 型
ACTACTAATCCCTGGATGCTGG	C 型
ATAGCTGACTACTAATCCCTGG	C 型
CAGCTATGTCAATGTTAATATGG	C 型
AGCTATGTCAATGTTAATATGGG	C 型
ACAATAGTTGTCTGATTTTATAGG	C 型
TAAAAATCAGACAACTATTGTGG	C 型
CACATTCCCTGTCTTACTTTTGG	C 型
TTCTCTTCCAAAAGTAAGACAGG	C 型
AGAAACTGTTCTTGAGTATTTGG	C 型
CTTGAGTATTTGGTATCTTTTGG	C 型
ATTTGGTATCTTTTGGAGTGTGG	C 型
TTTGGTGGTCTGTAAGCGGGAGG	C 型

GCATTGGTGGTCTGTAAGCGGG	C 型
GGCATTGGTGGTCTGTAAGCGG	C 型
GATAAGATAGGGGCATTGGTGG	C 型
GTTGATAAGATAGGGGCATTGG	C 型
CGGAAGTGTTGATAAGATAGGGG	C 型
CCGGAAGTGTTGATAAGATAGGG	C 型
CCCTATCTTATCAACACTTCCGG	C 型
TCCGGAAGTGTTGATAAGATAGG	C 型
CGTCTAACAACAGTAGTTTCCGG	C 型
ACTACTGTTGTTAGACGACGAGG	C 型
CTGTTGTTAGACGACGAGGCAGG	C 型
CGAGGGAGTTCTTCTTCTAGGGG	C 型
GCGAGGGAGTTCTTCTTCTAGGG	C 型
GGCGAGGGAGTTCTTCTTCTAGG	C 型
GATCTTCGTCTGCGAGGCGAGGG	C 型
AGATCTTCGTCTGCGAGGCGAGG	C 型
GATTGAGATCTTCGTCTGCGAGG	C 型
GATTGAGATCTTCTGCGACGCGG	C 型
GTCGCAGAAGATCTCAATCTCGG	C 型
TCGCAGAAGATCTCAATCTCGGG	C 型
ATCTCAATGTTAGTATCCCTTGG	C 型
TAGTATCCCTTGGAATCATAAGG	C 型
TATCCCTTGGAATCATAAGGTGG	C 型
ATCCCTTGGAATCATAAGGTGGG	C 型
TTCCACCTTATGAGTCCAAGGG	C 型
TTTCCACCTTATGAGTCCAAGG	C 型
CATAAGGTGGGAACTTTACTGG	C 型
ATAAGGTGGGAACTTTACTGGG	C 型
TACCTGTCTTTAATCCCGAGTGG	C 型
TGCCACTCGGGATTAAAGACAGG	C 型
AAGGAGGGAGTTTGCCACTCGGG	C 型
AAAGGAGGGAGTTTGCCACTCGG	C 型
AAATGAATGTGAGGAAAGGAGGG	C 型
TAAATGAATGTGAGGAAAGGAGG	C 型
CTGTAAATGAATGTGAGGAAAGG	C 型
CTTTCCTCACATTCATTTACAGG	C 型
TCCTCACATTCATTTACAGGAGG	C 型
TCCTCCTGTAAATGAATGTGAGG	C 型
TAATAGATGTCAACAATATGTGG	C 型
AATAGATGTCAACAATATGTGGG	C 型
TTTTTTCATTAAGTGAAGAGGG	C 型
CTTTTTTCATTAAGTGAAGAGG	C 型
CTCTTACAGTTAATGAAAAAGG	C 型
TAAAATTAATTATGCCTGCTAGG	C 型

AGGTTAGGATAGAACCTAGCAGG	C 型
GGCAAATATTTGGTAAGGTTAGG	C 型
CTAAGGGCAAATATTTGGTAAGG	C 型
TTTGTCTAAGGGCAAATATTTGG	C 型
AAATATTTGCCCTTAGACAAAGG	C 型
GGTTTAATGCCTTTGTCTAAGGG	C 型
CGGTTTAATGCCTTTGTCTAAGG	C 型
ACTGCATGTTTCAGGATAATACGG	C 型
TAATGATTAAGTGCATGTTTCAGG	C 型
TTAATCATTAAGTCAAACTAGG	C 型
GGCATTATTTACATACTCTGTGG	C 型
TTATTTACATACTCTGTGGAAGG	C 型
TTACATACTCTGTGGAAGGCTGG	C 型
ACACGCAGCGCCTCATTTTGTGG	C 型
CACGCAGCGCCTCATTTTGTGGG	C 型
ATATGGTGACCCACAAAATGAGG	C 型
TTTGTGGGTCACCATATTCTTGG	C 型
TTGTGGGTCACCATATTCTTGGG	C 型
AGCTCTTGTCCCAAGAATATGG	C 型
TGGGAACAAGAGCTACAGCATGG	C 型
GGGAACAAGAGCTACAGCATGGG	C 型
AACAAGAGCTACAGCATGGGAGG	C 型
AGAGCTACAGCATGGGAGGTTGG	C 型
TGGTCTTCCAAACCTCGACAAGG	C 型
TTCCAAACCTCGACAAGGCATGG	C 型
TCCAAACCTCGACAAGGCATGGG	C 型
CCCCATGCCTTGTGAGGTTTGG	C 型
CCAAACCTCGACAAGGCATGGGG	C 型
TTCGTCCCCATGCCTTGTGAGG	C 型
TCTTTCTGTTCCCAATCCTCTGG	C 型
CTTTCTGTTCCCAATCCTCTGGG	C 型
GGAAAGAATCCCAGAGGATTGGG	C 型
GGGAAAGAATCCCAGAGGATTGG	C 型
TGATCGGGAAAGAATCCCAGAGG	C 型
ATTCTTTCCCGATCACCAGTTGG	C 型
GCAGGGTCCAAGTGGTGATCGGG	C 型
CGCAGGGTCCAAGTGGTGATCGG	C 型
CACCAGTTGGACCCTGCGTTCCG	C 型
CTCCGAACGCAGGGTCCAAGTGG	C 型
TTGAGTTGGCTCCGAACGCAGGG	C 型
TTTGAGTTGGCTCCGAACGCAGG	C 型
CCAATCTGGATTGTTGAGTTGG	C 型
CCAAGTCAAACAATCCAGATTGG	C 型
CAAGTCAAACAATCCAGATTGGG	C 型

TTGGGGTTGAAGTCCCAATCTGG	C 型
TTGGGACTTCAACCCCAACAAGG	C 型
TCAACCCCAACAAGGATCACTGG	C 型
CTGGCCAGTGATCCTTGTTGGGG	C 型
TCTGGCCAGTGATCCTTGTTGGG	C 型
CTCTGGCCAGTGATCCTTGTTGG	C 型
CAACAAGGATCACTGGCCAGAGG	C 型
TCACTGGCCAGAGGCAAATCAGG	C 型
TGGCCAGAGGCAAATCAGGTAGG	C 型
GCTCCTACCTGATTGTCCTCTGG	C 型
AGAGGCCAAATCAGGTAGGAGCGG	C 型
GAGGCCAAATCAGGTAGGAGCGGG	C 型
CAGGTAGGAGCGGGAGCATTCTGG	C 型
AGGTAGGAGCGGGAGCATTCTGGG	C 型
GGAGCGGGAGCATTCGGGCCAGG	C 型
GAGCGGGAGCATTCGGGCCAGGG	C 型
CCGTGTGGTGGGGTGAACCCTGG	C 型
CCAGGGTTCACCCACACACCGG	C 型
GGGTTCACCCACACACGCGCGG	C 型
CAAAAGACCGCCGTGTGGTGGGG	C 型
CCCACCACACGGCGGTCTTTTGG	C 型
CCAAAAGACCGCCGTGTGGTGGG	C 型
CCCAAAAGACCGCCGTGTGGTGG	C 型
CCACCACACGGCGGTCTTTTGGG	C 型
CACCACACGGCGGTCTTTTGGGG	C 型
CACCCAAAAGACCGCCGTGTGG	C 型
CACACGGCGGTCTTTTGGGGTGG	C 型
TCTTTTGGGGTGGAGCCCTCAGG	C 型
GGGGTGGAGCCCTCAGGCTCAGG	C 型
GGGTGGAGCCCTCAGGCTCAGGG	C 型
TCAATATGCCCTGAGCCTGAGGG	C 型
GTCAATATGCCCTGAGCCTGAGG	C 型
GAGGCAGGAGGAGGTGCTGCTGG	C 型
CGATTGGTGGAGGCAGGAGGAGG	C 型
CTCCTCCTGCCTCCACCAATCGG	C 型
TGCCGATTGGTGGAGGCAGGAGG	C 型
GACTGCCGATTGGTGGAGGCAGG	C 型
GCCTCCACCAATCGGCAGTCAGG	C 型
TCCTGACTGCCGATTGGTGGAGG	C 型
TCTTCCTGACTGCCGATTGGTGG	C 型
CTGTCTTCCTGACTGCCGATTGG	C 型
AGAGGTGGAGAGATGGGAGTAGG	C 型
TCTCTTAGAGGTGGAGAGATGGG	C 型
GTCTCTTAGAGGTGGAGAGATGG	C 型

GGATGACTGTCTCTTAGAGGTGG	C 型
TGAGGATGACTGTCTCTTAGAGG	C 型
TCTAAGAGACAGTCATCCTCAGG	C 型
GTCATCCTCAGGCCATGCAGTGG	C 型
GAGTTCCACTGCATGGCCTGAGG	C 型
TGTTGTGGAGTTCCACTGCATGG	C 型
GGTCTCCATGCGACGTGCAGAGG	D 型
CTTCACCTCTGCACGTTCGCATGG	D 型
GAGGTGAAGCGAAGTGCACACGG	D 型
GAAGCGAAGTGCACACGGTCCGG	D 型
CATTCGGTGGGCGTTCACGGTGG	D 型
CAACATTCGGTGGGCGTTCACGG	D 型
CGCCCACCGAATGTTGCCCAAGG	D 型
ACACGGTCCGGCAGATGAGAAGG	D 型
GACCTTGGGCAACATTCGGTGGG	D 型
AGACCTTGGGCAACATTCGGTGG	D 型
GTAAGACCTTGGGCAACATTCGG	D 型
GTCTGTGCCTTCTCATCTGCCGG	D 型
GGCAGATGAGAAGGCACAGACGG	D 型
GCAGATGAGAAGGCACAGACGGG	D 型
CAGATGAGAAGGCACAGACGGGG	D 型
TGCCCAAGGTCTTACATAAGAGG	D 型
GTCCTCTTATGTAAGACCTTGGG	D 型
AGTCCTCTTATGTAAGACCTTGG	D 型
GTCTTACATAAGAGGACTCTTGG	D 型
CGGGGAGTCCGCGTAAAGAGAGG	D 型
GGGGCGCACCTCTCTTTACGCGG	D 型
GTAAAGAGAGGTGCGCCCCGTGG	D 型
AGAGAGGTGCGCCCCGTGGTCGG	D 型
AGGTGCGCCCCGTGGTCGGTCGG	D 型
AATGTCAACGACCGACCTTGAGG	D 型
CGCCCCGTGGTCGGTCGGAACGG	D 型
TGCCGTTCGACCGACCGGGG	D 型
CTGCCGTTCGACCGACCGGG	D 型
TCTGCCGTTCGACCGACCGG	D 型
TGGTCGGTCGGAACGGCAGACGG	D 型
TTTGAAGTATGCCTCAAGGTCGG	D 型
GTCGGAACGGCAGACGGAGAAGG	D 型
AGTCTTTGAAGTATGCCTCAAGG	D 型
TCGGAACGGCAGACGGAGAAGGG	D 型
CGGAACGGCAGACGGAGAAGGGG	D 型
AAGACTGTTTGTTTAAAGACTGG	D 型
AGACTGTTTGTTTAAAGACTGGG	D 型
CTGTTTGTTTAAAGACTGGGAGG	D 型

GTTTAAAGACTGGGAGGAGTTGG	D 型
TTTAAAGACTGGGAGGAGTTGGG	D 型
TTAAAGACTGGGAGGAGTTGGGG	D 型
TAAAGACTGGGAGGAGTTGGGGG	D 型
AGTCCCAAGCGACCCCGAGAAGG	D 型
GTCCCAAGCGACCCCGAGAAGGG	D 型
AGACTGGGAGGAGTTGGGGGAGG	D 型
GACCCTTCTCGGGGTCGCTTGGG	D 型
CGACCCTTCTCGGGGTCGCTTGG	D 型
CCCCGAGAAGGGTCGTCCGCAGG	D 型
CCTGCGGACGACCCTTCTCGGGG	D 型
TCCTGCGGACGACCCTTCTCGGG	D 型
ATCCTGCGGACGACCCTTCTCGG	D 型
GGGGGAGGAGATTAGATTAAAGG	D 型
TCCGCAGGATTCAGCGCCGACGG	D 型
CCCGTCGGCGCTGAATCCTGCGG	D 型
CCGCAGGATTCAGCGCCGACGGG	D 型
AGATTAAAGGTCTTTGTACTAGG	D 型
TTAAAGGTCTTTGTACTAGGAGG	D 型
TCTTTGTACTAGGAGGCTGTAGG	D 型
CGCCGACGGGACGTAAACAAAGG	D 型
GTCCTTTGTTTACGTCCCGTCGG	D 型
AGGAGGCTGTAGGCATAAATTGG	D 型
AAACAAAGGACGTCCCGCGCAGG	D 型
CGTCCCGCGCAGGATCCAGTTGG	D 型
CTGCCAACTGGATCCTGCGCGGG	D 型
GCTGCCAACTGGATCCTGCGCGG	D 型
TGCTAGGCTGTGCTGCCAACTGG	D 型
GTGAAAAAGTTGCATGGTGCTGG	D 型
GCAGCACAGCCTAGCAGCCATGG	D 型
GCAGAGGTGAAAAAGTTGCATGG	D 型
ACATCGTATCCATGGCTGCTAGG	D 型
AACAAGAGATGATTAGGCAGAGG	D 型
GCAAATATACATCGTATCCATGG	D 型
ATGGATACGATGTATATTTGCGG	D 型
TGGATACGATGTATATTTGCGGG	D 型
GACATGAACAAGAGATGATTAGG	D 型
ACGATGTATATTTGCGGGAGAGG	D 型
AGCTTGGAGGCTTGAACAGTAGG	D 型
CAAGCCTCCAAGCTGTGCCTTGG	D 型
AAGCCTCCAAGCTGTGCCTTGGG	D 型
CCTCCAAGCTGTGCCTTGGGTGG	D 型
CCACCCAAGGCACAGCTTGGAGG	D 型
AAGCCACCCAAGGCACAGCTTGG	D 型

AGCTGTGCCTTGGGTGGCTTTGG	D 型
GCTGTGCCTTGGGTGGCTTTGGG	D 型
CTGTGCCTTGGGTGGCTTTGGGG	D 型
GGTCTGGAGCAAACATTATCGGG	D 型
AGGTCTGGAGCAAACATTATCGG	D 型
CCTTGGGTGGCTTTGGGGCATGG	D 型
CCATGCCCCAAAGCCACCCAAGG	D 型
TGTTTTGCTCGCAGCAGGTCTGG	D 型
CCTGCTGCGAGCAAAACAAGCGG	D 型
CCGCTTGTTTTGCTCGCAGCAGG	D 型
TGCGAGCAAAACAAGCGGCTAGG	D 型
ATCGACCCTTATAAAGAATTTGG	D 型
TAGCTCCAAATTCTTTATAAGGG	D 型
GTAGCTCCAAATTCTTTATAAGG	D 型
CGGCTAGGAGTTCCGCAGTATGG	D 型
TAAAGAATTTGGAGCTACTGTGG	D 型
AGGAGTTCCGCAGTATGGATCGG	D 型
TCCGCAGTATGGATCGGCAGAGG	D 型
TCCTCTGCCGATCCATACTGCGG	D 型
GATCGGCAGAGGAGCCGAAAAGG	D 型
GCGCGTGCGTGGAACCTTTTCGG	D 型
ACTGAAGGAAAGAAGTCAGAAGG	D 型
AGGTTCCACGCACGCGCTGATGG	D 型
ATGGGGCCATCAGCGCGTGCGTGG	D 型
TCTAGAAGATCTCGTACTGAAGG	D 型
ACTGGCTGGGGCTTGGTCATGGG	D 型
CACTGGCTGGGGCTTGGTCATGG	D 型
CATGACCAAGCCCCAGCCAGTGG	D 型
ATGACCAAGCCCCAGCCAGTGGG	D 型
TGACCAAGCCCCAGCCAGTGGGG	D 型
GACCAAGCCCCAGCCAGTGGGGG	D 型
AACCCCCACTGGCTGGGGCTTGG	D 型
GACGCAACCCCCACTGGCTGGGG	D 型
ATACCGCCTCAGCTCTGTATCGG	D 型
TACCGCCTCAGCTCTGTATCGGG	D 型
TGACGCAACCCCCACTGGCTGGG	D 型
CTGACGCAACCCCCACTGGCTGG	D 型
TTCCCGATACAGAGCTGAGGCGG	D 型
GGCTTCCCGATACAGAGCTGAGG	D 型
TTTGCTGACGCAACCCCCACTGG	D 型
GGGTTGCGTCAGCAAACACTTGG	D 型
GCAAACACTTGGCACAGACCTGG	D 型
ACAATGCTCAGGAGACTCTAAGG	D 型
GGCACAGACCTGGCCGTTGCCGG	D 型

GCACAGACCTGGCCGTTGCCGGG	D 型
TGGTGAGGTGAACAATGCTCAGG	D 型
ACCTGGCCGTTGCCGGGCAACGG	D 型
CCTGGCCGTTGCCGGGCAACGGG	D 型
CCCGTTGCCCGGCAACGGCCAGG	D 型
CTGGCCGTTGCCGGGCAACGGGG	D 型
TTTACCCCGTTGCCCGGCAACGG	D 型
GTTGCCGGGCAACGGGGTAAAGG	D 型
CACCTCACCATACTGCACTCAGG	D 型
TGAACCTTTACCCCGTTGCCCGG	D 型
TGCCTGAGTGCAGTATGGTGAGG	D 型
GGGCAACGGGGTAAAGGTTCAGG	D 型
TTGCTTGCCTGAGTGCAGTATGG	D 型
TCAGGCAAGCAATTCTTTGCTGG	D 型
CAGGCAAGCAATTCTTTGCTGGG	D 型
AGGCAAGCAATTCTTTGCTGGGG	D 型
GGCAAGCAATTCTTTGCTGGGGG	D 型
GCAAGCAATTCTTTGCTGGGGGG	D 型
CAGGTATTGTTTACACAGAAAGG	D 型
CACAGAAAGGCCTTGTAAGTTGG	D 型
AACTAATGACTCTAGCTACCTGG	D 型
ACTAATGACTCTAGCTACCTGGG	D 型
CACTTTCTCGCCAACTTACAAGG	D 型
AATGACTCTAGCTACCTGGGTGG	D 型
ATGACTCTAGCTACCTGGGTGGG	D 型
TACCTGGGTGGGTGTTAATTTGG	D 型
TTCCAAATTAACACCCACCCAGG	D 型
TGCATGTATTCAATCTAAGCAGG	D 型
AGATTGAATACATGCATACAAGG	D 型
GATTGAATACATGCATACAAGGG	D 型
ACTACTAGGTCTCTAGATGCTGG	D 型
GCATACAAGGGCATTAAACGCAGG	D 型
TGTTGACATAACTGACTACTAGG	D 型
CAGTTATGTCAACACTAATATGG	D 型
AGTTATGTCAACACTAATATGGG	D 型
GGATAACCACATTGTGTAAATGG	D 型
GATAACCACATTGTGTAAATGGG	D 型
ATAACCACATTGTGTAAATGGGG	D 型
GCTGCCCCATTTACACAATGTGG	D 型
CTAATATGGGCCTAAAGTTCAGG	D 型
ACAAGAGTTGCCTGAACTTTAGG	D 型
TAAAGTTCAGGCAACTCTTGTGG	D 型
CAACGAATTGTGGGTCTTTTGGG	D 型
TCAACGAATTGTGGGTCTTTTGG	D 型

AAAGTATGTCAACGAATTGTGGG	D 型
GAAAGTATGTCAACGAATTGTGG	D 型
CACATTTCTTGTCTCACTTTTGG	D 型
TGACATACTTTCCAATCAATAGG	D 型
CTATTAACAGGCCTATTGATTGG	D 型
CAATCAATAGGCCTGTTAATAGG	D 型
AGAAACCGTTATAGAGTATTTGG	D 型
TTAGAAAACCTTCTATTAACAGG	D 型
AGACACCAAATACTCTATAACGG	D 型
ATAGAGTATTTGGTGTCTTTCGG	D 型
ATTTGGTGTCTTTCGGAGTGTGG	D 型
TTTTGTATGATGTGTTCTTGTGG	D 型
TATGATGTGTTCTTGTGGCAAGG	D 型
TTTGGTGGTCTATAAGCTGGAGG	D 型
GCATTGGTGGTCTATAAGCTGG	D 型
GATAGGATAGGGGCATTGGTGG	D 型
GTTATGTCATTGGAAGTTATGGG	D 型
GTTGATAGGATAGGGGCATTGG	D 型
GGTTATGTCATTGGAAGTTATGG	D 型
CGGAAGTGTTGATAGGATAGGGG	D 型
CCGGAAGTGTTGATAGGATAGGG	D 型
CCCTATCCTATCAACACTTCCGG	D 型
TCCGGAAGTGTTGATAGGATAGG	D 型
AATTTTATGGGTTATGTCATTGG	D 型
TAGTTTCCGGAAGTGTTGATAGG	D 型
GGTACTCTCTGAATTTTATGGG	D 型
GGGTTACTCTCTGAATTTTATGG	D 型
CGTCTAACAACAGTAGTTTCCGG	D 型
ACTACTGTTGTTAGACGACGAGG	D 型
CTGTTGTTAGACGACGAGGCAGG	D 型
CCCCATCTCTTGTTTTGTAGG	D 型
CCTAACAAAACAAAGAGATGGGG	D 型
CCCTAACAAAACAAAGAGATGGG	D 型
CCCATCTCTTGTTTTGTAGGG	D 型
ACCCTAACAAAACAAAGAGATGG	D 型
CGAGGGAGTTCTTCTTCTAGGGG	D 型
GCGAGGGAGTTCTTCTTCTAGGG	D 型
GGCGAGGGAGTTCTTCTTCTAGG	D 型
CTCCCTCGCCTCGCAGACGAAGG	D 型
GACCTTCGTCTGCGAGGCGAGGG	D 型
ACCCAAAGACAAAAGAAAATTGG	D 型
ACCAATTTTCTTTTGTCTTTGGG	D 型
AGACCTTCGTCTGCGAGGCGAGG	D 型
TACCAATTTTCTTTTGTCTTTGG	D 型

GATTGAGACCTTCGTCTGCGAGG	D 型
CAAAAGAAAATTGGTAACAGCGG	D 型
AATTGGTAACAGCGGTAAAAAGG	D 型
ATTGGTAACAGCGGTAAAAAGGG	D 型
GATTGAGATCTTCTGCGACGCGG	D 型
GTCGCAGAAGATCTCAATCTCGG	D 型
TCGCAGAAGATCTCAATCTCGGG	D 型
CTCAAGATGCTGTACAGACTTGG	D 型
ACCTCAATGTTAGTATTCCTTGG	D 型
TCCAAGGAATACTAACATTGAGG	D 型
TATGGATGATGTGGTATTGGGGG	D 型
ATATGGATGATGTGGTATTGGGG	D 型
TATATGGATGATGTGGTATTGGG	D 型
TTATATGGATGATGTGGTATTGG	D 型
TAGTATTCCTTGGACTCATAAGG	D 型
TATTCCTTGGACTCATAAGGTGG	D 型
ATTCCTTGGACTCATAAGGTGGG	D 型
TTTCAGTTATATGGATGATGTGG	D 型
TTCCCTTGGACTCATAAGGTGGGG	D 型
TTCCCCACCTTATGAGTCCAAGG	D 型
CTGTTTGGCTTTCAGTTATATGG	D 型
CATAAGGTGGGGAACTTTACTGG	D 型
TATAACTGAAAGCCAAACAGTGG	D 型
ATAACTGAAAGCCAAACAGTGGG	D 型
TAACTGAAAGCCAAACAGTGGGG	D 型
AACTGAAAGCCAAACAGTGGGGG	D 型
TAGGGCTTTCCCCCACTGTTTGG	D 型
CATTTGTTTCAGTGGTTCGTAGGG	D 型
CCATTTGTTTCAGTGGTTCGTAGG	D 型
CCTACGAACCACTGAACAAATGG	D 型
TACTAGTGCCATTTGTTTCAGTGG	D 型
TACCTGTCTTTAATCCTCATTGG	D 型
TTCCAATGAGGATTAAAGACAGG	D 型
TGGCACTAGTAAACTGAGCCAGG	D 型
AAAGATGGTGTTTTCCAATGAGG	D 型
GTAAACTGAGCCAGGAGAAACGG	D 型
TAAACTGAGCCAGGAGAAACGGG	D 型
GAGCCAGGAGAAACGGGCTGAGG	D 型
AAATGTATATTAGGAAAAGATGG	D 型
GGGCCTCAGCCCGTTTCTCCTGG	D 型
TCTTGGTGTAATGTATATTAGG	D 型
GGGCTGAGGCCCACTCCCATAGG	D 型
CGGAAAATTCCTATGGGAGTGGG	D 型
TCGGAAAATTCCTATGGGAGTGG	D 型

CACATTTTTTGATAATGTCTTGG	D 型
GGCTTTCGGAAAATTCCTATGGG	D 型
GGGCTTTCGGAAAATTCCTATGG	D 型
TAGGAATTTTCCGAAAGCCCAGG	D 型
TTTCCGAAAGCCCAGGATGATGG	D 型
AAAAAATGTGAACAGTTTGTAGG	D 型
TTCCGAAAGCCCAGGATGATGGG	D 型
ATCCCATCATCCTGGGCTTTCGG	D 型
GAAAGCCCAGGATGATGGGATGG	D 型
AAAGCCCAGGATGATGGGATGGG	D 型
TATTCCCATCCCATCATCCTGGG	D 型
GTATTCCCATCCCATCATCCTGG	D 型
GGATGATGGGATGGGAATACAGG	D 型
TTTTCTCATTAAGTGAAGTGGG	D 型
CTTTTCTCATTAAGTGAAGTGG	D 型
CAGGTGCAATTTCCGTCCGAAGG	D 型
GCAATTTCCGTCCGAAGGTTTGG	D 型
TGCTGTACCAAACCTTCGGACGG	D 型
CTGTTGCTGTACCAAACCTTCGG	D 型
CGAAGGTTTGGTACAGCAACAGG	D 型
AGGTTTGGTACAGCAACAGGAGG	D 型
TGCAATTGATTATGCCTGCTAGG	D 型
GGTTTGGTACAGCAACAGGAGGG	D 型
GCAACAGGAGGGATACATAGAGG	D 型
GCCTGCTAGGTTTTATCCAAAGG	D 型
ACCTTTGGATAAAACCTAGCAGG	D 型
GGTAAATATTTGGTAACCTTTGG	D 型
GGTTACCAAATATTTACCATTGG	D 型
TCCTTGAGCAGTAGTCATGCAGG	D 型
ACCTGCATGACTACTGCTCAAGG	D 型
CTTATCCAATGGTAAATATTTGG	D 型
CAAATATTTACCATTGGATAAGG	D 型
GAGCAGTAGTCATGCAGGTTCGG	D 型
AAATATTTACCATTGGATAAGGG	D 型
GTAGTCATGCAGGTTCGGCATGG	D 型
GGTTAATACCCCTATCCAATGG	D 型
GGTTCGGCATGGTCCCGTGCTGG	D 型
TCGGCATGGTCCCGTGCTGGTGG	D 型
TGGTCCCGTGCTGGTGGTTGAGG	D 型
GGATCCTCAACCACCAGCACGGG	D 型
ACTAGATGTTCTGGATAATAAGG	D 型
AGGATCCTCAACCACCAGCACGG	D 型
GTGCTGGTGGTTGAGGATCCTGG	D 型
TAATGATTAAGTAGATGTTCTGG	D 型

GTTGAGGATCCTGGAATTAGAGG	D 型
TCCTGGAATTAGAGGACAAACGG	D 型
CCTGGAATTAGAGGACAAACGGG	D 型
CCCGTTTGTCTCTAATTCCAGG	D 型
TGTGTAAATAGTGTCTAGTTTGG	D 型
GACACTATTTACACACTCTATGG	D 型
CTATTTACACACTCTATGGAAGG	D 型
TTGGTTCTTCTGGACTATCAAGG	D 型
TTTACACACTCTATGGAAGGCGG	D 型
TTACACACTCTATGGAAGGCGGG	D 型
CATCTTCTTGTGGTTCTTCTGG	D 型
GAAGAACCAACAAGAAGATGAGG	D 型
GCTATGCCTCATCTTCTTGTGG	D 型
AGAAGATGAGGCATAGCAGCAGG	D 型
GGCATAGCAGCAGGATGAAGAGG	D 型
ACACATAGCGCCTCATTTTGTGG	D 型
CACATAGCGCCTCATTTTGTGGG	D 型
ATATGGTGACCCACAAAATGAGG	D 型
TTTGTGGGTCACCATATTCTTGG	D 型
TTGTGGGTCACCATATTCTTGGG	D 型
GTTATCGCTGGATGTGTCTGCGG	D 型
AGATCTTGTCCCAAGAATATGG	D 型
AGACACATCCAGCGATAACCAGG	D 型
TGGGAACAAGATCTACAGCATGG	D 型
CAACTTGTCTGGTTATCGCTGG	D 型
GGGAACAAGATCTACAGCATGGG	D 型
CAGCGATAACCAGGACAAGTTGG	D 型
GGAACAAGATCTACAGCATGGGG	D 型
CGATAACCAGGACAAGTTGGAGG	D 型
ACCAGGACAAGTTGGAGGACAGG	D 型
TCCTGTCTCCAACCTTGTCTGG	D 型
AGGACAAGTTGGAGGACAGGAGG	D 型
CAAGTTGGAGGACAGGAGGTTGG	D 型
ACAGGAGGTTGGTGAGTGATTGG	D 型
TCTTTCCACCAGCAATCCTCTGG	D 型
CTTTCCACCAGCAATCCTCTGGG	D 型
GGAGGTTGGTGAGTGATTGGAGG	D 型
GAATCCCAGAGGATTGCTGGTGG	D 型
GTTGGTGAGTGATTGGAGGTTGG	D 型
TTGGTGAGTGATTGGAGGTTGGG	D 型
TGGTGAGTGATTGGAGGTTGGGG	D 型
AAAGAATCCCAGAGGATTGCTGG	D 型
TGGTCGGGAAAGAATCCCAGAGG	D 型
AGGTTGGGGACTGCGAATTTTGG	D 型

ATTCTTTCCCGACCACCAGTTGG	D 型
GCTGGATCCAACTGGTGGTCGGG	D 型
GGCTGGATCCAACTGGTGGTCGG	D 型
CGAATTTTGCCAAGACACACGG	D 型
TGAAGGCTGGATCCAACTGGTGG	D 型
CTCTGAAGGCTGGATCCAACTGG	D 型
GGGGGAACCTACCGTGTGTCTTGG	D 型
GCTGTGTTTGCTCTGAAGGCTGG	D 型
ATTTGCTGTGTTTGCTCTGAAGG	D 型
ACTTCTCTCAATTTTCTAGGGGG	D 型
CAAACACAGCAAATCCAGATTGG	D 型
AAACACAGCAAATCCAGATTGGG	D 型
GACTTCTCTCAATTTTCTAGGGG	D 型
GGACTTCTCTCAATTTTCTAGGG	D 型
TGGACTTCTCTCAATTTTCTAGG	D 型
TTGGGATTGAAGTCCCAATCTGG	D 型
TTGGGACTTCAATCCCAACAAGG	D 型
CGCAGAGTCTAGACTCGTGGTGG	D 型
CACCACGAGTCTAGACTCTGCGG	D 型
TACCGCAGAGTCTAGACTCGTGG	D 型
TCAATCCCAACAAGGACACCTGG	D 型
TCTGGCCAGGTGTCCTTGTTGGG	D 型
GTCTGGCCAGGTGTCCTTGTTGG	D 型
CTAGACTCTGCGGTATTGTGAGG	D 型
CACCTGGCCAGACGCCAACAAGG	D 型
TACCTTGTGTCGTCTGGCCAGG	D 型
TGGCCAGACGCCAACAAGGTAGG	D 型
GCTCCTACCTTGTGTCGTCTGG	D 型
GACGCCAACAAGGTAGGAGCTGG	D 型
TGCTCCAGCTCCTACCTTGTTGG	D 型
AAGGTAGGAGCTGGAGCATTCGG	D 型
AGGTAGGAGCTGGAGCATTCGGG	D 型
AGGAGCTGGAGCATTCGGGCTGG	D 型
GGAGCTGGAGCATTCGGGCTGGG	D 型
ACCCCGCCTGTAACACGAGAAGG	D 型
CCCTTCTCGTGTTACAGGCGGGG	D 型
CCCCGCCTGTAACACGAGAAGGG	D 型
CCCGCCTGTAACACGAGAAGGGG	D 型
CCCCTTCTCGTGTTACAGGCGGG	D 型
ACCCCTTCTCGTGTTACAGGCGG	D 型
AGGACCCCTTCTCGTGTTACAGG	D 型
GTAACACGAGAAGGGGTCCTAGG	D 型
CTGGGTTTCACCCACCGCACGG	D 型
GGTTTCACCCACCGCACGGAGG	D 型

CAAAAGGCCTCCGTGCGGTGGGG	D 型
CCAAAAGGCCTCCGTGCGGTGGG	D 型
CCCACCGCACGGAGGCCTTTTGG	D 型
CCCAAAAGGCCTCCGTGCGGTGG	D 型
CCACCGCACGGAGGCCTTTTGGG	D 型
CACCGCACGGAGGCCTTTTGGGG	D 型
CACCCCAAAGGCCTCCGTGCGG	D 型
CGCACGGAGGCCTTTTGGGGTGG	D 型
AACATCACATCAGGATTCCTAGG	D 型
AACATGGAGAACATCACATCAGG	D 型
CCTTTTGGGGTGGAGCCCTCAGG	D 型
CCTGAGGGCTCCACCCCAAAGG	D 型
GGGGTGGAGCCCTCAGGCTCAGG	D 型
GGGTGGAGCCCTCAGGCTCAGGG	D 型
ATGTTCCTCATGTTCAAGCGCAGG	D 型
TGTTCTCCATGTTCAAGCGCAGGG	D 型
GTAGTATGCCCTGAGCCTGAGGG	D 型
TGGGGACCCTGCGCTGAACATGG	D 型
TGTAGTATGCCCTGAGCCTGAGG	D 型
GTCAATCTTCTCGAGGATTGGGG	D 型
CGTCAATCTTCTCGAGGATTGGG	D 型
TCGTCAATCTTCTCGAGGATTGG	D 型
CCTTATCGTCAATCTTCTCGAGG	D 型
CCTCGAGAAGATTGACGATAAGG	D 型
CTCGAGAAGATTGACGATAAGGG	D 型
GAAGATTGACGATAAGGGAGAGG	D 型
GAGGCAGGAGGCGGATTGCTGG	D 型
CGATTGGTGGAGGCAGGAGGCGG	D 型
GATAAGGGAGAGGCAGTAGTCGG	D 型
TGGCGATTGGTGGAGGCAGGAGG	D 型
GTCTGGCGATTGGTGGAGGCAGG	D 型
GGAGAGGCAGTAGTCGGAACAGG	D 型
GAGAGGCAGTAGTCGGAACAGGG	D 型
GCCTCCACCAATCGCCAGACAGG	D 型
TCCTGTCTGGCGATTGGTGGAGG	D 型
CCTTCCTGTCTGGCGATTGGTGG	D 型
CCACCAATCGCCAGACAGGAAGG	D 型
CTGCCTTCCTGTCTGGCGATTGG	D 型
GGGTAGGCTGCCTTCCTGTCTGG	D 型
AGGGTTTACTGCTCCTGAACTGG	D 型
AAAGGTGGAGACAGCGGGGTAGG	D 型
CCTGAACTGGAGCCACCAGCAGG	D 型
CCTGCTGGTGGCTCCAGTTCAGG	D 型
CTGAACTGGAGCCACCAGCAGGG	D 型

TCTCAAAGGTGGAGACAGCGGGG	D 型
TTCTCAAAGGTGGAGACAGCGGG	D 型
TTTCTCAAAGGTGGAGACAGCGG	D 型
AGCCACCAGCAGGGAAATACAGG	D 型
GGCCTGTATTTCCCTGCTGGTGG	D 型
GGATGAGTGTTTCTCAAAGGTGG	D 型
AGAGGCCTGTATTTCCCTGCTGG	D 型
TGAGGATGAGTGTTTCTCAAAGG	D 型
TTTGAGAAACACTCATCCTCAGG	D 型
GAAATACAGGCCTCTCACTCTGG	D 型
AAATACAGGCCTCTCACTCTGGG	D 型
CTCATCCTCAGGCCATGCAGTGG	D 型
CTGCAAGATCCCAGAGTGAGAGG	D 型
CTCTGGGATCTTGCAGAGTTTGG	D 型
ATCTTGCAGAGTTTGGTGAAAGG	D 型
CAGAGTTTGGTGAAAGGTTGTGG	D 型
GAATTCCACTGCATGGCCTGAGG	D 型
GGTTGTGGAATTCCACTGCATGG	D 型

- [0155] (2)sgRNA功能性筛选
- [0156] 利用293FT构建针对B型、C型和D型HBV表面抗原(HBsAg)、核心抗原(HBcAg)、X蛋白(HBx)和聚合酶(Pol)的报告系统(reporter)细胞系，其示意图如图1A所示。保留蛋白转录起始位点之前包含CpG岛和调控元件的基因组序列，并分别在HBV表面抗原(HBsAg)第38个氨基酸、核心抗原(HBcAg)第129个氨基酸、X蛋白(HBx)第77个氨基酸和聚合酶(Pol)第177个氨基酸之后融合自切肽P2A和EGFP的基因序列，使用PiggyBac转座子系统将报告基因整合进293FT基因组，构建稳转细胞系，利用荧光的强弱来反应蛋白的表达水平，以此快速对CpG岛附近的sgRNA进行功能验证和初步筛选。
- [0157] 本实施例中采用的自切肽P2A的编码序列为：
GCTACTAACTTCAGCCTGCTGAAGCAGGCTGGCGACGTGGAGGAGAACCCTGGACCTGCCACC (SEQ ID NO: 1188)。
- [0158] 所采用的EGFP的编码序列为：

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGTGGTGCCCATCCTGGTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGC
CCACCCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCAC
ATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGG
CGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACACAGCCACAACGTCTATATCATGGCCGA
CAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGC
TGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCCAACGA
GAAGCGCGATCACATGGTCCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCA
TGGACGAGCT (SEQ ID NO: 1189)。

[0159] 利用上述的reporter细胞系进行sgRNA功能验证的筛选方案设计如下(图1B):
在reporter细胞铺板以后,将包含不同sgRNA的本申请所述的组合物EPIREG(该
组合物结构域为DNMT3A-hDNMT3L-xten80-ZIM3-xten80-dCas9,其氨基酸
序列如SEQ ID NO:1194所示,其核酸序列分别如SEQ ID NO:1192(CDS序列)和
1193(质粒序列)所示)通过转染递送到细胞内,二者分别携带红色荧光和蓝色荧
光。利用流式细胞仪分选红色荧光和蓝色荧光的双阳细胞群并长期培养,每隔
一周进行一次流式分析,检测reporter细胞的荧光强度变化,即可反应表观编辑
以后蛋白表达水平的变化。

[0160] (3)实验结果分析

[0161] 图2A-2F所示的结果表明,将包含不同sgRNA的本申请所述的组合物进行转
染14天之后,B型HBV的HBx reporter细胞系的荧光强度的出现了明显的下降,
说明HBx的表达受到了明显的抑制。在转染后的第21天,各个细胞系的结果基
本得到了维持,说明表观编辑的抑制效果得到了维持。本实施例综合了以上在
不同时间点各个sgRNA在该细胞系中的编辑能力,其中,BSG30、BSG35、BS
G43、BSG50均具有极高的编辑效果,且在基因组中靶向相对集中的区域。图2G
为测试sgRNA在B型HBV基因组中的靶向位置与转染后第28天可敲降细胞比例
的关系,该结果显示,敲低效果表现较好的sgRNA主要集中在BSG35和BSG29靶序

列两侧附近的区域，说明对这一部分区域的表观编辑可以较好地实现对HBx表达的抑制。

[0162] 图3A-3B所示的C型HBV的HBx reporter细胞系转染结果表明，在转染后的第14天，其荧光强度出现了明显的下降，说明HBx的表达受到了明显的抑制。图3C为测试sgRNA在C型HBV基因组中的靶向位置与转染后第14天可敲降细胞比例的关系，该结果显示，敲低效果表现较好的sgRNA主要集中在BSG35、SG75和BSG29靶序列两侧附近的区域，说明对这一部分区域的表观编辑可以较好地实现对HBx表达的抑制。

[0163] 图4A-4F所示的D型HBV的HBx reporter细胞系转染结果表明，转染后的第16天，其荧光强度出现了明显的下降，说明HBx的表达受到了明显的抑制。在转染后的第21天和第28天，各个细胞系的结果基本得到了维持，说明表观编辑的抑制效果得到了维持。图4G为测试sgRNA在D型HBV基因组中的靶向位置与转染后第28天可敲降细胞比例的关系，该结果显示，敲低效果表现较好的sgRNA主要集中在BSG35和BSG29靶序列两侧附近的区域，说明对这一部分区域的表观编辑可以较好地实现对HBx表达的抑制。

[0164] 综合三种基因型的HBx reporter细胞系转染结果，使用本申请提供的表观编辑工具靶向B型、C型和D型HBV基因组均可显著地抑制HBx的表达，并且这种抑制作用是长期的，至少可以维持28天。可同时靶向三个基因组的BSG29、BSG31、BSG34、BSG35对于三种基因型HBV的HBx表达都有接近最好的抑制效果。三个基因型的细胞系中，编辑效果最好的sgRNA都主要集中分布于BSG35和BSG29靶序列两侧区域，说明这一部分区域可能是对HBV基因进行表观调控的关键区域。

[0165] 本实施例经过功能性验证的sgRNA总结如下表所示(按各sgRNA的靶序列起始位置排列)：

实验编号	SEQ ID NO:	靶序列的互补核苷酸序列	基因组	靶序列在基因组中位置
BSG42	345	CAGGCCGTTGCCGAGCAACG	B 型	1149-1168
BSG39	657	GCAAACACTTGGCACAGACC	B 型	1168-1187
BSG44	231	ATAGGCCATCAGCGCATGCG	B 型	1219-1238
BSG50	642	GATCGGTAGAGGAGACACAA	B 型	1247-1266
BSG43	954	TCCTCTACCGATCCATACTG	B 型	1254-1273
BSG35	182	AGGAGTTCCGCAGTATGGAT	B 型	1264-1283
BSG30	451	CCTGCTGCGAGCAAAACAAG	B 型	1289-1308
BSG54	253	ATGGAAATGATGTATACTTG	B 型	1356-1375
BSG36	671	GCAGCACAGCCTAGCAGCCA	B 型	1375-1394
BSG28	484	CGCCGACGGGACGTAAACAA	B 型	1421-1440
BSG49	408	CCCGTCGGCGCTGAATCCCG	B 型	1431-1450
BSG52	693	GCGGGCGGTAGAGTCCCAAG	B 型	1473-1492
BSG45	72	ACAATAGGCGGAGAAGCGGG	B 型	1488-1507
BSG55	790	GGTACAATAGGCGGAGAAGC	B 型	1491-1510
BSG48	1063	TGGTCGGCCGGTACAATAGG	B 型	1500-1519
BSG51	441	CCTATTGTACCGGCCGACCA	B 型	1500-1519
BSG53	922	TATTGTACCGGCCGACCACG	B 型	1502-1521
BSG29	810	GTAAAGAGAGGTGCGCCCCG	B 型	1520-1539
BSG26	498	CGGGGAGTCCGCGTAAAGAG	B 型	1532-1551
BSG31	335	CAGATGAGAAGGCACAGACG	B 型	1550-1569
BSG34	591	GAAGCGAAGTGACACCGGTC	B 型	1573-1592
SG76	537	CTGACCGTTGCCGGGCAACG	C 型	1150-1169
SG89	667	GCAGAGACCTGACCGTTGCC	C 型	1158-1177
SG90	743	GGCCAAGCCCCATCCAGTGG	C 型	1200-1219

SG87	1048	TGGCCAAGCCCCATCCAGTG	C 型	1201-1220
SG78	232	ATAGGCCATCGGCGCATGCG	C 型	1220-1239
SG88	640	GATCGGCAGAGGAGCCACAA	C 型	1248-1267
DSG42	955	TCCTCTGCCGATCCATACTG	C 型	1255-1274
BSG35	182	AGGAGTTCCGCAGTATGGAT	C 型	1265-1284
BSG79	1034	TGCGAGCGAAACAAGCTGCT	C 型	1285-1304
BSG85	706	GGAAATACACCTCCTTTCCA	C 型	1356-1375
SG75	417	CCCTAGCAGCCATGGAAAAGG	C 型	1368-1387
SG82	77	ACACCCTAGCAGCCATGGAA	C 型	1371-1390
SG80	485	CGCCGACGGGACGTAGACAA	C 型	1422-1441
BSG49	408	CCCGTCGGCGCTGAATCCCG	C 型	1432-1451
SG84	116	ACGGCAGATGAAGAAGGGGA	C 型	1489-1508
SG86	492	CGGAACGGCAGATGAAGAAG	C 型	1493-1512
SG81	986	TCTGCCGTTCCGGCCGACCA	C 型	1501-1520
BSG29	810	GTAAAGAGAGGTGCGCCCCG	C 型	1521-1540
SG74	497	CGGGGAGACCGCGTAAAGAG	C 型	1533-1552
BSG31	335	CAGATGAGAAGGCACAGACG	C 型	1551-1570
BSG33	81	ACACGGTCCGGCAGATGAGA	C 型	1562-1581
BSG34	591	GAAGCGAAGTGCACACGGTC	C 型	1574-1593
DSG43	650	GATTGAATACATGCATACAA	D 型	1060-1079
BSG39	657	GCAAACACTTGGCACAGACC	D 型	1168-1187
BSG41	788	GGGTTGCGTCAGCAAACT	D 型	1179-1198
DSG45	601	GACCAAGCCCCAGCCAGTGG	D 型	1199-1218
DSG50	1006	TGACCAAGCCCCAGCCAGTG	D 型	1200-1219
DSG49	257	ATGGGCCATCAGCGCGTGCG	D 型	1219-1238
DSG42	955	TCCTCTGCCGATCCATACTG	D 型	1254-1273
BSG35	182	AGGAGTTCCGCAGTATGGAT	D 型	1264-1283
BSG30	451	CCTGCTGCGAGCAAAACAAG	D 型	1289-1308
BSG36	671	GCAGCACAGCCTAGCAGCCA	D 型	1375-1394
BSG27	698	GCTGCCAACTGGATCCTGCG	D 型	1390-1409
BSG28	484	CGCCGACGGGACGTAAACAA	D 型	1421-1440
DSG53	450	CCTGCGGACGACCTTCTCG	D 型	1447-1466
DSG55	491	CGGAACGGCAGACGGAGAAG	D 型	1492-1511
DSG52	1028	TGCCGTTCCGACCGACCACG	D 型	1502-1521
BSG29	810	GTAAAGAGAGGTGCGCCCCG	D 型	1520-1539
BSG26	498	CGGGGAGTCCGCGTAAAGAG	D 型	1532-1551
BSG31	335	CAGATGAGAAGGCACAGACG	D 型	1550-1569
BSG38	669	GCAGATGAGAAGGCACAGAC	D 型	1551-1570
BSG33	81	ACACGGTCCGGCAGATGAGA	D 型	1561-1580
BSG34	591	GAAGCGAAGTGCACACGGTC	D 型	1573-1592

[0166] 实施例2

[0167] 本申请组合物对原代肝细胞(PHH)中HBV标记物的敲低效果

[0168] 本实施例使用的是D型乙肝病毒(GenBank: U95551)感染的原代肝细胞(PHH)细胞系。根据实施例1所述的HBx reporter细胞系筛选的结果, BSG29、BS

G31、BSG34、BSG35位于基因组保守区并表现最好的抑制作用，因此使用这四条sgRNA配制本申请提供的药物组合物并在原代肝细胞中进行表观编辑，以此验证其对HBV基因表达的抑制作用。

- [0169] 实验方案如图5A所示，在原代肝细胞感染HBV两天之后，利用脂质纳米颗粒(LNP)(LNP制备参考文献：<https://doi.org/10.1038/s41586-021-03534-y>)将编码EPIREG的mRNA(SEQ ID NO:1191)和每一种sgRNA(BSG29、BSG31、BSG34、BSG35)的组合物递送到原代肝细胞中(给药剂量为2.5ug/ml，EPIREG和sgRNA的质量比为1: 1)，给药后每两天收集培养基上清液，检测细胞分泌的HBV表面抗原(HBsAg)和核心抗原(HBcAg)的表达水平，直至给药后14天收集细胞样品，抽提细胞中DNA和RNA，检测HBV的mRNA的水平以及HBV基因的cccDNA的变化。两组阳性对照分别使用Myrcludex B(药明康德，C9023GE070-1/PE0723)和RG7834(药明康德，ET25747-14-P1)，Myrcludex B可抑制HBV感染肝细胞，RG7834可抑制病毒蛋白的合成，并且在实验第13天停止RG7834，以观察病毒抗原蛋白的反弹情况。(具体检测方法按以下试剂盒说明书操作：HBsAg ELISA kit，Antu-CL 0310；HBeAg ELISA kit，Antu-CL 0312；RNeasy 96Kit(12)，QIAGEN#74182)
- [0170] 给药后2天、4天、6天、8天和10天的实验结果总结如图5B所示，分别递送包含BSG29、BSG31、BSG34、BSG35、BSG34+BSG35(BSG34+BSG35中EPIREG和sgRNA的质量比为1:0.5:0.5)的组合物至HBV感染的原代肝细胞之后，上清液中HBsAg以及HBeAg/HBcAg的表达水平均表现出不同程度的下降，其中BSG35效果最好，可以同时使HBsAg以及HBeAg/HBcAg的表达水平下降80%左右，而且，两个sgRNA结合(BSG34+BSG35)可以进一步提高抑制效果，使HBsAg以及HBeAg/HBcAg的表达水平进一步下降。对照组HBsAg以及HBeAg/HBcAg的表达水平随着感染天数的增加会持续上升，并分别在感染后第14天和第12天左右达到峰值，而BSG35实验组在给药后的第4天，HBsAg和HBeAg/HBcAg的表达水平基本趋近稳定；BSG35实验组的抑制效果与两个参照抑制剂Myrcludex B和RG7834接近，并且，RG7834停药后可观察到HBsAg以及HBeAg/HBcAg的表达都有明显的反弹(上升)趋势，而BSG35实验组在给药当天药物处理4小时之后就没有后续

给药，病毒抗原蛋白的水平在整个实验周期内一直稳定维持于较低的水平。此部分结果表明，表观编辑可以显著降低HBV感染的PHH中病毒蛋白的表达，并且，使用单个sgRNA的表观编辑可以同时实现对多个病毒蛋白表达的调控。

[0171] 实施例3

[0172] sgRNA慢病毒文库构建及其功能筛选

[0173] 本实施例使用D型HBV X蛋白(HBx)、核心抗原(HBcAg)和表面抗原(HBsAg)reporter细胞系进行文库筛选。首先根据实施例1所述的方法，利用reporter细胞系构建D型HBV基因的sgRNA细胞文库。将该sgRNA文库包装进入慢病毒，获得sgRNA慢病毒文库(携带红色荧光)，将慢病毒文库感染reporter细胞系，分选红色荧光获得sgRNA细胞文库。该分选的细胞文库中每个细胞整合文库中一条sgRNA，同时其本身携带reporter序列(携带绿色荧光)。将实施例2所述的EPIREG工具(该工具氨基酸序列为：SEQ ID NO:1194)(携带蓝色荧光)转染递送到同时携带sgRNA和reporter序列的细胞文库中，利用流式分析仪分选红/绿/蓝色荧光三阳细胞并长期培养，转染进入细胞的EPIREG即可与sgRNA一起沉默reporter中的GFP。在转染后第14天分选细胞文库中GFP阴性细胞群，提取阴性细胞群的基因组DNA进行NGS测序，以未处理的细胞文库作为对照参考，比较阴性细胞群中测序读数变高的sgRNA，从而分别获得各个reporter对应的可发挥编辑作用的候选sgRNA(本筛选方案示意图如图6所示，分析方法参考自文献DOI:10.1186/s13059-014-0554-4)。各个reporter细胞系经过功能性验证的sgRNA总结如下表所示(按各sgRNA的靶序列起始位置排列)。

[0174] HBx reporter筛选得到的sgRNA(如图7上图所示)信息总结如下：

SEQ ID NO:	靶序列的互补核苷酸序列	靶序列在基因组中位置	Log ₂ FC
650	GATTGAATACATGCATACAA	1056-1075	3.726
305	CACAGAAAGGCCCTTGTAAGT	1101-1120	4.9316
762	GGGCAACGGGGTAAAGGTTC	1133-1152	4.8522
1132	TTTACCCCGTTGCCCGGCAA	1141-1160	4.252
553	CTGGCCGTTGCCGGGCAACG	1145-1164	3.9905
735	GGCACAGACCTGGCCGTTGC	1154-1173	4.08
657	GCAAACACTTGGCACAGACC	1164-1183	3.2226
788	GGGTTGCGTCAGCAAACACT	1175-1194	4.4031
1146	TTTGCTGACGAACCCCCAC	1182-1201	4.3908
539	CTGACGCAACCCCCACTGGC	1186-1205	3.1536
1008	TGACGCAACCCCCACTGGCT	1187-1206	4.9717
612	GACGCAACCCCCACTGGCTG	1188-1207	6.7125
601	GACCAAGCCCCAGCCAGTGG	1195-1214	7.1876
1006	TGACCAAGCCCCAGCCAGTG	1196-1215	7.0191
249	ATGACCAAGCCCCAGCCAGT	1197-1216	5.9752
357	CATGACCAAGCCCCAGCCAG	1198-1217	6.2753
327	CACTGGCTGGGGCTTGGTCA	1199-1218	5.8386
127	ACTGGCTGGGGCTTGGTCAT	1200-1219	7.231
257	ATGGGCCATCAGCGCGTGCG	1218-1237	6.5564

204	AGGTTCCACGCACGCGCTGA	1223-1242	3.0712
641	GATCGGCAGAGGAGCCGAAA	1243-1262	7.351
955	TCCTCTGCCGATCCATACTG	1253-1272	7.6855
944	TCCGCAGTATGGATCGGCAG	1254-1273	7.2163
182	AGGAGTTCCGCAGTATGGAT	1260-1279	7.3504
496	CGGCTAGGAGTTCCGCAGTA	1265-1284	7.3529
1033	TGCGAGCAAAACAAGCGGCT	1280-1299	3.1795
451	CCTGCTGCGAGCAAAACAAG	1285-1304	5.3958
197	AGGTCTGGAGCAAAACATTAT	1305-1324	3.6817
792	GGTCTGGAGCAAAACATTATC	1306-1325	5.6925
254	ATGGATACGATGTATATTTG	1352-1371	3.4564
658	GCAAATATACATCGTATCCA	1354-1373	4.7952
88	ACATCGTATCCATGGCTGCT	1362-1381	3.4448
506	CGTCCCGCGCAGGATCCAGT	1393-1412	5.3827
484	CGCCGACGGGACGTAAACAA	1417-1436	3.3079
450	CCTGCGGACGACCCCTTCTCG	1447-1466	4.242
958	TCCTGCGGACGACCCCTTCTC	1446-1465	3.9999
467	CGACCCCTTCTCGGGGTCGCT	1454-1473	2.9245
825	GTCCCAAGCGACCCCGAGAA	1457-1476	5.2457
211	AGTCCCAAGCGACCCCGAGA	1458-1477	4.6655
491	CGGAACGGCAGACGGAGAAG	1488-1507	6.0674
973	TCGGAACGGCAGACGGAGAA	1489-1508	6.3619
834	GTCGGAACGGCAGACGGAGA	1490-1509	4.7704
1065	TGGTCGGTCGGAACGGCAGA	1496-1515	6.0793
985	TCTGCCGTTCCGACCGACCA	1499-1518	6.0276
483	CGCCCCGTGGTCGGTCGGAA	1503-1522	6.4685
200	AGGTGCGCCCCGTGGTCGGT	1508-1527	7.535
150	AGAGAGGTGCGCCCCGTGGT	1512-1531	3.8671
810	GTAAAGAGAGGTGCGCCCCG	1516-1535	5.1825
774	GGGGCGCACCTCTCTTTACG	1520-1539	3.0494
498	CGGGGAGTCCGCGTAAAGAG	1528-1547	2.939
335	CAGATGAGAAGGCACAGACG	1546-1565	4.44
669	GCAGATGAGAAGGCACAGAC	1547-1566	4.1676
591	GAAGCGAAGTGCACACGGTC	1569-1588	5.7107

注：Log₂ FC中的FC指fold change，为对两样品(组)间表达量的比值取以2为底的对数后的所得值，其用于表示两者之间的差异表达水平。例如，通常默认取Log₂ FC绝对值大于1为差异基因的筛选标准。

[0175] HBcAg reporter筛选得到的sgRNA(如图7中图所示)信息总结如下：

SEQ ID NO:	靶序列的互补核苷酸序列	靶序列在基因组中位置	Log ₂ FC
305	CACAGAAAGGCCTTGTAAGT	1101-1120	2.1583
1132	TTTACCCCGTTGCCCGGCAA	1141-1160	2.1043
553	CTGGCCGTTGCCGGGCAACG	1145-1164	2.4462

657	GCAAACACTTGGCACAGACC	1164-1183	1.9192
1008	TGACGCAACCCCCACTGGCT	1187-1206	1.8835
612	GACGCAACCCCCACTGGCTG	1188-1207	2.8035
39	AACCCCCACTGGCTGGGGCT	1193-1212	3.4538
601	GACCAAGCCCCAGCCAGTGG	1195-1214	4.6508
1006	TGACCAAGCCCCAGCCAGTG	1196-1215	4.2194
249	ATGACCAAGCCCCAGCCAGT	1197-1216	2.8648
357	CATGACCAAGCCCCAGCCAG	1198-1217	3.1523
327	CACTGGCTGGGGCTTGGTCA	1199-1218	1.8664
127	ACTGGCTGGGGCTTGGTCAT	1200-1219	2.0792
641	GATCGGCAGAGGAGCCGAAA	1243-1262	5.5293
955	TCCTCTGCCGATCCATACTG	1253-1272	5.1075
944	TCCGCAGTATGGATCGGCAG	1254-1273	4.8272
182	AGGAGTTCCGCAGTATGGAT	1260-1279	5.6721
496	CGGCTAGGAGTTCCGCAGTA	1265-1284	2.9031
451	CCTGCTGCGAGCAAAACAAG	1285-1304	2.9406
792	GGTCTGGAGCAAAACATTATC	1306-1325	2.103
113	ACGATGTATATTTGCGGGAG	1346-1365	2.0632
658	GCAAAATATACATCGTATCCA	1354-1373	2.2171
88	ACATCGTATCCATGGCTGCT	1362-1381	3.4446
506	CGTCCCGCGCAGGATCCAGT	1393-1412	3.3238
427	CCGCAGGATTCAGCGCCGAC	1430-1449	3.4136
958	TCCTGCGGACGACCCTTCTC	1446-1465	3.1878
467	CGACCCCTTCTCGGGGTCGCT	1454-1473	2.0481
825	GTCCCAAGCGACCCCGAGAA	1457-1476	3.3064
211	AGTCCCAAGCGACCCCGAGA	1458-1477	4.4175
491	CGGAACGGCAGACGGAGAAG	1488-1507	4.5168
973	TCGGAACGGCAGACGGAGAA	1489-1508	3.9552
834	GTCGGAACGGCAGACGGAGA	1490-1509	2.7177
1065	TGGTCGGTCGGAACGGCAGA	1496-1515	2.8985
985	TCTGCCGTTCCGACCGACCA	1499-1518	3.1824
1028	TGCCGTTCCGACCGACACG	1501-1520	3.8548
483	CGCCCCGTGGTCGGTCGGAA	1503-1522	4.8386
810	GTAAAGAGAGGTGCGCCCCG	1516-1535	2.5258
669	GCAGATGAGAAGGCACAGAC	1547-1566	3.0945
81	ACACGGTCCGGCAGATGAGA	1557-1576	2.0776
591	GAAGCGAAGTGACACCGGTC	1569-1588	4.4134
628	GAGGTGAAGCGAAGTGACACA	1574-1593	3.2225
791	GGTCTCCATGCGACGTGCAG	1593-1612	5.2397
481	CGCCCACCGAATGTTGCCCA	1624-1643	4.5414
1025	TGCCCCAAGGTCTTACATAAG	1638-1657	2.0852
62	AATGTCAACGACCGACCTTG	1678-1697	2.2035
215	AGTCTTTGAAGTATGCCTCA	1693-1712	2.1589
872	TAAAGACTGGGAGGAGTTGG	1723-1742	3.0933

145	AGACTGGGAGGAGTTGGGGG	1726-1745	3.5473
161	AGATTAAAGGTCTTTGTACT	1754-1773	2.5156
998	TCTTTGTACTAGGAGGCTGT	1764-1783	4.2906
180	AGGAGGCTGTAGGCATAAAAT	1774-1793	2.7289
841	GTGAAAAAAGTTGCATGGTGC	1805-1824	2.0112
668	GCAGAGGTGAAAAAGTTGCA	1811-1830	1.9744
598	GACATGAACAAGAGATGATT	1833-1852	4.6761
372	CCACCCAAGGCACAGCTTGG	1868-1887	4.0382
50	AAGCCACCCAAGGCACAGCT	1871-1890	2.927
560	CTGTGCCTTGGGTGGCTTTG	1876-1895	2.7695
379	CCATGCCCCAAAGCCACCCA	1881-1900	3.0989
981	TCTAGAAGATCTCGTACTGA	1972-1991	2.1673
225	ATACCGCCTCAGCTCTGTAT	1994-2013	3.966
1104	TTCCCGATACAGAGCTGAGG	1997-2016	3.9205
750	GGCTTCCCGATACAGAGCTG	2000-2019	4.6799
74	ACAATGCTCAGGAGACTCTA	2021-2040	2.1202
1068	TGGTGAGGTGAACAATGCTC	2032-2051	2.5736
311	CACCTCACCATACTGCACTC	2045-2064	3.495
120	ACTACTAGGTCTCTAGATGC	2134-2153	2.9841
70	ACAAGAGTTGCCTGAACTTT	2181-2200	2.5329
139	AGACACCAAATACTCTATAA	2240-2259	2.8391

[0176] HBsAg reporter筛选得到的sgRNA(如图7下图所示)信息总结如下：

SEQ ID NO:	靶序列的互补核苷酸序列	靶序列在基因组中位置	Log ₂ FC
410	CCCGTTGCCCGGCAACGGCC	1146-1165	1.9735
601	GACCAAGCCCCAGCCAGTGG	1195-1214	1.3073
127	ACTGGCTGGGGCTTGGTCAT	1200-1219	1.3983
641	GATCGGCAGAGGAGCCGAAA	1243-1262	1.58
955	TCCTCTGCCGATCCATACTG	1253-1272	2.3817
944	TCCGCAGTATGGATCGGCAG	1254-1273	1.6439
182	AGGAGTTCCGCAGTATGGAT	1260-1279	1.3776
792	GGTCTGGAGCAAACATTATC	1306-1325	1.5074
658	GCAAATATACATCGTATCCA	1354-1373	1.8084
88	ACATCGTATCCATGGCTGCT	1362-1381	1.2095
671	GCAGCACAGCCTAGCAGCCA	1371-1390	1.2918
698	GCTGCCAACTGGATCCTGCG	1389-1408	1.3233
506	CGTCCCGCGCAGGATCCAGT	1393-1412	1.5863
242	ATCCTGCGGACGACCCCTTCT	1444-1463	0.88088
467	CGACCCCTTCTCGGGGTCGCT	1454-1473	1.8988
825	GTCCCAAGCGACCCCGAGAA	1457-1476	1.0352
211	AGTCCCAAGCGACCCCGAGA	1458-1477	1.5263
491	CGGAACGGCAGACGGAGAAG	1488-1507	2.6046
973	TCGGAACGGCAGACGGAGAA	1489-1508	1.056
834	GTCGGAACGGCAGACGGAGA	1490-1509	1.2666

985	TCTGCCGTTCCGACCGACCA	1499-1518	1.8695
1028	TGCCGTTCCGACCGACCACG	1501-1520	0.89219
483	CGCCCCGTGGTCGGTCGGAA	1503-1522	2.4519
150	AGAGAGGTGCGCCCCGTGGT	1512-1531	1.453
810	GTAAAGAGAGGTGCGCCCCG	1516-1535	1.7059
774	GGGGCGCACCTCTCTTTACG	1520-1539	1.1527
669	GCAGATGAGAAGGCACAGAC	1547-1566	1.1259
791	GGTCTCCATGCGACGTGCAG	1593-1612	3.4863
1025	TGCCCCAAGGTCTTACATAAG	1638-1657	1.4524
829	GTCTCTTATGTAAGACCTT	1640-1659	0.90289
839	GTCTTACATAAGAGGACTCT	1646-1665	1.3376
62	AATGTCAACGACCGACCTTG	1678-1697	1.5364
215	AGTCTTTGAAATATGCCTCA	1693-1712	1.3799
145	AGACTGGGAGGAGTTGGGGG	1726-1745	1.8314
161	AGATTAAGGTCTTTGTACT	1754-1773	1.8238
1084	TTAAAGGTCCTTGTACTAGG	1757-1776	1.0831
998	TCTTTGTACTAGGAGGCTGT	1764-1783	2.1989
180	AGGAGGCTGTAGGCATAAAT	1774-1793	1.7547
668	GCAGAGGTGAAAAAGTTGCA	1811-1830	1.297
27	AACAAGAGATGATTAGGCAG	1827-1846	1.3821
598	GACATGAACAAGAGATGATT	1833-1852	1.4722
930	TCAGGCAAGCAATTCCTTGC	2063-2082	1.8744
227	ATAGAGTATTTGGTGTCTTT	2245-2264	0.92755
121	ACTACTGTTGTIAGACGACG	2335-2354	1.0553
112	ACCTTTGGATAAAACCTAGC	2639-2658	1.203
593	GAATCCCAGAGGATTGCTGG	2863-2882	0.91237
270	ATTTGCTGTGTTTGCTCTGA	2911-2930	1.397
1072	TGTAGTATGCCCTGAGCCTG	3048-3067	1.1325

- [0177] 实施例4
- [0178] 文库筛选的候选sgRNA效果验证实验
- [0179] 在原代肝细胞(PHH)中进行效果验证
- [0180] 本验证实验使用的是D型乙肝病毒(GenBank: U95551)感染的原代肝细胞(PHH)细胞系(药明康德, M00995-P)。综合实施例3中三个文库的筛选结果, 结合保守性等因素, 挑选出下表所示的20条sgRNA来配制本申请提供的药物组合物并在所述原代肝细胞中进行表观编辑, 以此验证筛选所得候选sgRNA对HBV基因表达的抑制作用。
- [0181] 本实施例用于验证的20条sgRNA序列如下:

实验编号	SEQ ID NO:	靶序列的互补核苷酸序列	靶序列在基因组中位置
D113(DSG45)	601	GACCAAGCCCCAGCCAGTGG	1195-1214
D111	249	ATGACCAAGCCCCAGCCAGT	1197-1216

D108	127	ACTGGCTGGGGCTTGGTCAT	1200-1219
D101(DSG42)	955	TCCTCTGCCGATCCATACTG	1253-1272
D100	944	TCCGCAGTATGGATCGGCAG	1254-1273
D99(BSG35)	182	AGGAGTTCCGCAGTATGGAT	1260-1279
D97	496	CGGCTAGGAGTTCCGCAGTA	1265-1284
D63	506	CGTCCCGCGCAGGATCCAGT	1393-1412
D14(BSG38)	669	GCAGATGAGAAAGGCACAGAC	1547-1566
D1	791	GGTCTCCATGCGACGTGCAG	1593-1612
D16	1025	TGCCCAAGGTCTTACATAAG	1638-1657
D33	215	AGTCTTTGAAGTATGCCTCA	1693-1712
D45	145	AGACTGGGAGGAGTTGGGGG	1726-1745
D58	998	TCTTTGTACTAGGAGGCTGT	1764-1783
D61	180	AGGAGGCTGTAGGCATAAAT	1774-1793
D69	668	GCAGAGGTGAAAAAGTTGCA	1811-1830
D75	598	GACATGAACAGAGATGATT	1833-1852
D199	121	ACTACTGTTGTAGACGACG	2335-2354
D340	593	GAATCCCAGAGGATTGCTGG	2863-2882
D409	1072	TGTAGTATGCCCTGAGCCTG	3048-3067

[0182] 实验方案如图5A所示，在原代肝细胞感染HBV两天之后，利用脂质纳米颗粒(LNP)(LNP制备参考文献：<https://doi.org/10.1038/s41586-021-03534-y>)将编码EPIREG的mRNA(序列为SEQ ID NO:1191)和上述每一种sgRNA的组合物递送到原代肝细胞中(给药剂量为2.5ug/ml，EPIREG和sgRNA的质量比为1: 1)，阳性对照使用Myrcludex B(药明康德提供，C9023GE070-1/PE0723)，Myrcludex B可抑制HBV感染肝细胞。各组给药后每两天收集培养基上清液，检测细胞分泌的HBV表面抗原(HBsAg)、核心抗原(HBcAg/HBeAg)和HBV DNA的表达水平，直至感染后14天收集细胞样品，抽提细胞中DNA和RNA，检测HBV的mRNA的水平。(具体检测方法按以下试剂盒说明书操作：HBsAg ELISA kit，Antu-CL 0310；HBeAg ELISA kit，Antu-CL 0312；RNeasy 96 Kit(12)，QIAGEN#74182)

[0183] 感染后14天的实验结果总结如图8所示，分别递送20条sgRNA和EPIREG的组合物至HBV感染的原代肝细胞之后，上清液中HBsAg，HBeAg/HBcAg以及HBV DNA的表达水平均表现出不同程度的下降，其中17条sgRNA可将HBV病毒抗原以及DNA的表达降低50%以上(除D409、D199、D16)，9条sgRNA抑制率可达到接近80%及以上(D340、D75、D97、D99、D100、D101、D108、D111、D113)，其中，D99(即BSG35)表现出最高的抑制水平；EPIREG对于HBV总RNA和

pgRNA的抑制表现出与HBV抗原和DNA相同的效率和趋势，表明该表观编辑工具有效地抑制了HBV的转录；同时，20条sgRNA和EPIREG的组合物均未影响PHH细胞活力，说明了该表观编辑工具的安全性。综上，通过在PHH细胞系中验证，大部分文库筛选获得的候选sgRNA对于HBV表现出明显的抑制效果。

[0184] 组合物对转基因HBV小鼠中HBV标记物的敲低效果

[0185] 本验证实验使用的是转基因HBV小鼠(北京维通达生物技术有限公司，C57BL/6-HBV，后续饲养和检测由北京维通达完成)，该品系小鼠基因组中插入了1.28倍长度的HBV基因组(A型，GenBank：AF305422.1)。综合所有筛选结果，编号D99(BSG35)的sgRNA位于基因组保守区并表现出最好的抑制作用，因此使用此条sgRNA配制本申请提供的药物组合物并在转基因HBV小鼠中进行表观编辑，以此验证本药物在体内模型中对整合型HBV基因表达的抑制作用。

[0186] 通过尾静脉注射脂质纳米颗粒(LNP)(LNP制备参考文献：<https://doi.org/10.1038/s41586-021-03534-y>)的方式将编码EPIREG的mRNA(序列为SEQ ID NO:1191)和编号D99(BSG35)的sgRNA的组合物(mRNA和sgRNA的质量比为1:1)递送到转基因HBV小鼠中(剂量为10mg/kg)，阴性对照注射PBS(200ul)，阳性对照使用反义寡核苷酸(ASO)药物Bepirovirsen(GSK3228836，剂量为200mg/kg)，其可靶向所有HBV RNA，从而降低HBV抗原的表达。各组给药后定期采血，检测血清中分泌的HBV表面抗原(HBsAg)以及核心抗原(HBcAg/HBeAg)的水平，利用qPCR检测HBV DNA的水平。(乙肝病毒e抗原定量试剂盒：迈克生物，货号IM4403003；乙肝病毒表面抗原定量试剂盒：迈克生物，货号IM4403001；乙肝病毒核酸定量试剂盒：圣湘生物，2015340008；检测方法按试剂盒说明书操作)。

[0187] 给药后实验结果总结如图9A-9C所示，转基因HBV小鼠接受单针EPIREG注射以后，血清中HBsAg，HBeAg以及HBV DNA的表达水平可在一周内迅速降低，降低幅度接近或者优于ASO组，但ASO组HBV标记物会逐渐恢复至原始水平，而EPIREG组HBV标记物可得到稳定持久抑制，其中HBsAg和HBV DNA可在给药约2个月之后继续降低(图9A)。经过分析EPIREG组中每只小鼠的HBV标记物表达曲线(图9B)可知，EPIREG单针给药后，三只小鼠在2个月左右HBsAg

和HBV DNA出现自发地进一步下降，其中HBsAg可以降低至检测下限水平，并且，此三只小鼠血清中可检测到HBsAb的逐步表达(图9C)。此结果表明，在转基因HBV小鼠模型中，包含编码EPIREG的mRNA和本申请提供的sgRNA的表观编辑组合物可以在体内显著持久地沉默整合型HBV基因组的表达，并且由于HBV的长期沉默，小鼠可以逐渐恢复对表面抗原HBsAg的免疫反应，从而进一步降低HBV标记物的表达。

[0188] 组合物对AAV-HBV小鼠中HBV标记物的敲低效果

[0189] 本验证实验使用的是AAV-HBV小鼠模型(造模、后续饲养和检测由北京维通达生物技术有限公司完成)，利用携带1.3倍长度HBV基因组(D型，GenBank: U95551)的腺病毒相关病毒(Adenovirus-associated viruses/AAV)感染小鼠，模拟游离HBV基因组在体内的表达。综合所有筛选结果，编号D99(BSG35)的sgRNA位于基因组保守区并表现出最好的抑制作用，因此使用此条sgRNA配制本申请提供的药物组合物并在AAV-HBV小鼠中进行表观编辑，以此验证本药物在体内模型中对游离型HBV基因表达的抑制作用。

[0190] 通过尾静脉注射的方式，将两个滴度(1E10与1E11)的携带1.3倍长度HBV基因组的AAV(北京五加和基因科技有限公司，AMV-002)递送至小鼠体内，实现HBV基因组在体内不同程度的稳定表达；之后，分别在首次给药当天(D0)和第15天(D15)，通过尾静脉注射脂质纳米颗粒(LNP)(LNP制备参考文献：<https://doi.org/10.1038/s41586-021-03534-y>)将编码EPIREG的mRNA(序列为SEQ ID NO:1191)和编号D99(BSG35)的sgRNA的组合物(mRNA和sgRNA的质量比为1: 1)递送到AAV-HBV小鼠中(两次注射剂量同为5mg/kg)，阴性对照注射PBS(200ul)。各组给药后定期采血，检测血清中分泌的HBV表面抗原(HBsAg)以及核心抗原(HBcAg/HBeAg)水平，利用qPCR检测HBV DNA的水平(所有检测试剂盒如本实施例上部分所述)。

[0191] 给药后实验结果总结如图10所示，AAV-HBV小鼠接受两针EPIREG注射以后，血清中HBsAg，HBeAg以及HBV DNA的表达水平可迅速降低，1E10 AAV组病毒表达水平低，给药后短期内病毒标志物可以降低至接近检测下限；1E11 AAV组病毒表达水平高，给药后可实现病毒标志物更大幅度的降低，最终

该表观编辑工具在 $1E10$ 与 $1E11$ AAV组中均可稳定持久地降低HBV标记物。此结果表明，在AAV-HBV小鼠模型中，该表观编辑可以在体内显著持久地沉默游离型HBV基因组的表达。

权利要求书

[权利要求 1]

一种组合物，所述组合物包含：

(1)融合分子或编码所述融合分子的核酸序列，其中所述融合分子包含至少一种DNA结合蛋白和至少一种基因表达调节剂，以及
(2)至少一种单一引导RNA(sgRNA)，或编码所述sgRNA的核酸序列，

其中所述sgRNA与乙型肝炎病毒(HBV)基因附近和/或HBV基因调控元件内的靶DNA序列互补，所述HBV基因为包含SEQ ID NO:1所示的核苷酸序列的B型HBV基因、包含SEQ ID NO:2所示的核苷酸序列的C型HBV基因或者包含SEQ ID NO:3所示的核苷酸序列的D型HBV基因，所述HBV基因调控元件包含转录起始位点、核心启动子、启动子、增强子、沉默子、绝缘子元件、边界元件和/或基因座控制区，且所述靶DNA序列位于所述HBV基因的第1056位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。

[权利要求 2]

根据权利要求1所述的组合物，所述靶DNA序列位于所述HBV基因的第1056位至第1900位核苷酸之间、第1972位至第2082位核苷酸之间、第2134位至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。

[权利要求 3]

根据权利要求1或2所述的组合物，所述靶DNA序列位于所述HBV基因的第1060位至1079位核苷酸之间、第1149位至第1612位核苷酸之间、第1693位至第1852位核苷酸之间、和/或第2863位至第2882位核苷酸之间。

[权利要求 4]

根据权利要求1-3中任一项所述的组合物，所述靶DNA序列位于下列所述B型HBV基因的一个或多个区域：第1149至第1190位核苷酸之间、第1210位至第1310位核苷酸之间、第1350位至第1400位核苷酸之间、

酸之间、第1420位至第1450位核苷酸之间和第1470位至第1592位核苷酸之间。

[权利要求 5] 根据权利要求1-3中任一项所述的组合物，所述靶DNA序列位于下列所述C型HBV基因的一个或多个区域：第1150位至第1180位核苷酸之间、第1200位至第1310位核苷酸之间、第1350位至第1390位核苷酸之间、第1420位至第1460位核苷酸之间和第1480位至第1593位核苷酸之间。

[权利要求 6] 根据权利要求1-3中任一项所述的组合物，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1056位至第1079位核苷酸之间、第1101位至第1900位核苷酸之间、第1972位至第2082位核苷酸之间、第2134位至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2882位核苷酸之间、第2911位至第2930位核苷酸之间、和第3048位至第3067位核苷酸之间。

[权利要求 7] 根据权利要求1-3和6中任一项所述的组合物，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1060位至第1079位核苷酸之间、第1160位至第1310位核苷酸之间、第1253位至第1284位核苷酸之间、第1370位至第1470位核苷酸之间、第1490位至第1612位核苷酸之间、第1693位至第1852位核苷酸之间、和第2863位至第2882位核苷酸之间。

[权利要求 8] 根据权利要求1-7中任一项所述的组合物，所述sgRNA包含SEQ ID NOs: 4-1165中任一项所示的核苷酸序列。

[权利要求 9] 根据权利要求1-7中任一项所述的组合物，所述sgRNA包含SEQ ID NOs: 4-1165中任一项所示的核苷酸序列的部分序列，所述部分序列的长度为15-20个碱基对。

[权利要求 10] 根据权利要求1-9中任一项所述的组合物，所述至少一种DNA结合蛋白是CRISPR酶、锌指核酸酶(ZNF)、转录激活子样效应因子核

酸酶(TALEN)、归巢核酸内切酶、dCas9-FokI核酸酶或MegaTal核酸酶。

- [权利要求 11] 根据权利要求10所述的组合物，所述CRISPR酶为2类Cas蛋白和/或其突变体。
- [权利要求 12] 根据权利要求10或11所述的组合物，所述CRISPR酶为下列所述的一种或多种Cas蛋白：2类II-A型Cas蛋白、2类II-B型Cas蛋白、2类II-C型Cas蛋白、2类V-A型Cas蛋白、2类V-B型Cas蛋白、2类V-C型Cas蛋白、2类V-U型Cas蛋白、和以上所述的突变体。
- [权利要求 13] 根据权利要求10-12中任一项所述的组合物，所述CRISPR酶为Cas9蛋白和/或其突变体。
- [权利要求 14] 根据权利要求1-13中任一项所述的组合物，所述至少一种DNA结合蛋白是dCas9。
- [权利要求 15] 根据权利要求14所述的组合物，所述dCas9包括金黄色葡萄球菌(*Staphylococcus aureus*)dCas9、化脓性链球菌(*Streptococcus pyogenes*)dCas9、空肠弯曲菌(*Campylobacter jejuni*)dCas9、白喉棒状杆菌(*Corynebacterium diphtheria*)dCas9、凸腹真杆菌(*Eubacterium ventriosum*)dCas9、巴氏链球菌(*Streptococcus pasteurianus*)dCas9、香肠乳杆菌(*Lactobacillus farciminis*)dCas9、球形球毛菌(*Sphaerochaeta globus*)dCas9、固氮螺菌属(*Azospirillum*，例如，菌株B510)dCas9、嗜重氮葡萄糖醋杆菌(*Gluconacetobacter diazotrophicus*)dCas9、灰色奈瑟球菌(*Neisseria cinerea*)dCas9、肠道罗斯拜瑞氏菌(*Roseburia intestinalis*)dCas9、食清洁剂细小棒菌(*Parvibaculum lavamentivorans*)dCas9、卤水硝酸盐裂解菌(*Nitratifractor salinus*)，例如，菌株DSM 16511)dCas9、海鸥弯曲菌(*Campylobacter lari*，例如，菌株CF89-12)dCas9、嗜热链球菌(*Streptococcus thermophilus*，例如，菌株LMD-9)dCas9。
- [权利要求 16] 根据权利要求14或15所述的组合物，所述dCas9包含SEQ ID NO：1170-1187所示的氨基酸序列。

- [权利要求 17] 根据权利要求1-16中任一项所述的组合物，所述至少一种基因表达调节剂提供所述HBV基因附近和/或HBV基因调控元件内至少一个核苷酸的修饰。
- [权利要求 18] 根据权利要求1-17中任一项所述的组合物，所述至少一种基因表达调节剂包含选自DNA甲基转移酶、DNA羟甲基化酶、DNA去甲基化酶、组蛋白甲基转移酶、组蛋白去甲基化酶、组蛋白乙酰转移酶、组蛋白去乙酰化酶、磷酸酶、激酶、转录活化物、转录阻遏物、以上所述的部分、以及以上所述的任意组合中的一种或多种。
- [权利要求 19] 根据权利要求17所述的组合物，所述至少一个核苷酸的所述修饰是DNA甲基化。
- [权利要求 20] 根据权利要求1-19中任一项所述的组合物，所述至少一种基因表达调节剂包含选自DNA甲基转移酶(DNMT)、基于锌指蛋白的转录因子、以上所述的部分、以及以上所述的任意组合中的一种或多种。
- [权利要求 21] 根据权利要求1-20中任一项所述的组合物，所述至少一种基因表达调节剂包含DNA甲基转移酶(DNMT)或其部分，以及基于锌指蛋白的转录因子或其部分。
- [权利要求 22] 根据权利要求18、20和21所述的组合物，所述DNA甲基转移酶是DNMT3A、DNMT3B、DNMT3L、DNMT1或DNMT2。
- [权利要求 23] 根据权利要求22所述的组合物，所述DNMT3A包含SEQ ID NO:1166所示的氨基酸序列，以及所述DNMT3L包含SEQ ID NO:1167或1195所示的氨基酸序列。
- [权利要求 24] 根据权利要求20或21所述的组合物，所述基于锌指蛋白的转录因子是Krüppel相关抑制因子(KRAB)或衍生自ZIM3的KRAB结构域(ZIM3 KRAB)。
- [权利要求 25] 根据权利要求24所述的组合物，所述基于锌指蛋白的转录因子包含SEQ ID NO:1168或1196所示的氨基酸序列。

- [权利要求 26] 根据权利要求18和20-25中任一项所述的组合物，所述DNA甲基转移酶选自DNMT3A和DNMT3L及其组合，且所述基于锌指蛋白的转录因子是KRAB或ZIM3 KRAB。
- [权利要求 27] 根据权利要求1-26中任一项所述的组合物，所述融合分子包含与所述至少一种DNA结合蛋白的C端、N端或两端融合的所述至少一种基因表达调节剂。
- [权利要求 28] 根据权利要求1-27中任一项所述的组合物，所述至少一种基因表达调节剂与所述至少一种DNA结合蛋白直接融合。
- [权利要求 29] 根据权利要求1-27中任一项所述的组合物，所述至少一种基因表达调节剂通过非调节剂、第二调节剂或接头与所述至少一种DNA结合蛋白间接融合。
- [权利要求 30] 根据权利要求1-29中任一项所述的组合物，所述融合分子包含结构域DNMT3A-DNMT3L-dCas9-KRAB或结构域DNMT3A-DNMT3L-ZIM3 KRAB-dCas9，其中，-表示所述融合分子各部分结构域直接和/或间接地连接，且所述各部分结构域按照从N端到C端的顺序。
- [权利要求 31] 根据权利要求1-30中任一项所述的组合物，所述融合分子包含SEQ ID NO:1169或1194所示的氨基酸序列。
- [权利要求 32] 根据权利要求1-31中任一项所述的组合物，所述融合分子还包含至少一个核定位序列(NLS)。
- [权利要求 33] 根据权利要求32所述的组合物，所述至少一个核定位序列与所述至少一种DNA结合蛋白的C端、N端或两端直接或间接融合。
- [权利要求 34] 根据权利要求1-33中任一项所述的组合物，编码所述融合分子的所述核酸序列是脱氧核糖核酸(DNA)或信使核糖核酸(mRNA)。
- [权利要求 35] 根据权利要求1-34中任一项所述的组合物，所述融合分子包装在脂质体或脂质纳米颗粒中。
- [权利要求 36] 根据权利要求1-35中任一项所述的组合物，所述融合分子和所述sgRNA包装在脂质体或脂质纳米颗粒中。

- [权利要求 37] 根据权利要求1-36中任一项所述的组合物，所述融合分子和所述sgRNA包装在相同的脂质体或脂质纳米颗粒中，或者包装在不同的脂质体或脂质纳米颗粒中。
- [权利要求 38] 根据权利要求1-37中任一项所述的组合物，所述脂质体或所述脂质纳米颗粒包含可电离脂质(20%–70%，摩尔比)、聚乙二醇化脂质(0%–30%，摩尔比)、支撑脂质(30%–50%，摩尔比)和胆固醇(10%–50%，摩尔比)。
- [权利要求 39] 根据权利要求38所述的组合物，所述可电离脂质选自pH响应性可电离脂质、热响应性可电离脂质和光响应性可电离脂质。
- [权利要求 40] 根据权利要求1-34中任一项所述的组合物，所述融合分子包装在AAV载体中。
- [权利要求 41] 根据权利要求1-34和40中任一项所述的组合物，所述融合分子和所述sgRNA包装在AAV载体中。
- [权利要求 42] 根据权利要求1-34、40和41中任一项所述的组合物，所述融合分子和所述sgRNA包装在相同的AAV载体或不同的AAV载体中。
- [权利要求 43] 根据权利要求1-42中任一项所述的组合物，所述组合物是包含药学上可接受的载体的药物组合物。
- [权利要求 44] 一种单一引导RNA(sgRNA)，所述sgRNA包含与靶DNA序列互补的序列，所述靶DNA序列位于乙型肝炎病毒(HBV)基因附近和/或HBV基因调控元件内；所述HBV基因为包含SEQ ID NO:1所示的核苷酸序列的B型HBV基因、包含SEQ ID NO:2所示的核苷酸序列的C型HBV基因或者包含SEQ ID NO:3所示的核苷酸序列的D型HBV基因，所述HBV基因调控元件包含转录起始位点、核心启动子、启动子、增强子、沉默子、绝缘子元件、边界元件和/或基因座控制区，且所述靶DNA序列位于所述HBV基因的第1056位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。

- [权利要求 45] 根据权利要求44所述的sgRNA，所述靶DNA序列位于所述HBV基因的第1056位至第1900位核苷酸之间、第1972位至第2082位核苷酸之间、第2134位至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2930位核苷酸之间、和/或第3048位至第3067位核苷酸之间。
- [权利要求 46] 根据权利要求44或45所述的sgRNA，所述靶DNA序列位于所述HBV基因的第1060位至第1079位核苷酸之间、第1149位至第1612位核苷酸之间、第1693位至第1852位核苷酸之间、和/或第2863位至第2882位核苷酸之间。
- [权利要求 47] 根据权利要求44-46中任一项所述的sgRNA，所述靶DNA序列位于下列所述B型HBV基因的一个或多个区域：第1149至第1190位核苷酸之间、第1210位至第1310位核苷酸之间、第1350位至第1400位核苷酸之间、第1420位至第1450位核苷酸之间和第1470位至第1592位核苷酸之间。
- [权利要求 48] 根据权利要求44-46中任一项所述的sgRNA，所述靶DNA序列位于下列所述C型HBV基因的一个或多个区域：第1150位至第1180位核苷酸之间、第1200位至第1310位核苷酸之间、第1350位至第1390位核苷酸之间、第1420位至第1460位核苷酸之间和第1480位至第1593位核苷酸之间。
- [权利要求 49] 根据权利要求44-46中任一项所述的组合物，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1056位至第1079位核苷酸之间、第1101位至第1900位核苷酸之间、第1972位至第2085位核苷酸之间、第2134位至第2264位核苷酸之间、第2335位至第2354位核苷酸之间、第2639位至第2658位核苷酸之间、第2863位至第2882位核苷酸之间、第2911位至第2930位核苷酸之间、和第3048位至第3067位核苷酸之间。
- [权利要求 50] 根据权利要求44-46和49中任一项所述的sgRNA，所述靶DNA序列位于下列所述D型HBV基因的一个或多个区域：第1060位至第1079

位核苷酸之间、第1160位至第1310位核苷酸之间、第1253位至第1284位核苷酸之间、第1370位至第1470位核苷酸之间、第1490位至第1612位核苷酸之间、第1693位至第1852位核苷酸之间、和第2863位至第2882位核苷酸之间。

- [权利要求 51] 根据权利要求44-50中任一项所述的sgRNA，所述sgRNA包含SEQ ID NOs:4-1165中任一项所示的核苷酸序列。
- [权利要求 52] 根据权利要求44-50中任一项所述的sgRNA，所述sgRNA包含所述SEQ ID NOs:4-1165中任一项所示的核苷酸序列的部分序列，所述部分序列的长度为15-20个碱基对。
- [权利要求 53] 编码权利要求44-52中任一项所述的sgRNA的核酸分子。
- [权利要求 54] 减少或消除细胞中表达乙型肝炎病毒(HBV)基因产物的方法，所述方法包括将权利要求1-43中任一项所述的组合物引入所述细胞中，从而减少或消除所述细胞中表达所述HBV基因产物的步骤。
- [权利要求 55] 体内减少或消除受试者中表达乙型肝炎病毒(HBV)基因产物的方法，所述方法包括将权利要求1-43中任一项所述的组合物引入所述受试者的细胞中，从而减少或消除所述受试者中表达所述HBV基因产物的步骤。
- [权利要求 56] 用于治疗受试者乙型肝炎病毒(HBV)感染相关疾病或减轻受试者HBV感染相关疾病症状的方法，所述方法包括将有效量的权利要求1-43中任一项所述的组合物引入到所述受试者的细胞的步骤。
- [权利要求 57] 根据权利要求55或56所述的方法，所述受试者为哺乳动物，诸如人、猴、小鼠、大鼠、兔、猪、马、猫和狗。
- [权利要求 58] 根据权利要求55-57中任一项所述的方法，所述方法包括向所述受试者施用一次或一次以上的所述组合物。
- [权利要求 59] 根据权利要求55-58中任一项所述的方法，所述方法包括向所述受试者施用至少两次所述组合物。

- [权利要求 60] 根据权利要求55-59中任一项所述的方法，所述施用一次以上或施用至少两次所述组合物的间隔为1天、2天、3天、4天、5天、6天、7天、8天、9天、10天、11天、12天、13天、14天或15天。
- [权利要求 61] 根据权利要求56所述的方法，所述HBV感染相关疾病包括由HBV感染引起的肝炎、肝硬化、肝纤维化和肝细胞癌。
- [权利要求 62] 权利要求1-43中任一项所述的组合物，所述组合物用于治疗受试者乙型肝炎病毒(HBV)感染相关疾病或减轻受试者HBV感染相关疾病症状。
- [权利要求 63] 一种试剂盒，所述试剂盒包含权利要求1-43中任一项所述的组合物以及用于放置所述组合物的容器和/或用法说明书。