202100748

(19)中華民國智慧財產局

(12)發明說明書公開本　(11)公開編號：TW 202100748 A

(43)公開日：中華民國 110 (2021) 年 01 月 01 日

(21)申請案號：109112554　(22)申請日：中華民國 109 (2020) 年 04 月 14 日

(51)Int. Cl.：*C12N15/113 (2010.01)*　*C12N15/79 (2006.01)*

C12N15/63 (2006.01)　*A61P21/00 (2006.01)*

(30)優先權：2019/04/14　美國　62/833,759

(71)申請人：美國公爵大學 (美國) DUKE UNIVERSITY　(US)

美國

(72)發明人：格斯巴赫　查爾斯　A GERSBACH, CHARLES A. (US)；皮卡　奧利弗　亞德里安 PICKAR OLIVER, ADRIAN (US)

(74)代理人：陳長文

申請實體審查：無　申請專利範圍項數：35 項　圖式數：14　共 129 頁

(54)名稱

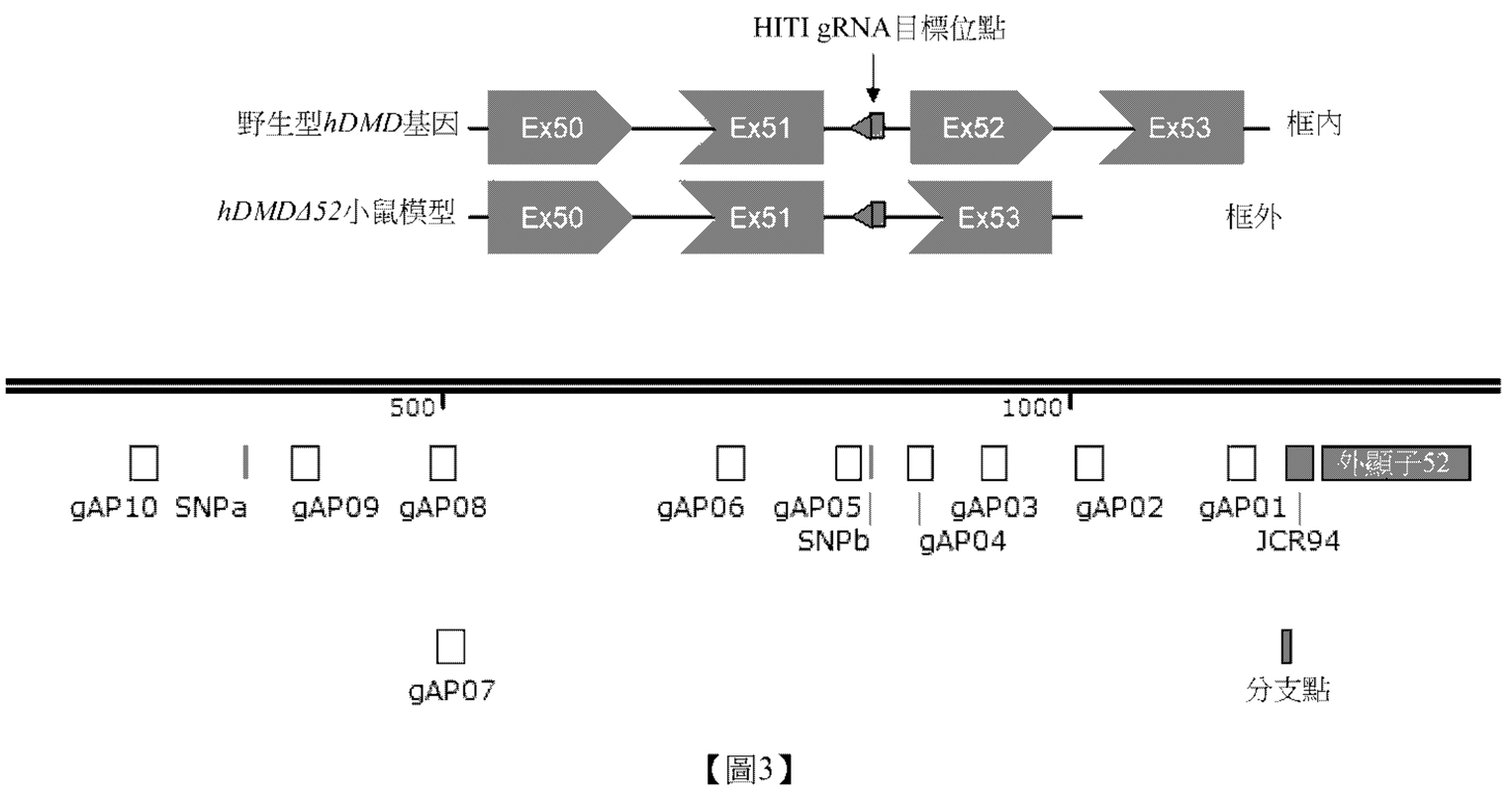
恢復肌縮蛋白功能之基於ＣＲＩＳＰＲ／ＣＡＳ之基因組編輯組合物

(57)摘要

本文揭示基於 CRISPR/Cas 之基因組編輯組合物及藉由恢復肌縮蛋白功能來治療杜興氏肌肉萎縮症(Duchenne Muscular Dystrophy)之方法。

Disclosed herein are CRISPR/Cas-based genome editing compositions and methods for treating Duchenne Muscular Dystrophy by restoring dystrophin function.

指定代表圖：



【圖3】

202100748

【發明摘要】

【中文發明名稱】

恢復肌縮蛋白功能之基於CRISPR/CAS之基因組編輯組合物

【英文發明名稱】

CRISPR/CAS-BASED GENOME EDITING COMPOSITION FOR RESTORING DYSTROPHIN FUNCTION

【中文】

本文揭示基於CRISPR/Cas之基因組編輯組合物及藉由恢復肌縮蛋白功能來治療杜興氏肌肉萎縮症(Duchenne Muscular Dystrophy)之方法。

【英文】

Disclosed herein are CRISPR/Cas-based genome editing compositions and methods for treating Duchenne Muscular Dystrophy by restoring dystrophin function.

【指定代表圖】

圖3

【代表圖之符號簡單說明】

無

【發明說明書】

【中文發明名稱】

恢復肌縮蛋白功能之基於CRISPR/CAS之基因組編輯組合物

【英文發明名稱】

CRISPR/CAS-BASED GENOME EDITING COMPOSITION FOR RESTORING DYSTROPHIN FUNCTION

【技術領域】

【0001】 本發明係關於基於CRISPR/Cas之基因組編輯組合物及藉由恢復肌縮蛋白功能來治療杜興氏肌肉萎縮症(Duchenne Muscular Dystrophy)之方法。

【先前技術】

【0002】 杜興氏肌肉萎縮症(DMD)為最普遍的致死性遺傳性兒童疾病，其在新生兒中發生率為約1:5000。導致20多歲患者死亡的進行性肌肉無力為肌縮蛋白基因突變之結果。在大多數情況下(約60%)，突變由來自肌縮蛋白基因之79個外顯子中之一或多者的缺失(導致閱讀框架破壞)組成。先前的治療性策略通常旨在產生截短但部分功能性肌縮蛋白質之表現，該表現再現了對應於貝氏肌肉萎縮症(其伴有相比於DMD而言更溫和的症狀)之基因型。舉例而言，若干組已採用CRISPR/Cas9技術對經培養人類DMD細胞及DMD之mdx小鼠模型進行基因編輯，以恢復因缺失特定外顯子之肌縮蛋白閱讀框架。然而，仍需要研發用以恢復完整的完全功能性肌縮蛋白質的基因編輯策略。

【發明內容】

【0003】 在一態樣中，本發明係關於一種基於CRISPR/Cas之基因

組編輯系統。該系統可包括一或多種編碼組合物之載體，該組合物包含：(a) 引導RNA (gRNA)，其靶向突變肌縮蛋白基因之片段； (b) Cas蛋白質或包含該Cas蛋白質之融合蛋白；及(c) 供體序列，其包含野生型肌縮蛋白基因之片段。在另一態樣中，該系統可包括(a) 引導RNA (gRNA)，其靶向突變肌縮蛋白基因之片段；(b) Cas蛋白質或包含該Cas蛋白質之融合蛋白；及(c) 供體序列，其包含野生型肌縮蛋白基因之片段。在一些實施例中，該野生型肌縮蛋白基因之該片段由兩個gRNA間隔序列及/或PAM序列側接。在一些實施例中，該gRNA靶向與該突變肌縮蛋白基因之外顯子並列的內含子，且其中該外顯子選自該突變肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59及61-66。在一些實施例中，該供體序列包含該野生型肌縮蛋白基因之外顯子或其功能性等效物，且其中該外顯子選自該野生型肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59、61-66。在一些實施例中，該突變肌縮蛋白基因之該外顯子經突變或自該肌縮蛋白基因至少部分地缺失，或其中該突變肌縮蛋白基因之該外顯子缺失且該內含子與相應野生型肌縮蛋白基因中之缺失的外顯子的位置並列。在一些實施例中，該外顯子為外顯子52。在一些實施例中，該gRNA結合及靶向包含以下之聚核苷酸序列： a) SEQ ID NO: 17或SEQ ID NO: 18；b) SEQ ID NO: 17或SEQ ID NO: 18之片段； c) SEQ ID NO: 17或SEQ ID NO: 18之互補序列或其片段； d) 與SEQ ID NO: 17或SEQ ID NO: 18或其互補序列實質上一致的核酸；或e) 在嚴格條件下雜交至SEQ ID NO: 17或SEQ ID NO: 18、其互補序列或與其實質上一致之序列的核酸。在一些實施例中，該gRNA包含聚核苷酸序列SEQ ID NO: 19或SEQ ID NO: 20或其變體或由其編碼。在一些實施例中， 該Cas蛋白質

為化膿性鏈球菌(*Streptococcus pyogenes*)Cas9蛋白質或金黃色葡萄球菌(*Staphylococcus aureus*)Cas9蛋白質。在一些實施例中，該Cas蛋白質包含胺基酸序列SEQ ID NO: 1、2、3或4。在一些實施例中，該兩個gRNA間隔序列獨立地包含選自SEQ ID NO: 5-8及25-45之序列。在一些實施例中，該兩個gRNA間隔序列為相同的。在一些實施例中，該兩個gRNA間隔序列為不同的。在一些實施例中，該兩個gRNA間隔序列中之至少一者包含序列SEQ ID NO: 25或SEQ ID NO: 26。在一些實施例中，該供體序列包含SEQ ID NO: 21或SEQ ID NO: 22之聚核苷酸。在一些實施例中，該載體為病毒載體。在一些實施例中，該載體為腺相關病毒(AAV)載體。在一些實施例中，該AAV載體為AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV-10、AAV-11、AAV-12、AAV-13或AAVrh.74載體。在一些實施例中，該一或多個載體中之一者包含聚核苷酸序列SEQ ID NO: 23或24。在一些實施例中，gRNA與供體序列之間的莫耳比為1:1、或1:15、或5:1至1:10、或1:1至1:5。

【0004】 在另一態樣中，本發明係關於重組聚核苷酸，其編碼包含野生型肌縮蛋白基因之片段或其功能性等效物的供體序列，且其中該片段或其功能性等效物由兩個gRNA間隔序列側接。在一些實施例中，該供體序列包含該肌縮蛋白基因之該外顯子，且其中該外顯子選自外顯子1-8、10、11、12、14、16-22、43-59、61-66。在一些實施例中，該重組聚核苷酸包含序列SEQ ID NO: 23或24。

【0005】 本發明之另一態樣提供一種載體，其包含如本文所詳述之重組聚核苷酸。在一些實施例中，該載體包含驅動該重組聚核苷酸表現之異源啟動子。

【0006】 本發明之另一態樣提供一種細胞，其包含如本文所詳述之重組聚核苷酸或如本文所詳述之載體。

【0007】 本發明之另一態樣提供一種用於恢復具有突變肌縮蛋白基因之細胞的肌縮蛋白功能的組合物，該組合物包含如本文所詳述之系統、如本文所詳述之重組聚核苷酸或如本文所詳述之載體。

【0008】 本發明之另一態樣提供一種套組，其包含如本文所詳述之系統、如本文所詳述之重組聚核苷酸或如本文所詳述之載體或如本文所詳述之組合物。

【0009】 本發明之另一態樣提供一種用於恢復具有突變肌縮蛋白基因之細胞或個體的肌縮蛋白功能的方法。該方法可包括使細胞或個體與如本文所詳述之系統、如本文所詳述之重組聚核苷酸或如本文所詳述之載體或如本文所詳述之組合物接觸。在一些實施例中，該肌縮蛋白功能藉由插入該野生型肌縮蛋白基因之外顯子52來恢復。在一些實施例中，該個體患有杜興氏肌肉萎縮症。

【0010】 本發明之另一態樣提供一種用於恢復具有由一或多個缺失或突變型外顯子引起的破壞的肌縮蛋白基因之細胞或個體的肌縮蛋白功能的方法。該方法可包括使細胞或個體與如本文所詳述之系統、如本文所詳述之重組聚核苷酸或如本文所詳述之載體或如本文所詳述之組合物接觸。在一些實施例中，肌縮蛋白功能藉由插入對應於該一或多個缺失或突變型外顯子之肌縮蛋白基因的一或多個野生型外顯子來恢復。在一些實施例中，該等缺失或突變型外顯子中之一者為外顯子52。

【圖式簡單說明】

【0011】 **圖1**為細胞中編碼肌縮蛋白質之外顯子及各種相互作用的

示意圖。

【0012】 **圖2**為肌縮蛋白質之示意圖。

【0013】 **圖3**為用於產生靶向hDMD-外顯子52上游之hDMD-內含子51之gRNA的策略之圖。

【0014】 **圖4A**展示**圖4B**所示之測量員分析的引物數目及預期條帶大小。

【0015】 **圖4B**為展示靶向hDMD-外顯子52上游之hDMD-內含子51之gRNA的編輯效率的凝膠。

【0016】 **圖5A**展示基於引物位置之HEK293T SNP結果，凝膠展示於**圖5B**中。

【0017】 **圖6A**為展示用編碼gRNA之質體電穿孔的肌母細胞的凝膠，該等質體經19-23 bp間隔序列重新設計，進一步結果展示於**圖6B**中。

【0018】 **圖7**為展示gRNA表現質體或AAV-HITI供體質體之HEK293T轉染結果的凝膠。

【0019】 **圖8A**為展示用以偵測AAV質體(AAV-CMV-Cas9質體及AAV-U6-gRNA-Ex52質體)之HITI介導之整合的巢式PCR結果之凝膠，使該等AAV質體電穿孔至hDMDΔ52/mdx小鼠之原代肌母細胞中。**圖8B**為具有預期HITI介導之插入的桑格測序(Sanger sequencing)結果。

【0020】 **圖9**為用於證實活體內編輯、確定最佳gRNA/供體序列組合及確定AAV-Cas9與AAV供體質體之最佳比率的實驗之示意圖。

【0021】 **圖10A**為展示具有靶向切割位點下游之引物的小鼠之基因組DNA中靶向Ex52插入的凝膠。**圖10B**為展示靶向切割位點上游之引物的小鼠之基因組DNA中靶向Ex52插入的凝膠。

【0022】 **圖11**為展示經處理hDMDΔ52/mdx小鼠之mRNA中靶向Ex52插入之凝膠。

【0023】 **圖12**為用以證實經處理小鼠之蛋白質恢復的西方墨點分析。

【0024】 **圖13**為經編輯小鼠的AAV-ITR基因組整合之伊路米那(Illumina)深度測序定量的結果。

【0025】 **圖14A**為展示自外顯子45至外顯子69之cDNA擴增的凝膠。**圖14B**來自mRNA之PacBio測序分析，展示了外顯子51與外顯子53之間的118 bp的測序讀段與外顯子52序列匹配。

【實施方式】

【0026】

相關申請案之交叉引用

本申請案主張2019年4月14日申請之美國臨時專利申請第62/833,759號之優先權；該案以全文引用的方式併入本文中。

【0027】

關於聯邦資助研究之聲明

本發明係在政府支持下在由美國國家衛生研究院(National Institutes of Health)授予的授權R01AR069085下進行。政府在本發明中具有某些權利。

【0028】 本發明提供基於CRISPR/Cas之基因/基因組編輯組合物及藉由恢復肌縮蛋白功能來治療杜興氏肌肉萎縮症之方法。DMD通常由破壞閱讀框架之肌縮蛋白基因的缺失引起。許多治療DMD之策略旨在藉由移除或跳過額外外顯子來恢復閱讀框架，因為已表明內部截短肌縮蛋白質

可仍具有部分功能性。本文詳述了用於校正肌縮蛋白基因之基於AAV之同源非依賴性靶向整合序列(HITI)介導的基因編輯療法。特定言之，吾人採用CRISPR/Cas9基因編輯技術，以引導缺失的外顯子靶向插入至肌縮蛋白基因中。作為治療相關目標，HITI介導之基因組編輯策略在DMD之人類化小鼠模型中經最佳化，在該模型中，外顯子52已在攜帶全長人類肌縮蛋白基因之小鼠(*hDMDΔ52/mdx*小鼠)中移除。為了實現靶向整合，使含有包括外顯子52之缺失型基因組序列的AAV載體與編碼AAV之Cas9/gRNA表現盒一起共遞送。證實了經培養細胞中靶向外顯子52整合。結合AAV遞送，用於靶向插入缺失的外顯子之HITI介導之策略提供一種用以恢復全長肌縮蛋白之方法及改進的功能性結果。

1. 定義

【0029】如本文所用，術語「包含(comprise(s))」、「包括(include(s))」、「具有(having)」、「具有(has)」、「可」、「含有」及其變體意欲為並不排除額外動作或結構之可能性的開放式過渡片語、術語或詞語。除非在上下文中以其他方式明確指定，否則單數形式「一(a/an)」及「該」包括複數個參考物。無論是否明確闡述，本發明亦涵蓋其他實施例「包含」、「由以下組成」及「基本上由以下組成」：本文所呈現之實施例或要素。

【0030】對於本文數值範圍之敘述，明確涵蓋具有相同精確程度之在其之間的各間插數值。舉例而言，對於6-9之範圍，除6與9以外涵蓋數值7與8，且對於6.0-7.0之範圍，明確涵蓋數值6.0、6.1、6.2、6.3、6.4、6.5、6.6、6.7、6.8、6.9及7.0。

【0031】如本文所用，術語「約」或「近似地」意味在如一般熟習

此項技術者所確定之特定值的可接受誤差範圍內，其將部分取決於如何量測或確定該值，亦即，量測系統之限制。舉例而言，「約」可意謂依據此項技術中之實務，在3個或3個以上標準差之範圍內。或者，「約」可意謂在給定值之至多20%，較佳至多10%，更佳至多5%且再更佳至多1%之範圍內。或者，特別是就生物系統或方法而言，該術語可意謂在一定值之一個數量級範圍內，較佳在5倍範圍內且更佳在2倍範圍內。

【0032】 除非另外定義，否則本文所用之所有技術及科學術語均具有與一般熟習此項技術者通常理解相同之含義。如有衝突，以本文件包括的定義為主。儘管可使用類似於或等效於本文所述之方法及材料的方法及材料來實踐或測試本發明，但下文描述較佳方法及材料。本文提及之所有公開案、專利申請案、專利及其他參考案均以全文引用的方式併入本文中。本文所揭示之材料、方法及實例僅為說明性的且並不意欲具限制性。

【0033】 如本文中可互換使用，「腺相關病毒」或「AAV」係指一種屬於細小病毒科之依賴病毒屬的細小病毒，其感染人類及一些其他靈長類動物物種。目前尚不清楚AAV是否會引起疾病，且因此該病毒是否會引起極輕度免疫反應。

【0034】 如本文所用，「結合區域」係指由基於CRISPR/Cas之基因組編輯系統識別及結合的目標區域內之區域。

【0035】 如本文所用，「染色體」係指與組蛋白相關之染色體DNA的組織複合物。

【0036】 如本文中可互換使用，「成簇規律間隔短回文重複序列」及「CRISPR」係指含有多個較短的直接重複序列之基因座，其發現於約40%的定序細菌及90%的定序古菌之基因組中。

【0037】 如本文所用，「編碼序列」或「編碼核酸」意味包含編碼蛋白質之核苷酸序列的核酸(RNA或DNA分子)。編碼序列可進一步包括可操作地連接至調節元件之起始及終止信號，該等調節元件包括能夠引導在投與核酸之個體或哺乳動物的細胞中表現的啟動子及聚腺苷酸化信號。編碼序列可經密碼子最佳化。

【0038】 如本文所用，「互補序列」或「互補」意味核酸可意謂核酸分子之核苷酸或核苷酸類似物之間的華特生-克里克(Watson-Crick)(例如A-T/U及C-G)或胡斯坦(Hoogsteen)鹼基配對。「互補性」係指兩個核酸序列之間共有的特性，使得當其反向平行於彼此對準時，各位置處之核苷酸鹼基將互補。

【0039】 如本文中可互換使用，「杜興氏肌肉萎縮症」或「DMD」係指引起肌肉退化及最終死亡之隱性的致命性X性聯病症。DMD為常見的遺傳性單基因疾病且人群發病率為3500/1。DMD為導致肌縮蛋白基因之無意義或框移突變之遺傳性或自發性突變的結果。引起DMD之大部分肌縮蛋白突變為外顯子之缺失，其破壞閱讀框架且引起肌縮蛋白基因之過早轉譯終止。DMD患者通常在兒童時期失去身體支持能力，在青少年時期逐漸變得虛弱，且在二十多歲時死亡。

【0040】 如本文所用，「肌縮蛋白」係指一種桿狀細胞質蛋白，其為經由細胞膜將肌纖維之細胞骨架與周圍細胞外基質連接起來的蛋白質複合物之重要部分。肌縮蛋白為負責調節肌細胞完整性及功能的細胞膜之肌縮蛋白聚糖複合物提供結構穩定性。如本文中可互換使用，肌縮蛋白基因或「DMD基因」在基因座Xp21處為2.2兆鹼基。原代轉錄量測值為約2,400 kb，成熟mRNA為約14 kb。79個外顯子編碼超過3500個胺基酸之

蛋白質。

【0041】 如本文所用，「外顯子52」係指肌縮蛋白基因之第52個外顯子。外顯子52緊鄰於DMD患者之框架破壞缺失。外顯子52可包含SEQ ID NO: 21之聚核苷酸。外顯子52可包含於SEQ ID NO: 22之聚核苷酸內。

【0042】 如本文所用，「強化子」係指含有多個活化子及抑制子結合位點之非編碼DNA序列。強化子之長度範圍為200 bp至1 kb，且可能在近端(啟動子上游5'或經調節基因之第一內含子內)，或可能遠端(相鄰基因或遠離基因座之基因間區域之內含子內)。經DNA循環，活性強化子獨立於核心DNA結合基元啟動子特異性接觸啟動子。4至5個強化子可與啟動子相互作用。類似地，強化子可調節超過一種基因而無鍵限制，且可「跳過」相鄰基因以調節更遠的基因。轉錄調節可涉及位於不同於啟動子所在的染色體中的元件。鄰近基因之近段強化子或啟動子可充當募集更多遠端元件之平台。

【0043】 如本文所用，「功能性」或「完全功能性」描述具有生物活性之蛋白質。「功能性基因」係指轉錄至mRNA之基因，使該基因轉譯成功能性蛋白質。

【0044】 如本文所用，「融合蛋白」係指經由最初編碼單獨蛋白質之兩個或更多個基因的接合所產生的嵌合蛋白質。融合基因之轉譯產生具有由原始蛋白質中之每一者衍生的功能性特性的單個多肽。

【0045】 如本文所用，「基因構築體」係指DNA或RNA分子，其包含編碼蛋白質之核苷酸序列。編碼序列包括可操作地連接至調節元件之起始及終止信號，該等調節元件包括能夠引導在投與核酸分子之個體之細胞

中表現的啟動子及聚腺苷酸化信號。如本文所用，術語「可表現形式」係指基因構築體，其含有可操作地連接至編碼蛋白質之編碼序列的必需調節元件，使得當存在於個體之細胞中時，將表現該編碼序列。

【0046】 如本文所用，「基因組編輯」係指改變基因。基因組編輯可包括校正或恢復突變基因。基因組編輯可改變剪接受體位點。基因組編輯可用於藉由改變相關基因來治療疾病或增強肌肉修復。

【0047】 如本文所用，術語「異源」係指包含兩個或更多個在自然界中彼此沒有相同關係之子序列的核酸。舉例而言，以重組方式產生之核酸通常具有兩個或更多個來自不相關基因之序列，該等不相關基因以合成方式佈置以產生新功能性核酸，例如來自一種來源之啟動子與來自另一種來源之編碼區。因此，在此情形下，兩種核酸彼此異源。當添加至細胞時，重組核酸亦將與細胞之內源基因異源。因此，在染色體中，異源核酸將包括已整合於染色體或非原生(非天然存在)染色體外核酸中之非原生(非天然存在)核酸。類似地，異源蛋白質指示蛋白質包含兩個或更多個在自然界中彼此沒有相同關係的子序列(例如「融合蛋白」，當兩個子序列由單核酸序列編碼時)。

【0048】 如本文所用，在兩個或更多個核酸或多肽序列之情形下，「相同/一致」或「一致性」意謂序列具有指定百分比之殘基，該等殘基在指定區域上相同。可藉由以下計算該百分比：最佳比對兩個序列，比較指定區域上之兩個序列，確定兩個序列中相同殘基出現的位置數，得到匹配位置數，將匹配位置數除以指定區域中之位置的總數目，且將結果乘以100，得到序列一致性之百分比。在兩個序列具有不同長度或比對產生一或多個交錯末端且指定比較區域僅包括單個序列之情況下，單個序列之殘

基包括於計算之分母中，但不包括於分子中。當比較DNA與RNA時，可將胸(腺)嘧啶(T)及尿嘧啶(U)視為等效的。一致性可人工地或藉由使用諸如BLAST或BLAST 2.0之電腦序列演算法來執行。

【0049】 如本文中可互換使用，「突變基因」或「突變型基因」係指已經歷可偵測突變之基因。突變基因已經歷變化，諸如基因物質之丟失、獲得或交換，其影響到基因之正常傳遞及表現。如本文所用，「破壞的基因」係指具有引起過早終止密碼子之突變的突變基因。相對於全長未破壞基因產物，破壞的基因產物被截短。

【0050】 如本文所用，「正常基因」係指尚未經歷變化，諸如基因物質之丟失、獲得或交換的基因。正常基因經歷正常基因傳遞及基因表現。

【0051】 如本文所用，「核酸」或「寡核苷酸」或「聚核苷酸」意謂共價連接在一起的至少兩種核苷酸。單鏈之描述亦界定互補鏈之序列。因此，核酸亦涵蓋所描繪之單鏈的互補鏈。核酸之諸多變體可用於與給定核酸相同的目的。因此，核酸亦涵蓋實質上相同的核酸及其互補序列。單鏈提供可在嚴格雜交條件下與目標序列雜交之探針。因此，核酸亦涵蓋在嚴格雜交條件下雜交之探針。

【0052】 核酸可為單鏈或雙鏈，或可含有雙鏈及單鏈序列之部分。核酸可為基因組及cDNA之DNA、RNA或雜交體，其中核酸可含有去氧核苷酸與核糖核苷酸之組合及鹼基之組合，該等鹼基包括尿嘧啶、腺嘌呤、胸(腺)嘧啶、胞嘧啶、鳥嘌呤、肌苷、黃嘌呤、次黃嘌呤、異胞嘧啶及異鳥嘌呤。核酸可藉由化學合成方法或藉由重組方法得到。

【0053】 「開放閱讀框架」係指以起始密碼子開始且在終止密碼子

處結束之一段密碼子。在具有多個外顯子之真核基因中，移除內含子且隨後在轉錄之後使外顯子接合在一起，得到用於蛋白質轉譯之最終mRNA。開放閱讀框架可為一段連續的密碼子。在一些實施例中，用於表現蛋白質之開放閱讀框架僅適用於剪接mRNA，而非基因組DNA。

【0054】 如本文所用，「可操作地連接」意謂基因之表現在與其空間連接之啟動子的控制下。啟動子可位於在其控制下的基因之5'(上游)或3'(下游)。啟動子與基因之間的距離可與啟動子與其控制之基因(啟動子所源自之基因)之間的距離大致相同。如此項技術中已知，此距離之變化可在不損失啟動子功能之情況下調節。

【0055】 當置於功能性關係中時，核酸或胺基酸序列與彼此「可操作地連接」(或「以操作方式連接」)。舉例而言，若啟動子或強化子調節編碼序列之轉錄或有助於調節編碼序列之轉錄，則啟動子或強化子可操作地連接至該編碼序列。可操作地連接之DNA序列通常相鄰，且可操作地連接之胺基酸序列通常相鄰且在相同閱讀框架中。然而，由於強化子一般在與啟動子隔開至多若干千鹼基或更高時起作用且內含子序列可具有可變的長度，因此一些聚核苷酸元件可操作地連接但不相鄰。類似地，在原代多肽序列中不相鄰的某些胺基酸序列仍然可由於例如多肽鏈之摺疊而可操作地連接。關於融合多肽，術語「以操作方式連接」及「可操作地連接」可指代組件中之每一者在與另一組件連接時執行與若未如此連接相同之功能。

【0056】 如本文所用，「部分功能性」描述由突變基因編碼且生物活性比功能性蛋白質低但比非功能性蛋白質高之蛋白質。

【0057】 如本文中可互換地使用，「過早終止密碼子」或「框外終

止密碼子」係指DNA序列中之無意義突變，其通常未在野生型基因中發現的位置處產生終止密碼子。與蛋白質之全長版本相比，過早終止密碼子可使得蛋白質被截短或縮短。

【0058】 如本文所用，「啟動子」意謂能夠賦予、活化或增強細胞之核酸表現的合成或天然衍生型分子。啟動子可包含一或多個特定轉錄調節序列，以進一步增強表現及/或改變其空間表現及/或時間表現。啟動子亦可包含遠端強化子或抑制子元件，該等元件可位於距轉錄之起始位點多達若干千鹼基對處。啟動子可來源於包括病毒、細菌、真菌、植物、昆蟲及動物等來源。啟動子可相對於發生表現之細胞、組織或器官或相對於發生表現之發育階段，或響應於諸如生理應激、病原體、金屬離子或誘導劑之外部刺激，組成性或差異地調節基因組件之表現。啟動子之代表性實例包括噬菌體T7啟動子、噬菌體T3啟動子、SP6啟動子、lac操作者啟動子、tac啟動子、SV40晚期啟動子、SV40早期啟動子、RSV-LTR啟動子、CMV IE啟動子、SV40早期啟動子或SV40晚期啟動子及CMV IE啟動子。

【0059】 在用於指代例如細胞或核酸、蛋白質或載體時，術語「重組」指示細胞、核酸、蛋白質或載體已藉由引入異源核酸或蛋白質或改變天然核酸或蛋白質進行修飾，或細胞係衍生自如此修飾之細胞。因此，舉例而言，重組細胞表現未在細胞之原生(天然存在)形式內發現之基因或表現原生基因之第二複本，該原生基因以其他方式正常或異常表現，完全表現或未完全表現。

【0060】 如本文所用，「骨胳肌」係指一類條紋肌肉，其受肌神經系統控制且藉由稱為肌腱之膠原蛋白纖維束附接至骨骼。骨胳肌由稱為肌

細胞或「肌細胞」(有時稱為「肌纖維」)之個別組件構成。肌細胞在稱為肌生成之過程中由發育的肌母細胞(一類產生肌細胞之胚胎母細胞)之融合形成。此等長圓柱形多核細胞亦稱為肌纖維。

【0061】 如本文所用，「骨胳肌病狀」係指與骨胳肌相關之病狀，諸如肌肉萎縮、衰老、肌肉退化、傷口癒合及肌肉無力或萎縮。

【0062】 如本文可互換使用，術語「個體」及「患者」係指任何脊椎動物，包括但不限於哺乳動物(例如牛、豬、駱駝、駱馬、馬、山羊、兔、綿羊、倉鼠、天竺鼠、貓、犬、大鼠及小鼠，非人類靈長類動物(例如猴，諸如獼猴或恆河猴、黑猩猩等)及人類)。在一些實施例中，個體可為人類或非人類。個體或患者可經歷其他形式之治療。

【0063】 「治療(treat/treating/treatment)」在本文中各自可互換使用以描述逆轉、緩解或抑制此類術語所適用之疾病的發展或此類疾病之一或多種症狀。治療可以短期或長期方式進行。該術語亦係指在患上疾病之前降低該疾病之嚴重程度與該疾病相關之一或多種症狀。此類在患病之前降低疾病之嚴重程度係指向投與時未患該疾病之個體投與抗體或醫藥組合物。「預防」亦係指預防疾病或與此類疾病相關之一或多種症狀的復發。「治療」及「治療上」係指治療行為，如上文所定義之「治療」。

【0064】 本文中關於核酸所用之「變體」意謂(i)所提及之核苷酸序列的一部分或片段；(ii)所提及之核苷酸序列或其部分的互補序列；(iii)與所提及之核酸或其互補序列實質上一致的核酸；或(iv)在嚴格條件下與所提及之核酸、其互補序列或與其實質上一致之序列雜交的核酸。

【0065】 對於因胺基酸之插入、缺失或保守取代而在胺基酸序列上不同的肽或多肽之「變體」，但保留至少一種生物活性。變體亦可意謂胺

基酸序列與保持至少一種生物活性之參考蛋白質之胺基酸序列實質上一致的蛋白質。胺基酸之保守取代(亦即，胺基酸被具有類似特性(例如帶電區域之親水性程度及分佈)之不同胺基酸置換)在此項技術中公認為典型地涉及微小變化。如此項技術中所理解，此等微小變化可部分地藉由考慮胺基酸之親水指數來識別。Kyte等人，J. Mol. Biol. 157:105-132 (1982)胺基酸之親水指數係基於其疏水性及電荷之考慮。此項技術中已知具有類似親水指數之胺基酸可經取代且仍保留蛋白質功能。在一個態樣中，親水指數為±2之胺基酸經取代。胺基酸之親水性亦可用於展現將產生保留生物功能之蛋白質的取代。考慮在肽之情形下胺基酸之親水性准許計算彼肽之最大局部平均親水性。可使用親水值在彼此±2內的胺基酸進行取代。胺基酸之疏水性指數及親水性值均受該胺基酸之特定側鏈影響。與彼觀測結果一致，與生物功能相容之胺基酸取代理解為視胺基酸且尤其彼等胺基酸之側鏈的相對相似性(如由疏水性、親水性、電荷、尺寸及其他特性所展現)而定。

【0066】 如本文所用，「載體」意味含有複製起點之核酸序列。載體可為病毒載體、噬菌體、細菌人工染色體或酵母人工染色體。載體可為DNA或RNA載體。載體可為自我複製染色體外載體，且較佳為DNA質體。舉例而言，載體可編碼本文所述之基於CRISPR/Cas之基因組編輯系統，該系統包括編碼Cas蛋白質或融合蛋白之聚核苷酸序列，及/或至少一個SEQ ID NO: 19或SEQ ID NO: 20之gRNA核苷酸序列或靶向包含SEQ ID NO: 17或SEQ ID NO: 18之核苷酸序列的gRNA。

2. 用於恢復肌縮蛋白之基於CRISPR/Cas之基因組編輯系統

【0067】 本文提供適用於恢復肌縮蛋白基因功能之基於

CRISPR/Cas之基因組編輯系統。在一些實施例中，基於CRISPR/Cas之基因組編輯系統包括Cas蛋白質或融合蛋白及結合及靶向對應於SEQ ID NO: 17或SEQ ID NO: 18之聚核苷酸序列的至少一個引導RNA (gRNA)。在一些實施例中，至少一個引導RNA (gRNA)包含SEQ ID NO: 19或SEQ ID NO: 20之聚核苷酸或由其編碼。融合蛋白可包含兩個異源多肽域。在一些實施例中，融合蛋白包含Cas蛋白質及鹼基編輯域，或具有其他酶促功能之域。在一些實施例中，至少一個gRNA結合及靶向對應於以下之聚核苷酸序列：a) SEQ ID NO: 17或SEQ ID NO: 18之片段； b) SEQ ID NO: 17或SEQ ID NO: 18之互補序列或其片段； c) 與SEQ ID NO: 17或SEQ ID NO: 18或其互補序列實質上一致的核酸；或d) 在嚴格條件下雜交至SEQ ID NO: 17或SEQ ID NO: 18、其互補序列或與其實質上一致之序列的核酸。

a) 肌縮蛋白基因

【0068】 肌縮蛋白為一種桿狀細胞質蛋白，其為經由細胞膜將肌纖維之細胞骨架與周圍細胞外基質連接起來的蛋白質複合物之一部分（**圖1**)。肌縮蛋白為細胞膜之肌縮蛋白聚糖複合物提供結構穩定性。肌縮蛋白基因在基因座Xp21處為2.2兆鹼基。原代轉錄量測值為約2,400 kb，成熟mRNA為約14 kb。79個外顯子包括約220萬個核苷酸且編碼超過3500個胺基酸之蛋白質 (**圖2**)。正常骨架肌肉組織僅含有少量肌縮蛋白但不存在導致嚴重及不可治癒的症狀的異常表現。肌縮蛋白基因之一些突變使得在受感染患者中產生缺陷性肌縮蛋白及嚴重的肌縮蛋白性表現型。肌縮蛋白基因之一些突變在受感染患者中產生部分功能性肌縮蛋白質及溫和得多的肌縮蛋白性表現型。

【0069】 DMD為導致肌縮蛋白基因之無意義或框移突變之遺傳性或自發性突變的結果。DMD為最普遍的致死性遺傳性兒童疾病且影響約5,000/1的新生兒。DMD之特徵在於進行性肌肉無力(由於缺乏功能性肌縮蛋白基因)，通常導致二十多歲患者死亡。大部分突變為破壞閱讀框架之肌縮蛋白基因的缺失。DMD的天然存在之突變及其結果相對充分理解。已知發生在含於桿狀域內之外顯子45-55區域中的框內缺失可產生高度功能性肌縮蛋白質，且許多載劑無症狀或呈現輕度症狀。肌縮蛋白之外顯子45-55為突變熱點。此外，超過60%患者可理論上藉由靶向肌縮蛋白基因之此區域中的外顯子得以治療。已努力在mRNA剪接期間藉由跳過非必需外顯子(例如跳過外顯子45)來恢復DMD患者之破壞的肌縮蛋白閱讀框架，從而產生內部缺失但具有功能性的肌縮蛋白質。缺失內部肌縮蛋白外顯子(例如缺失外顯子45)保留了恰當閱讀框架且可產生內部截短但具有部分功能性的肌縮蛋白質。肌縮蛋白之外顯子45-55之間的缺失產生與DMD相比更溫和的表現型。

【0070】 肌縮蛋白基因可為突變肌萎縮蛋白基因。肌縮蛋白基因可為野生型肌萎縮蛋白基因。肌縮蛋白基因可具有在功能上與野生型肌萎縮蛋白基因相同之序列，舉例而言，序列可經密碼子優化但仍編碼與野生型肌縮蛋白相同之蛋白質。突變肌縮蛋白基因可包括一或多種相對於野生型肌縮蛋白基因之突變。突變可包括例如核苷酸缺失、取代、添加、置換或其組合。突變可包括至少一個內含子或外顯子之全部或部分的缺失。突變肌縮蛋白基因之外顯子可經突變或自肌縮蛋白基因至少部分地缺失。突變肌縮蛋白基因之外顯子可完全缺失。突變肌縮蛋白基因可具有對應於野生型肌縮蛋白基因中之相應序列的其一部分或片段。在一些實施例中，由缺

失或突變型外顯子引起的破壞肌縮蛋白基因可在DMD患者中藉由添加回相應野生型外顯子來恢復。在一些實施例中，由缺失或突變型外顯子52引起的破壞的肌縮蛋白可在DMD患者中藉由添加回野生型外顯子52來恢復。在某些實施例中，添加外顯子52以恢復閱讀框架改善了DMD個體(包括具有缺失突變之DMD個體)的表現型。在某些實施例中，可添加一或多個外顯子且將其插入破壞的肌縮蛋白基因中。可添加及插入一或多個外顯子以便恢復肌縮蛋白中之相應突變或缺失型外顯子。除添加回及插入外顯子52以外，可添加一或多個外顯子且將其插入破壞的肌縮蛋白基因中。在某些實施例中，肌縮蛋白基因之外顯子52係指肌縮蛋白基因之第52個外顯子。外顯子52常緊鄰於DMD患者之框架破壞缺失且已靶向用於跳過基於寡核苷酸之外顯子的臨床試驗中。

【0071】 本發明之基因構築體(例如載體)可介導將外顯子52高效添加至肌縮蛋白基因(例如人類肌縮蛋白基因)中。本發明之基因構築體(例如載體)可恢復來自DMD患者之細胞的肌縮蛋白質表現。外顯子52緊鄰於DMD之框架破壞缺失。向肌縮蛋白轉錄物中添加外顯子52可用於治療DMD患者。可將本發明之基因構築體(例如載體)轉染至人類DMD細胞中且介導有效的基因修飾及轉化成正確的閱讀框架。蛋白質恢復可伴隨有框架恢復且在經基於CRISPR/Cas之基因組編輯系統處理細胞的主體群體中偵測到。

b) 融合蛋白

【0072】 基於CRISPR/Cas之基因編輯系統可包括融合蛋白或編碼融合蛋白之核酸序列。融合蛋白可包括Cas蛋白質及基因/基因組編輯域，或具有其他酶促功能之域。在一些實施例中，編碼融合蛋白之核酸序列為

DNA。在一些實施例中，編碼融合蛋白之核酸序列為RNA。

i) Cas蛋白質

【0073】 基於CRISPR/Cas之基因編輯系統可包括Cas蛋白質。Cas蛋白質與gRNA之3'端形成複合物。基於CRISPR系統之特異性取決於兩種因素：目標序列及前間隔序列鄰近基元(PAM)。目標序列位於gRNA之5'端上且經設計以在稱為前間隔序列之正確的DNA序列處與宿主DNA上之鹼基對結合。藉由簡單交換gRNA之識別序列，可將Cas蛋白質引導至新基因組目標。PAM序列位於待改變之DNA上且由Cas蛋白質識別。Cas蛋白質之PAM識別序列可具有物種特異性。

【0074】 Cas9蛋白質為一種使核酸裂解且由CRISPR基因座編碼及涉及II型CRISPR系統的核酸內切酶。Cas9分子可與一或多個gRNA分子相互作用，且與gRNA分子一起定位至包含目標域且在某些實施例中，PAM序列之位點。Cas9分子識別PAM序列之能力可例如使用如此項技術中已知之轉型分析來確定。在一些實施例中，基於CRISPR/Cas之基因編輯系統包括來自化膿性鏈球菌之Cas9蛋白質。在一些實施例中，Cas9蛋白質包含SEQ ID NO: 1之胺基酸序列。在一些實施例中，基於CRISPR/Cas之基因編輯系統包括來自金黃色葡萄球菌之Cas9蛋白質。在一些實施例中，Cas9蛋白質包含SEQ ID NO: 2之胺基酸序列。

【0075】 在一些實施例中，基於CRISPR/Cas之基因編輯系統包括催化死亡的dCas9。在一些實施例中，Cas9蛋白質可經突變以使得核酸酶活性失活的。不具有核酸內切酶活性之失活的Cas9蛋白質(「iCas9」，亦被稱作「dCas9」)可藉由gRNA靶向細菌、酵母及人類細胞中之基因，以經由位阻使基因表現沉默。關於使核酸酶活性失活之化膿性鏈球菌Cas9

序列的例示性突變包括：D10A、E762A、H840A、N854A、N863A及/或D986A。具有D10A突變之化膿性鏈球菌Cas9蛋白質可包含SEQ ID NO: 3之胺基酸序列。具有D10A及H849A突變之化膿性鏈球菌Cas9蛋白質可包含SEQ ID NO: 4之胺基酸序列。關於使核酸酶活性失活之金黃色葡萄球菌Cas9序列的例示性突變包括D10A及N580A。

【0076】Cas9蛋白質或突變Cas9蛋白質可來自任何細菌或古菌物種，諸如化膿性鏈球菌、金黃色葡萄球菌、嗜熱鏈球菌(*Streptococcus thermophiles*)或腦膜炎雙球菌(*Neisseria meningitides*)。在一些實施例中，Cas蛋白質或突變Cas9蛋白質為衍生自以下之細菌屬的Cas9蛋白質：鏈球菌屬(*Streptococcus*)、葡萄球菌屬(*Staphylococcus*)、芽孢桿菌屬(*Brevibacillus*)、棒狀桿菌屬(*Corynebacter*)、薩特菌屬(*Sutterella*)、軍團菌屬(*Legionella*)、弗朗西斯氏菌屬(*Francisella*)、螺旋體屬(*Treponema*)、產線菌屬(*Filifactor*)、真桿菌屬(*Eubacterium*)、乳桿菌屬(*Lactobacillus*)、擬桿菌屬(*Bacteroides*)、黃沃拉菌屬(*Flaviivola*)、黃桿菌屬(*Flavobacterium*)、球毛菌屬(*Sphaerochaeta*)、固氮螺旋菌屬(*Azospirillum*)、葡糖醋桿菌屬(*Gluconacetobacter*)、奈瑟氏菌屬(*Neisseria*)、羅氏菌屬(*Roseburia*)、紅菌屬(*Parvibaculum*)、葡萄球菌(*Staphylococcus*)、硝化裂化器菌(*Nitratifractor*)、黴漿菌屬(*Mycoplasma*)或曲桿菌類(*Campylobacter*)。在一些實施例中，Cas9蛋白質或突變Cas9蛋白質選自包括但不限於以下之群：化膿性鏈球菌(*Streptococcus pyogenes*)、新兇手弗朗西斯氏菌(*Francisella novicida*)、金黃色葡萄球菌(*Staphylococcus aureus*)、腦膜炎雙球菌(*Neisseria meningitides*)、嗜熱鏈球菌(*Streptococcus thermophiles*)、齒垢密螺旋體

(*Treponema denticola*)、枯草芽孢桿菌(*Brevibacillus laterosporus*)、空腸彎曲桿菌(*Campylobacter jejuni*)、白喉棒狀桿菌(*Corynebacterium diphtheria*)、凸腹真桿菌(*Eubacterium ventriosum*)、巴氏鏈球菌(*Streptococcus pasteurianus*)、香腸乳桿菌(*Lactobacillus farciminis*)、球狀球毛菌(*Sphaerochaeta globus*)、固氮螺旋菌屬(*Azospirillum*)、重氮營養葡糖醋桿菌屬(*Gluconacetobacter diazotrophicus*)、灰色奈瑟球菌(*Neisseria cinerea*)、腸道羅斯氏菌(*Roseburia intestinalis*)、食清潔劑細小棒菌(*Parvibaculum lavamentivorans*)、碳源硝化桿菌(*Nitratifractor salsuginis*)及曲桿菌類(*Campylobacter lari*)。

【0077】 在某些實施例中，Cas9分子或突變Cas9蛋白質與目標核酸相互作用及裂解目標核酸之能力具有PAM序列相關性。PAM序列為目標核酸中之序列。在某些實施例中，目標核酸之裂解發生於PAM序列上游。來自不同細菌物種之Cas9分子可識別不同序列基元(例如PAM序列)。在某些實施例中，化膿性鏈球菌之Cas9分子識別序列基元NGG (SEQ ID NO: 10)且引導對彼序列上游之目標核酸序列1至10，諸如3至5 bp進行裂解(參見例如Mali 2013)。在某些實施例中，金黃色葡萄球菌之Cas9分子識別序列基元NNGRR (R=A或G) (SEQ ID NO: 12)且引導對彼序列上游之目標核酸序列1至10，諸如3至5 bp進行裂解。在某些實施例中，金黃色葡萄球菌之Cas9分子識別序列基元NNGRRN (R=A或G) (SEQ ID NO: 13)且引導對彼序列上游之目標核酸序列1至10，諸如3至5 bp進行裂解。在某些實施例中，金黃色葡萄球菌之Cas9分子識別序列基元NNGRRT (R=A或G) (SEQ ID NO: 14)且引導對彼序列上游之目標核酸序列1至10，諸如3至5 bp進行裂解。在某些實施例中，金黃色葡萄球菌之Cas9分子識

別序列基元NNGRRT (R=A或G) (SEQ ID NO: 15)且引導對彼序列上游之目標核酸序列1至10，諸如3至5 bp進行裂解。在前述實施例中，N可為任何核苷酸殘基，例如A、G、C或T中之任一者。Cas9分子可經工程改造以改變Cas9分子之PAM特異性。

【0078】 在一些實施例中，Cas9蛋白質或突變Cas9蛋白質可識別PAM序列NGG (SEQ ID NO: 10)或NGA (SEQ ID NO: 16)。在一些實施例中，Cas9蛋白質或突變Cas9蛋白質可識別PAM序列NNNRRT (SEQ ID NO: 11)。在一些實施例中，Cas9蛋白質或突變Cas9蛋白質可識別PAM序列ATTCCT (SEQ ID NO: 9)。在一些實施例中，Cas9蛋白質或突變Cas9蛋白質為金黃色葡萄球菌之Cas9蛋白質且識別序列基元NNGRR (R=A或G) (SEQ ID NO: 12)、NNGRRN (R=A或G) (SEQ ID NO: 13)、NNGRRT (R=A或G) (SEQ ID NO: 14)或NNGRRV (R=A或G) (SEQ ID NO: 15)。在前述實施例中，N可為任何核苷酸殘基，例如A、G、C或T中之任一者。Cas9分子可經工程改造以改變Cas9分子之PAM特異性。

【0079】 另外或替代地，編碼Cas9分子或Cas9多肽之核酸可包含核定位序列(NLS)。核定位序列為此項技術中已知的。

c) gRNA

【0080】 基於CRISPR/Cas之基因組編輯系統可包括至少一個gRNA。gRNA可靶向肌縮蛋白基因之片段。gRNA可靶向突變肌縮蛋白基因之片段。gRNA可靶向野生型肌縮蛋白基因之片段。片段之長度可為約5至約200、約10至約200、約5至約300或約10至約300個核苷酸。片段之長度可為至少約5個、至少約10個、至少約15個、至少約20個、至少約30個、至少約40個、至少約50個或至少約100個核苷酸。gRNA可靶向包含

突變或缺失之肌縮蛋白基因的片段或部分，或靶向接近或與其並列的序列。gRNA可靶向與肌縮蛋白基因之外顯子並列的內含子。gRNA可靶向與突變肌縮蛋白基因之外顯子並列的內含子。野生型肌縮蛋白基因之片段可由本文所詳述之兩個gRNA間隔序列及/或PAM序列側接。每個gRNA間隔序列可包含選自SEQ ID NO: 5-8及25-45之胺基酸序列。兩個gRNA間隔序列可為相同的。兩個gRNA間隔序列可為不同的。在一些實施例中，該兩個gRNA間隔序列中之至少一者包含序列SEQ ID NO: 25或SEQ ID NO: 26。外顯子可選自肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59及61-66。在一些實施例中，該外顯子為外顯子52。gRNA提供靶向基於CRISPR/Cas之基因編輯系統。gRNA為兩個非編碼RNA之融合物：crRNA及tracrRNA。sgRNA可藉由交換編碼20 bp前間隔序列之序列靶向任何所需DNA序列，該序列經由與所需DNA目標之互補鹼基配對賦予靶向特異性。gRNA模擬參與II型效應子系統之天然存在的crRNA:tracrRNA雙工。可包括(例如)42-核苷酸crRNA與75-核苷酸tracrRNA之此雙工充當Cas9之引導。

【0081】在一些實施例中，至少一個gRNA可靶向且結合目標區域。在一些實施例中，可使用在1個與20個之間的gRNA改變目標基因，例如改變剪接受體位點。舉例而言，1個gRNA與20個gRNA之間、1個gRNA與15個gRNA之間、1個gRNA與10個gRNA之間、1個gRNA與5個gRNA之間、2個gRNA與20個gRNA之間、2個gRNA與15個gRNA之間、2個gRNA與10個gRNA之間、2個gRNA與5個gRNA之間、5個gRNA與20個gRNA之間、5個gRNA與15個gRNA之間或5個gRNA與10個gRNA之間的gRNA可包括於基於CRISPR/Cas之基因編輯系統中且用於改變剪接受

體位點。在一些實施例中，至少1個gRNA、至少2個gRNA、至少3個gRNA、至少4個gRNA、至少5個基因、至少6個gRNA、至少7個gRNA、至少8個gRNA、至少9個基因、至少10個gRNA、至少11個gRNA、至少12個gRNA、至少13個基因、至少14個gRNA、至少15個gRNA或至少20個gRNA可包括於基於CRISPR/Cas之基因編輯系統中且用於改變剪接受體位點。

【0082】基於CRISPR/Cas之基因編輯系統可使用具有不同序列及長度之gRNA。gRNA可包含目標DNA序列之互補聚核苷酸序列，諸如包含SEQ ID NO: 17或SEQ ID NO: 18之目標序列或包含SEQ ID NO: 17或SEQ ID NO: 18之目標序列的互補聚核苷酸序列，接著為NGG。gRNA可包含互補聚核苷酸序列之5'端處的「G」。gRNA可包含目標DNA序列之至少10個鹼基對、至少11個鹼基對、至少12個鹼基對、至少13個鹼基對、至少14個鹼基對、至少15個鹼基對、至少16個鹼基對、至少17個鹼基對、至少18個鹼基對、至少19個鹼基對、至少20個鹼基對、至少21個鹼基對、至少22個鹼基對、至少23個鹼基對、至少24個鹼基對、至少25個鹼基對、至少30個鹼基對互補聚核苷酸序列，接著為NGG。gRNA可靶向目標基因之啟動子區、強化子區或轉錄區中的至少一者。

【0083】至少一個gRNA可結合及靶向包含SEQ ID NO: 17或SEQ ID NO: 18之核酸序列。目標序列可包含SEQ ID NO: 17或SEQ ID NO: 18之聚核苷酸或其片段或其截短，諸如5'-截短。截短可比序列SEQ ID NO: 17或SEQ ID NO: 18短1、2、3、4、5或6個核苷酸。gRNA可包含對應於SEQ ID NO: 17或SEQ ID NO: 18之聚核苷酸、其互補序列、其變體或其片段。gRNA可由包含SEQ ID NO: 17、SEQ ID NO: 18之聚核苷酸

序列編碼。靶向基因組中之目標序列的gRNA部分可稱為gRNA間隔序列或前間隔序列。前間隔序列可定義為與基因組中之靶向序列互補的gRNA部分。前間隔序列可包含SEQ ID NO: 17或SEQ ID NO 18之聚核苷酸，或其片段，或其截短，或其互補序列。gRNA可包括gRNA骨架。gRNA骨架促進Cas9結合至gRNA及核酸內切酶活性。gRNA骨架為跟在對應於gRNA靶向之序列的gRNA部分後的聚核苷酸序列。gRNA靶向部分與gRNA骨架共同形成一個聚核苷酸。在一些實施例中，gRNA靶向部分與gRNA骨架共同可包含聚核苷酸序列SEQ ID NO: 19或SEQ ID NO: 20或其互補序列。在一些實施例中，gRNA靶向部分與gRNA骨架共同由聚核苷酸序列SEQ ID NO: 19或SEQ ID NO: 20或其互補序列編碼。gRNA可由SEQ ID NO: 19之聚核苷酸、其互補序列、其變體或其片段或SEQ ID NO: 20之聚核苷酸、其互補序列、其變體或其片段編碼。

d) 供體序列

【0084】 基於CRISPR/Cas之基因編輯系統可包括至少一個供體序列。供體序列可包含肌縮蛋白基因之片段。舉例而言，供體序列可包含編碼肌縮蛋白基因之外顯子或外顯子之任何組合的核酸序列。供體序列可包含野生型肌縮蛋白基因之外顯子或其功能性等效物。外顯子可選自肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59、61-66。在一些實施例中，外顯子為肌縮蛋白基因之外顯子52。供體序列可包括野生型肌縮蛋白基因之片段或其功能性等效物，且其片段或功能性等效物可由兩個gRNA間隔序列側接。供體序列可進一步包括至少一個額外聚核苷酸，其對應於待插入之外顯子周圍或附近的內含子序列。供體序列可進一步包括至少一個額外聚核苷酸，其對應於外顯子52周圍或附近之內含子序列。供

體序列可包括SEQ ID NO: 21或SEQ ID NO: 22中之至少一者的核酸序列、其互補序列、其變體或其片段。

【0085】 gRNA及供體序列可以多種莫耳比存在。gRNA與供體序列之間的莫耳比可為1:1、或1:15、或5:1至1:10、或1:1至1:5。gRNA與供體序列之間的莫耳比可為至少1:1、至少1:2、至少1:3、至少1:4、至少1:5、至少1:6、至少1:7、至少1:8、至少1:9、至少1:10、至少1:15或至少1:20。gRNA與供體序列之間的莫耳比可低於20:1、低於15:1、低於10:1、低於9:1、低於8:1、低於7:1、低於6:1、低於5:1、低於4:1、低於3:1、低於2:1或低於1:1。

3. 用於恢復肌縮蛋白功能之組合物

【0086】 本文揭示用於恢復肌縮蛋白功能之組合物。組合物藉由添加一或多個外顯子以恢復肌縮蛋白之閱讀框架來恢復肌縮蛋白功能。舉例而言，待添加之外顯子可為外顯子52。組合物可包括上文所揭示之基於CRISPR/Cas之基因編輯系統。組合物亦可包括病毒遞送系統。舉例而言，病毒遞送系統可包括腺相關病毒載體或經修飾豆狀病毒載體。

【0087】 將核酸引入宿主細胞之方法為此項技術中已知，且可使用任何已知方法將核酸(例如表現構築體)引入細胞中。適合之方法包括例如病毒或噬菌體感染、轉染、共軛、原生質體融合、聚陽離子或脂質與核酸共軛物、脂質體轉染、電穿孔、核轉染、免疫微脂囊、磷酸鈣沈澱、聚乙二亞胺(PEI)介導之轉染、DEAE-聚葡萄糖介導之轉染、脂質體介導之轉染、粒子槍技術、磷酸鈣沈澱、直接微注射、奈米粒子介導之核酸傳遞及其類似者。在一些實施例中，組合物可藉由mRNA遞送及核糖核蛋白(RNP)複合物遞送來遞送。

a)構築體及質體

【0088】如上文所述，組合物可包含編碼本文所揭示之基於CRISPR/Cas之基因編輯系統的基因構築體。基因構築體，諸如質體或表現載體可包含編碼基於CRISPR/Cas之基因編輯系統及/或gRNA中之至少一者的核酸。如上文所述，組合物可包含編碼經修飾腺相關病毒(AAV)載體之基因構築體及編碼本文所揭示之基於CRISPR/Cas之基因編輯系統的核酸序列。在一些實施例中，如上文所述，組合物可包含編碼經修飾腺病毒載體之基因構築體及編碼本文所揭示之基於CRISPR/Cas之基因編輯系統的核酸序列。基因構築體，諸如質體可包含編碼基於CRISPR/Cas之基因編輯系統的核酸。如上文所述，組合物可包含編碼經修飾豆狀病毒載體之基因構築體。基因構築體，諸如質體可包含編碼Cas蛋白質或融合蛋白及至少一個gRNA之核酸。基因構築體可作為起作用之染色體外分子存在於細胞中。基因構築體可為線性微型染色體，包括著絲點、端粒或質體或黏質體。

【0089】基因構築體亦可為重組病毒載體之基因組的一部分，包括重組慢病毒、重組腺病毒及重組腺病毒相關病毒。基因構築體可為存活於細胞中的減毒活微生物或重組微生物載體中基因物質的一部分。基因構築體可包含用於核酸之編碼序列之基因表現的調節元件。調節元件可為啟動子、強化子、起始密碼子、終止密碼子或聚腺苷酸化信號。

【0090】核酸序列可構成可作為載體之基因構築體。載體可能能夠在哺乳動物之細胞中表現Cas蛋白質或融合蛋白，諸如基於CRISPR/Cas之基因編輯系統。載體可為重組載體。載體可包含編碼Cas蛋白質或融合蛋白(諸如基於CRISPR/Cas之基因編輯系統)之異源核酸。載體可為質

體。載體可適用於用編碼基於CRISPR/Cas之基因編輯系統的核酸轉染細胞，在表現基於CRISPR/Cas之基因編輯系統的條件下培養及維持轉型的宿主細胞。

【0091】 可針對穩定性及較高表現量使編碼序列最佳化。在一些情況下，密碼子經選擇以減少RNA之二級結構形成，諸如由於分子內結合而形成之二級結構形成。

【0092】 載體可包含編碼基於CRISPR/Cas之基因編輯系統的異源核酸且可進一步包含可位於基於CRISPR/Cas之基因編輯系統編碼序列上游的起始密碼子及可位於基於CRISPR/Cas之基因編輯系統編碼序列下游的終止密碼子。起始及終止密碼子可在基於CRISPR/Cas之基因編輯系統編碼序列的框內。載體亦可包含可操作地連接至基於CRISPR/Cas之基因編輯系統編碼序列的啟動子。啟動子可為普遍存在的啟動子。啟動子可為組織特異性啟動子。組織特異性啟動子可為肌肉特異性啟動子。基於CRISPR/Cas之基因編輯系統可在光誘導性或化學誘導性的控制下，使得能夠在空間及時間上動態控制基因/基因組編輯。可操作地連接至基於CRISPR/Cas之基因編輯系統編碼序列的啟動子可為來自猴病毒40 (SV40)之啟動子、小鼠乳腺腫瘤病毒(MMTV)啟動子、人類免疫缺陷病毒(HIV)啟動子諸如牛免疫缺陷病毒(BIV)長末端重複序列(LTR)啟動子、莫洛尼病毒(Moloney virus)啟動子、禽類白血病病毒(ALV)啟動子、巨細胞病毒(CMV)啟動子諸如CMV極早期啟動子(CMV-IE)、埃-巴二氏病毒(Epstein Barr virus，EBV)啟動子，或勞斯肉瘤病毒(Rous sarcoma virus，RSV)啟動子。啟動子亦可為來自人類基因，諸如人類泛素C (hUbC)、人類肌動蛋白、人類肌凝蛋白、人類血紅素、人類肌肉肌酸或人類金屬硫蛋白之

啟動子。啟動子亦可為天然或合成的組織特異性啟動子，諸如肌肉或皮膚特異性啟動子。此類啟動子之實例描述於美國專利申請公開案第US20040175727號中，其內容以全文併入本文中。啟動子可為例如CK8啟動子、Spc512啟動子、MHCK7啟動子。

【0093】載體亦可包含聚腺苷酸化信號，其可位於基於CRISPR/Cas之基因編輯系統下游。聚腺苷酸化信號可為SV40聚腺苷酸化信號、LTR聚腺苷酸化信號、牛生長激素(bGH)聚腺苷酸化信號、人類生長激素(hGH)聚腺苷酸化信號或人類β-血球蛋白聚腺苷酸化信號。SV40聚腺苷酸化信號可為來自pCEP4載體之聚腺苷酸化信號(Invitrogen, San Diego, CA)。

【0094】載體亦可包含基於CRISPR/Cas之基因編輯系統或sgRNA上游之強化子。強化子對於DNA表現可為必需的。強化子序列可為人類肌動蛋白、人類肌凝蛋白、人類血紅蛋白、人類肌肉肌酸，或病毒強化子，諸如來自CMV、HA、RSV或EBV之強化子。聚核苷酸功能強化子描述於美國專利第5,593,972、5,962,428及WO94/016737號中，其內容各自以引用的方式完全併入。載體亦可包含哺乳動物複製起點，以便在染色體外維持載體且在細胞中產生載體之多個複本。載體亦可包含調節序列，其可能極適合於投與載體之哺乳動物或人類細胞的基因表現。載體亦可包含報導基因，諸如綠色螢光蛋白質(「GFP」)及/或可選標記，諸如潮黴素(「Hygro」)。

【0095】載體可為用以藉由常規技術及可易於獲得之起始物質產生蛋白質之表現載體或系統，該等物質包括Sambrook等人, Molecular Cloning and Laboratory Manual,第二版, Cold Spring Harbor (1989)，其

以完全引用之方式併入。在一些實施例中，載體可包含編碼基於CRISPR/Cas之基因編輯系統的核酸序列，包括編碼Cas蛋白質或融合蛋白的核酸序列及編碼至少一個gRNA的核酸序列，該至少一個gRNA包含SEQ ID NO: 19之核酸序列、其互補序列、其變體或其片段或SEQ ID NO: 20之核酸序列、其互補序列、其變體或其片段，或靶向SEQ ID NO: 17或SEQ ID NO: 18之核酸序列的gRNA、其變體或其片段。在一些實施例中，兩個載體可包含編碼基於CRISPR/Cas之基因編輯系統的核酸序列，包括包含編碼Cas蛋白質或融合蛋白之核酸序列的第一載體及包含編碼至少一個gRNA之核酸序列的第二載體。

【0096】 在一些實施例中，組合物由mRNA及蛋白質/RNA複合物(核糖核蛋白(RNP))遞送。舉例而言，經純化Cas蛋白質或融合蛋白可與引導RNA組合，以形成RNP複合物。本文所述之方法亦可需要遞送本文所述之DNA供體序列。

b)經修飾豆狀病毒載體

【0097】 用於添加或插入外顯子52之組合物可包括經修飾豆狀病毒載體。經修飾豆狀病毒載體包括編碼Cas蛋白質或融合蛋白之第一聚核苷酸序列及編碼至少一個gRNA之第二聚核苷酸序列。第一聚核苷酸序列可操作地連接至啟動子。啟動子可為組成性啟動子、誘導性啟動子、可抑制啟動子或可調節啟動子。

【0098】 第二聚核苷酸序列編碼至少1個gRNA。舉例而言，第二聚核苷酸序列可編碼至少1個gRNA、至少2個gRNA、至少3個gRNA、至少4個gRNA、至少5個gRNA、至少6個gRNA、至少7個gRNA、至少8個gRNA、至少9個gRNA、至少10個gRNA、至少11 gRNA、至少12個

gRNA、至少13個gRNA、至少14個gRNA、至少15個gRNA、至少16個gRNA、至少17個gRNA、至少18個gRNA、至少19個gRNA或至少20個gRNA。第二聚核苷酸序列可操作地連接至啟動子。啟動子可為組成性啟動子、誘導性啟動子、可抑制啟動子或可調節啟動子。至少一個gRNA可結合至目標基因或基因座，諸如對應於外顯子52之目標區域。

c) 腺相關病毒載體

【0099】 AAV可用於使用各種構築體組態將組合物遞送至細胞。舉例而言，AAV可在單獨載體上遞送Cas蛋白質或融合蛋白及gRNA表現盒。或者，可將Cas蛋白質或融合蛋白與至多兩個gRNA表現盒兩者組合於4.7 kb封裝限制內的單一AAV載體中。

【0100】 如上文所述，組合物包括經修飾腺相關病毒(AAV)載體。經修飾AAV載體可能能夠在哺乳動物細胞中遞送及表現位點特異性核酸酶。舉例而言，經修飾AAV載體可為AAV-SASTG載體(Piacentino等人(2012) Human Gene Therapy 23:635-646)。經修飾AAV載體可基於若干衣殼類型中之一或多者，包括AAV1、AAV2、AAV5、AAV6、AAV8及AAV9。經修飾AAV載體可基於具有替代性向肌肉性AAV衣殼的AAV2假型，諸如AAV2/1、AAV2/6、AAV2/7、AAV2/8、AAV2/9、AAV2.5及AAV/SASTG載體，其藉由全身性及局部遞送以有效轉導骨胳肌或心肌(Seto等人，Current Gene Therapy (2012) 12:139-151)。構築體可包含SEQ ID NO: 23之聚核苷酸序列。構築體可包含SEQ ID NO: 24之聚核苷酸序列。

4. 用以恢復具有突變肌縮蛋白基因之個體的肌縮蛋白功能的方法

【0101】 本發明之主題提供用於恢復患有DMD及/或具有突變肌縮

蛋白基因之細胞及/或個體的肌縮蛋白功能(例如突變肌縮蛋白基因，例如突變人類肌縮蛋白基因)的方法。該方法可包括向細胞或個體投與基於CRISPR/Cas之基因編輯系統、編碼該基於CRISPR/Cas之基因編輯系統的聚核苷酸或載體或如上文所述之該基於CRISPR/Cas9之基因編輯系統的組合物。在一些實施例中，該個體患有杜興氏肌肉萎縮症(Duchenne Muscular Dystrophy)。在一些實施例中，肌縮蛋白功能藉由插入對應於該一或多個缺失或突變型外顯子之肌縮蛋白基因的一或多個野生型外顯子來恢復。在一些實施例中，該肌縮蛋白功能藉由插入該野生型肌縮蛋白基因之外顯子52來恢復。

【0102】 該方法可包括向細胞或個體投與如上文所述之本發明之基因構築體(例如載體)或包含其之組合物。如上文所述，該方法可包含向個體之骨胳肌及/或心肌投與本發明之基因構築體(例如載體)或包含其之組合物以用於骨胳肌及/或心肌之基因組編輯。使用本發明之基因構築體(例如載體)或包含其之組合物將基於CRISPR/Cas之基因編輯系統遞送至骨胳肌或心肌可恢復完全功能性或部分功能性蛋白質之表現。基於CRISPR/Cas之基因編輯系統由於其較高成功率及有效的基因修飾而具有高級基因組編輯之優勢。

【0103】 該方法可包括投與基於CRISPR/Cas之基因編輯系統，諸如投與Cas蛋白質或融合蛋白，編碼該Cas蛋白質或融合蛋白之核苷酸序列及/或至少一個由SEQ ID NO: 19、其互補序列、其變體或其片段或SEQ ID NO: 20、其互補序列、其變體或其片段編碼或對應於其之gRNA或靶向SEQ ID NO: 17或SEQ ID NO: 18之核酸序列的gRNA、其變體或其片段

【0104】 使用本發明之基因構築體(例如載體)或包含其之組合物將基於CRISPR/Cas之基因編輯系統遞送至目標肌肉，例如可使用修復模板或供體DNA來恢復完全功能性或部分功能性蛋白質之表現，該修復模板或供體DNA可替換整個基因或含有突變之區域。基於CRISPR/Cas之基因編輯系統可用於在靶向基因組基因座處引入位點特異性雙鏈斷裂。當基於CRISPR/Cas之基因編輯系統結合至目標DNA序列時，產生位點特異性雙鏈斷裂，由此允許目標DNA之裂解。此DNA裂解可刺激天然DNA修復機構，其可產生兩種可能的修復路徑中之一者：例如同源定向修復(HDR)或非同源末端接合(NHEJ)路徑。

【0105】 本發明基於CRISPR/Cas之基因編輯系統可涉及使用基於同源定向修復或核酸酶介導之非同源末端接合(NHEJ)的校正方法，其使得能夠在可能無法進行同源重組或基於選擇之基因校正的增殖受限型原代細胞株中進行有效校正。此策略對基於活性CRISPR/Cas之基因編輯系統的快速且穩固組裝與用於治療基因疾病之有效的基因編輯方法進行整合，該等基因疾病由引起框移、過早終止密碼子、異常剪接供體位點或異常剪接受體位點之非必需編碼區域的突變引起。

【0106】 可經由無模板NHEJ介導之DNA修復來恢復來自內源性突變基因之蛋白質表現。相比於靶向目標基因RNA之短暫方法，藉由短暫地表現基於CRISPR/Cas之基因編輯系統校正基因組中之目標基因閱讀框架可永久地恢復藉由每個經修飾細胞及所有其子代之目標基因表現。在某些實施例中，NHEJ為核酸酶介導之NHEJ (在某些實施例中指代起始Cas分子之NHEJ)切斷雙鏈DNA。該方法可包含向個體之骨胳肌或心肌投與本發明之基因構築體(例如載體)或包含其之組合物以用於骨胳肌或心肌之

基因組編輯。

【0107】 核酸酶介導之NHEJ基因校正可校正突變目標基因且具有相比於HDR路徑之若干潛在優勢。舉例而言，NHEJ不需要供體模板，該模板可引起非特異性插入型突變誘發。相比於HDR，NHEJ在細胞週期之所有階段中有效運行，且因此可有效地用於循環及有絲分裂後細胞，諸如肌纖維。此為跳過基於寡核苷酸之外顯子或藥理性強制通讀終止密碼子提供了一種穩固、持久的基因恢復替代方案，且理論上可能只需要一種藥物治療。使用基於CRISPR/Cas之基因編輯系統之基於NHEJ之基因校正以及包括大範圍核酸酶及鋅指核酸酶之其他工程改造的核酸酶可與除本文所述之質體電穿孔方法之外的基於細胞及基因療法的其他現有活體外及活體內平台組合。舉例而言，藉由基於mRNA之基因轉移或以經純化細胞可滲透蛋白質的形式遞送基於CRISPR/Cas之基因編輯系統可實現無DNA基因組編輯方法，其將避免出現插入型突變誘發之任何可能性。

【0108】 最近，用於同源依賴性靶向整合(HITI)之CRISPR/Cas9策略的AAV遞送已產生活體內神經元之基因組編輯。參見Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R.等人。經由CRISPR/Cas9之活體內基因組編輯介導了同源依賴性靶向整合。Nature 540, 144-149 (2016)。本文描述用於校正DMD之基於AAV之HITI介導的基因編輯療法。舉例而言，此類AAV CRISPR/Cas9遞送系統可用於在DMD之人類化動物模型中提供有效及功能性的校正。

5. 醫藥組合物

【0109】 基於CRISPR/Cas之基因編輯系統可呈醫藥組合物形式。醫藥組合物可包含約1 ng至約10 mg DNA，其編碼基於CRISPR/Cas之基

因編輯系統。根據待使用之投與模式調配本文所詳述之醫藥組合物。在醫藥組合物為可注射醫藥組合物之情況下，其為無菌的、無熱原質的及無微粒的。較佳使用等張調配物。一般而言，等張性添加劑可包括氯化鈉、右旋糖、甘露醇、山梨糖醇及乳糖。在一些情況下，等張溶液，諸如磷酸鹽緩衝生理鹽水為較佳的。穩定劑包括明膠及白蛋白。在一些實施例中，向調配物中添加血管收縮劑。

【0110】 含有基於CRISPR/Cas之基因編輯系統的醫藥組合物可進一步包含醫藥學上可接受之賦形劑。醫藥學上可接受之賦形劑可為功能性分子，其呈媒劑、佐劑、載劑或稀釋劑形式。醫藥學上可接受之賦形劑可為轉染促進劑，其可包括表面活性劑，諸如免疫刺激性複合物(ISCOMS)、弗氏不完全佐劑(Freunds incomplete adjuvant)、包括單磷醯基脂質A之LPS類似物、胞壁醯基肽、醌類似物、囊泡(諸如角鯊烯及角鯊烯)、玻尿酸、脂質、脂質體、鈣離子、病毒蛋白質、聚陰離子、聚陽離子或奈米粒子，或其他已知之轉染促進劑。

【0111】 轉染促進劑為聚陰離子、聚陽離子，包括聚-L-麩胺酸酯(LGS)或脂質。轉染促進劑為聚-L-麩胺酸酯，且更佳地，聚-L-麩胺酸酯以低於6 mg/ml之濃度存在於含有基於CRISPR/Cas之基因編輯系統的醫藥組合物中。轉染促進劑亦可包括表面活性劑，諸如免疫刺激性複合物(ISCOMS)、弗氏不完全佐劑、包括單磷醯基脂質A之LPS類似物、胞壁醯基肽、醌類似物及囊泡(諸如角鯊烯及角鯊烯)及玻尿酸亦可與基因構築體結合投與。在一些實施例中，編碼基於CRISPR/Cas之基因編輯系統的DNA載體亦可包括轉染促進劑，諸如脂質、脂質體(包括此項技術中已知之脂質體或其他脂質體，如DNA-脂質體混合物(參見例如W09324640))、

鈣離子、病毒蛋白質、聚陰離子、聚陽離子或奈米粒子或其他已知轉染促進劑。較佳地，轉染促進劑為聚陰離子、聚陽離子，包括聚-L-麩胺酸(LGS)或脂質。

6. 遞送方法

【0112】本文提供一種用於遞送用以提供基因構築體之基於CRISPR/Cas之基因編輯系統的醫藥調配物及/或基於CRISPR/Cas之基因編輯系統的蛋白質的方法。基於CRISPR/Cas之基因編輯系統的遞送可為基於CRISPR/Cas之基因編輯系統之轉染或電穿孔，如在細胞中表現且遞送至細胞表面的一或多個核酸分子。可將基於CRISPR/Cas之基因編輯系統蛋白質遞送至細胞。核酸分子可使用BioRad Gene Pulser Xcell或Amaxa Nucleofector IIb裝置或其他電穿孔裝置進行電穿孔。可使用若干不同緩衝劑，包括BioRad電穿孔溶液，Sigma磷酸鹽緩衝生理鹽水產品D8537號(PBS)、Invitrogen OptiMEM I (OM)或Amaxa Nucleofector溶液V (N.V.)。轉染可包括轉染試劑，諸如脂染胺2000。

【0113】可在具有及不具有活體內電穿孔、脂質體介導、奈米粒子促進及/或重組載體之情況下藉由DNA注射(亦稱為DNA疫苗接種)將編碼基於CRISPR/Cas之基因編輯系統蛋白質的載體遞送至哺乳動物。重組載體可藉由任何病毒模式遞送。病毒模式可為重組慢病毒、重組腺病毒及/或重組腺相關病毒。

【0114】可將編碼基於CRISPR/Cas之基因編輯系統蛋白質的核苷酸引入細胞中以誘導目標基因之基因表現。舉例而言，可將編碼基於CRISPR/Cas之基因編輯系統的一或多個核苷酸序列朝向目標基因引入哺乳動物細胞中。在將基於CRISPR/Cas之基因編輯系統遞送至細胞且隨即

將載體遞送至哺乳動物之後，經轉染細胞將表現基於CRISPR/Cas之基因編輯系統。可向哺乳動物投與基於CRISPR/Cas之基因編輯系統來誘導或調節哺乳動物之目標基因的基因表現。哺乳動物可為人類、非人類靈長類動物、母牛、豬、綿羊、山羊、羚羊、野牛、水牛、牛類、鹿、刺蝟、食蟻獸、鴨嘴獸、象、駱馬、羊駝、小鼠、大鼠或雞，且較佳為人類、母牛、豬或雞。

【0115】 在將本發明之基因構築體或組合物遞送至組織且隨即將載體遞送至哺乳動物之細胞後，經轉染細胞將表現gRNA分子及Cas9分子。可向哺乳動物投與基因構築體或組合物以改變基因表現或再工程改造或改變基因組。舉例而言，可向哺乳動物投與基因構築體或組合物以恢復哺乳動物之肌縮蛋白功能。哺乳動物可為人類、非人類靈長類動物、母牛、豬、綿羊、山羊、羚羊、野牛、水牛、牛類、鹿、刺蝟、食蟻獸、鴨嘴獸、象、駱馬、羊駝、小鼠、大鼠或雞，且較佳為人類、母牛、豬或雞。

【0116】 可在具有及不具有活體內電穿孔、脂質體介導、奈米粒子促進及/或重組載體之情況下藉由DNA注射(亦稱為DNA疫苗接種)將編碼gRNA分子及Cas9分子之基因構築體(例如載體)遞送至哺乳動物。重組載體可藉由任何病毒模式進行遞送。病毒模式可為重組慢病毒、重組腺病毒及/或重組腺相關病毒。

【0117】 可將本發明之基因構築體(例如載體)或包含其之組合物引入細胞中，以便以基因方式恢復肌縮蛋白基因(例如人類肌縮蛋白基因)之肌縮蛋白功能。在某些實施例中，將本發明之基因構築體(例如載體)或包含其之組合物引入來自DMD患者之肌母細胞中。在某些實施例中，將基因構築體(例如載體)或包含其之組合物引入來自DMD患者之纖維母細胞

中，且可用MyoD處理經基因方式校正之纖維母細胞來誘導分化成肌母細胞，該等肌母細胞可植入至個體中，諸如個體之受損肌肉中，以驗證經校正肌縮蛋白質具功能及/或治療個體。經修飾細胞亦可為幹細胞，諸如經誘導多能幹細胞、骨髓衍生之祖細胞、骨胳肌祖細胞、來自DMD患者之人類骨胳肌母細胞、CD 133^{+}細胞、中胚層血管母細胞及經MyoD或Pax7轉導細胞或其他肌原性祖細胞。舉例而言，基於CRISPR/Cas之基因編輯系統可引起經誘導多能幹細胞之神經元或肌原性分化。

7. 投與途徑

【0118】 基於CRISPR/Cas之基因編輯系統及其組合物可藉由不同途徑投與給個體，包括經口、非經腸、舌下、經皮、經直腸、經黏膜、局部、經由吸入、經由頰內投與、胸膜內、靜脈內、動脈內、腹膜內、皮下、肌肉內、鼻內鞘內及關節內或其組合。對於獸醫學用途，組合物可作為適合的可接受之調配物根據正常獸醫學實踐投與。獸醫可容易地確定最適合於特定動物之給藥方案及投與途徑。基於CRISPR/Cas之基因編輯系統及其組合物可藉由傳統注射器、無針注射裝置、「微彈轟擊基因槍(microprojectile bombardment gene gun)」或其他物理方法(諸如電穿孔(「EP」)、「流體動力方法」)或超音波投與。可在具有及不具有活體內電穿孔、脂質體介導、奈米粒子促進、重組載體(諸如重組慢病毒、重組腺病毒及重組腺病毒相關病毒)之情況下藉由包括DNA注射(亦稱為DNA疫苗接種)之若干技術將組合物遞送至哺乳動物。

【0119】 本發明之基因構築體(例如載體)或包含其之組合物可藉由不同途徑投與給個體，包括經口、非經腸、舌下、經皮、經直腸、經黏膜、局部、經由吸入、經由頰內投與、胸膜內、靜脈內、動脈內、腹膜

內、皮下、肌肉內、鼻內鞘內及關節內或其組合。在某些實施例中，將本發明之基因構築體(例如載體)或組合物以肌肉內、靜脈內或其組合形式投與至個體(例如患有DMD之個體)。對於獸醫學用途，本發明之基因構築體(例如載體)或組合物可作為適合的可接受之調配物根據正常獸醫學實踐投與。獸醫可容易地確定最適合於特定動物之給藥方案及投與途徑。該等組合物可藉由傳統注射器、無針注射裝置、「微彈轟擊基因槍」或其他物理方法(諸如電穿孔(「EP」)、「流體動力方法」)或超音波投與。

【0120】 可在具有及不具有活體內電穿孔、脂質體介導、奈米粒子促進、重組載體(諸如重組慢病毒、重組腺病毒及重組腺病毒相關病毒)之情況下藉由包括DNA注射(亦稱為DNA疫苗接種)之若干技術將本發明之基因構築體(例如載體)或組合物遞送至哺乳動物。可將組合物注射至骨骼肌或心肌中。舉例而言，可將組合物注射至脛前肌或尾部。

【0121】 在一些實施例中，藉由以下投與本發明之基因構築體(例如載體)或其組合物：1)尾部靜脈注射(全身性)至成年小鼠中；2)肌肉內注射，例如局部注射至例如成年小鼠的TA或腓腸肌中；3)腹膜內注射至P2小鼠中；或4)面部靜脈注射(全身性)至P2小鼠中。

8. 細胞類型

【0122】 此等遞送方法及/或投與途徑中之任一者可與多種細胞類型一起使用，例如當前研究的基於細胞之DMD療法的彼等細胞類型，包括但不限於永生化肌母細胞，諸如野生型及DMD患者衍生細胞株、原代DMD真皮纖維母細胞、經誘導性多能幹細胞、骨髓衍生之祖細胞、骨胳肌祖細胞、來自DMD患者之人類骨骼肌母細胞、CD 133^{+}細胞、中胚層血管母細胞、心肌細胞、肝細胞、軟骨細胞、間葉細胞祖細胞、造血幹細

胞、平滑肌細胞及經MyoD或Pax7轉導細胞或其他肌原性祖細胞。人類肌原細胞之永生化可用於經基因方式校正之肌原細胞的純系衍生。可對細胞進行活體外修飾以分離及擴增永生化DMD肌母細胞之純系群體，該群體包括經基因方式校正或恢復之肌縮蛋白基因且在基因組之蛋白質編碼區中不含其他核酸酶引入型突變。或者，藉由非病毒或非整合病毒基因轉移或藉由直接遞送含有細胞穿透基元之經純化蛋白質及gRNA而短暫活體內遞送基於CRISPR/Cas之系統可能能夠在極少或沒有外源性DNA整合之風險下原位高度特異性校正及/或恢復。

9. 套組

【0123】 本文提供一種套組，其可用於校正突變型肌縮蛋白基因及/或恢復肌縮蛋白功能。該套組包含用於恢復肌縮蛋白功能之至少由SEQ ID NO: 19之聚核苷酸序列、其互補序列、其變體或其片段編碼的gRNA，或由SEQ ID NO: 20之聚核苷酸序列、其互補序列、其變體或其片段編碼的gRNA，或靶向SEQ ID NO: 17或SEQ ID NO: 18之聚核苷酸序列、其互補序列、其變體或其片段的gRNA及基於CRISPR/Cas之編輯系統的使用說明書。本文亦提供一種套組，其可用於編輯骨胳肌或心肌之肌縮蛋白基因。該套組包含如上文所述之用於骨胳肌或心肌之基因組編輯的基因構築體(例如載體)或包含其之組合物，及該組合物之使用說明書。

【0124】 套組中包括之說明書可附至包裝材料或可作為藥品說明書包括在內。雖然該等說明書通常為書面或印刷材料，但不限於此。本發明涵蓋能夠儲存此等說明書且將其傳達至終端使用者之任何媒介。該等媒介包括但不限於電子儲存媒介(例如磁盤、膠帶、濾筒、晶片)、光學媒介(例如CD ROM)及其類似物。如本文所用，術語「說明書」可包括提供該

等說明書之網際網路站點的地址。

【0125】 用於恢復骨胳肌或心肌之肌縮蛋白功能的基因構築體(例如載體)或包含其之組合物可包括特異性地結合及裂解肌縮蛋白基因之區域的經修飾AAV載體，其包括如上文所述之gRNA分子及Cas蛋白質或融合蛋白。如上文所述，基於CRISPR/Cas之基因編輯系統可包括於套組中，以特異性地結合及靶向突變型肌縮蛋白基因之特定區域，例如外顯子52。

10. 實例

【0126】 參考以下實施例可較好地理解前文，其係出於說明之目的呈現且並不意欲限制本發明之範疇。本發明具有由隨附非限制性實例說明之多個態樣及實施例。

實例1

SaCas9 gRNA設計及HEK293T細胞之hDMD-內含子51目標的篩選

【0127】 靶向hDMD-外顯子52上游之hDMD-內含子51的SaCas9 gRNA經設計具有21 bp間隔序列且選殖至個別表現質體中(**圖3**)。在CMV啟動子(375 ng)下用表現SaCas9之質體及在U6啟動子(125 ng)下用表現個別gRNA之質體轉染24孔盤中之HEK293T細胞。在轉染後3天提取基因組DNA。藉由測量員分析評估編輯效率(**圖4A**及**圖4B**)。陰性對照(NC)不含gRNA。HEK293T基因組DNA中亦觀測到與單核苷酸多態性(SNP)相關的額外條帶。此等條帶對應於基於SNP位置之預期大小(**圖5A**、**圖5B**)。繼續測試gRNA 03、gRNA 06、gRNA 07及gRNA 09之19-23 bp間隔序列。

實例2

SaCas9 gRNA設計及人類8036肌母細胞之hDMD-內含子51目標的篩選

【0128】 基於HEK293T中測試的gRNA的編輯活性，用19-23 bp間

隔序列重新設計頂部gRNA且選殖至個別表現質體中。在人類8036肌母細胞中篩選重新設計的gRNA。在CMV啟動子(10 μg)下用表現SaCas9之質體及在U6啟動子(10 μg)下用表現個別gRNA之質體轉染6孔盤中之人類8036肌母細胞。在電穿孔後3天分離基因組DNA。藉由測量員分析評估編輯效率(**圖6A**、**圖6B**)。陰性對照(NC)不含gRNA。亦藉由肽分析評估編輯效率。選擇gRNA g12、g16及g7，以產生AAV整合載體。

實例3

HEK293T細胞中AAV-HITI供體質體之SaCas9 gRNA篩選

【0129】基於在人類8036肌母細胞中測試的gRNA的編輯活性，將頂部gRNA選殖至AAV-HITI供體質體(gRNA g12、g16及g7)中。在CMV啟動子(375 ng)下用表現SaCas9之質體及在U6啟動子(125 ng)下用表現個別gRNA之質體或在U6啟動子(125 ng)下用表現個別gRNA之AAV-HITI供體質體轉染24孔盤中之HEK293T細胞。在轉染後3天提取基因組DNA。藉由測量員分析評估編輯效率(**圖7**)。基於表現個別gRNA之AAV-HITI供體質體的編輯活性，未來實驗中使用g7及g12供體，以產生AAV-HITI供體質體。

實例4

hDMD-外顯子52的活體外HITI介導之整合

【0130】自hDMDΔ52/mdx小鼠分離原代肌母細胞。用表現個別gRNA之AAV-CMV-Cas9質體(10 μg)及AAV-U6-gRNA-Ex52供體質體(10 μg)電穿孔6孔盤中之此等細胞。在電穿孔後6天提取基因組DNA。使用巢式PCR偵測HITI介導之hDMD-外顯子52整合(**圖8A**)。切除gRNA12供體樣品中之方形條帶且送至桑格測序(**圖8B**)，以證實在靶向位點處之

hDMD-外顯子52供體的整合。

實例5

hDMDΔ52/mdx小鼠模型中hDMD-外顯子52的活體內HITI介導之整合

【0131】 **圖9**中展示了用於證實活體內編輯、確定最佳gRNA/供體序列組合及確定AAV-Cas9與AAV供體質體之最佳比率的實驗之示意圖。經由局部肌肉內注射向雄性6至8週齡hDMDΔ52/mdx小鼠之脛前(TA)肌中注射AAV-Cas9及AAV-HITI供體。在注射後4週，收集TA肌用於處理，以評估HITI介導之編輯。注射有PBS之小鼠充當陰性對照；N=4。

【0132】 用靶向切割位點下游之引物(**圖10A**)及用靶向切割位點上游之引物(**圖10B**)檢查hDMDΔ52/mdx小鼠之基因組DNA的靶向Ex52插入。自TA肌提取基因組DNA。PCR分析證實了在靶向位點處存在Ex52插入。

【0133】 檢測經處理hDMDΔ52/mdx小鼠之mRNA的靶向Ex52插入(**圖11**)。自TA肌提取總RNA且用於產生cDNA。PCR分析經由RNA轉錄物中之剪接證實了Ex52插入的存在。

實例6

經處理hDMDΔ52/mdx小鼠之肌縮蛋白質恢復

【0134】 自TA肌提取蛋白質且用於西方墨點分析。用25 μg總蛋白質裝載PBS及經處理TA肌。為了充當陽性對照，用3.125 μg總蛋白質裝載hDMD/mdx TA肌。用抗肌縮蛋白(純系2c6；MANDYS106)或抗GAPDH(純系14C10)對膜進行染色。西方墨點分析證實了經處理小鼠的蛋白質恢復(**圖12**)。

實例7

經編輯hDMDΔ52/mdx小鼠的AAV-ITR序列整合之深度測序定量

【0135】 **圖13**為經編輯小鼠之外顯子52基因組整合的伊路米那深度測序定量的結果。自TA肌提取基因組DNA。為了測序分析公平，使用Nextera Tn5轉位子標記基因組DNA。為了富集靶向序列，使用內含子51目標位點上游之基因組特異性引物完成PCR且將其與對由轉位子插入之標籤序列具有特異性的反向引物配對。使用第二PCR添加實驗性條碼及伊路米那銜接子序列。藉由第二代測序法偵測ITR整合。使用回旋分析偵測在基因組特異性引物與內含子51目標位點之間的與AAV載體匹配之ITR序列及與基因組DNA序列匹配之基因組特異性序列的存在。

實例8

經編輯hDMDΔ52/mdx小鼠之mRNA的外顯子52插入之PacBio測序定量

【0136】 自TA肌提取總RNA且用於產生cDNA (**圖14A**)。為了富集靶向序列，使用外顯子45及外顯子69中之引物完成PCR。使用第二PCR添加實驗性條碼及PacBio銜接子序列。藉由PacBio測序來偵測外顯子52插入(**圖14B**)。定量3'-外顯子51與5'-外顯子53序列之間118nt的讀段。將此等序列與外顯子52供體進行比對，以證實預期編輯。在外顯子51與外顯子53之間的118 bp的測序讀段與外顯子52序列匹配。

* * *

【0137】 出於完整性原因，各種態樣在以下編號之條項中陳述：

【0138】 條項1. 一種基於CRISPR/Cas之基因組編輯系統，其包含一或多種編碼組合物之載體，該組合物包含： (a) 引導RNA (gRNA)，其靶向突變肌縮蛋白基因之片段； (b) Cas蛋白質或包含該Cas蛋白質之融合蛋白；及(c) 供體序列，其包含野生型肌縮蛋白基因之片段。

【0139】 條項2. 一種基於CRISPR/Cas之基因組編輯系統，其包含： (a) 引導RNA (gRNA)，其靶向突變肌縮蛋白基因之片段； (b) Cas蛋白質或包含該Cas蛋白質之融合蛋白；及(c) 供體序列，其包含野生型肌縮蛋白基因之片段。

【0140】 條項3. 如條項1或2之系統，其中該野生型肌縮蛋白基因之該片段由兩個gRNA間隔序列及/或PAM序列側接。

【0141】 條項4. 如條項1至3中任一項之系統，其中該gRNA靶向與該突變肌縮蛋白基因之外顯子並列的內含子，且其中該外顯子選自該突變肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59及61-66。

【0142】 條項5. 如條項1至3中任一項之系統，其中該供體序列包含該野生型肌縮蛋白基因之外顯子或其功能性等效物，且其中該外顯子選自該野生型肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59、61-66。

【0143】 條項6. 如條項4之系統，其中該突變肌縮蛋白基因之該外顯子經突變或自該肌縮蛋白基因至少部分地缺失，或其中該突變肌縮蛋白基因之該外顯子缺失且該內含子與相應野生型肌縮蛋白基因中之缺失的外顯子的位置並列。

【0144】 條項7. 如條項4或5之系統，其中該外顯子為外顯子52。

【0145】 條項8. 如條項1至7中任一項之系統，其中該gRNA結合及靶向包含以下之聚核苷酸序列： a) SEQ ID NO: 17或SEQ ID NO: 18；b) SEQ ID NO: 17或SEQ ID NO: 18之片段； c) SEQ ID NO: 17或SEQ ID NO: 18之互補序列或其片段； d) 與SEQ ID NO: 17或SEQ ID NO: 18或其互補序列實質上一致的核酸；或e) 在嚴格條件下雜交至SEQ ID NO: 17

或SEQ ID NO: 18、其互補序列或與其實質上一致之序列的核酸。

【0146】 條項9. 如條項1至8中任一項之系統，其中該gRNA包含聚核苷酸序列SEQ ID NO: 19或SEQ ID NO: 20或其變體或由其編碼。

【0147】 條項10. 如條項1至9中任一項之系統，其中該Cas蛋白質為化膿性鏈球菌(*Streptococcus pyogenes*)Cas9蛋白質或金黃色葡萄球菌(*Staphylococcus aureus*)Cas9蛋白質。

【0148】 條項11. 如條項1至10中任一項之系統，其中該Cas蛋白質包含胺基酸序列SEQ ID NO: 1、2、3或4。

【0149】 條項12. 如條項3至11中任一項之系統，其中該兩個gRNA間隔序列獨立地包含選自SEQ ID NO: 5-8及25-45之序列。

【0150】 條項13. 如條項12之系統，其中該兩個gRNA間隔序列為相同的。

【0151】 條項14. 如條項12之系統，其中該兩個gRNA間隔序列為不同的。

【0152】 條項15. 如條項3至14中任一項之系統，其中該兩個gRNA間隔序列中之至少一者包含序列SEQ ID NO: 25或SEQ ID NO: 26。

【0153】 條項16. 如條項1至15中任一項之系統，其中該供體序列包含SEQ ID NO: 21或SEQ ID NO: 22之聚核苷酸。

【0154】 條項17. 如條項1及3至16中任一項之系統，其中該載體為病毒載體。

【0155】 條項18. 如條項17之系統，其中該載體為腺相關病毒(AAV)載體。

【0156】 條項19. 如條項18之系統，其中該AAV載體為AAV1、

AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV-10、AAV-11、AAV-12、AAV-13或AAVrh.74載體。

【0157】 條項20. 如條項18之系統，其中該一或多個載體中之一者包含聚核苷酸序列SEQ ID NO: 23或24。

【0158】 條項21. 如條項1至20中任一項之系統，其中gRNA與供體序列之間的莫耳比為1:1、或1:15、或5:1至1:10、或1:1至1:5。

【0159】 條項22. 一種重組聚核苷酸，其編碼包含野生型肌縮蛋白基因之片段或其功能性等效物的供體序列，且其中該片段或其功能性等效物由兩個gRNA間隔序列側接。

【0160】 條項23. 如條項22之重組聚核苷酸，其中該供體序列包含該肌縮蛋白基因之該外顯子，且其中該外顯子選自外顯子1-8、10、11、12、14、16-22、43-59、61-66。

【0161】 條項24. 如條項22或23之重組聚核苷酸，其中該重組聚核苷酸包含序列SEQ ID NO: 23或24。

【0162】 條項25. 一種載體，其包含如條項22至24中任一項所述之重組聚核苷酸。

【0163】 條項26. 如條項25之載體，其中該載體包含驅動該重組聚核苷酸表現之異源啟動子。

【0164】 條項27. 一種細胞，其包含如條項22至24中任一項所述之重組聚核苷酸或如條項25或26所述之載體。

【0165】 條項28. 一種用於恢復具有突變肌縮蛋白基因之細胞的肌縮蛋白功能的組合物，該組合物包含如條項1至21中任一項所述之系統、如條項22至24中任一項所述之重組聚核苷酸或如條項25至26所述之載

體。

【0166】 條項29. 一種套組，其包含如條項1至21中任一項所述之系統、如條項22至24中任一項所述之重組聚核苷酸或如條項25或26所述之載體或如條項28所述之組合物。

【0167】 條項30. 一種用於恢復具有突變肌縮蛋白基因之細胞或個體的肌縮蛋白功能的方法，該方法包含使該細胞或該個體與如條項1至21中任一項所述之系統、如條項22至24中任一項所述之重組聚核苷酸或如條項25或26所述之載體或如條項28所述之組合物接觸。

【0168】 條項31. 如條項30之方法，其中該肌縮蛋白功能藉由插入該野生型肌縮蛋白基因之外顯子52來恢復。

【0169】 條項32. 如條項30或31之方法，其中該個體患有杜興氏肌肉萎縮症(Duchenne Muscular Dystrophy)。

【0170】 條項33. 一種用於恢復具有由一或多個缺失或突變型外顯子引起的破壞的肌縮蛋白基因之細胞或個體的肌縮蛋白功能的方法，該方法包含使該細胞或該個體與如條項1至21中任一項所述之系統、如條項22至24中任一項所述之重組聚核苷酸或如條項25或26所述之載體或如條項28所述之組合物接觸。

【0171】 條項34. 如條項33之方法，其中肌縮蛋白功能藉由插入對應於該一或多個缺失或突變型外顯子之肌縮蛋白基因的一或多個野生型外顯子來恢復。

【0172】 條項35. 如條項34之方法，其中該等缺失或突變型外顯子中之一者為外顯子52。

序列

化膿性鏈球菌Cas 9 (SEQ ID NO: 1)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

金黃色葡萄球菌Cas9分子 (SEQ ID NO: 2)

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKL
LFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISR
NSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRT
YYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEK
FQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQ
IAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR
LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKM
INEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIP
RSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER
DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKH
HAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYK
YSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKL
KLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLK
PYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYR
VIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQII
KKG

化膿性鏈球菌 Cas 9 (具有D10A) (SEQ ID NO: 3)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE

LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

*化膿性鏈球菌*Cas 9 (具有D10A、H849A) (SEQ ID NO: 4)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDA
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI
ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV
KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

PAM (SEQ ID NO: 9)

ATTCCT

PAM (SEQ ID NO: 10)

NGG

PAM (SEQ ID NO: 11)

NNNRRT

PAM (SEQ ID NO: 12)

NNGRR (R=A或G)

PAM (SEQ ID NO: 13)

NNGRRN (R=A或G)

PAM (SEQ ID NO: 14)

NNGRRT (R=A或G)

PAM (SEQ ID NO: 15)

NNGRRV (R=A或G；V=A、C或G)

PAM (SEQ ID NO: 16)

NGA

gRNA7之目標 (SEQ ID NO: 17)

TCATTTATAATACAGGGGAAT

gRNA12之目標(SEQ ID NO: 18)

TTAAGTAATCCGAGGTACTC

gRNA7 (包括目標序列及骨架)(SEQ ID NO: 19)

TCATTTATAATACAGGGGAATGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGT
GTTTATCTCGTCAACTTGTTGGCGAGA

gRNA12 (包括目標序列及骨架)(SEQ ID NO: 20)

TTAAGTAATCCGAGGTACTCGTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTG
TTTATCTCGTCAACTTGTTGGCGAGA

外顯子 52 (SEQ ID NO: 21)

GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGA
AAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA

具有一些內含子之外顯子52(**SEQ ID NO: 22**)

GTTAAATTGTTTTCTATAAACCCTTATACAGTAACATCTTTTTTATTTCTAAAAGTGTTTTGGCTGGTCT
CACAATTGTACTTTACTTTGTATTATGTAAAAGGAATACACAACGCTGAAGAACCCTGATACTAAGGGAT
ATTTGTTCTTACAGGCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGC
TGCCCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAAGTAAGTTT
TTTAACAAGCATGGGACACACAAAGCAAGATGCATGACAAGTTTCAATAAAAACTTAAGTTCATATATCC
CCCTCACATTTATAAAAATAATGTGAAATAATTGTAAATGATAACAATTGTGCTGAGATTTTCAGTCCAT
AATGTTACCTTTTAATAAATGAATGTAATTCCATTGAATAGAAGAAATAC

gRNA7的AAV (SEQ ID NO: 23)

具有gRNA7之外顯子52供體序列的AAV基因組：

AAV ITR、U6啟動子、gRNA7間隔序列、PAM、SaCas9 gRNA骨架、供體序列、外顯子52

GGGGGGGGGGGGGGGGGGGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTG
*GCCAACTCCATCACTAGGG*GTTCCTCAGATCTGAATTCGGTACCTTCCT*AGGGCCTATTTCCCATGATTCCT*
TCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA
GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGAC
*TATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGAC*GAAACAC
CG*TCATTTATAATACAGGGGAAT*GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAAT
GCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTG
GCGTTTAAACATTCCT*ATTCCCCTGTATTATAAATGA*GTTAAATTGTTTTCTATAAACCCTTATACAG
TAACATCTTTTTTATTTCTAAAAGTGTTTTGGCTGGTCTCACAATTGTACTTTACTTTGTATTATG
TAAAAGGAATACACAACGCTGAAGAACCCTGATACTAAGGGATATTTGTTCTTACAGGCAACAAT
GCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAA
AACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAAGTAAGTTTTTTAACAAGC
ATGGGACACACAAAGCAAGATGCATGACAAGTTTCAATAAAAACTTAAGTTCATATATCCCCCTC
ACATTTATAAAAATAATGTGAAATAATTGTAAATGATAACAATTGTGCTGAGATTTTCAGTCCAT
AATGTTACCTTTTAATAAATGAATGTAATTCCATTGAATAGAAGAAATACATTCCT*ATTCCCCTGTA*
*TTATAAATGA*GCTAGCTCGGAGAGACGACATCTAGAATTAAACTGTCACTATCGATTACTAATTTTTTG
CTCATAATAGAAGCAGCGATTAAAGGAATAGAATCAACAGTTCCAGTAACATCTCTTAGTGCATACAT
TTTTTTATCAGCAGGAACAATATCATCTGACTGACCTGTGATGCTCATTCCAACTTCATTAATTGTTTTA
ATGAATTTTTCTTTAGATAATTCACTTGTTCAACCTTTACATGACTCTAATTTATCAATAGTACCACCAG
TTACTCCAAGTCCTCTACCAGAAAGTTTACAAACCTTTACTCCATAACTTGCAACTAACGGACTATATA
CTAAACTTGTTTTGTCTCCAACTCCGCCAGTTGAATGCTTATCAGCTTTTAAACCTGTAACCTCACTGAC
ATCATAAACATATCCTGATTCAACATAAGATTGGGTTAATGCTAAAGTTTCTGCTTTGGTCATTCCATT
AAAATAAACCGCCATAGCAAAAGCAGCCATTTGATAATCTGTTACATTATTTTTAACGTAACTGTTTAT
CAATCATTTAATTTCTTCAGCTGATAATTCTATTGAATGTTTCTTTTTTTCTATAATTTCACTAAAACTGT
AGTTCATAAGTCTCCTTTTGTAAGAGTGCACAATATTTACACCATTACTCTTTCTACTATATTATAATAG
AATAGACATATAAAAAACATAAGGAGTACAAATGGTTTTTGATAAAAATAACAAAGTTTATAGTGAAT
GAATAAATAGCCAAAAATTGGATGATGAGTTGAAAAGCCTTTTAGTAAATGCTACTGATGATGAATTG
CATGCAGCATTTGAAGGAATAGAGTTAGAATTTGGAACAGCAGGTATAAGAGGTATTCTTGGAGCAG
GACCTGGAAGATTTAACGTTTACACTGTTAAAAAAGTTACTATTGCATTTGCAGAATTATTAAAACAA
AATTACCCAAATAGGTTGAATGATGGAATAGTTGTTGGTCATGATAACCGTCATAATTCTAAACAGTTT
GCAAAAGTTGTAGCCGAAGTTTTATCAAGCTTGTGAAATAGCTGTTGAAGCTGGATTAGAATTTGTTA
AAACATCAACAGGATTTTCAAAATCAGGTGCAACATTTGAAGATGTTAAACTAATGAAGTCAGTTGTT
AAAGACAATGCTTTAGTTAAAGCAGCTGGTGGAGTTAGAACATTTGAAGATGCTCAAAAAATGATTGA
AGCAGGAGCTGACCGCTTAGGAACAAGTGGTGGAGTAGCTATTATTAAAGGTGAAGAAAACAACGCG
AGTTACTAAAACTAGCGTTTTTTTATTTTGCTCATTTTTATTAAAAGTTTGCAAAAAGGAACATAAAAA
TTCTAATTATTGATACTAAAGTTATTAAAAAGAAGATTTTGGTTGATTTTATAAAGGTCATAGAATATA
ATATTTTAGCATGTGTATTTTGTGTGCTCATTTACAACCGTCTCGCGGCCGCGGGGATCCAGACATGAT
AAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAA
TTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCA
TTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAGTCGACCTCGAGCAGTGTGGTTT
TGCAAGAGGAAGCAAAAAGCCTCTCCACCCAGGCCTGGAATGTTTCCACCCAAGTCGAAGGCAGTGT
GGTTTTGCAAGAGGAAGCAAAAAGCCTCTCCACCCAGGCCTGGAATGTTTCCACCCAATGTCGAGCAA
CCCCGCCCAGCGTCTTGTCATTGGCGAATTCGAACACGCAGATGCAGTCGGGGCGGCGCGGTCCCAGG
TCCACTTCGCATATTAAGGTGACGCGTGTGGCCTCGAACACCGAGCGACCCTGCAGCCAATATGGGAT
CGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTAT
GACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCC

GGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTAT
CGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGAC
TGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGT
ATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACC
AAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTG
GACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACG
GCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTT
TCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCG
TGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCC
CGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGGATCCGTCGACTAGAGC
TCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCC
TTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTG
AGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA
ATAGCAGGCATGCTGGGGAGAGATCTAGGAAC*CCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGC*
TCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTG
AGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCCCCCCCCCCCCCCCCC

gRNA12的AAV (SEQ ID NO: 24)

具有gRNA12之外顯子52供體序列的AAV基因組：

AAV ITR、*U6*啟動子、gRNA12間隔序列、PAM、SaCas9 gRNA骨架、

供體序列、外顯子52

GGGGGGGGGGGGGGGGGGGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCA
AAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTG
*GCCAACTCCATCACTAGGG*GTTCCTCAGATCTGAATTCGGTACCTTCCT*AGGGCCTATTTCCCATGATTCCT*
TCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA
GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGAC
*TATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGAC*GAAACAC
CG*TTAAGTAATCCGAGGTACTC*GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAAT
GCCGTGTTTATCTCGTCAACTTGTTGGCGAGATTTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTG
GCGTTTAAACATTCCT*GAGTACCTCGGATTACTTAA*GTTAAATTGTTTTCTATAAACCCTTATACAGT
AACATCTTTTTTATTTCTAAAAGTGTTTTGGCTGGTCTCACAATTGTACTTTACTTTGTATTATGT
AAAAGGAATACACAACGCTGAAGAACCCTGATACTAAGGGATATTTGTTCTTACAGGCAACAAT
GCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAA
AACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAAGTAAGTTTTTTAACAAGC
ATGGGACACACAAAGCAAGATGCATGACAAGTTTCAATAAAAACTTAAGTTCATATATCCCCCTC
ACATTTATAAAAATAATGTGAAATAATTGTAAATGATAACAATTGTGCTGAGATTTTCAGTCCAT
AATGTTACCTTTTAATAAATGAATGTAATTCCATTGAATAGAAGAAATACATTCCT*GAGTACCTCG*
*GATTACTTAA*GCTAGCTCGGAGAGACGACATCTAGAATTAAACTGTCACTATCGATTACTAATTTTTTG
CTCATAATAGAAGCAGCGATTAAAGGAATAGAATCAACAGTTCCAGTAACATCTCTTAGTGCATACAT
TTTTTTATCAGCAGGAACAATATCATCTGACTGACCTGTGATGCTCATTCCAACTTCATTAATTGTTTTA
ATGAATTTTTCTTTAGATAATTCACTTGTTCAACCTTTACATGACTCTAATTTATCAATAGTACCACCAG
TTACTCCAAGTCCTCTACCAGAAAGTTTACAAACCTTTACTCCATAACTTGCAACTAACGGACTATATA
CTAAACTTGTTTTGTCTCCAACTCCGCCAGTTGAATGCTTATCAGCTTTTAAACCTGTAACCTCACTGAC
ATCATAAACATATCCTGATTCAACATAAGATTGGGTTAATGCTAAAGTTTCTGCTTTGGTCATTCCATT
AAAATAAACCGCCATAGCAAAAGCAGCCATTTGATAATCTGTTACATTATTTTTAACGTAACTGTTTAT
CAATCATTTAATTTCTTCAGCTGATAATTCTATTGAATGTTTCTTTTTTTCTATAATTTCACTAAAACTGT
AGTTCATAAGTCTCCTTTTGTAAGAGTGCACAATATTTACACCATTACTCTTTCTACTATATTATAATAG
AATAGACATATAAAAAACATAAGGAGTACAAATGGTTTTTGATAAAAATAACAAAGTTTATAGTGAAT
GAATAAATAGCCAAAAATTGGATGATGAGTTGAAAAGCCTTTTAGTAAATGCTACTGATGATGAATTG
CATGCAGCATTTGAAGGAATAGAGTTAGAATTTGGAACAGCAGGTATAAGAGGTATTCTTGGAGCAG
GACCTGGAAGATTTAACGTTTACACTGTTAAAAAAGTTACTATTGCATTTGCAGAATTATTAAAACAA

AATTACCCAAATAGGTTGAATGATGGAATAGTTGTTGGTCATGATAACCGTCATAATTCTAAACAGTTT
GCAAAAGTTGTAGCCGAAGTTTTATCAAGCTTGTGAAATAGCTGTTGAAGCTGGATTAGAATTTGTTA
AAACATCAACAGGATTTTCAAAATCAGGTGCAACATTTGAAGATGTTAAACTAATGAAGTCAGTTGTT
AAAGACAATGCTTTAGTTAAAGCAGCTGGTGGAGTTAGAACATTTGAAGATGCTCAAAAAATGATTGA
AGCAGGAGCTGACCGCTTAGGAACAAGTGGTGGAGTAGCTATTATTAAAGGTGAAGAAAACAACGCG
AGTTACTAAAACTAGCGTTTTTTTATTTTGCTCATTTTTATTAAAAGTTTGCAAAAAGGAACATAAAAA
TTCTAATTATTGATACTAAAGTTATTAAAAAGAAGATTTTGGTTGATTTTATAAAGGTCATAGAATATA
ATATTTTAGCATGTGTATTTTGTGTGCTCATTTACAACCGTCTCGCGGCCGCGGGGATCCAGACATGAT
AAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAA
TTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCA
TTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAGTCGACCTCGAGCAGTGTGGTTT
TGCAAGAGGAAGCAAAAAGCCTCTCCACCCAGGCCTGGAATGTTTCCACCCAAGTCGAAGGCAGTGT
GGTTTTGCAAGAGGAAGCAAAAAGCCTCTCCACCCAGGCCTGGAATGTTTCCACCCAATGTCGAGCAA
CCCCGCCCAGCGTCTTGTCATTGGCGAATTCGAACACGCAGATGCAGTCGGGGCGGCGCGGTCCCAGG
TCCACTTCGCATATTAAGGTGACGCGTGTGGCCTCGAACACCGAGCGACCCTGCAGCCAATATGGGAT
CGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTAT
GACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCC
GGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTAT
CGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGAC
TGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGT
ATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACC
AAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTG
GACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACG
GCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTT
TCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCG
TGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCC
CGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGGATCCGTCGACTAGAGC
TCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCC
TTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTG
AGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA
ATAGCAGGCATGCTGGGGAGAGATCTAGGAAC*CCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGC*
TCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTG
AGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCCCCCCCCCCCCCCCC

SEQ ID NO: 25 gRNA7間隔序列

ATTCCCCTGTATTATAAATGA

SEQ ID NO: 26: gRNA12間隔序列

GAGTACCTCGGATTACTTAA

引導#	目標/原始 gRNA	Cas9	間隔序列	(nt)
第一輪篩選				
gAP1	Dyst (內含子 51)	SaCas9	CTTTACTTTGTATTATGTAAA (SEQ ID NO: 27)	21
gAP2	Dyst (內含子 51)	SaCas9	TTTGAAATATTTTTGATATCT (SEQ ID NO: 28)	21
gAP3	Dyst (內含子 51)	SaCas9	TTTAAGTAATCCGAGGTACTC (SEQ ID NO: 29)	21
gAP4	Dyst (內含子 51)	SaCas9	TTTAAATACATTGTCGTAATT (SEQ ID NO: 30)	21
gAP5	Dyst (內含子 51)	SaCas9	TACCTTAATTTTGACGTCACA (SEQ ID NO: 31)	21

gAP6	Dyst (內含子51)	SaCas9	ATTTGACAGGTGAGAAATCTC (SEQ ID NO: 32)	21
gAP7	Dyst (內含子51)	SaCas9	TCATTTATAATACAGGGGAAT (SEQ ID NO: 33)	21
gAP8	Dyst (內含子51)	SaCas9	TTAAAGTCATTTATAATACAG (SEQ ID NO: 34)	21
gAP9	Dyst (內含子51)	SaCas9	AAATAGACACTGAAGAAAGGG (SEQ ID NO: 35)	21
gAP10	Dyst (內含子51)	SaCas9	CCCCAATTAAAATAAAATTTA (SEQ ID NO: 36)	21
第二輪篩選				
gAP11	g3	SaCas9	TAAGTAATCCGAGGTACTC (SEQ ID NO: 37)	19
gAP12	g3	SaCas9	TTAAGTAATCCGAGGTACTC (SEQ ID NO: 38)	20
gAP13	g3	SaCas9	GTTTAAGTAATCCGAGGTACTC (SEQ ID NO: 39)	22
gAP14	g3	SaCas9	GGTTTAAGTAATCCGAGGTACTC (SEQ ID NO: 40)	23
gAP15	g6	SaCas9	TTGACAGGTGAGAAATCTC (SEQ ID NO: 41)	19
gAP16	g6	SaCas9	TTTGACAGGTGAGAAATCTC (SEQ ID NO: 42)	20
gAP17	g6	SaCas9	CATTTGACAGGTGAGAAATCTC (SEQ ID NO: 43)	22
gAP18	g6	SaCas9	TCATTTGACAGGTGAGAAATCTC (SEQ ID NO: 44)	23
gAP19	g7	SaCas9	ATTTATAATACAGGGGAAT (SEQ ID NO: 45)	19
gAP20	g7	SaCas9	CATTTATAATACAGGGGAAT (SEQ ID NO: 5)	20
gAP21	g7	SaCas9	GTCATTTATAATACAGGGGAAT (SEQ ID NO: 6)	22
gAP22	g7	SaCas9	AGTCATTTATAATACAGGGGAAT (SEQ ID NO: 7)	23
gAP23	零亂	SaCas9	GCACTACCAGAGCTAACTCA (SEQ ID NO: 8)	20

【補充序列表】

<110> 美國公爵大學(DUKE UNIVERSITY)

<120> 恢復肌縮蛋白功能之基於CRISPR/CAS之基因組編輯組合物

<140> TW109112554
<141> 2020-04-14

<150> 62/833,759
<151> 2019-04-14

<160> 45

<170> PatentIn版本3.5

<210> 1
<211> 1368
<212> PRT
<213> 化膿性鏈球菌

<400> 1

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 2
<211> 1053
<212> PRT
<213> 金黃色葡萄球菌

<400> 2

Met Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
50 55 60

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
65 70 75 80

Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
85 90 95

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
100 105 110

Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr
115 120 125

Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
130 135 140

Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
145 150 155 160

Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
165 170 175

Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
180 185 190

Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
195 200 205

Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
210 215 220

Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
225 230 235 240

Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
245 250 255

Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
260 265 270

Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
275 280 285

Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
290 295 300

Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
305 310 315 320

Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
325 330 335

Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
340 345 350

Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
355 360 365

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
370 375 380

Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
385 390 395 400

Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
405 410 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
420 425 430

Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
435 440 445

Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
450 455 460

Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
465 470 475 480

Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
485 490 495

Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
500 505 510

Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
515 520 525

Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
530 535 540

Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
545 550 555 560

Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
565 570 575

Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
580 585 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
595 600 605

Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
610 615 620

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
625 630 635 640

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
645 650 655

Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
660 665 670

Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
675 680 685

Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
690 695 700

Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
705 710 715 720

Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
725 730 735

Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
740 745 750

Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
755 760 765

Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
770 775 780

Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
785 790 795 800

Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
805 810 815

Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
820 825 830

Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
835 840 845

Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
850 855 860

Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile
865 870 875 880

Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
885 890 895

Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
900 905 910

Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
915 920 925

Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
930 935 940

Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
945 950 955 960

Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
965 970 975

Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
980 985 990

Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
995 1000 1005

Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
1010 1015 1020

Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
1025 1030 1035

Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 3
<211> 1368
<212> PRT
<213> 人工序列

<220>

<223> 合成

<400> 3

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 4
<211> 1368
<212> PRT
<213> 人工序列

<220>
<223> 合成

<400> 4

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 5
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 5
catttataat acaggggaat 20

<210> 6
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 6
gtcatttata atacagggga at 22

<210> 7
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 7
agtcatttat aatacagggg aat 23

<210> 8
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 8
gcactaccag agctaactca 20

<210> 9
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 9
attcct 6

<210> 10
<211> 3
<212> DNA
<213> 人工序列

<220>
<223> 合成

<220>
<221> misc_feature
<222> (1)..(1)
<223> n爲a、c、g或t

<400> 10
ngg 3

<210> 11
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> 合成

<220>
<221> misc_feature
<222> (1)..(3)

<223> n爲a、c、g或t

<220>
<221> r
<222> (4)..(4)
<223> r爲a或g

<220>
<221> r
<222> (5)..(5)
<223> r爲a或g

<400> 11
nnnrrt 6

<210> 12
<211> 5
<212> DNA
<213> 人工序列

<220>
<223> 合成

<220>
<221> misc_feature
<222> (1)..(2)
<223> n爲a、c、g或t

<220>
<221> r
<222> (4)..(4)
<223> r爲a或g

<220>
<221> r
<222> (5)..(5)
<223> r爲a或g

<400> 12
nngrr 5

<210> 13
<211> 6
<212> DNA
<213> 人工序列

<220>

<223> 合成

<220>
<221> misc_feature
<222> (1)..(2)
<223> n爲a、c、g或t

<220>
<221> r
<222> (4)..(4)
<223> r爲a或g

<220>
<221> r
<222> (5)..(5)
<223> r爲a或g

<220>
<221> misc_feature
<222> (6)..(6)
<223> n爲a、c、g或t

<400> 13
nngrrn

<210> 14
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> 合成

<220>
<221> misc_feature
<222> (1)..(2)
<223> n爲a、c、g或t

<220>
<221> r
<222> (4)..(4)
<223> r爲a或g

<220>
<221> r
<222> (5)..(5)
<223> r爲a或g

<400> 14
nngrrt 6

<210> 15
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> 合成

<220>
<221> misc_feature
<222> (1)..(2)
<223> n爲a、c、g或t

<220>
<221> r
<222> (4)..(4)
<223> r爲a或g

<220>
<221> r
<222> (5)..(5)
<223> r爲a或g

<220>
<221> v
<222> (6)..(6)
<223> v爲a、c或g

<400> 15
nngrrv 6

<210> 16
<211> 3
<212> DNA
<213> 人工序列

<220>
<223> 合成

<220>
<221> misc_feature
<222> (1)..(1)

<223> n爲a、c、g或t

<400> 16
nga 3

<210> 17
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 17
tcatttataa tacaggggaa t 21

<210> 18
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 18
ttaagtaatc cgaggtactc 20

<210> 19
<211> 97
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 19
tcatttataa tacaggggaa tgttttagta ctctggaaac agaatctact aaaacaaggc 60

aaaatgccgt gtttatctcg tcaacttgtt ggcgaga 97

<210> 20
<211> 96
<212> DNA
<213> 人工序列

<220>

<223> 合成

<400> 20
ttaagtaatc cgaggtactc gttttagtac tctggaaaca gaatctacta aaacaaggca 60

aaatgccgtg tttatctcgt caacttgttg gcgaga 96

<210> 21
<211> 118
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 21
gcaacaatgc aggatttgga acagaggcgt ccccagttgg aagaactcat taccgctgcc 60

caaaatttga aaaacaagac cagcaatcaa gaggctagaa caatcattac ggatcgaa 118

<210> 22
<211> 470
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 22
gttaaattgt tttctataaa cccttataca gtaacatctt ttttatttct aaaagtgttt 60

tggctggtct cacaattgta ctttactttg tattatgtaa aaggaataca caacgctgaa 120

gaaccctgat actaagggat atttgttctt acaggcaaca atgcaggatt tggaacagag 180

gcgtccccag ttggaagaac tcattaccgc tgcccaaaat ttgaaaaaca agaccagcaa 240

tcaagaggct agaacaatca ttacggatcg aagtaagttt tttaacaagc atgggacaca 300

caaagcaaga tgcatgacaa gtttcaataa aaacttaagt tcatatatcc ccctcacatt 360

tataaaaata atgtgaaata attgtaaatg ataacaattg tgctgagatt ttcagtccat 420

aatgttacct tttaataaat gaatgtaatt ccattgaata gaagaaatac 470

<210> 23
<211> 4294

<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 23
ggggggggggg gggggggggtt ggccactccc tctctgcgcg ctcgctcgct cactgaggcc 60

gggcgaccaa aggtcgcccg acgcccgggc tttgcccggg cggcctcagt gagcgagcga 120

gcgcgcagag agggagtggc caactccatc actaggggtt cctcagatct gaattcggta 180

ccttcctagg gcctatttcc catgattcct tcatatttgc atatacgata caaggctgtt 240

agagagataa ttggaattaa tttgactgta aacacaaaga tattagtaca aaatacgtga 300

cgtagaaagt aataatttct tgggtagttt gcagttttaa aattatgttt taaaatggac 360

tatcatatgc ttaccgtaac ttgaaagtat ttcgatttct tggctttata tatcttgtgg 420

aaaggacgaa acaccgtcat ttataataca ggggaatgtt ttagtactct ggaaacagaa 480

tctactaaaa caaggcaaaa tgccgtgttt atctcgtcaa cttgttggcg agatttttt 540

ctagacccag ctttcttgta caaagttggc gtttaaacat tcctattccc ctgtattata 600

aatgagttaa attgttttct ataaaccctt atacagtaac atcttttta tttctaaaag 660

tgttttggct ggtctcacaa ttgtacttta ctttgtatta tgtaaaagga atacacaacg 720

ctgaagaacc ctgatactaa gggatatttg ttcttacagg caacaatgca ggatttggaa 780

cagaggcgtc cccagttgga agaactcatt accgctgccc aaaatttgaa aaacaagacc 840

agcaatcaag aggctagaac aatcattacg gatcgaagta agttttttaa caagcatggg 900

acacacaaag caagatgcat gacaagtttc aataaaaact taagttcata tatccccctc 960

acatttataa aaataatgtg aaataattgt aaatgataac aattgtgctg agattttcag 1020

tccataatgt tacctttaa taaatgaatg taattccatt gaatagaaga aatacattcc 1080

tattcccctg tattataaat gagctagctc ggagagacga catctagaat taaactgtca 1140

ctatcgatta ctaatttttt gctcataata gaagcagcga ttaaaggaat agaatcaaca 1200

gttccagtaa catctcttag tgcatacatt tttttatcag caggaacaat atcatctgac 1260

tgacctgtga tgctcattcc aacttcatta attgttttaa tgaatttttc tttagataat 1320

tcacttgttc aacctttaca tgactctaat ttatcaatag taccaccagt tactccaagt 1380

cctctaccag aaagtttaca aacctttact ccataacttg caactaacgg actatatact 1440

aaacttgttt tgtctccaac tccgccagtt gaatgcttat cagcttttaa acctgtaacc 1500

tcactgacat cataaacata tcctgattca acataagatt gggttaatgc taaagtttct 1560

gctttggtca ttccattaaa ataaaccgcc atagcaaaag cagccatttg ataatctgtt 1620

acattatttt taacgtaact gtttatcaat catttaattt cttcagctga taattctatt 1680

gaatgtttct ttttttctat aatttcacta aaactgtagt tcataagtct ccttttgtaa 1740

gagtgcacaa tatttacacc attactcttt ctactatatt ataatagaat agacatataa 1800

aaaacataag gagtacaaat ggtttttgat aaaaataaca aagtttatag tgaatgaata 1860

aatagccaaa aattggatga tgagttgaaa agccttttag taaatgctac tgatgatgaa 1920

ttgcatgcag catttgaagg aatagagtta gaatttggaa cagcaggtat aagaggtatt 1980

cttggagcag gacctggaag atttaacgtt tacactgtta aaaaagttac tattgcattt 2040

gcagaattat taaaacaaaa ttacccaaat aggttgaatg atggaatagt tgttggtcat 2100

gataaccgtc ataattctaa acagtttgca aaagttgtag ccgaagtttt atcaagcttg 2160

tgaaatagct gttgaagctg gattagaatt tgttaaaaca tcaacaggat tttcaaaatc 2220

aggtgcaaca tttgaagatg ttaaactaat gaagtcagtt gttaaagaca atgctttagt 2280

taaagcagct ggtggagtta gaacatttga agatgctcaa aaaatgattg aagcaggagc 2340

tgaccgctta ggaacaagtg gtggagtagc tattattaaa ggtgaagaaa acaacgcgag 2400

ttactaaaac tagcgttttt ttattttgct cattttttatt aaaagtttgc aaaaaggaac 2460

ataaaaattc taattattga tactaaagtt attaaaaaga agattttggt tgattttata 2520

aaggtcatag aatataatat tttagcatgt gtattttgtg tgctcattta caaccgtctc 2580

gcggccgcgg ggatccagac atgataagat acattgatga gtttggacaa accacaacta 2640

gaatgcagtg aaaaaaatgc tttatttgtg aaatttgtga tgctattgct ttatttgtaa 2700

ccattataag ctgcaataaa caagttaaca acaacaattg cattcatttt atgtttcagg 2760

ttcaggggga ggtgtgggag gtttttttagt cgacctcgag cagtgtggtt ttgcaagagg 2820

aagcaaaaag cctctccacc caggcctgga atgtttccac ccaagtcgaa ggcagtgtgg 2880

ttttgcaaga ggaagcaaaa agcctctcca cccaggcctg gaatgtttcc acccaatgtc 2940

gagcaacccc gcccagcgtc ttgtcattgg cgaattcgaa cacgcagatg cagtcggggc 3000

ggcgcggtcc caggtccact tcgcatatta aggtgacgcg tgtggcctcg aacaccgagc 3060

gaccctgcag ccaatatggg atcggccatt gaacaagatg gattgcacgc aggttctccg 3120

gccgcttggg tggagaggct attcggctat gactgggcac aacagacaat cggctgctct 3180

gatgccgccg tgttccggct gtcagcgcag gggcgccccgg ttctttttgt caagaccgac 3240

ctgtccggtg ccctgaatga actgcaggac gaggcagcgc ggctatcgtg gctggccacg 3300

acgggcgttc cttgcgcagc tgtgctcgac gttgtcactg aagcgggaag ggactggctg 3360

ctattgggcg aagtgccggg gcaggatctc ctgtcatctc accttgctcc tgccgagaaa 3420

gtatccatca tggctgatgc aatgcggcgg ctgcatacgc ttgatccggc tacctgccca 3480

ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta ctcggatgga agccggtctt 3540

gtcgatcagg atgatctgga cgaagagcat caggggctcg cgccagccga actgttcgcc 3600

aggctcaagg cgcgcatgcc cgacggcgag gatctcgtcg tgacccatgg cgatgcctgc 3660

ttgccgaata tcatggtgga aaatggccgc ttttctggat tcatcgactg tggccggctg 3720

ggtgtggcgg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt 3780

ggcggcgaat gggctgaccg cttcctcgtg ctttacggta tcgccgctcc cgattcgcag 3840

cgcatcgcct tctatcgcct tcttgacgag ttcttctgag gggatccgtc gactagagct 3900

cgctgatcag cctcgactgt gccttctagt tgccagccat ctgttgtttg cccctccccc 3960

gtgccttcct tgaccctgga aggtgccact cccactgtcc tttcctaata aaatgaggaa 4020

attgcatcgc attgtctgag taggtgtcat tctattctgg ggggtggggt ggggcaggac 4080

agcaaggggg aggattggga agacaatagc aggcatgctg gggagagatc taggaacccc 4140

tagtgatgga gttggccact ccctctctgc gcgctcgctc gctcactgag gccgcccggg 4200

caaagcccgg gcgtcgggcg acctttggtc gcccggcctc agtgagcgag cgagcgcgca 4260

gagagggagt ggccaacccc ccccccccccc cccc 4294

<210> 24
<211> 4291
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 24

gggggggggg gggggggggtt ggccactccc tctctgcgcg ctcgctcgct cactgaggcc 60

gggcgaccaa aggtcgcccg acgcccgggc tttgcccggg cggcctcagt gagcgagcga 120

gcgcgcagag agggagtggc caactccatc actaggggtt cctcagatct gaattcggta 180

ccttcctagg gcctatttcc catgattcct tcatatttgc atatacgata caaggctgtt 240

agagagataa ttggaattaa tttgactgta aacacaaaga tattagtaca aaatacgtga 300

cgtagaaagt aataatttct tgggtagttt gcagttttaa aattatgttt taaaatggac 360

tatcatatgc ttaccgtaac ttgaaagtat ttcgatttct tggctttata tatcttgtgg 420

aaaggacgaa acaccgttaa gtaatccgag gtactcgttt tagtactctg gaaacagaat 480

ctactaaaac aaggcaaaat gccgtgttta tctcgtcaac ttgttggcga gatttttttc 540

tagacccagc tttcttgtac aaagttggcg tttaaacatt cctgagtacc tcggattact 600

taagttaaat tgttttctat aaacccttat acagtaacat ctttttttatt tctaaaagtg 660

ttttggctgg tctcacaatt gtactttact ttgtattatg taaaaggaat acacaacgct 720

gaagaaccct gatactaagg gatatttgtt cttacaggca acaatgcagg atttggaaca 780

gaggcgtccc cagttggaag aactcattac cgctgcccaa aatttgaaaa acaagaccag 840

caatcaagag gctagaacaa tcattacgga tcgaagtaag ttttttaaca agcatgggac 900

acacaaagca agatgcatga caagtttcaa taaaaactta agttcatata tccccctcac 960

atttataaaa ataatgtgaa ataattgtaa atgataacaa ttgtgctgag attttcagtc 1020

cataatgtta ccttttaata aatgaatgta attccattga atagaagaaa tacattcctg 1080

agtacctcgg attacttaag ctagctcgga gagacgacat ctagaattaa actgtcacta 1140

tcgattacta atttttttgct cataatagaa gcagcgatta aaggaataga atcaacagtt 1200

ccagtaacat ctcttagtgc atacattttt ttatcagcag gaacaatatc atctgactga 1260

cctgtgatgc tcattccaac ttcattaatt gttttaatga atttttcttt agataattca 1320

cttgttcaac ctttacatga ctctaattta tcaatagtac caccagttac tccaagtcct 1380

ctaccagaaa gtttacaaac ctttactcca taacttgcaa ctaacggact atatactaaa 1440

cttgttttgt ctccaactcc gccagttgaa tgcttatcag cttttaaacc tgtaacctca 1500

ctgacatcat aaacatatcc tgattcaaca taagattggg ttaatgctaa agtttctgct 1560

ttggtcattc cattaaaata aaccgccata gcaaaagcag ccatttgata atctgttaca 1620

ttatttttaa cgtaactgtt tatcaatcat ttaatttctt cagctgataa ttctattgaa 1680

tgtttctttt tttctataat ttcactaaaa ctgtagttca taagtctcct tttgtaagag 1740

tgcacaatat ttacaccatt actctttcta ctatattata atagaataga catataaaaa 1800

acataaggag tacaaatggt ttttgataaa aataacaaag tttatagtga atgaataaat 1860

agccaaaaat tggatgatga gttgaaaagc cttttagtaa atgctactga tgatgaattg 1920

catgcagcat ttgaaggaat agagttagaa tttggaacag caggtataag aggtattctt 1980

ggagcaggac ctggaagatt taacgtttac actgttaaaa aagttactat tgcatttgca 2040

gaattattaa aacaaaatta cccaaatagg ttgaatgatg gaatagttgt tggtcatgat 2100

aaccgtcata attctaaaca gtttgcaaaa gttgtagccg aagttttatc aagcttgtga 2160

aatagctgtt gaagctggat tagaatttgt taaaacatca acaggatttt caaaatcagg 2220

tgcaacattt gaagatgtta aactaatgaa gtcagttgtt aaagacaatg ctttagttaa 2280

agcagctggt ggagttagaa catttgaaga tgctcaaaaa atgattgaag caggagctga 2340

ccgcttagga acaagtggtg gagtagctat tattaaaggt gaagaaaaca acgcgagtta 2400

ctaaaactag cgttttttta ttttgctcat ttttattaaa agtttgcaaa aaggaacata 2460

aaaattctaa ttattgatac taaagttatt aaaaagaaga ttttggttga ttttataaag 2520

gtcatagaat ataatatttt agcatgtgta ttttgtgtgc tcatttacaa ccgtctcgcg 2580

gccgcgggga tccagacatg ataagataca ttgatgagtt tggacaaacc acaactagaa 2640

tgcagtgaaa aaaatgcttt atttgtgaaa tttgtgatgc tattgcttta tttgtaacca 2700

ttataagctg caataaacaa gttaacaaca acaattgcat tcattttatg tttcaggttc 2760

aggggggaggt gtgggaggtt ttttagtcga cctcgagcag tgtggttttg caagaggaag 2820

caaaaagcct ctccacccag gcctggaatg tttccaccca agtcgaaggc agtgtggttt 2880

tgcaagagga agcaaaaagc ctctccaccc aggcctggaa tgtttccacc caatgtcgag 2940

caaccccgcc cagcgtcttg tcattggcga attcgaacac gcagatgcag tcgggggcggc 3000

gcggtcccag gtccacttcg catattaagg tgacgcgtgt ggcctcgaac accgagcgac 3060

cctgcagcca atatgggatc ggccattgaa caagatggat tgcacgcagg ttctccggcc 3120

gcttgggtgg agaggctatt cggctatgac tgggcacaac agacaatcgg ctgctctgat 3180

gccgccgtgt tccggctgtc agcgcagggg cgcccggttc tttttgtcaa gaccgacctg 3240

tccggtgccc tgaatgaact gcaggacgag gcagcgcggc tatcgtggct ggccacgacg 3300

ggcgttcctt gcgcagctgt gctcgacgtt gtcactgaag cgggaaggga ctggctgcta 3360

ttgggcgaag tgccggggca ggatctcctg tcatctcacc ttgctcctgc cgagaaagta 3420

tccatcatgg ctgatgcaat gcggcggctg catacgcttg atccggctac ctgcccattc 3480

gaccaccaag cgaaacatcg catcgagcga gcacgtactc ggatggaagc cggtcttgtc 3540

gatcaggatg atctggacga agagcatcag gggctcgcgc cagccgaact gttcgccagg 3600

ctcaaggcgc gcatgcccga cggcgaggat ctcgtcgtga cccatggcga tgcctgcttg 3660

ccgaatatca tggtggaaaa tggccgcttt tctggattca tcgactgtgg ccggctgggt 3720

gtggcggacc gctatcagga catagcgttg gctacccgtg atattgctga agagcttggc 3780

ggcgaatggg ctgaccgctt cctcgtgctt tacggtatcg ccgctcccga ttcgcagcgc 3840

atcgccttct atcgccttct tgacgagttc ttctgagggg atccgtcgac tagagctcgc 3900

tgatcagcct cgactgtgcc ttctagttgc cagccatctg ttgtttgccc ctccccccgtg 3960

ccttccttga ccctggaagg tgccactccc actgtccttt cctaataaaa tgaggaaatt 4020

gcatcgcatt gtctgagtag gtgtcattct attctggggg gtggggtggg gcaggacagc 4080

aaggggggagg attgggaaga caatagcagg catgctgggg agagatctag gaacccctag 4140

tgatggagtt ggccactccc tctctgcgcg ctcgctcgct cactgaggcc gcccgggcaa 4200

agcccgggcg tcgggcgacc tttggtcgcc cggcctcagt gagcgagcga gcgcgcagag 4260

agggagtggc caacccccccc ccccccccccc c 4291

<210> 25
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 25
attcccctgt attataaatg a 21

<210> 26
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 26
gagtacctcg gattacttaa 20

<210> 27
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 27
ctttactttg tattatgtaa a 21

<210> 28
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 28
tttgaaatat ttttgatatc t 21

<210> 29
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 29
tttaagtaat ccgaggtact c 21

<210> 30
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 30
tttaaataca ttgtcgtaat t 21

<210> 31
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 31
taccttaatt ttgacgtcac a 21

<210> 32
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 32
atttgacagg tgagaaatct c 21

<210> 33
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 33
tcatttataa tacaggggaa t 21

<210> 34
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 34
ttaaagtcat ttataataca g 21

<210> 35
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 35
aaatagacac tgaagaaagg g 21

<210> 36
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 36
ccccaattaa ataaaattt a 21

<210> 37
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 37
taagtaatcc gaggtactc 19

<210> 38
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 38
ttaagtaatc cgaggtactc 20

<210> 39
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 39
gtttaagtaa tccgaggtac tc 22

<210> 40
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 40
ggtttaagta atccgaggta ctc 23

<210> 41
<211> 19

<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 41
ttgacaggtg agaaatctc 19

<210> 42
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 42
tttgacaggt gagaaatctc 20

<210> 43
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 43
catttgacag gtgagaaatc tc 22

<210> 44
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 44
tcatttgaca ggtgagaaat ctc 23

<210> 45
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 合成

<400> 45
atttataata caggggaat 19

【發明申請專利範圍】

【請求項1】

一種基於CRISPR/Cas之基因組編輯系統，其包含一或多種編碼組合物之載體，該組合物包含：

(a) 引導RNA (gRNA)，其靶向突變肌縮蛋白基因之片段；

(b) Cas蛋白質或包含該Cas蛋白質之融合蛋白；及

(c) 供體序列，其包含野生型肌縮蛋白基因之片段。

【請求項2】

一種基於CRISPR/Cas之基因組編輯系統，其包含：

(a) 引導RNA (gRNA)，其靶向突變肌縮蛋白基因之片段；

(b) Cas蛋白質或包含該Cas蛋白質之融合蛋白；及

(c) 供體序列，其包含野生型肌縮蛋白基因之片段。

【請求項3】

如請求項1或2之系統，其中該野生型肌縮蛋白基因之該片段由兩個gRNA間隔序列及/或PAM序列側接。

【請求項4】

如請求項1至3中任一項之系統，其中該gRNA靶向與該突變肌縮蛋白基因之外顯子並列的內含子，且其中該外顯子選自該突變肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59及61-66。

【請求項5】

如請求項1至3中任一項之系統，其中該供體序列包含該野生型肌縮蛋白基因之外顯子或其功能性等效物，且其中該外顯子選自該野生型肌縮蛋白基因之外顯子1-8、10、11、12、14、16-22、43-59、61-66。

【請求項6】

如請求項4之系統，其中該突變肌縮蛋白基因之該外顯子經突變或自該肌縮蛋白基因至少部分地缺失，或其中該突變肌縮蛋白基因之該外顯子缺失且該內含子與相應野生型肌縮蛋白基因中之缺失的外顯子的位置並列。

【請求項7】

如請求項4或5之系統，其中該外顯子為外顯子52。

【請求項8】

如請求項1至7中任一項之系統，其中該gRNA結合及靶向包含以下之聚核苷酸序列：

a) SEQ ID NO: 17或SEQ ID NO: 18；

b) SEQ ID NO: 17或SEQ ID NO: 18之片段；

c) SEQ ID NO: 17或SEQ ID NO: 18之互補序列或其片段；

d) 與SEQ ID NO: 17或SEQ ID NO: 18或其互補序列實質上一致的核酸；或

e) 在嚴格條件下雜交至SEQ ID NO: 17或SEQ ID NO: 18、其互補序列或與其實質上一致之序列的核酸。

【請求項9】

如請求項1至8中任一項之系統，其中該gRNA包含聚核苷酸序列SEQ ID NO: 19或SEQ ID NO: 20或其變體或由其編碼。

【請求項10】

如請求項1至9中任一項之系統，其中該Cas蛋白質為化膿性鏈球菌(*Streptococcus pyogenes*)Cas9蛋白質或金黃色葡萄球菌(*Staphylococcus*

aureus)Cas9蛋白質。

【請求項11】

如請求項1至10中任一項之系統，其中該Cas蛋白質包含胺基酸序列SEQ ID NO: 1、2、3或4。

【請求項12】

如請求項3至11中任一項之系統，其中該兩個gRNA間隔序列獨立地包含選自SEQ ID NO: 5-8及25-45之序列。

【請求項13】

如請求項12之系統，其中該兩個gRNA間隔序列為相同的。

【請求項14】

如請求項12之系統，其中該兩個gRNA間隔序列為不同的。

【請求項15】

如請求項3至14中任一項之系統，其中該兩個gRNA間隔序列中之至少一者包含序列SEQ ID NO: 25或SEQ ID NO: 26。

【請求項16】

如請求項1至15中任一項之系統，其中該供體序列包含SEQ ID NO: 21或SEQ ID NO: 22之聚核苷酸。

【請求項17】

如請求項1及3至16中任一項之系統，其中該載體為病毒載體。

【請求項18】

如請求項17之系統，其中該載體為腺相關病毒(AAV)載體。

【請求項19】

如請求項18之系統，其中該AAV載體為AAV1、AAV2、AAV3、

AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV-10、AAV-11、AAV-12、AAV-13或AAVrh.74載體。

【請求項20】

如請求項18之系統，其中該一或多個載體中之一者包含聚核苷酸序列SEQ ID NO: 23或24。

【請求項21】

如請求項1至20中任一項之系統，其中gRNA與供體序列之間的莫耳比為1:1、或1:15、或5:1至1:10、或1:1至1:5。

【請求項22】

一種重組聚核苷酸，其編碼包含野生型肌縮蛋白基因之片段或其功能性等效物的供體序列，且其中該片段或其功能性等效物由兩個gRNA間隔序列側接。

【請求項23】

如請求項22之重組聚核苷酸，其中該供體序列包含該肌縮蛋白基因之外顯子，且其中該外顯子選自外顯子1-8、10、11、12、14、16-22、43-59、61-66。

【請求項24】

如請求項22或23之重組聚核苷酸，其中該重組聚核苷酸包含序列SEQ ID NO: 23或24。

【請求項25】

一種載體，其包含如請求項22至24中任一項之重組聚核苷酸。

【請求項26】

如請求項25之載體，其中該載體包含驅動該重組聚核苷酸表現之異

源啟動子。

【請求項27】

一種細胞，其包含如請求項22至24中任一項之重組聚核苷酸或如請求項25或26之載體。

【請求項28】

一種用於恢復具有突變肌縮蛋白基因之細胞的肌縮蛋白功能的組合物，該組合物包含如請求項1至21中任一項之系統、如請求項22至24中任一項之重組聚核苷酸或如請求項25或26之載體。

【請求項29】

一種套組，其包含如請求項1至21中任一項之系統、如請求項22至24中任一項之重組聚核苷酸或如請求項25或26之載體或如請求項28之組合物。

【請求項30】

一種用於恢復具有突變肌縮蛋白基因之細胞或個體的肌縮蛋白功能的方法，該方法包含使該細胞或該個體與如請求項1至21中任一項之系統、如請求項22至24中任一項之重組聚核苷酸或如請求項25或26之載體或如請求項28之組合物接觸。

【請求項31】

如請求項30之方法，其中該肌縮蛋白功能藉由插入該野生型肌縮蛋白基因之外顯子52來恢復。

【請求項32】

如請求項30或31之方法，其中該個體患有杜興氏肌肉萎縮症(Duchenne Muscular Dystrophy)。

【請求項33】

一種用於恢復具有由一或多個缺失或突變型外顯子引起的破壞的肌縮蛋白基因之細胞或個體的肌縮蛋白功能的方法，該方法包含使該細胞或該個體與如請求項1至21中任一項之系統、如請求項22至24中任一項之重組聚核苷酸或如請求項25或26之載體或如請求項28之組合物接觸。

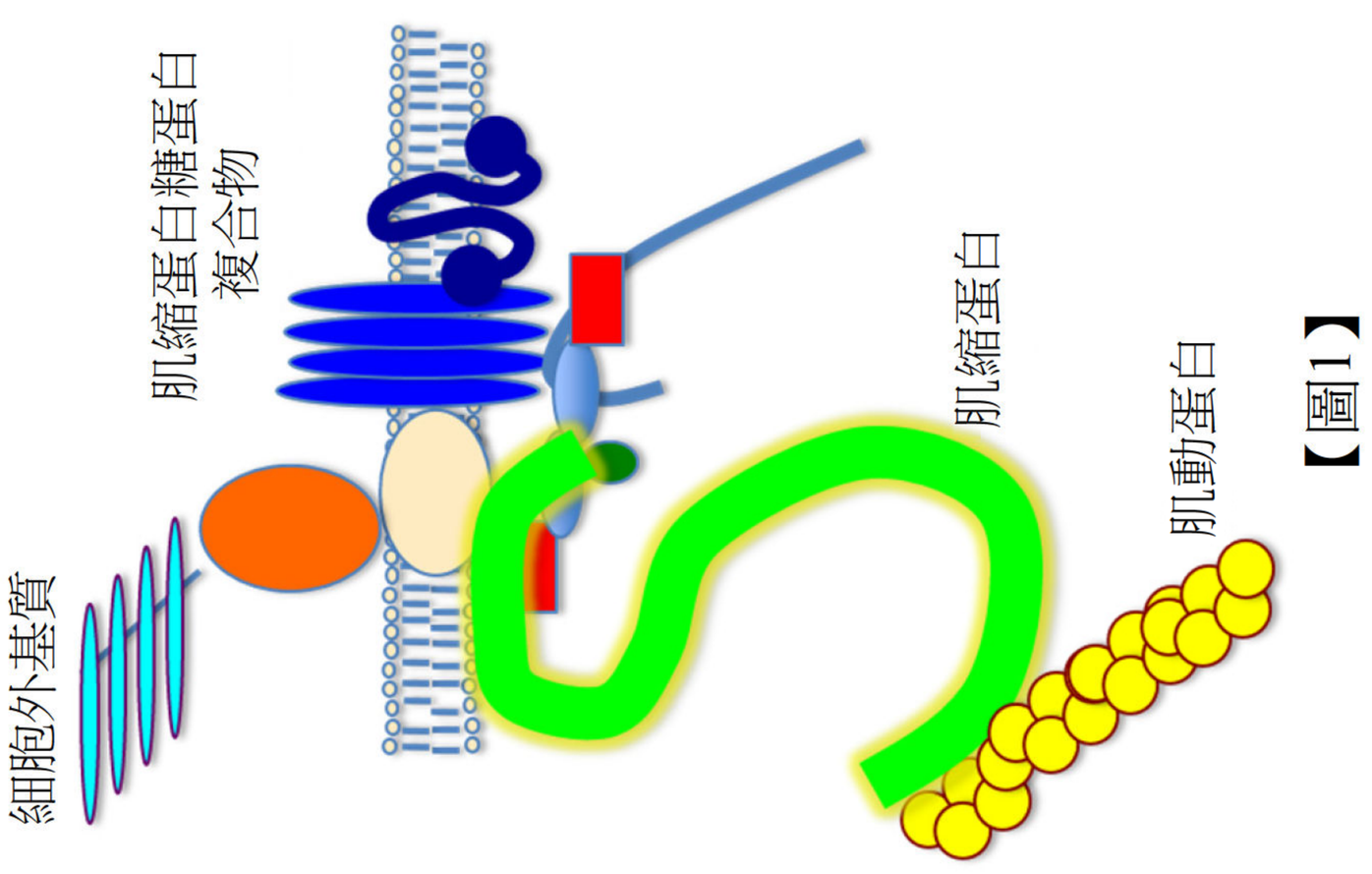
【請求項34】

如請求項33之方法，其中肌縮蛋白功能藉由插入對應於該一或多個缺失或突變型外顯子之肌縮蛋白基因的一或多個野生型外顯子來恢復。

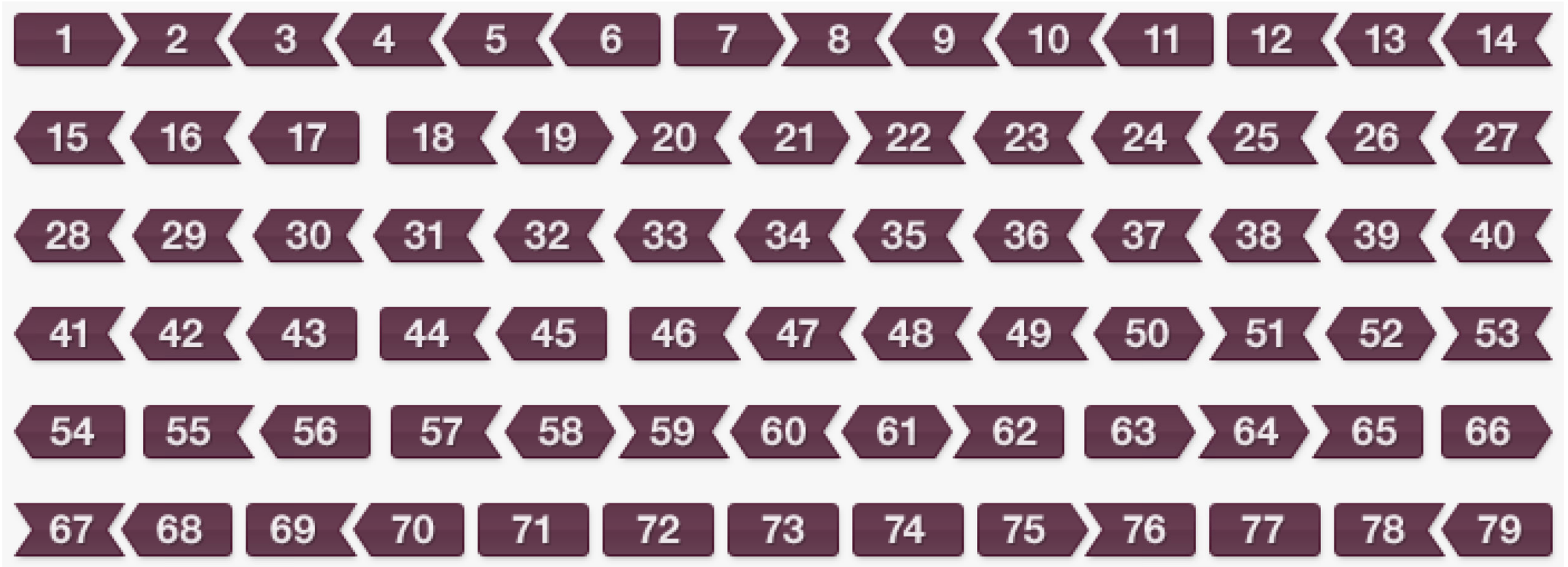
【請求項35】

如請求項34之方法，其中該等缺失或突變型外顯子中之一者為外顯子52。

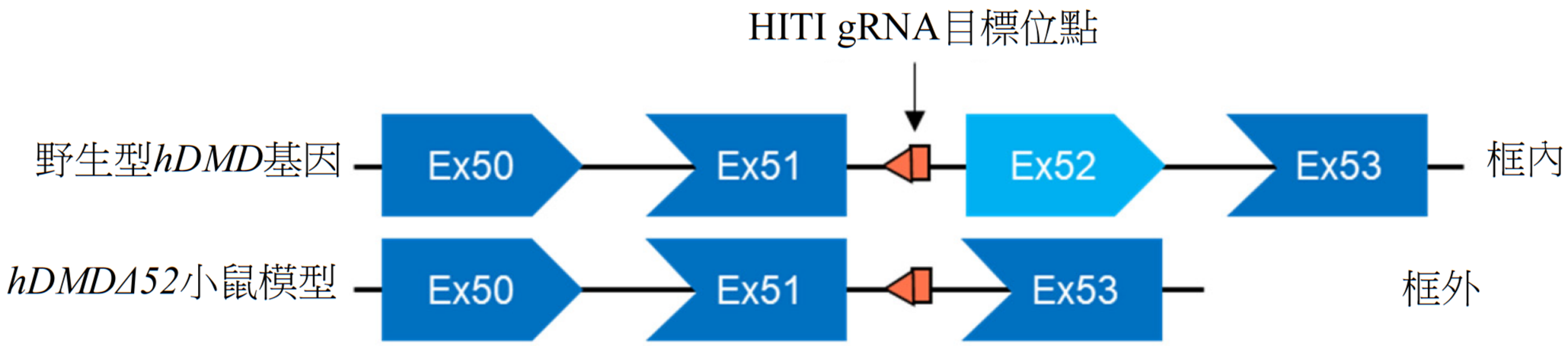
【發明圖式】

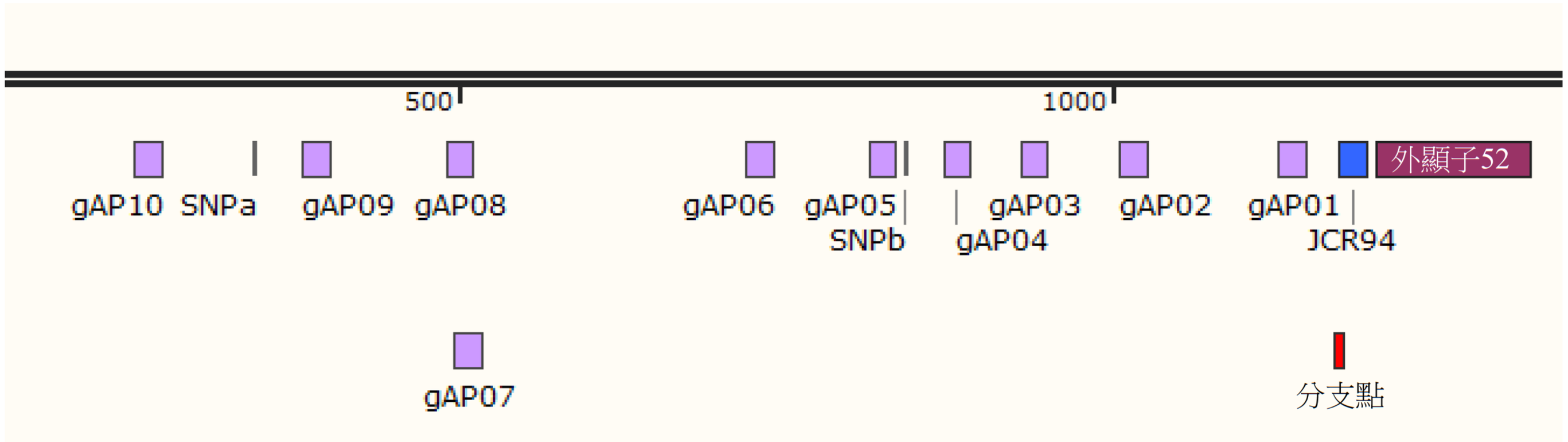


【圖1】


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
41 42 43 44 45 46 47 48 49 50 51 52 53
54 55 56 57 58 59 60 61 62 63 64 65 66
67 68 69 70 71 72 73 74 75 76 77 78 79

【圖2】


HITI gRNA目標位點
野生型hDMD基因
Ex50
Ex51
Ex52
Ex53
框內
hDMDΔ52小鼠模型
Ex50
Ex51
Ex53
框外

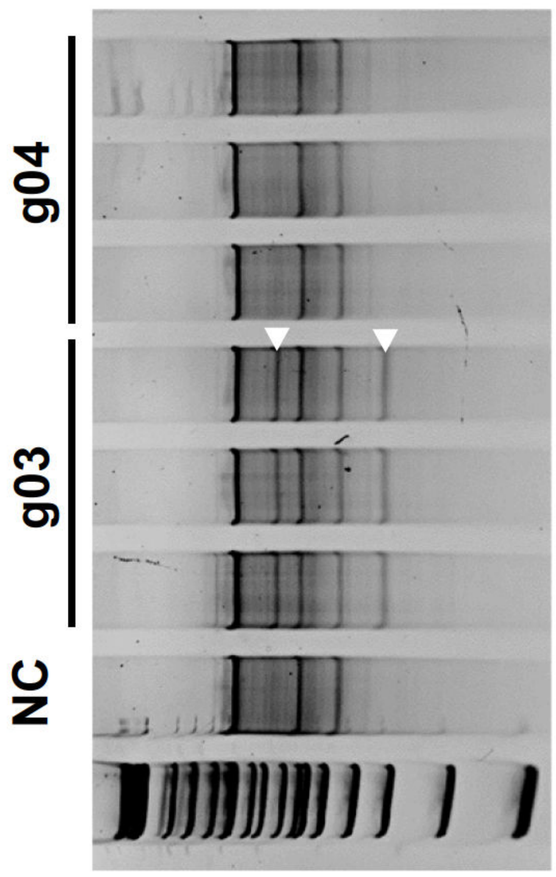

500
1000
gAP10
SNPa
gAP09
gAP08
gAP06
gAP05
SNPb
gAP04
gAP03
gAP02
gAP01
JCR94
外顯子52
gAP07
分支點

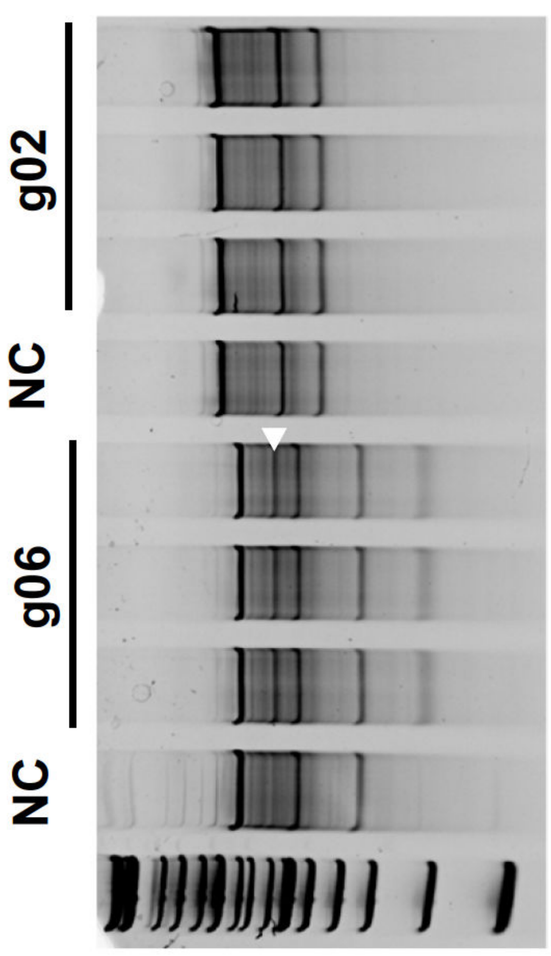
【圖3】

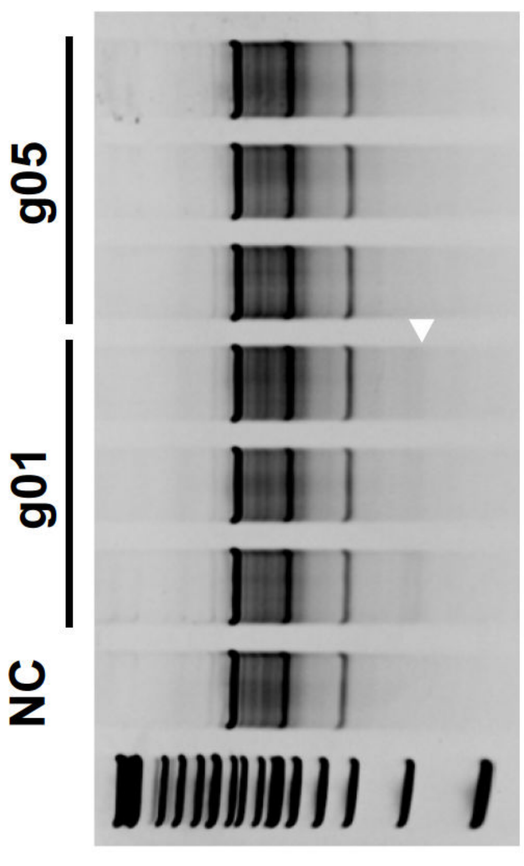
C229168DA.pdf
109112554

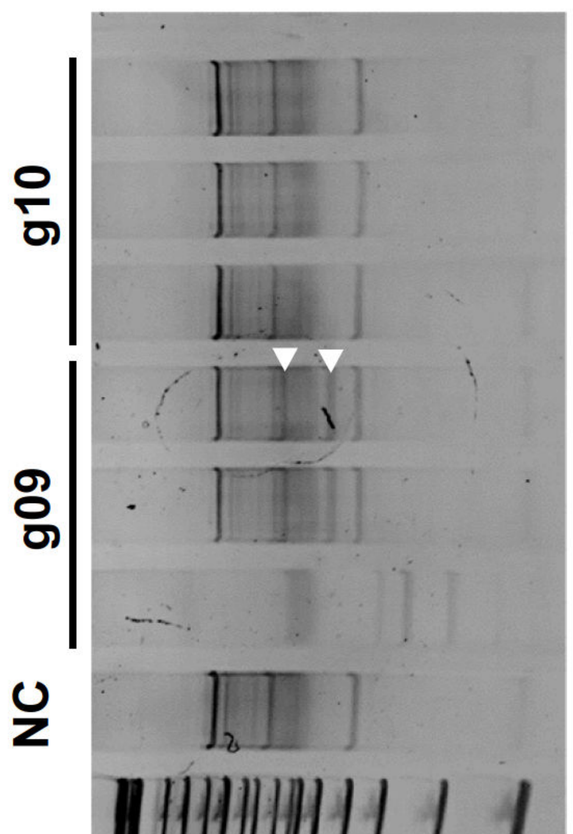
引導 #	F 引物	R 引物	親本 (bp)	L 臂 (BP)	R 臂 (BP)
gAP1	AP-578F	AP-579R	817	608	209
gAP2	AP-580F	AP-581R	933	694	239
gAP3	AP-580F	AP-581R	933	618	315
gAP4	AP-580F	AP-581R	933	559	374
gAP5	AP-578F	AP-579R	817	295	522
gAP6	AP-578F	AP-579R	817	201	616
gAP7	AP-582F	AP-583R	903	244	659
gAP8	AP-582F	AP-583R	903	238	665
gAP9	AP-584F	AP-585R	943	387	556
gAP10	AP-584F	AP-585R	943	259	684

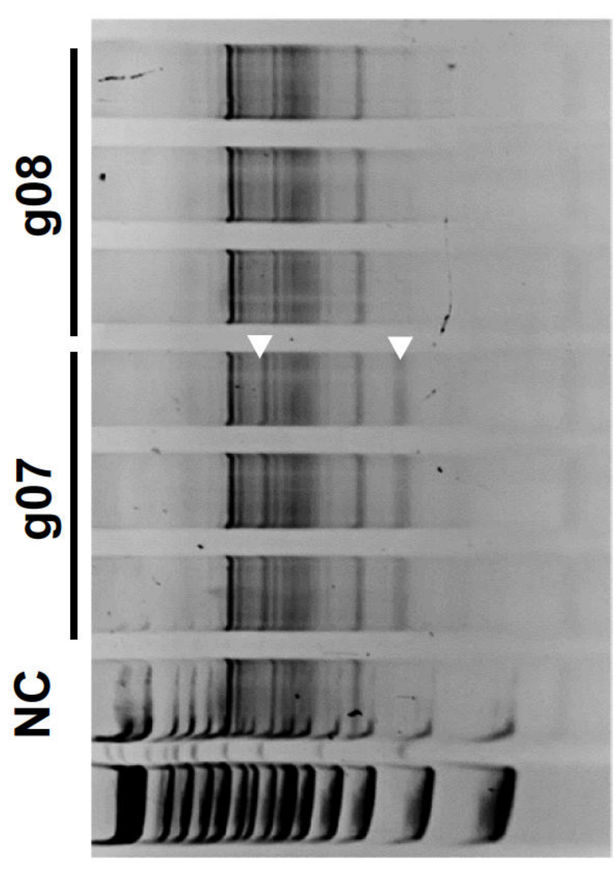
【圖4A】









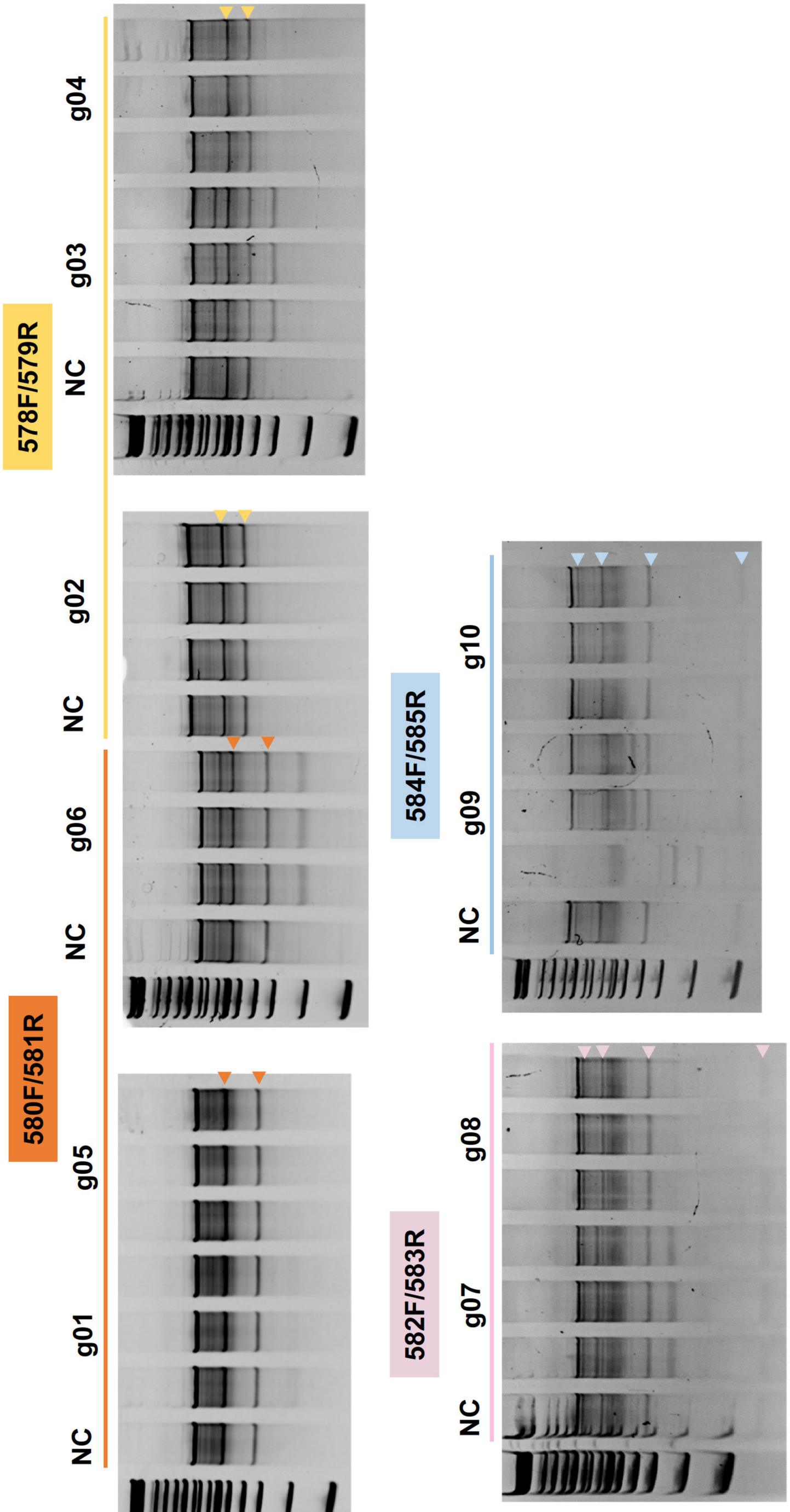


【圖4B】

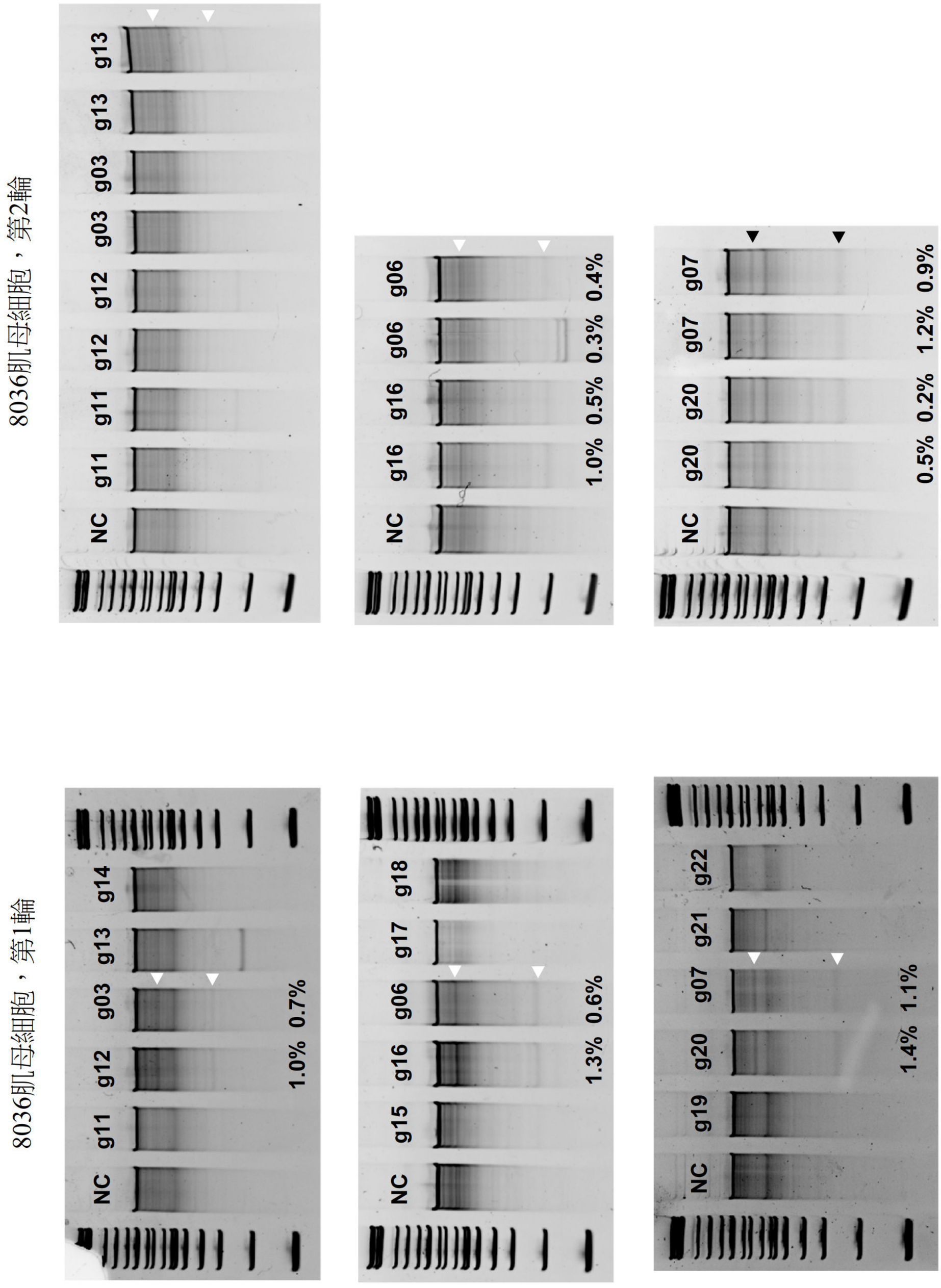
基於引物位置之293T SNP結果		
	SNPa	**SNPb**
578F/579R	外部引物	**306/511**
580F/581R	外部引物	**512/421**
582F/583R	**73/830**	**572/330**
584F/585R	**333/610**	**830/112**

引導 #	F 引物	R 引物	親本 (bp)	L 臂 (BP)	R 臂 (BP)
gAP1	AP-578F	AP-579R	817	608	209
gAP2	AP-580F	AP-581R	933	694	239
gAP3	AP-580F	AP-581R	933	618	315
gAP4	AP-580F	AP-581R	933	559	374
gAP5	AP-578F	AP-579R	817	295	522
gAP6	AP-578F	AP-579R	817	201	616
gAP7	AP-582F	AP-583R	903	244	659
gAP8	AP-582F	AP-583R	903	238	665
gAP9	AP-584F	AP-585R	943	387	556
gAP10	AP-584F	AP-585R	943	259	684

【圖5A】


578F/579R
NC
g03
g04
580F/581R
NC
g06
g05
NC
g01
582F/583R
NC
g07
g08
584F/585R
NC
g09
g10
NC
g02

【圖5B】

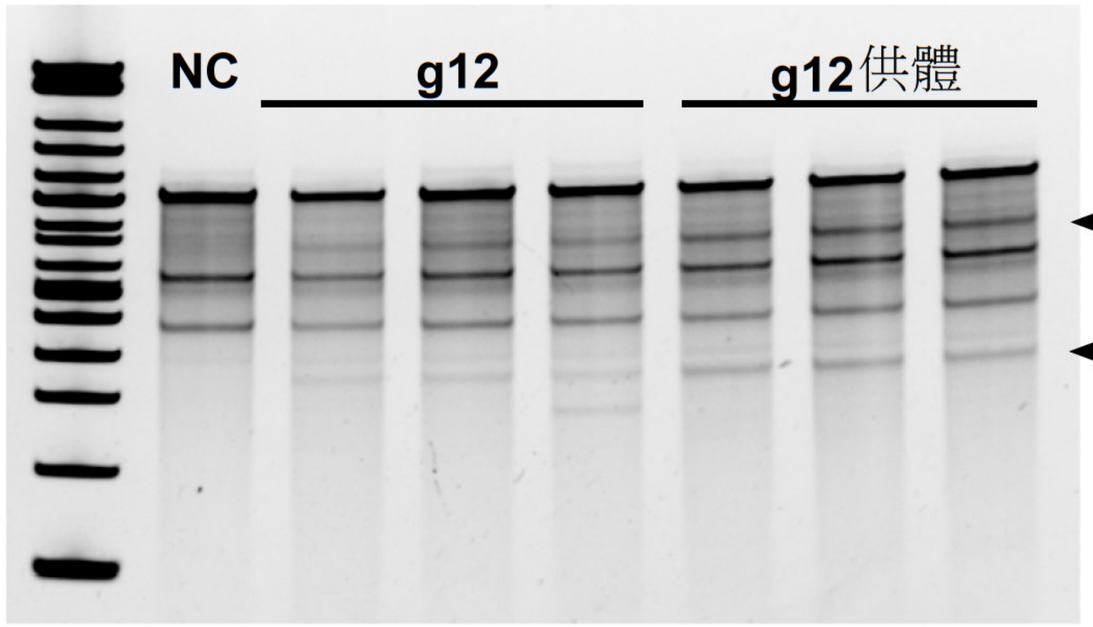

8036肌母細胞，第2輪
NC g11 g11 g12 g12 g03 g13
NC g16 g16 g06 g06 g13 g13
1.0% 0.5% 0.3% 0.4%
NC g20 g20 g07 g07 g12 g12
0.5% 0.2% 1.2% 0.9%
8036肌母細胞，第1輪
NC g11 g12 g03 g13 g14
1.0% 0.7%
NC g15 g16 g06 g17 g18
1.3% 0.6%
NC g19 g20 g07 g21 g22
1.4% 1.1%

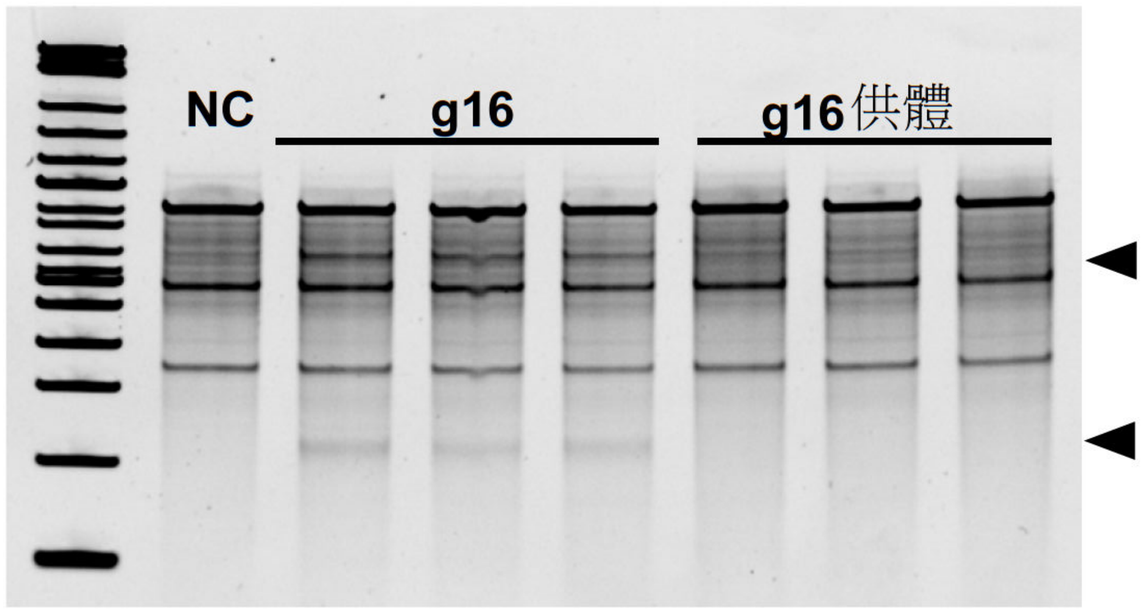
【圖6A】

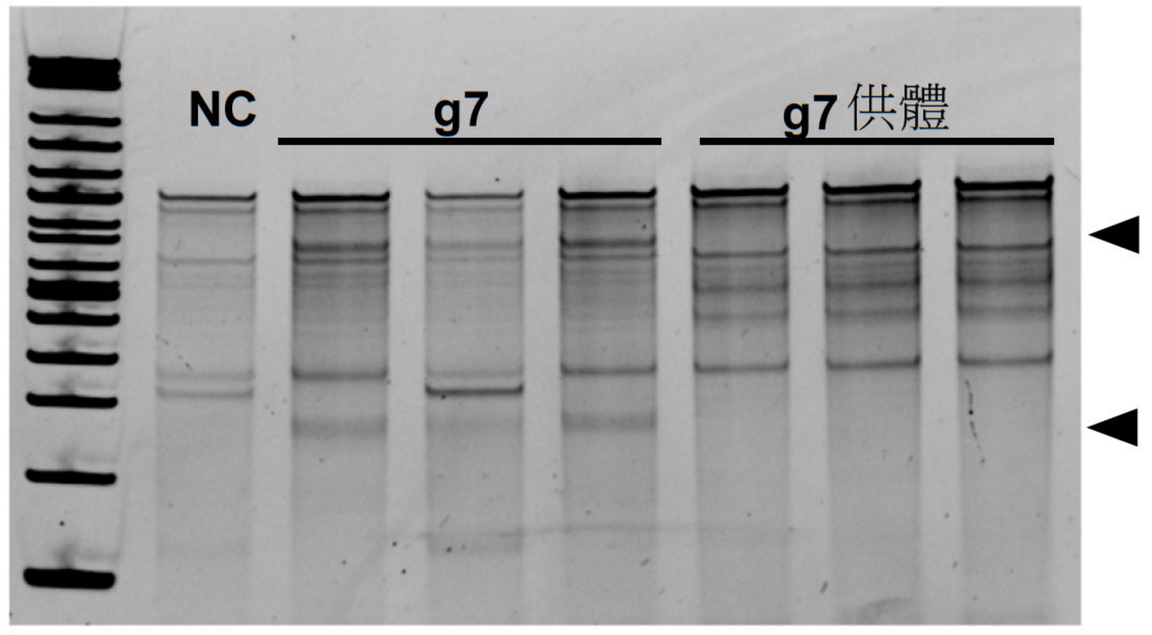
gRNA	間隔序列 (nt)	第一輪測量員	第一輪肽	第二輪測量員	
g12	**20**	最佳的為 **g3**	**7.7%**		***
g3	21		1.6%		
g13	22		1.7%		
g16	**20**	最佳的為 **g6**	**5.5%**	最佳的為 **g6**	***
g6	21		2.6%		
g17	22		1.3%		
g20	20	最佳的為 g7	6.4%		
g7	**21**		**7.0%**	最佳的為 **g7**	***
g21	22		0.5%		

【圖6B】

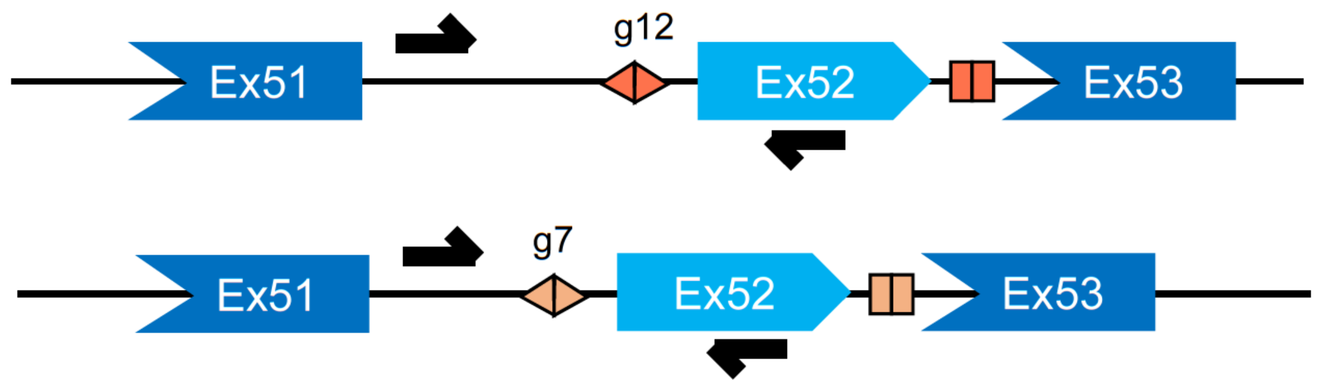
引導 #	F 引物	R引物	親本(bp)	L 臂 (BP)	R 臂 (BP)
gAP12/g12 供體	AP-580F	AP-581R	933	618	315
gAP16/g16 供體	AP-578F	AP-579R	817	201	616
gAP7/g7 供體	AP-582F	AP-583R	903	244	659

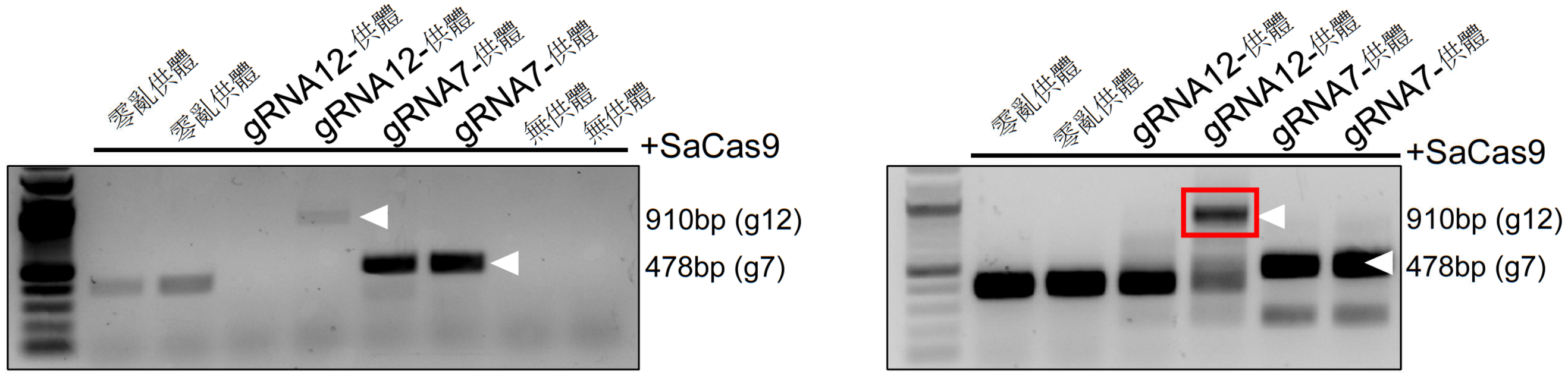






【圖7】


g12
Ex51
Ex52
Ex53
g7
Ex51
Ex52
Ex53


雲亂供體
雲亂供體
gRNA12-供體
gRNA12-供體
gRNA7-供體
gRNA7-供體
無供體
無供體
+SaCas9
910bp (g12)
478bp (g7)
雲亂供體
雲亂供體
gRNA12-供體
gRNA12-供體
gRNA7-供體
gRNA7-供體
+SaCas9
910bp (g12)
478bp (g7)

【圖8A】


caaggcattcagacagtggtttaagtaatccgaggtatacctcggattacttaagttaaatTGTTTTCTATAAACCCTTATACAGTAACATCTTTTTTATTTCTAAAAGTGTTTTGGCT
gttccgtaagtctgtcaccaaattcattaggctccatatggagcctaatgaattcaatttaACAAAAGATATTTGGGAATATGTCATTGTAGAAAAAATAAAGATTTTCACAAAACCGA
gAP03
HITI供體插入
gAP12 (g3)
內含子51
caaggcattcagacagtggtttaagtaatccgaggtatacctcggattacttaagttaaatTGTTTTCTATAAACCCTTATACAGTAACATCTTTTTTATTTCTAAAAGTGTTTTGGCT
CAAGGCATTCAGACAGTGGTTTAAGTAATCCGAGGTATACCTCGGATTACTTAAGTTAGATTGTTTTCTATAAACCCTTATACAGTAACATCTTTTTTATTTCTAAGAGTGTTTTGGCT
350
360
370
380
390
400
410
420
430
440
450
460

【圖8B】



hDMDΔ52小鼠
雄性6至8週齡
總共16隻小鼠

第0天

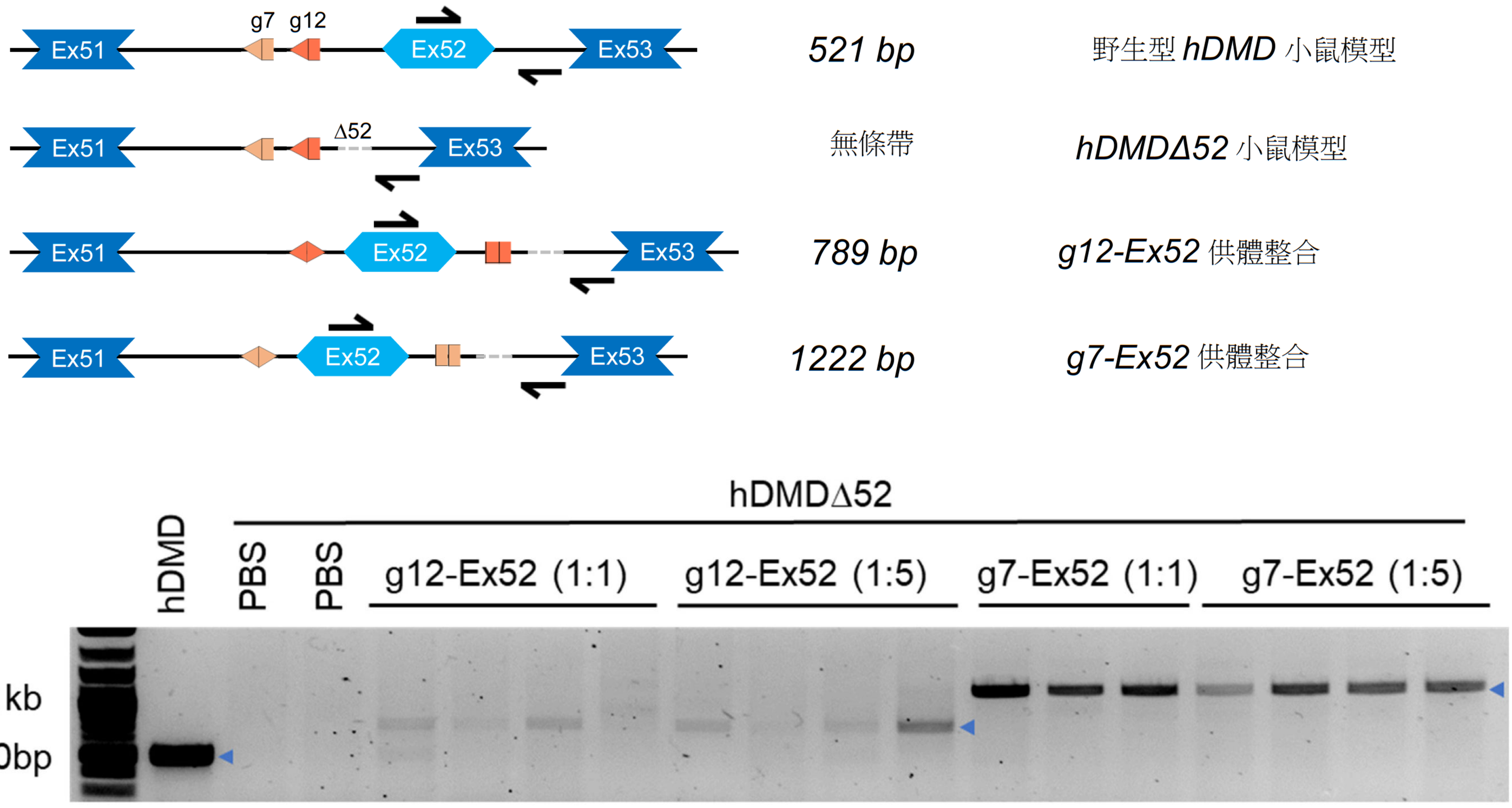
AAV-Cas9 + AAV-gRNA7-供體
N=4 1:1 (4.8e11 vg + 4.8e11 vg)
N=4 1:5 (1.85e11 vg + 9.23e11 vg)

AAV-Cas9 + AAV-gRNA12-供體
N=4 1:1 (4.8e11 vg + 4.8e11 vg)
N=4 1:5 (1.85e11 vg + 9.23e11 vg)

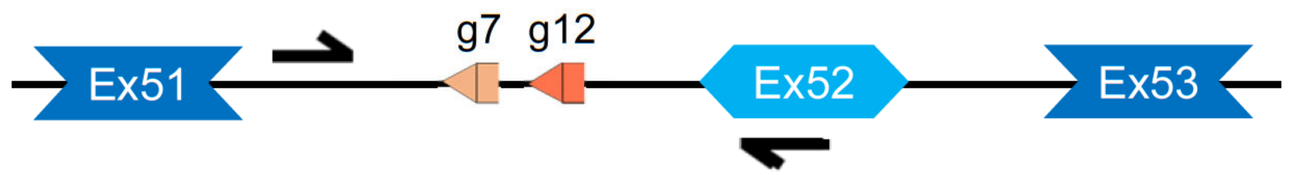
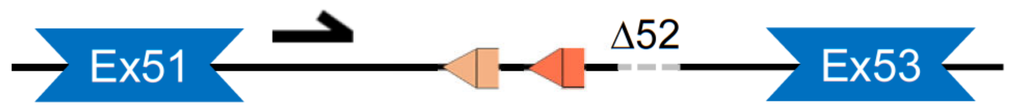
注射後4週

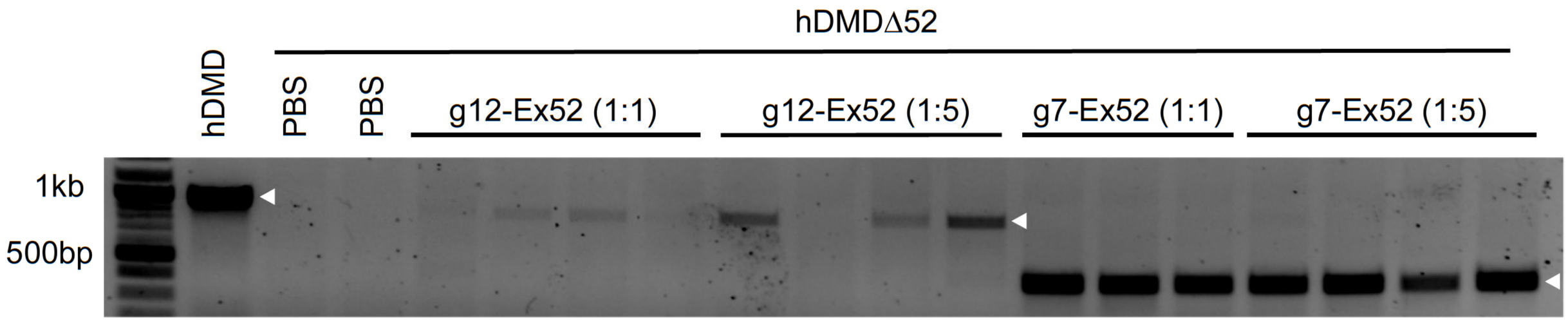
終止
收集TA
gDNA
RNA
蛋白質

【圖9】


擴增子大小
樣品
g7
g12
Ex51
Ex52
Ex53
521 bp
野生型 hDMD 小鼠模型
Δ52
無條帶
hDMDΔ52 小鼠模型
789 bp
g12-Ex52 供體整合
1222 bp
g7-Ex52 供體整合
hDMDΔ52
hDMD
PBS
PBS
g12-Ex52 (1:1)
g12-Ex52 (1:5)
g7-Ex52 (1:1)
g7-Ex52 (1:5)
1kb
500bp

【圖10A】

樣品	擴增子大小	
野生型 *hDMD* 小鼠模型	*978 bp*	
hDMDΔ52 小鼠模型	無條帶	
g12-Ex52 供體整合	*896 bp*	
g7-Ex52 供體整合	*465 bp*	

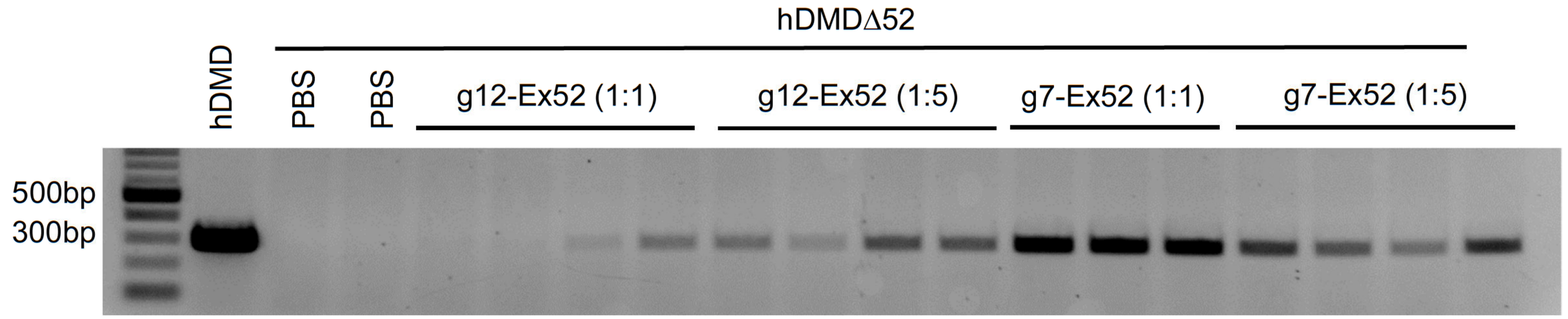


【圖10B】

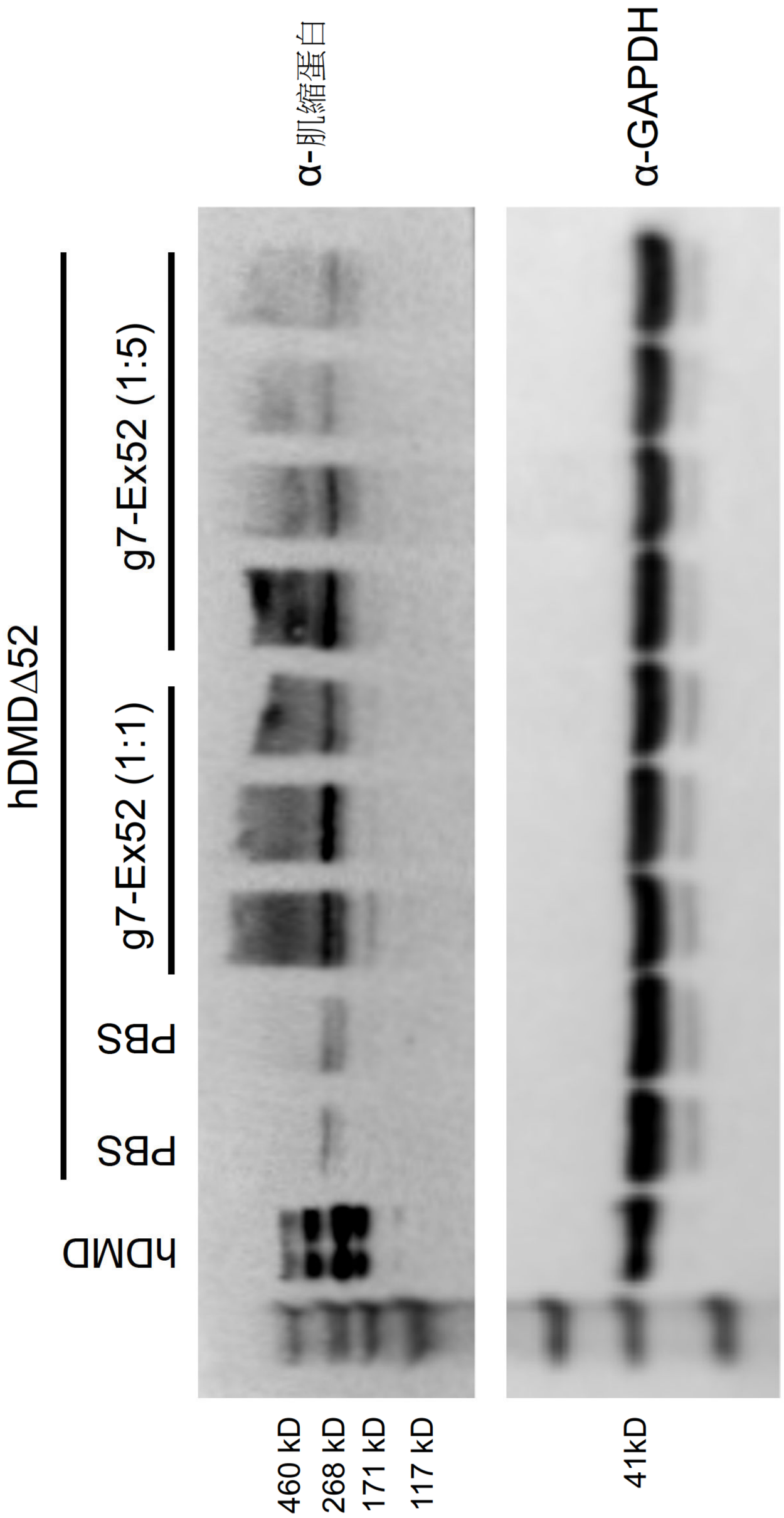


預期大小:

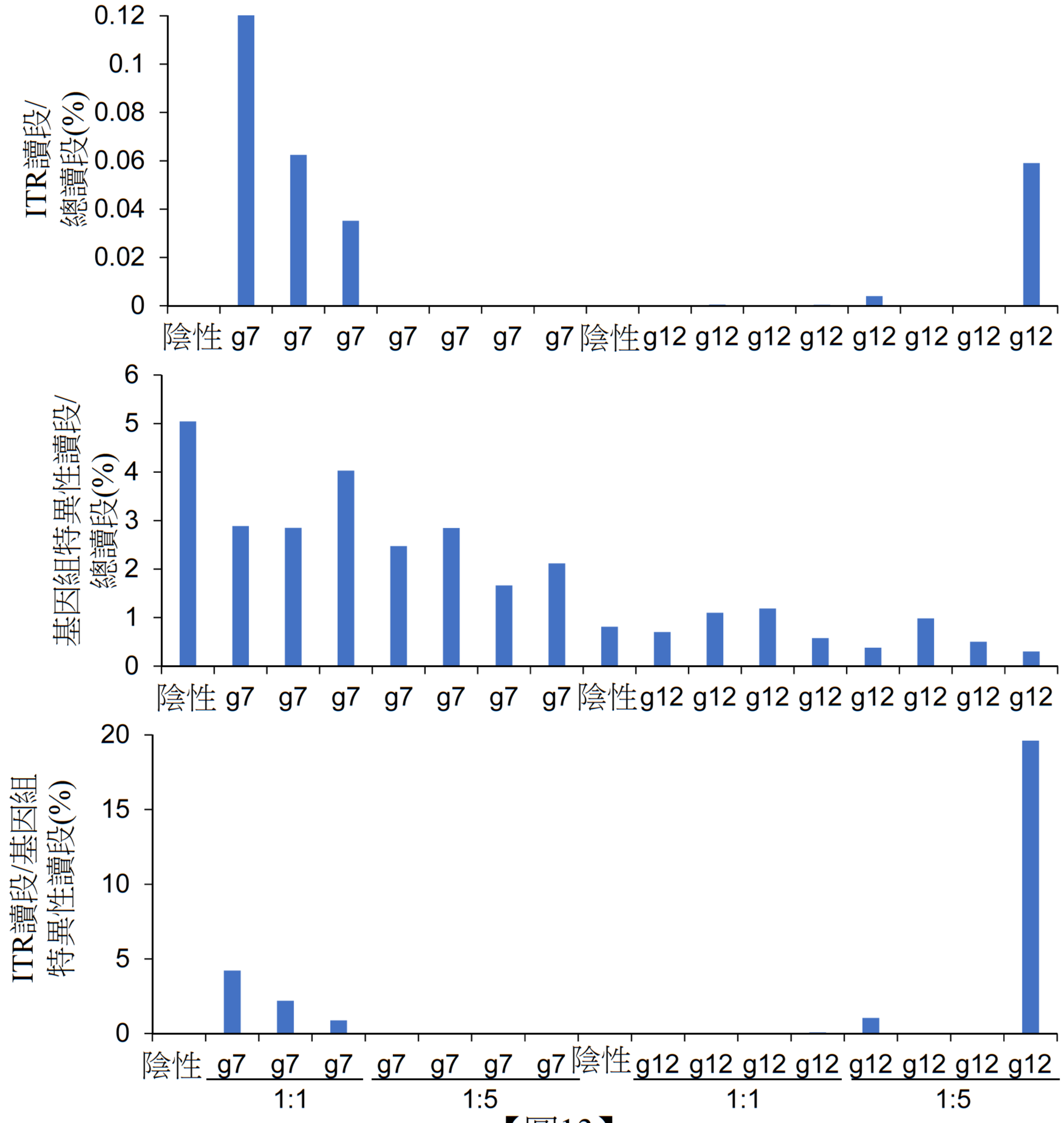
300bp



【圖11】


α-肌縮蛋白
α-GAPDH
hDMDΔ52
g7-Ex52 (1:1)
g7-Ex52 (1:5)
hDMD
PBS
PBS
g7-Ex52 (1:1)
g7-Ex52 (1:5)
460 kD
268 kD
171 kD
117 kD
41kD

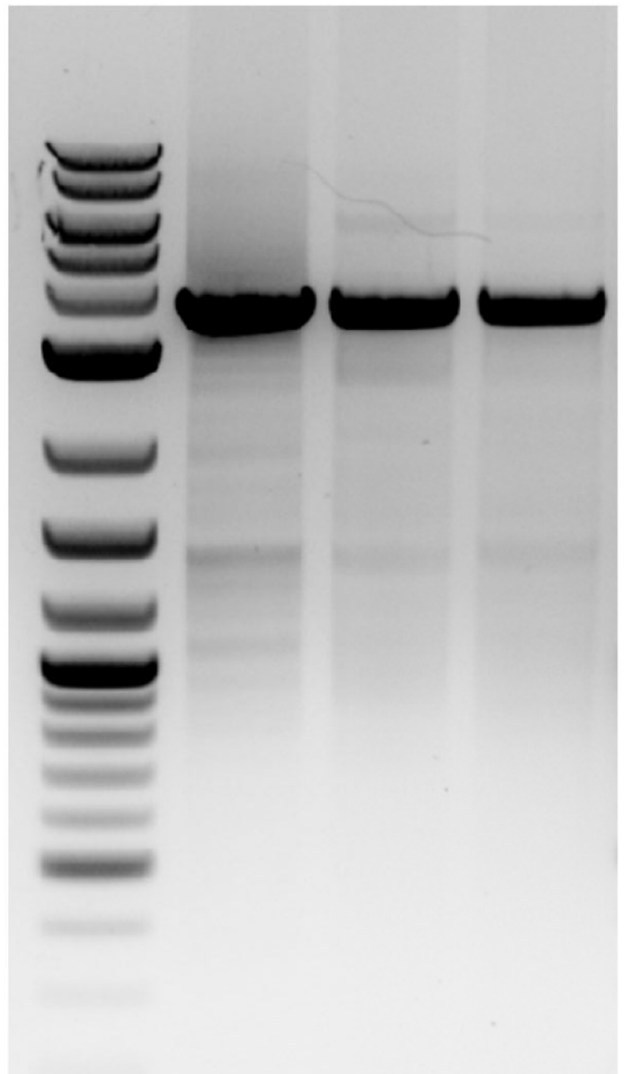
【圖12】

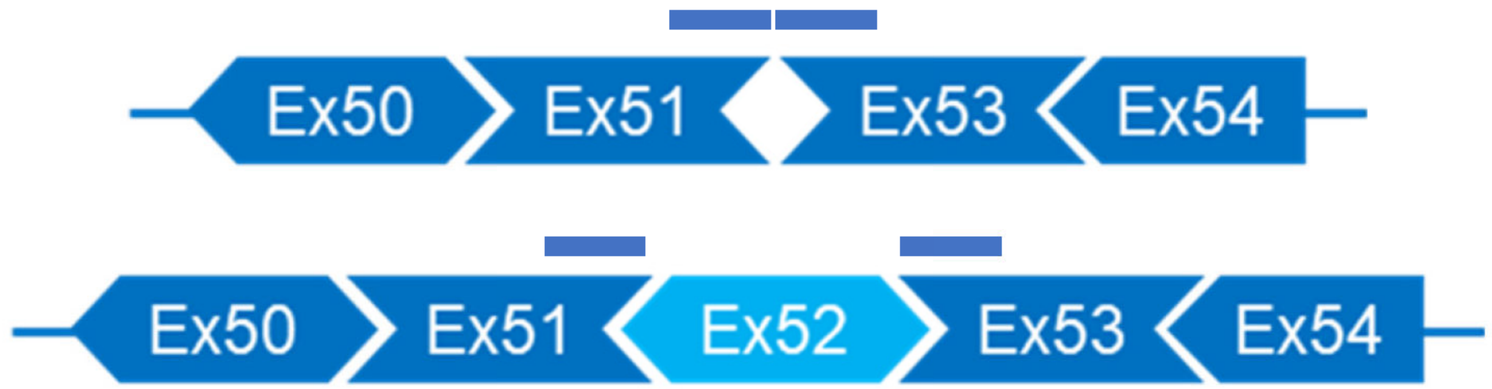

ITR讀段/總讀段(%)
0.12
0.1
0.08
0.06
0.04
0.02
0
陰性 g7 g7 g7 g7 g7 g7 g7 陰性 g12 g12 g12 g12 g12 g12 g12 g12
基因組特異性讀段/總讀段(%)
6
5
4
3
2
1
0
陰性 g7 g7 g7 g7 g7 g7 g7 陰性 g12 g12 g12 g12 g12 g12 g12 g12
ITR讀段/基因組特異性讀段(%)
20
15
10
5
0
陰性 g7 g7 g7 g7 g7 g7 g7 陰性 g12 g12 g12 g12 g12 g12 g12 g12
1:1
1:5
1:1
1:5

【圖13】

202100748

外顯子 45 – 外顯子 69
3843 bp





小鼠	**Ex52** 插入 (讀段數)	無插入 (讀段數)	Ex52插入 (百分比)
最佳編輯的(gRNA7)	232	21692	1.07%
次最佳編輯的(gRNA7)	50	17004	0.29%
最佳編輯的gRNA12	35	21377	0.16%
未編輯的對照	0	23082	0%

【圖14A】


atcaagcagaaggcaacaatgcaggatttggaacagaggcgtccccagttggaagaactcattaccgctgcccaaaatttgaaaaacaagaccagcaatcaagaggctagaacaatcattacggatcgaattgaaagaattcagaat
tagttcgtcttccgttgttacgtcctaaaccttgtctccgcaggggtcaaccttcttgagtaatggcgacgggttttaaacttttttgttctggtcgttagttctccgatcttgttagtaatgcctagcttaactttcttaagtcttc
I K Q K A T M Q D L E Q R R P Q L E E L I T A A Q N L K N K T S N Q E A R T I I T D R I E R I Q N
(起始密碼子框內)
外顯子51
外顯子52
外顯子53
atcaagcagaaggcaacaatgcaggatttggaacagaggcgtccccagttggaagaactcattaccgctgcccaaaatttgaaaaacaagaccagcaatcaagaggctagaacaatcattacggatcgaattgaaagaattcagaat
GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA
GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA
GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA
GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA
GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA
GCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAA

【圖14B】