一种靶向编辑 RNA 的新方法

技术领域

本申请属于基因编辑治疗领域，具体地，本申请创造了一种名为 CUSPER(C to U Specific Programmable Editing of RNA)的靶向编辑 RNA 的方法，其包括利用 CUSPER 技术进行 RNA 上由 C 到 U 碱基的精准位点编辑，可用于治疗由于 T 到 C 突变导致的疾病。

背景技术

近年来，以 CRISPR(Clustered regularly interspaced short palindromic repeats)为首的基因组编辑技术正在飞速发展，并对生物以及医学诸多领域产生了深远的影响。许多科研工作者和生物技术公司也在致力于将该技术推上临床。2019 年 9 月，北京大学邓宏魁教授与其合作者发表文章 11 首次报道了利用 CRISPR 技术编辑干细胞并将其回输至患者，以治疗其艾滋病和白血病的临床试验结果，为 CRISPR 技术在基因治疗方向的转化做出了巨大的贡献。

尽管 CRISPR 技术存在极大的应用前景，但该技术也存在一系列缺陷，导致该技术从科研阶段向临床治疗应用的转化步履维艰。问题之一便是 CRISPR 技术用到的核心作用酶：Cas9。基于 CRISPR 的 DNA 编辑技术，必须外源表达 Cas9 或拥有相似功能的其它核酸酶，从而造成了以下几个问题。首先，需要外源表达的核酸酶通常具有较大的分子量，这使得通过病毒载体将其递送至体内的效率急剧下降。其次，由于核酸酶的外源表达，使这种方法存在潜在的核酸酶脱靶可能，这将导致其应用中的潜在致癌风险。最后，外源表达的 Cas9 等类似核酸酶是从细菌中发现的，而非人类或哺乳动物天然存在的，这使其可能引起患者体内的免疫反应，这一方面可能会对患者自身造成损伤，另一方面也可能会使外源表达的核酸酶被中和，从而失去应有的活性，影响治疗效果。

2017 年，麻省理工学院张锋教授及其课题组报道了一种名为 REPAIR (RNA Editing for Programmable A to I Replacement)的 RNA 编辑技术，通

过外源表达Cas13-ADAR融合蛋白及单向导RNA (single guide RNA, sgRNA)可以实现靶向目标RNA的A到I的编辑2，但是该方法同CRISPR技术一样，仍需要外源蛋白的表达。无法解决外源蛋白表达造成的问题。

2019年1月，Thorsten Stafforst课题组报道了名为RESTORE (recruiting endogenous ADAR to specific trans for oligonucleotide-mediated RNA editing, Merkle et al., 2019)的RNA单碱基编辑技术。RESTORE能够摆脱对外源蛋白的依赖，但RESTORE技术需要在IFN-γ存在的前提下才能有较高的编辑效率，而IFN-γ是决定自体免疫发展和严重程度的关键因子9，这使得该技术在医学领域的应用大打折扣。另一方面，RESTORE技术中同样也用到一段向导RNA，而其使用的向导RNA是化学合成的寡核苷酸，并且其合成的寡核苷酸需要人为引入大量的化学修饰以保证其稳定性。在这些化学修饰中，有一部分是非天然的修饰，使得该寡核苷酸可能存在毒性或免疫原性；还有一部分修饰会导致相同碱基链的不同构象，使得对于相同的RNA序列，可能有数十种不同的构象组合，这增加了其向细胞内递送的难度。

2019年7月，北京大学魏文胜教授课题组在Nature Biotechnology上发表的文章[4]中，首次报道了一种核酸编辑技术：LEAPER (Leveraging Endogenous ADAR for Programmable Editing of RNA)。与CRISPR技术不同的是，一方面该技术从原理上摆脱了对外源核酸酶过表达的依赖，并且可以通过化学合成RNA完成，也可以通过腺相关病毒(AAV)、慢病毒等载体递送至患者发挥功能，这使得其递送手段的选择上更加灵活多变，使得该技术在向医学领域转化的过程中，具有更大的优势；另一方面该技术只能实现腺苷A到肌酐I（肌酐I在蛋白质翻译过程中会被识别为鸟酐G）的编辑，使得对其他突变，例如T到C的突变无能为力。此外，与CRISPR技术类似，该技术同样需要一段RNA作为向导，以便将内源的核酸酶招募至所需编辑的位点。该段向导RNA被命名为arRNA (adar-recruiting RNA)。

2019年7月，张锋教授课题组报道了一种名为RESCUE（RNA Editing for Specific C to U Exchange)的新技术[1]。该技术基于2017年该课题组报道的Cas13-ADAR基本骨架，在负责反应的ADAR催化结构域上做了不同的突变尝试。最终将ADAR催化结构域的A到I编辑活性修改为C到U编辑活性，从而实现了特定位点RNA上C到U的编辑，进一步拓展了对碱基

的精准编辑范围。然而，该技术仍需要外源表达 Cas13 与 ADAR 突变后的融合蛋白，无法解决细菌源蛋白表达造成的问题。

发明内容

为了解决上述基因编辑技术中的问题，以便将基因编辑技术更好地应用于医学领域，迫切地需要找到一种易于递送且可高效精准纠正 T 至 C 突变的定向基因编辑技术。

本申请创造了一种新的 RNA 编辑技术 CUSPER (C to U Specific Programmable Editing of RNA)，该技术无需依赖细菌来源的 Cas13b 的表达，并且将 RNA 编辑的应用范围从 A 到 I 的编辑拓展到 C 到 U 的编辑。

在本申请中，一方面由于无需依赖细菌来源的大分子蛋白的表达，因此减少了潜在的免疫系统风险，和因免疫系统对外源蛋白的攻击而对系统编辑效率带来的影响；另一方面，随着需要引入的蛋白分子量的显著缩减，其递送方式也变得更加灵活，可以包括化学转化及生物学递送，例如 AAV 递送等。因此，本申请提供的技术方案，不仅解决了 C 到 U 的单碱基编辑技术问题，而且提高了编辑系统的安全性、稳定性和使用的灵活性，更加利于体内应用，更具生物医药领域的应用前景。

具体地，本申请涉及：

1. 一种用于 RNA 编辑的工程化组合物或系统，包含：

1） 经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体，其中，所述腺苷脱氨酶蛋白或其催化结构域经过修饰具有催化胞苷脱氨基的活性，和

2） 将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到靶标 RNA 的 arRNA 或包含该 arRNA 或包含其编码序列的构建体；

其中所述 arRNA 包含与所述靶标 RNA 杂交的互补 RNA 序列，并且所述 arRNA 募集所述腺苷脱氨酶蛋白或其催化结构域至靶标 RNA 致使靶标 RNA 中的靶标胞苷脱氨基。

在一些实施方案中，所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体和包含所述 arRNA 的编码序列的构建体为同一构建体。在一些实施方案中，所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体和包

含所述 arRNA 的编码序列的构建体为分开的构建体。

在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如 ADAR2 蛋白）经过缺失、添加或取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如 ADAR2 蛋白）或其催化结构域是经过取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含如下突变修饰： E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含以其它 ADAR2 同源蛋白作为参考序列并具有 E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T 相应突变的蛋白。

在一些实施方案中，本申请用于 RNA 编辑的工程化组合物或系统包含上述经修饰的腺苷脱氨酶蛋白的催化结构域。在一些实施方案中，所述催化结构域为上述经修饰的腺苷脱氨酶蛋白的催化结构域。

2. 项 1 的工程化组合物或系统，其中所述腺苷脱氨酶在一个或多个位点发生突变修饰从而具有对胞苷脱氨使其转化成尿苷的活性。

3. 项 2 的工程化组合物或系统，其中所述腺苷脱氨酶蛋白或其催化结构域为 ADAR2 蛋白或其同源蛋白或其催化结构域。

4. 项 3 中的工程化组合物或系统，其中所述突变修饰包含： E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。

在一些实施方案中，所述经修饰的腺苷脱氨酶蛋白或其催化结构域为 NP_001103.1 蛋白的同源蛋白，具有与如下突变相应的突变：E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T。

5. 项 1-4 中任一项的工程化组合物或系统，其中所述 arRNA 与靶标 RNA 杂交时，其与靶标胞苷相对的靶向碱基为 A、U、C、或 G。

6. 项 5 中所述的工程化组合物或系统，其中所述靶向碱基为 U 或 C。

7. 项 1-6 中任一项的工程化组合物或系统，其中：

所述 arRNA 在对应于靶标 RNA 的靶标碱基的上游、下游、或上游和下游的一个或多个位置包含非配对核苷酸，以形成和靶标碱基上游、下游、或上游和下游的一个或多个位置的核苷酸错配。

8. 项 7 中的工程化组合物或系统，其中所述 arRNA 靶向碱基的 3’最近邻碱基与靶标 RNA 形成错配。

9. 项 8 中所述的工程化组合物或系统，其中所述 arRNA 与靶标 RNA 杂交时，所述靶向碱基的 3’最近邻碱基与靶标 RNA 形成 G-G 错配。

10. 项 1-9 中任一项的工程化组合物或系统，所述 arRNA 与靶标 RNA 杂交时，所述靶向碱基的 5’最近邻碱基与靶标 RNA 不形成错配。

11. 项 10 中的工程化组合物或系统，其中所述靶向碱基的 5’最近邻碱基为 U。

12. 项 9 中的工程化组合物或系统，其中当所述 arRNA 与靶标 RNA 杂交时，与所述 arRNA 的靶向碱基的 5’最近邻碱基相对的靶标 RNA 中的碱基优选次序从高到低为 G 或 C、U 或 A（G≈C>U≈A）。

13. 项 1-6 中任一项的工程化组合物或系统，其中当所述 arRNA 与靶标 RNA 杂交时，所述靶标碱基及其 5’和 3’相邻碱基形成的靶标碱基三联体仅在靶标碱基处形成错配，其中所述靶标碱基三联体选自：ACG、ACC、UCC、UCG、CCC、CCG、UCA、UCU。

14. 项 1-13 任一项的工程化组合物或系统，其中所述 arRNA 长度>50 nt、>55nt、>60nt、>65nt、>70nt、>75nt、>80nt、>85nt、>90nt、>95nt、>100nt、>105nt、>110nt、>115nt、>120nt。在一些实施方案中，所述 arRNA 长约 151-53 nt、131-61 nt、121-61 nt、111-65、101-71 nt、91-71 nt、81-71 nt。在一些具体实施方案中，所述 arRNA 的长度是该项定义的长度范围内的任一正整数。

15. 项 1-14 任一项的工程化组合物或系统，其中所述 arRNA 中靶向碱基距离 3’端及 5’端的长度相等。

16. 项 1-14 中任一项的组合物，其中所述 arRNA 中靶向碱基距离 3’端的长度为 45-5nt，40-5nt，35-10nt，25nt-15nt，24nt-11nt。在一些具体实

施方案中，所述 arRNA 的长度是该项定义的长度范围内的任一正整数。

17. 项 1-14 任一项的工程化组合物或系统，其中所述 arRNA 中靶向碱基距离 5’端的长度为 80-30nt，70-35nt，60-40nt，55nt-35nt，55nt-45nt。在一些具体实施方案中，所述 arRNA 的长度是该项定义的长度范围内的任一正整数。

18. 项 1-17 任一项的工程化组合物或系统，其中所述 arRNA 是化学修饰的。

19. 项 18 的工程化组合物或系统，其中所述化学修饰包括 2’-O-甲基修饰或核苷酸间 3’硫代修饰。

20. 一种使细胞中靶标 RNA 中的靶标胞嘧啶脱氨基的方法，包括将如下 1）和 2）导入所述细胞：

1） 经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体，和

2） 将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到所述靶标 RNA 的 arRNA 或包含该 arRNA 或包含其编码序列的构建体，其中，所述腺苷脱氨酶蛋白或其催化结构域经过修饰具有催化胞苷脱氨基的活性，所述 arRNA 包含与所述靶标 RNA 杂交的互补 RNA 序列，并且所述 arRNA 募集所述腺苷脱氨酶蛋白或其催化结构域至靶标 RNA 致使靶标 RNA 中的靶标胞苷脱氨基。

在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如 ADAR2 蛋白）经过缺失、添加或取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如 ADAR2 蛋白）或其催化结构域是经过取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含如下突变修饰： E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含以其它 ADAR2 蛋白作为参考序列并具有 E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418

E/S661T 相应突变的蛋白。

在一些实施方案中，本申请用于 RNA 编辑的工程化组合物或系统包含上述经修饰的腺苷脱氨酶蛋白的催化结构域。

21. 项 20 的方法，其中，所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体和包含所述 arRNA 的编码序列的构建体为同一构建体，被同时导入所述细胞，或者所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体和包含所述 arRNA 的编码序列的构建体为分开的构建体，所述分开的构建体被同时或分开地导入所述细胞。

22. 项 20 或 21 的方法，其中所述腺苷脱氨酶在一个或多个位点发生突变修饰从而具有对胞苷脱氨使其转化成尿苷的活性。

23. 项 22 的方法，其中所述腺苷脱氨酶蛋白或其催化结构域为 ADAR 2 蛋白或其同源蛋白或其催化结构域。

24. 项 23 中的方法，其中所述突变修饰包含：E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。

在一些实施方案中，所述经修饰的腺苷脱氨酶蛋白或其催化结构域为 NP_001103.1 蛋白的同源蛋白，具有与如下突变相应的突变：E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T。

25. 项 21-24 中任一项方法，其中所述 arRNA 与靶标 RNA 杂交时，其与靶标胞苷相对的靶向碱基为 A、U、C、或 G。

26. 项 25 中所述的方法，其中所述靶向碱基为 U 或 C。

27. 项 21-26 中任一项的方法，其中：

所述 arRNA 在对应于靶标 RNA 的靶标碱基的上游、下游、或上游和下游的一个或多个位置包含非配对核苷酸，以形成和靶标碱基上游、下游、或上游和下游的一个或多个位置的核苷酸错配。

28. 项 27 中的方法，其中所述 arRNA 靶向碱基的 3’最近邻碱基与靶标 RNA 形成错配。

29. 项 28 中所述的方法，其中所述 arRNA 与靶标 RNA 杂交时，所述

靶向碱基的3’最近邻碱基与靶标RNA形成G-G错配。

30. 项21-29中任一项的方法，所述arRNA与靶标RNA杂交时，所述靶向碱基的5’最近邻碱基与靶标RNA不形成错配。

31. 项30中的方法，其中所述靶向碱基的5’最近邻碱基为U。

32. 项29中的方法，其中当所述arRNA与靶标RNA杂交时，与所述arRNA的靶向碱基的5’最近邻碱基相对的靶标RNA中的碱基优选次序从高到低为G或C、U或A（G≈C>U≈A）。

33. 项21-26中任一项的方法，其中当所述arRNA与靶标RNA杂交时，所述靶标碱基及其5’和3’相邻碱基形成的靶标碱基三联体仅在靶标碱基处形成错配，其中所述靶标碱基三联体选自：ACG、ACC、UCC、UCG、CCC、CCG、UCA、UCU。

34. 项21-33任一项的方法，其中所述arRNA长度>50nt、>55nt、>60nt、>65nt、>70nt、>75nt、>80nt、>85nt、>90nt、>95nt、>100nt、>105nt、>110nt、>115nt、>120nt。在一些实施方案中，所述arRNA长约151-53 nt、131-61 nt、121-61 nt、111-65、101-71 nt、91-71 nt、81-71 nt。在一些具体实施方案中，所述arRNA的长度是该项定义的长度范围内的任一正整数。

35. 项21-34任一项的方法，其中所述arRNA中靶向碱基距离3’端及5’端的长度相等。

36. 项21-34中任一项的方法，其中所述arRNA中靶向碱基距离3’端的长度为45-5nt，40-5nt，35-10nt，25nt-15nt，24nt-11nt。在一些具体实施方案中，所述arRNA的长度是该项定义的长度范围内的任一正整数。

37. 项21-34任一项的方法，其中所述arRNA中靶向碱基距离5’端的长度为80-30nt，70-35nt，60-40nt，55nt-35nt，55nt-45nt。在一些具体实施方案中，所述arRNA的长度是该项定义的长度范围内的任一正整数。

38. 项21-37任一项的方法，其中所述arRNA是化学修饰的。

39. 项38的方法，其中所述化学修饰包括2’-O-甲基修饰或核苷酸间3’硫代修饰。

40. 项21-39中任一项的方法，其中所述细胞为哺乳动物细胞。

41. 一种治疗由T到C突变引起的疾病的方法，包括使用如项21-39中任一项的方法使包含所述T到C突变转录形成的信使RNA中靶标碱基C

脱氨基，以校正所述突变。

42. 一种经修饰的腺苷脱氨酶蛋白，其中所述腺苷脱氨酶蛋白为 ADAR2，其包含 E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T 突变修饰，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致，所述 ADAR2 蛋白经过所述突变修饰而具有催化胞苷脱氨基的活性。

在一些实施方案中，所述经修饰的腺苷脱氨酶蛋白或其催化结构域为 NP_001103.1 蛋白的同源蛋白，具有与如下突变相应的突变：E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T。

43. 如项 42 中所述的经修饰的腺苷脱氨酶蛋白用于催化胞苷脱氨基转化为尿苷的用途。

附图说明

图 1 显示了 BFP 报告体系以及待进行 C 到 U 编辑的靶标碱基。

图 2 显示了当靶向为点为 U 时，CUSPER 和 RESCUE 技术设计向导 RNA 时靶标碱基的选择以及靶向为点相邻碱基的设计原则。

图 3 显示了对 CUSPER 编辑体系的测试。图中“/”代表未添加对应质粒或 arRNA，仅添加相同体积水。

图 4 显示了 LEAPER 技术对靶标碱基相邻上下游碱基的偏好性。图中箭头所示为与图 3 对应的当靶标碱基的 3’相邻碱基为 A 时，LEAPER 技术对 5’相邻上游的偏好性。

图 5 显示了 RESCUE 技术对靶标碱基相邻上下游碱基的偏好性。图中箭头所示为与图 3 对应的当 3’相邻碱基为 A 时，RESCUE 技术对 5’相邻碱基的偏好性。

图 6 显示了对 CUSPER 编辑体系的重复测试结果。图中“/”代表未添加对应质粒或 arRNA 而仅添加相同体积水。

图 7 显示靶标 C 对应 U，及其相邻碱基无错配的情况下，mRNA 与对应 arRNA 序列。

图 8 显示靶标 C 对应 U，及其相邻碱基无错配的情况下全部 16 种 3’和

5’相邻碱基组合的 C 到 U RNA 编辑效率

发明详述

为解决现有基因编辑技术中普遍依赖异种外源蛋白的困境，在更多种类的碱基上实现精准的单个碱基编辑，本申请创造了一种 RNA 编辑技术 CUSPER(C to U Specific Programmable Editing of RNA)。CUSPER 将 RNA 编辑的应用范围从 A 到 I 的编辑拓展到 C 到 U 的编辑，且由于使用了哺乳动物本身表达的酶蛋白进行编辑，避免了向细胞内引入异种外源蛋白，从而提高了安全性，并使整个基因编辑过程更加高效便利。

定义

如本文所用，“CUSPER 技术”是本申请技术方案中的原创技术，其中的 CUSPER 是“C to U Specific Programmable Editing of RNA”的缩写，即“将胞苷 C 转变为尿苷 U 的特异性可编程 RNA 编辑”。该技术使用一段可与靶标 RNA 互补杂交的短 RNA，将经修饰的腺苷脱氨酶蛋白或包含其催化结构域的蛋白募集到靶标 RNA，以对靶标胞苷脱氨基，使其转变为尿苷。其中所述经修饰的腺苷脱氨酶蛋白或包含其催化结构域的蛋白经修饰而具有催化胞嘧啶脱氨基的活性。所述可与靶标 RNA 互补杂交的短 RNA 即 arRNA。本申请中所使用的“arRNA(adar-recruiting RNA)”是指可以招募腺苷脱氨酶蛋白或包含其催化结构域至靶标 RNA 的单链 RNA，该 arRNA 募集所述腺苷脱氨酶蛋白或其催化结构域至靶标RNA致使靶标RNA中的靶标胞苷脱氨基。如本文所用，核酸的“互补”是指一条核酸通过传统的 Watson-Crick 碱基配对与另一条核酸形成氢键的能力。百分比互补性表示核酸分子中可与另一核酸分子形成氢键(即，Watson-Crick 碱基配对)的残基的百分比(例如，10 个中的约 5、6、7、8、9、10 个分别为约 50%，60%，70%，80%，90%和 100%互补)。“完全互补”是指核酸序列的所有连续残基与第二核酸序列中相同数量的连续残基形成氢键。如本文所用，“基本上互补”是指在约 40、50、60、70、80、100、150、200、250 或更多个核苷酸的区域内，至少约 70%，75%，80%，85%，90%，95%，97%，98%，99%或 100%中的任何一个的互补程度，或指在严格条件下杂交的两条核酸。对于单个碱基或单个核苷酸，按照

Watson-Crick 碱基配对原则，A 与 T 或 U、C 与 G 或 I 配对时，被称为互补或匹配，反之亦然；而除此以外的碱基配对都称为不互补或不匹配。

“杂交”是指其中一种或多种多核苷酸反应形成复合物的反应，所述复合物通过核苷酸残基的碱基之间的氢键稳定。所述氢键可以通过 Watson Crick 碱基配对，Hoogstein 结合或以任何其他序列特异性的方式发生。能够与给定序列杂交的序列称为给定序列的“互补序列”。

如本文所用，术语“递送”是指将核酸、蛋白等生物大分子通过某些途径从细胞膜外引入细胞膜内。所述“递送”例如电转染、脂质体转染、脂质-纳米颗粒递送、病毒递送、外泌体递送等方式。

如本文所用，术语“靶标 RNA”是指待编辑的目标 RNA，其包含待编辑的胞苷。所述靶标 RNA 可以是成熟 mRNA 或 mRNA 前体。所述待编辑的胞苷被称为“靶标碱基”、“靶标胞苷”或“靶标 C”。在靶标 RNA 的 5’端与靶标胞苷相邻的碱基称为“5’相邻碱基”；在靶标 RNA 的 3’端与靶标胞苷相邻的碱基称为“3’相邻碱基”；靶标碱基及其 3’和 5’ 相邻碱基组成的碱基三联体在本文中被称为“靶标碱基三联体”。当 arRNA 与靶标 RNA 杂交时，在 arRNA 上与所述靶标碱基相对的碱基称为“靶向碱基”，在 arRNA 的 5’端与靶向碱基相邻的碱基称为“5’最近邻碱基”；在 arRNA 的 3’端与靶向碱基相邻的碱基称为“3’最近邻碱基”；靶向碱基及其 3’和 5’ 最近邻碱基组成的碱基三联体在本文中被称为“靶向碱基三联体”。

在本文中，靶向碱基距离 3’端的长度是指靶向碱基的 3’最近邻碱基至 3’最末端碱基的所有碱基个数；所述靶向碱基距离 5’端的长度是指靶向碱基的 5’最近邻碱基至 5’最末端碱基的所有碱基个数。

如本文所用，术语“腺苷脱氨酶（Adenosine to inosine acting on RNA enzyme，ADAR)”是指是一类在真核生物（包括人等哺乳动物）各组织中广泛表达的腺苷脱氨酶，能够催化 RNA 分子中腺苷 A 到肌苷 I 的转换。

在本申请中，E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T 突变是指发生于 ADAR2 蛋白的一系列突变，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致，即使用 NP_001103.1 作为参考序列。本领域技术人员应当知晓，对于不同的 ADAR2 蛋白的参考序列，所述突变中的氨基酸

编号可能发生变化。因此，如本申请所用，E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T 突变包括不同ADAR2蛋白参考序列中与所述突变氨基酸位置对应的突变，所述 ADAR2 蛋白经过所述突变修饰而具有催化胞苷脱氨基的活性。相应的，本申请提供的经修饰的腺苷脱氨酶蛋白，包括以 NP_001103.1 作为参考序列并包含 E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T 突变修饰的 ADAR2，也涵盖以不同 ADAR2 蛋白作为参考序列并包含相应突变修饰的 ADAR2。

如本文所用，术语“构建体”是指包含某一核酸序列的核酸载体，所述核酸载体可以是线性核酸分子、质粒或病毒载体等。所述核酸分子可以是单链也可以是双链分子。所述特定的核酸序列可以是 DNA 序列也可以是 RNA 序列。在一些实施方案中，所述核酸序列不经过转录、翻译或表达而直接发挥其功能。在一些实施方案中，所述核酸序列为 DNA 序列，其经过转录形成 RNA 后以 RNA 分子形式发挥功能。在一些实施方案中，所述核酸序列为 RNA，其经过翻译后以多肽或蛋白质的形式发挥作用。在一些实施方案中，所述核酸序列为 DNA，其经过转录和翻译步骤形成蛋白质后以蛋白质的形式发挥功能。所述构建体可以通过包装为病毒、脂质纳米颗粒或外泌体等形式进入细胞，亦可以通过电转化、显微注射、化学转化等方式进入细胞。

本申请所用术语“修饰”是指通过化学方法例如基因工程方法改变核酸或蛋白的组成或结构，从而使所述核酸或蛋白的一项或多项特性或功能发生改变。例如，在本申请中，腺苷脱氨酶蛋白或其催化结构域经过修饰，例如添加、缺失和/或突变一个或多个氨基酸后具有催化胞苷脱氨基的作用。

工程化组合物或系统

本申请提供了一种用于 RNA 编辑的工程化组合物或系统，其包含：

1） 经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体，其中，所述腺苷脱氨酶蛋白或其催化结构域经过修饰具有催化胞苷脱氨基的活性，和

2） 将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到靶标 RNA

的 arRNA 或包含该 arRNA 或包含其编码序列的构建体；

其中所述 arRNA 包含与所述靶标 RNA 杂交的互补 RNA 序列，并且所述 arRNA 募集所述腺苷脱氨酶蛋白或其催化结构域至靶标 RNA 致使靶标 RNA 中的靶标胞苷脱氨基。

在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如 ADAR2 蛋白）经过缺失、添加或取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如 ADAR2 蛋白）或其催化结构域是经过取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含如下突变修饰： E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含以其它 ADAR2 同源蛋白作为参考序列并具有 E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T 相应突变的蛋白。

在一些实施方案中，本申请用于 RNA 编辑的工程化组合物或系统包含上述经修饰的腺苷脱氨酶蛋白或其催化结构域。在一些实施方案中，所述催化结构域为上述经修饰的腺苷脱氨酶蛋白的催化结构域。

在一些实施方案中，所述腺苷脱氨酶在一个或多个位点发生突变修饰从而具有对胞苷脱氨使其转化成尿苷的活性。在一些实施方案中，所述腺苷脱氨酶蛋白或其催化结构域为 ADAR2 蛋白或其同源蛋白或所述 ADAR2 蛋白的催化结构域或所述 ADAR2 蛋白同源蛋白的催化结构域。在一些实施方案中，所述突变修饰包含：E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。

在一些实施方案中，所述经修饰的腺苷脱氨酶蛋白或其催化结构域通过构建体导入细胞进行表达，所述构建体选自线性核酸、质粒及病毒载体中的任意。在一些实施方案中，所述细胞为真核细胞。在一些实施方案中，所述细胞为哺乳动物细胞。在一些实施方案中，所述细胞为人细胞。

在一些实施方案中，所述 arRNA 与靶标 RNA 杂交时，其与靶标胞苷相对的靶向碱基为 A、U、C、或 G。在一些实施方案中，优选的靶向碱基顺序为 U＞C＞A≈G，即，当多条 arRNA 序列相比，其中除靶向碱基以外其他的所有碱基均相同时，靶向碱基为 U 或 C 的 arRNA 通常具有更高的编辑效率。因此优选地，在一些实施方案中，所述靶向碱基为 U 或 C。

在一些实施方案中，所述 arRNA 在对应于靶标 RNA 的靶标碱基的上游、下游、或上游和下游的一个或多个位置包含非配对核苷酸，以形成和靶标碱基上游、下游、或上游和下游的一个或多个位置的核苷酸错配。在一些实施方案中，所述 arRNA 靶向碱基的 3’最近邻碱基与靶标 RNA 形成错配。在一些实施方案中，所述 arRNA 与靶标 RNA 杂交时，所述靶向碱基的 3’最近邻碱基与靶标 RNA 形成 G-G 错配。在一些实施方案中，所述 arRNA 与靶标 RNA 杂交时，所述靶向碱基的 5’最近邻碱基与靶标 RNA 不形成错配。在一些实施方案中，所述靶向碱基的 5’最近邻碱基与靶标 RNA 不形成错配，且其中所述靶向碱基的 5’最近邻碱基为 U。在一些实施方案中，当所述 arRNA 与靶标 RNA 杂交时，与所述 arRNA 的靶向碱基的 5’最近邻碱基相对的靶标 RNA 中的碱基优选次序从高到低为 G 或 C、U 或 A。在一些实施方案中，当所述 arRNA 与靶标 RNA 互补杂交时，所述 arRNA 靶向三联体以外的一个或多个碱基与靶标 RNA 形成错配。并且在一些实施方案中，所述错配可以进一步提高基于所述 arRNA 的靶向编辑效率。

在一些实施方案中，当所述 arRNA 与靶标 RNA 杂交时，所述靶标碱基及其 5’和 3’相邻碱基形成的靶标碱基三联体仅在靶标碱基处形成错配，并且，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度相等，此时，优选地其中所述靶标碱基三联体选自：ACG、ACC、UCC、UCG、CCC、CCG、UCA、UCU。在一些实施方案中，当所述 arRNA 与靶标 RNA 杂交时，所述靶标碱基及其 5’和 3’相邻碱基形成的靶标碱基三联体仅在靶标碱基处形成错配，并且，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度不相等，此时，其中所述靶标碱基三联体选自：ACG、ACC、UCC、UCG、CCC、CCG、UCA、UCU。

在一些实施方案中，所述 arRNA 长度>50nt、>55nt、>60nt、>65nt、>70nt、>75nt、>80nt、>85nt、>90nt、>95nt、>100nt、>105nt、>110nt、>115n

t、>120nt。在一些实施方案中，所述 arRNA 长约 151-53 nt、131-61 nt、121-61 nt、111-65、101-71 nt、91-71 nt、81-71 nt。在一些具体实施方案中，所述 arRNA 的长度是该项定义的长度范围内的任一正整数。

在一些实施方案中，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度相等。在一些实施方案中，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度不相等。在一些实施方案中，所述 arRNA 中靶向碱基距离 3’端的长度为 45-5nt，40-5nt，35-10nt，25nt-15nt，24nt-11nt。在一些具体实施方案中，所述长度选自本项定义的长度范围内的任一正整数。

在一些实施方案中，所述 arRNA 中靶向碱基距离 5’端的长度为 80-30nt，70-35nt，60-40nt，55nt-35nt，55nt-45nt。在一些具体实施方案中，所述长度选自本项定义的长度范围内的任一正整数。在一些实施方案中，所述 arRNA 中靶向碱基距离 5’端的长度大于 80。

在一些实施方案中，所述 arRNA 由化学合成。在一些实施方案中，所述 arRNA 为寡核苷酸。在一些实施方案中，所述 arRNA 是化学修饰的。在一些实施方案中，所述化学修饰包括 2’-O-甲基修饰或核苷酸间 3’硫代修饰。在一些实施方案中，所述化学修饰选自如下的一项或多项：

序列前 3 个和后 3 个核苷酸分别被 2`-OMe 修饰，

前 3 个和后 3 个核苷酸间连接均为硫代磷酸酯键连接，

序列中全部 U 均被 2`-OMe 修饰，

靶向碱基的 3’最近邻碱基为 2`-OMe 修饰的 A，

靶向碱基的 5’最近邻碱基为 2`-OMe 修饰的 C，

靶向碱基与其 3’最近邻碱基和 5’最近邻碱基分别以硫代磷酸酯键连接，

前 5 个和后 5 个核苷酸分别被 2`-OMe 修饰，和

前 5 个和后 5 个核苷酸间连接为硫代磷酸酯键连接。

在一些实施方案中，所述 arRNA 由一种构建体编码并转录生成。在一些实施方案中，所述构建体选自线性核酸链、病毒载体或质粒。

编辑 RNA 的方法

本申请提供了一种使细胞中靶标 RNA 中的靶标胞嘧啶脱氨基的方法——CUSPER（程序化的 C 到 U 的 RNA 编辑），包括将如下 1）和 2）导入所

述细胞：

1） 经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体，和

2） 将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到所述靶标RNA的arRNA或包含该arRNA或包含其编码序列的构建体。其中，所述腺苷脱氨酶蛋白或其催化结构域经过修饰具有催化胞苷脱氨基的活性，所述arRNA包含与所述靶标RNA杂交的互补RNA序列，并且所述arRNA募集所述腺苷脱氨酶蛋白或其催化结构域至靶标RNA致使靶标RNA中的靶标胞苷脱氨基。

在一些实施方案中，所述细胞为真核细胞。在一些实施方案中，所述细胞为哺乳动物细胞。在一些实施方案中，所述细胞为人细胞。在一些实施方案中，所述细胞为小鼠细胞。

在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如ADAR2蛋白）经过缺失、添加或取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请的经修饰的腺苷脱氨酶蛋白是腺苷脱氨酶蛋白（例如ADAR2蛋白）或其催化结构域是经过取代一个或多个氨基酸而具有胞苷脱氨基的活性的蛋白质。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含如下突变修饰： E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与NP_001103.1中的氨基酸编号一致。在一些实施方案中，本申请经修饰的腺苷脱氨酶蛋白包含以其它ADAR2蛋白作为参考序列并具有E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T相应突变的蛋白。

本领域技术人员应当知晓，所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体包含所述经修饰的腺苷脱氨酶蛋白或其催化结构域的编码序列。在一些实施方案中，所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体与包含所述arRNA的编码序列的构建体为同一构建体。在一些实施方案中，所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的编码序列的构建体与包含所述arRNA的编码序列的构建体为不同构建体，即所述经

修饰的腺苷脱氨酶蛋白或其催化结构域的编码序列与所述 arRNA 的编码序列分别位于不同构建体上。在一些实施方案中，所述不同构建体为两个或多于两个。在一些实施方案中，所述不同的构建体同时或分开地导入所述细胞。应当理解，本段所说的不同构建体指的是非同一个构建体，并不代表所述不同构建体分别属于不同的构建体类别。在一些实施方案中，将所述表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体与所述将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到所述靶标 RNA 的 arRNA 或包含该 arRNA 的构建体或包含该 arRNA 的编码序列的构建体导入同一细胞。在一些实施方案中，将经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体，和多种 arRNA 或包含所述 arRNA 的构建体或包含所述 arRNA 的编码序列的构建体导入细胞，以实现对靶标 RNA 的高通量编辑，其中所述的多种 arRNA 为靶向不同靶标 RNA 的 arRNA，或为靶向同一靶标 RNA 不同靶标碱基（例如 C）位点的 arRNA。

因此，本申请还涵盖一种高通量编辑细胞中靶标 RNA 中的靶标胞嘧啶的方法，包括将如下 1）和 2）导入所述细胞：

1） 经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体，和

将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到所述靶标 RNA 的多种（例如 2 种以上、5 种以上、5 种以上、10 种以上）arRNA 或包含该 arRNA 或包含其编码序列的构建体，其中，所述腺苷脱氨酶蛋白或其催化结构域经过修饰具有催化胞苷脱氨基的活性，所述多种（例如 2 种以上、5 种以上、5 种以上、10 种以上）arRNA 包含分别与所述靶标 RNA 杂交的互补 RNA 序列，并且所述 arRNA 募集所述腺苷脱氨酶蛋白或其催化结构域至靶标 RNA 致使靶标 RNA 中的靶标胞苷脱氨基。

在一些实施方案中，所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体和包含所述 arRNA 的编码序列的构建体为同一构建体，或者所述表达经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体和包含所述 arRNA 的编码序列的构建体为分开的构建体，所述分开的构建体被同时或分开地导入所述细胞。

在一些实施方案中，所述腺苷脱氨酶在一个或多个位点发生突变修饰从

而具有对胞苷脱氨使其转化成尿苷的活性。

在一些实施方案中，所述腺苷脱氨酶蛋白或其催化结构域为 ADAR2 蛋白或其同源蛋白或其催化结构域。

在一些实施方案中，所述突变修饰包含： E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。在一些实施方案中，所述构建体选自病毒载体、质粒和线性核酸。在一些实施方案中，所述经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体以及包含将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到所述靶标 RNA 的 arRNA 的寡核苷酸或转录该 arRNA 的构建体通过病毒侵染、化学转染、电转染、外泌体递送、或纳米脂质颗粒递送等方式被导入所述细胞。

在一些实施方案中，所述腺苷脱氨酶在一个或多个位点发生突变修饰从而具有对胞苷脱氨使其转化成尿苷的活性。在一些实施方案中，所述腺苷脱氨酶蛋白或其催化结构域为 ADAR2 蛋白或其同源蛋白或其催化结构域。在一些实施方案中，所述突变修饰包含： E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。

在一些实施方案中，所述 arRNA 与靶标 RNA 杂交时，其与靶标胞苷相对的靶向碱基为 A、U、C、或 G。在一些实施方案中，所述 arRNA 与靶标 RNA 杂交时，其与靶标胞苷相对的靶向碱基优选为 U 或 C。

在一些实施方案中，所述 arRNA 在对应于靶标 RNA 的靶标碱基的上游、下游、或上游和下游的一个或多个位置包含非配对核苷酸，以形成和靶标碱基上游、下游、或上游和下游的一个或多个位置的核苷酸错配。在一些实施方案中，所述 arRNA 靶向碱基的 3’最近邻碱基与靶标 RNA 形成错配。在一些实施方案中，所述 arRNA 与靶标 RNA 杂交时，所述靶向碱基的 3’最近邻碱基与靶标 RNA 形成 G-G 错配。在一些实施方案中，所述 arRNA 与靶标 RNA 杂交时，所述靶向碱基的 5’最近邻碱基与靶标 RNA 不形成错配。在一些实施方案中，所述靶向碱基的 5’最近邻碱基为 U。在一些实施方案

中，当所述 arRNA 与靶标 RNA 杂交时，与所述 arRNA 的靶向碱基的 5’最近邻碱基相对的靶标 RNA 中的碱基优选次序从高到低为 G 或 C、U 或 A（G≈C>U≈A）。在一些实施方案中，当所述 arRNA 与靶标 RNA 杂交时，与所述 arRNA 的靶向碱基的 5’最近邻碱基相对的靶标 RNA 中的碱基优选为 G 或 C。在一些实施方案中，当所述 arRNA 与靶标 RNA 杂交时，与所述 arRNA 的靶向碱基的 5’最近邻碱基相对的靶标 RNA 中的碱基最优选为 G。在一些实施方案中，当所述 arRNA 与靶标 RNA 互补杂交时，所述 arRNA 靶向三联体以外的一个或多个碱基与靶标 RNA 形成错配。并且在一些实施方案中，所述错配可以进一步提高基于所述 arRNA 的靶向编辑效率。

在一些实施方案中，当所述 arRNA 与靶标 RNA 杂交时，所述靶标碱基及其 5’和 3’相邻碱基形成的靶标碱基三联体仅在靶标碱基处形成错配，并且，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度相等，此时，优选地其中所述靶标碱基三联体选自：ACG、ACC、UCC、UCG、CCC、CCG、UCA、UCU。在一些实施方案中，当所述 arRNA 与靶标 RNA 杂交时，所述靶标碱基及其 5’和 3’相邻碱基形成的靶标碱基三联体仅在靶标碱基处形成错配，并且，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度不相等，此时，其中所述靶标碱基三联体选自：ACG、ACC、UCC、UCG、CCC、CCG、UCA、UCU。

在一些实施方案中，所述 arRNA 长度>50nt、>55nt、>60nt、>65nt、>70nt、>75nt、>80nt、>85nt、>90nt、>95nt、>100nt、>105nt、>110nt、>115nt、>120nt。在一些实施方案中，所述 arRNA 长约 151-53 nt、131-61 nt、121-61 nt、111-65、101-71 nt、91-71 nt、81-71 nt。在一些具体实施方案中，所述 arRNA 的长度是该项定义的长度范围内的任一正整数。

在一些实施方案中，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度相等。在一些实施方案中，所述 arRNA 中靶向碱基距离 3’端及 5’端的长度不相等。在一些实施方案中，所述 arRNA 中靶向碱基距离 3’端的长度为 45-5nt，40-5nt，35-10nt，25nt-15nt，24nt-11nt。在一些具体实施方案中，所述长度为本项定义的长度范围内的任一正整数。

在一些实施方案中，所述 arRNA 中靶向碱基距离 5’端的长度为 80-30 nt，70-35nt，60-40nt，55nt-35nt，55nt-45nt。在一些具体实施方案中，所述

长度为本项定义的长度范围内的任一正整数。在一些实施方案中，所述 arRNA 中靶向碱基距离 5’端的长度大于 80。

在一些实施方案中，所述 arRNA 为寡核苷酸或包含于寡核苷酸。在一些实施方案中，所述寡核苷酸是化学修饰的。在一些实施方案中，所述化学修饰包括 2’-O-甲基修饰或核苷酸间 3’硫代修饰。在一些实施方案中，所述化学修饰选自如下的一项或多项：

序列前 3 个和后 3 个核苷酸分别被 2`-OMe 修饰，

前 3 个和后 3 个核苷酸间连接均为硫代磷酸酯键连接，

序列中全部 U 均被 2`-OMe 修饰，

靶向碱基的 3’最近邻碱基为 2`-OMe 修饰的 A，

靶向碱基的 5’最近邻碱基为 2`-OMe 修饰的 C，

靶向碱基与其 3’最近邻碱基和 5’最近邻碱基分别以硫代磷酸酯键连接，

前 5 个和后 5 个核苷酸分别被 2`-OMe 修饰，和

前 5 个和后 5 个核苷酸间连接为硫代磷酸酯键连接。

在一些实施方案中，所述 arRNA 由一种构建体编码并转录生成。在一些实施方案中，所述构建体选自线性核酸链、病毒载体或质粒。

RNA 编辑相关酶蛋白及其用途

本申请提供了一种经修饰的腺苷脱氨酶蛋白，其中所述腺苷脱氨酶蛋白为 ADAR2，其包含对应于 Genebank 登记号为 NP_001103.1 的 ADAR2 中如下氨基酸突变：E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T，或所述 NP_001103.1 的同源 ADAR2 蛋白相应位置的氨基酸突变，所述 ADAR2 蛋白经过所述突变修饰而具有催化胞苷脱氨基的活性。

本申请同时还提供了所述经修饰的腺苷脱氨酶蛋白用于催化胞苷脱氨基转化为尿苷的用途。在一些实施方案中，所述催化胞苷脱氨基转化为尿苷的用途发生于细胞内。在一些实施方案中，所述催化胞苷脱氨基转化为尿苷的用途发生于细胞外。

疾病治疗方法

本申请提供了一种治疗由T到C突变引起的疾病的方法，包括使用如前所述的编辑RNA的方法使包含所述T到C突变转录形成的信使RNA中靶标碱基C脱氨基，以校正所述突变。

在一些实施方案中，所述治疗由T到C突变引起的疾病的方法包括将如前所述的工程化组合物或系统注射入受试者体内。在一些实施方案中，所述治疗将包括将如下1）和2）注射入受试者体内。

1） 经修饰的腺苷脱氨酶蛋白或其催化结构域或表达该经修饰的腺苷脱氨酶蛋白或其催化结构域的构建体，和

2） 将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到所述靶标RNA的arRNA或包含该arRNA或包含其编码序列的构建体。其中，所述腺苷脱氨酶蛋白为如前所述的RNA编辑相关酶蛋白。

在一些实施方案中，所述注射为静脉注射、动脉灌注、肌肉注射、皮下注射或瘤内注射。

在一些实施方案中，所述由T到C突变引起的疾病包括遗传性疾病和癌症。

试剂盒及制剂

本申请还提供了一种试剂盒，其用于催化胞苷脱氨基转化为尿苷。在一些实施方案中，所述试剂盒包含如前所述的工程化组合物或系统。在一些实施方案中，所述试剂盒包含编码或表达所述工程化组合物系统中所述经修饰的腺苷脱氨酶蛋白或其催化结构域和/或将所述经修饰的腺苷脱氨酶蛋白或其催化结构域募集到靶标RNA的arRNA的构建体。

本申请提供的利用CUSPER技术对RNA进行靶向编辑的方法，具有以下优点：

一方面，新的技术招募编辑蛋白(例如腺苷脱氨酶（ADAR）)时，不像RESCUE技术，需要设计含Cas13b招募骨架的向导RNA，也不需要细菌来源的Cas13b过表达，减小了外源表达蛋白的长度使得通过病毒载体进行装载及人体内递送容易、多样，同时，也可以减小外源表达的核酸酶被中和导致

的基因编辑失败的可能性，这使得其在应用至医疗领域时比 RESCUE 技术具有显著优势。

另一方面，新的体系不像 LEAPER 技术，LEAPER 技术仅能实现 RNA 编辑的应用范围从 A 到 I 的编辑，而新的体系将该编辑拓宽至 C 到 U，这使得很多在基因组上呈现 T 到 C 突变的遗传病，以及其它需要将 C 转变为 T/U 的应用，可以用本申请的技术治疗。

因此，本申请的技术相比于现有技术而言，应用范围更广，更加安全有效。

以上详细描述了本发明的优选实施方式，但是，本发明并不限于此。在本发明的技术构思范围内，可以对本发明的技术方案进行多种简单变型，包括各个技术特征以任何其它的合适方式进行组合，这些简单变型和组合同样应当视为本发明所公开的内容，均属于本发明的保护范围。

下面结合具体实施例，对本发明的技术方案作进一步的详述，但本发明并不限于以下实施例。如未特别指出，以下所涉及的试剂均可通过商业途径购得。为简便起见，部分操作未详述操作的参数、步骤及所使用的仪器，应当理解，这些都是本领域技术人员所熟知并可重复再现的。

实施例

实施例 1：经修饰的 ADAR2 及 BFP 报告体系的分子构建

1. 突变型 ADAR2-r16-293T 的构建

参照参考文献 1 中报道的 RESCUE 技术对 ADAR2(RNA 腺苷脱氨酶 2)催化结构域进行诱变，突变位点与文献中 r16 相同(dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) r16, https://benchling.com/s/seq-19Ytwwh0i0vSIbyXYZ95)，其中所述突变修饰的氨基酸编号与 NP_001103.1 中的氨基酸编号一致。使用本领域常规 DNA 合成技术体外合成包含上述突变的 ADAR2 编码序列片段，并通过酶切连接插入 pLenti-ADAR2 质粒载体（pLenti-ADAR2 质粒骨架由魏文胜教授试验室惠赠）的 ADAR2 XmaI 酶切位点及 AscI 酶切位点之间。通过以上步骤构建的质粒命名为 pLenti-ADAR2-r16，其包含上述突变的 ADAR2 基因命名为 ADAR2-r16。ADAR2-r16 全长 cDNA 序列

为：SEQ ID NO 1。通过第二代慢病毒包装体系（pCAG-VSVG 由 Arthur Nienhuis & Patrick Salmon 惠赠(Addgene plasmid # 35616 ; http://n2t.net/addgene:35616 ; RRID:Addgene_35616); pCMVR8.74 由 Didier Trono 惠赠 (Addgene plasmid # 22036 ; http://n2t.net/addgene:22036; RRID:Addgene_22036)）将 pLenti-ADAR2-r16 包装成慢病毒，并使用所述慢病毒侵染 293T 细胞并于 48 小时后使用 10ug/ml 终浓度的 Blasticidin (Solarbio B9300)进行抗性筛选。筛选过后存活的细胞称为 ADAR2-r16-293T。

SEQ ID NO 1:

ATGGATATAGAAGATGAAGAAAACATGAGTTCCAGCAGCACTGATGTGAAGGAAAACCGC
AATCTGGACAACGTGTCCCCCAAGGATGGCAGCACACCTGGGCCTGGCGAGGGCTCTCAGCTC
TCCAATGGGGGTGGTGGTGGCCCCGGCAGAAAGCGGCCCCTGGAGGAGGGCAGCAATGGCCA
CTCCAAGTACCGCCTGAAGAAAAGGAGGAAAACACCAGGGCCCGTCCTCCCCAAGAACGCCCT
GATGCAGCTGAATGAGATCAAGCCTGGTTTGCAGTACACACTCCTGTCCCAGACTGGGCCCGTG
CACGCGCCTTTGTTTGTCATGTCTGTGGAGGTGAATGGCCAGGTTTTTGAGGGCTCTGGTCCCA
CAAAGAAAAAGGCAAAACTCCATGCTGCTGAGAAGGCCTTGAGGTCTTTCGTTCAGTTTCCTAA
TGCCTCTGAGGCCCACCTGGCCATGGGGAGGACCCTGTCTGTCAACACGGACTTCACATCTGAC
CAGGCCGACTTCCCTGACACGCTCTTCAATGGTTTTGAAACTCCTGACAAGGCGGAGCCTCCCT
TTTACGTGGGCTCCAATGGGGATGACTCCTTCAGTTCCAGCGGGGACCTCAGCTTGTCTGCTTC
CCCGGTGCCTGCCAGCCTAGCCCAGCCTCCTCTCCCTGCCTTACCACCATTCCCACCCCCGAGTG
GGAAGAATCCCGTGATGATCTTGAACGAACTGCGCCCAGGACTCAAGTATGACTTCCTCTCCGA
GAGCGGGGAGAGCCATGCCAAGAGCTTCGTCATGTCTGTGGTCGTGGATGGTCAGTTCTTTGAA
GGCTCGGGGAGAAACAAGAAGCTTGCCAAGGCCCGGGCTGCGCAGTCTGCCCTGGCCGCCATT
TTTAACTTGCACTTGGATCAGACGCCATCTCGCCAGCCTATTCCCAGTGAGGGTCTTCAGCTGC
ATTTACCGCAGGTTTTAGCTGACGCTGTCTCACGCCTGGTCATAGGTAAGTTTGGTGACCTGAC
CGACAACTTCTCCTCCCCTCACGCTCGCAGAATAGGTCTGGCTGGAGTCGTCATGACAACAGGC
ACAGATGTTAAAGATGCCAAGGTGATATGTGTTTCTACAGGATCTAAATGTATTAATGGTGAAT
ACCTAAGTGATCGTGGCCTTGCATTAAATGACTGCCATGCAGAAATAGTATCTCGGAGATCCTT
GCTCAGATTTCTTTATACACAACTTGAGCTTTACTTAAATAACGAGGATGATCAAAAAAGATCC
ATCTTTCAGAAATCAGAGCGAGGGGGGTTTAGGCTGAAGGAGAATATACAGTTTCATCTGTAC
ATCAGCACCTCTCCCTGTGGAGATGCCAGAATCTTCTCACCACATGAGGCAATCCTGGAAGAAC

CAGCAGATAGACACCCAAATCGTAAAGCAAGAGGACAGCTACGGACCAAAATAGAGGCTGGT
CAGGGGACGATTCCAGTGCGCAACAATGCGAGCATCCAAACGTGGGACGGGGTGCTGCAAGG
GGAGCGGCTGCTCACCATGTCCTGCAGTGACAAGATTGCACGCTGGAACGTGGTGGGCATCCA
GGGATCACTGCTCAGCATTTTCGTGGAGCCCATTTACTTCTCGAGCATCATCCTGGGCAGCCTTT
ACCACGGGGACCACCTTTCCAGGGCCATGTACCAGCGGATCTCCAACATAGAGGACCTGCCAC
CTCTCTACACCCTCAACAAGCCTTTGCTCACAGGCATCAGCAATGCAGAAGCACGGCAGCCAG
GGAAGGCCCCCATATTCAGTGTCAACTGGACGGTAGGCGACTCCGCTATTGAGGTCATCAACG
CCACGACTGGGAAGGGAGAGCTGGGCCGCGCGTCCCGCCTGTGTAAGCACGCGTTGTACTGTC
GCTGGATGCGTGTGCACGGCAAGGTTCCCTCCCACTTACTACGCTCCAAGATTACCAAGCCCAA
CGTGTACCATGAGACAAAGCTGGCGGCAAAGGAGTACCAGGCCGCCAAGGCGCGTCTGTTCAC
AGCCTTCATCAAGGCGGGGCTGGGGGCCTGGGTGGAGAAGCCCACCGAGCAGGACCAGTTCTC
ACTCACGCCCGATTACAAGGATGACGACGATAAGTAG

2. BFP 报告体系的构建

BFP 报告体系参照参考文献 7 构建，全部 BFP（蓝色荧光蛋白）cDNA 序列在体外合成，具体序列为：SEQ ID NO 2。BFP cDNA 序列通过 CMV 启动子后的多克隆位点克隆至 pCDH-CMV 质粒载体(pCDH-CMV 质粒骨架由 Kazuhiro Oka 惠赠，Addgene plasmid # 72265; http://n2t.net/addgene:72265; RRID: Addgene_72265)。报告体系中待编辑的靶标碱基为 BFP 序列第 199 位的碱基 C，对应第 66 位组氨酸，参见图 1。

SEQ ID NO 2:

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC
GGCGACGTAAACGGCCACAAGTTCAGCGTGTCTGGCGAGGGCGAGGGCGATGCCACCTACGGC
AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA
CCACCCTGAC**CCA**CGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACT
TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCT
GAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACA
ACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGA
TCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA
TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCA

AAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCA
CTCTCGGCATGGACGAGCTGTACAAGTGA

该序列第 198,199,200 位的碱基依次为 CCA，命名为 BFP-CCA，缩略为 C*。当 199 位的碱基 C 在 RNA 水平通过脱氨基化被编辑为 U 之后，BFP 荧光蛋白会从原先的蓝色荧光变为绿色荧光，从而可以通过流式细胞仪 FITC（Fluorescein isothiocyanate，异硫氰酸荧光素）通道检测到信号。由于 198 位核苷酸从 C 突变为 A、T、G 后，对应密码子（196、197、198 位碱基）ACC、ACA、ACT、ACG 均编码苏氨酸，因此该位突变为同义突变。这使得该报告体系可以同时测定和比较在 mRNA 上靶标碱基的 5'上游相邻碱基不同时，C 到 U 的编辑效率。可据此使用定点诱变试剂盒（Q5® Site-Directed Mutagenesis Kit, NEB E0554S)向 198 位置碱基中引入突变，从而 198,199,200 三个位置碱基分别为：GCA，命名为 BFP-GCA，缩略为 G*；ACA，命名为 BFP-ACA，缩略为 A*；TCA 命名为 BFP-UCA，缩略为 U*。

实施例 2：CUSPER 体系的初步测试

1.arRNA 的设计与合成

本实施例中使用的 arRNA 包含与靶标 mRNA 互补的反义 RNA，靶标碱基位于 arRNA 的中间位置，5'上游以及 3'下游按照相同长度向两边延伸。由于合成长度的限制，本实施例选取长度为 91nt 的 RNA 进行体外合成。如图 3 所示，当所述 arRNA 第 46 位核苷酸分别为 A、U、G、C 时，则分别缩写为 A^、U^、G^、C^。四种合成的 arRNA 具体序列见下表 1。如图 2 所示，与 LEAPER 技术设计方法不同的是，本批试验中四条 arRNA 的设计，由于试验目的是确定靶标碱基与 arRNA 上的靶向碱基错配时的编辑效率，因此使用了不同的靶向碱基进行了测试，即对第 46 位分别为 A、U、G、C 的 arRNA 进行了测试。而在 arRNA 第 47 位的 5'最近邻碱基(对应报告体系第 198 位)则按照引入突变前的 BFP 序列的靶标三联体碱基（即 CCA）进行设计，即所述 arRNA 靶向碱基的 5'最近邻碱基（第 47 位）为与 A 互补的 U。这意味着，当且仅当后续测试的报告体系为 BFP-CCA 的时候，arRNA 的设计方才与 LEAPER 技术一致，即：当 arRNA 与靶标 RNA 杂交时，仅在靶标碱基存在错配；而当报告体系为 BFP-GCA，BFP-TCA，BFP-ACA 的时候，arR

NA 的设计不仅在靶标碱基存在错配，而且在 arRNA 靶标碱基的 3’相邻碱基处也存在错配。

表 1

arRNA 名称	arRNA 序列
A^	SEQ ID NO 3: gcugcuucauguggucgggguagcggcugaagcacugcacgccguAggucagggugguc acgagggugggccagggcacgggcagcuugcc
U^	SEQ ID NO 4: gcugcuucauguggucgggguagcggcugaagcacugcacgccguUggucagggugguc acgagggugggccagggcacgggcagcuugcc
C^	SEQ ID NO 5: gcugcuucauguggucgggguagcggcugaagcacugcacgccguCggucagggugguc acgagggugggccagggcacgggcagcuugcc
G^	SEQ ID NO 6: gcugcuucauguggucgggguagcggcugaagcacugcacgccguGggucagggugguc acgagggugggccagggcacgggcagcuugcc

注：大写字母只为突出序列间差异，同一字母的大小写不同不代表碱基的差异。

2. C 到 U 编辑测试

ADAR2-r16-293T 以 300000 个细胞/孔的密度铺板至 6 孔板，铺板后 24 小时，用 Lipofectamine 3000 转染 (Invitrogen L3000015)，转染步骤按照说明书进行。按照说明书采用不同浓度的 Lipofectamine 3000 转染试剂进行两次重复试验，重复 1 每孔使用 3.75μL 的 Lipofectamine 3000，重复每孔 2 使用 7.5μL 的 Lipofectamine 3000。每孔添加 2.5μg BFP 报告质粒（选自：BFP-GCA，缩略为 G*； BFP-ACA，缩略为 A*; BFP-TCA，缩略为 T*; BFP-CCA，缩略为 C*)以及 25pmol 化学合成的 arRNA。转染后 48h 通过 FACS 检测 FITC 通道信号强度。阳性细胞平均荧光强度(Mean Fluorescent Intensity, MFI)统计结果如图 3 所示。

图 3 中 mRNA 行表示对应孔中添加的 BFP 报告体系质粒，arRNA 行表示对应孔中添加的 arRNA。BFP 报告体系中，198,199,200 三个碱基在原始序列中为 CCA，而当 198 位 C 变为 A、T 或 G 的时候，对应的 65 位氨基酸均为苏氨酸，所以 BFP-GCA、BFP-CCA、BFP-ACA、BFP-TCA 四种不同报告体系 198 位的改变不会造成原本蛋白功能的改变。而当 199 位 C 被编辑变为 U 时，199、200、201 形成的密码子从 CAC 变为 UAC，对应的 66 位氨

基酸从组氨酸(Histidine, His, H)变为酪氨酸(Tyrosine, Tyr, Y)，从而实现BFP到GFP的荧光转变。如图3所示，当不加入任何arRNA时，报告体系本底GFP信号MFI约为5×10^4 (mRNA行标注为U*, arRNA行标注为/的报告体系；以及，mRNA行标注为A*, arRNA标注为/的报告体系)。而当通过DNA水平的点突变，使199位C突变为T时，GFP信号MFI约为2.4×10^6~3.1×10^6，比本底值高约100倍。因此，如果199位的C在RNA水平全部变为U，则可导致GFP MFI的约100倍提升。

在DNA水平199位C不变的基础上,如图3所示,而当加入arRNA后，GFP的MFI最多能提升至超过5×10^5，荧光强度超过了DNA水平的199位C点突变成T后荧光强度的20%。这说明，本申请的技术可以在不改变DNA序列的前提下，通过在转录水平上将199位的C转变为U，从而改变最终的蛋白功能。

在LEAPER技术文献中(Qu et al., 2019，原图2f，对应于本申请图4)，当靶标碱基3'相邻碱基（如N^2位置所示）为A时，LEAPER技术对靶标碱基5'相邻碱基（如N1位置所示）的偏好(指在5'上游碱基为A、U、G或C时，相同arRNA所得到的编辑效率)是：U > C≈A > G。在RESCUE技术文献中(Abudayyeh et al., 2019，原图1c，对应于本申请图5）当靶标碱基3'相邻碱基为A时，RESCUE技术对靶标碱基5'相邻碱基的偏好是：U≈A >> C≈G。而在本申请的CUSPER技术中，我们意外地发现，该技术在当靶标碱基3'相邻碱基为A时，对靶标碱基5'上游碱基的偏好与LEAPER技术和RESCUE技术两者均不相同，如图3所示，如果固定arRNA为编辑效率较高的U^或者C^，可以看到，本申请的技术对5'上游碱基的偏好是：G > C >> U≈A。

3. 对CUSPER碱基偏好性的进一步确定

为进一步确定CUSPER的编辑能力及其碱基偏好性，我们对实施例2第2部分做了进一步重复，如图6所示。相较于实施例2第2部分，本部分中补充了实施例2第2部分中未涉及的仅有BFP-GCA和BFP-CCA两个报告体系质粒而不添加任何arRNA情况下的对照试验。并且对图3相关测试

中 MFI 超过本底值 2 倍以上的条件做了重复。同时，试验中重复 1 和重复 2 对应两株不同批次制作的 ADAR2-r16-293T。

如图 6 所示，重复试验中该体系对靶标碱基 5’相邻碱基的偏好呈现出与图 3 显示的实施例 2 第 2 部分的相关测试类似的模式，均为在 5’相邻碱基为 G 或 C 时有较好编辑效率，且 G 大于 C。

以上结果表明，本申请的技术可以在不改变 DNA 序列的前提下，从转录水平影响最终蛋白功能。同时，该技术对靶标碱基 5’相邻碱基的偏好可能与 LERPER 和 RESCUE 技术均不相同，具体参见表 2。值得注意的是，本研究中对于 BFP-GCA、BFP-TCA、BFP-ACA 的测试中，arRNA 的设计不仅在靶标碱基处存在一个错配，在 arRNA 的 3’下游也存在另外一个错配 (如图 2 所示)。因此，后续还需进一步测试当这些位置不存在错配时，该技术对 5’上游碱基的偏好是否还存在类似的模式。

表 2. 三种 RNA 编辑技术靶标碱基相邻碱基偏好性对比

技术名称	错配数量	当靶标碱基 3’相邻碱基为 A 时，对 5’相邻碱基的偏好性
CUSPER 技术	CCA 一个错配 其余两个错配	G > C >> A ≈ U
LEAPER [4]	均一个错配	U > C ≈ A > G
RESCUE [1]	均一个错配	U≈A >> C≈G

实施例 3：3’及 5’相邻碱基均无错配的情况下，对 CUSPER 编辑体系的测试

通过对 GFP 信号的读取，我们可以快速粗略地判断不同 arRNA 的编辑效率，但如果需要进一步确认编辑效率，则需通过二代测序（Next Generation Sequencing, NGS)来最终确认 mRNA 中有多少比例的 C 被编辑成 U。同时，RNA 单碱基编辑体系无论是 A 到 I (Qu et al., 2019) 还是 C 到 U (Abudayyeh et al., 2019)其靶标碱基 A 或者 C 的 5’相邻碱基和 3’相邻碱基对其编辑效率影响较大。由于 BFP 到 GFP 报告体系的限制，如果以 BFP 的 DNA 序列(SEQ ID NO 2)中第 199 位的 C 为靶标碱基，当 5’相邻碱基（第 198 位）C 变为 A、T、G 时均不影响对应的 65 位氨基酸；但当 3’相邻碱基（第 200 位）的 A 变为 T、C 或 G 时，则会导致其完全失去 GFP 信号。因此，通过

报告体系读取GFP的试验无法对3'相邻碱基为T、C、G的体系进行测试。而如果通过二代测序则可直接读取编辑后其mRNA中U占A、U、C、G的百分比，从而较为方便地比较4种不同5'相邻碱基和4种不同3'相邻碱基共16种情况下，其编辑效率的高低。

参照实施例1.2的步骤，根据BFP的DNA序列(SEQ ID NO 2)中第198、199、200位的DNA序列不同，我们构建了16个不同的报告体系，即ACA、ACT（对应mRNA：ACU)、ACC、ACG、TCA（对应mRNA：UCA)、TCT（对应mRNA：UCU)、TCC（对应mRNA：UCC)、TCG（对应mRNA：UCG)、CCA、CCT（对应mRNA：CCU)、CCC、CCG、GCA、GCT（对应mRNA：GCU)、GCC、GCG。并且参照实施例1.1的慢病毒包装、侵染步骤，侵染293T，使其稳定整合入293T细胞中。

与以上16个不同的报告体系相对应，在mRNA靶标碱基C对应arRNA靶向碱基为U的基础上，使arRNA靶向碱基的3'及5'最近邻碱基可与靶标RNA中靶标碱基的5'及3'相邻碱基以沃森克里克碱基配对原则互补配对。所合成arRNA序列见表3，其中mRNA中靶标以及其5'和3'相邻碱基与arRNA靶向碱基以及其5'和3'最近邻碱基的对应关系表见图7。

按照以下步骤，向16个不同的报告体系细胞中，转染入对应的arRNA，并进行RNA提取和二代测序。

1.细胞培养采用含有10%FBS (Vistech SE100-011)的DMEM (Hyclone SH30243.01)。报告体系细胞以15000个细胞/孔，传至12孔板。此时记为0小时。

2.细胞传代后24小时，用RNAi MAX(Invitrogen, 13778150)试剂将12.5pmol的arRNA转入每个孔。转染步骤参照供应商说明书。

3.细胞传代后72小时，用胰酶(Invitrogen 25300054)消化整孔细胞，用800μL TRIzol (Invitrogen, 15596018)收样，采用Direct-zol RNA Miniprep试剂盒(Zymo Resaerch, R2052)进行RNA提取，每个样品取1000ng提取好的总RNA用TransScript® One-Step gDNA Removal and cDNA Synthesis SuperMix试剂盒(全式金，AT311)进行反转录合成cDNA，取1μL反转录产物，用序列为ggagtgagtacggtgtgcCTACGGCAAGCTGACCCTGAAGTT（SEQ ID NO：7）和gagttggatgctggatggGTAGTTGCCGTCGTCCTTGAAGAAG

（SEQ ID NO：8）的两个引物以及Q5热启动酶(NEB, M0494L)进行PCR。PCR产物用Hi-TOM试剂盒(诺禾致源，REF PT045)进行建库并按照以下步骤完成二代测序和数据分析。

i. Illumina测序

将构建好的测序文库，通过NovaSeq6000平台以PE150方式进行高通量测序。

ii.测序数据处理

高通量测序得到的原始数据以fastp (v0.19.6)进行质控，过滤掉低质量、带有接头序列、及含有polyG等的序列。将得到的高质量测序数据以自主开发的拆分脚本按照相应的Barcode序列拆分到每个样本，使用BWA (v0.7.17-r1188) 软件与扩增的目标区域序列进行比对，通过SAMtools (v1.9) 进行格式转换以生成BAM文件、统计比对信息并重新排序和建立索引。

iii. 编辑效率分析

使用JACUSA (v1.3.0)软件检测所有的mRNA靶标碱基，所用参数为：call-1 -a B,R,D,I,Y,M:4 -C ACGT -c 2 -p 1 -P UNSTRANDED -R -u DirMult-CE。过滤掉同时在对照和处理样本中出现的高频突变之后，以C->U突变之外的平均突变频率的三倍作为阈值，将靶标碱基C到U的突变频率在阈值之上的部分作为真实的靶标C突变为U的频率。

二代测序结果如图8显示。通过二代测序可以证实，CUSPER体系确实可以实现mRNA的C到U单碱基编辑。相较于实施例2可以看出，如果靶向碱基的5’及3’最近邻碱基与靶标RNA均无错配，则编辑效率相对较低，而当靶向碱基的5’最近邻碱基与靶标RNA形成G-G错配时，便可以提高编辑效率。此外，当除靶标碱基的C-U错配以外，靶向碱基的5’及3’最近邻碱基均与靶标碱基的3’及5’相邻碱基以沃森-克里克原则配对时，当mRNA上靶标碱基三联体为下述序列时可以获得较高的编辑效率：ACG、ACC、UCC、UCG、CCC、CCG、UCA、UCU。

表3. 3’及5’相邻碱基均无错配的arRNA序列

arRNA名称	**arRNA序列**
arRNA-GCA	gcugcuucauguggucgggguagcggcugaagcacugcacgccgUUCgucaggguggucacgagggugggccagggcacgggcagcuugcc （SEQ ID NO：9）

arRNA-GCU	gcugcuucauguggucgggguagcggcugaagcacugcacgccgAUCgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 10)
arRNA-GCC	gcugcuucauguggucgggguagcggcugaagcacugcacgccgGUCgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 11)
arRNA-GCG	gcugcuucauguggucgggguagcggcugaagcacugcacgccgCUCgucaggguggucacga gggugggccagggcacgggcagcuugcc (SEQ ID NO: 12)
arRNA-CCA	gcugcuucauguggucgggguagcggcugaagcacugcacgccgUUGgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 13)
arRNA-CCU	gcugcuucauguggucgggguagcggcugaagcacugcacgccgAUGgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 14)
arRNA-CCC	gcugcuucauguggucgggguagcggcugaagcacugcacgccgGUGgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 15)
arRNA-CCG	gcugcuucauguggucgggguagcggcugaagcacugcacgccgCUGgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 16)
arRNA-UCA	gcugcuucauguggucgggguagcggcugaagcacugcacgccgUUAgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 17)
arRNA-UCU	gcugcuucauguggucgggguagcggcugaagcacugcacgccgAUAgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 18)
arRNA-UCC	gcugcuucauguggucgggguagcggcugaagcacugcacgccgGUAgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 19)
arRNA-UCG	gcugcuucauguggucgggguagcggcugaagcacugcacgccgCUAgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 20)
arRNA-ACA	gcugcuucauguggucgggguagcggcugaagcacugcacgccgUUUgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 21)
arRNA-ACU	gcugcuucauguggucgggguagcggcugaagcacugcacgccgAUUgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 22)
arRNA-ACC	gcugcuucauguggucgggguagcggcugaagcacugcacgccgGUUgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 23)
arRNA-ACG	gcugcuucauguggucgggguagcggcugaagcacugcacgccgCUUgucaggguggucacg agggugggccagggcacgggcagcuugcc (SEQ ID NO: 24)

arRNA-随机	uaauccugaauaucgcgcaauuccccagcagagaacaucgcggugugaacgucccuuuauaccgg gcagguauagcugaaaucagcguggc（SEQ ID NO：25）

注：大小写字母无区别，大写字母只为突出序列间差异。其中arRNA，以其靶向的靶标碱基三联体命名。

参考文献

1. Abudayyeh, O. O., Gootenberg, J. S., Franklin, B., Koob, J., Kellner, M. J., Ladha, A., ... & Zhang, F. (2019). A cytosine deaminase for programmable single-base RNA e diting. *Science, 365*(6451), 382-386.

2. Cox, D. B., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joun g, J., & Zhang, F. (2017). RNA editing with CRISPR-Cas13. *Science, 358*(6366), 1019-1027.

3. Eckstein, F. (2014). Phosphorothioates, essential components of therapeutic oligonucle otides. *Nucleic acid therapeutics, 24*(6), 374-387.

4. Merkle, T., Merz, S., Reautschnig, P., Blaha, A., Li, Q., Vogel, P., ... & Stafforst, T. (2019). Precise RNA editing by recruiting endogenous ADARs with antisense oligon ucleotides. *Nature biotechnology, 37*(2), 133.

5. Pollard, K. M., Cauvi, D. M., Toomey, C. B., Morris, K. V., & Kono, D. H. (201 3). Interferon-γ and systemic autoimmunity. *Discovery medicine, 16*(87), 123.

6. Qu, L., Yi, Z., Zhu, S., Wang, C., Cao, Z., Zhou, Z., ... & Bao, Y. (2019). Progra mmable RNA editing by recruiting endogenous ADAR using engineered RNAs. *Nature biotechnology, 37*(9), 1059-1069.

7. Vu, L. T., Nguyen, T. T. K., Md Thoufic, A. A., Suzuki, H., & Tsukahara, T. (201 6). Chemical RNA editing for genetic restoration: the relationship between the structure and deamination efficiency of carboxyvinyldeoxyuridine oligodeoxynucleotides. *Chemica l biology & drug design, 87*(4), 583-593.

8. Xu, L., Wang, J., Liu, Y., Xie, L., Su, B., Mou, D., ... & Zhao, L. (2019). CRISP R-edited stem cells in a patient with HIV and acute lymphocytic leukemia. *New Engla nd Journal of Medicine, 381*(13), 1240-1247.