



(12) 发明专利

(10) 授权公告号 CN 110944674 B

(45) 授权公告日 2024. 07. 19

(21) 申请号 201880048355.4

(22) 申请日 2018.05.18

(65) 同一申请的已公布的文献号
申请公布号 CN 110944674 A

(43) 申请公布日 2020.03.31

(30) 优先权数据
62/508,968 2017.05.19 US

(85) PCT国际申请进入国家阶段日
2020.01.19

(86) PCT国际申请的申请数据
PCT/US2018/033515 2018.05.18

(87) PCT国际申请的公布数据
W02018/213786 EN 2018.11.22

(73) 专利权人 编码治疗公司
地址 美国加利福尼亚州

(72) 发明人 卡提克·拉马穆尔蒂

斯蒂芬妮·塔利亚泰拉
安妮·坦依豪斯 安德鲁·杨
陈思颖 张弛 斯蒂芬妮·马丁
大卫·奥伯科夫勒

(74) 专利代理机构 北京市金杜律师事务所
11256
专利代理师 陈文平

(51) Int.Cl.
A61K 48/00 (2006.01)
C12N 15/85 (2006.01)
C12N 15/861 (2006.01)

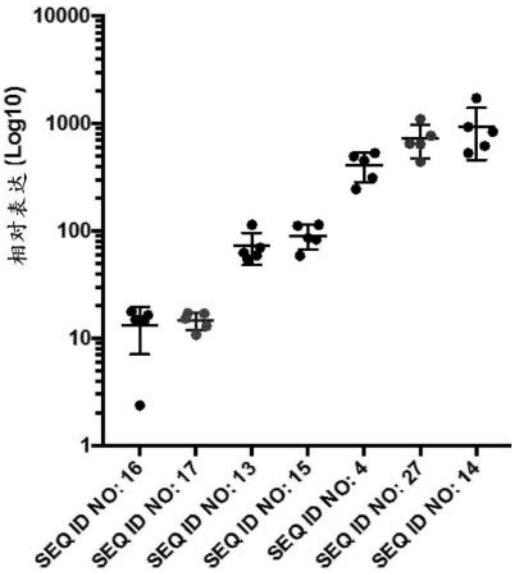
(56) 对比文件
CA 2985624 A1, 2016.06.23
CN 104520428 A, 2015.04.15

审查员 崔娟娟

权利要求书6页 说明书53页
序列表12页 附图22页

(54) 发明名称
高活性调控元件

(57) 摘要
本文提供了用于驱动转基因的高表达的组合物和方法。用于驱动转基因的高表达的组合物和方法包括一个或多个来源于人类的调控元件，该调控元件在与转基因可操作地连接时，可导致转基因在一种或多种细胞类型或组织中的高表达。



1. 一种表达盒,其包含与治疗性转基因可操作地连接的调控元件,其中所述调控元件包含SEQ ID NO:2的序列,所述调控元件导致转基因表达增加。
2. 根据权利要求1所述的表达盒,其中所述调控元件是非天然存在的。
3. 根据权利要求1所述的表达盒,其中所述调控元件包含内含子序列。
4. 根据权利要求3所述的表达盒,其中所述调控元件位于启动子和所述治疗性转基因之间。
5. 根据权利要求1所述的表达盒,其中所述调控元件包含启动子序列。
6. 根据权利要求5所述的表达盒,其中所述调控元件是所述盒中唯一的启动子。
7. 根据权利要求1所述的表达盒,其中所述调控元件不超过100bp。
8. 根据权利要求1所述的表达盒,其中所述调控元件不超过60bp。
9. 根据权利要求1所述的表达盒,其中所述调控元件不超过50bp。
10. 根据权利要求1所述的表达盒,其中所述表达盒是重组腺伴随病毒(rAAV)的一部分。
11. 根据权利要求1所述的表达盒,其中所述rAAV是rAAV8。
12. 根据权利要求1-11中任一项所述的表达盒,其中所述治疗性转基因是ATP7B或其变体。
13. 根据权利要求1-11中任一项所述的表达盒,其中所述治疗性转基因是因子VIII或其变体。
14. 根据权利要求1-11中任一项所述的表达盒在制备用于治疗威尔逊病的药物中的用途,其中所述转基因是ATP7B或其变体。
15. 根据权利要求1-11中任一项所述的表达盒在制备用于治疗ATP7B中的遗传缺陷的药物中的用途,其中所述转基因是ATP7B或其变体。
16. 根据权利要求1-11中任一项所述的表达盒在制备用于治疗凝血障碍的药物中的用途,其中所述转基因是因子VIII或其变体。
17. 根据权利要求1-11中任一项所述的表达盒在制备用于治疗单倍性不足或遗传缺陷的药物中的用途。
18. 根据权利要求17所述的用途,其中所述治疗性转基因是ATP7A或其变体。
19. 根据权利要求17所述的用途,其中所述治疗性转基因是ATP7B或其变体。
20. 根据权利要求17所述的用途,其中所述治疗性转基因是ATP8B1或其变体。
21. 根据权利要求17所述的用途,其中所述治疗性转基因是ABCB4或其变体。
22. 根据权利要求17所述的用途,其中所述治疗性转基因是ABCB11或其变体。
23. 根据权利要求17所述的用途,其中所述治疗性转基因是CDKL5或其变体。
24. 根据权利要求17所述的用途,其中所述治疗性转基因是CNTNAP2或其变体。
25. 根据权利要求17所述的用途,其中所述治疗性转基因是ZEB2或其变体。
26. 根据权利要求17所述的用途,其中所述治疗性转基因是因子V或其变体。
27. 根据权利要求17所述的用途,其中所述治疗性转基因是因子VII或其变体。
28. 根据权利要求17所述的用途,其中所述治疗性转基因是因子VIII或其变体。
29. 根据权利要求17所述的用途,其中所述治疗性转基因是因子IX或其变体。
30. 根据权利要求17所述的用途,其中所述治疗性转基因是因子XI或其变体。

31. 根据权利要求1-11中任一项所述的表达盒在制备用于治疗单倍性不足或遗传缺陷的药物中的用途,其中所述治疗性转基因是调节内源基因表达的转录调节因子。

32. 根据权利要求31所述的用途,其中所述转录调节因子是所述内源基因的转录激活因子。

33. 根据权利要求32所述的用途,其中所述内源基因是ATP7A。

34. 根据权利要求32所述的用途,其中所述内源基因是ATP7B。

35. 根据权利要求32所述的用途,其中所述内源基因是ATP8B1。

36. 根据权利要求32所述的用途,其中所述内源基因是ABCB4。

37. 根据权利要求32所述的用途,其中所述内源基因是ABCB11。

38. 根据权利要求32所述的用途,其中所述内源基因是CDKL5。

39. 根据权利要求32所述的用途,其中所述内源基因是CNTNAP2。

40. 根据权利要求32所述的用途,其中所述内源基因是ZEB2。

41. 根据权利要求32所述的用途,其中所述转基因是因子V。

42. 根据权利要求32所述的用途,其中所述转基因是因子VII。

43. 根据权利要求32所述的用途,其中所述转基因是因子VIII。

44. 根据权利要求32所述的用途,其中所述转基因是因子IX。

45. 根据权利要求32所述的用途,其中所述转基因是因子XI。

46. 一种AAV表达盒,其包含与至少3kb的转基因可操作地连接的不超过120bp的来源于人类的调控元件,其中与当可操作地连接CMV启动子时的所述转基因的表达相比,所述调控元件导致转基因表达增加至少2倍,并且其中所述调控元件是SEQ ID NO:2。

47. 根据权利要求46所述的AAV表达盒,其中所述增加的转基因表达是与当可操作地连接CMV启动子时的所述转基因的表达相比至少50倍。

48. 根据权利要求46所述的AAV表达盒,其中所述增加的转基因表达是至少100倍。

49. 根据权利要求46所述的AAV表达盒,其中与可操作地连接所述转基因的CMV启动子的大小标准化表达活性相比,所述调控元件表现出至少1.5倍的大小标准化表达活性。

50. 根据权利要求46所述的AAV表达盒,其中所述增加的转基因表达在至少2种不同的细胞类型中发生。

51. 根据权利要求50所述的AAV表达盒,其中所述至少2种不同的细胞类型选自肾细胞、神经元和肝脏细胞。

52. 根据权利要求46所述的AAV表达盒,其中所述调控元件位于启动子的下游。

53. 根据权利要求46所述的AAV表达盒,其中所述转基因是以下中的任一个:ATP7a; ATP7B; ATP8B1; ABCB4; ABCB11; CDKL5; CNTNAP2; ZEB2; 因子1、2、3、4、5、6、7、8、9、10、11和12; 或其变体或功能片段。

54. 根据权利要求46所述的AAV表达盒,其中所述转基因是ATP7B或其变体。

55. 根据权利要求46所述的AAV表达盒,其中所述转基因是FVIII或其变体。

56. 根据权利要求46所述的AAV表达盒,其中所述转基因是基因编辑蛋白。

57. 根据权利要求56所述的AAV表达盒,其中所述基因编辑蛋白是Cas。

58. 根据权利要求46所述的AAV表达盒,其中所述转基因是DNA结合蛋白。

59. 根据权利要求58所述的AAV表达盒,其中所述DNA结合蛋白是增加内源基因表达的

转录激活因子。

60. 根据权利要求58所述的AAV表达盒,其中所述转基因是减少内源基因表达的转录抑制因子。

61. 根据权利要求59所述的AAV表达盒,其中所述转录激活因子增加以下内源基因中的任一种的表达:ATP7A;ATP7B;ATP8B1;ABCB4;ABCB11;CDKL5;CNTNAP2;ZEB2;以及因子1、2、3、4、5、6、7、8、9、10、11和12。

62. 根据权利要求59所述的AAV表达盒,其中所述转录激活因子增加内源性ATP7A的表达。

63. 根据权利要求59所述的AAV表达盒,其中所述转录激活因子增加内源性ATP7B的表达。

64. 根据权利要求59所述的AAV表达盒,其中所述转录激活因子增加内源性ATP8B1的表达。

65. 根据权利要求59所述的AAV表达盒,其中所述转录激活因子增加内源性ABCB4的表达。

66. 根据权利要求59所述的AAV表达盒,其中所述转录激活因子增加内源性ABCB11的表达。

67. 根据权利要求59所述的AAV表达盒,其中所述转录激活因子增加内源性FVIII的表达。

68. 根据权利要求46-67中任一项所述的AAV表达盒,其中所述AAV选自:AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ和scAAV。

69. 根据权利要求68所述的AAV表达盒,其中所述AAV是AAV8。

70. 一种产生重组蛋白的方法,所述方法包括将编码所述蛋白质的序列与SEQ ID NO:2的序列可操作地连接。

71. 调控元件在制备用于治疗肝脏疾病或病况的基因疗法药物中的用途,其中所述调控元件包含SEQ ID NO:2的序列且与治疗性转基因可操作性地连接。

72. 根据权利要求71所述的用途,其中所述肝脏疾病或病况是威尔逊病。

73. 根据权利要求71所述的用途,其中所述肝脏疾病或病况是凝血障碍。

74. 根据权利要求72所述的用途,其中所述转基因是ATP7A或ATP7B,或其变体或功能片段。

75. 根据权利要求73所述的用途,其中所述转基因是因子1、2、3、4、5、6、7、8、9、10、11、12,或其变体或功能片段。

76. 根据权利要求75所述的用途,其中所述转基因是因子8,或其变体或功能片段。

77. 根据权利要求71所述的用途,其中所述调控元件导致至少2种细胞类型中增加的转基因表达。

78. 根据权利要求71所述的用途,其中所述调控元件导致至少3种细胞类型中增加的转基因表达。

79. 根据权利要求71所述的用途,其中所述调控元件导致肝细胞中增加的转基因表达。

80. 根据权利要求71-79中任一项所述的用途,其中所述调控元件导致与当可操作地连接CMV启动子时的转基因表达相比至少2倍水平的转基因表达增加。

81. 根据权利要求72所述的用途,其中所述基因疗法利用AAV。
82. 根据权利要求81所述的用途,其中所述AAV是AAV8。
83. 一种表达载体,其包含:与转基因可操作地连接的具有小于或等于100bp的来源于人类的调控元件,其中所述调控元件是SEQ ID NO:2,并且其中与不具有所述调控元件的第二表达载体相比,所述调控元件使所述转基因的全局表达增加至少两倍。
84. 根据权利要求83所述的表达载体,其中所述第二表达载体包含与CMV启动子可操作地连接的所述转基因。
85. 一种表达载体,其包含与转基因可操作地连接的具有小于或等于100bp的来源于人类的调控元件,其中所述调控元件是SEQ ID NO:2,并且借此所述转基因的表达高于由CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子表达的相同转基因的表达。
86. 根据权利要求85所述的表达载体,其中所述转基因的所述表达高于由UCL-HLP启动子表达的相同转基因的表达。
87. 一种表达载体,其包含与转基因可操作地连接的来源于人类的调控元件,其中所述调控元件是SEQ ID NO:2,并且其中由所述转基因编码的蛋白质具有(i)通过被配置用于检测所述转基因的ELISA测定所测量的 $>1.0\text{IU/mL}$ 的浓度;或(ii)通过Coatest测定所测量的 $>25\%$ 的活性。
88. 一种载体,其包含来源于人类的调控元件,所述调控元件具有大小为小于或等于100bp的序列,与在细胞中或体内未发现与所述调控元件相关的转基因可操作地连接,并且其中所述调控元件是SEQ ID NO:2。
89. 根据权利要求83、85、87或88所述的载体,其中所述调控元件是内含子序列。
90. 根据权利要求83、85或87所述的表达载体,其中使用荧光素酶作为所述转基因导致在整个小鼠中测量的大于 1×10^8 光子/秒的全局表达。
91. 根据权利要求87所述的表达载体,其中所述活性或转基因表达对应于每只小鼠 $16\mu\text{g}$ 表达载体的剂量。
92. 根据权利要求91所述的表达载体,其中所述活性或转基因表达对应于每只小鼠 $12\mu\text{g}$ 表达载体的剂量。
93. 根据权利要求83、85或87所述的表达载体,其中所述转基因的内源形式在体内不与所述调控元件连接。
94. 根据权利要求83所述的表达载体,其中所述转基因的所述全局表达处于高于使用带有选自以下的调控元件的载体的所述转基因表达的水平:CMV启动子、CMVe启动子、超级核心启动子、TTR启动子、Proto 1启动子和UCL-HLP启动子。
95. 根据权利要求83、85、87或88所述的表达载体,其中在小鼠体内的至少3、4、5、6或7种不同的细胞类型中可检测到表达。
96. 根据权利要求85所述的表达载体,其中所述不同的细胞类型选自:肺泡细胞、心肌细胞、上皮细胞、肝细胞、肠细胞、肌细胞、神经元和肾脏细胞。
97. 根据权利要求83、85、87或88所述的表达载体,其中所述调控元件是增强子。
98. 根据权利要求97所述的表达载体,进一步包含启动子。
99. 根据权利要求98所述的表达载体,其中所述启动子是CMV启动子、CMV、启动子、超级

核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子、AAT启动子、KAR启动子、EF1 α 启动子、EFS启动子或CMVe增强子/CMV启动子组合。

100. 根据权利要求97所述的表达载体,进一步包含一个或多个转录后修饰位点。

101. 根据权利要求83、85、87或88所述的表达载体,其中所述转基因是Cas9。

102. 根据权利要求101所述的表达载体,其中所述转基因是saCas9。

103. 一种表达载体,其包含与调控序列可操作地连接的因子VIII转基因,所述调控序列能够驱动所述因子VIII的表达至>1.0IU/mL的浓度,如通过被配置用于在细胞中或在体内检测因子VIII的ELISA测定所测量的,并且其中所述调控元件是SEQ ID NO:2。

104. 一种病毒颗粒,其包含权利要求83-103中任一项所述的表达载体。

105. 根据权利要求104所述的病毒颗粒,其中所述病毒颗粒是AAV。

106. 根据权利要求105所述的病毒颗粒,其中所述AAV选自:AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ和scAAV。

107. 根据权利要求104所述的病毒颗粒,其中所述病毒颗粒是慢病毒。

108. 根据权利要求104所述的病毒颗粒,其中所述病毒颗粒是腺病毒。

109. 根据权利要求83、85、87或88所述的表达载体,其中所述转基因是治疗性转基因。

110. 根据权利要求109所述的表达载体,其中所述治疗性转基因是因子VIII、Cas9、DNA结合蛋白、激素、生长或分化因子、胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体、囊性纤维化跨膜调节因子、与粘多糖贮积症I、II、III或IV型相关的基因、 β 珠蛋白或脂蛋白脂肪酶。

111. 根据权利要求109所述的表达载体,其中所述治疗性转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体或片段。

112. 根据权利要求109所述的表达载体,其中所述治疗性转基因是CDKL5、CNTNAP2、ZEB2,或其变体或片段。

113. 根据权利要求83-112中任一项所述的表达载体在制备用于将转基因递送至动物中的多个不同组织的药物中的用途。

114. 一种用于产生蛋白质、抗体或其他生物制剂的方法,其包括使细胞与权利要求83-112中任一项所述的表达载体接触。

115. 根据权利要求114所述的方法,其中所述细胞是CHO细胞或HEK293T细胞。

116. 一种用于产生转基因动物或植物的方法,其包括向动物或植物施用权利要求83-112中任一项所述的表达载体。

117. 根据权利要求83、85、87或88所述的表达载体,其中所述调控元件为40-50bp。

118. 根据权利要求83、85、87或88所述的表达载体,其中所述调控元件为50-60bp。

119. 一种表达载体,其包含与转基因可操作地连接的来源于人类的调控元件,其中由所述转基因编码的蛋白质具有(i)通过被配置用于检测所述转基因的ELISA测定所测量的>1.0IU/mL的浓度;或(ii)通过Coatest测定所测量的>10%的活性,并且其中所述调控元件是SEQ ID NO:2。

120. 一种表达载体,其包含与转基因可操作地连接的具有小于或等于120bp的来源于人类的调控元件,借此所述转基因的表达高于由UCL-HLP启动子表达的相同转基因的表达,

并且其中所述调控元件是SEQ ID NO:2。

121. 根据权利要求119或120所述的表达载体,其中所述调控元件是启动子。

122. 根据权利要求119或120所述的表达载体,其中所述治疗性转基因是包含DNA结合域和转录调节域的融合蛋白。

高活性调控元件

[0001] 交叉引用

[0002] 本申请要求于2017年5月19日提交的美国临时申请号62/508,968的权益,其通过引用整体并入本文。

[0003] 序列表

[0004] 本申请包含序列表,该序列表已经以ASCII格式电子提交,并且通过引用整体并入于此。所述ASCII副本创建于2018年5月17日,被命名为46482-707_601_SL.txt,并且大小为16,836字节。

背景技术

[0005] 基因工程在治疗疾病和产生用于治疗或其他用途的蛋白质方面具有巨大潜力。患者,尤其是那些具有潜在遗传因素的患者,可以直接针对潜在病因进行治疗,而无需依靠药物或手术。此外,通过针对潜在病因本身,基因疗法具有有效治愈患者的潜力。然而,尽管如此,基因疗法的临床应用仍在几个方面需要改进。一个挑战是在一个或多个靶组织中实现转基因的高表达率,或在一个或多个靶组织中实现基因编辑的高效率。另一个挑战是由常用载体的包装限制(或克隆能力)所引入的。需要能够驱动可操作地连接的转基因的高表达的短调控元件。

发明内容

[0006] 本文描述了用于驱动转基因的高表达的组合物和方法。本文还描述了用于驱动包含一个或多个人类或来源于人类的调控元件(RE)的转基因的高表达的组合物和方法,该调控元件当与所述转基因可操作地连接时可促进或导致所述转基因在一种或多种靶细胞类型或组织中的高(或增加的)表达。在一些情况下,本文公开的任何实施方案的高或增加的转基因表达与对照相比进行确定,该对照例如,组成型启动子、CAG、CMV启动子、超级核心启动子(SCP)、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子。可用于比较以确定本文公开的调控元件的高或增加的转基因表达的其他对照包括单独的缓冲液、minCMV、EFS、单独的载体或不驱动表达的序列。在一些实施方案中,本文公开的一个或多个调控元件在细胞中或体内、体外和/或离体驱动转基因的高表达。在一些实施方案中,包含与转基因可操作地连接的本文所述的一个或多个调控元件的表达盒或载体可适于在任何表达系统或基因疗法中使用,以驱动转基因在细胞、受试者或动物模型中的高表达,或导致在细胞、受试者或动物模型中治疗有效水平的转基因表达。在一些实施方案中,本公开内容的一个或多个RE与较大转基因(例如,其序列超过1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb的转基因)可操作地连接,以导致所述转基因在细胞中或体内的高表达。在一些情况下,所述转基因在细胞中或体内的这样的高表达是相对于不具有所述RE的所述转基因的表达,其中与不具有所述RE的转基因表达相比,具有所述RE的转基因的表达是至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍或大

于100倍。在一些情况下,一个或多个RE在至少2种、至少3种、至少4种、至少5种、至少6种、至少7种、至少8种、至少9种或至少10种不同的细胞类型中导致这样的高转基因表达。在一些情况下,本公开内容的一个或多个RE与转基因可操作地连接,用于适于全身施用的基因疗法治疗。在一些情况下,本公开内容的一个或多个RE与转基因可操作地连接,用于适于在肝脏或肝细胞中表达的基因疗法治疗。在一些情况下,转基因是DNA结合蛋白,例如,转录调节因子,其调控内源基因。在一些情况下,本文公开的任何实施方案的RE包含启动子序列、内含子序列、5' UTR序列或其任何组合。在一些情况下,本文公开的任何实施方案的RE是来源于人类的序列(或与人类参考基因组中的序列具有至少80%、90%、95%或99%序列同一性的序列)。在一些情况下,本文公开的任何实施方案的RE的长度小于或等于300bp、250bp、200bp、150bp、140bp、130bp、120bp、110bp、100bp、70bp或50bp。在一些情况下,表达构建体(例如,载体、AAV或病毒载体)包含与本公开内容的转基因可操作地连接的一个或多个外源RE序列,其中每个外源RE序列选自表1列出的或SEQ ID NO:1-2、13-17和22-41的序列;或者其中每个外源RE序列包含表1或SEQ ID NO:1-2、13-17和22-41的序列;或者其中每个外源RE序列包含与表1或SEQ ID NO:1-2、13-17和22-41的序列具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性(例如,局部序列同一性)的序列(例如,通过BLAST测量的序列同一性)。在一些情况下,表达盒(例如,载体、AAV或病毒载体)包含与任何转基因可操作地连接的一个或多个RE序列,其中每个RE序列不超过49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp,并且包含根据表1或SEQ ID NO:1-2、13-17和22-41中的任一个的序列。在一些情况下,表达构建体(例如,载体、AAV或病毒载体)包含与本公开内容的转基因可操作地连接的一个或多个外源RE序列,其中每个RE序列不超过49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp,并且包含与表1或SEQ ID NO:1-2、13-17和22-41中的任一个具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性的序列。在一些情况下,使用BLAST测量局部序列同一性。在一些情况下,序列同一性百分比是指全局序列同一性,覆盖表达盒中的整个外源RE。在其他情况下,序列同一性百分比是指局部序列同一性,覆盖与表1列出的任一序列对应的区域或部分区域。在一些情况下,序列同一性百分比是指局部序列同一性,覆盖包含表达盒的不超过49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp的区域,不包括转基因。如本文所用,当描述本公开内容的RE时,“相对较短”优选是不超过45bp、49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、

259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp的RE序列。相对较短的RE的实例包括表1列出的或SEQ ID NO:1-2、13-17和22-41的序列;或者与其具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性,并且同样不超过45bp、49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp的序列。

[0007] 在一些情况下,本文公开的任何实施方案的RE序列位于转基因的上游或下游。在一些情况下,本文公开的任何实施方案的两个或更多个RE与转基因可操作地连接,其中所述RE中的至少一个或多个位于所述转基因的上游或下游。在一些情况下,本公开内容的任何实施方案的一个或多个RE与转基因可操作地连接,以驱动所述转基因的全局表达,其中全局表达是指在至少2、3、4、5、6、7、8、9或10种不同细胞类型的表达。在一些情况下,本公开内容的任何实施方案的一个或多个RE与转基因可操作地连接,以驱动转基因的全局表达,其中包含所述RE和所述转基因的构建体可以全身递送,例如,通过口服、静脉内、肌肉内、腹膜内、鞘内、肠内或肠胃外施用。在各种实施方案中,本文公开的RE是来源于人类的(或是与人类参考基因组中的序列具有至少80%、90%、95%或99%序列同一性的序列)。在一些情况下,本文公开的任何实施方案的RE是非天然存在的。

[0008] 在一些情况下,表达载体包含来源于人类的调控元件,其具有小于或等于400bp、300bp、250bp、200bp、150bp、140bp、130bp、120bp、110bp、100bp、70bp或50bp,与转基因可操作地连接,其中与不具有所述调控元件的第二表达载体相比,所述调控元件使所述转基因的全局表达(例如,在至少2、3、4、5、6或更多种不同细胞类型中的表达)增加至少两倍。在一些情况下,所述第二表达载体包含与对照启动子(如CMV启动子)可操作地连接的转基因。

[0009] 在一些情况下,本公开内容的RE包含来源于人类的内含子序列,例如,SEQ ID NO:1-2。在一些情况下,本公开内容的表达盒包含与转基因可操作地连接的SEQ ID NO:1-2中的一个或多个。在一些情况下,包含SEQ ID NO:1-2中的任一个或多个的RE位于所述转基因的上游、下游或上游和下游两者。

[0010] 在一些情况下,本公开内容的RE包含来源于人类的启动子序列,例如,SEQ ID NO:13-17和22-41。在一些情况下,本公开内容的表达盒包含与转基因可操作地连接的SEQ ID NO:13-17和22-41中的一个或多个。在一些情况下,包含SEQ ID NO:13-17和22-41中的任一个或多个的RE位于所述转基因的上游、下游或上游和下游两者。

[0011] 在一些情况下,本公开内容的RE包含来源于人类的内含子序列和/或启动子序列,例如,SEQ ID NO:1-2、13-17和22-41中的两个或更多个;或者表1列出的序列中的两个或更多个。在一些情况下,本公开内容的表达盒包含与转基因可操作地连接的表1列出的序列中的一个或多个、两个或更多个、三个或更多个、四个或更多个或者五个或更多个。在一些情况下,表1列出的一个或多个RE在表达盒中位于转基因的上游、下游或上游和下游两者。在一些情况下,表1列出的一个或多个RE在本文公开的表达盒中位于转基因的上游。

[0012] 在一些情况下,本文公开的任何实施方案的RE由选自表1或SEQ ID NO:1-2、13-17和22-41及其任何组合的序列组成。在一些情况下,表1列出的两个或更多个RE在没有连接

体序列的情况下组合。在一些情况下,使用1-50bp的连接体序列将表1列出的两个或更多个RE组合。在一些情况下,将表1列出的两个或更多个RE放置在表达构建体中,使得RE序列的5'起始间隔小于2bp、5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb。

[0013] 在一些情况下,表达载体包含与转基因可操作地连接的具有小于或等于300bp、250bp、200bp、150bp、140bp、130bp、120bp、110bp、100bp、70bp或50bp的来源于人类的调控元件,借此转基因的表达高于由对照调控元件如组成型启动子、CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子表达的相同转基因。在一些情况下,与RE可操作地连接的转基因的表达高于连接至UCL-HLP启动子的无RE的相同转基因的表达。

[0014] 在一些情况下,表达载体包含与转基因可操作地连接的来源于人类的调控元件,其中由所述转基因编码的蛋白质具有(i)>1.0IU/mL、>1.5IU/mL、>2.0IU/mL、>2.5IU/mL或>3.0IU/mL的浓度,通过被配置用于检测所述转基因的ELISA测定所测量;和/或(ii)>20%、>25%、>30%、>35%、>40%、>45%、>50%、>55%、>60%、>65%、>70%、>75%、>80%、>85%、>90%或>95%的活性,通过Coatest测定所测量。在一些情况下,这样的活性或转基因表达在动物模型如小鼠中测定。在一些情况下,将本文所述的转基因表达或活性相对于递送至细胞或小鼠的表达盒、载体、病毒或DNA的剂量,如每只小鼠16μg(或10-20μg)表达载体的剂量标准化。

[0015] 在一些情况下,载体包含来源于人类的调控元件,其具有大小为小于或等于300bp、250bp、200bp、150bp、140bp、130bp、120bp、110bp、100bp、70bp或50bp的序列,与在细胞中或体内未发现与所述调控元件相关或者通常与所述调控元件连接的转基因可操作地连接。

[0016] 在一些情况下,本文所述的表达盒或载体包含调控元件,其是内含子序列、启动子序列、增强子序列或其组合。在一些情况下,所述表达盒或载体包含调控元件,其为SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41或其组合,或与其具有至少80%、至少85%、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%或至少99%序列同一性的序列。在一些情况下,所述表达盒或载体包含转基因,其为报道基因,例如萤光素酶。在一些情况下,使用报道基因(例如,萤光素酶)作为转基因导致在整个小鼠中测量的大于 1×10^8 光子/秒的全局表达,或导致在动物(例如,小鼠)中的至少2、3、4、5、6、7、8、9、10或更多种细胞类型中的转基因表达增加。在一些情况下,将本文所述的表达盒或载体的活性或转基因表达相对于递送至小鼠的表达盒、载体、病毒或DNA的剂量,如每只小鼠12μg的剂量或10-20μg的表达载体的剂量标准化。在一些情况下,本文公开的转基因的内源形式在细胞中或体内不与调控元件连接,或者在细胞中或体内通常不与调控元件连接或相关联。在一些情况下,所述转基因的全局表达处于高于使用具有选自以下的调控元件的载体的所述转基因表达的水平:CMV启动子、CMVe启动子、超级核心启动子、TTR启动子、Proto 1启动子和UCL-HLP启动子。在一些情况下,可在体内(例如,在小鼠或在不同的人体细胞系中)的至少2、3、4、5、6、7种或更多种不同的细胞类型中检测到表达。在一些情况下,所述不同的细胞类型选自:肺泡

细胞、心肌细胞、上皮细胞、肝细胞、肠细胞、肌细胞、神经元和肾脏细胞。在一些情况下,本文公开的调控元件是或包含增强子、启动子、5' UTR序列、内含子序列或其任何组合。在一些情况下,本文公开的调控元件是或包含启动子序列。在一些情况下,调控元件包含启动子,其为CMV启动子、CMV、启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子、AAT启动子、KAR启动子、EF1 α 启动子、EFS启动子或CMVe增强子/CMV启动子组合,或其任何组合、变体或片段。

[0017] 在一些情况下,本文公开的表达载体或盒进一步包含一个或多个转录后修饰位点。在一些情况下,所述表达载体或盒包含转基因,其为Cas9、saCas9或dCas9。在一些情况下,所述表达载体或盒包含转基因,其为DNA结合蛋白或转录调节因子,例如转录激活因子。在一些情况下,这样的转录调节蛋白作用于内源基因,以增加所述内源基因在细胞中或体内的表达。

[0018] 在其他情况下,表达载体包含与本文公开的调控序列可操作地连接的因子VIII转基因,该调控序列能够驱动因子VIII的表达至>1.0IU/mL的浓度,如通过被配置用于在体内检测因子VIII的ELISA测定所测量的。

[0019] 在一些情况下,本文公开的任何实施方案中的转基因是与单倍性不足相关联的基因。在一些情况下,这样的单倍性不足是在肝脏或肝细胞内。在一些情况下,所述转基因是以下中的任一种:因子VIII、Cas9、激素、生长或分化因子、胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体、囊性纤维化跨膜调节因子、与粘多糖贮积症I、II、III或IV型相关的基因、 β 珠蛋白或脂蛋白脂肪酶,或其变体、亚基或功能片段。在一些情况下,所述治疗性转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体、亚基或功能片段。在一些情况下,所述治疗性转基因是CDKL5、CNTNAP2、ZEB2,或其变体或功能片段。在一些情况下,所述转基因是纤维蛋白原、凝血酶原、凝固因子或凝血因子(例如,因子1、2、3、4、5、6、7、8、9、10、11或12),或其变体或功能片段。

[0020] 在一些情况下,本文所述的表达载体或盒中的任一种是病毒颗粒,或使用病毒颗粒递送,或被衣壳化或提供在病毒颗粒中。在一些情况下,所述病毒颗粒是腺伴随病毒或AAV。在一些情况下,所述AAV选自:AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ和scAAV。在其他情况下,所述病毒颗粒是慢病毒。在一些情况下,所述病毒颗粒是腺病毒。

[0021] 在一些情况下,本文公开的表达载体或盒中的任一种包含转基因,其为治疗性转基因。在一些情况下,所述治疗性转基因是因子VIII、Cas9、DNA结合蛋白、激素、生长或分化因子、胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体、囊性纤维化跨膜调节因子、与粘多糖贮积症I、II、III或IV型相关的基因、 β 珠蛋白或脂蛋白脂肪酶,或其变体或功能片段。在一些情况下,所述治疗性转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体、亚基或功能片段。在一些情况下,所述转基因是以下中的任一种:ATP7A、ATP7B、AT8B1(或ATP8B1)、MDR3(或ABCB4)、ABCB11(或ABCB11)、CDKL5、CNTNAP2(或CNTNAP2)、ZEB2、因子V、因子VIII、因子IX或因子X,或其变体、亚基或功能片段。在一些情况下,所述治疗性转基因是CDKL5、CNTNAP2、ZEB2,或其变体、亚基或功能片段。在一些情况下,

所述转基因是纤维蛋白原、凝血酶原、凝固因子或凝血因子(例如,因子1、2、3、4、5、6、7、8、9、10、11或12),或其变体或功能片段。

[0022] 在一些情况下,用于将转基因递送至动物(例如,哺乳动物、人、小鼠等)中的多种不同组织或细胞类型的方法包括向所述动物施用本文公开的表达载体中的任一种。在一些情况下,所述施用包括全身施用,例如通过口服、静脉内、肌肉内、腹膜内、鞘内、肠内或肠胃外施用。

[0023] 在一些情况下,用于产生蛋白质、抗体或其他生物制剂的方法包括使细胞与本文公开的表达载体中的任一种接触。在一些情况下,用于产生蛋白质的所述细胞是CHO细胞或HEK293T细胞。在一些情况下,本文公开的任何实施方案的表达构建体被用于基因疗法,例如,在肝细胞、肾细胞和/或神经元(或在中枢神经系统(CNS))中表达。在其他情况下,本文公开的任何实施方案的表达构建体被用于离体产生蛋白质,例如,在任何哺乳动物或任何人类细胞系中,如在肾细胞、上皮细胞、肝脏细胞、肝细胞、神经元、成纤维细胞或CNS细胞中。

[0024] 在一些情况下,用于产生转基因动物或植物的方法包括向动物或植物施用本文公开的表达载体或盒中的任一种。在一些情况下,在本文公开的表达载体或盒中的任一种中使用的调控元件相对较短,或为40-50bp、50-60bp、60-70bp、70-80bp、80-90bp、90-100bp、100-110bp、110-120bp或120-130bp。在一些情况下,本文公开的表达盒中的RE为49bp、56bp、100bp、117bp、259bp或266bp。在一些情况下,在本文公开的表达载体或盒中的任一种中使用的调控元件为45-50bp、50-60bp、45-60bp、90-117bp、95-110bp、115-120bp、250-260bp或260-270bp。在一些情况下,在本文公开的表达载体或盒中的任一种中使用的调控元件为约100bp。在一些情况下,在本文公开的表达载体或盒中的任一种中使用的调控元件小于或等于300bp、250bp、200bp、150bp、140bp、130bp、120bp、110bp、100bp、70bp或50bp。在一些情况下,在本文公开的表达载体或盒中的任一种中使用的调控元件为100bp或更小。在一些情况下,本文公开的两个或更多个相对较短的RE被组合,并与转基因可操作地连接以驱动所述转基因的高表达。在一些情况下,可以将两个或更多个RE组合而没有任何间插核苷酸,或使用1-50bp的连接体进行组合。

[0025] 在一些方面,本文公开的表达盒包含与治疗性转基因可操作地连接的来源于人类的调控元件,其中所述调控元件包含以下一个或多个:(i) SEQ ID NO:1-2、13-17和22-41;(ii) 其片段或组合;或(iii) 与(i)和(ii)中的任一个具有至少80%序列同一性的序列。在一些情况下,所述调控元件来源于hg19的人类序列。在一些情况下,所述调控元件是非天然存在的。在一些情况下,所述调控元件包含内含子序列。在一些情况下,所述内含子序列位于启动子和所述治疗性转基因之间。在一些情况下,所述调控元件包含启动子序列。在一些情况下,所述调控元件是所述盒中唯一的启动子。在一些情况下,所述调控元件不超过100bp。在一些情况下,所述调控元件不超过60bp。在一些情况下,所述调控元件不超过50bp。在一些情况下,所述表达盒是rAAV的一部分。在一些情况下,所述rAAV是rAAV8。在一些情况下,所述治疗性转基因是ATP7B。在一些情况下,所述治疗性转基因是因子VIII。在一些方面,治疗威尔逊病(Wilson's disease)的方法包括施用本文公开的表达盒,并且其中所述转基因是ATP7B。在一些情况下,治疗单倍性不足或ATP7B中的遗传缺陷的方法包括施用本文公开的表达盒,并且其中所述转基因是ATP7B。在一些情况下,治疗凝血障碍的方法

包括施用本文公开的表达盒,并且其中所述转基因是因子VIII。在一些情况下,治疗单倍性不足或遗传缺陷的方法包括施用本文公开的表达盒,并且其中所述转基因选自:ATP7A;ATP7B;ATP8B1;ABCB4;ABCB11;CDKL5;CNTNAP2;ZEB2;因子1、2、3、4、5、6、7、8、9、10、11和12;及其变体或功能片段。在一些情况下,治疗单倍性不足或遗传缺陷的方法包括施用本文公开的表达盒,并且其中所述转基因是调节内源基因表达的转录调节因子。在一些情况下,治疗单倍性不足或遗传缺陷的方法包括施用本文公开的表达盒,并且其中所述转基因是以下内源基因中的任一种的转录激活因子:ATP7A;ATP7B;ATP8B1;ABCB4;ABCB11;CDKL5;CNTNAP2;ZEB2;以及因子1、2、3、4、5、6、7、8、9、10、11和12。

[0026] 在一些方面,AAV表达盒包含与至少3kb的转基因可操作地连接的不超过120bp的来源于人类的调控元件,其中与当可操作地连接CMV启动子时的转基因的表达相比,所述调控元件导致转基因表达增加至少2倍。在一些情况下,所述调控元件选自:SEQ ID NO:1-2、13-17和22-41。在一些情况下,所述调控元件包含(i) SEQ ID NO:1-2、13-17和22-41;(ii) 其组合;或(iii) 与(i)和(ii)中的任一个具有至少80%序列同一性的序列。在一些情况下,所述增加的转基因表达是至少50倍。在一些情况下,所述增加的转基因表达是至少100倍。在一些情况下,所述增加的转基因表达在至少2种不同的细胞类型(例如,肝细胞和肾细胞)中发生。在一些情况下,所述增加的转基因表达在至少3种不同的细胞类型(例如,肝细胞、肾细胞、神经元和/或上皮细胞)中发生。在一些情况下,所述调控元件包含SEQ ID NO:22-41中的任一个或多个,并且其中所述表达盒中不存在其他启动子序列。在一些情况下,所述调控元件位于所述转基因的上游。在一些情况下,所述调控元件包含SEQ ID NO:1和SEQ ID NO:2中的一个或多个。在一些情况下,所述调控元件位于启动子的下游。在一些情况下,所述转基因选自:ATP7A;ATP7B;ATP8B1;ABCB4;ABCB11;CDKL5;CNTNAP2;ZEB2;因子1、2、3、4、5、6、7、8、9、10、11和12;及其变体或功能片段。在一些情况下,所述转基因是ATP7A或ATP7B,或其变体或功能片段。在一些情况下,所述转基因是因子1、2、3、4、5、6、7、8、9、10、11和12中的任一个,或其变体或功能片段。在一些情况下,所述转基因是因子8,或其变体或功能片段。在一些情况下,所述转基因是基因编辑蛋白。在一些情况下,所述基因编辑蛋白是Cas(例如,Cas9)。在一些情况下,所述转基因是DNA结合蛋白。在一些情况下,所述转基因是增加内源基因表达的转录激活因子。在一些情况下,所述转基因是减少内源基因表达的转录抑制因子。在一些情况下,所述转录激活因子增加以下内源基因中的任一种的表达:ATP7A;ATP7B;ATP8B1;ABCB4;ABCB11;CDKL5;CNTNAP2;ZEB2;以及因子1、2、3、4、5、6、7、8、9、10、11和12。在一些情况下,所述AAV选自:AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ和scAAV。

[0027] 在一些方面,产生重组蛋白的方法包括将编码所述蛋白质的序列与以下一种或多种可操作地连接:(i) SEQ ID NO:1-2、13-17和22-41;(ii) 其组合;或(iii) 与(i)和(ii)中的任一个具有至少80%序列同一性的序列。

[0028] 在一些方面,治疗肝脏疾病或病况的方法包括施用基因疗法,所述基因疗法包含与选自以下的一个或多个调控元件可操作地连接的治疗性转基因:(i) SEQ ID NO:1-2、13-17和22-41;(ii) 其组合;或(iii) 与(i)和(ii)中的任一个具有至少80%序列同一性的序列。在一些情况下,所述肝脏疾病或病况是威尔逊病。在一些情况下,所述肝脏疾病或病况是凝血障碍。在一些情况下,所述转基因是ATP7A或ATP7B,或其变体或功能片段。在一些情

况下,所述转基因是因子1、2、3、4、5、6、7、8、9、10、11、12,或其变体或功能片段。在一些情况下,所述转基因是因子8,或其变体或功能片段。在一些情况下,所述调控元件导致至少2种细胞类型中增加的转基因表达。在一些情况下,所述调控元件导致至少3种细胞类型中增加的转基因表达。在一些情况下,所述调控元件导致肝细胞中增加的转基因表达。在一些情况下,所述调控元件导致与当可操作地连接CMV启动子时的转基因表达相比至少2倍水平的转基因表达增加。在一些情况下,所述基因疗法是AAV。在一些情况下,所述AAV是AAV8。

[0029] 在一些方面,表达载体包含与转基因可操作地连接的具有小于或等于100bp的来源于人类的调控元件,其中与不具有所述调控元件的第二表达载体相比,所述调控元件使所述转基因的全局表达增加至少两倍。在一些情况下,所述第二表达载体包含与CMV启动子可操作地连接的转基因。

[0030] 在一些方面,表达载体包含与转基因可操作地连接的具有小于或等于100bp的来源于人类的调控元件,借此所述转基因的表达高于由CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子表达的相同转基因的表达。在一些情况下,所述转基因的表达高于由UCL-HLP启动子表达的相同转基因的表达。在一些方面,表达载体包含与转基因可操作地连接的来源于人类的调控元件,其中由所述转基因编码的蛋白质具有(i)通过被配置用于检测所述转基因的ELISA测定所测量的 $>1.0\text{IU/mL}$ 的浓度;和/或(ii)通过Coatest所测量的 $>25\%$ 的活性。在一些方面,本文公开的载体包含来源于人类的调控元件,其具有大小为小于或等于100bp大小的序列,与在细胞中或体内未发现与所述调控元件相关的转基因可操作地连接。在一些情况下,本文公开的载体包含调控元件,其为内含子序列。在一些情况下,所述调控元件是SEQ ID NO:1或SEQ ID NO:2,或与其具有至少80%同源性的序列。在一些情况下,使用萤光素酶作为转基因导致在整个小鼠中测量的大于 1×10^8 光子/秒的全局表达。在一些情况下,所述活性或转基因表达对应于每只小鼠 $16\mu\text{g}$ 表达载体的剂量。在一些情况下,所述活性或转基因表达对应于每只小鼠 $12\mu\text{g}$ 表达载体的剂量。在一些情况下,所述转基因的内源形式在体内不与所述调控元件连接。在一些情况下,所述转基因的全局表达处于高于使用带有选自以下的调控元件的载体的转基因表达的水平:CMV启动子、CMVe启动子、超级核心启动子、TTR启动子、Proto 1启动子和UCL-HLP启动子。在一些情况下,在小鼠体内或细胞系中的至少2、3、4、5、6或7种不同的细胞类型中可检测到表达。在一些情况下,所述不同的细胞类型选自:肺泡细胞、心肌细胞、上皮细胞、肝细胞、肠细胞、肌细胞、神经元和肾脏细胞。在一些情况下,所述调控元件是增强子。在一些情况下,本文公开的载体进一步包含启动子。在一些情况下,所述启动子是CMV启动子、CMV、启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子、AAT启动子、KAR启动子、EF1 α 启动子、EFS启动子或CMVe增强子/CMV启动子组合。在一些情况下,本文公开的载体进一步包含一个或多个转录后修饰位点。在一些情况下,所述转基因是Cas9。在一些情况下,所述转基因是saCas9。

[0031] 在一些方面,表达载体包含与调控序列可操作地连接的因子VIII转基因,所述调控序列能够驱动因子VIII的表达至 $>1.0\text{IU/mL}$ 的浓度,如通过被配置用于在体内检测因子VIII的ELISA测定所测量的。在一些情况下,病毒颗粒包含本文公开的表达载体。在一些情况下,所述病毒颗粒是AAV。在一些情况下,所述AAV选自:AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ和scAAV。在一些情况下,所述病

毒颗粒是慢病毒。在一些情况下,所述病毒颗粒是腺病毒。在一些情况下,所述转基因是治疗性转基因。在一些情况下,所述治疗性转基因是因子VIII、Cas9、DNA结合蛋白、激素、生长或分化因子、胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体、囊性纤维化跨膜调节因子、与粘多糖贮积症I、II、III或IV型相关的基因、 β 珠蛋白或脂蛋白脂肪酶。在一些情况下,所述治疗性转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体或功能片段。在一些情况下,所述治疗性转基因是CDKL5、CNTNAP2、ZEB2,或其变体或功能片段。在一些情况下,用于将转基因递送至动物中的多种不同组织或细胞类型的方法包括向所述动物施用本文公开的表达载体。在一些情况下,用于产生蛋白质的方法包括用本文公开的表达载体转染细胞。在一些情况下,所述细胞是CHO细胞或HEK293T细胞。在一些情况下,用于产生转基因动物或植物的方法包括向动物或植物施用本文公开的任何实施方案的表达载体。在一些情况下,所述调控元件为40-50bp。在一些情况下,所述调控元件为50-60bp。

[0032] 在一些方面,表达载体包含与转基因可操作地连接的来源于人类的调控元件,其中由所述转基因编码的蛋白质具有(i)通过被配置用于检测所述转基因的ELISA测定所测量的 $>1.0\text{IU/mL}$ 的浓度;和/或(ii)通过Coatest所测量的 $>10\%$ 的活性。在其他方面,表达载体包含与转基因可操作地连接的具有 $<120\text{bp}$ 的来源于人类的调控元件,借此所述转基因的表达高于由UCL-HLP启动子表达的相同转基因的表达。在一些情况下,所述调控元件是启动子。在一些情况下,所述治疗性转基因是包含DNA结合域和转录调节域的融合蛋白。

[0033] 援引并入

[0034] 本说明书中提到的所有出版物、专利和专利申请均通过引用并入本文,其程度如同明确且单独地指出每个单独的出版物、专利或专利申请通过引用并入。

附图说明

[0035] 本发明的新颖性特征在所附的权利要求书中具体阐述。通过参考以下对利用到本发明原理的说明性实施方案加以阐述的详细描述以及附图,将获得对本发明的特征和优点的更好的理解,在这些附图中:

[0036] 图1A示出了包含不同调控元件(例如,SEQ ID NO:1和SEQ ID NO:2)的质粒的标准化萤光素酶值,证明了RE对HEK293T细胞中萤光素酶表达的影响。例如,SEQ ID NO:3(包含与最小CMV(minCMV)启动子组合的SEQ ID NO:1)驱动的萤光素酶表达的水平比单独的minCMV启动子驱动的表达高约1.4倍,并比SCP启动子驱动的表达高约60倍。

[0037] 图1B示出了在HEK293T细胞中,与SCP、minCMV和CAG相比,每个调控元件(例如,SEQ ID NO:3和SEQ ID NO:4)的大小标准化活性(通过用标准化萤光素酶活性除以调控元件的碱基对长度计算)。SEQ ID NO:4(包含连接至minCMV启动子的SEQ ID NO:2)导致大小标准化活性的水平比单独的minCMV启动子高约3.5倍,并比SCP启动子高约140倍。

[0038] 图1C示出了与阴性对照和阳性对照(SEQ ID NO:4)相比,调控元件SEQ ID NO:17和22组合、SEQ ID NO:16和23组合、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:32和SEQ ID NO:33的标准化萤光素酶表达。所有RE均导致标准化萤光素酶表达的高水平,与SEQ ID NO:4驱动的表达的水平相当。

[0039] 图1D示出了在HEK293T细胞中,与阴性对照、阳性对照(SEQ ID NO:4)、CMV+CMVe和单独的CMV相比,调控元件SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:26或SEQ ID NO:27的标准化萤光素酶表达。每个调控元件都驱动比单独的CMV和CMV+CMVe更高的萤光素酶表达。

[0040] 图1E示出了在HEK293T细胞中,与阴性对照相比,调控元件SEQ ID NO:17和22组合、SEQ ID NO:16和23组合、SEQ ID NO:34、SEQ ID NO:35、SEQ ID NO:36、SEQ ID NO:37、SEQ ID NO:38、SEQ ID NO:39、SEQ ID NO:40或SEQ ID NO:41的标准化萤光素酶表达。测试的每个调控元件都驱动比阴性对照更高的萤光素酶表达。

[0041] 图2A示出了在给予表达盒的小鼠中的萤光素酶的表达,该表达盒包含在不同调控元件(SCP、SerpE_TTR、Proto1、minCMV、UCL-HLP、CMVe、SEQ ID NO:3或SEQ ID NO:4)控制下的萤光素酶。

[0042] 图2B示出了在注射了图2A的表达盒的3只小鼠中观察到的来自萤光素酶表达的平均总发光(光子/秒)的定量。

[0043] 图3A示出了通过酶联免疫吸附测定所测量的各种调控元件(UCL-HLP、SEQ ID NO:3或SEQ ID NO:4)对小鼠中人类因子8(FVIII)表达的影响。包含SEQ ID NO:3或SEQ ID NO:4的表达盒导致人FVIII的水平(IU/mL)升高。

[0044] 图3B示出了通过测定人类因子8(FVIII)的活性百分比的Coatest测定所测量的各种调控元件(UCL-HLP、SEQ ID NO:3或SEQ ID NO:4)对小鼠中FVIII表达的影响。

[0045] 图4A示出了在小鼠脑的齿状回(dente gyrus)中用包含Cas9酶、指导RNA和几种调控元件(minCMV、EFS或SEQ ID NO:4)之一的单个AAV进行的基因编辑,以删除tdTomato(也称为红色荧光蛋白(RFP))上游提供的终止序列。RFP用浅灰色表示。与阴性对照、minCMV和EFS相比,包含SEQ ID NO:4的表达盒产生导致RFP更高表达的基因编辑。

[0046] 图4B示出了图4A中表达RFP的细胞的平均数目的定量。

[0047] 图5A示出了用于使用包含本公开内容的调控元件的rAAV表达盒在FVIII敲除小鼠的体内治疗凝血缺陷的基因疗法的实例。将以 3^{11} 个基因组拷贝/kg(gc/kg)的剂量或 3^{12} gc/kg的剂量用PBS对照或包含本公开内容的调控元件的rAAV治疗的小鼠的失血(g)结果与未治疗的非敲除小鼠进行比较。与对照(例如,PBS对照小鼠)相比,以 3^{12} gc/kg的剂量用rAAV基因疗法治疗小鼠导致失血减少。

[0048] 图5B示出了直至凝血的以分钟(min)为单位的小鼠的出血时间。

[0049] 图6A示出了本公开内容的示例性表达盒、载体或质粒,其包含AAV ITR-L和ITR-R以及与转基因可操作地连接的一个或多个调控元件,例如增强子、启动子、稳定元件和polyA信号。这样的rAAV载体可用于基因疗法。

[0050] 图6B示出了表达盒的特写视图,其中调控元件SEQ ID NO:1与转基因可操作地连接,并且其中调控元件位于启动子下游以及基因和多聚腺苷酸化信号上游。

[0051] 图6C示出了表达盒,其中调控元件SEQ ID NO:2与转基因可操作地连接,并且其中调控元件位于启动子下游以及基因和多聚腺苷酸化信号上游。

[0052] 图7示出了本公开内容的示例性表达盒、载体或质粒,其包含AAV ITR-L和ITR-R以及与较大转基因可操作地连接的本公开内容的一个或多个相对较短的来源于人类的调控元件(例如,SEQ ID NO:13-17和22-41),其中转基因的大小为至少3kb或3-5kb。这样的rAAV载体可用于基因疗法治疗。

[0053] 图8示出了与未治疗的敲除小鼠和SEQ ID NO:9 (UCL-HLP) 相比,在用本公开内容的各种表达盒治疗的小鼠中表达的人FVIII的浓度(IU/mL),该表达盒例如包含调控元件SEQ ID NO:13-17中的任一个的表达盒。

[0054] 图9示出了用本公开内容的各种表达盒治疗的小鼠中人FVIII的活性百分比,其中SEQ ID NO:13-17是指调控元件。

[0055] 图10示出了用本公开内容的各种表达盒治疗的小鼠的失血实验,以克(g)为单位测量,其中SEQ ID NO:13-17是指调控元件。

[0056] 图11示出了用本公开内容的各种表达盒治疗的小鼠直至凝血的以分钟(min)为单位的出血时间,其中SEQ ID NO:13-17是指调控元件。

[0057] 图12示出了以 1×10^{10} (或 10^{10}) 个基因组拷贝/小鼠(gc/小鼠)、 1×10^{11} (或 10^{11}) gc/小鼠或 1×10^{12} (或 10^{12}) gc/小鼠的剂量用表达盒治疗的小鼠中ATP7B的相对表达($\log_2$),该表达盒包含具有SEQ ID NO:13序列的调控元件。数据显示,包含SEQ ID NO:13的表达盒的剂量与体内ATP7B的表达之间呈正相关。

[0058] 图13示出了用表达盒治疗并以 5×10^{11} gc/小鼠的剂量施用的小鼠中EGFP的相对表达($\log_{10}$),该表达盒包含具有SEQ ID NO:4、13、14、15、16、17或27的序列的调控元件。数据显示,RE导致体内小鼠的肝脏中EGFP表达的水平升高,如通过RNA转录物所测量的。

具体实施方式

[0059] 本公开内容涵盖用于转基因的高表达的组合物和方法。本文还描述了用于包含一个或多个个人或来源于人类的调控元件(RE)的转基因的高表达的组合物和方法,该RE当与转基因可操作地连接时可以促进或导致转基因在一种或多种靶细胞类型或组织中的高(或增加的)表达。在一些实施方案中,本文公开的一个或多个调控元件驱动转基因在细胞中或体内、体外和/或离体的高表达。在一些实施方案中,包含与转基因可操作地连接的一个或多个本文所述的调控元件的表达盒或载体可适用于任何表达系统或基因疗法中,从而驱动转基因在细胞、受试者或动物模型中的高表达,或者导致在细胞、受试者或动物模型中的治疗有效水平的转基因表达。在一些实施方案中,本公开内容的一个或多个RE与较大转基因(例如,其序列大于1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb的转基因)可操作地连接,以导致转基因在细胞中或体内的高表达。在一些情况下,这样的转基因在细胞中或体内的高表达是相对于不具有所述RE的转基因的表达,其中与不具有所述RE的转基因表达相比,或与阴性对照(例如,单独的缓冲液、单独的载体或包含已知不具有表达活性的序列的载体)的转基因表达相比,具有所述RE的转基因的表达是至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少150倍、至少200倍、至少250倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍、至少1000倍、至少1010倍、至少1020倍、至少1030倍、至少1040倍或至少1050倍。

[0060] 在一些情况下,一个或多个RE在至少2种、至少3种、至少4种、至少5种、至少6种、至少7种、至少8种、至少9种或至少10种不同的细胞类型中导致高转基因表达。在一些情况下,本公开内容的一个或多个RE与转基因可操作地连接,以用于适于全身施用的基因疗法治疗。在一些情况下,本公开内容的一个或多个RE与转基因可操作地连接,以用于适于在肝脏

或肝细胞中表达的基因疗法治疗。在一些情况下,转基因是在肝脏或肝细胞中表达的基因,或者是与肝脏疾病或病况相关的基因,例如,与威尔逊病相关联的ATP7B;与进行性家族性肝内胆汁淤积3型相关联的ABCB4;与遗传性果糖不耐受相关联的ALDOB;与糖原贮积病IV型相关联的GBE1;与酪氨酸血症I型相关联的FAH;与精氨琥珀酸裂合酶缺乏症相关联的ASL;与citrin缺乏症(CTLN2、NICCD)相关联的SLC25A13;与胆固醇酯贮积病相关联的LIPA;与 α -1抗胰蛋白酶缺乏症相关联的SERPINA1;与囊性纤维化相关联的CFTR;与遗传性血色素沉着症相关联的HFE;或与阿尔斯特伦综合征(**Alström syndrome**)相关联的ALMS1。在一些情况下,本文公开的一种或多种调控元件被用于基因疗法治疗,以治疗遗传性肝病,例如,胆汁酸合成障碍、碳水化合物代谢障碍、氨基酸代谢障碍、尿素循环障碍或脂质代谢障碍。在一些情况下,转基因是调节内源基因的DNA结合蛋白,例如,转录调节因子。在一些情况下,本文公开的任何实施方案的RE包含启动子序列、内含子序列、5' UTR序列或其任何组合。在一些情况下,本文公开的任何实施方案的RE是来源于人类的序列(或与人类参考基因组中的序列具有至少80%、90%、95%或99%序列同一性的序列)。在一些情况下,本文公开的任何实施方案的RE的长度小于或等于50bp、60bp、100bp、120bp、260bp或270bp。在一些情况下,本文公开的任何实施方案的RE小于或等于300bp、250bp、200bp、150bp、140bp、130bp、120bp、110bp、100bp、70bp或50bp。在一些情况下,表达盒(例如,载体、AAV或病毒载体)包含与本公开内容的转基因可操作地连接的一个或多个RE序列,其中每个RE序列是表1所列的序列中的任一个,或是与表1的序列具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性(例如,局部序列同一性)的序列。在一些情况下,表达构建体(例如,载体、AAV或病毒载体)包含与本公开内容的转基因可操作地连接的一个或多个RE序列,其中每个RE序列不超过50bp、60bp、70bp、80bp、90bp、100bp、110bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、260bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp,并且包含根据表1的序列中的任一个或多个的序列。在一些情况下,表达盒(例如,载体、AAV或病毒载体)包含与本公开内容的转基因可操作地连接的一个或多个RE序列,其中每个RE序列不超过50bp、60bp、70bp、80bp、90bp、100bp、110bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、260bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp,并且包含与表1所列的序列中的任一个具有至少70%、至少75%、至少80%、至少85%、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%或至少99%局部序列同一性的序列。在一些情况下,使用BLAST测量局部序列同一性。在一些情况下,本文公开的任何实施方案的RE序列位于转基因的上游或下游。在一些情况下,本文公开的任何实施方案的两个或更多个RE与转基因可操作地连接,其中至少一个或多个RE位于转基因的上游、下游或者上游和下游。

[0061] 在一些情况下,本公开内容的任何实施方案的一个或多个RE与转基因可操作地连接,以驱动转基因的全局表达,其中全局表达是指至少在2、3、4、5、6、7、8、9或10种不同细胞类型中的转基因表达。在一些情况下,本公开内容的任何实施方案的一个或多个RE与转基因可操作地连接,以在全身递送至受试者或动物中时驱动转基因的全局表达。在一些情况

下,本文公开的表达盒的全身递送或施用是通过口服、静脉内、肌肉内、腹膜内、鞘内、肠和/或肠胃外施用的递送。在各种实施方案中,本文公开的任何实施方案的RE是来源于人类的(或是与人类参考基因组中的序列具有至少80%、90%、95%或99%序列同一性的序列)。在一些情况下,本文公开的任何实施方案的RE是非天然存在的。

[0062] 基因工程方法可以替代基因、修饰基因或添加基因以实现期望的结果。基因疗法是利用基因工程来治疗疾病或病症。基因工程也可以用于修饰细胞或生物体,以产生一种或多种感兴趣的蛋白质。基因工程的一个挑战是具有调控元件,该调控元件以足够高的水平驱动表达以实现治疗效果,或者以足够高的水平驱动表达以实现工业/生产水平。基因工程的另一个挑战是设计表达盒,其足够小以配合在期望的载体内,但包含感兴趣的蛋白质的完整编码序列和足够的调控序列以确保高表达。载体或病毒表达载体的克隆能力是较大转基因表达的特殊挑战。例如,AAV载体具有~4.8kb的包装能力,慢病毒具有~8kb的能力,腺病毒具有~7.5kb的能力,并且甲病毒具有~7.5kb的能力。一些病毒具有更大的包装能力,例如疱疹病毒具有>30kb的能力,而牛痘具有~25kb的能力,但是使用其他病毒可能有一定优势。例如,AAV的优势包括低致病性、整合到宿主基因组中的频率极低以及感染分裂和非分裂细胞的能力。

[0063] 基因疗法的一个挑战是确保转基因在至少一种或多种靶细胞类型或组织中以足够高的水平表达,以在体内产生治疗效果。在一些情况下,期望在至少2、3、4、5、6、7、8、9、10种或更多种不同的细胞类型或组织中增加转基因表达,该细胞类型或组织例如,肾细胞、上皮细胞、肝脏细胞、肝细胞、神经元、成纤维细胞和/或CNS细胞。在一些情况下,期望在离体的至少1、2、3、4、5、6、7、8、9或10种细胞系中增加转基因表达以供蛋白质合成,该细胞系例如,哺乳动物或人类细胞系,如肾细胞、上皮细胞、肝脏细胞、肝细胞、神经元、成纤维细胞、免疫细胞和/或CNS细胞。用于基因疗法的传统方法通常依赖于递送方法和/或媒介物(例如,改变所使用的病毒或病毒的衣壳序列)。该领域的一项技术挑战是提高基因表达在一种或多种细胞类型或组织中的水平,特别是当转基因较大时,以发挥治疗作用。较大转基因的表达与技术挑战相关联,因为载体(例如,病毒或AAV载体)具有有限的克隆能力。为了表达较大的转基因(例如,序列大于1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb的转基因),必须使用较小的调控元件以在载体中容纳较大的转基因。因此,需要新型短调控元件(例如,小于或等于300bp、250bp、200bp、150bp、140bp、130bp、120bp、110bp、100bp、70bp或50bp的RE),其与没有RE的表达、阴性对照(例如,已知不驱动任何表达的序列、空载体或单独的缓冲液对照),或与可操作地连接至转基因的启动子(例如,CMV启动子、SCP、UCL-HLP等)相比,导致在细胞中或体内的转基因表达增加。

[0064] 如本文所用,除非上下文另外明确指出,否则单数形式“一个”、“一种”和“该”意在包括复数个指代物。此外,就具体实施方式和/或权利要求中使用的术语“包括”、“具有”、“带有”或其变体而言,这些术语旨在是包含性的,类似于术语“包含”。

[0065] 术语“约”或“大约”是指在本领域普通技术人员确定的特定值的可接受误差范围内,该误差范围将部分取决于如何测量或确定该值,即测量系统的局限性。例如,按照本领域的实践,“约”可以指在一个或多个标准偏差内。或者,“约”可以指给定值的至多20%、至多15%、至多10%、至多5%或至多1%的范围。

[0066] 术语“确定”、“测量”、“评价”、“评估”、“测定”、“分析”及其语法等同物在本文中可

以互换使用,以指代任何形式的测量,并且包括确定元素是否存在(例如检测)。这些术语可以包括定量和/或定性确定。评估可以是相对的,也可以是绝对的。

[0067] 术语“表达”是指从DNA模板转录核酸序列或多核苷酸的过程(诸如转录成mRNA或其他RNA转录物)和/或转录的mRNA被随后翻译成肽、多肽或蛋白质的过程。转录物和编码的多肽可统称为“基因产物”。如果多核苷酸来源于基因组DNA,则表达可包括在真核细胞中剪接mRNA。在一些情况下,全局表达是期望的,并且是指基因(例如,转基因)在至少2、3、4、5、6、7、8、9、10种或更多种不同的细胞类型或组织中的表达,或者基因在多种细胞类型中的表达。通过具有不同的细胞标志物、形态、表型、基因型、功能和/或用于分类细胞类型的任何其他方式来辨别细胞类型。在一些情况下,不同的细胞类型或组织是指体内的细胞类型或组织。在一些情况下,不同的细胞类型或组织是指离体的细胞类型或组织。在一些情况下,不同的细胞类型或组织包括,但不限于,不同的哺乳动物细胞系、人类细胞系、肾细胞、上皮细胞、肝脏细胞、肝细胞、神经元、成纤维细胞和/或CNS细胞。

[0068] 如本文所用,“可操作地连接”、“可操作的连接”或其语法等同物是指例如启动子、增强子、多聚腺苷酸化序列等遗传元件的并列,其中元件之间的关系允许它们以预期的方式发挥作用。例如,如果可包含启动子和/或增强子序列的调控元件帮助起始编码序列的转录,则该调控元件可操作地连接至编码区。只要维持这种功能关系,在调控元件与编码区之间可以有间隔的残基。

[0069] 如本文所用,“载体”是指包含多核苷酸或与多核苷酸相关联,且可用于介导多核苷酸向细胞递送的大分子或大分子缔合物。载体的实例包括质粒、病毒载体、脂质体和其他基因递送媒介物。载体通常包含与基因可操作地连接的遗传元件,例如调控元件,以促进基因在靶标中的表达。调控元件和与其可操作地连接以供表达的一个或多个基因的组合被称为“表达盒”。

[0070] 术语“AAV”是腺伴随病毒的缩写,可用于指代该病毒本身或其衍生物。该术语涵盖所有血清型、亚型、以及天然存在形式和重组形式,除非另有要求。缩写“rAAV”是指重组腺伴随病毒,也称为重组AAV载体(或“rAAV载体”)。术语“AAV”包括AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、rh10及其杂合物,禽AAV、牛AAV、犬AAV、马AAV、灵长类动物AAV、非灵长类动物AAV和羊AAV。AAV的各种血清型的基因组序列,以及天然末端重复序列(TR)、Rep蛋白和衣壳亚基的序列是本领域已知的。这样的序列可以在文献中或公共数据库如GenBank中找到。如本文所用,“rAAV载体”是指包含非AAV来源的多核苷酸序列(即与AAV异源的多核苷酸)的AAV载体,该序列通常是用于细胞遗传转化的感兴趣的序列。通常,异源多核苷酸的侧翼是至少一个,并且通常是两个,AAV反向末端重复序列(ITR)。术语rAAV载体涵盖rAAV载体颗粒和rAAV载体质粒二者。rAAV载体可以是单链的(ssAAV)或自互补的(scAAV)。“AAV病毒”或“AAV病毒颗粒”或“rAAV载体颗粒”是指由至少一个AAV衣壳蛋白和衣壳化的多核苷酸rAAV载体组成的病毒颗粒。如果颗粒包含异源多核苷酸(即野生型AAV基因组以外的多核苷酸,如要递送至哺乳动物细胞的转基因),则其通常称为“rAAV载体颗粒”或简称为“rAAV载体”。因此,rAAV颗粒的产生必然包括rAAV载体的产生,因为这种载体包含在rAAV颗粒内。

[0071] 如本文所用,术语“治疗”、“疗法”等是指获得所需的药理和/或生理作用,包括但不限于缓和、延迟或减慢进展,减少影响或症状,预防发作,抑制、缓解疾病或病症的发作,

获得关于疾病、病症或医学状况的有益或期望的结果,如治疗益处和/或预防益处。如本文所用,“治疗”涵盖对哺乳动物特别是人类中的疾病的任何治疗,并且包括:(a)在可能易患疾病或有获得该疾病的风险但尚未被诊断为患有该疾病的受试者中预防该疾病的发生;(b)抑制疾病,即阻止其发展;以及(c)减轻疾病,即引起疾病消退。治疗益处包括根除或缓解所治疗的潜在病症。同样,通过消除或缓解与潜在病症相关的一种或多种生理症状而获得治疗益处,使得在受试者中观察到改善,尽管受试者可能仍患有该潜在病症。在一些情况下,为了预防益处,将组合物施用于有发生特定疾病的风险的受试者,或报告疾病的一种或多种生理症状的受试者,即使可能尚未对该疾病作出诊断。本公开内容的方法可以用于任何哺乳动物。在一些情况下,该治疗可导致症状减少或停止(例如,癫痫发作的频率或持续时间减少)。预防作用包括延迟或消除疾病或病况的出现,延迟或消除疾病或病况的症状的发作,减慢、停止或逆转疾病或病况的进展,或其任何组合。

[0072] 术语“有效量”或“治疗有效量”是指本文所述的组合物的量,其足以影响预期的应用,包括但不限于如下定义的疾病治疗。治疗有效量可以根据预期的治疗应用(细胞中或体内)或所治疗的受试者和疾病状况而变化,例如,受试者的体重和年龄、疾病状况的严重程度、施用方式等,这很容易由本领域普通技术人员确定。该术语还适用于将在靶细胞中诱导特定应答的剂量。具体剂量将根据所选的特定组合物、要遵循的给药方案、是否与其他化合物联合施用、施用时间、该剂量所施用至的组织以及携带该剂量的物理递送系统而变化。

[0073] 核苷酸或肽序列的“片段”是指小于被认为是“全长”序列的序列。

[0074] 分子的“变体”是指此类序列的等位基因变异,即,在结构和生物活性上与整个分子或其片段基本相似的序列。如本文所用,基因的变体是指与最常见的野生型DNA序列(例如,cDNA,或由其GenBank引用的序列或与由其UniProt登录号引用的蛋白质序列相对应的序列)相比具有遗传改变或突变的序列。基因变体包括突变体,如与野生型序列相比具有增强的活性或功能的突变体;蛋白质的任何已知亚基;以及由可变剪接产生的基因的已知同种型。

[0075] 术语“功能片段”旨在包括分子的“片段”、“变体”、“类似物”或“化学衍生物”。DNA或蛋白质序列的“功能片段”具有该序列的至少生物活性片段,该生物活性片段是指保留与全长DNA或蛋白质序列的生物活性基本相似的生物活性(在功能或结构上)的片段。DNA序列的生物活性可以是其以已知归因于全长序列的方式影响表达的能力。例如,调控元件的功能片段将保留其作为全长RE影响转录的能力。

[0076] 术语“受试者”和“个体”在本文可互换使用,以指脊椎动物,优选为哺乳动物,更优选为人类。“受试者”是指动物,诸如哺乳动物,例如人类。本文所述的方法可用于人类治疗、兽医应用和/或疾病或病况的动物模型的临床前研究。在一些情况下,受试者是哺乳动物,并且在一些情况下,受试者是人类。

[0077] 术语“体内”是指在受试者体内发生的事件。

[0078] 术语“体外”是指在受试者体外发生的事件。例如,体外测定涵盖在受试者体外运行的任何测定。体外测定包括其中使用活细胞或死细胞的基于细胞的测定。体外测定还包括其中不使用完整细胞的无细胞测定。

[0079] 诸如出于评估同一性、突变或测试序列的一个或多个位置相对于参考序列的一个或多个指定位置所处的位置的目的,可以通过任何合适的比对算法来进行序列比较,包

括但不限于Needleman-Wunsch算法(参见例如, www.ebi.ac.uk/Tools/psa/emboss_needle/上可用的EMBOSS Needle比对器, 任选使用默认设置)、BLAST算法(参见例如, blast.ncbi.nlm.nih.gov/Blast.cgi上可用的BLAST比对工具, 任选使用默认设置)以及Smith-Waterman算法(参见例如, www.ebi.ac.uk/Tools/psa/emboss_water/上可用的EMBOSS Water比对器, 任选使用默认设置)。可以使用所选算法的任何合适参数, 包括默认参数, 来评估最佳比对。

[0080] 通常, 可以互换使用的“序列同一性”或“序列同源性”是指两个多核苷酸或多肽序列的精确核苷酸-核苷酸或氨基酸-氨基酸对应性。通常, 用于确定序列同一性的技术包括确定多核苷酸的核苷酸序列和/或确定由此编码的氨基酸序列, 并将这些序列与第二核苷酸或氨基酸序列进行比较。两个或更多个序列(多核苷酸或氨基酸)可以通过确定它们的“同一性百分比”, 也称为“同源性百分比”, 来进行比较。与可能是较长分子(例如, 多核苷酸或多肽)内的序列的参考序列(例如, 核酸或氨基酸序列)的同一性百分比可以计算为两个最佳比对序列之间精确匹配的数目除以参考序列的长度并乘以100。同一性百分比还可以例如通过使用高级BLAST计算机程序比较序列信息来确定, 该计算机程序包括2.2.9版, 可从美国国立卫生研究院获得。BLAST程序基于Karlin和Altschul, *Proc. Natl. Acad. Sci. USA* 87:2264-2268 (1990)的比对方法, 有关讨论见Altschul等人, *J. Mol. Biol.* 215:403-410 (1990); Karlin和Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5877 (1993); 和Altschul等人, *Nucleic Acids Res.* 25:3389-3402 (1997)。简而言之, BLAST程序将同一性定义为相同比对符号(即核苷酸或氨基酸)的数目除以两个序列中较短者的符号总数。该程序可以用于确定被比较序列的整个长度上的同一性百分比。提供默认参数以优化短查询序列的搜索, 例如使用blastp程序。该程序还允许使用SEG过滤器来屏蔽查询序列区段, 如由Wootton和Federhen, *Computers and Chemistry* 17:149-163 (1993)的SEG程序确定的。所需的序列同一性程度的范围约为80%至100%, 以及介于两者之间的整数值。通常, 所公开的序列与要求保护的序列之间的同一性百分比为至少80%、至少85%、至少90%、至少95%、至少98%或至少99%。通常, 精确匹配表示在参考序列的长度上100%的同一性。在一些情况下, 提及序列同一性百分比是指使用BLAST(基本局部比对搜索工具)测量的序列同一性。在其他情况下, ClustalW可用于多序列比对。

[0081] 除非另有说明, 否则本文所用的所有术语具有对于本领域技术人员而言相同的含义, 并且本发明的实践将采用分子生物学、微生物学和重组DNA技术中的常规技术, 这些技术均属于本领域技术人员的知识范畴。

[0082] 调控元件

[0083] 调控元件可以在DNA和/或RNA水平上起作用。调控元件可以在感兴趣的细胞类型中起到调节基因表达选择性的作用。调控元件可以在基因表达的转录阶段、转录后阶段或翻译阶段起到调节基因表达的作用。调控元件包括但不限于启动子、增强子、内含子或其他非编码序列。在RNA水平下, 调控可以在翻译(例如, 稳定mRNA以进行翻译的稳定性元件)、RNA切割、RNA剪接和/或转录终止的水平下发生。在一些情况下, 调控元件可以将转录因子募集到编码区, 该转录因子增加感兴趣的细胞类型中基因表达的选择性。在一些情况下, 调控元件可以提高产生RNA转录物的速率, 提高产生的RNA的稳定性, 和/或提高从RNA转录物合成蛋白质的速率。

[0084] 调控元件是能够影响(例如,增加)基因(例如,报道基因,如EGFP或萤光素酶;转基因;或治疗基因)在一种或多种细胞类型或组织中的表达的核酸序列或遗传元件。在一些情况下,调控元件可以是转基因、内含子、启动子、增强子、UTR、反向末端重复序列(ITR)、长末端重复序列(LTR)、稳定性元件、翻译后响应元件或polyA序列或其组合。在一些情况下,调控元件是启动子、增强子、内含子序列或其组合。在一些情况下,调控元件来源于人类序列(例如,hg19)。

[0085] 一种分离调控元件的方法包括从来自动物模型的一种或多种细胞类型分离细胞核,这可以通过以下过程来实现:使用分离一种或多种组织或细胞类型的亲和纯化方法(例如,结合某些细胞类型的珠子);使用高通量天然引发和DNA合成以从细胞核中的开放染色质区域生成序列库;对序列库进行测序以鉴别在一种或多种组织或细胞类型中驱动基因表达的推定序列;以及在一种或多种体外细胞系和/或动物模型的报道基因系统中表达。在一些情况下,将两个或更多个调控元件组合,以形成更大的RE。在一些情况下,组合的调控元件在多种细胞类型,例如,至少2、3、4、5、6或更多种细胞类型中表现出增强的基因表达。在一些情况下,将调控元件一次截短一个或多个碱基,以确定保留其增加转基因表达能力的最小序列量。保留转基因表达活性的较小调控元件有助于形成包含较大转基因的基因疗法,或者有助于解决载体或质粒的克隆能力因希望使用基因疗法递送的转基因大小而受到限制的情况。例如,本公开内容的一个或多个RE当与转基因可操作地连接时驱动增加的转基因表达,其中该转基因为至少1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb。在一些情况下,与阴性对照、组成型启动子、CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子相比,这样的转基因表达的增加是至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍。

[0086] 在一些情况下,本公开内容的调控元件导致可操作地连接的转基因的较高或增加的表达,其中该较高或增加的表达与对照例如组成型启动子、CMV启动子、CAG、超级核心启动子(SCP)、TTR启动子、Proto 1启动子、UCL-HLP启动子、minCMV、EFS或CMVe启动子相比而确定。可通过本文公开的调控元件确定较高或增加的转基因表达的其他对照包括单独的缓冲液或单独的载体。在一些情况下,阳性对照是指具有已知表达活性的RE,如SEQ ID NO:4,其可用于比较。在一些情况下,调控元件驱动与阳性对照(例如,SEQ ID NO:4或与转基因可操作地连接的已知启动子)相当的或更高的转基因表达。

[0087] 在一些情况下,可以去除SEQ ID NO:13-17的5'最末端的17个核苷酸,以形成保留表达活性的较短RE。例如,不具有5'最末端的17个核苷酸的SEQ ID NO:13-17分别是SEQ ID NO:30、26、28、23和22。

[0088] 在其他情况下,两个或更多个相对较短的RE可以组合,以形成具有高转基因表达活性和/或大小标准化的基因表达活性的较大RE。例如,可以将SEQ ID NO:17与SEQ ID NO:22组合;或可以将SEQ ID NO:16与SEQ ID NO:22组合。

[0089] 本公开内容考虑到相对较短的,来源于人类的调控元件,其增加了可操作地连接的转基因的表达。在一些情况下,短的来源于人类的RE是表1或者SEQ ID NO:1-2、13-17和

22-41中的任一个。

[0090] 表1:调控元件 (RE) 的实例

[0091]	SEQ ID NO:	序列 (5'-至-3')	长度
	1	GTAAGGTAAGAATTGAATTTCTCAGTTGAAGGATGCTTACAC TCTTGTCCATCTAG	56bp
	2	GTGTGTATGCTCAGGGGCTGGGAAAGGAGGGGAGGGAGCTC CGGCTCAG	49bp
	3	GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCG GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGT CAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCC	266bp

[0092]

SEQ ID NO:	序列 (5'-至-3')	长度
	AAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCG GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGT ACCGTAAGGTAAGAATTGAATTTCTCAGTTGAAGGATGCTTA CACTCTTGTCATCTAG	
4	GTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCG GTTTGACTCACGGGGATTTCCTCAAGTCTCCACCCCATTGACGT CAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCC AAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCG GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGT ACCGTGTGTATGCTCAGGGGCTGGGAAAGGAGGGGAGGGAG CTCCGGCTCAG	259bp
13	GTGATGACGTGTCCCATAAGGCCCTCGGTCTAAGGCTTCCC TATTTCTGTTTCGCCGGCGGCCATTTTGGGTGGAAGCGATA GCTGAGTGGCGGCGGCTGCTGATTGTGTTCTAG	117bp
14	GTGATGACGTGTCCCATACTTCCGGGTGAGGTGGGCCGGCTG TCTTGACCTTCTTTGCGGCTCGGCCATTTTGTCCAGTCAGTC CGGAGGCTGCGGCTGCAGAAGTACCGCCTGCG	117bp
15	GTGATGACGTGTCCCATAATTTTCATCTCGCGAGACTTGTGAG CGGCCATCTTGGTCCTGCCCTGACAGATTCTCCTATCGGGGT CACAGGGACGCTAAGATTGCTACCTGGACTTTC	117bp
16	GTGATGACGTGTCCCATGGCCTCATTGGATGAGAGGTCCCAC CTCACGGCCCCGAGGCGGGGCTTCTTTGCGCTTAAAGCCGAG CCGGGCCAATGTTCAAATGCGCAGCTCTTAGTC	117bp
17	GTGATGACGTGTCCCATCCCCCTCCACCCCCTAGCCCGCGG AGCACGCTGGGATTTGGCGCCCCCTCCTCGGTGCAACCTAT ATAAGGCTCACAGTCTGCGCTCCTGGTACACGC	117bp
22	CCCCCTCCACCCCCTAGCCCGCGGAGCACGCTGGGATTTGG CGCCCCCTCCTCGGTGCAACCTATATAAGGCTCACAGTCTG CGCTCCTGGTACACGC	100bp
23	GGCCTCATTGGATGAGAGGTCCCACCTCACGGCCCCGAGGCG GGGCTTCTTTGCGCTTAAAGCCGAGCCGGGCCAATGTTCAA ATGCGCAGCTCTTAGTC	100bp
24	GGGTGGGGCCCCGCGCGTATAAAGGGGGCGCAGGCGGGCTGG GCGTTCCACAGGCCAAGTGCCTGTGCTCGAGGGGTGCCGG CCAGGCCTGAGCGAGCGA	100bp
25	GGTGCGATATTCGGATTGGCTGGAGTCGGCCATCACGCTCCA GCTACGCCACTTCCTTTTCGTGGCACTATAAAGGGTGCTGCA CGGCGCTTGCACTCTCT	100bp
26	ACTTCCGGGTGAGGTGGGCCGGCTGTCTTGACCTTCTTTGCG GCTCGGCCATTTTGTCCAGTCAGTCCGGAGGCTGCGGCTGC AGAAGTACCGCCTGCG	100bp
27	GCTGAGCGCGCGCGATGGGGCGGGAGGTTTGGGGTCAAGGA GCAAACCTCTGCACAAGATGGCGGCGGTAGCGGCAGTGGCGG CGCGTAGGAGGCGGTGAG	100bp
28	ATTTTCATCTCGCGAGACTTGTGAGCGGCCATCTTGGTCCTG CCCTGACAGATTCTCCTATCGGGGTACAGGGACGCTAAGAT TGCTACCTGGACTTTC	100bp

SEQ ID NO:	序列 (5'-至-3')	长度
29	TGGGACCCCCGGAAGGCGGAAGTTCTAGGGCGGAAGTGGCC GAGAGGAGAGGAGAATGGCGGCGGAAGGCTGGATTGCGCT TGGGGCTGGGGCCGCGG	100bp
30	AAGGCCCTCGGTCTAAGGCTTCCCTATTTCCTGGTTCGCCG GCGGCCATTTTGGGTGGAAGCGATAGCTGAGTGGCGGCGGC TGCTGATTGTGTTCTAG	100bp
31	AGTGACCCGGAAGTAGAAGTGGCCCTTGCAGGCAAGAGTGC TGGAGGGCGGCAGCGGCGACCGGAGCGGTAGGAGCAGCAA TTTATCCGTGTGCAGCCCC	100bp
32	GGGAGGGGCGCGCTGGGGAGCTTCGGCGCATGCGCGCTGAG GCCTGCCTGACCGACCTTCAGCAGGGCTGTGGCTACCATGTT CTCTCGCGCGGGTGTCTG	100bp
33	ACTGCGCACGCGCGCGGTTCGCACCGATTACGCCCCCTTCCG GCGCCTAGAGCACCGCTGCCGCCATGTTGAGGGGGGGACCG CGACCAGCTGGGCCCCCT	100bp
34	CCCTCGAGGGGCGGAGCAAAAAGTGAGGCAGCAACGCCTCC TTATCCTCGCTCCCGCTTTCAGTTCTCAATAAGGTCCGATGTT CGTGTATAAATGCTCG	100bp
35	CTTGGTGACCAAATTTGAAAAAAAAAAAAACCGCGCCAAC TCATGTTGTTTTCAATCAGGTCCGCCAAGTTTGTATTTAAGG AACTGTTTCAGTTCATA	100bp
36	GGCTGAGCTATCCTATTGGCTATCGGGACAAAATTTGCTTGA GCCAATCAAAGTGCTCCGTGGACAATCGCCGTTCTGTCTATA AAAAGGTGAAGCAGCG	100bp
37	GGAAGTGCCAGACCGGAGGTGCGTCATTACCGGCGACGCC GATACGGTTCCTCCACCGAGGCCCATGCGAAGCTTCCACTA TGGCTTCCAGCACTGTC	100bp
38	CCCTCGAGGGGCGGAGCAAAAAGTGAGGCAGCAACGCCTCC TTATCCTCGCTCCCGCTTTCAGTTCTCAATAAGGTCCGATGTT CGTGTATAAATGCTCG	100bp
39	CTTGGTGACCAAATTTGAAAAAAAAAAAAACCGCGCCAAC TCATGTTGTTTTCAATCAGGTCCGCCAAGTTTGTATTTAAGG AACTGTTTCAGTTCATA	100bp
40	GGCTGAGCTATCCTATTGGCTATCGGGACAAAATTTGCTTGA GCCAATCAAAGTGCTCCGTGGACAATCGCCGTTCTGTCTATA AAAAGGTGAAGCAGCG	100bp
41	GGAAGTGCCAGACCGGAGGTGCGTCATTACCGGCGACGCC GATACGGTTCCTCCACCGAGGCCCATGCGAAGCTTCCACTA TGGCTTCCAGCACTGTC	100bp

[0094] 在一些情况下,相对较短的来源于人类的RE包含表1或SEQ ID NO:1-2、13-17和22-41的序列;或包含与表1或SEQ ID NO:1-2、13-17和22-41的序列具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性(例如,局部序列同一性)的序列。在一些情况下,较短的来源于人类的RE包括与hg19的人类序列具有至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性的非天然存在的序列。在一些情况下,相对较短的来源于人类的RE包含与位于表2所列的以下基因组或染色体位置之一处的人类序列具有至少50%、至少55%、至少60%、至少65%、至少

70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性的非天然存在的序列。

[0095] 表2:用于得到RE的基因组位置的实例

SEQ ID NO:	基因组位置	相对于阴性对照的倍数增加(标准化)
22	PEG10-cs;PEG10-he_chr7:94285387-94285887-3	154
23	SAT1-he_chrX:23801204-23801704-1	207
24	CDKN1C-he;CDKN1C-cs_chr11:2906751-2907251-3	532
25	BTG1-cs;BTG1-he_chr12:92539423-92539923-3	501
26	ATP5A1-he_chr18:43678011-43678511-3	1049
27	GANAB-he_chr11:62413828-62414328-3	54
28	ATP5F1-he_chr1:111991856-111992356-3	675
29	LMAN2-cs_chr5:176778403-176778903-3	
[0096] 30	RAB1A-he_chr2:65356988-65357488-3	1049
31	NEDD8-he_chr14:24701287-24701787-3	253
32	ATP5C1-he_chr10:7829892-7830392-3	
33	SSR4-he_chrX:153059879-153060379-1	
34	HIST1H4B-he_chr6:26027230-26027730(-)-2	
35	HIST1H4C-he_chr6:26103854-26104354(+)-3	313
36	HIST2H2AA3-he_chr1:149814228-149814728(-)-4	
37	TOMM6-he_chr6:41755153-41755653(+)-3	
38	HIST1H4B-he_chr6:26027230-26027730(-)-2	
39	HIST1H4C-he_chr6:26103854-26104354(+)-3	
40	HIST2H2AA3-he_chr1:149814228-149814728(-)-4	
41	TOMM6-he_chr6:41755153-41755653(+)-3	

[0097] 在一些情况下,相对较短的来源于人类的RE来源于表2所列的基因组位置处的启动子,并且与阴性对照(例如,单独载体对照、缓冲液对照或已知不具有表达活性的序列)相比,或者与组成型启动子、CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子相比,表现出至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍的转基因表达。

[0098] 在一些情况下,本公开内容的相对较短的RE包含内含子序列(例如,SEQ ID NO:1或SEQ ID NO:2),其当与表达盒中的转基因可操作地连接时,与在SCP控制下的转基因表达或者阴性对照(例如,单独的载体、单独的缓冲液或没有任何表达活性的序列)相比,导致至少50倍、100倍、500倍、1000倍或5000倍或超过1000倍的标准基因(例如,萤光素酶、ATP7B或FVIII)表达。在一些情况下,本公开内容的相对较短的RE(例如,SEQ ID NO:1或SEQ ID NO:2)当与表达盒中的转基因可操作地连接时,与SCP或CAG的大小标准化活性相比,导致至少1.5倍、2.5倍、5倍、10倍、15倍或20倍或更多倍的大小标准化活性。在一些情况下,本公开内容的相对较短的RE(例如,SEQ ID NO:1或SEQ ID NO:2)当与表达盒中的转基因(例如,萤光素酶)可操作地连接时,与在SCP、SerpE_TTR、Proto1、minCMV、UCL-HLP或CMVe控制下的萤光素酶相比;或与缓冲液对照相比,导致至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、10倍或更多倍的总萤光素酶发光(以光子/秒计)。在一些情况下,本公开内容的相对较短的RE(例如,

SEQ ID NO:1或SEQ ID NO:2) 当与表达盒中的人FVIII可操作地连接时,导致体内人FVIII浓度增加,其中该浓度增加是与缓冲液对照或与FVIII可操作地连接的UCL-HLP相比,至少1IU/mL、1.2IU/mL、1.5IU/mL或大于1IU/mL的FVIII。

[0099] 在一些情况下,本公开内容的相对较短的RE来源于人类启动子序列(例如,SEQ ID NO:13-17和22-41),其当与表达盒中的转基因(例如,萤光素酶、ATP7B或FVIII)可操作地连接时,与阴性对照(例如,单独的载体、单独的缓冲液或已知不具有表达活性的序列)相比,导致至少50倍、100倍、200倍、300倍、400倍、500倍、600倍、700倍、800倍、900倍或者1000或更多倍的标准化转基因表达。在一些情况下,与在CMV或CMV+CMVe控制下的萤光素酶表达相比,RE导致至少5倍、10倍或者100或更多倍的标准化萤光素酶表达。在一些实施方案中,两个或更多个RE可以组合以形成更大的RE。在一些情况下,RE位于转基因的上游。

[0100] 在一些情况下,本文公开的相对较短的来源于人类的RE包括不超过40bp、45bp、49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp,并且包括根据表1或SEQ ID NO:1-2、13-17和22-41中的任一个的序列。在一些情况下,表1列出的或SEQ ID NO:1-2、13-17和22-41的序列中的两个或更多个、三个或更多个、四个或更多个、五个或更多个、六个或更多个、七个或更多个、八个或更多个、九个或更多个或者十个或更多个可以在表达盒或载体中组合或串联使用。

[0101] 在各个方面,当载体上存在大小限制(如在AAV基因疗法中)或需要非常高水平的表达(如在离体蛋白质产生中)时,相对较短的来源于人类的调控元件尤其有用。本文公开的调控元件可以是组成型的或诱导型的。因此,本公开内容提供了高表达以及短调控元件两者的益处。在一些情况下,与对照调控元件如组成型启动子、CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子相比,本文公开的RE表现出每个碱基对或每个核苷酸更高的表达水平。在一些情况下,将SEQ ID NO:4用作用于比较基因表达水平的对照。

[0102] 在一些情况下,本公开内容的一个或多个RE适于增加较大转基因的表达(例如,在基因疗法或表达盒中),这是由于RE增强了转基因表达,而不占用显著的克隆能力。在一些情况下,本公开内容的RE不超过49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp,并包含根据表1或SEQ ID NO:1-2、13-17和22-41中的任一个的序列,或包含与其具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性的序列。在一些情况下,与没有RE的转基因表达相比,或与阴性对照(例如,单独载体对照、单独的缓冲液或已知不具有表达活性的序列)相比,本公开内容的相对较短的RE使转基因表达增加至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍。在一些情况下,对照是与相同转基因(例如,萤光素酶)可操作

地连接的组成型启动子。可以用作对照以与可操作地连接至转基因的RE的转基因表达进行比较的其他启动子的实例包括但不限于,CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子、AAT启动子、KAR启动子、EF1 α 启动子、EFS启动子或CMVe增强子/CMV启动子组合。

[0103] 本文的调控元件可以是启动子,并且当包含在表达盒中时将驱动下游序列的转录,该启动子可以与下游序列(例如,转基因)紧密相关或直接接触。启动子可以驱动连接的转基因的高、中或低表达。在一些情况下,启动子可以对一种或多种细胞类型具有特异性,或者可以在许多或多种细胞类型(例如,在至少2、3、4、5、6、7、8、9或10种细胞类型)中以相似的水平表达。

[0104] 在一些情况下,本文公开的RE包含来源于人类的序列。在一些情况下,本公开内容的RE是非天然存在的。如本文所用,术语“来源于人类的”是指与人类参考基因组(或人类基因组构建,如hg19)的序列具有至少80%、90%、95%或99%序列同一性的序列。同源序列可以是具有与人类基因组区域相比有至少80%序列同一性(例如,通过BLAST测量)的区域的序列。例如,与人类序列具有至少80%、至少85%、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%或至少99%同源性的序列被认为是来源于人类的序列。

[0105] 在一些情况下,调控元件包含来源于人类的序列和附加的序列,使得总体上该调控元件与人类基因组中的序列具有较低的序列同一性,而调控元件的一部分与人类基因组具有100%序列同一性(或局部序列同一性)。在一些情况下,序列同一性百分比是指覆盖表1列出的或SEQ ID NO:1-2、13-17和22-41中的任何序列中的至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的区域内的同源性。

[0106] 在一些情况下,来源于人类的调控元件是与人类序列100%相同的序列。在一些情况下,调控元件的序列是100%来源于人类的,其中RE与人类序列的区别为至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、45、50、55、60、65、70、75、80、85、90或95个核苷酸或碱基对。

[0107] 在其他情况下,至少5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、98%或99%的调控元件序列是来源于人类的。例如,调控元件的50%序列可以是来源于人类的,而其余50%是来源于非人类的(例如,来源于小鼠的或完全合成的)。进一步举例来说,被认为是50%来源于人类并且具有300bp的调控元件可与人类基因组中的序列具有45%的总体序列同一性,同时RE的碱基对1-150可能与人类基因组中大小相似的区域具有90%的同一性(例如,局部序列同一性)。

[0108] 在一些情况下,本公开内容的RE包含启动子序列、内含子序列、增强子序列、5' UTR序列或其组合。在一些情况下,本公开内容的RE是非天然存在的。在一些情况下,本公开内容的RE是来源于人类的(或是与人类参考基因组中的序列具有至少80%、90%、95%或99%序列同一性的序列)。

[0109] 在一些情况下,本文公开的相对较短的调控元件可以来源于基因组启动子序列(例如,表2中的基因组位置)。在一些情况下,本文公开的调控元件可以来源于基因组启动子序列和3' 非翻译区(3' UTR)。在一些情况下,本文公开的调控元件可以来源于基因间序

列。在一些情况下,本文公开的调控元件可以来源于基因下游的基因组序列,或来自5' UTR序列,或5' UTR与下游序列的混合物。

[0110] 本文中的调控元件可以是增强子,并且与没有增强子的情况下启动子对相同转基因的表达相比,其连同启动子一起存在于表达载体或盒中增加了可操作地连接的转基因的表达。增强子可以通过转录机制、转录后机制或两者来增加连接的转基因的表达。

[0111] 在一些情况下,本文的调控元件是内含子序列(例如,SEQ ID NO:1-2),或包含内含子,并且与没有内含子序列情况下启动子对相同转基因的表达相比,其连同启动子一起存在于表达载体或盒中增加了可操作地连接的转基因的表达。内含子序列或调控元件可通过转录机制、转录后机制或两者来增加连接的转基因的表达。

[0112] 在一些情况下,本文的调控元件是启动子序列(例如,SEQ ID NO:13-17和22-41),或包含启动子序列,并且其可以与表达盒中的转基因可操作地连接,而无需任何其他启动子序列和/或增强子序列以表达该转基因。

[0113] 在各种实施方案中,表1中的一个或多个RE可以驱动可操作地连接的转基因的表达增加,而无需表达盒中的启动子。在一些实施方案中,在较大转基因(例如,其序列大于1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb的转基因)太大而不能在常规表达盒中表达的情况下,这样的转基因可以与表1的一个或多个RE可操作地连接而不需要常规启动子,并且与不含RE的表达盒中的转基因表达相比,或与对照(例如,空载体、组成型启动子、CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子)相比,其以更高的水平表达。

[0114] 本文公开的调控元件的益处之一是其相对较短的长度。在一些情况下,本发明的调控元件的长度小于或等于40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp或400bp。在一些情况下,调控元件的长度小于或等于100bp、小于90bp、小于80bp、小于70bp、小于60bp或小于50bp,或在48-57bp之间,或在40-60bp之间,或在50-100bp之间。在一些情况下,本文所述的调控元件为40-50bp、45-55bp、50-60bp或55-65bp。在一些情况下,调控元件为45-60bp。在一些情况下,本文所述的调控元件为49bp或56bp。例如,调控元件SEQ ID NO:1的长度为56bp,并且调控元件SEQ ID NO:2的长度为49bp。在一些情况下,调控元件可以在100bp与150bp之间、在110bp与140bp之间、在110bp与130bp之间或在115bp与125bp之间。例如,调控元件SEQ ID NO:13-17的长度为117bp。在一些情况下,调控元件为或为约100bp。例如,SEQ ID NO:22-41的长度各自为100bp。

[0115] 本公开内容的调控元件优选具有来源于人类的序列。调控元件可以来源于转录起始位点上游的基因组序列、5' UTR序列、外显子序列、内含子序列或3' UTR序列。在一些实施方案中,来源于人类的调控元件是来源于人类的内含子序列。在一些情况下,来源于人类的调控元件是来源于人类序列的启动子序列。在一些实施方案中,来源于人类的调控元件是SEQ ID NO:1、SEQ ID NO:2或其组合的内含子序列,或者与其具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性(例如,通过BLAST测量)的序列。在一些实施方案中,来源于人类的调控元件是SEQ ID NO:13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41或其组合的启动子序列,或与其具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性(例如,通过BLAST测量)的序列。

[0116] 在一些情况下,调控元件包含5'非翻译区(5' UTR)的部分或全部。5' UTR调控元件可以通过几种不同方式影响基因的表达。5' UTR调控元件可以包含RNA结合蛋白的结合位点。此外,由5' UTR中的调控元件形成的二级结构可影响翻译所需RNA结合蛋白的结合。在一些实例中,调控元件可以具有高度的二级结构。在一些情况下,调控元件可以有很少的二级结构或没有二级结构。调控元件还可以包含内部核糖体进入位点(IRES),允许5'帽独立翻译。调控元件可以包含上游翻译起始密码子(uAUG)。在一些实施方案中,调控元件不包含上游翻译起始密码子。在一些实施方案中,调控元件在AUG密码子的一个碱基内不包含任何密码子,或包含的与AUG密码子相似的密码子少于偶然预期。在一些情况下,调控元件可以包含上游开放阅读框,其当存在上游AUG(或足够相似的序列)时发生,然后是框内终止密码子。在一些实例中,调控元件不包含uORF。在一些情况下,调控元件包含微小RNA结合位点,或RNA结合蛋白的结合位点。

[0117] 在一些情况下,本公开内容的调控元件包含与来源于人类的序列同源的序列。如果序列与人类序列具有至少80%、85%、90%、95%、98%或99%序列同一性,则该序列“与来源于人类的序列同源”。优选地,与来源于人类的调控序列同源的序列与人类序列至少90%相同。在一些实施方案中,本文的调控元件是与SEQ ID NO:1-2、13-17和22-41具有至少80%、85%、90%、95%、98%或99%序列同一性的调控元件。当调控元件与SEQ ID NO:1-2中的任一个同源时,与没有调控元件的类似载体相比,这样的调控元件仍导致可操作地连接的转基因的较高表达(当启动子也存在于表达载体或盒中时)。可以例如在HEK293T或CHO细胞的表达中观察到这样的转基因的较高表达。在一些情况下,可以在多种细胞类型中,例如两种或更多种哺乳动物细胞系、人类细胞系、肾细胞、上皮细胞、肝脏细胞、肝细胞、神经元、成纤维细胞和/或CNS细胞中观察到这样的转基因的较高表达。

[0118] 在一些情况下,当调控元件与SEQ ID NO:13-17和22-41中的任一个同源时,与没有调控元件的类似载体相比,这样的调控元件导致可操作地连接的转基因的较高表达(即使表达载体或盒中不存在其他启动子)。可以例如在HEK293T或CHO细胞的表达中观察到这样的转基因的较高表达。在一些情况下,可以在多种细胞类型中,例如两种或更多种哺乳动物细胞系、人类细胞系、肾细胞、上皮细胞、肝脏细胞、肝细胞、神经元、成纤维细胞和/或CNS细胞中观察到这样的转基因的较高表达。

[0119] 本公开内容的调控元件还可以是上述任一个的功能片段,例如,SEQ ID NO:1-2、13-17和22-41的片段,或其任何组合。与没有调控元件的类似表达盒或载体相比,这样的功能片段也增加了表达盒或载体中转基因的表达。当功能片段是增强子、内含子序列、启动子序列或其组合时,与没有功能片段的类似载体或盒相比,当该功能片段与转基因可操作地连接时,观察到更高的表达。片段的长度优选小于或等于25bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp或110bp。在一些情况下,本公开内容的来自人类启动子序列的RE可以增加可操作地连接的转基因的表达,而不需要在表达盒或载体中的第二启动子。在一些情况下,常规表达盒或载体中的启动子可以被以下取代:一个或多个SEQ ID NO:13-17和22-41、其组合,或者与SEQ ID NO:13-17和22-41或其组合具有至少70%、至少75%、至少80%、至少85%、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%或至少99%序列同一性(例如,通过BLAST测量)的序列。

[0120] 在一些实施方案中,调控元件是以下任一种:(i):SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41,(ii)与SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41中的任一个具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的核酸序列,(iii):(i)或(ii)的任何序列的功能片段,或(iv):(i)、(ii)和(iii)的任何序列的组合。在一些实施方案中,调控元件的两个或更多个拷贝被用于增强转基因表达,例如,SEQ ID NO:1-2、13-17和22-41中的一个或多个的两个或更多个拷贝。在一些实施方案中,将SEQ ID NO:1和2的组合或其功能片段的组合用作调控元件,以增加转基因表达。在其他实施方案中,SEQ ID NO:1和/或2中的一个或多个与另一调控元件如启动子或增强子耦合或可操作地连接,以增加转基因表达。在其他实施方案中,SEQ ID NO:1和/或2中的一个或多个与SEQ ID NO:1-2、13-17和22-41中的一个或多个耦合或可操作地连接,以增加转基因表达。在一些实施方案中,当调控元件与任何启动子耦合或可操作地连接时,内含子序列的调控元件可以增加转基因表达。在一些情况下,当调控元件与转基因耦合或可操作地连接而没有任何其他启动子序列时,来源于人类启动子序列的调控元件可以增加转基因表达。在一些情况下,当组合的调控元件与转基因耦合或可操作地连接而没有任何其他启动子序列时,包含来源于人类的启动子序列和内含子序列的调控元件可以增加转基因表达。

[0121] 在一些实施方案中,可以通过添加本文公开的一个或多个调控元件(例如,SEQ ID NO:1-2、13-17和22-41)以改善或增加基因疗法的转基因表达,从而增强任何基因疗法或已知基因疗法。

[0122] 在一些情况下,调控元件包含来源于人类的序列和来源于非人类的序列,使得总体上调控元件与人类基因组具有低序列同一性。然而,调控元件与人类基因组具有100%序列同一性。在其他情况下,至少50%、60%、70%、80%、90%、95%、98%或99%的调控元件序列是来源于人类的,或者至少10、20、30、40或50个相邻的核苷酸是来源于人类的。例如,调控元件可以使其序列的50%是来源于人类的,其余的50%是来源于非人类的(例如,来源于小鼠的、来源于病毒的或完全合成的)。

[0123] 表达的增加可以发生在转录或转录后水平。例如,在转录水平,调控元件可以通过募集转录因子和/或RNA聚合酶、增加转录起始或募集增加转录水平的DNA和/或组蛋白修饰来增加表达。这样的表达的增加可以通过测量代表转基因的RNA转录物的量的增加来检测。在转录后水平,调控元件可以通过增加翻译成蛋白质的蛋白质的量来增加表达。这可以通过各种机制例如通过增加mRNA的稳定性或增加翻译所需的蛋白质的募集和组装来实现。这样的表达的增加可以通过测量代表转基因的表达的蛋白质的量来检测。产生的蛋白质的量可以直接测量,例如通过酶联免疫吸附测定(ELISA),或者间接测量,例如通过功能测定。在功能测定中通常测量的蛋白质是萤光素酶。然而,其他蛋白质,如因子VIII也可以通过功能测定进行测量。优选地,与使用相同表达载体减去调控元件的转基因的表达相比,本文的调控元件使表达载体或盒的基因、转基因或治疗性转基因的表达增加至少1.4、1.5、1.6、1.7、1.8、1.9、2、2.5、3、3.5、4、5、6、7、8、9或10倍,或大于5倍,或大于10倍。在一些情况下,与使用相同表达载体减去调控元件的转基因的表达相比,本文的调控元件使表达载体或盒的基

因、转基因或治疗性转基因的表达增加至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍。这样的表达载体或盒可以包括与调控元件分离的启动子。启动子的实例包括: minCMV启动子、超级核心启动子(SCP)、TTR启动子、Proto 1启动子、UCL-HLP启动子、AAT启动子、KAR启动子、EF1 α 启动子和EFS启动子。在一些情况下, 包含与转基因可操作地连接的SEQ ID NO: 13-17和22-41中的一个或多个的表达载体或盒不需要单独的启动子进行转基因表达。

[0124] 在一些情况下, 与本文的任何调控元件连同minCMV启动子可操作地连接的转基因的表达高于连接至minCMV启动子而没有本文所述的调控元件的转基因的表达。在一些情况下, 当连接至如本文所述的minCMV启动子和调控元件时, 转基因的表达高于当连接至以下启动子中的任一个而没有调控元件时的转基因表达: 超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子、AAT启动子、KAR启动子、EF1 α 启动子或EFS启动子。

[0125] 在一些情况下, 与本文的任何调控元件(例如, SEQ ID NO: 1-2、13-17和22-41)可操作地连接的转基因表达高于当连接至UCL-HLP启动子时的转基因表达。

[0126] 在一些情况下, 本文所述的调控元件驱动相对于其长度的高表达。例如, 长度是CMV启动子一半的本公开内容的调控元件驱动大于CMV启动子水平的一半的表达。在一些情况下, 调控元件由每碱基对的活性(或转基因表达)定义, 或称为大小标准化活性。例如, 驱动与CMV启动子相同水平的表达但长度为CMV启动子的一半的调控元件具有两倍的每碱基对活性。如图1A所示, SEQ ID NO: 3(SEQ ID NO: 1与minCMV启动子组合)产生5552的标准化萤光素酶值, 其产生约21的大小标准化活性值, 如图1B所示(SEQ ID NO: 3为266bp)。SEQ ID NO: 4(SEQ ID NO: 2与minCMV启动子组合)产生53的大小标准化值。单独的minCMV启动子产生19的大小标准化值, 并且SCP启动子和CAG启动子分别产生1和10的大小标准化值, 如图1B所示。

[0127] 表达, 特别是高基因表达, 可以使用本领域已知的任何技术测量。在一些情况下, 使用RNA定量/测序技术如定量PCR、northern印迹或下一代测序来测量当转基因受本公开内容的调控元件控制时的相对和较高表达。

[0128] 在一些情况下, 高表达可以通过从感兴趣的转基因/基因产生/表达的蛋白质的浓度测量。蛋白质的浓度可以通过本领域已知的任何方法测量。用于测量蛋白质表达的方法的非限制性实例包括, 但不限于, ELISA、放射免疫测定、western印迹或高效液相色谱法。参见, 例如, Noble JE, Quantification of protein concentration using UV absorbance and Coomassie dyes, *Methods Enzymol.* 2014; 536: 17-26; Kurien BT, Scofield RH. A brief review of other notable protein detection methods on acrylamide gels, *Methods Mol Biol.* 2012; 869: 617-20; 以及 Daniel M. Bollag, Michael D. Rozycki 和 Stuart J. Edelstein, *Protein Methods*, 第二版, Wiley Publishers (1996)。蛋白质表达可以在体外或体内测量。

[0129] 在各种实施方案中, 可以通过使用PCR方法测量从转基因产生的mRNA或RNA转录物的量来确定转基因表达。在一些情况下, 可以通过与基因产物相互作用的蛋白质或免疫测定来测量转基因基因的表达。在一些情况下, Northern印迹分析和/或原位杂交也可以用于

分析体内转基因表达。还可以监测转基因表达的蛋白质的水平(例如,来自荧光报道转基因的荧光),以确定在细胞中或体内转基因表达在转录方面的增加以及蛋白质合成(即翻译)的增加。还可以使用多种方法测定蛋白质水平,该方法包括,但不限于,western印迹分析、免疫组织化学、免疫荧光组织化学和/或ELISA测定。

[0130] 在一些实例中,如通过ELISA所测量的,本公开内容的调控元件导致可操作地连接的转基因在相关细胞类型中以至少0.5、1.0、1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2.0、2.5或3IU/ml的水平表达。相关细胞类型的实例包括,但不限于,肝细胞、肝脏细胞、肌肉细胞、肺细胞、上皮细胞和肾细胞,例如293T和CHO细胞。在一些实施方案中,调控元件增加转基因表达的能力可以在小鼠中评估,其中测量整个小鼠中的转基因表达的总量和/或具有转基因表达的细胞类型或组织类型的总数目。在一些情况下,如通过ELISA所测量的,本公开内容的调控元件可以导致可操作地连接的转基因在小鼠或其他生物体的血液中以至少0.05、0.06、0.07、0.08、0.09、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2.0、2.5或3.0IU/ml的总体水平表达。

[0131] 当评估表达盒或载体的活性时,可以将活性或表达表示为每单位剂量的活性或表达水平,或相对于施用或递送至细胞或小鼠中的表达盒或载体的剂量标准化。在一些实施方案中,将转基因的表达或活性相对于使用的质粒或DNA(例如,每只小鼠的 $\mu\text{g}/\text{kg}$)或者病毒颗粒的量标准化(例如,相对于每只小鼠的基因组拷贝/ kg 的量标准化),以允许在具有或不具有调控元件的不同表达载体或盒之间进行比较。例如,当评估小鼠中调控元件的活性时,测定的表达或活性可以相对于每只小鼠约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或者大于10或20 μg 表达载体、盒或质粒的剂量标准化。在一些实施方案中,表达水平或活性可以相对于每只小鼠 10^{10} 、 10^{11} 、 10^{12} 、 10^{13} 、 10^{14} 或 10^{15} gc/kg个包含本文公开的表达载体或盒的病毒颗粒标准化。

[0132] 在其他实例中,连接的转基因的表达通过表达的蛋白质活性的功能测定来测量。功能测定的实例包括用于定量萤光素酶活性的萤光素酶测定和用于定量因子VIII活性的Coatest测定。在Coatest测定中,活性因子VIII(一种辅因子)的表达促进释放发色基团pNA的反应。在405nm处光度读取与释放的发色基团相关的颜色,并且颜色的强度与因子VIII活性的量成比例。因子VIII活性可以测量为包含1.0IU/mL因子VIII的血浆所显示的活性水平的百分比。当蛋白质的浓度为1.0IU/mL时,与minCMV和因子VIII可操作地连接的本公开内容的调控元件例如SEQ ID NO:1或2可导致与没有调控元件的类似表达载体(如没有调控元件的UCL-HLP)的活性水平的相比超过10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%、250%或大于250%的因子VIII活性水平。在一些情况下,例如当蛋白质的浓度为1.0IU/mL时,与因子VIII可操作地连接的本公开内容的调控元件例如SEQ ID NO:13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41,可以导致与具有UCL-HLP启动子的类似表达载体的活性水平相比超过10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%、200%、250%或大于250%的因子VIII活性水平。

[0133] 本文公开的调控元件可以驱动或增加转基因在一种或多种不同细胞类型中的表达。当调控元件能够增加转基因在多种细胞类型中的表达时,这种表达被认为是全局表达。

全局表达不需要转基因在生物体的每个细胞中表达。全局调控元件可以在不同表达的细胞类型中以类似的水平表达转基因,也可以在不同细胞类型中以一定范围的表达水平表达转基因。在一些情况下,全局调控元件可以驱动或增加转基因在至少两种、三种、四种、五种、十种、十五种或二十种不同细胞类型中的表达。例如,具有全局表达的调控元件可以驱动或增加转基因在至少2、3、4、5、6、7、8、9、10、11或12个不同细胞系中的表达。细胞系的实例包括293细胞、CHO细胞、癌细胞系和永生化细胞系。在一些情况下,调控元件驱动或增加转基因在动物体内的一种或多种细胞类型中的表达。

[0134] 在一些实施方案中,本文的调控元件驱动基因在肝细胞中的表达。在一些实施方案中,调控元件驱动基因在以下细胞类型中的任一种或多种中的表达:肺泡细胞、心肌细胞、上皮细胞、肝细胞、肾细胞、肠细胞、肌细胞、神经元、卵巢细胞和肾脏细胞。在一些实施方案中,调控元件驱动基因在以下细胞系中的任一种或多种中的表达:中国仓鼠卵巢细胞(CHO)、小鼠骨髓瘤细胞(NSO)、幼仓鼠肾细胞(BHK)、B16黑素瘤细胞、HEK293T细胞、HeLa细胞、HT-1080细胞和PER.C5细胞。在一些情况下,调控元件驱动基因在离体或体外的细胞系中的表达。

[0135] 本公开内容的调控元件可以组合在一起,或与其他调控元件组合。此类其他调控元件包括,例如,组成型启动子、诱导型启动子、抑制因子、增强子或转录后稳定性元件。在一个实施方案中,载体包含SEQ ID NO:3的调控元件,其包含SEQ ID NO:1和minCMV启动子,在连接的转基因的上游。本文考虑的启动子的实例包括:minCMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子、CMVe增强子/CMV启动子组合、AAT启动子、KAR启动子、EF1 α 启动子、EFS启动子或CBA启动子。

[0136] 本公开内容的组合调控元件可以仍然很短。组合调控元件可具有长度为约50至500、100至400、50至200、100至200或150至300个碱基对的总长度。组合调控元件可具有长度为小于约500、450、400、350、300、250、200、150或100个碱基对的组合长度。在一些实施方案中,调控元件为40-50bp、45-50bp、49bp、50-60bp、56bp、50-55bp或45-60bp。

[0137] 组合的调控元件可以来自不同物种。在优选的实例中,组合调控元件的至少一部分是来源于人类的。非人类来源的元件可以来源于哺乳动物、病毒或合成序列。

[0138] 表达盒

[0139] 术语“表达盒”和“核酸盒”可互换使用,是指多核苷酸分子或核酸序列。在一些情况下,表达盒包含与转基因可操作地连接的一个或多个本文公开的调控元件。在一些情况下,表达盒包含一个或多个调控元件。在一些情况下,表达盒还包含启动子。在一些情况下,表达盒包含SEQ ID NO:1-2、13-17和22-41的一个或多个序列和/或其任何组合。在一些情况下,表达盒包含(i):SEQ ID NO:1-2、13-17和22-41中的一个或多个,(ii):与SEQ ID NO:1-2、13-17和22-41中的任一个具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的核酸序列;(iii):(i)或(ii)的任何序列的功能片段;或(iv):(i)、(ii)和/或(iii)的任何序列的组合。在一些情况下,序列同一性通过BLAST测量。在一些情况下,调控元件位于表达盒中转基因的上游。在一些情况下,调控元件位于表达盒中转基因的下游。

[0140] 在一些方面,本文所述的一个或多个调控元件(例如,SEQ ID NO:1-2、13-17和22-41)与表达盒中的转基因(例如,报道基因或治疗性转基因)可操作地连接。在一些情况下,

基因疗法包括表达盒,该表达盒包含与本公开内容的一个或多个、两个或更多个、三个或更多个、四个或更多个或五个或更多个调控元件可操作地连接的转基因,导致该转基因的表达增加,该转基因例如,报道基因、eGFP、RFP、因子VIII、Cas9、DNA结合蛋白、激素、生长或分化因子、胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体、囊性纤维化跨膜调节因子,与粘多糖贮积症I、II、III或IV型相关的基因、 β 珠蛋白或脂蛋白脂肪酶或其变体或片段。在一些情况下,转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体或功能片段,或与其具有至少80%、至少85%、至少90%、至少95%或至少99%序列同一性的序列。在一些情况下,转基因是CDKL5、CNTNAP2、ZEB2,或其变体或功能片段,或与其具有至少80%、至少85%、至少90%、至少95%或至少99%序列同一性的序列。在一些情况下,转基因是纤维蛋白原、凝血酶原、凝固因子或凝血因子(例如,因子1、2、3、4、5、6、7、8、9、10、11或12),或其变体或功能片段。

[0141] 在一些情况下,表达盒适于通过基因疗法递送。在一些情况下,表达盒是线性或环状构建体。在一些情况下,表达盒是质粒、载体、病毒载体或rAAV的一部分。在一些情况下,包含与转基因可操作地连接的本公开内容的一个或多个RE的表达盒适于在多种细胞中表达,例如,在任何哺乳动物或人类细胞系、肾细胞、上皮细胞、肝脏细胞、肝细胞、神经元、成纤维细胞和/或CNS细胞中。在一些情况下,包含与转基因可操作地连接的本公开内容的一个或多个RE(例如,SEQ ID NO:1-2、13-17和22-41)的表达盒适于在2种或更多种、3种或更多种、4种或更多种或者5种或更多种哺乳动物或人类细胞系或细胞类型中表达。在一些情况下,转基因表达在以下细胞类型中的2种或更多种、3种或更多种、4种或更多种或者5种或更多种中发生:肾细胞、上皮细胞、肝细胞、神经元、成纤维细胞、心肌细胞、CNS细胞、肌肉细胞、血细胞和/或干细胞。

[0142] 在一些情况下,包含与转基因可操作地连接的本公开内容的一个或多个RE(例如,SEQ ID NO:1-2、13-17和22-41)的表达盒适于在肝脏或肝细胞中进行转基因表达。在一些情况下,包含与转基因可操作地连接的本公开内容的一个或多个RE的表达盒适于转基因的全身递送,例如,通过口服、静脉内、肌肉内、腹膜内、鞘内、肠内或肠胃外施用。在一些情况下,转基因的全身递送涉及肝脏或肝细胞中转基因表达的增加。在一些情况下,包含与转基因可操作地连接的本公开内容的一个或多个RE的表达盒与没有RE的表达盒相比,或与对照(例如,单独的载体对照、缓冲液对照或不具有任何表达活性的序列)相比,导致转基因在肝脏细胞中以至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍的水平表达。在一些情况下,由相对较短的RE驱动的表达与由相同转基因可操作地连接的组成型启动子、CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子驱动的表达进行比较。

[0143] 在一些情况下,包含本公开内容的调控元件的表达盒导致可操作地连接的转基因的较高或增加的表达,其中与对照例如组成型启动子、CMV启动子、超级核心启动子(SCP)、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子相比确定了较高或增加的表达。可以用于比较以确定较高或增加的转基因表达的其他对照包括单独的缓冲液、minCMV、

EFS、单独的载体或已知不具有表达活性的序列。在一些情况下,将CAG或SEQ ID NO:4用作阳性对照以比较表达水平和/或大小标准化活性。

[0144] 感兴趣的转基因或基因的表达可以通过表达载体或表达盒发生。这样的表达载体或盒可以包含以下附加元件中的任一个或多个:启动子、增强子或转录后调控元件。本公开内容的调控元件可以位于转基因的上游或下游,并且可以被转录或不被转录。在一些实例中,本公开内容的调控元件位于启动子的下游和转基因的上游。在一些实例中,调控元件被定位为转基因的启动子。在一些情况下,本公开内容的一个或多个RE与转基因可操作地连接,而在表达载体或盒中没有任何其他启动子。

[0145] 表达载体或盒可以是环状或线性核酸分子。在一些情况下,表达载体或盒递送至细胞(例如,多种不同的细胞或细胞类型,包括靶细胞或细胞类型和/或非靶细胞或细胞类型)。载体或盒可以是整合载体或非整合载体,该整合是指载体或盒将表达载体或盒的一部分或全部和/或转基因整合到细胞基因组中的能力。整合载体或非整合载体都可以用于递送与调控元件可操作地连接的转基因。载体或盒的实例包括但不限于:(a)非病毒载体,如核酸载体,包括线性寡核苷酸和环状质粒;人工染色体,如人类人工染色体(HAC)、酵母人工染色体(YAC)和细菌人工染色体(BAC或PAC);游离型载体;转座子(例如PiggyBac);和(b)病毒载体,如逆转录病毒载体、慢病毒载体、腺病毒载体和腺伴随病毒载体(AAV)。病毒对递送核酸具有若干优点,包括高感染性。在一些实施方案中,使用病毒递送核酸分子或表达盒,该核酸分子或表达盒包含与转基因可操作地连接的本文所述的一个或多个调控元件。

[0146] 本文所述的载体可以直接或通过递送系统递送至感兴趣的细胞。递送系统的实例包括但不限于,病毒载体(例如,逆转录病毒、腺病毒、腺伴随病毒(AAV)、辅助病毒依赖型腺病毒系统、杂交腺病毒系统、单纯疱疹、痘病毒、慢病毒和EB病毒)、物理系统(裸DNA、DNA轰击、电穿孔、流体动力、超声和磁转染)和化学系统(阳离子脂质、不同的阳离子聚合物和脂质聚合物)。

[0147] 可以使用任何已知的技术将调控元件和可操作地连接的转基因或者包含调控元件和转基因的组合物递送至感兴趣的细胞,以在感兴趣的细胞类型、组织或生物体中赋予或诱导转基因的体外、体内或离体表达。

[0148] 病毒基因疗法载体或基因递送载体的优选特征包括可再现、稳定繁殖和纯化至高滴度的能力;介导靶向递送的能力(例如,将转基因特异性递送至感兴趣的组织或器官而没有在他处的广泛载体散播);以及介导基因递送和转基因表达而不会引起有害的副作用的能力。

[0149] 已将几种类型的病毒,例如非致病性细小病毒、腺伴随病毒工程化用于基因递送的目的。用于各种基因疗法应用的基于病毒的载体可以利用病毒感染途径,但避免随后表达可导致复制和毒性的病毒基因。这样的基于病毒的载体可以通过从病毒基因组删除全部或部分编码区而获得,但完整保留对于诸如将表达载体包装到病毒衣壳中或者将载体核酸或异源基因或DNA整合到宿主基因组中等功能所必需的那些序列(例如,末端重复序列)。例如,可以将包含转基因的表达载体或盒克隆到病毒骨架中,如缺少病毒基因的经修饰或工程化的病毒骨架中,并与其他载体(例如包装载体)结合使用,该载体在例如共转染时产生重组病毒载体颗粒。

[0150] 已将AAV、非致病性细小病毒的几种血清型工程化用于基因递送的目的,其中一些

已知对某些组织或细胞类型具有向性。可将用于各种基因疗法应用的病毒工程化为复制缺陷型,或在受试者或宿主中具有低毒性和低致病性。可以通过从病毒基因组中删除全部或部分编码区,并完整保留对于例如将载体基因组包装到病毒衣壳中或将载体核酸(例如异源DNA序列)整合至宿主基因组中等功能而言所必需的序列(例如,反向末端重复序列),从而获得此类基于病毒的载体。例如,可以将包含转基因的表达载体或盒克隆到病毒骨架,例如缺少病毒基因的经修饰或工程化的病毒骨架中,并与其他载体(例如包装载体)结合使用,该载体在例如共转染时产生重组病毒载体颗粒。

[0151] 在一些实施方案中,用于在体内、离体或体外向细胞、细胞类型或组织中递送一种或多种调控元件和转基因的AAV载体或AAV病毒颗粒或病毒粒子优选是复制缺陷的。在一些实施方案中,AAV病毒被工程化或基因修饰,以使其仅在辅助因子存在时才能复制并生成病毒粒子。

[0152] 在一些实施方案中,表达载体或盒被设计成由AAV或重组AAV(rAAV)递送。在一些实施方案中,使用慢病毒或慢病毒载体递送表达载体或盒。在一些实施方案中,优选使用慢病毒或慢病毒载体递送较大的转基因,即超过AAV克隆能力的基因。

[0153] 表达载体或表达盒可以被设计用于由AAV递送。AAV可以是任何血清型,例如,AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ或者嵌合、杂合或变体AAV。AAV还可以是自互补AAV(scAAV)。设计用于通过AAV递送的表达载体或盒可以包含5' ITR、一个或多个调控元件、任选的最小启动子、转基因、任选的一个或多个内含子、任选的polyA信号和3' ITR。在一些情况下,表达载体或盒还包含一个或多个转录后RNA调控元件。

[0154] 包含SEQ ID NO:1或SEQ ID NO:2的示例性表达载体或盒分别如图6B和图6C所示。可以在启动子位置包含SEQ ID NO:13-17和22-41中的任一个或多个的示例性表达载体,如图7所示。这样的表达载体或盒可以使用本文所述的递送系统或基因疗法中的任一种进行转基因表达,该递送系统或基因疗法包括但不限于,AAV、慢病毒、腺病毒或逆转录病毒。

[0155] 表达载体或盒可以设计成通过优化的治疗性逆转录病毒载体来递送。逆转录病毒载体可以是慢病毒,包含左侧的(5')LTR;有助于病毒的包装和/或核输入的序列、至少一个核酸细胞类型选择性调控元件、任选的慢病毒反向应答元件(RRE);任选的启动子或其活性部分;与各种调控元件可操作地连接的治疗性转基因;任选的绝缘子;以及右侧的(3')逆转录病毒ITR。

[0156] 在一些实施方案中,表达载体或盒包含一个或多个调控元件。在一些实例中,载体或盒包含组合的两个或更多个调控元件。在一些实例中,载体或盒包含未组合的两个或更多个调控元件,例如,转基因上游的启动子和位于转基因下游的增强子。在一些实施方案中,内含子序列位于启动子的下游。

[0157] 在一些情况下,表达载体或表达盒包含高驱动调控元件、全局调控元件或短调控元件。在一些情况下,高驱动调控元件驱动全局表达和/或是短的。在一些情况下,调控元件是短的全局调控元件。在一些实施方案中,包含调控元件的表达载体或盒被用于病毒载体或包装在病毒中,该病毒可用于体内感染动物模型或者离体或体外转染细胞,以评估调控元件在整体的动物模型中或某些细胞类型中或靶细胞类型中的活性。

[0158] 在一些实施方案中,包含本公开内容的调控元件的表达盒也包含一种或多种转基

因。转基因可以是蛋白质编码基因。在一些情况下,表达盒包含治疗性转基因。治疗性转基因可以替代缺失或有缺陷的基因,或补偿蛋白质的表达缺陷。治疗性转基因可参与细胞信号传导途径。在一些情况下,表达载体或表达盒包含用于商业生产的转基因。在一些实施方案中,用于商业生产的转基因可以产生治疗性蛋白质,或者改善宿主细胞的功能,或者恢复或拯救宿主细胞的表型。

[0159] 本文公开的调控元件可以位于表达载体或表达盒内的任何位置。例如,调控元件可以位于增强子的上游、增强子的下游但在启动子的上游、转基因的5' UTR内、转基因中的内含子内、转基因的3' UTR内或转基因的下游。在一些情况下,一个或多个调控元件位于可操作地连接的转基因的上游或下游。例如,SEQ ID NO:3和4分别在minCMV启动子下游包含SEQ ID NO:1和2的序列。调控元件可以从表达盒或载体内的不同位置对转基因表达具有不同的影响。每个调控元件可以在表达载体或盒中具有优选的位置,从而导致该调控元件的最佳表达。

[0160] 利用具有高活性的较短调控元件可以允许在相同包装能力内递送较大的转基因,而不牺牲转基因表达或高表达。在一些情况下,与没有调控元件的相似载体或盒相比,使用较短的调控元件允许在单个载体中递送两个或更多个转基因,或以相对较高的表达水平递送较大的转基因。两个或更多个转基因可以包含两个或更多个不同的蛋白质编码基因或者蛋白质编码基因和非编码元件。在一个实例中,载体包含Cas9编码序列和一个或多个指导RNA,如实施例4中所述。在一些情况下,调控元件的长度为约20至100、30至100、40至80或50至60个碱基对。在一些情况下,调控元件的长度小于150、120、110、100、90、80、70或60个碱基对。本公开内容的组合调控元件可以是短的。组合调控元件可以具有长度为约50至500、100至400、50至200、100至200或150至300个碱基对的总长度。组合调控元件可以具有长度为小于约500、450、400、350、300、250、200、150或100个碱基对的组合长度。

[0161] 在一些情况下,表达盒或载体包含与较大转基因可操作地连接的本公开内容的一个或多个RE,该较大转基因的序列大于1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb。在一些情况下,较大转基因与一个或多个RE可操作地连接,其中每个RE不超过49bp、50bp、56bp、60bp、70bp、80bp、90bp、100bp、110bp、117bp、120bp、130bp、140bp、150bp、160bp、170bp、180bp、190bp、200bp、210bp、220bp、230bp、240bp、250bp、259bp、260bp、265bp、270bp、280bp、290bp、300bp、310bp、320bp、330bp、340bp、350bp、360bp、370bp、380bp、390bp或400bp,并且包含根据表1或SEQ ID NO:1-2、13-17和22-41中的任一个的序列。在一些情况下,相对于没有RE的表达载体或盒,或与阴性对照(例如,单独的载体对照或包含不具有已知表达活性的序列的载体)相比,RE使转基因的表达增加至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍。

[0162] 当评估表达盒或载体的活性时,该活性或表达可表示为单位剂量的活性或表达水平,或相对于所施用或递送至细胞、小鼠或受试者的表达盒或载体的剂量标准化。在一些情况下,将转基因的表达或活性相对于所用的质粒或DNA(例如,每只小鼠的 $\mu\text{g/kg}$)或病毒颗粒标准化(例如,相对于每个小鼠或受试者的基因拷贝数/kg的量标准化),以允许在具有或不具有调控元件的不同表达载体或表达盒之间进行比较。例如,当评估小鼠中的调控元件

的活性时,所测定的细胞中的转基因表达或转基因活性可相对于每只小鼠约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20 μg ,或大于10 μg ,或大于20 μg 表达载体、表达盒或质粒的剂量标准化。在一些情况下,表达水平或活性可相对于每只小鼠 10^{10} 、 10^{11} 、 10^{12} 、 10^{13} 、 10^{14} 或 10^{15} 个gc/kg病毒颗粒(包含本文公开的表达载体或表达盒的)标准化。

[0163] 使用基因疗法(例如,rAAV)递送本公开内容的表达盒的一个优势是,随着时间的推移,这样的疗法可以提供更加靶向和持续的治疗效果。此外,病毒基因疗法可被工程化为对一种或多种感兴趣的细胞类型或组织(例如,肝细胞或肝脏)具有向性。例如,可以对病毒基因疗法工程化,以将有效负载或治疗剂例如转录调节因子或转基因在体内感染并递送至一个或多个区域、组织或细胞类型。

[0164] 转基因

[0165] 本文公开的构建体(例如,表达盒或载体)可用于驱动转基因在感兴趣的细胞或多个细胞中的表达。在一些情况下,本文公开的构建体用于重组蛋白表达。例如,本文公开的调控元件可用于驱动重组蛋白的表达。用本公开内容的方法产生的重组蛋白可被产生用于治疗目的、研究目的或商业目的。可以作为体内、离体或体外蛋白质表达系统的一部分表达的转基因的实例包括,但不限于:Cas9、因子VIII、DNA结合蛋白、激素、生长/分化因子,例如胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;调控或调节免疫系统的治疗性蛋白质,例如,细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体;囊性纤维化跨膜调节因子;与粘多糖贮积症I、II、III和IV型相关的基因; β 珠蛋白;以及脂蛋白脂肪酶。在一些情况下,表达的转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体或功能片段。在一些情况下,表达的转基因是CDKL5、CNTNAP2、ZEB2,或其变体或功能片段。在一些情况下,转基因是纤维蛋白原、凝血酶原、凝固因子或凝血因子(例如,因子1、2、3、4、5、6、7、8、9、10、11或12),其变体或功能片段。

[0166] 在一些情况下,包含与转基因可操作地连接的本公开内容的一个或多个RE的表达盒适于基因疗法治疗。在一些情况下,基因疗法治疗被施用或适于全身递送,如通过静脉内输注或注射。在一些情况下,基因疗法适于在肝脏或肝细胞中的表达。在一些情况下,可操作地连接的转基因是在肝脏或肝细胞中表达的基因,或者是与肝脏疾病或病况相关的基因,例如,与威尔逊病相关联的ATP7B;与进行性家族性肝内胆汁淤积3型相关联的ABCB4;与遗传性果糖不耐受相关联的ALDOB;与糖原贮积病IV型相关联的GBE1;与酪氨酸血症I型相关联的FAH;与精氨琥珀酸裂合酶缺乏症相关联的ASL;与citrin缺乏症(CTLN2、NICCD)相关联的SLC25A13;与胆固醇酯贮积病相关联的LIPA;与 α -1抗胰蛋白酶缺乏症相关联的SERPINA1;与囊性纤维化相关联的CFTR;与遗传性血色素沉着症相关联的HFE;或与阿尔斯特伦综合征相关联的ALMS1。在一些情况下,本文公开的一种或多种调控元件或包含RE的表达盒被用于基因疗法治疗,以治疗遗传性肝病,例如,胆汁酸合成障碍、碳水化合物代谢障碍、氨基酸代谢障碍、尿素循环障碍或脂质代谢障碍。在一些情况下,本文公开的一种或多种调控元件或包含RE的表达盒被用于基因疗法治疗(例如,AAV基因疗法)以治疗威尔逊病;进行性家族性肝内胆汁淤积3型;遗传性果糖不耐受;糖原贮积病IV型;酪氨酸血症I型;精氨琥珀酸裂合酶缺乏症;citrin缺乏症(CTLN2、NICCD);胆固醇酯贮积病; α -1抗胰蛋白酶缺乏症;囊性纤维化;遗传性血色素沉着症;或阿尔斯特伦综合征。在一些情况下,本文公开的一种或多种调控元件或包含RE的表达盒被用于基因疗法治疗(例如,AAV基因疗法)以治疗

凝血障碍(例如,血友病、血友病A或血友病B)或威尔逊病。

[0167] 在一些情况下,表达重组蛋白的方法包括将表达盒或载体引入细胞或多个细胞中,其中表达盒或载体包含与转基因可操作地连接的本公开内容的一个或多个RE (SEQ ID NO:1-2、13-17和22-41)。在一些情况下,转基因是以下中的任一种:Cas9、因子VIII、DNA结合蛋白、激素、生长/分化因子,例如胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;调控或调节免疫系统的治疗性蛋白质,例如,细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体;囊性纤维化跨膜调节因子;与粘多糖贮积症I、II、III和IV型相关的基因; β 珠蛋白;以及脂蛋白脂肪酶。在一些情况下,表达的转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体或片段。在一些情况下,表达的转基因是CDKL5、CNTNAP2、ZEB2,或其变体或片段。在一些情况下,转基因是纤维蛋白原、凝血酶原、凝固因子或凝血因子(例如,因子1、2、3、4、5、6、7、8、9、10、11或12)。

[0168] 本文公开的人类基因的核苷酸和蛋白质序列可以在各种公共数据库中找到,包括GenBank (NCBI) 或UniProKB数据库。例如,人类UniProtKB登录号包括:Q04656 (ATP7A); P35670 (ATP7B); 043520 (AT8B1); P21439 (MDR3,也称为ABCB4); 095342 (ABCB11); 076039 (CDKL5); Q9UHC6 (CNTNAP2,也称为CNTNAP2); 060315 (ZEB2); P12259 (F5或因子V); P00451 (F8或因子VIII); P00740 (F9或因子IX); P00742 (F10或因子X)。此类蛋白质序列中的任一个都可以由本文公开的表达盒中的核酸序列(或转基因)编码。

[0169] 重组蛋白可以在细菌细胞、酵母细胞、昆虫细胞、植物细胞或哺乳动物细胞中产生。在一些实例中,重组蛋白在哺乳动物细胞中产生。在一些情况下,蛋白质在动物模型中产生。可用于产生蛋白质的细胞系的实例包括中国仓鼠卵巢细胞(CHO)、小鼠骨髓瘤细胞(NSO)、幼仓鼠肾细胞(BHK)和B16黑素瘤细胞。在一些实例中,使用来源于人类的细胞系。可用于产生蛋白质的人类细胞系的实例包括HEK293T、HeLa、HT-1080和PER.C5。重组蛋白还可以在动物或动物产物中产生。重组蛋白还可以在植物中产生。重组蛋白可以在植物的叶、茎、花、果实或种子中表达。

[0170] 在一些实施方案中,本公开内容的调控元件可用于驱动或增加报道基因如萤光素酶的表达。在一些实施方案中,根据实施例2的方法测量,一个或多个调控元件可以导致可操作地连接的萤光素酶转基因以产生大于 10^{17} 、 10^{18} 、 10^{19} 、 10^{20} 或 10^{21} 光子/秒/小鼠的输出水平表达。

[0171] 本文还考虑到治疗单倍性不足或任何与基因表达和/或蛋白质活性或合成不足(或基因低表达)相关的病症或病况的方法,包括在有需要的受试者中施用表达盒或载体,其中该表达盒或载体包含与转基因可操作地连接的本公开内容的一个或多个RE (SEQ ID NO:1-2、13-17和22-41)。在一些情况下,转基因是DNA结合蛋白或导致内源基因表达增加的转录调节因子(例如,转录激活因子)。在一些情况下,表达盒或载体中的转基因是治疗性转基因。在一些情况下,治疗性转基因是以下中的任一种:Cas9、因子VIII、DNA结合蛋白、激素、生长/分化因子,例如胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;调控或调节免疫系统的治疗性蛋白质,例如,细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体;囊性纤维化跨膜调节因子;与粘多糖贮积症I、II、III和IV型相关的基因; β 珠蛋白;以及脂蛋白脂肪酶。在一些情况下,表达的转基因是ATP7A、ATP7B、AT8B1(或ATP8B1)、MDR3(或ABCB4)、ABCB11、CDKL5、

CNTP2 (或CNTNAP2)、ZEB2、因子V、因子VIII、因子IX、或因子X,或其变体、亚基、同种型或功能片段。进行性家族性肝内胆汁淤积 (PFIC) 包括PFIC1、PFIC2和PFIC3,其与ATP8B1、ABCB11和ABCB4基因突变相关联。进行性家族性肝内胆汁淤积2型 (PFIC2) 与ABCB11的缺陷或突变相关联。ATP8B1基因突变可导致PFIC1。ABCB11基因中的突变可导致PFIC2。ABCB4基因突变可导致PFIC3。在一些情况下,本文公开的转基因是CDKL5,并且用于基因疗法以治疗CDKL5缺陷。CNTNAP2中的突变或缺陷可导致自闭症谱系障碍。在一些情况下,转基因是纤维蛋白原、凝血酶原、凝固因子或凝血因子 (例如,因子1、2、3、4、5、6、7、8、9、10、11或12),或其变体或功能片段。在一些情况下,转基因优先在肝细胞中表达。在一些情况下,施用是全身的,例如,通过口服、静脉内、肌肉内、腹膜内、鞘内、肠内或肠胃外施用。在一些情况下,该方法包括通过静脉内输注或注射递送表达盒或载体。在一些情况下,施用是局部的,例如,通过注射到肝脏、CNS或体内器官中。在一些情况下,表达盒或载体是AAV,例如,AAV8。

[0172] 在一些情况下,本文考虑到治疗与基因的过表达相关联的病症或病况的方法,该方法包括在有需要的受试者中施用表达盒或载体,其中该表达盒或载体包含与转基因可操作地连接的本公开内容的一个或多个RE (SEQ ID NO:1-2、13-17和22-41)。在一些情况下,转基因是DNA结合蛋白或导致内源基因表达降低的转录调节因子 (例如,转录抑制因子)。在一些情况下,转基因优先在肝细胞中表达。在一些情况下,施用是全身的,例如,通过口服、静脉内、肌肉内、腹膜内、鞘内、肠内或肠胃外施用。在一些情况下,该方法包括通过静脉内输注或注射递送表达盒或载体。在一些情况下,施用是局部的,例如,通过注射到肝脏、CNS或体内器官中。在一些情况下,表达盒或载体是AAV,例如,AAV8。

[0173] 在一些情况下,本文考虑到治疗与以下基因中的任一个的缺陷相关联的遗传病症或病况的方法:ATP7A、ATP7B、AT8B1 (或ATP8B1)、MDR3 (或ABCB4)、ABCB11 (或ABCB11)、CDKL5、CNTP2 (或CNTNAP2)、ZEB2、因子V、因子VIII、因子IX、因子X、纤维蛋白原、凝血酶原、凝固因子或凝血因子 (例如,因子1、2、3、4、5、6、7、8、9、10、11或12)。在一些情况下,本文公开了治疗威尔逊病的方法。在一些情况下,治疗威尔逊病或者与ATP7A、ATP7B、AT8B1 (或ATP8B1)、MDR3 (或ABCB4)、ABCB11 (或ABCB11)、CDKL5、CNTP2 (或CNTNAP2)、ZEB2、因子V、因子VIII、因子IX、因子X、纤维蛋白原、凝血酶原、凝固因子或凝血因子 (例如,因子1、2、3、4、5、6、7、8、9、10、11或12) 相关的任何其他遗传病症或病况的方法包括将本文公开的表达盒或载体施用于有需要的受试者 (例如,动物或人类),其中该表达盒或载体包含与本公开内容的转基因可操作地连接的本公开内容的一个或多个RE,该转基因例如,ATP7A、ATP7B、AT8B1 (或ATP8B1)、MDR3 (或ABCB4)、ABCB11 (或ABCB11)、CDKL5、CNTP2 (或CNTNAP2)、ZEB2、因子V、因子VIII、因子IX、因子X、纤维蛋白原、凝血酶原、凝固因子、凝血因子 (例如,因子1、2、3、4、5、6、7、8、9、10、11或12)、DNA结合蛋白或转录调节因子。在一些情况下,转基因是编辑内源基因的Cas蛋白 (例如,Cas9或saCas9蛋白) 和/或指导RNA (例如,Cas从内源基因去除突变)。在一些情况下,表达盒或载体优先在肝脏中表达转基因。在一些情况下,施用是全身的,例如,通过静脉内输注或注射。在一些情况下,施用是局部的,例如,直接注射到组织或器官中,例如,注射到肝脏或CNS中。

[0174] 在一些情况下,本文公开的表达盒或载体用于治疗威尔逊病,其中该表达盒或载体包含与治疗性转基因可操作地连接的一个或多个RE (例如,SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41),其中该

转基因是ATP7B或者作用于内源ATP7B基因的转录调节因子或DNA结合蛋白,以增加ATP7B在细胞中或体内的表达。在一些情况下,将这样的表达盒或载体递送至诊断患有威尔逊病或有威尔逊病风险的受试者。在一些情况下,这样的表达盒或载体通过静脉内输注或注射递送至受试者。在一些情况下,递送是全身的。在一些情况下,治疗威尔逊病的方法包括对有需要的受试者施用基因疗法(例如,AAV基因疗法),其中该基因疗法包含与治疗性转基因可操作地连接的一个或多个RE(例如,SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41),其中该转基因是ATP7B或作用于内源ATP7B基因的转录调节因子,以增加ATP7B在细胞中或体内的表达。在一些情况下,表达盒或载体是AAV,例如,AAV8。在一些情况下,表达盒或载体优先在肝细胞或肝脏中表达。

[0175] 在一些情况下,本文公开的表达盒或载体用于治疗凝血障碍,其中该表达盒或载体包含与治疗性转基因可操作地连接的一个或多个RE(例如,SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41),其中该转基因是纤维蛋白原、凝血酶原、凝固因子或凝血因子(例如,因子1、2、3、4、5、6、7、8、9、10、11或12),或者作用于与纤维蛋白原、凝血酶原、凝固因子或凝血因子(例如,因子1、2、3、4、5、6、7、8、9、10、11或12)相对应的内源基因的转录调节因子。在一些情况下,该转基因是因子VIII。在一些情况下,将这样的表达盒或载体递送至诊断患有凝血障碍或有凝血障碍风险的受试者。在一些情况下,这样的表达盒或载体通过静脉内输注或注射递送至受试者。在一些情况下,递送是全身的。在一些情况下,治疗凝血障碍的方法包括向有需要的受试者施用基因疗法(例如,AAV基因疗法),其中该基因疗法包括与治疗性转基因可操作地连接的本公开内容的一个或多个RE(例如,SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41),其中该转基因是因子VIII或作用于内源因子VIII基因的转录调节因子,以增加其在细胞中或体内的表达。在一些情况下,该转基因是因子1、2、3、4、5、6、7、8、9、10、11或12中的任一个。在一些情况下,表达盒或载体是AAV。在一些情况下,表达盒或载体优先在肝细胞或肝脏中表达。

[0176] 在一些情况下,该转基因可以是DNA结合蛋白,或与DNA结合相关联的蛋白。在一些情况下,DNA结合蛋白可以是转录因子,或转录因子的部分。在一些情况下,DNA结合蛋白可以是Cas9、Cas家族蛋白、saCas9、dCas9、dCas家族蛋白、转录激活因子样效应因子(TALE)蛋白、或锌指蛋白、或融合蛋白。在一些情况下,治疗遗传疾病的方法包含施用表达盒或载体,该表达盒或载体包含基因编辑蛋白(例如,Cas9和指导RNA)以及本公开内容的一个或多个RE(例如,SEQ ID NO:1、2、3、4、13、14、15、16、17、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41)。在一些情况下,指导RNA将Cas蛋白靶向于感兴趣的内源基因或本文公开的内源基因。

[0177] 在一些情况下,本文公开的构建体被用于递送基因疗法。特别地,本公开内容的相对较短的调控元件可以对需要表达较大转基因如因子VIII或ATP7B的基因疗法特别有用。在一些实施方案中,较大转基因在肝脏细胞如肝细胞中表达。在其他实施方案中,较大转基因在中枢神经系统(CNS)如神经元中表达。在一些实施方案中,本文公开的转基因中的任一种与本文公开的一个或多个调控元件可操作地连接,以增加转基因的表达。在一些情况下,本文公开的转基因在细胞中或体内、体外和/或离体表达。在一些情况下,此类较大转基因使用基因疗法、重组表达载体或质粒或者病毒载体或质粒递送至细胞中。可作为基因疗法

或者离体或体外表达系统的一部分表达的较大转基因的实例包括,但不限于:门克斯(Menkes)蛋白(ATP7A)、ATP结合盒亚家族b成员4(ABCB4)、ATP酶磷脂转运8B1(ATP8B1)、ATP酶铜转运 β (ATP7B)、细胞周期蛋白依赖性激酶样5(CDKL5)、接触蛋白相关蛋白样2(CNTNAP2)、ATP结合盒亚家族B成员11(ABCB11)和锌指E-box结合同源异型框2(ZEB2),及其任何变体、亚基、突变体或功能片段,包括密码子优化的变体。

[0178] 在一些实施方案中,本文公开的构建体被用于基因编辑。例如,构建体可包含本文公开的调控元件、Cas9编码序列和指导RNA。在一些情况下,Cas9编码序列编码saCas9。具有Cas9或saCas9编码序列和指导RNA的构建体可用于校正功能异常的基因,例如,通过校正突变(例如,去除过早的终止密码子)。

[0179] 本文还考虑到包含本文公开的一个或多个调控元件的基因疗法或表达载体,以治疗各种疾病,包括各种遗传病症,如重症联合免疫缺陷(ADA-SCID)、门克斯病、家族性肝内胆汁淤积、拜勒综合症、威尔逊病、皮质发育不良-局灶性癫痫综合症、Mowat-Wilson综合征、PFIC2、PFIC3、慢性肉芽肿病、血友病、先天性失明、代谢紊乱、溶酶体贮积病、肌肉营养不良症、癌症、病毒感染、心脏病和糖尿病。在一些情况下,包含本文公开的一个或多个调控元件的基因疗法可用于治疗威尔逊病。

[0180] 在一些实施方案中,本文公开的调控元件被用于治疗患有凝血障碍的受试者。例如,含有本文公开的调控元件和FVIII的编码序列的表达盒可用于拯救具有FVIII缺陷的受试者中的凝血障碍或疾病。如图5A和图10所示,将含有本公开内容的调控元件和FVIII的编码序列的表达盒转染到患有凝血障碍的受试者中可导致受伤后的失血减少。将含有本文公开的调控元件和FVIII的编码序列的表达盒转染到患有凝血障碍的受试者中可导致至少约10%、20%、30%的失血。如图5B和图11所示,将含有本文公开的调控元件和FVIII的编码序列的表达盒转染到患有凝血障碍的受试者中可导致受伤后出血到停止的时间减少。

[0181] 可以用本公开内容的构建体、基因疗法或表达载体/盒中的任一种治疗的受试者包括人类、灵长类动物、狗、猫、啮齿动物、马、牛、羊、猪和其他家畜。在一些情况下,受试者患有疾病/病症或者有患上疾病/病症的风险。例如,受试者可以患有遗传疾病,或具有已知与遗传疾病或该疾病的易感性相联系的DNA序列。在一些情况下,受试者患有血友病。在一些情况下,受试者患有血友病A。在一些实施方案中,受试者具有因子VIII的缺陷。在一些实施方案中,包含与因子VIII可操作地连接的本文所述的一个或多个调控元件的基因疗法被用于治疗诊断患有血友病A的受试者。

[0182] 实施例

[0183] 提供这些实施例仅出于说明的目的,并非限制本文提供的权利要求书的范围。

[0184] 实施例1

[0185] HEK293T细胞中的高表达

[0186] 用含有在几种不同调控元件之一的控制下的萤光素酶基因的质粒DNA转染HEK293T细胞,该调控元件即无启动子对照、SCP、CMV、SEQ ID NO:3(与minCMV可操作地连接的SEQ ID NO:1)、SEQ ID NO:4(与minCMV可操作地连接的SEQ ID NO:2)和CAG。图1A示出了来自每个构建体的标准化萤光素酶值。图1B示出了来自每个构建体的大小标准化活性值。本文中使用的调控元件和启动子的序列提供于以上表1和以下表3。调控元件连接至minCMV启动子的SEQ ID NO:1和调控元件连接至minCMV启动子的SEQ ID NO:2驱动比单独的

minCMV和单独的SCP更高水平的萤光素酶表达。与对照、SCP或没有调控元件(即,SEQ ID NO:1或2)的minCMV启动子相比,当连接至minCMV启动子时,SEQ ID NO:1和SEQ ID NO:2都驱动HEK293T肾细胞中萤光素酶的高表达。

[0187] 使用其他对照和调控序列进行了相似的标准化萤光素酶表达实验,如图1C、图1D和图1E所示。还测定了萤火虫的表达,以确保在所有测试样品中转染效率相似。在图1C中,将调控元件SEQ ID NO:4、SEQ ID NO:17和22组合(或SEQ ID NO:17/22)、SEQ ID NO:16和23组合(或SEQ ID NO:16/23)、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:32和SEQ ID NO:33的标准化萤光素酶表达与两个阴性对照(即已知不驱动任何表达的序列)进行比较。每一个测试的调控元件都导致高水平的萤光素酶表达,例如,与阴性对照相比,为至少10、50或100或更多倍。

[0188] 在图1D中,将来自包含调控元件SEQ ID NO:4、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:26或SEQ ID NO:27的质粒的标准化萤光素酶表达与阴性对照以及包含连接至CMVe的CMV或与萤光素酶可操作地连接的CMV的相似质粒进行比较。每一个测试的调控元件(即,SEQ ID NO:4、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:26和SEQ ID NO:27)都导致比阴性对照、单独的CMV以及CMV+CMVe更高的萤光素酶表达。与包含连接至萤光素酶的CMV启动子或CMV+CMVe的质粒相比,测试的相对较短的RE显示出至少10倍、至少15倍、至少20倍、至少30倍、至少40倍或大于50倍的标准化萤光素酶表达。

[0189] 在图1E中,将来自包含与萤光素酶可操作地连接的调控元件SEQ ID NO:17/22、SEQ ID NO:16/23、SEQ ID NO:34、SEQ ID NO:35、SEQ ID NO:36、SEQ ID NO:37、SEQ ID NO:38、SEQ ID NO:39、SEQ ID NO:40或SEQ ID NO:41的质粒的标准化萤光素酶表达与阴性对照(即,已知不具有表达活性的序列)进行比较。每一个调控元件(即,SEQ ID NO:17/22、16/23和34-41)都驱动比阴性对照更高的萤光素酶表达,而SEQ ID NO:35和SEQ ID NO:39中的每一个都可驱动高于 10^2 的标准化萤光素酶表达,或比阴性对照高至少100倍的萤光素酶表达。

[0190] 表3:本文公开的附加序列

SEQ ID NO.	序列 (5'-至-3')	名称
5	GTACTTATATAAGGGGGTGGGGGCGCGTTTCGTCCTCAGTCGCGATCGAACACTCGAGCCGAGCAGACGTGCCTACGGACC	SCP
6	GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCCAGTTATCGGAGGAGCAAACAGGGGCTAAGTCCACGCTAGCGTCTGTCTGCACATTTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCTAGGCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAGCCGTC	SerpE_TTR
7	GTTTGCTGCTTGCAATGTTTGCCCATTTTAGGGTGGACACAGGACGCTGTGGTTTCTGAGCCAGGGCTAGCGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGTTGCCCTCTGGATCCACTGCTTAAATACGGACGAGGAC	Proto1

SEQ ID NO.	序列 (5'-至-3')	名称
	AGGGCCCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTG GGACAGTGAATCGCCAC	
8	TGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCG GTTTGACTCACGGGGATTTC AAGTCTCCACCCCATTTGACG TCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTT CCAAAATGTCGTAACA ACTCCGCCCATTTGACGCAAAATGG GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGC T	minCMV
9	GTTTGCTGCTTGCAATGTTTGCCCATTTTAGGGTGGACACA GGACGCTGTGGTTTCTGAGCCAGGGGGCGACTCAGATCCC AGCCAGTGGACTTAGCCCTGTTTGCTCCTCCGATAACTGG GGTGACCTTGTTAATATTACCAGCAGCCTCCCCCGTTGC CCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGC CCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTGGGACA GTGAATC	UCL-HLP
10	CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC CCAACGACCCCCGCCCATTTGACGTCAATAATGACGTATGTT CCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATG GGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAAT GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA CCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATG	CMVe
11	GTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCC CAACGACCCCCGCCCATTTGACGTCAATAATGACGTATGTT CCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGG GTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCA AGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATG ACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC CTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG TCATCGCTATTACCATGGGTGAGGTGAGCCCCACGTTCTG CTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTT GTATTTATTTATTTTAAATTATTTTGTGCAGCGATGGGGGC GGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGG GCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGG CAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTATG GCGAGGCGGCGGGCGGGCGGCGGCCCTATAAAAAGCGAAGC GCGCGGCGGGCGGGAGTCGCTGCGTTGCCTTCGCCCCGTG CCCCGCTCCGCGCCGCCTCGCGCCCGCCCGCCCGGCTCTGA CTGACCGCGTTACTCCACAGGTGAGCGGGCGGGACGGCC CTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGG CTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTAAAGGGCT CCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGG TGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGC CCGCGCTGCCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGC GGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGG CCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAGGGG AACAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAG CAGGGGGTGTGGGCGCGGCGGTGCGGCTGTAACCCCCCCC	CAG

[0192]

SEQ ID NO.	序列 (5'-至-3')	名称
[0193]	TGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCG GGTGCGGGGCTCCGTGCGGGGCGTGCGCGGGGGCTCGCCG TGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCG GGGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGG GGCGCGGCGGCCCCGAGCGCCGCGGCTGTCGAGGCGCG GCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAG GGCGCAGGGACTTCCTTTGTCCCAAATCTGGCGGAGCCGA AATCTGGGAGGCGCCGCCGCACCCCTCTAGCGGGCGCGG GCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGG GAGGGCCTTCGTGCGTGCCTCGCCGCGCCCGTCCCTTCTCCA TCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTTC GGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGT GACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATGCCT TCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTGTTG TGCTGTCTCATCATTTTGGCAAAGAATT	
12	GCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCAC AGTCCCCGAGAAGTTGGGGGGAGGGGTTCGGCAATTGAACC GGTGCCTAGAGAAGGTGGCGCGGGGTAACTGGGAAAGTG ATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGGGGGA GAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTT TCGCAACGGGTTTGCCGCCAGAACACAGG	EFS

[0194] 实施例2

[0195] 全局高驱动

[0196] 为了评估调控元件SEQ ID NO:1和SEQ ID NO:2的表达模式,制备了重组AAV载体,其包含与萤光素酶连接的每个调控元件。使用的调控元件是:SCP、SERpE_TTR、Proto1、CMV、UCL-HLP、CMVe、SEQ ID NO:1和SEQ ID NO:2。通过流体动力尾静脉注射将质粒DNA递送至SCID-beige小鼠(每只小鼠12 μ g表达载体),并通过在注射后48小时向小鼠注射5ug (~0.25mg/kg)的furimazine来测定萤光素酶的表达。图2A示出了在不同调控元件控制下的萤光素酶的表达,并示出了当每个RE连接至minCMV时,SEQ ID NO:1和2能够在至少一种附加细胞类型以及肾细胞中驱动可操作地连接的转基因的表达,如白色区域所示,该区域指示萤光素酶在小鼠各种组织中的表达。如图2B所示,与其他调控元件相比,调控元件SEQ ID NO:3和SEQ ID NO:4各自导致高水平的萤光素酶表达,如通过以光子/秒为单位的萤光素酶发光所测量的。

[0197] 实施例3

[0198] 人类因子VIII的表达

[0199] 向SCID-beige小鼠注射16 μ g表达盒DNA,其含有与UCL-HLP可操作地连接的因子VIII(NIH蛋白质登录ID ABV90867.1)的编码序列;minCMV与SEQ ID NO:1;或minCMV与SEQ ID NO:2。至少八只小鼠注射了这3种质粒中的每一种,并且至少另外八只小鼠注射了缓冲液。在注射质粒DNA后24小时,获取血液样品并使用已建立的方法(COATEST®SP4 FVIII,Chromogenix;Visualize™ DVIII抗原试剂盒,Affinity Biologicals INC.)分析血浆的VIII因子活性和浓度。如图3A所示,minCMV与SEQ ID NO:1和minCMV与SEQ ID NO:2表达盒都导致因子VIII的较高血浆浓度,为1.5IU/ml或更高。图3B示出,SEQ ID NO:3(minCMV与SEQ ID NO:1)和SEQ ID NO:4(minCMV与SEQ ID NO:2)导致血浆因子VIII活性水平为

100%或更高。调控元件SEQ ID NO:1和/或2当连同minCMV与因子VIII编码序列可操作地连接时,导致血液因子VIII活性水平分别为114%和166%,比UCL-HLP启动子所见的活性高约10倍和15倍。

[0200] 实施例4

[0201] 基因编辑

[0202] 开发了多合一重组AAV病毒粒子以删除转基因小鼠Ai9 (Jax库存号:007909) 中tdTomato上游的终止盒。一旦终止盒被saCas9删除,tdTomato在CAG启动子的控制下表达。rAAV被开发为含有在EFS启动子、CMV启动子或CMV启动子与SEQ ID NO:2的控制下的saCas9。rAAV还包含在U6启动子控制下的两个Sa指导RNA支架。将这些rAAV病毒粒子注射到上述转基因小鼠的齿状回中。3周后将大脑固定并切片,并评估tdTomato在齿状回中的表达。如图4B所示,CMV与SEQ ID NO:2 (或SEQ ID NO:4) 比单独的CMV或EFS导致更多的tdTomato表达细胞。图4A示出了SEQ ID NO:2当与CMV连接时 (或SEQ ID NO:4) 至少在脑细胞 (例如,神经元) 的亚群中驱动连接的转基因 (RFP) 的表达。

[0203] 实施例5

[0204] 凝血障碍的基因疗法治疗

[0205] 基因工程化的FVIII敲除小鼠 (F8tm1Srcmo, F8基因的外显子16-19缺失;获自中国上海的上海模式生物中心有限公司) 具有凝血缺陷,该凝血缺陷可以通过测量通过剪掉尾的末端并使其流血直至发生凝血所释放的血量或者直至发生凝血所经过的时间来测定。向FVIII敲除小鼠注射磷酸盐缓冲盐水 (PBS对照) 或含有驱动FVIII转基因表达的SEQ ID NO:4的重组AAV病毒粒子。以两种不同剂量注射rAAV病毒粒子: 3^{11} gc/kg或 3^{12} gc/kg。注射后三周,将经治疗的FVIII敲除小鼠剪尾,并测量直至凝血的时间和失血量。一组非敲除小鼠也被剪尾,并评估了直至凝血的血量和时间。如图5A所示,PBS处理的FVIII敲除小鼠的失血量是未治疗的非敲除小鼠的超过两倍。然而,rAAV治疗的FVIII敲除小鼠失去比PBS治疗的小鼠更少的血液。在较高的rAAV剂量下,治疗的小鼠与未治疗的非敲除小鼠的失去相同的血量。如图5B所示,出血时间测定观察到相似的结果。虽然较低的rAAV剂量未显示对出血时间的影响,但含有与FVIII可操作地连接的SEQ ID NO:4的较高剂量的rAAV将出血时间减少到几乎与非敲除小鼠的时间相同。

[0206] 实施例6

[0207] 凝血障碍的基因疗法治疗

[0208] 向FVIII敲除小鼠 (如实施例5所述) 注射磷酸盐缓冲盐水或含有驱动FVIII表达的SEQ ID NO:13-17中的每一个的重组AAV病毒粒子,或含有驱动FVIII表达的SEQ ID NO:9 (UCL-HLP) 的对照病毒粒子。图8示出了未治疗的敲除小鼠 (KO) 和用各种病毒粒子治疗的小鼠在转染三周后血液中FVIII的浓度。包含SEQ ID NO:13、SEQ ID NO:14或SEQ ID NO:17的病毒粒子导致FVIII在体内的最高表达;与其他调控元件相比,SEQ ID NO:15导致FVIII的中等表达;并且与其他序列相比,SEQ ID NO:16导致FVIII的低表达。然而,SEQ ID NO:16仍然导致比UCL-HLP启动子高约2倍的FVIII浓度。

[0209] 还如实施例3所述测定表达的FVIII的活性,如图9和表4所示。不同病毒粒子的活性水平遵循与表达水平相同的模式。

[0210] 表4:转染小鼠血液中FVIII的活性

[0211]

	KO	SEQ ID NO: 13	SEQ ID NO: 14	SEQ ID NO: 15	SEQ ID NO: 16	SEQ ID NO: 17	SEQ ID NO: 9
不同小鼠中 FVIII 的活性 水平 (野生型的百分比)	-1.4019	-1.1497	172.7369	19.9399	1.0077	5.2719	-0.7564
	-1.412	93.1133	46.0407	14.038	0.8576	6.1928	-0.4091
	-1.0884	62.7729	198.6102	3.0212	0.8175	20.677	-0.2897
	-1.9783	2.2538	4.9381	9.5587	0.6174	23.1995	-0.7022
	-2.1097	86.574	6.5522	12.6458	2.6694	12.9859	-0.8216
	-1.8427	14.8857	35.0801	4.5446	3.4701	154.5709	-0.4308
	-1.9392	64.9174	13.806	12.3936	2.0187	53.9879	0.3508
	-1.6585	12.4836	31.4121	2.9909	-0.8641	15.8733	-0.7347
	-1.5796	1.2449	26.6502	18.2246	-0.794	44.1145	-0.5285
平均值	-1.6678	37.4551	53.7062	10.62227	1.0818	40.00547	-0.45142

[0212] 注射后三周,将经治疗的FVIII敲除小鼠剪尾,并测量直至凝血的时间和血量。一组非敲除小鼠也被剪尾,并评估直至凝血的血量和时间。如图10和表5所示,PBS治疗的FVIII敲除小鼠的失血量是未治疗的非敲除小鼠(或WT小鼠)的超过两倍。然而,rAAV治疗的FVIII敲除小鼠失去比PBS治疗的小鼠更少的血液。接受含有SEQ ID NO:17的病毒粒子的小鼠失去与未治疗的非敲除小鼠相似的血量。出血时间测定中观察到相似的结果,如图11和表6所示。用包含SEQ ID NO:15的病毒粒子治疗的小鼠的出血时间显示出血时间和失血最强的减少。

[0213] 表5:不同病毒粒子转染的小鼠中凝血前的失血量

[0214]

	KO	SEQ ID NO: 13	SEQ ID NO: 14	SEQ ID NO: 15	SEQ ID NO: 16	SEQ ID NO: 17	SEQ ID NO: 9	WT
来自不同小鼠的失血量 (g)	0.0814	0.3264	0.1107	0.0223	0.3606	-0.0114	0.5145	0.0447
	0.4122	0.05	0.1479	0.1471	0.1367	0.0145	1.1232	0.0001
	-0.0522	0.6822	0.2897	0.0136	0.6372	0.0341	0.4819	-0.0357
	0.6397	0.0178	0.0526	0.0449	0.219	0.0575	0.486	0.0836
	0.3505	0.0356	0.2819	0.3934	0.2606	0.0963	0.4708	0.1065
	0.4982	0.1057	0.1665	0.0331	0.201	0.0837	0.0806	0.0693
	-0.0116	0.2129	0.0934	0.0919	0.0827	0.0288	0.34	0.061
	0.1827	0.0133	0.197	0.0748	0.438	0.103	0.2327	0.0423
	0.2793	0.0573	0.0355	0.0782	0.0275	0.0223	0.4931	0.0919
平均值	0.264	0.167	0.154	0.094	0.251	0.062	0.425	0.060

[0216] 表6:不同病毒粒子转染的小鼠直至凝血的出血时间

[0217]

	KO	SEQ ID NO: 13	SEQ ID NO: 14	SEQ ID NO: 15	SEQ ID NO: 16	SEQ ID NO: 17	SEQ ID NO: 9	WT
不同小鼠的出血时间 (min)	6.42	13.25	3.28	6.78	11.22	2.38	27.63	4.55
	25.23	8.3	11.02	9.5	15.6	1.95	26.88	0.38
	0	28.4	14.78	1.28	24.27	18.93	20.4	0.63
	26.17	0.92	2.52	3.62	27.5	17.73	20.2	5.15
	19.98	4.22	19.02	13.28	19.17	6.7	30	4.6
	30	16.95	3.27	1.15	23.5	9.95	13.15	2.13
	14.65	30	6.38	14.1	13.23	3.87	29.67	3.7
	21.93	7.58	8.17	1.58	22.45	21.43	14.43	4.68
	22.35	17.82	4.73	11.52	12.43	8.5	30	2.33
			14.6	0.35	13.32	13.33	1.66	8.63
平均值	18.52	14.16	8.13	6.97	18.81	10.16	23.59	3.12

[0218] 实施例7

[0219] 肝脏中的ATP7B表达

[0220] 向C57BL6/J小鼠静脉内注射含有在SEQ ID NO:13控制下的ATP7B的AAV8载体。病毒递送后两周,收获肝脏。将肝穿孔打碎,并在补充有10u1/mLβ-巯基乙醇的600u1 RLT中匀浆,并使用RNeasy Mini试剂盒(Qiagen)根据制造商方案,包括柱上DNA酶处理,提取RNA。对于每个样品,使用OligoDT引物,用Superscript IV(Invitrogen)逆转录RNA(3ug)(65°C 5m,55°C 10m,85°C 10m)。针对人类ATP7B(5'-CATTCCAGGACTGTCCATTCT-3'(SEQ ID NO:18);5'-GGCCTGAACGTAGAAGTACCA-3'(SEQ ID NO:19))以及GAPDH(5'-ACCACAGTCCATGCCATCAC-3'(SEQ ID NO:20);5'-TCCACCACCCTGTTGCTGTA-3'(SEQ ID NO:21))用标准曲线进行验证以确保可靠性,并在以下扩增条件(98°C 2m,40X[98°C 10s,67°C 30s,92°C 15s])下用于qPCR(Phusion,SYBR Green)。将ATP7B表达相对于GAPDH标准化,并使用δ-δCt法表示为相对于基线的倍数变化。如图12和表7所示,SEQ ID NO:13以三种不同剂量驱动肝脏中ATP7B的表达,并显示出最高病毒剂量导致最高ATP7B表达的剂量响应。

[0221] 表7:转染小鼠肝脏中的ATP7B表达

[0222]

	缓冲液	SEQ ID: 13; 1E10 gc/小鼠	SEQ ID: 13; 1E11 gc/小鼠	SEQ ID: 13; 1E12 gc/小鼠
转染小鼠肝脏中的 相对 ATP7B 表达 (log2)	3.214475452	111.7850412	87.30041862	5639.103082
	1.992504985	7.333913922	375.1308235	6142.385453
	2.857162979	20.64780715	306.8202728	1207.591395
	3.20705701	27.24494487	183.2814118	4273.64098
	0.339445061	1.698879272	3.813855015	2376.435176
	0.45520343	0.501590988	2.734448707	1352.347313
	0.405540152		19.17626705	26577.47941
	0.271919214		0.5721968	128.407293
平均值	1.592913535	28.20202957	122.3537118	5962.17376

[0223] 实施例7说明,包含与ATP7B转基因可操作地连接的本文公开的一个或多个RE的表达盒可用于在动物、哺乳动物或有需要的人类受试者中治疗与ATP7B缺陷相关的病症或病况,例如威尔逊病。此外,这样的表达盒可以使用rAAV8在体内全身递送,例如,通过静脉内注射或输注。

[0224] 实施例8

[0225] 肝脏中的高EGFP转录

[0226] 向C57BL6/J小鼠静脉内注射含有在各种RE的控制下的EGFP的AAV8载体,该RE为:SEQ ID NO:4、13、14、15、16、17和27。每个AAV8载体以 5×10^{11} gc/小鼠的剂量施用。病毒递送两周后,收获肝脏。将肝穿孔打碎,并在补充有10u1/m L β -巯基乙醇的600u1 RLT中匀浆,并使用RNeasy Mini试剂盒(Qiagen)根据制造商协议,包括柱上DNA酶处理,提取RNA。对于每个样品,使用OligoDT引物用Superscript IV(Invitrogen)逆转录RNA(3ug) (65°C 5m, 55°C 10m, 85°C 10m)。针对EGFP (5' -GCTACCCCGACCACATGAAG-3' (SEQ ID NO:42); 5' -TCTTGTAGTTGCCGTCGTCC-3' (SEQ ID NO:43)和GAPDH(SEQ ID NO:20和SEQ ID NO:21)的引物在以下扩增条件(98°C 2m, 40X[98°C 10s, 65°C 30s, 92°C 15s])下用于qPCR(Phusion, SYBR Green)。EGFP表达相对于GAPDH标准化,并使用 δ - δ Ct法表示为相对于基线的倍数变化(图13)。如图13所示,每个测试的RE在肝脏中增加EGFP表达,如通过RNA转录物所测量的。此外,SEQ ID NO:4、SEQ ID NO:27和SEQ ID NO:14导致相当水平的EGFP。

[0227] 实施例8说明,本公开内容的相对较短的RE可导致肝脏中高水平的转基因表达,如通过RNA转录物所测量的。

[0228] 尽管本文已经示出和描述了本发明的优选实施方案,但对于本领域技术人员显而易见的是,这些实施方案仅以实例的方式提供。在不偏离本发明的情况下,本领域技术人员现在将想到许多变化、改变和替代。应当理解,本文所述的本发明的实施方案的各种替代方案可以用于实施本发明。以下权利要求定义了本发明的范围,并且由此涵盖这些权利要求范围内的方法和结构及其等同项。

[0229] 本公开内容的各种实施方案参考以下编号的条款定义。

[0230] 1. 一种表达盒,其包含与治疗性转基因可操作地连接的调控元件,其中所述调控元件包含以下一个或多个:(i) SEQ ID NO:1-2、13-17和22-41;或(ii)与(i)中的任一个具有至少80%、85%、90%或95%序列同一性的序列。

[0231] 2. 根据条款1所述的表达盒,其中所述调控元件是非天然存在的。

[0232] 3. 根据条款1-2中任一项所述的表达盒,其中所述调控元件包含内含子序列。

[0233] 4. 根据条款1-3中任一项所述的表达盒,其中所述调控元件位于启动子和所述治疗性转基因之间。

[0234] 5. 根据条款1-2中任一项所述的表达盒,其中所述调控元件包含启动子序列。

[0235] 6. 根据条款5所述的表达盒,其中所述调控元件是所述盒中唯一的启动子。

[0236] 7. 根据条款1-6中任一项所述的表达盒,其中所述调控元件相对较短。

[0237] 8. 根据条款1-6中任一项所述的表达盒,其中所述调控元件不超过100bp。

[0238] 9. 根据条款1-8中任一项所述的表达盒,其中所述调控元件不超过60bp。

[0239] 10. 根据条款1-9中任一项所述的表达盒,其中所述调控元件不超过50bp。

[0240] 11. 根据条款1-10中任一项所述的表达盒,其中所述表达盒是rAAV的一部分。

[0241] 12. 根据条款11所述的表达盒,其中所述rAAV是rAAV8。

[0242] 13. 根据条款1-10中任一项所述的表达盒,其中所述表达盒是腺病毒的一部分。

[0243] 14. 根据条款1-10中任一项所述的表达盒,其中所述表达盒是慢病毒的一部分。

[0244] 15. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因大于1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb。

[0245] 16. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是ATP7B,或其

变体或功能片段。

[0246] 17. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是ATP7A,或其变体或功能片段。

[0247] 18. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是ATP8B1,或其变体或功能片段。

[0248] 19. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是ABCB4,或其变体或功能片段。

[0249] 20. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是ABCB11,或其变体或功能片段。

[0250] 21. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是CDKL5,或其变体或功能片段。

[0251] 22. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是CNTNAP2,或其变体或功能片段。

[0252] 23. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是ZEB2,或其变体或功能片段。

[0253] 24. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是因子V,或其变体或功能片段。

[0254] 25. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是因子VIII,或其变体或功能片段。

[0255] 26. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是因子IX,或其变体或功能片段。

[0256] 27. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是因子X,或其变体或功能片段。

[0257] 28. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是凝血因子。

[0258] 29. 根据条款28所述的表达盒,其中所述凝血因子是因子1、2、3、4、5、6、7、8、9、10、11或12中的任一个,或其变体或功能片段。

[0259] 30. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是DNA结合蛋白。

[0260] 31. 根据条款30所述的表达盒,其中所述DNA结合蛋白是内源基因的转录调节因子。

[0261] 32. 根据条款31所述的表达盒,其中所述转录调节因子是转录激活因子。

[0262] 33. 根据条款32所述的表达盒,其中所述内源基因在体内低表达。

[0263] 34. 根据条款30-33中任一项所述的表达盒,其中所述内源基因与肝脏疾病或病况相关联。

[0264] 35. 根据条款30-33中任一项所述的表达盒,其中所述内源基因与进行性家族性肝内胆汁淤积(PFIC)1型、2型或3型相关联。

[0265] 36. 根据条款30-33中任一项所述的表达盒,其中所述内源基因与自闭症谱系障碍相关联。

[0266] 37. 根据条款30-33中任一项所述的表达盒,其中所述内源基因与血友病相关联。

- [0267] 38. 根据条款30-33中任一项所述的表达盒,其中所述内源基因与威尔逊病相关联。
- [0268] 39. 根据条款30-33中任一项所述的表达盒,其中所述内源基因与门克斯病或枕骨角综合征相关联。
- [0269] 40. 根据条款30-33中任一项所述的表达盒,其中所述内源基因是CDKL5。
- [0270] 41. 根据条款35所述的表达盒,其中所述内源基因是ABCB11,其与进行性家族性肝内胆汁淤积2型(PFIC2)相关联。
- [0271] 42. 根据条款35所述的表达盒,其中所述内源基因是ATP8B1,其与进行性家族性肝内胆汁淤积1型(PFIC1)相关联。
- [0272] 43. 根据条款35所述的表达盒,其中所述内源基因是ABCB4,其与进行性家族性肝内胆汁淤积3型(PFIC3)相关联。
- [0273] 44. 根据条款36所述的表达盒,其中所述内源基因是CNTNAP2,其与自闭症谱系障碍相关联。
- [0274] 45. 根据条款37所述的表达盒,其中所述内源基因是因子VIII或其变体,其与血友病A相关联。
- [0275] 46. 根据条款37所述的表达盒,其中所述内源基因是因子IX或其变体,其与血友病B相关联。
- [0276] 47. 根据条款37所述的表达盒,其中所述内源基因是因子1、2、3、4、5、6、7、8、9、10、11或12中的任一个。
- [0277] 48. 根据条款38所述的表达盒,其中所述内源基因是ATP7B或其变体,其与威尔逊病相关联。
- [0278] 49. 根据条款39所述的表达盒,其中所述内源基因是ATP7A或其变体,门克斯病或枕骨角综合征。
- [0279] 50. 根据条款1-14中任一项所述的表达盒,其中所述治疗性转基因是基因编辑蛋白。
- [0280] 51. 根据条款50所述的表达盒,其中所述基因编辑蛋白是Cas蛋白。
- [0281] 52. 根据条款51所述的表达盒,进一步包含指导RNA,所述指导RNA将Cas蛋白靶向于包含内源基因的基因组位点。
- [0282] 53. 根据条款52所述的表达盒,其中所述内源基因包含突变或异常表达。
- [0283] 54. 根据条款53所述的表达盒,其中所述突变导致内源基因的低表达。
- [0284] 55. 根据条款51-54中任一项所述的表达盒,其中所述内源基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11、CDKL5、CNTNAP2、ZEB2或因子VIII。
- [0285] 56. 根据条款51-54中任一项所述的表达盒,其中所述内源基因是因子1、2、3、4、5、6、7、8、9、10、11或12。
- [0286] 57. 根据条款1-56中任一项所述的表达盒,其中与没有所述调控元件的表达盒相比,所述调控元件导致所述治疗性转基因或所述内源基因的表达增加至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍。

[0287] 58. 根据条款57所述的表达盒,其中所述增加的表达通过使用PCR测定的来自所述治疗性转基因或所述内源性的RNA转录物的量来确定。

[0288] 59. 根据条款57所述的表达盒,其中所述增加的表达通过使用免疫测定的从所述治疗性转基因或所述内源基因合成的蛋白质的水平来确定。

[0289] 60. 根据条款30-56中任一项所述的表达盒,其中与没有所述调控元件的表达盒相比,所述调控元件导致所述内源基因的表达增加至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍。

[0290] 61. 根据条款30-56中任一项所述的表达盒,其中所述调控元件导致与健康受试者中所述内源基因的表达相当的水平的所述内源基因增加的表达。

[0291] 62. 根据条款60-61中任一项所述的表达盒,其中所述增加的表达通过使用PCR测定的来自所述内源基因的RNA转录物的量来确定。

[0292] 63. 根据条款60-61中任一项所述的表达盒,其中所述增加的表达通过使用免疫测定的从所述内源基因合成的蛋白质的水平来确定。

[0293] 64. 一种治疗威尔逊病的方法,其包括施用根据条款1-16、30-34、38、48、50-55和57-63中任一项所述的表达盒。

[0294] 65. 一种治疗ATP7B中的遗传缺陷的方法,其包括施用根据条款1-16、30-34、38、48、50-55和57-63中任一项所述的表达盒。

[0295] 66. 一种治疗凝血障碍的方法,其包括施用根据条款1-15、24-34、37、45-47和50-63中任一项所述的表达盒。

[0296] 67. 一种治疗单倍性不足或遗传缺陷的方法,其包括施用根据条款1-66中任一项所述的表达盒。

[0297] 68. 根据条款67所述的方法,其中所述治疗性转基因或所述内源基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11、CDKL5、CNTNAP2、ZEB2或因子VIII,或其变体。

[0298] 69. 根据条款67所述的方法,其中所述治疗性转基因或所述内源基因是因子1、2、3、4、5、6、7、8、9、10、11或12,或其变体。

[0299] 70. 根据条款67或68所述的方法,其中所述转基因是ATP7B,或其变体或功能片段。

[0300] 71. 根据条款67或68所述的方法,其中所述转基因是调节内源基因表达的转录调节因子。

[0301] 72. 根据条款71所述的方法,其中所述内源基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11、CDKL5、CNTNAP2、ZEB2或因子VIII,或其变体。

[0302] 73. 一种产生重组蛋白或生物制剂的方法,其包括将根据条款1-63中任一项所述的表达盒引入细胞中。

[0303] 74. 一种AAV表达盒,其包含与至少3kb的转基因可操作地连接的不超过120bp的来源于人类的调控元件,其中与当可操作地连接CMV启动子时的所述转基因的表达相比,所述调控元件导致转基因表达增加至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少

900倍或至少1000倍。

[0304] 75. 一种AAV表达盒,其包含与至少3kb的转基因可操作地连接的相对较短的来源于人类的调控元件,其中与没有所述调控元件的相当的表达盒相比,所述相对较短的调控元件导致转基因表达增加至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少600倍、至少700倍、至少800倍、至少900倍或至少1000倍。

[0305] 76. 根据条款74或75所述的AAV表达盒,其中所述调控元件包含(i) SEQ ID NO:1-2、13-17和22-41;(ii) 其组合;或(iii) 或与(i)和(ii)中的任一个具有至少80%、85%、90%或95%序列同一性的序列。

[0306] 77. 根据条款74-76中任一项所述的AAV表达盒,其中所述增加的转基因表达是与当可操作地连接CMV启动子时的所述转基因的表达相比至少50倍。

[0307] 78. 根据条款77所述的AAV表达盒,其中所述增加的转基因表达是至少100倍。

[0308] 79. 根据条款74-78中任一项所述的AAV表达盒,其中与可操作地连接所述转基因的CMV启动子的大小标准化表达活性相比,所述调控元件表现出至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少25倍、至少50倍或至少100倍的大小标准化表达活性。

[0309] 80. 根据条款74-79中任一项所述的AAV表达盒,其中所述增加的转基因表达在至少2种细胞类型中发生。

[0310] 81. 根据条款80所述的AAV表达盒,其中所述至少2种细胞类型选自:肾细胞、上皮细胞、神经元和肝脏细胞。

[0311] 82. 根据条款74-81中任一项所述的AAV表达盒,其中所述调控元件包含SEQ ID NO:22-41中的任一个或多个,并且其中所述表达盒中不存在其他启动子序列。

[0312] 83. 根据条款74-82中任一项所述的AAV表达盒,其中所述调控元件包含SEQ ID NO:1或SEQ ID NO:2。

[0313] 84. 根据条款83所述的AAV表达盒,其中所述调控元件位于启动子的下游。

[0314] 85. 根据条款74-84中任一项所述的AAV表达盒,其中所述调控元件位于所述转基因的上游。

[0315] 86. 根据条款74-85中任一项所述的AAV表达盒,其中所述转基因是ATP7A;ATP7B;ATP8B1;ABCB4;ABCB11;CDKL5;CNTNAP2;ZEB2;因子1、2、3、4、5、6、7、8、9、10、11或12;或其变体或功能片段。

[0316] 87. 根据条款86所述的AAV表达盒,其中所述转基因是ATP7B,或其变体或功能片段。

[0317] 88. 根据条款86所述的AAV表达盒,其中所述转基因是FVIII,或其变体或功能片段。

[0318] 89. 根据条款74-85中任一项所述的AAV表达盒,其中所述转基因是基因编辑蛋白。

[0319] 90. 根据条款89所述的AAV表达盒,其中所述基因编辑蛋白是Cas。

[0320] 91. 根据条款74-85中任一项所述的AAV表达盒,其中所述转基因是DNA结合蛋白。

[0321] 92. 根据条款91所述的AAV表达盒,其中所述DNA结合蛋白是增加内源基因表达的

转录激活因子。

[0322] 93. 根据条款91所述的AAV表达盒,其中所述DNA结合蛋白是减少内源基因表达的转录抑制因子。

[0323] 94. 根据条款92所述的AAV表达盒,其中所述转录激活因子增加以下的表达: ATP7A; ATP7B; ATP8B1; ABCB4; ABCB11; CDKL5; CNTNAP2; ZEB2; 或因子1、2、3、4、5、6、7、8、9、10、11或12,或其变体。

[0324] 95. 根据条款94所述的AAV表达盒,其中所述转录激活因子增加内源性ATP7A的表达。

[0325] 96. 根据条款94所述的AAV表达盒,其中所述转录激活因子增加内源性ATP7B的表达。

[0326] 97. 根据条款94所述的AAV表达盒,其中所述转录激活因子增加内源性ATP8B1的表达。

[0327] 98. 根据条款94所述的AAV表达盒,其中所述转录激活因子增加内源性ABCB4的表达。

[0328] 99. 根据条款94所述的AAV表达盒,其中所述转录激活因子增加内源性ABCB11的表达。

[0329] 100. 根据条款94所述的AAV表达盒,其中所述转录激活因子增加内源性FVIII的表达。

[0330] 101. 根据条款74-100中任一项所述的AAV表达盒,其中所述AAV选自: AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ和scAAV。

[0331] 102. 根据条款101所述的AAV表达盒,其中所述AAV是AAV8。

[0332] 103. 一种产生重组蛋白的方法,所述方法包括将编码所述蛋白质的序列与以下一个或多个可操作地连接: (i) SEQ ID NO: 1-2、13-17和22-41; (ii) 其组合; 或 (iii) 与 (i) 和 (ii) 中的任一个具有至少80%、85%、90%或95%序列同一性的序列。

[0333] 104. 根据条款103所述的方法,其中编码所述蛋白质的所述序列大于1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb、5kb、5.5kb、6kb、6.5kb、7kb或7.5kb。

[0334] 105. 一种治疗肝脏疾病或病况的方法,其包括施用基因疗法,所述基因疗法包含与选自以下的一个或多个调控元件可操作地连接的治疗性转基因: (i) SEQ ID NO: 1-2、13-17和22-41; (ii) 其组合; 或 (iii) 与 (i) 和 (ii) 中的任一个具有至少80%、85%、90%或95%序列同一性的序列。

[0335] 106. 根据条款105所述的方法,其中所述肝脏疾病或病况是威尔逊病。

[0336] 107. 根据条款105所述的方法,其中所述肝脏疾病或病况是凝血障碍。

[0337] 108. 根据条款105所述的方法,其中所述转基因是ATP7A或ATP7B,或其变体或功能片段。

[0338] 109. 根据条款105所述的方法,其中所述转基因是因子1、2、3、4、5、6、7、8、9、10、11、12,或其变体或功能片段。

[0339] 110. 根据条款109所述的方法,其中所述转基因是因子8,或其变体或功能片段。

[0340] 111. 根据条款105-110中任一项所述的方法,其中所述调控元件导致至少2种细胞类型中增加的转基因表达。

[0341] 112. 根据条款150-110中任一项所述的方法, 其中所述调控元件导致至少3种细胞类型中增加的转基因表达。

[0342] 113. 根据条款105-112中任一项所述的方法, 其中所述调控元件导致肝细胞中增加的转基因表达。

[0343] 114. 根据条款105-113中任一项所述的方法, 其中所述调控元件导致与当可操作地连接CMV启动子时的转基因表达相比至少2倍水平的转基因表达增加。

[0344] 115. 根据条款105-114中任一项所述的方法, 其中所述基因疗法利用AAV。

[0345] 116. 根据根据条款115所述的方法, 其中所述AAV是AAV8。

[0346] 117. 根据条款105-114中任一项所述的方法, 其中所述基因疗法利用腺病毒。

[0347] 118. 根据条款105-114中任一项所述的方法, 其中所述基因疗法利用慢病毒。

[0348] 119. 一种表达载体, 其包含: 与转基因可操作地连接的具有小于或等于100bp的来源于人类的调控元件, 其中与不具有所述调控元件的第二表达载体相比, 所述调控元件使所述转基因的全局表达增加至少两倍。

[0349] 120. 根据条款119所述的表达载体, 其中所述第二表达载体包含与CMV启动子可操作地连接的所述转基因。

[0350] 121. 一种表达载体, 其包含与转基因可操作地连接的具有小于或等于100bp的来源于人类的调控元件, 借此所述转基因的表达高于由CMV启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子或CMVe启动子表达的相同转基因的表达。

[0351] 122. 根据条款121所述的表达载体, 其中所述转基因的所述表达高于由UCL-HLP启动子表达的相同转基因的表达。

[0352] 123. 一种表达载体, 其包含与转基因可操作地连接的来源于人类的调控元件, 其中由所述转基因编码的蛋白质具有(i) 通过被配置用于检测所述转基因的ELISA测定所测量的 $>1.0\text{IU/mL}$ 的浓度; 或(ii) 通过Coatest测定所测量的 $>25\%$ 的活性。

[0353] 124. 一种载体, 其包含来源于人类的调控元件, 所述调控元件具有大小为小于或等于100bp的序列, 与在细胞中或体内未发现与所述调控元件相关的转基因可操作地连接。

[0354] 125. 根据条款119-124中任一项所述的载体, 其中所述调控元件是内含子序列。

[0355] 126. 根据条款125所述的载体, 其中所述调控元件是SEQ ID NO:1或SEQ ID NO:2, 或与其具有至少80%、85%、90%或95%同源性的序列。

[0356] 127. 根据条款1-63、74-102和119-126中任一项所述的表达载体或盒, 其中使用荧光素酶作为所述转基因导致在整个小鼠中测量的大于 1×10^8 光子/秒的全局表达。

[0357] 128. 根据条款127所述的表达载体或盒, 其中所述活性或转基因表达对应于每只小鼠 $16\mu\text{g}$ 表达载体的剂量。

[0358] 129. 根据条款127所述的表达载体或盒, 其中所述活性或转基因表达对应于每只小鼠 $12\mu\text{g}$ 表达载体的剂量。

[0359] 130. 根据条款1-63、74-102和119-129中任一项所述的表达载体或盒, 其中所述转基因的内源形式在体内不与所述调控元件连接。

[0360] 131. 根据条款127所述的表达载体或盒, 其中所述转基因的所述全局表达处于高于使用带有选自以下的调控元件的载体或盒的所述转基因表达的水平: CMV启动子、CMVe启动子、超级核心启动子、TTR启动子、Proto 1启动子和UCL-HLP启动子。

[0361] 132. 根据条款1-63、74-102和119-131中任一项所述的表达载体或盒,其中在小鼠体内的至少3、4、5、6或7种不同的细胞类型中可检测到表达。

[0362] 133. 根据条款132所述的表达载体或盒,其中所述不同的细胞类型选自:肺泡细胞、心肌细胞、上皮细胞、肝细胞、肠细胞、肌细胞、神经元和肾脏细胞。

[0363] 134. 根据条款119-133中任一项所述的表达载体或盒,其中所述调控元件是增强子。

[0364] 135. 根据条款134所述的表达载体或盒,进一步包含启动子。

[0365] 136. 根据条款135所述的表达载体或盒,其中所述启动子是CMV启动子、CMV、启动子、超级核心启动子、TTR启动子、Proto 1启动子、UCL-HLP启动子、AAT启动子、KAR启动子、EF1 α 启动子、EFS启动子或CMVe增强子/CMV启动子组合。

[0366] 137. 根据条款119-136中任一项所述的表达载体,进一步包含一个或多个转录后修饰位点。

[0367] 138. 根据条款119-137中任一项所述的表达载体,其中所述转基因是Cas9。

[0368] 139. 根据条款138所述的表达载体,其中所述转基因是saCas9。

[0369] 140. 一种表达载体,其包含与调控序列可操作地连接的因子VIII转基因,所述调控序列能够驱动所述因子VIII的表达至 $>1.0\text{IU/mL}$ 的浓度,如通过被配置用于在细胞中或在体内检测因子VIII的ELISA测定所测量的。

[0370] 141. 根据条款140所述的表达载体,其中所述调控序列相对较短。

[0371] 142. 根据条款140或141所述的表达载体,其中所述调控序列包含以下中的一个或多个:SEQ ID NO:1-2、13-17和22-41或其组合;或与其具有至少80%、85%、90%或95%序列同一性的序列。

[0372] 143. 一种病毒颗粒,其包含根据条款119-142中任一项所述的表达载体。

[0373] 144. 根据条款143所述的病毒颗粒,其中所述病毒颗粒是AAV。

[0374] 145. 根据条款144所述的病毒颗粒,其中所述AAV选自:AAV1、AAV2、AAV3、AAV3b、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV-DJ和scAAV。

[0375] 146. 根据条款143所述的病毒颗粒,其中所述病毒颗粒是慢病毒。

[0376] 147. 根据条款143所述的病毒颗粒,其中所述病毒颗粒是腺病毒。

[0377] 148. 根据条款119-147所述的表达载体,其中所述转基因是治疗性转基因。

[0378] 149. 根据条款148所述的表达载体,其中所述治疗性转基因是因子VIII、Cas9、DNA结合蛋白、激素、生长或分化因子、胰岛素、生长激素、VEGF、神经营养因子、成纤维细胞上皮因子;细胞因子、白细胞介素、淋巴因子、肿瘤坏死因子、抗体、免疫球蛋白、干扰素、嵌合T细胞受体;脂蛋白受体、囊性纤维化跨膜调节因子、与粘多糖贮积症I、II、III或IV型相关的基因、 β 珠蛋白或脂蛋白脂肪酶。

[0379] 150. 根据条款148所述的表达载体,其中所述治疗性转基因是ATP7A、ATP7B、ATP8B1、ABCB4、ABCB11,或其变体或功能片段。

[0380] 151. 根据条款148所述的表达载体,其中所述治疗性转基因是CDKL5、CNTNAP2、ZEB2,或其变体或功能片段。

[0381] 152. 一种用于将转基因递送至动物中的多个不同组织的方法,其包括向所述动物施用条款119-151中任一项所述的表达载体。

- [0382] 153.一种用于产生蛋白质、抗体或其他生物制剂的方法,其包括使细胞与条款119-151中任一项所述的表达载体接触。
- [0383] 154.根据条款153所述的方法,其中所述细胞是CHO细胞或HEK293T细胞。
- [0384] 155.一种用于产生转基因动物或植物的方法,其包括向动物或植物施用条款119-151中任一项所述的表达载体。
- [0385] 156.根据条款119-151中任一项所述的表达载体,其中所述调控元件为40-50bp。
- [0386] 157.根据条款119-151中任一项所述的表达载体,其中所述调控元件为50-60bp。
- [0387] 158.一种表达载体,其包含与转基因可操作地连接的来源于人类的调控元件,其中由所述转基因编码的蛋白质具有(i)通过被配置用于检测所述转基因的ELISA测定所测量的>0.1IU/mL的浓度;或(ii)通过Coatest测定所测量的>10%的活性。
- [0388] 159.一种表达载体,其包含与转基因可操作地连接的具有小于或等于120bp的来源于人类的调控元件,借此所述转基因的表达高于由UCL-HLP启动子表达的相同转基因的表达。
- [0389] 160.根据条款158-159中任一项所述的表达载体,其中所述调控元件是启动子。
- [0390] 161.根据条款158-160中任一项所述的表达载体,其中所述治疗性转基因是包含DNA结合域和转录调节域的融合蛋白。
- [0391] 162.一种表达盒,其包含与治疗性转基因可操作地连接的相对较短的调控元件(RE)。
- [0392] 163.根据条款162所述的表达盒,其中所述相对较短的RE包含以下中的一个或多个:(i) SEQ ID NO:1-2、13-17和22-41;或(ii)与(i)中的任一个具有至少80%、85%、90%或95%序列同一性的序列。
- [0393] 164.根据条款64-72和105-118中任一项所述的方法,其中所述施用是全身的。
- [0394] 165.根据条款164所述的方法,其中所述施用包括在受试者中静脉内注射或输注。
- [0395] 166.根据条款165所述的方法,其中所述受试者是人类。
- [0396] 167.根据条款165所述的方法,其中所述受试者是动物。

- [0001] 序列表
- [0002] <110> 编码基因组学公司
- [0003] <120> 高活性调控元件
- [0004] <130> 46482-707.601
- [0005] <140>
- [0006] <141>
- [0007] <150> 62/508,968
- [0008] <151> 2017-05-19
- [0009] <160> 43
- [0010] <170> PatentIn version 3.5
- [0011] <210> 1
- [0012] <211> 56
- [0013] <212> DNA
- [0014] <213> 人工序列
- [0015] <220>
- [0016] <223> 人工序列的描述:合成
- [0017] 寡核苷酸
- [0018] <400> 1
- [0019] gtaaggtaag aattgaattt ctcagttgaa ggatgcttac actcttgtcc atctag 56
- [0020] <210> 2
- [0021] <211> 49
- [0022] <212> DNA
- [0023] <213> 人工序列
- [0024] <220>
- [0025] <223> 人工序列的描述:合成
- [0026] 寡核苷酸
- [0027] <400> 2
- [0028] gtgtgtatgc tcaggggctg ggaaaggagg ggaggagct ccggctcag 49
- [0029] <210> 3
- [0030] <211> 266
- [0031] <212> DNA
- [0032] <213> 人工序列
- [0033] <220>
- [0034] <223> 人工序列的描述:合成
- [0035] 多核苷酸
- [0036] <400> 3
- [0037] gtgatgcggt tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt 60
- [0038] ccaagtctcc accccattga cgtcaatggg agttttgttt ggcacaaaa tcaacgggac 120
- [0039] tttccaaaat gtcgtaacaa ctccgcccc ttgacgaaa tgggcggtag gcgtgtacgg 180
- [0040] tgggaggtct atataagcag agctggtacc gtaaggtaag aattgaattt ctcagttgaa 240
- [0041] ggatgcttac actcttgtcc atctag 266

- [0042] <210> 4
[0043] <211> 259
[0044] <212> DNA
[0045] <213> 人工序列
[0046] <220>
[0047] <223> 人工序列的描述:合成
[0048] 多核苷酸
[0049] <400> 4
[0050] gtgatgcggt tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt 60
[0051] ccaagtctcc accccattga cgtcaatggg agtttgtttt ggcacaaaa tcaacgggac 120
[0052] tttccaaaat gtcgtaacaa ctccgcccc ttgacgcaaa tgggcggtag gcgtgtacgg 180
[0053] tgggaggtct atataagcag agctggtagc gtgtgtatgc tcaggggctg ggaaaggagg 240
[0054] ggaggagct ccggctcag 259
[0055] <210> 5
[0056] <211> 80
[0057] <212> DNA
[0058] <213> 人工序列
[0059] <220>
[0060] <223> 人工序列的描述:合成
[0061] 寡核苷酸
[0062] <400> 5
[0063] gtacttatat aagggggtgg gggcgcttc gtcctcagtc gcgatcgaac actcgagccg 60
[0064] agcagacgtg cctacggacc 80
[0065] <210> 6
[0066] <211> 279
[0067] <212> DNA
[0068] <213> 人工序列
[0069] <220>
[0070] <223> 人工序列的描述:合成
[0071] 多核苷酸
[0072] <400> 6
[0073] ggggaggctg ctggtgaata ttaaccaagg tcacccagc tatcggagga gcaaacaggg 60
[0074] gctaagtcca cgctagcgtc tgtctgcaca tttcgtagag cgagtgttcc gataactctaa 120
[0075] tctccctagg caaggttcat atttgtgtag gttacttatt ctcttttgt tgactaagtc 180
[0076] aataatcaga atcagcaggt ttggagtcag cttggcaggg atcagcagcc tgggttggaa 240
[0077] ggagggggta taaaagcccc ttcaccagga gaagccgtc 279
[0078] <210> 7
[0079] <211> 262
[0080] <212> DNA
[0081] <213> 人工序列
[0082] <220>
[0083] <223> 人工序列的描述:合成

- [0084] 多核苷酸
- [0085] <400> 7
- [0086] gtttgctgct tgcaatgttt gcccatTTTA gggTggacac aggacgtgt ggTttctgag 60
- [0087] ccagggctag cgggcgactc agatcccagc cagtggactt agcccctgtt tgctcctccg 120
- [0088] ataactgggg tgacctgggt taatattcac cagcagcctc ccccgTTGCC cctctggatc 180
- [0089] cactgcttaa atacggacga ggacaggGCC ctgtctcctc agcttcaggc accaccactg 240
- [0090] acctgggaca gtgaatcgcc ac 262
- [0091] <210> 8
- [0092] <211> 203
- [0093] <212> DNA
- [0094] <213> 人工序列
- [0095] <220>
- [0096] <223> 人工序列的描述:合成
- [0097] 多核苷酸
- [0098] <400> 8
- [0099] tgatgcggtt ttggcagtac atcaatgggc gtggatagcg gtttgactca cggggatttc 60
- [0100] caagtctcca cccattgac gtcaatggga gtttgTTTTg gcacaaaaat caacgggact 120
- [0101] ttccaaaatg tcgtaacaac tccgccccat tgacgcaaat gggcggtagg cgtgtacggt 180
- [0102] gggaggtcta tataagcaga gct 203
- [0103] <210> 9
- [0104] <211> 251
- [0105] <212> DNA
- [0106] <213> 人工序列
- [0107] <220>
- [0108] <223> 人工序列的描述:合成
- [0109] 多核苷酸
- [0110] <400> 9
- [0111] gtttgctgct tgcaatgttt gcccatTTTA gggTggacac aggacgtgt ggTttctgag 60
- [0112] ccagggggcg actcagatcc cagccagtgg acttagcccc tgTttgctcc tccgataact 120
- [0113] ggggtgacct tggTTaatat tcaccagcag cctcccccgT tgcctctctg gatccactgc 180
- [0114] ttaaatacgg acgaggacag ggccctgtct cctcagcttc aggcaccacc actgacctgg 240
- [0115] gacagtgaat c 251
- [0116] <210> 10
- [0117] <211> 304
- [0118] <212> DNA
- [0119] <213> 人工序列
- [0120] <220>
- [0121] <223> 人工序列的描述:合成
- [0122] 多核苷酸
- [0123] <400> 10
- [0124] cgTTacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60
- [0125] gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

[0126]	atgggtggag tatttacggt aaactgcca cttggcagta catcaagtgt atcatatgcc	180
[0127]	aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	240
[0128]	catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac	300
[0129]	catg	304
[0130]	<210>	11
[0131]	<211>	1643
[0132]	<212>	DNA
[0133]	<213>	人工序列
[0134]	<220>	
[0135]	<223>	人工序列的描述:合成
[0136]	多核苷酸	
[0137]	<400>	11
[0138]	gttacataac ttacggtaaa tggcccgctt ggctgaccgc ccaacgaccc ccgcccattg	60
[0139]	acgtcaataa tgacgtatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa	120
[0140]	tgggtggagt atttacggta aactgcccac ttggcagtac atcaagtgt tcatatgcc	180
[0141]	agtacgccc ctattgacgt caatgacggt aaatggccc cctggcatta tgcccagta	240
[0142]	atgaccttat gggactttcc tacttggcag tacatctacg tattagtcac cgctattacc	300
[0143]	atgggtcgag gtgagcccca cgttctgctt cactctcccc atctcccccc cctccccacc	360
[0144]	cccaattttg tatttattta ttttttaatt attttgtgca gcgatggggg cggggggggg	420
[0145]	gggggcgcgc gccagcgagg gcggggcggg gcgaggggcg gggcggggcg aggcggagag	480
[0146]	gtgcggcggc agccaatcag agcggcggcg tccgaaagt tccctttatg gcgaggcggc	540
[0147]	ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc gggagtcgct gcgttgccct	600
[0148]	cgccccgtgc cccgctccgc gccgcctcgc gccgcccgc ccggctctga ctgaccgcgt	660
[0149]	tactcccaca ggtgagcggg cgggacggcc cttctcctcc gggctgtaat tagcgcttgg	720
[0150]	tttaatgacg gctcgtttct tttctgtggc tgcgtgaaag ccttaaaggg ctccgggagg	780
[0151]	gccctttgtg cgggggggag cggctcgggg ggtgcgtgcg tgtgtgtgtg cgtggggagc	840
[0152]	gccgcgtgcg gccgcgctg cccggcggtt gtgagcgtg cgggcgcggc gcggggcttt	900
[0153]	gtgcgtccg cgtgtgcgcg aggggagcgc ggccgggggc ggtgccccgc ggtgcggggg	960
[0154]	ggctgcgagg ggaacaaagg ctgcgtgcgg ggtgtgtgcg tgggggggtg agcagggggt	1020
[0155]	gtgggcgcgg cggtcgggct gtaaccccc cctgcacccc cctccccgag ttgctgagca	1080
[0156]	cggcccggtt tcgggtgcgg ggctccgtgc ggggcgtggc gcggggctcg ccgtgccggg	1140
[0157]	cgggggggtg cggcaggtg gggtgccggg cggggcgggg ccgcctcggg ccggggaggg	1200
[0158]	ctcgggggag gggcgcggcg gccccggagc gccggcggtt gtcgaggcgc ggcgagccgc	1260
[0159]	agccattgcc ttttatggta atcgtgcgag agggcgagg gacttccttt gtcccaaata	1320
[0160]	tggcggagcc gaaatctggg aggcgccgcc gcaccccctc tagcgggcgc gggcgaagcg	1380
[0161]	gtgcggcgcc ggcaggaagg aaatgggcgg ggagggcctt cgtgcgtcgc cgcgccgccg	1440
[0162]	tcccctctc catctccagc ctcggggctg ccgcaggggg acggctgcct tcggggggga	1500
[0163]	cggggcaggg cggggttcgg cttctggcgt gtgaccggcg gctctagagc ctctgctaac	1560
[0164]	catgttcatt cttctctct tttctacag ctcttgggca acgtgctggt tgttgtgctg	1620
[0165]	tctcatcatt ttggcaaaga att	1643
[0166]	<210>	12
[0167]	<211>	231

- [0168] <212> DNA
[0169] <213> 人工序列
[0170] <220>
[0171] <223> 人工序列的描述:合成
[0172] 多核苷酸
[0173] <400> 12
[0174] gctccggtgc ccgtcagtgg gcagagcgca catcggccac agtccccgag aagttggggg 60
[0175] gaggggtcgg caattgaacc ggtgcctaga gaaggtggcg cggggtaaac tgggaaagtg 120
[0176] atgtcgtgta ctggctccgc ctttttcccg aggggtggggg agaaccgtat ataagtgcag 180
[0177] tagtcgccgt gaacgttctt tttcgcaacg ggtttgccgc cagaacacag g 231
[0178] <210> 13
[0179] <211> 117
[0180] <212> DNA
[0181] <213> 人工序列
[0182] <220>
[0183] <223> 人工序列的描述:合成
[0184] 多核苷酸
[0185] <400> 13
[0186] gtgatgacgt gtcccataag gccctcgggt ctaaggcttc cctatttcct ggttcgccgg 60
[0187] cggccatttt ggggtggaagc gatagctgag tggcgccggc tgctgattgt gttctag 117
[0188] <210> 14
[0189] <211> 117
[0190] <212> DNA
[0191] <213> 人工序列
[0192] <220>
[0193] <223> 人工序列的描述:合成
[0194] 多核苷酸
[0195] <400> 14
[0196] gtgatgacgt gtcccatact tccgggtcag gtgggccggc tgtcttgacc ttctttgcgg 60
[0197] ctcgccatt ttgtcccagt cagtccggag gctgcggctg cagaagtacc gcctgcg 117
[0198] <210> 15
[0199] <211> 117
[0200] <212> DNA
[0201] <213> 人工序列
[0202] <220>
[0203] <223> 人工序列的描述:合成
[0204] 多核苷酸
[0205] <400> 15
[0206] gtgatgacgt gtcccatatt ttcatctcgc gagacttggt agcggccatc ttggtcctgc 60
[0207] cctgacagat tctcctatcg gggtcacagg gacgctaaga ttgctacctg gactttc 117
[0208] <210> 16
[0209] <211> 117

[0210]	<212> DNA
[0211]	<213> 人工序列
[0212]	<220>
[0213]	<223> 人工序列的描述:合成
[0214]	多核苷酸
[0215]	<400> 16
[0216]	gtgatgacgt gtcccatggc ctcatggat gagaggtccc acctcacggc ccgaggcggg 60
[0217]	gcttctttgc gcttaaaagc cgagccgggc caatgttcaa atgcgcagct cttagtc 117
[0218]	<210> 17
[0219]	<211> 117
[0220]	<212> DNA
[0221]	<213> 人工序列
[0222]	<220>
[0223]	<223> 人工序列的描述:合成
[0224]	多核苷酸
[0225]	<400> 17
[0226]	gtgatgacgt gtcccatccc cctccacccc cctagcccgc ggagcacgct gggatttggc 60
[0227]	gccccctcc tcggtgcaac ctatataagg ctacagtct gcgctcctgg tacacgc 117
[0228]	<210> 18
[0229]	<211> 21
[0230]	<212> DNA
[0231]	<213> 人工序列
[0232]	<220>
[0233]	<223> 人工序列的描述:合成
[0234]	引物
[0235]	<400> 18
[0236]	cattccagga ctgtccattc t 21
[0237]	<210> 19
[0238]	<211> 21
[0239]	<212> DNA
[0240]	<213> 人工序列
[0241]	<220>
[0242]	<223> 人工序列的描述:合成
[0243]	引物
[0244]	<400> 19
[0245]	ggcctgaacg tagaagtacc a 21
[0246]	<210> 20
[0247]	<211> 20
[0248]	<212> DNA
[0249]	<213> 人工序列
[0250]	<220>
[0251]	<223> 人工序列的描述:合成

- [0252] 引物
[0253] <400> 20
[0254] accacagtcc atgccatcac 20
[0255] <210> 21
[0256] <211> 20
[0257] <212> DNA
[0258] <213> 人工序列
[0259] <220>
[0260] <223> 人工序列的描述:合成
[0261] 引物
[0262] <400> 21
[0263] tccaccaccc tgttgctgta 20
[0264] <210> 22
[0265] <211> 100
[0266] <212> DNA
[0267] <213> 人工序列
[0268] <220>
[0269] <223> 人工序列的描述:合成
[0270] 多核苷酸
[0271] <400> 22
[0272] cccccctcca cccctagcc cgcggagcac gctgggattt ggccccccc tctcgggtgc 60
[0273] aacctatata aggctcacag tctgcgctcc tggtacacgc 100
[0274] <210> 23
[0275] <211> 100
[0276] <212> DNA
[0277] <213> 人工序列
[0278] <220>
[0279] <223> 人工序列的描述:合成
[0280] 多核苷酸
[0281] <400> 23
[0282] ggcctcattg gatgagaggt cccacctcac ggcccagagc ggggcttctt tgcgcttaaa 60
[0283] agccgagccg ggccaatggt caaatgcgca gctcttagtc 100
[0284] <210> 24
[0285] <211> 100
[0286] <212> DNA
[0287] <213> 人工序列
[0288] <220>
[0289] <223> 人工序列的描述:合成
[0290] 多核苷酸
[0291] <400> 24
[0292] ggggtggggcc cgcgctata aagggggcgc aggcgggctg ggcgttcac aggccaagtg 60
[0293] cgctgtgctc gaggggtgcc ggccaggcct gagcgagcga 100

[0294] <210> 25
[0295] <211> 100
[0296] <212> DNA
[0297] <213> 人工序列
[0298] <220>
[0299] <223> 人工序列的描述:合成
[0300] 多核苷酸
[0301] <400> 25
[0302] ggtgcatat tcgattggc tggagtcggc catcacgtc cagctacgcc acttcctttt 60
[0303] cgtggcacta taaagggtgc tgcacggcgc ttgcatctct 100
[0304] <210> 26
[0305] <211> 100
[0306] <212> DNA
[0307] <213> 人工序列
[0308] <220>
[0309] <223> 人工序列的描述:合成
[0310] 多核苷酸
[0311] <400> 26
[0312] acttccgggt caggtgggcc ggctgtcttg accttctttg cggctcggcc attttgtccc 60
[0313] agtcagtccg gaggtgcgg ctgcagaagt accgcctgcg 100
[0314] <210> 27
[0315] <211> 100
[0316] <212> DNA
[0317] <213> 人工序列
[0318] <220>
[0319] <223> 人工序列的描述:合成
[0320] 多核苷酸
[0321] <400> 27
[0322] gctgagcgcg cgcgatggg cgggaggttt ggggtcaagg agcaaactct gcacaagatg 60
[0323] gcggcggtag cggcagtggc ggcgcgtagg aggcggtgag 100
[0324] <210> 28
[0325] <211> 100
[0326] <212> DNA
[0327] <213> 人工序列
[0328] <220>
[0329] <223> 人工序列的描述:合成
[0330] 多核苷酸
[0331] <400> 28
[0332] attttcatct cgcgagactt gtgagcggcc atcttgggcc tgccctgaca gatttcctta 60
[0333] tcggggtcac agggacgcta agattgctac ctggactttc 100
[0334] <210> 29
[0335] <211> 100

- [0336] <212> DNA
[0337] <213> 人工序列
[0338] <220>
[0339] <223> 人工序列的描述:合成
[0340] 多核苷酸
[0341] <400> 29
[0342] tgggaccccc ggaaggcgga agttctaggg cggaagtggc cgagaggaga ggagaatggc 60
[0343] ggcggaaggc tggatttggc gttggggctg gggccggcgg 100
[0344] <210> 30
[0345] <211> 100
[0346] <212> DNA
[0347] <213> 人工序列
[0348] <220>
[0349] <223> 人工序列的描述:合成
[0350] 多核苷酸
[0351] <400> 30
[0352] aaggccctc ggtctaaggc ttccctatct cctggttcgc cggcggccat tttgggtgga 60
[0353] agcgatagct gaggcggc ggctgctgat tgtgttctag 100
[0354] <210> 31
[0355] <211> 100
[0356] <212> DNA
[0357] <213> 人工序列
[0358] <220>
[0359] <223> 人工序列的描述:合成
[0360] 多核苷酸
[0361] <400> 31
[0362] agtgaccgga aagtagaagt ggcccttgca ggcaagagt ctggaggcg gcagcggcga 60
[0363] ccggagcggg aggcagcga atttatccgt gtgcagcccc 100
[0364] <210> 32
[0365] <211> 100
[0366] <212> DNA
[0367] <213> 人工序列
[0368] <220>
[0369] <223> 人工序列的描述:合成
[0370] 多核苷酸
[0371] <400> 32
[0372] gggagggcg cgctggggag cttcggcga tgcgcgtga ggcctgcctg accgacctc 60
[0373] agcagggtg tggctaccat gttctctcgc gcgggtgtcg 100
[0374] <210> 33
[0375] <211> 100
[0376] <212> DNA
[0377] <213> 人工序列

[0378] <220>
[0379] <223> 人工序列的描述:合成
[0380] 多核苷酸
[0381] <400> 33
[0382] actgcgcacg cgcgcggtcg caccgattca cgcccccttc cggcgcctag agcaccgctg 60
[0383] ccgccatgtt gaggggggga ccgcgaccag ctggggccct 100
[0384] <210> 34
[0385] <211> 100
[0386] <212> DNA
[0387] <213> 人工序列
[0388] <220>
[0389] <223> 人工序列的描述:合成
[0390] 多核苷酸
[0391] <400> 34
[0392] ccctcgaggg gcggagcaaa aagtgaggca gcaacgcctc cttatcctcg ctcccgttt 60
[0393] cagttctcaa taaggtccga tgttcgtgta taaatgctcg 100
[0394] <210> 35
[0395] <211> 100
[0396] <212> DNA
[0397] <213> 人工序列
[0398] <220>
[0399] <223> 人工序列的描述:合成
[0400] 多核苷酸
[0401] <400> 35
[0402] cttggtgacc aaatttgaaa aaaaaaaaa accgcgcca ctcattgtt tttcaatcag 60
[0403] gtccgccaag tttgtattta aggaactgtt tcagttcata 100
[0404] <210> 36
[0405] <211> 100
[0406] <212> DNA
[0407] <213> 人工序列
[0408] <220>
[0409] <223> 人工序列的描述:合成
[0410] 多核苷酸
[0411] <400> 36
[0412] ggctgagcta tcctattggc tatcgggaca aaatttgctt gagccaatca aagtgtccg 60
[0413] tggacaatcg ccgttctgtc tataaaaagg tgaagcagcg 100
[0414] <210> 37
[0415] <211> 100
[0416] <212> DNA
[0417] <213> 人工序列
[0418] <220>
[0419] <223> 人工序列的描述:合成

- [0420] 多核苷酸
[0421] <400> 37
[0422] ggaagtgcc a gaccggaggt gcgtcattca cggcgacgc cgatacggtt cctccaccga 60
[0423] ggcccatgcg aagctttcca ctatggcttc cagcactgtc 100
[0424] <210> 38
[0425] <211> 100
[0426] <212> DNA
[0427] <213> 人工序列
[0428] <220>
[0429] <223> 人工序列的描述:合成
[0430] 多核苷酸
[0431] <400> 38
[0432] ccctcgaggg gcggagcaaa aagtgaggca gcaacgcctc cttatcctcg ctcccgttt 60
[0433] cagttctcaa taaggtccga tgttcgtgta taaatgctcg 100
[0434] <210> 39
[0435] <211> 100
[0436] <212> DNA
[0437] <213> 人工序列
[0438] <220>
[0439] <223> 人工序列的描述:合成
[0440] 多核苷酸
[0441] <400> 39
[0442] cttggtgacc aaatttgaaa aaaaaaaaa accgcgcca ctcattgtgt tttcaatcag 60
[0443] gtccgccaag tttgtattta aggaactgtt tcagttcata 100
[0444] <210> 40
[0445] <211> 100
[0446] <212> DNA
[0447] <213> 人工序列
[0448] <220>
[0449] <223> 人工序列的描述:合成
[0450] 多核苷酸
[0451] <400> 40
[0452] ggctgagcta tcctattgga tatcgggaca aaatttgctt gagccaatca aagtgcctcg 60
[0453] tggacaatcg ccgttctgtc tataaaaagg tgaagcagcg 100
[0454] <210> 41
[0455] <211> 100
[0456] <212> DNA
[0457] <213> 人工序列
[0458] <220>
[0459] <223> 人工序列的描述:合成
[0460] 多核苷酸
[0461] <400> 41

[0462] ggaagtgccg gaccggaggt gcgtcattca ccggcgacgc cgatacggtt cctccaccga 60
[0463] ggcccatgcg aagctttcca ctatggcttc cagcactgtc 100
[0464] <210> 42
[0465] <211> 20
[0466] <212> DNA
[0467] <213> 人工序列
[0468] <220>
[0469] <223> 人工序列的描述:合成
[0470] 引物
[0471] <400> 42
[0472] gctaccccgga ccacatgaag 20
[0473] <210> 43
[0474] <211> 20
[0475] <212> DNA
[0476] <213> 人工序列
[0477] <220>
[0478] <223> 人工序列的描述:合成
[0479] 引物
[0480] <400> 43
[0481] tcttgtagtt gccgtcgctc 20

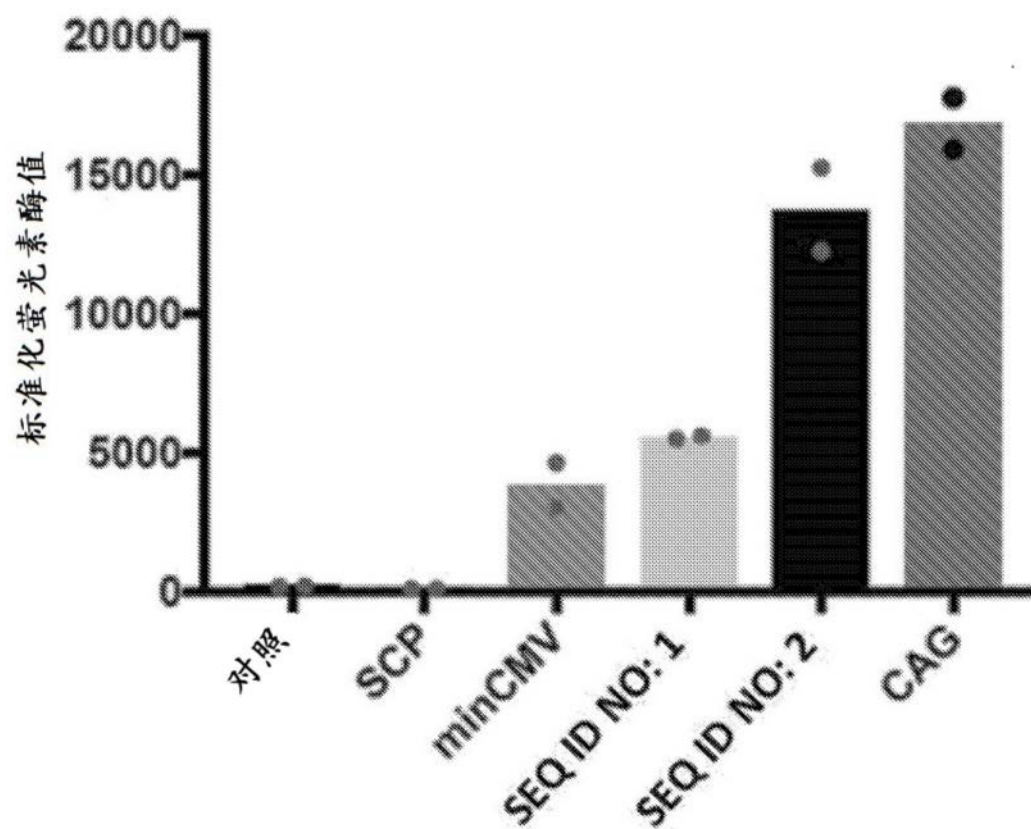


图1A

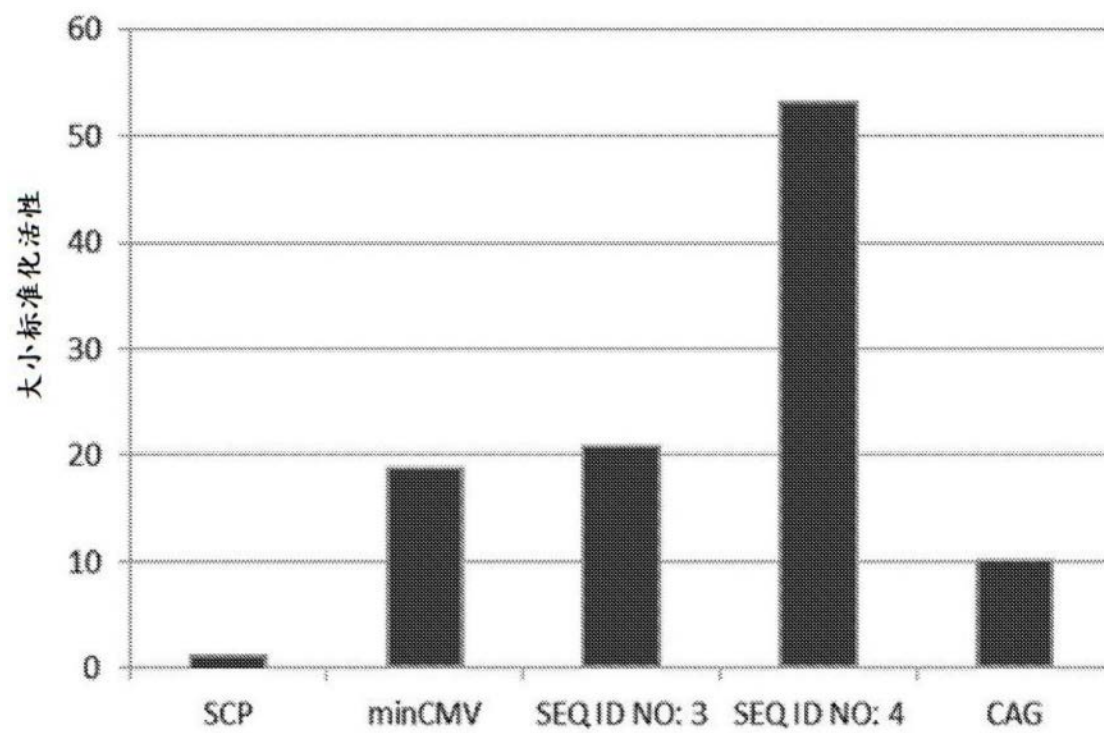


图1B

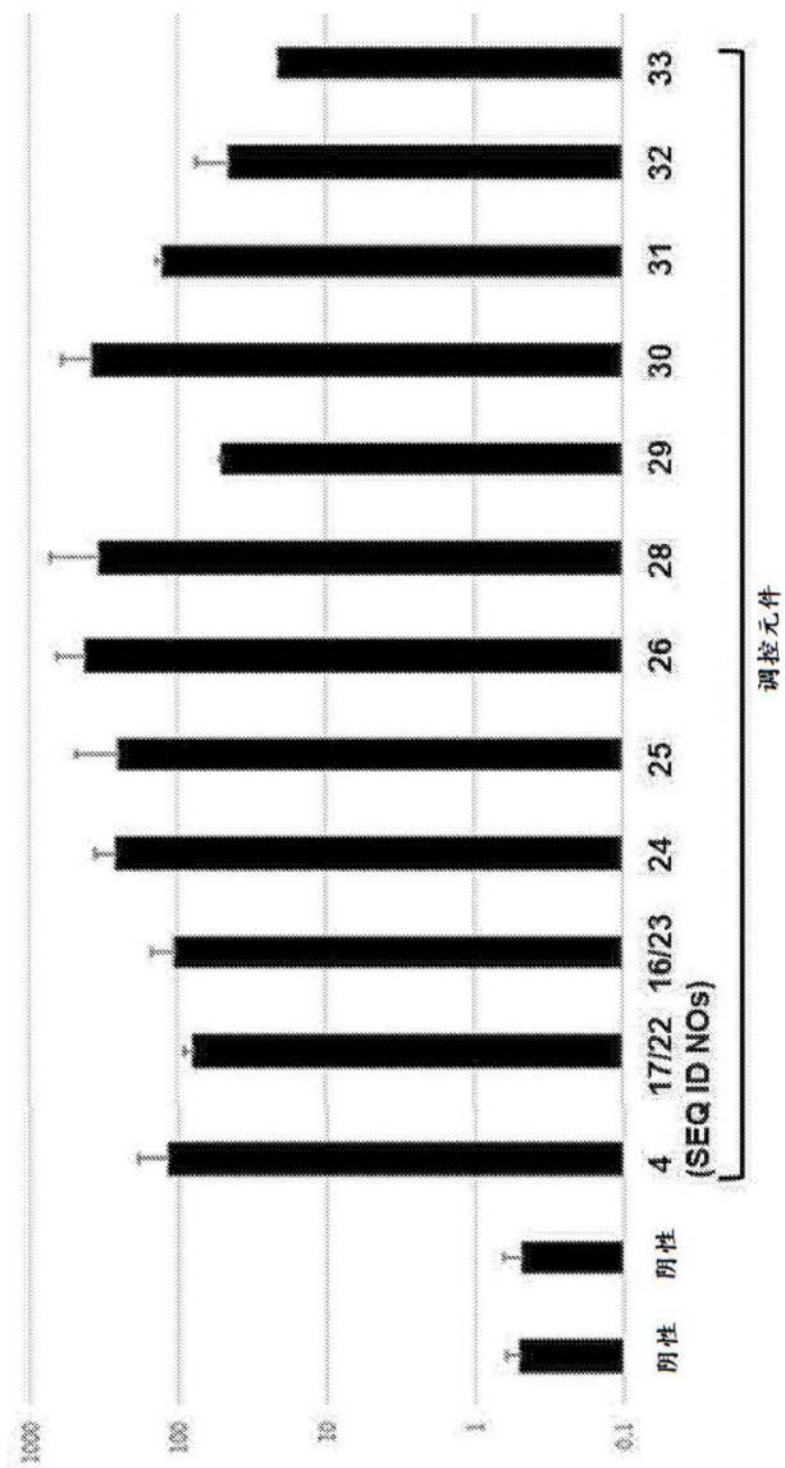


图1C

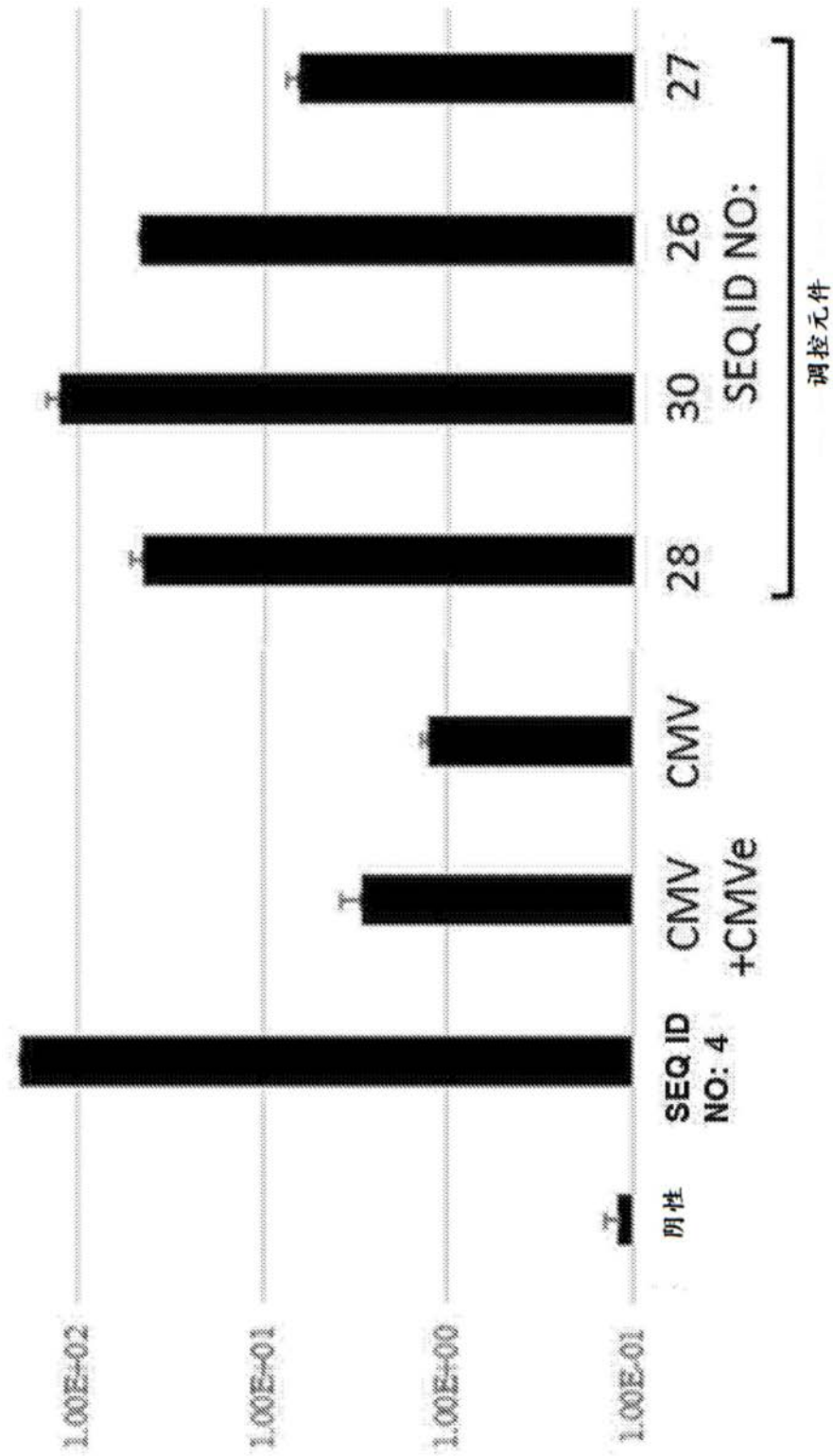


图1D

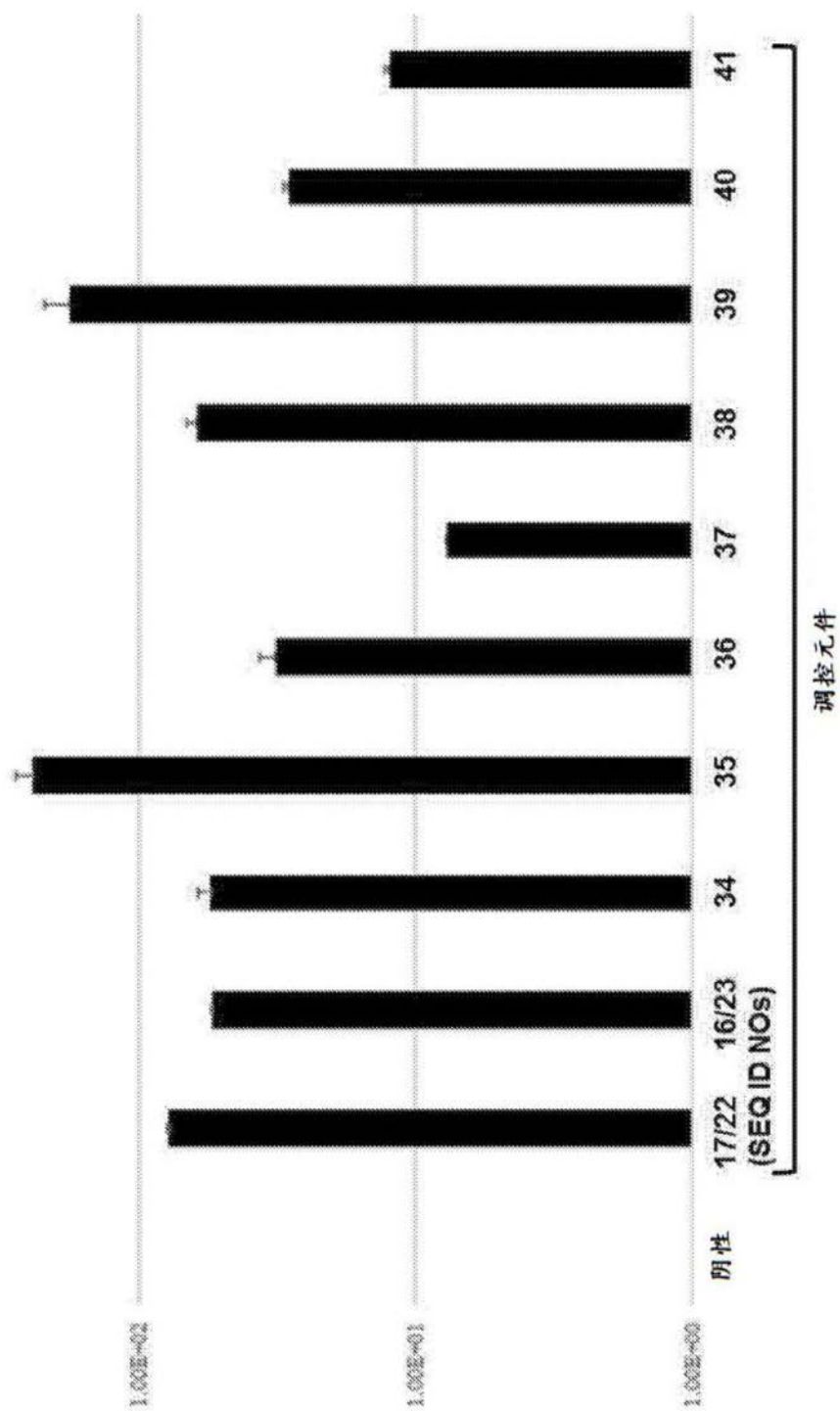


图1E

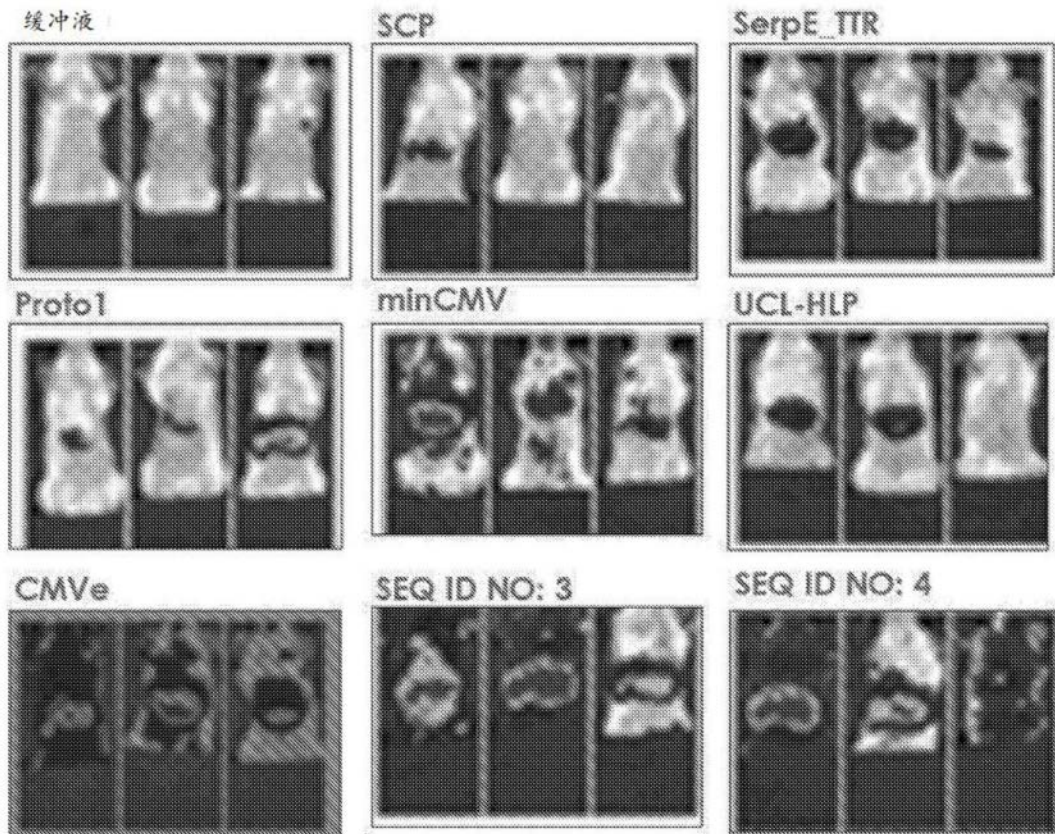


图2A

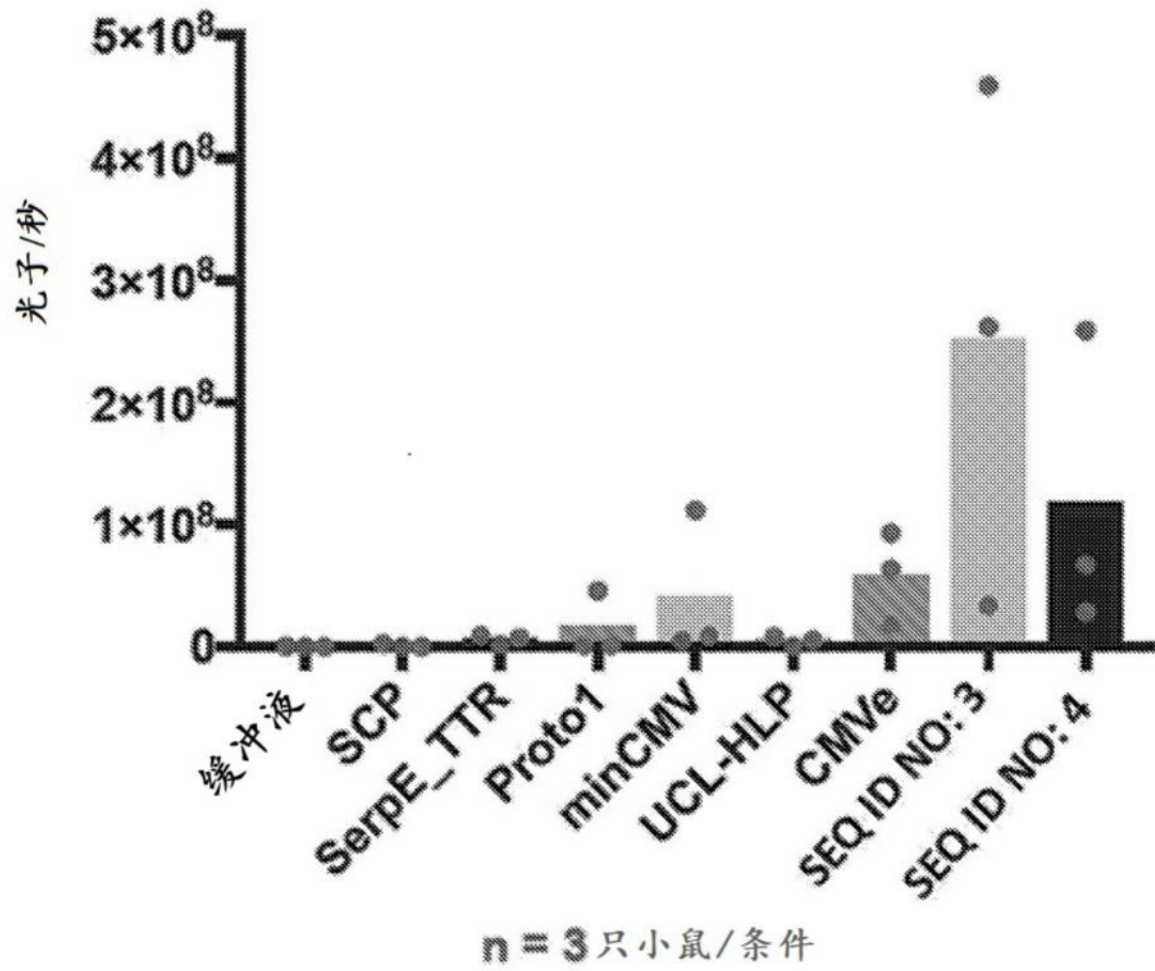


图2B

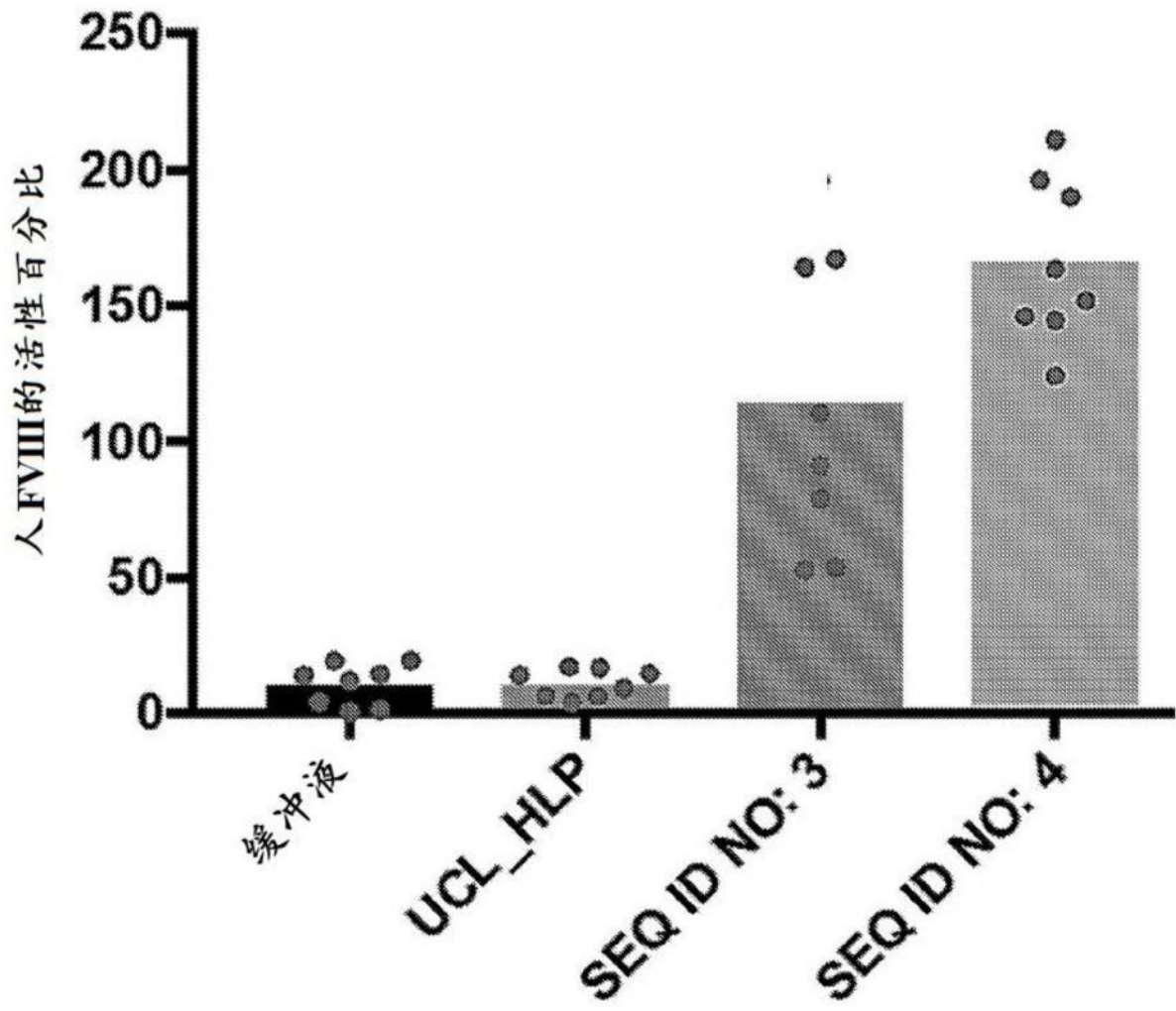


图3B

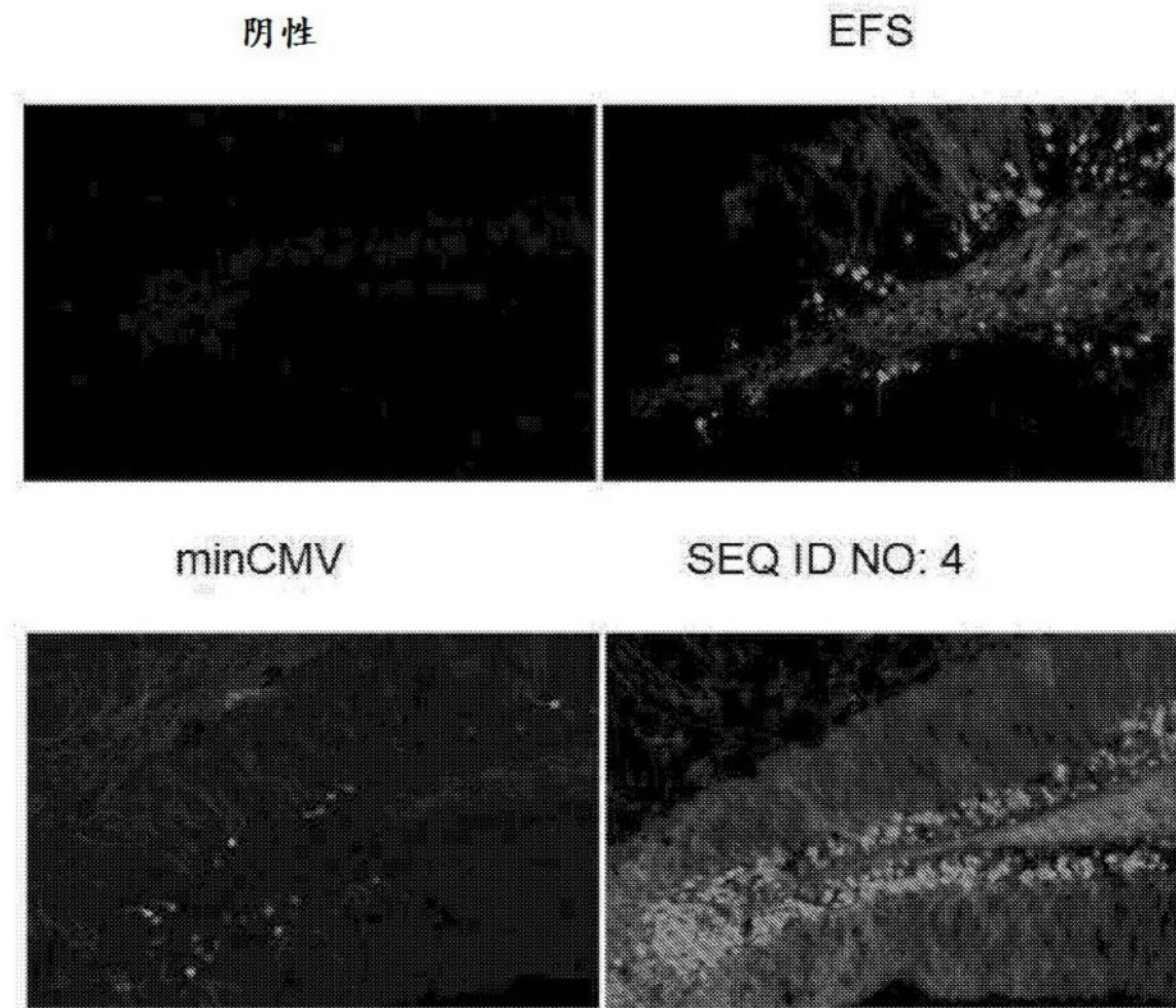


图4A

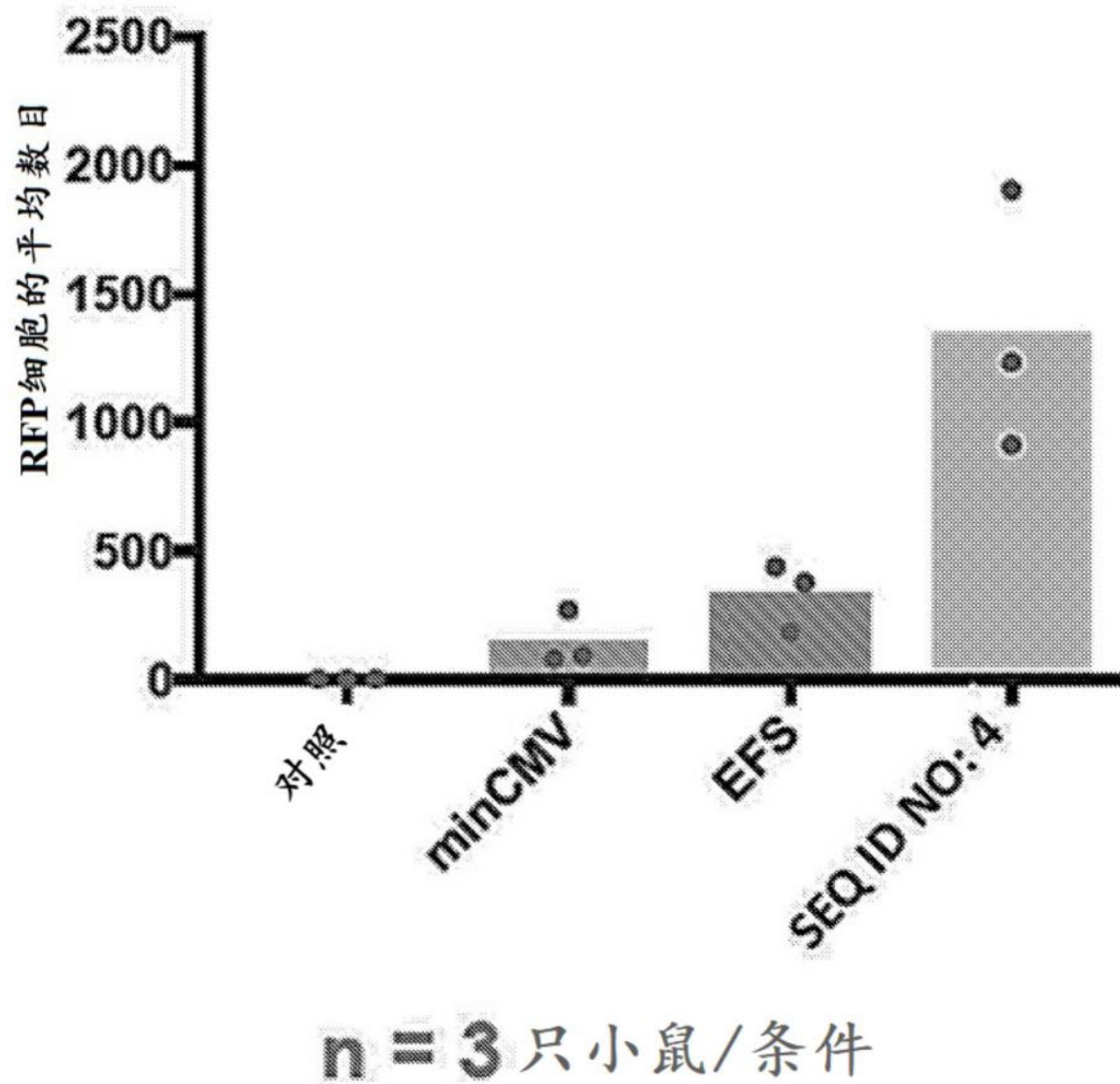


图4B

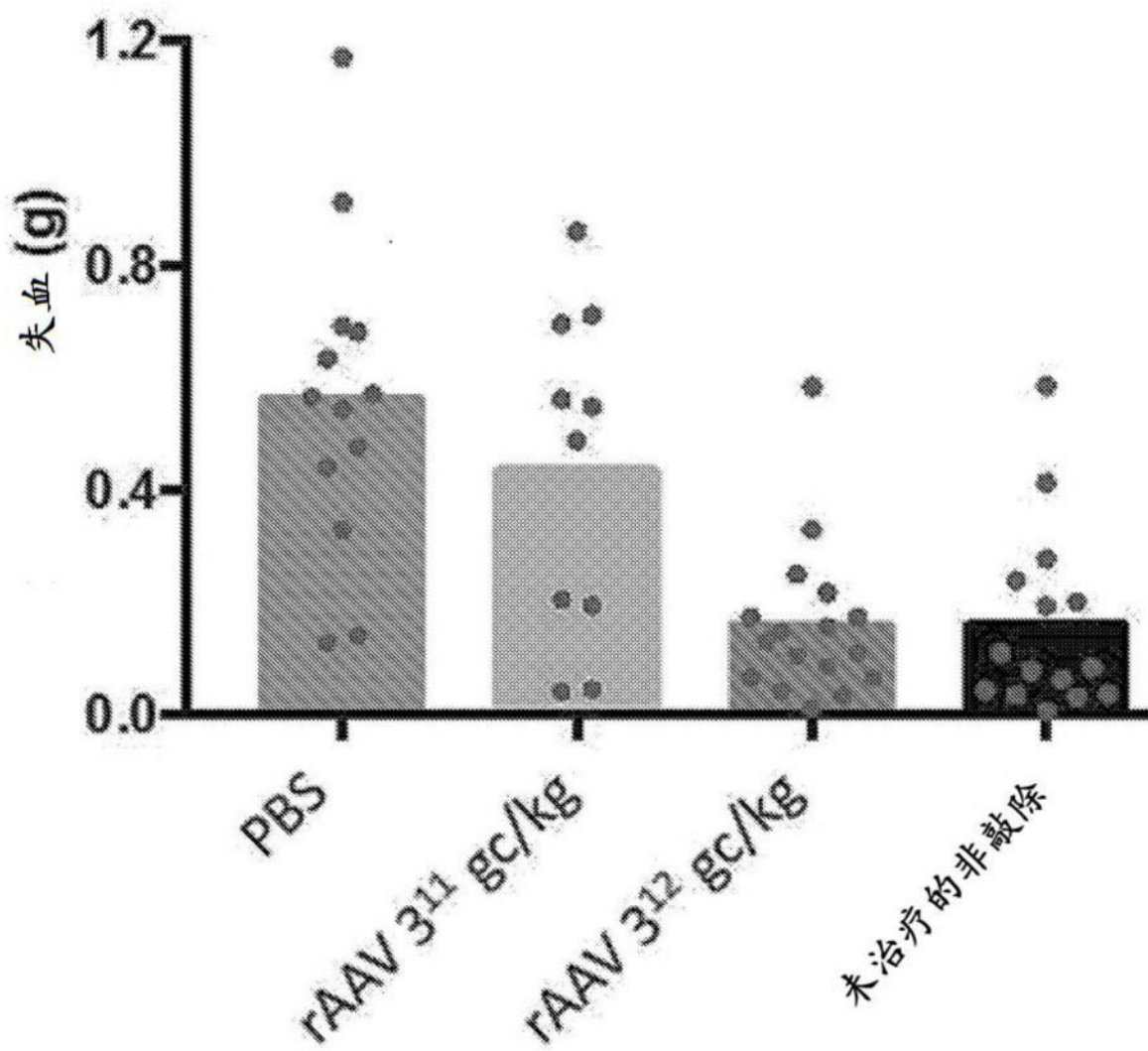


图5A

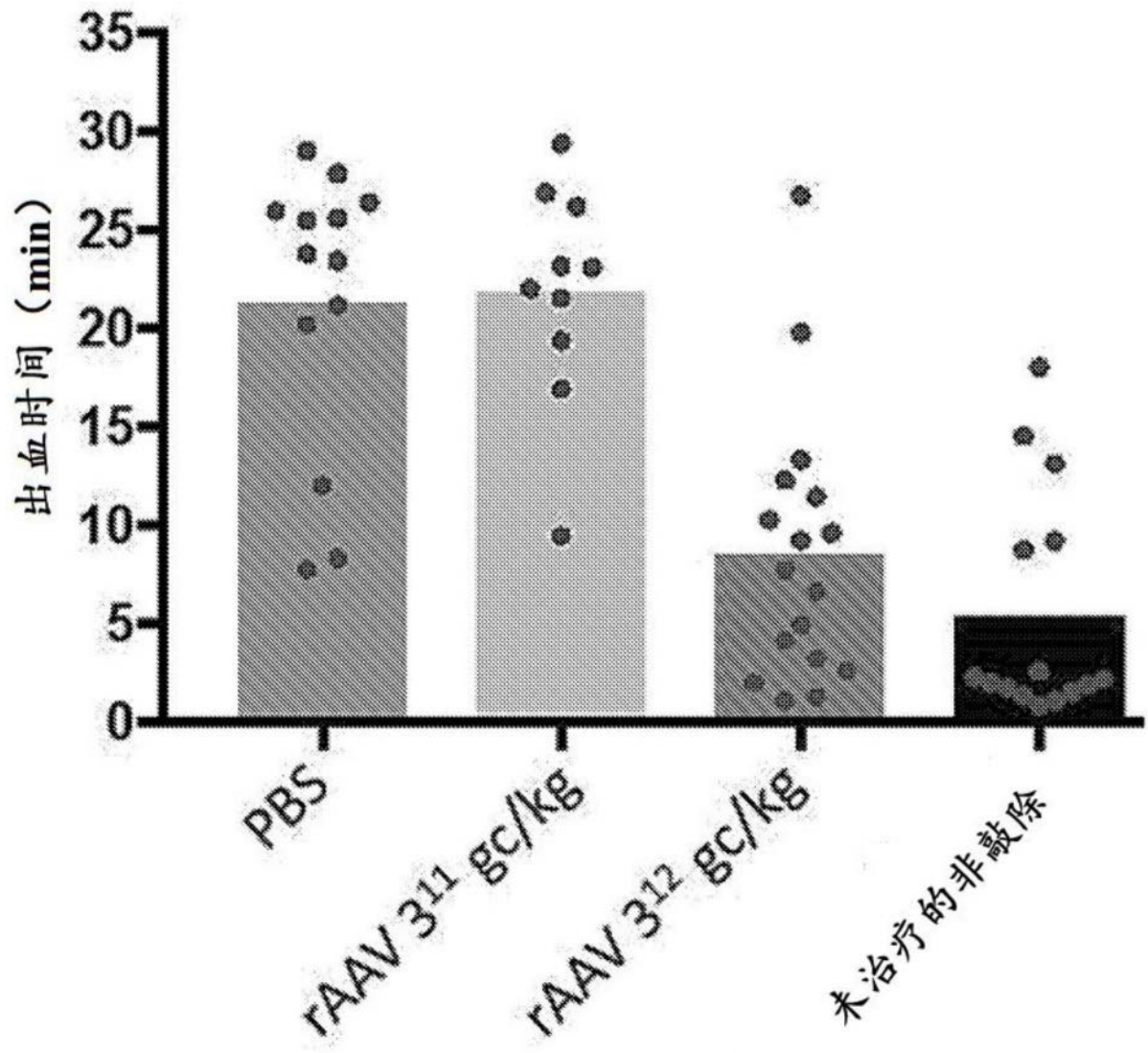


图5B

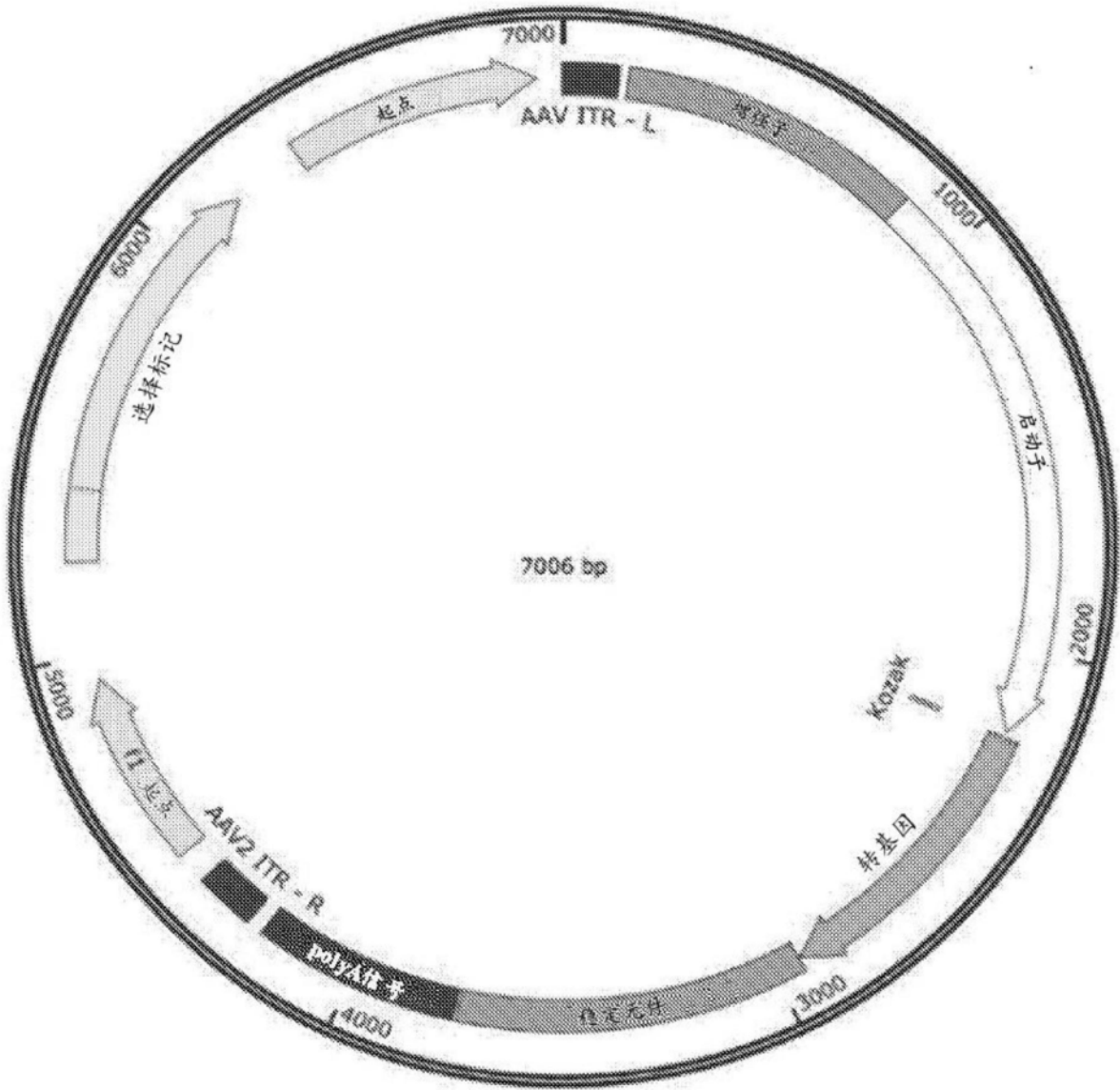


图6A

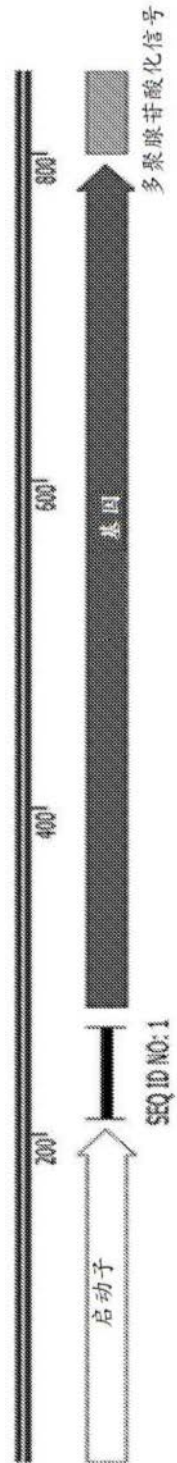


图6B

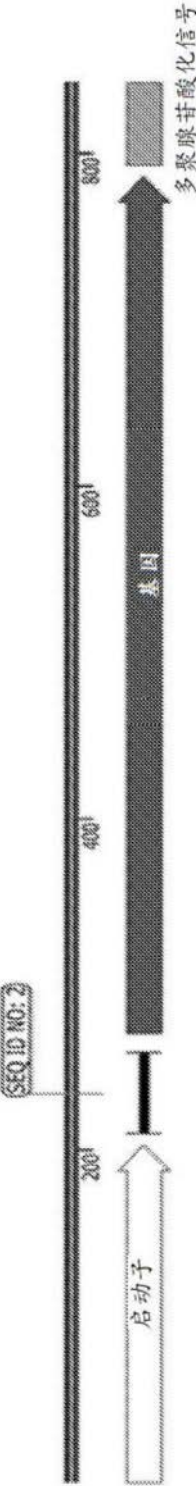


图6C



图7

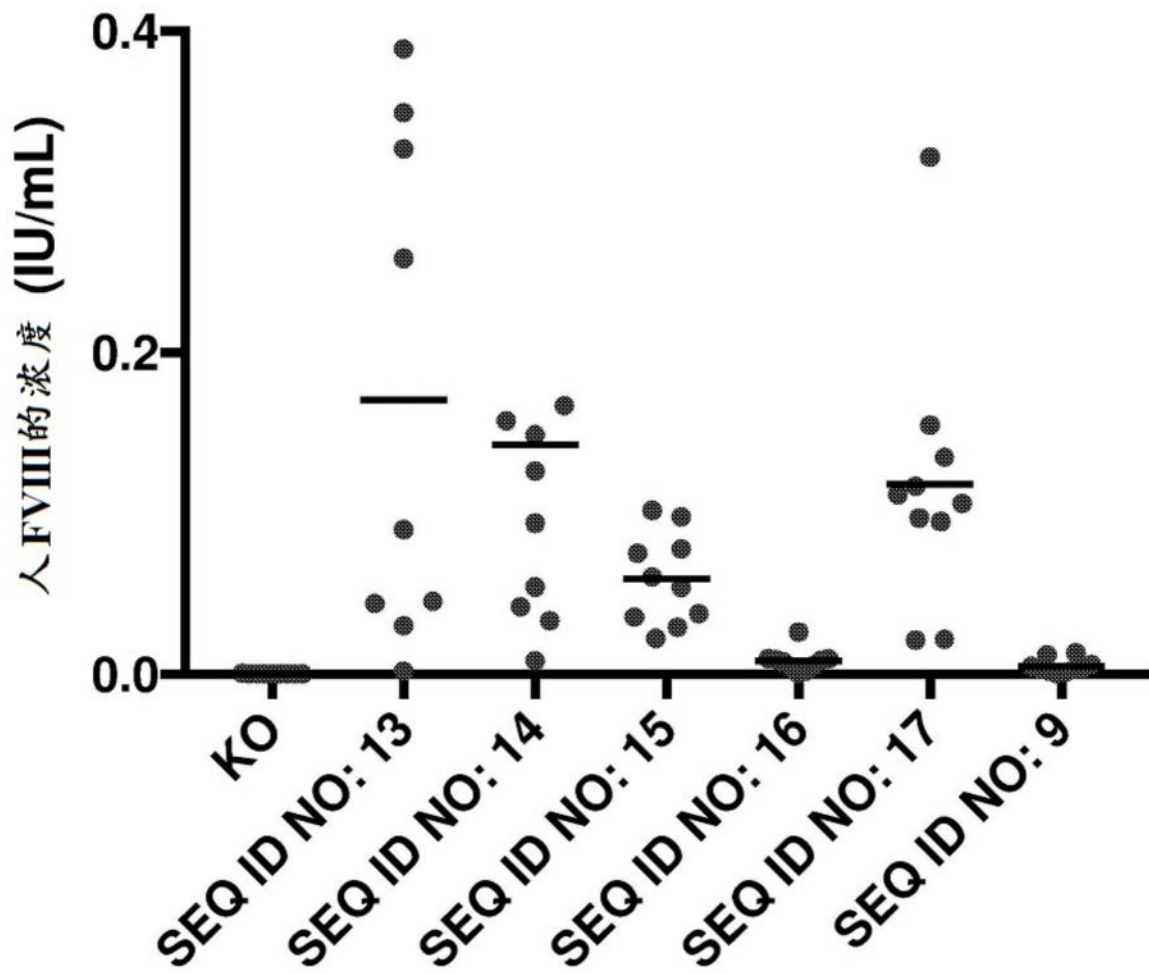


图8

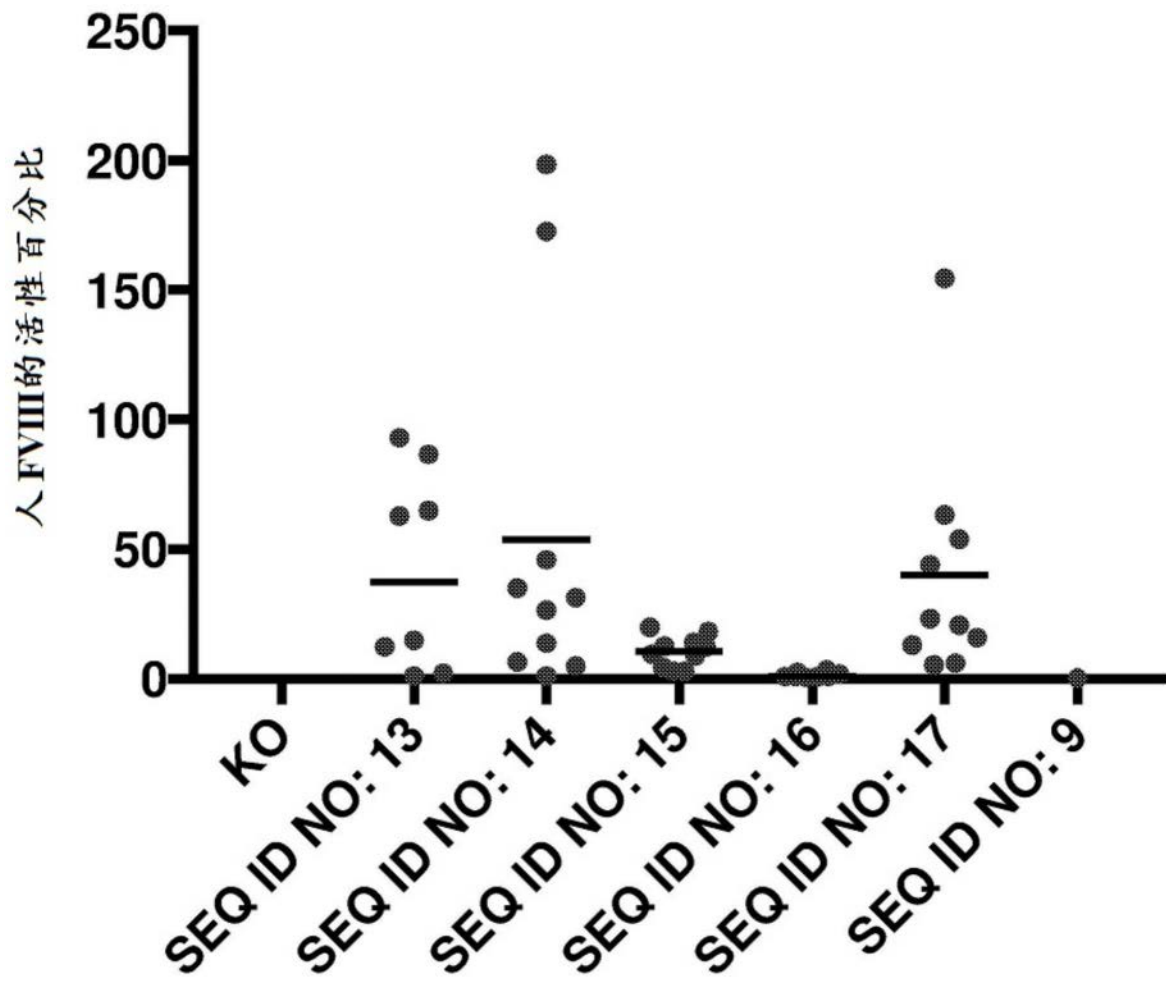


图9

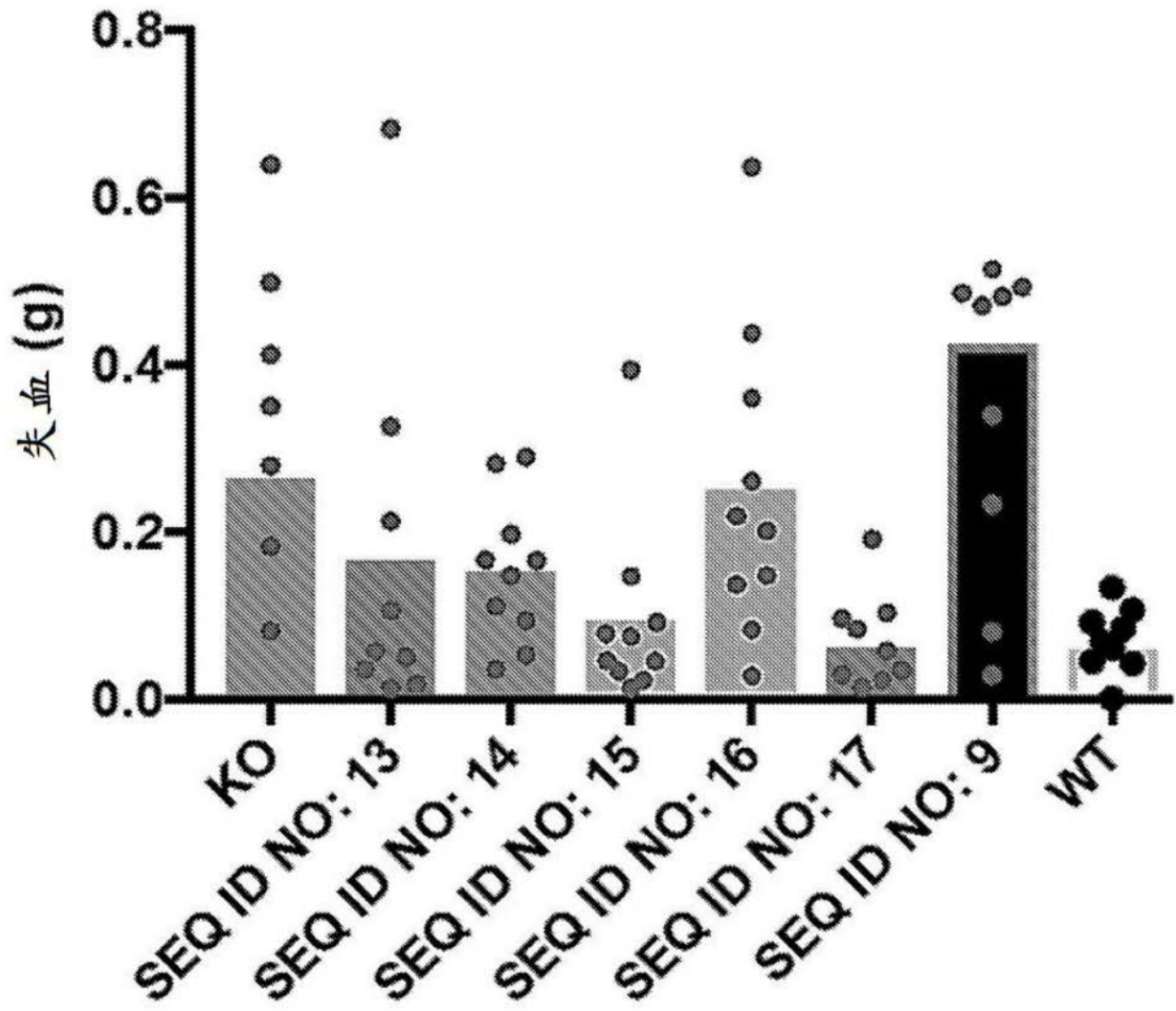


图10

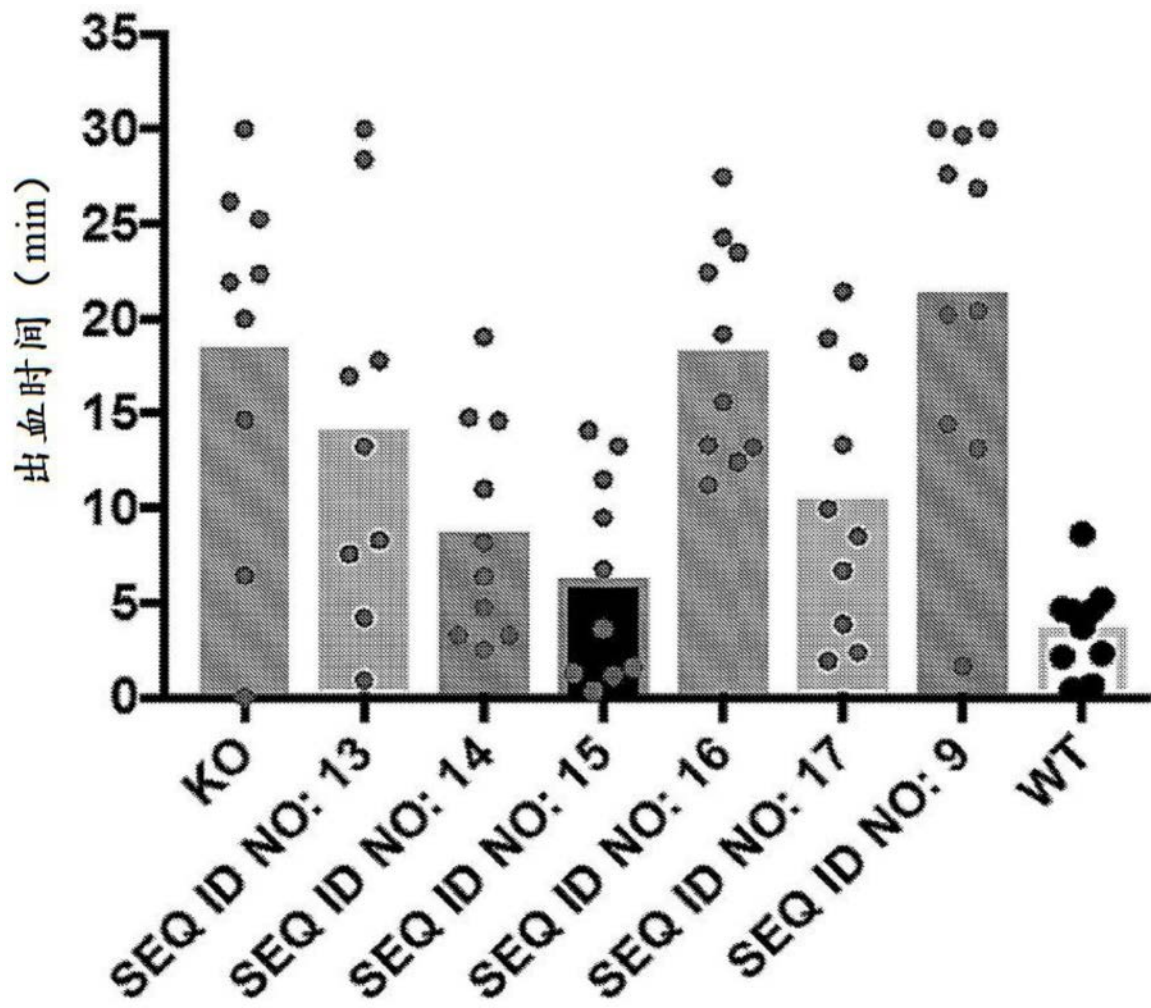


图11

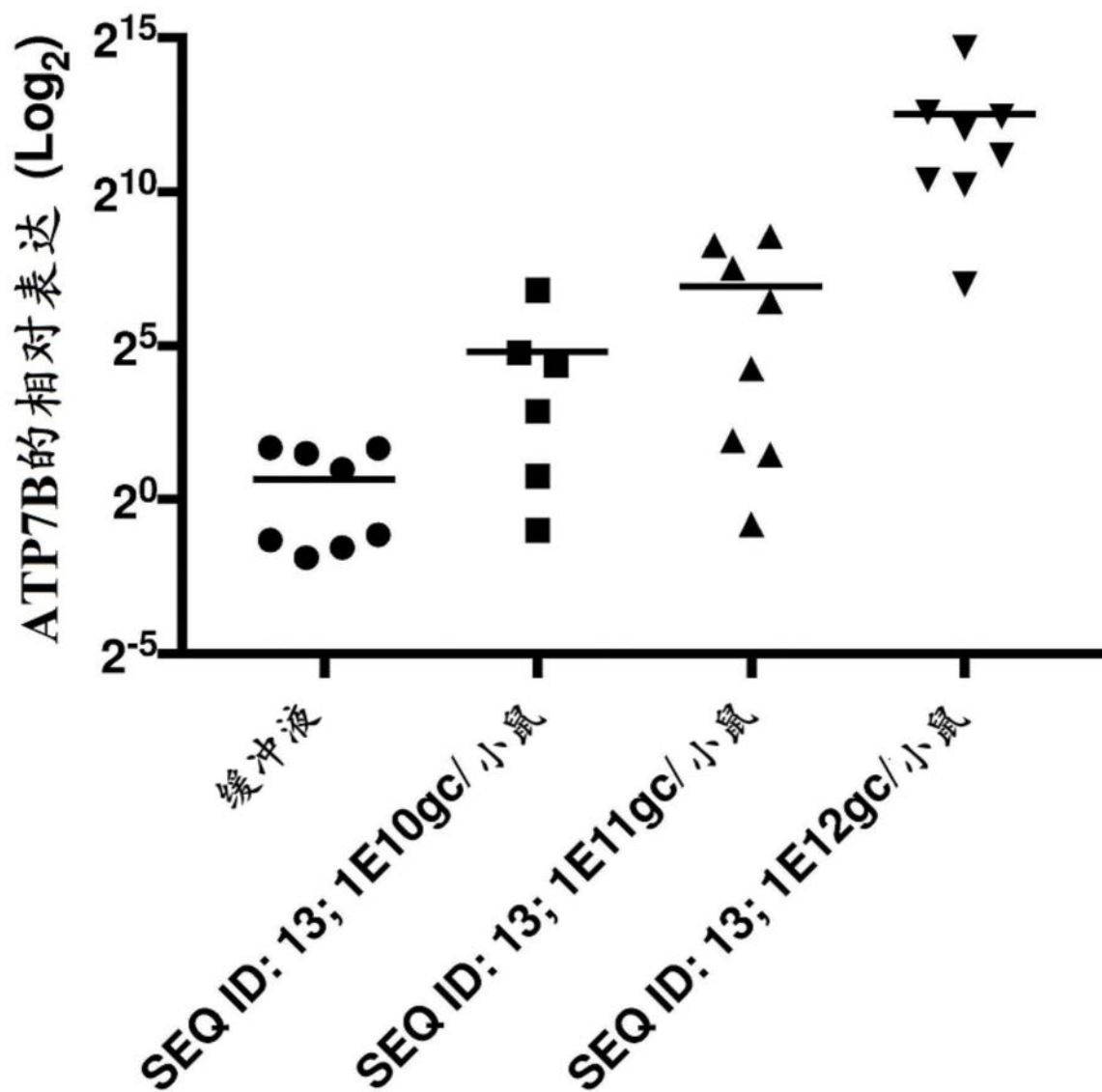


图12

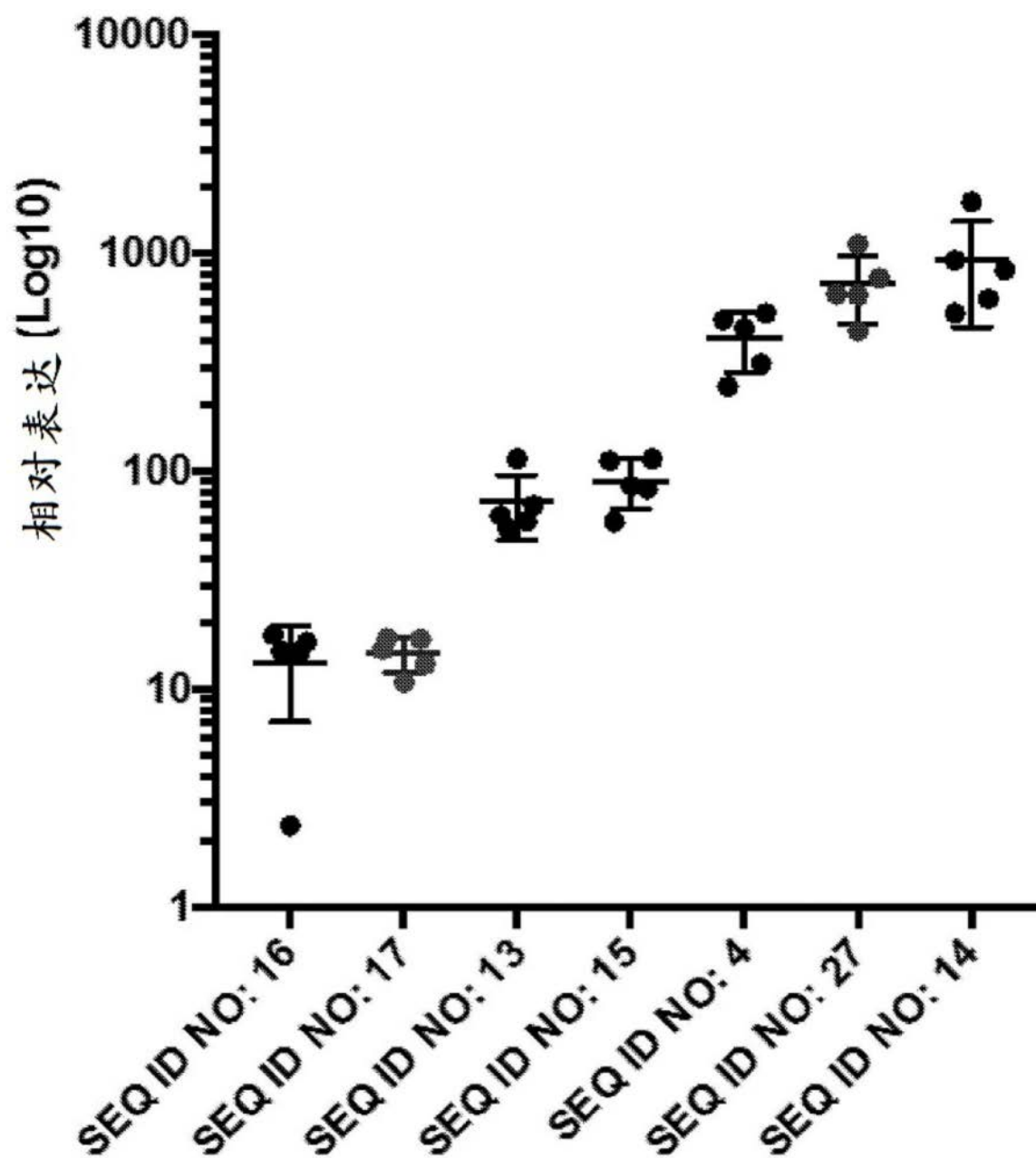


图13