(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102421902 A
(43)申请公布日 2012.04.18

(21)申请号 201080017915.3
(22)申请日 2010.02.25
(30)优先权数据
61/155,207 2009.02.25 US
(85)PCT申请进入国家阶段日
2011.10.21
(86)PCT申请的申请数据
PCT/US2010/025386 2010.02.25
(87)PCT申请的公布数据
WO2010/099301 EN 2010.09.02
(71)申请人 约翰·霍普金斯大学
地址 美国马里兰州
(72)发明人 N·L·克雷格
(74)专利代理机构 北京戈程知识产权代理有限公司 11314
代理人 程伟 韩文华
(51)Int.Cl.
C12N 15/12(2006.01)
C12N 15/63(2006.01)
C12N 5/10(2006.01)
A61K 48/00(2006.01)

权利要求书 2 页 说明书 24 页 附图 5 页

(54)发明名称

PIGGYBAC转座子变异及其使用方法

(57)摘要

本发明提供高活性piggyBac转座子,特别是来自粉斑夜蛾的高活性piggyBac转座子,它的转座频率比野生型高。本发明特征也在于合并缺陷的piggyBac转座子。可在基因转移系统中使用piggyBac转座子与转座酶,用于稳定导入核苷酸至细胞的DNA中。所述基因转移系统可用于方法例如但不限于基因治疗、插入突变或基因发现。

CN 102421902 A

1. 一种转座子，包括：一或多高活性 piggyBac 核苷酸序列以及保留转座子活性的变异、衍生物与片段。

2. 如权利要求 1 所述的转座子，其中相较于野生型 piggyBac 转座子，所述高活性 piggyBac 转座子具有较高程度的转座切割。

3. 如权利要求 1 所述的转座子，包括 2、3、4、5 或更多高活性 piggyBac 核苷酸序列以及保留转座子活性的变异、衍生物与片段。

4. 如权利要求 1 所述的转座子，其中所述高活性 piggyBac 核苷酸序列是来自 Noctuidae 属。

5. 如权利要求 1 所述的转座子，其中所述高活性 piggyBac 核苷酸序列是来自粉斑夜蛾物种。

6. 如权利要求 1 所述的转座子，包括选自 SEQ ID NO:34-SEQ ID NO:63 或 SEQ ID NO:70 SEQ ID NO:96 的核苷酸序列。

7. 如权利要求 6 所述的转座子，其中所述核苷酸序列编码选自 SEQ ID NO:3-SEQ ID NO:32 的胺基酸序列。

8. 一种转座子，包括：一或多合并缺陷的 piggyBac 核苷酸序列以及其变异、衍生物与片段。

9. 如权利要求 8 所述的转座子，其中相较于野生型 piggyBac 转座子，所述合并缺陷的 piggyBac 转座子具有较低的合并率。

10. 如权利要求 8 所述的转座子，其中所述合并缺陷的 piggyBac 核苷酸序列是来自 Noctuidae 属。

11. 如权利要求 8 所述的转座子，其中所述合并缺陷的 piggyBac 核苷酸序列是来自粉斑夜蛾物种。

12. 如权利要求 8 所述的转座子，包括选自 SEQ ID NO:67-SEQ ID NO:69 的核苷酸序列。

13. 如权利要求 12 所述的转座子，其中所述核苷酸序列编码选自 SEQ ID NO:64-SEQ ID NO:66 的胺基酸序列。

14. 如权利要求 2 或 9 所述的转座子，其中所述野生型 piggyBac 转座子包括相当于 SEQ ID NO:1 的核苷酸序列。

15. 如权利要求 1 或 8 所述的转座子，其中所述转座子可插入至细胞的 DNA 中。

16. 如权利要求 1 或 8 所述的转座子，其中所述转座子更包括标示蛋白质。

17. 如权利要求 1 或 8 所述的转座子，其中所述转座子插入质体中。

18. 如权利要求 17 所述的转座子，其中所述转座子更包括至少一部份的开放读架。

19. 如权利要求 17 所述的转座子，其中所述转座子更包括至少一表现控制区域。

20. 如权利要求 17 所述的转座子，其中所述表现控制区域是选自启动子、促进子或消音子。

21. 如权利要求 17 所述的转座子，其中所述转座子更包括启动子操作连结到至少一部份的开放读架。

22. 如权利要求 15 所述的转座子，其中所述细胞是来自于动物。

23. 如权利要求 22 所述的转座子，其中所述细胞是来自于脊椎动物或无脊椎动物。

24. 如权利要求 17 所述的转座子，其中所述脊椎动物是哺乳动物。

25. 一种基因转移系统，包括 ：根据权利要求 1 或权利要求 8 的转座子 ；以及 piggyBac 转座酶。

26. 如权利要求 25 所述的基因转移系统，其中所述 piggyBac 转座酶是来自 Noctuidae 属。

27. 如权利要求 26 所述的基因转移系统，其中所述 piggyBac 转座酶是来自粉斑夜蛾物种。

28. 如权利要求 27 所述的基因转移系统，其中所述 piggyBac 转座酶包括相当于 SEQ ID NO ：33 的胺基酸序列。

29. 如权利要求 25 所述的基因转移系统，其中所述 piggyBac 转座酶是哺乳动物 piggyBac 转座酶。

30. 如权利要求 25 所述的基因转移系统，其中所述转座子插入所述细胞的基因组中。

31. 如权利要求 30 所述的基因转移系统，其中所述细胞是来自动物。

32. 如权利要求 30 所述的基因转移系统，其中所述细胞是来自脊椎动物或无脊椎动物。

33. 如权利要求 32 所述的基因转移系统，其中所述脊椎动物是哺乳动物。

34. 一种包括权利要求 1 的所述转座子的细胞。

35. 一种包括权利要求 8 的所述转座子的细胞。

36. 一种医药组合物，包括 ：包括高活性 piggyBac 核苷酸序列的转座子与 piggyBac 转座酶，连同医药可接受的载体、赋形剂或运送体。

37. 一种用于导入外来 DNA 至细胞中的方法，包括 ：使所述细胞接触权利要求 25 所述的基因转移系统，因而导入外来 DNA 至细胞中。

38. 如权利要求 37 所述的方法，其中所述细胞是干细胞。

39. 一种套组，包括 ：包括高活性核苷酸序列的转座子以及用于导入 DNA 至细胞中的说明。

40. 如权利要求 39 所述的套组，其中所述高活性 piggyBac 核苷酸序列是来自 Noctuidae 属。

41. 如权利要求 39 所述的套组，其中所述高活性 piggyBac 核苷酸序列是来自粉斑夜蛾物种。

42. 如权利要求 39 所述的套组，包括选自 SEQ ID NO ：34-SEQ ID NO ：63 或 SEQ ID NO ：70-SEQ ID NO ：96 的核苷酸序列。

43. 一种套组，包括 ：包括合并缺陷的 piggyBac 核苷酸序列的转座子以及使用说明。

44. 如权利要求 43 所述的套组，包括选自 SEQ ID NO ：67-SEQ ID NO ：69 的核苷酸序列。

PIGGYBAC 转座子变异及其使用方法

背景技术

[0001] 本申请主张 2009 年 2 月 25 日申请的美国临时申请案案号 61/155,206 的优先权，所述美国临时申请案全文并入本申请做为参考。

[0002] 把 DNA 导入细胞的典型方法包含例如磷酸钙的 DNA 浓缩剂试剂、聚乙二醇、含脂质的试剂，例如磷脂质、多层脂质体，以及由病毒为媒介的策略。然而，所述方法可具有一些限制。例如，DNA 浓缩试剂与病毒媒介策略相关的大小限制。再者，在病毒策略中，可转染至细胞的核苷酸量有限制。并非所有方法都能促使进入的核苷酸插入细胞的核苷酸中，以及虽然 DNA 浓缩方法与含脂质试剂相对容易制备，但是插入核苷酸至病毒载体中需要大量劳力。病毒为媒介的策略可以是细胞形式或组织形式专一性，以及在体内使用病毒为媒介的策略可产生免疫问题。

[0003] 解决这些问题的一种合适工具是转座子。转座子 (transposons) 或转座因子 (transposable elements) 包含 (短的) 核苷酸序列，具有终端重复序列上游与下游。有活性的转座子编码的酵素会促使核苷酸的切除与插入至目标 DNA 序列中。

[0004] 转座因子代表许多真核基因组的实质部分。例如，～ 50％的人类基因组是衍生自转座因子序列，以及其它基因组，例如植物，可包含实质更高比例的转座因子衍生 DNA。转座因子典型分为两类，第 1 类与第 2 类。第 1 类表示为反转录转座子 (LINEs、SINEs、LTRs 与 ERVs)。第 2 类包含“剪与贴”DNA 转座子，由终端反转重复 (TIRs) 定性，以及被因子编码的转座酶 (transposase) 动员 (mobilized)。

[0005] 目前，在真核细胞中，辨识 10 个超级家族的剪与贴 DNA 转座子 (Feschotte and Pritham,2007)。

[0006] 虽然第 2 类在许多真核细胞中广为分布且具有活性，但是已经被认为在哺乳动物基因组中转座性地不活化。这结论主要是基于人类与老鼠基因组序列的初始分析。虽然两个物种有许多不同的 DNA 转座子，但没有近代活性的显示 (Lander et at.2001 ;Waterston et al.2002)。例如，在人类基因组中，有超过 300,000 个可辨识的 DNA 因子 (elements)，它们被分为 120 个家族，并且属于 5 个超级家族。这些因子大部分 (40 个家族 ；～ 98,000 复制 (copies)) 合并在至少 4-8 千万年，但是没有证据显示任何人类 DNA 转座子家族比 3 万 7 千年年轻 (Pace and Feschotte,2007)。

[0007] 可在实验室条件下模拟水平基因转换的自然过程。在植物中，Ac/Ds 与 Spm 家族已经时常转染至异质物种 (heterologous species) 中 (Osborne and Baker,1995 Curr. Opin.Cell Biol.7,406-413)。然而，在动物中，活性转座子系统从一物种转换至另一物种的障碍是转换的物种专一性，需要自然宿主产生的要素 (factors)。

[0008] 无脊椎动物与脊椎动物的转座子在模式生物体中皆有转基因 (transgenesis) 与插入突变 (insertional mutagenesis) 的潜力。特别是在相同物种中其它转座子系统的可得性开启遗传分析的新可能性。仍需要新方法导入 DNA 至细胞中，以及特别是促进不同大小的转座子有效插入细胞核苷酸或是插入 DNA 至细胞基因组的方法，同时比目前技艺中可得的架构 (construct) 具更有效率的转录 / 转译结果。

[0009] 发明概述

[0010] 如下详细说明，已知在许多昆虫、鼠、猪与包含人类的哺乳类动物细胞中，粉斑夜蛾 (Trichoplusia ni ;cabbage looper moth) 的 piggyBac 转座子是活性因子。本申请的发明人已经分离粉斑夜蛾 piggyBac 变异，比野生型 (wildtype) 的转座频率更高。高活性的转座子可用于基因转移统，把核苷酸稳定导入至细胞的 DNA 中。再者，本申请的发明人已经辨识合并缺陷 piggyBac 转移子 (integration defective piggyback transposons)。本发明的基因转移系统可用于例如但不限于基因治疗、插入突变或基因发现的方法中。

[0011] 因此，在第一方面，本发明特征在于包括一或多高活性 piggyBac 核苷酸序列的转座子，及其保留转座子活性的变异、衍生物与片段。在一实施例中，相较于野生型 piggyBac 转座子，高活性 piggyBac 转座子具有较高量的转座子切除 (transposon excision)。

[0012] 在另一实施例中，转座子包括 2、3、4、5 或更多的高活性 piggyBac 核苷酸序列以及其保留活性的变异、衍生物与片段的转座子。在另一实施例中，高活性 piggyBac 核苷酸序列是来自于 Noctuidae 属。在另一相关实施例中，高活性 piggyBac 核苷酸序列是来自于粉斑夜蛾物种。

[0013] 在另一实施例中，所述核苷酸序列是选自于 SEQ ID NO :34-SEQ ID NO :63 或 SEQ ID NO :70 SEQ ID NO :96。

[0014] 在另一实施例中，所述核苷酸序列编码的氨基酸序列是选自 SEQ ID NO :3-SEQ ID NO :32。

[0015] 在另一方面，本发明特征是包括一或多个合并缺陷的 (integration defective) piggyBac 核苷酸序列以及其变异、衍生物与片段的转座子。

[0016] 在一实施例中，相较于野生型 piggyBac 转座子，所述合并缺陷的 (integration defective)piggyBac 转座子具有较低的合并率 (rate of integration)。

[0017] 在另一实施例中，所述合并缺陷的 (integration defective)piggyBac 核苷酸序列是来自于 Noctuidae 属。在另一相关实施例中，所述合并缺陷的 (integration defective)piggyBac 核苷酸序列是来自于粉斑夜蛾物种。

[0018] 在一实施例中，所述核苷酸序列是选自于 SEQ ID NO :67-SEQ ID NO :69。

[0019] 在另一实施例中，所述核苷酸序列编码的氨基酸序列是选自 SEQ ID NO :64-SEQ ID NO :66。

[0020] 在另一实施例中，所述野生型 piggyBac 转座子包括对应于 SEQ ID NO :1 的核苷酸序列。

[0021] 在一些实施例中，高活性变异包括 SEQ ID NO :2 中胺基酸改变，选自于 G2C、Q40R、S3N、S26P、DOV、G165S、T43A、Q55R、T57A、S61R、I82V、I90V、S103P、S103T、N113S、M185L、M194V、S230N、R281G、M282V、G316E、P410L、I426V、Q497L、K501N、K565I、N505D、S573L、S509G、N570S、N538K、K575R、Q591P、Q591R、F594L。

[0022] 在上述方面的实施例中，所述转座子可插入至细胞的 DNA 中。

[0023] 在上述方面的另一实施例中，所述转座子可包括标示蛋白质 (marker protein)。

[0024] 在上述方面的另一实施例中，所述转座子是插入质体 (plasmid) 中。

[0025] 在一实施例中，所述转座子更包括至少一部份的开放读架 (open reading frame)。在另一实施例中，所述转座子更包括至少一表现控制区域。在另一实施例中，所述

表现控制区域是选自于启动子、促进子或消音子(silencer)。在另一相关实施例中,所述转座子更包括启动子操作连结至至少一部份的开放读架。

[0026] 在一实施例中,所述细胞是得自于动物。

[0027] 在另一实施例中,所述细胞是得自于脊椎或无脊椎动物。

[0028] 在另一实施例中,所述脊椎动物是哺乳动物。

[0029] 在一实施例中,本发明特征在于包括根据任一上述方面的转座子的基因转移系统;以及 piggyBac 转座酶。

[0030] 在一实施例中,piggyBac 转座酶是来自于 Noctuidae 属。在相关实施例中,所述 piggyBac 转座酶是来自粉斑夜蛾物种。

[0031] 在另一实施例中,所述 piggyBac 转座酶包括对应于 SEQ ID NO:33 的氨基酸序列。

[0032] 在另一实施例中,所述 piggyBac 转座酶是哺乳动物 piggyBac 转座酶。

[0033] 在一实施例中,所述转座子是插入细胞的基因组中。

[0034] 在另一实施例中,所述细胞是得自于动物。

[0035] 在另一实施例中,所述细胞是得自于脊椎或无脊椎动物。

[0036] 在另一实施例中,所述脊椎动物是哺乳动物。

[0037] 在一些实施例中,本发明特征也在于包括任依上述方面的转座子的细胞。

[0038] 在其它方面中,本发明特征在于包括转座子的医药组合物,所述转座子包括高活性的 piggyBac 核苷酸序列与 piggyBac 转座酶,以及医药可接受的载体、辅剂或运送体。

[0039] 本发明特征也在于导入外来 DNA 至细胞中的方法,所述方法包括使细胞接触上述方面的基因转移系统,因而导入外来 DNA 至细胞中。

[0040] 在一实施例中,所述细胞是干细胞。

[0041] 本发明特征也在于套组,所述套组包括:含有高活性 piggyBac 核苷酸序列以及导入 DNA 至细胞中的说明。

[0042] 在一实施例中,所述高活性 piggyBac 核苷酸序列是来自于 Noctuidae 属。在另一实施例中,所述高活性 piggyBac 核苷酸序列是来自于粉斑夜蛾物种。

[0043] 在另一实施例中,所述核苷酸序列是选自 SEQ ID NO:34-SEQ ID NO:63 或 SEQ ID NO:70-SEQ ID NO:96。

[0044] 在另一方面,本发明特征也在于套组,所述套组包括转座子与使用说明,所述转座子包括合并缺陷的 piggBac 核苷酸序列。

[0045] 在一实施例中,所述核苷酸序列是选自 SEQ ID NO:67-SEQ ID NO:69。

[0046] 由本申请的详细说明与权利要求书可清楚理解本申请的特征与优点。

[0047] 定义

[0048] 除非特别指出,否则本申请的实施是使用习知的化学、分子生物学、微生物学、重组 DNA 与免疫学的技术,这些都是熟知此技艺的人士能力范围之内。文献中已解释这些技术。例如,J.Sambrook,E.F.Fritsch 与 T.Maniatis,1989,Molecular Cloning:A Laboratory Manual,Second Edition,Books 1-3,Cold Spring Harbor Laboratory Press;Ausubel,F.M.et al.(1995 与 periodic supplements;Current Protocols in Molecular Biology,ch.9,13 与 16,John Wiley & Sons,New York,N.Y.);B.Roe,J.Crabtree 与 A.Kahn,1996,DNA Isolation and Sequencing:Essential Techniques,

John Wiley & Sons ;J.M.Polak and James O'D.McGee,1990, In Situ Hybridization : Principles and Practice ;Oxford University Press ;M.J.Gait(Editor),1984, Oligonucleotide Synthesis :A Practical Approach, IrI Press ;D.M.J.Lilley 与 J.E.Dahlberg,1992, Methods of Enzymology :DNA Structure Part A :Synthesis and Physical Analysis of DNA Methods in Enzymology, Academic Press ;Using Antibodies :A Laboratory Manual :Portable Protocol NO.I by Edward Harlow, David Lane, Ed Harlow(1999, Cold Spring Harbor Laboratory Press, ISBN 0-87969-544-7) ; Antibodies :A Laboratory Manual by Ed Harlow(Editor), David Lane(Editor)(1988, Cold Spring Harbor Laboratory Press, ISBN 0-87969-3,4-2),1855.Handbook of Drug Screening,由 Ramakrishna Seethala, Prabhavathi B.Fernandes 编辑 (2001, New York, N.Y., Marcel Dekker, ISBN 0-8247-0562-9) ;以及 Lab Ref :A Handbook of Recipes, Reagents,and Other Reference Tools for Use at the Bench,Jane Roskams and Linda Rodgers 编辑,2002,Cold Spring Harbor Laboratory,ISBN 0-87969-630-3。这些文献皆并入本申请做为参考。

[0049] 在本申请中,除非特别说明,否则以下语词的意义如下所述。

[0050] 在本申请中,"包括"与"具有"等类似语词可具有美国专利法中所定义的意义,以及可以代表"包含"与类似语词 ;"实质上由…组成"的意义如美国专利法中所描述,以及所述语词是开放式的,允许超出主张之外的新增内容存在,只要所主张的基本或新颖特征不受增加物的影响,但排除习知技艺的实施例。

[0051] 在本申请中,"合并缺陷的"语词是指相较于对应野生型转座子,以较低频率整合 (integrate) 至宿主基因组的转座子。在一些实施例中,本申请的转座子是通过习知的整合机制而整合。

[0052] 在本申请中,"核苷酸"或"聚核苷酸"语词是指单股或双股 DNA 与 RNA,以及其组合物。聚核苷酸可包含有不同功能的核苷酸序列,例如包含编码序列以及非编码序列,例如调节序列。聚核苷酸可直接得自于自然来源,或是可用重组、酵素或化学技术的辅助制备而得。聚核苷酸型态可以是线形或是环形。例如,聚核苷酸可以是载体的一部分或是片段。"编码序列"或是"编码区域"是编码多肽的聚核苷酸,以及当放在适当调节序列的控制下时,表现编码的多肽。编码区域的边界通常是由 5' 端的转译起始码与 3' 端的转译停止码决定。调节序列是核苷酸序列,调节其操作链接的编码区域的表现。调节序列的非限制范例包含启动子、转录起始位置、转译起始位置、转译停止位置、转录终结子 (包含例如聚腺苷酸化信号) 以及居中序列 (intervening sequence)(内含子 ;introns)。

[0053] 在本申请中,"操作连结"语词是指与另一核苷酸序列有功能关系的核苷酸序列。例如,如果编码序列是操作链接至启动子序列,则这通常是指所述启动子可启动编码序列的转录。操作连结是指连结的 DNA 序列典型是邻近的,以及视需要,连接两个蛋白质编码区域,读取架连续。由于当与启动子相距几千碱基时,促进子 (enhancer) 仍具有功能,以及内含子序列可以是不同长度,所以有些核苷酸序列可以是操作连结但并不相邻。

[0054] 在本申请中,"多肽"语词是指任何长度的胺基酸聚合物。因此,多肽的定义中包含例如胜肽、寡胜肽、蛋白质、抗体与酵素。这语词也包含多肽的表现后修饰,例如醣基化 (例如增加醣类)、乙酰基化、磷酸化等。

[0055] 在本申请中,“转座子”或“转座因子”是指多可从供应 (donor) 多核苷酸切割出来的核苷酸,所述供应多核苷酸例如载体,并且可合并至目标位置,例如细胞的基因组或是染色体外的 DNA。转座子包多核苷酸,所述多核苷酸包含由在转座子终端由 cis-acting 核苷酸序列侧边相接的核苷酸序列。如果至少一 cis-acting 核苷酸序列是位在核苷酸序列的 5’ 以及至少一 cis-acting 核苷酸序列是位在核苷酸序列的 3’,则核苷酸序列是被 cis-acting 核苷酸序列“侧边连接”。Cis-actging 核苷酸序列包含在转座子各端的至少一倒转重复 (在本申请中也是指倒转终端重复, ITR),较佳是哺乳动物 piggyBac 家族转座酶结合的转座子。在一些较佳实施例中,所述转座子是来自 Noctuidae 家族。在一些较佳实施例中,所述转座子是粉斑夜蛾 (Trichoplusia ni ;cabbage looper moth) 的 piggyBac 转座子。

[0056] 在本申请中,“粉斑夜蛾”是指 Noctuidae 蛾属的一员。

[0057] “分离的”多肽或多核苷酸是指已经从自然环境移除、使用重组技术产生的或是化学或酵素合成的多肽或多核苷酸。较佳地,本发明的多肽或多核苷酸是纯化的,亦即实质上没有任何其它的多肽或多核苷酸以及相关的细胞产物或其它杂质。

[0058] 在本申请中,“转座酶”语词是指催化从供应多核苷酸 (例如载体) 切除转座子以及后续合并转座子至目标细胞基因组或染色体外 DNA 的多肽。较佳地,转座酶结合倒转序列或直接重复。转座酶可以呈现为多肽。或者,所述转座酶呈现为多核苷酸,包含编码转座酶的编码序列。所述多核苷酸可以是 RNA,例如编码转座酶的 mRNA,或是 DNA,例如编码转座酶的编码序列。当转座酶呈现为编码转座酶的编码序列时,在本发明的一些方面中,所述编码序列可以存在包含转座子的相同载体上,亦即 in cis。在本发明的其它方面中,所述转座酶编码序列可存在第二载体上,亦即 in trans。在一些实施例中,所述转座酶是哺乳动物 piggyBac 转座酶。

附图说明

[0059] 图 1 是表格 (表 2) 说明在本发明高活性变异中,与野生型 (正常化至 1) 不同的胺基酸改变与折迭增加。

[0060] 图 2(A 与 B), (A) 部分说明野生型粉斑夜蛾转座子的核苷酸序列,对应于 SEQ ID NO :1 以及对应胺基酸序列 (SEQ ID NO :2,(B) 部分说明对应于野生型粉斑夜蛾转座酶的胺基酸序列,对应于 SEQ ID NO :33。

[0061] 图 3 说明切割高活性 piggyBac 突变株的辨识。

[0062] 图 4 说明合并缺陷的 piggyBac 转座子的胺基酸序列,对应于 SEQ ID NOs 64-66。

具体实施方式

[0063] 对于基因工程使用转座子的能力是高度取决于它可移动的频率。例如,如果 10 个子代中有 1 个具有转座子活动,则比 1000 个子代中有 1 个发生转座子活动更容易分离出所要的衍生型。本申请的发明人已经从粉斑夜蛾分离高活性 piggyBac 转座子,它比野生型具有更高的转座活性。在许多生物中使用插入突变与转殖,广用转座子例如 piggyBac 于基因组工程。来自粉斑夜蛾的 piggyBac 转座子已显示为在昆虫、老鼠、猪与包含人类的哺乳动物细胞中是活性因子。

[0064] 因此，本发明特征在于高活性 piggyBac 转座子。高活性 piggyBac 转座子是指比野生型 piggyBac 转座子具有更高频率的转座活动 (transposon event)。例如，在一些实施例中，高活性 piggyBac 转座子发生转座 0.5 倍、1 倍、2 倍、3 倍、4 倍、5 倍、6 倍、7 倍、8 倍、9 倍、10 倍、15 倍、20 倍、25 倍、30 倍、40 倍、45 倍、50 倍或更多。

[0065] 根据本发明的较佳实施例，高活性 piggyBac 转座子与转座酶结合，包含一对重复序列。在一些较佳实施例中，第一重复是典型位在核苷酸序列的上游，以及第二重复是典型位在核苷酸序列的下游。因此，第二重复与第一重复的序列相同，但是与第一重复的读取方向相反（互补双股序列的 5’与 3’端交换）。由于两重复是正好相反的重复序列，所以这些重复称为“倒转重复”(IRs)。在一些实施例中，在上述核苷酸序列的上游或下游，可发生多个重复。较佳地，位在上述核苷酸序列上游与下游的重复数量一样。在一些实施例中，所述重复是短的，在 10-20 碱基对之间，较佳是 15 个碱基对。

[0066] 本申请所述的重复 (IRs) 较佳是入细胞 DNA 的核苷酸序列。核苷酸序列可包含基因的全部或部分的开放读架（亦即部分的蛋白质编码基因）、一或多个表现控制序列（亦即核苷酸中的调节区域）或连同所有或部分的开放读架。较佳的表现控制序列包含但不限于启动子、促进子、边界控制因子、位置控制区域或消音子。在较佳实施例中，核苷酸序列包括启动子操作连接至至少一部份的开放读架。根据一些实施例，本发明的高活性转座子可较佳发生为线性转座子（从 5’端至 3’端延伸），可用作为线性片段或环化，例如在质体中。

[0067] 本发明特征在于高活性 piggyBac 核苷酸序列及其保留转座子活性的变异、衍生物与片段。在本发明的较佳实施例中，相较于野生型 piggyBac 转座子，高活性 piggyBac 转座子具有较高量的转座子切割 (transposon excision)。

[0068] 在一些较佳实施例中，所述高活性 piggyBac 核苷酸序列是来自于 Noctuidae 属。在其它实施例中，所述高活性 piggyBac 转座子核苷酸序列是来自粉斑夜蛾。

[0069] 本发明的较佳实施例是指编码本申请定义的高活性 piggyBac 转座子的核苷酸。

[0070] 熟知此技艺的人士可知由于遗传码的重复性 (degeneracy)，因此许多不同的多核苷酸可边码相同的多肽。除此之外，应理解熟知此技艺的人士可使用惯常技术制造核苷酸取代，而不影响本发明多核苷酸编码的多肽序列，用来反应任何特定宿主生物表现所述多肽的密码子使用。可用此技艺的任何方法修饰所述多核苷酸。可进行所述修饰，用来促进本发明多核苷酸的体内活性或生命期。

[0071] 可用重组、合成或熟知此技艺的人士知道的任何技术产生多核苷酸例如 DNA 多核苷酸。也可以用标准技术选殖产生。

[0072] 通常使用重组方法可产生较长的多核苷酸，例如使用 PCR（聚合酶链反应）选殖技术。这涉及制造一对引子（例如约 15 至 30 核苷酸）侧接要选殖的脂质目标序列区域，使引子接触动物或人类细胞的 mRNA 或 cDNA，在放大所要区域需要的条件下，进行聚合酶链反应，分离放大的片段（例如在洋菜胶体上纯化反应混合物）以及回收放大的 DNA。所述引子可被设计为包含合适的限制酵素辨识位置，因而被放大的 DNA 可被选殖到合适的选殖载体中。

[0073] 应理解本发明的多核苷酸可以只包含编码区域。然而，较佳是如果所述多核苷酸更包括一部分核苷酸在操作链接中，允许编码区域的有效转译。更佳是如果所述多核苷酸 (DNA 形式）更包括启动子在操作链接中，允许编码区域的转录，以及部分的核苷酸允许在

目标细胞中编码区域的有效转译。

[0074] 根据本发明，核苷酸典型包括核醣核酸，包含 mRNA、DNA、cDNA、染色体 DNA、染色体外 DNA、质体 DNA、病毒 DNA 或 RNA。在一些较佳实施例中，由于遗传密码的重复性，核苷酸较佳是选自编码高活性 piggyBac 转座子相同胺基酸序列的任何核酸序列。这些其它的核苷酸序列可在选择的宿主生物中造成编码融合蛋白质的改良表现。熟知此技艺的人士知道适当调整核苷酸序列的表格。通常用标准程序进行所述核苷酸与 / 或衍生物的制备与纯化（请见 Sambrook et al. 2001, Molecular Cloning : A Laboratory Manual, Cold Spring Harbor, N. Y.）。相较于自然的核苷酸序列，这些自然核苷酸的其它变异可具有一或多个插入、删除与 / 或取代的密码子。相较于转座子的自然核苷酸序列，这些序列变异较佳造成具有至少一胺基酸取代、删除与 / 或插入的合并缺陷与 / 高活性 piggyBac 转座子蛋白质。因此，本发明的核苷酸序列可编码修饰的（非自然的）转座子序列。再者，启动子或其它表现控制区域可操作链接编码本申请蛋白质的核苷酸，用来调节蛋白质表现的量或组织特性。

[0075] 在特定实施例中，粉斑夜蛾野生型 piggyBac 转座子包括对应于 SEQ ID NO : 1 的核苷酸序列。

[0076] SEQ ID NO : 1

[0077] 粉斑夜蛾 piggyBac 转座子

[0078]

CCCTAGAAAGATACTCTGCGTAAAATTGACGCATGCATTCTTGAAATATTGCTCTCTCTTTCTAAATAGCGCGAATCC
GTCGCTGTGCGTTTAGGACATCTCAGTCGCCGCTTGGAGCTCCCGTGAGGCGTGCTTGTCAATACGGTAAGTGTCACT
GATTTTGAACTATAACGACCGCGTGAGTCAAAATGACGCATGATTATCTTTTACGTGACTTTTAAGATTTAACTCATA
CGATAATTATATTGTTATTTCATGTTCTACTTACGTGATAACTTATTATATATATATTTTCTTGTTATAGATATCGTG
ACTAATATATAATAAAATGGGTAGTTCTTTTAGACGATGAGCATATCCTCTCTGCTCTTCTGCAAAGCGATGATGAGCT
TGTTGGTGAGGATTCTGACAGTGAAATATCAGATCACGTAAGTGAAGATGACGTCCAGAGCGATACAGAAGAAGCGTT
TATAGATGAGGTACATGAAGTGCAGCCAACGTCAAGCGGTAGTGAAATATTAGACGAACAAAATGTTATTGAACAACC
AGGTTCTTCATTGGCTTCTAACAAAATCTTGACCTTGCCACAGAGGACTATTAGAGGTAAGAATAAACATTGTTGGTC
AACTTCAAAGTCCACGAGGCGTAGCCGAGTCTCTGCACTGAATCATGTCAGATCTCAAAGAGGTCCGACGCGTATGTG
CCGCAATATATATGACCCACTTTTATGCTTCAAACTATTTTTTACTGATGAGATAATTTCGGAAATTGTAAAATGGAC
AAATGCTGAGATATCATTGAAACGTCGGGAATCTATGACAGGTGCTACATTTCGTGACACGAATGAAGATGAAATCTA
TGCTTTCTTTGGTATTCTGGTAATGACAGCAGTGAGAAAAGATAACCACATGTCCACAGATGACCTCTTTGATCGATC
TTTGTCAATGGTGTACGTCTCTCGTAATGAGTCGTGATGGTTTTGATTTTTGATACGATGTCTTAGAATGGATGACAA
AAGTATACGGCCCACACTTCGAGAAAACGATGTATTTACTCTGTTAGAAAAATATGGATCTCTTTATCCATCAGTG
CATACAAAATTACACTCCAGGGGCTCATTTGACCATAGATGAACAGTTACTTGGTTTTAGAGGACGGTGTCCGTTTAG
GATGTATATCCCAAACAAGCCAATTAAGTATGGAATAAAAATCCTCATGATGTGTGACAGTGGTACAAAGTATATGAT
AAATGGAATGCCTTATTTGGGAAGAGGAACACAGACCAACGGAGTACCACTCGGTGAATACTACGTGAAGGAGTTATC
AAAGCCTGTGCACGGTAGTTGTCGTAATATCACGTGTGACAATTGGTTCACCTCAATCCCTTTGGCAAAAAACTTACT
ACAAGAACCGTATAAGTTAACCATTGTGGGAACCGTGCGATCAAACAAACGCGAGATACCGGAAGTACTGAAAAACAG
TCGCTCCAGGCCAGTGGGAACATCGATGTTTTGCTTTGACGGACCCCTTACTCTCGTCTCATATAAACCGAAGCCAGC
TAAGATGGTATACTTATTATCATCTTGTGATGAGGATGCTTCTATCAACGAAAGTACCGTAAACCGCAAATGGTTAT
GTATTATAATCAAACTAAAGGCGCAGTGGACACGCTAGACCAAATTGTTCTTTGATGACCTGCAGTAGGAAGACGAA
TAGGTGGCCTATGGCATTATTGTACGGAATGATAAACATTGCCTGCATAAATTCTTTTATTATATACAGCCATAATGT
CAGTAGCAAGGGAGAAAAGGTTCAAAGTCGCAAAAAATTTATGAGAAACCTTTACATGAGCCTGACGTCATCGTTTAT
GCGTAAGCGTTTAGAAGCTCCTACTTTGAAGAGATATTTGCGCGATAATATCTCTAATATTTTGCCAAATGAAGTGCC
TGGTACATCGAATGACAGTACTGAAGAGCCAGTAACGAAAAAACGTACTTACTGTACTTACTGCCCTCTAAAATAAG
GCGAAAGGCAAATGCATCGTGCAAAAAATGCAAAAAAGTTATTTGTCGAGAGCATAATATTGATATGTGCCAAAGTTG
TTTCTGACTGACTAATAAGTATAATTTGTTTCTATTATGTATAAGTTAAGCTAATTACTTATTTTATAATACAACATG
ACTGTTTTTTAAAGTACAAAATAAGTTTATTTTTGTAAAAGAGAGAATGTTTAAAAGTTTTGTTACTTTATAGAAGAAA
TTTTGAGTTTTTGTTCTTTTTTAATAAATAAATAAACATAAATAAATAGTTTGTTGAATTTATTATTAGTATGTAAGT
GTAAATATAATAAAACTTAATATCTATTCAAATTAATAAATAAACCTCGATATACAGACCGATAAAACACATGCGTCA
ATTTTACGCATATTATCTTTAACGTACGTCACAATATGATTATCTTTCTAGGG

[0079] 在其它实施例中，粉斑夜蛾野生型 piggyBac 转座子胺基酸序列是相当于 SEQ ID NO : 2，如下所示：

[0080] MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNV
IEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISE
IVKWTNAEISLKRRESMTGATFRDTNEDEIYAFFGILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRC

LRMDDKSIRPTLRENDVFTPVRKIWDLF1HQCIQNYTPGAHLTIDEQLLGFRGRCPFRMYIPNKPSKYGIKILMMCD SGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNΓ Γ CDNWFTSIPLAKNLLQEPYKLTIVGTVRSNK REIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLDQM CSVMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLR DNISNILPNEVPGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

[0081] 在本申请中，在一些实施例中，本发明特征在合并缺陷的(integration defective)piggyBac转座子。合并缺陷是指相较于对应的野生型转座子，用较低频率合并至宿主基因组的转座子。在一些实施例中，本发明的转座子适用习知的合并机制进行合并。

[0082] 在一些实施例中，合并缺陷的piggyBac转座子是来自于野生型piggyBac序列SEQ ID NO:2。在一些实施例中，合并缺陷的piggyBac转座子包括SEQ ID NO:2选自R372A或K375A的改变。

[0083] 在一些较佳实施例中，合并缺陷的piggyBac转座子包括选自SEQ ID NO:64、SEQ ID NO:65或SEQ ID NO:66的氨基酸序列。

[0084] 在一些实施例中，SEQ ID NO:2胺基酸改变包括R372A，并且相当于SEQ ID NO:64。

[0085] SEQ ID NO:64

[0086] MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNV IEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISE IVKWTNAEISLKRRESMTGATFRDTNEDEIYAFFGILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRC LRMDDKSIRPTLRENDVFTPVRKIWDLF1HQCIQNYTPGAHLTIDEQLLGFRGRCPFRMYIPNKPSKYGIKILMMCD SGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQEPYKLTIVGTVASNKRE IPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLDQMCS VMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDN ISNILPNEVPGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICRE HNIDMCQSCF

[0087] SEQ ID NO:64编码的合并缺陷变异相当于SEQ ID NO:1中CGA变成GCA的核苷酸改变，以及相当于SEQ ID NO:67。

[0088] 在一些其它的实施例中，SEQ ID NO:2中胺基酸改变包括K375A以及相当于SEQ ID NO:65。

[0089] SEQ ID NO:65

[0090] MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNV IEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISE IVKWTNAEISLKRRESMTGATFRDTNEDEIYAFFGILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRC LRMDDKSIRPTLRENDVFTPVRKIWDLF1HQCIQNYTPGAHLTIDEQLLGFRGRCPFRMYIPNKPSKYGIKILMMCD SGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQEPYKLTIVGTVASNKRE IPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLDQMCS VMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDN ISNILPNEVPGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

[0091] SEQ ID NO:65编码的合并缺陷变异相当于SEQ ID NO:1中AAA变成GCA的核苷酸改变，以及相当于SEQ ID NO:68。

[0092] 在一些其它的实施例中，SEQ ID NO:2的氨基酸改变包括R372A、K375A以及相当

于 SEQ ID NO :66。

[0093] SEQ ID NO :66

[0094] MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNV IEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISE IVKWTNAEISLKRRESMTGATFRDTNEDEIYAFFGILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRC LRMDDKSIRPTLRENDVFTPVRKIWDLF1HQCIQNYTPGAHLTIDEQLLGFRGRCPFRMYIPNKPSKYGIKILMMCD SGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQEPYKLTIVGTVASNKRE IPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLDQMCS VMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDN ISNILPNEVPGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

[0095] SEQ ID NO :66 编码的合并缺陷变异相当于 SEQ ID NO :1 中 CGA 改变成 GCA/AAA 变成 GCA 的核苷酸改变，以及相当于 SEQ ID NO :69。

[0096] 在本申请中，本发明特征也在高活性 piggyBac 转座子。

[0097] 在一些较佳实施例中，所述高活性 piggyBac 转座子是从合并缺陷的 piggyBac 变异产生。亦即较佳是在合并缺陷 piggyBac 变异中制造一或多个突变。在其它实施例中，所述高活性 piggyBac 转座子是从野生型序列产生。亦即较佳是在野生型 piggyBac 转座子序列中制造一或多个突变。

[0098] 在实施例中，所述高活性 piggyBac 包括选自 SEQ ID NO :3-SEQ ID NO :32 的氨基酸序列。

[0099] 在其它实施例中，所述高活性 piggyBac 包括选自 SEQ ID NO :34-SEQ ID NO :63 的核苷酸序列。

[0100] 所述高活性 piggyBac 较佳可包括选自 SEQ ID NO :34-SEQ ID NO :63 的一或多个核苷酸序列。例如，所述高活性 piggyBac 较佳可包括选自 SEQ ID NO :34-SEQ ID NO :63 的 1、2、3、4、5 或更多的核苷酸序列。

[0101] 在一些实施例中，所述高活性变异包括 SEQ ID NO :2 的氨基酸改变，选自于 L15P、D19N/F395L、S31P/T164A、H33Y、E44K/K334R、E45G、C97R/T242I、S103P，R189K/G120G、R189R/D450N/R526R、M194T、M194V、S213SN436I、I221T、S373P、N384T、453S/N571S、T560A、N571S、S573A、S584P、M589V、M589V/D170D、S592G，F594L、Stop/WLESCN、Stop595ELESCN/H33H。

[0102] 在一些实施例中，所述高活性变异包括 SEQ ID NO :2 的氨基酸改变，选自于 G2C、Q40R、S3N、S26P、DOV、G165S、T43A、Q55R、T57A、S61R、I82V、I90V、S103P、S103T、N113S、M185L、M194V、S230N、R281G、M282V、G316E、P410L、I426V、Q497L、K501N、K565I、N505D、S573L、S509G、N570S、N538K、K575R、Q591P、Q591R、F594L。

[0103] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 L15P，以及相当于 SEQ ID NO :3。SEQ ID NO :4 编码的高活性变异相当于 SEQ ID NO :1 中 GAC 变成 AAC/UUU 变成 CUU 的核苷酸改变，以及相当于 SEQ ID NO :35 或 71。

[0104] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 S31P/T164A 以及相当于 SEQ ID NO :5。SEQ ID NO :5 编码的高活性变异相当于 SEQ ID NO :64 或 65 中 UCA 变成 CCA/ACA 变成 GCA 的核苷酸改变，以及相当于 SEQ ID NO :36 或 72。

[0105] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 H33Y，以及相当于 SEQ ID NO :6。SEQ ID NO :6 编码的高活性变异相当于 SEQ ID NO :64 或 65 中 CAC 变成 UAC 的核苷酸变异，以及相当于 SEQ ID NO :37 或 73。

[0106] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 E44K/K334R，以及相当于 SEQ ID NO :7。SEQ ID NO :7 编码的高活性变异相当于 SEQ ID NO :1 中 GAA 变成 AAA/AAG 变成 AGG 的核苷酸改变，以及相当于 SEQ ID NO :38 或 74。

[0107] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 E45G，以及相当于 SEQ ID NO :8。SEQ ID NO :8 编码的高活性变异相当于 SEQ ID NO :1 中 GAA 变成 GGA 的核苷酸改变，以及相当于 SEQ ID NO :39 或 75。

[0108] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 C97R/T242I，以及相当于 SEQ ID NO :9。SEQ ID NO :9 编码的高活性变异相当于 SEQ ID NO :1 中 UGU 变成 CGU/ACU 变成 AUU 的核苷酸改变，以及相当于 SEQ ID NO :40 或 76。

[0109] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 S103P，以及相当于 SEQ ID NO :10。SEQ ID NO :10 编码的高活性变异相当于 SEQ ID NO :1 中 UCC 变成 CCC 的核苷酸改变，以及相当于 SEQ ID NO :41 或 77。

[0110] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 R189K/G120G，以及相当于 SEQ ID NO :11。SEQ ID NO :11 编码的高活性变异相当于 SEQ ID NO :1 中 AGA 变成 AAA/GGU 变成 GGC 的核苷酸改变，以及相当于 SEQ ID NO :42 或 78。

[0111] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 R189R/D450N/R526R，以及相当于 SEQ ID NO :12。SEQ ID NO :12 编码的高活性变异相当于 SEQ ID NO :1 中 AGA 变成 AGG/GAC 变成 AAC 的核苷酸改变，以及相当于 SEQ ID NO :43 或 79。

[0112] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 M194T，以及相当于 SEQ ID NO :13。SEQ ID NO :13 编码的高活性变异相当于 SEQ ID NO :1 中 AUG 变成 ACG 的核苷酸改变，以及相当于 SEQ ID NO :44 或 80。

[0113] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 M194V，以及相当于 SEQ ID NO :14。SEQ ID NO :14 编码的高活性变异相当于 SEQ ID NO :1 中 AUG 变成 GUG 的核苷酸改变，以及相当于 SEQ ID NO :45 或 81。

[0114] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 S213S/V436I，以及相当于 SEQ ID NO :15。SEQ ID NO :15 编码的高活性变异相当于 SEQ ID NO :1 中 AGU 变成 AGC 的核苷酸改变，以及相当于 SEQ ID NO :46 或 82。

[0115] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 I221T，以及相当于 SEQ ID NO :16。SEQ ID NO :16 编码的高活性变异相当于 SEQ ID NO :1 中 AUA 变成 ACA 的核苷酸改变，以及相当于 SEQ ID NO :47 或 83。

[0116] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 S373P 在 M6+ 之间，以及相当于 SEQ ID NO :17。SEQ ID NO :17 编码的高活性变异相当于 SEQ ID NO :1 中 UCA 变成 CCA 的核苷酸改变，以及相当于 SEQ ID NO :48 或 84。

[0117] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 N384T，以及相当于 SEQ ID NO :18。SEQ ID NO :18 编码的高活性变异相当于 SEQ ID NO :1 中 AAC 变成 ACC 的核苷酸改变，以及相当于 SEQ ID NO :49 或 85。

[0118] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 C453S/N571S，以及相当于 SEQ ID NO :19。SEQ ID NO :19 编码的高活性变异相当于 SEQ ID NO :1 中 UGU 变成 AGU/AAU 变成 AGU 的核苷酸改变，以及相当于 SEQ ID NO :50 或 86。

[0119] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 T560A，以及相当于 SEQ ID NO :20。SEQ ID NO :20 编码的高活性变异相当于 SEQ ID NO :1 中 ACU 变成 GCU 的核苷酸改变，以及相当于 SEQ ID NO :51 或 87。

[0120] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 N571S，以及相当于 SEQ ID NO :21。SEQ ID NO :21 编码的高活性变异相当于 SEQ ID NO :1 中 AAU 变成 AAG 的核苷酸改变，以及相当于 SEQ ID NO :52 或 88。

[0121] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 S573A，以及相当于 SEQ ID NO :22。SEQ ID NO :22 编码的高活性变异相当于 SEQ ID NO :1 中 UCG 变成 GCG 的核苷酸改变，以及相当于 SEQ ID NO :53 或 89。

[0122] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 S584P，以及相当于 SEQ ID NO :23。SEQ ID NO :23 编码的高活性变异相当于 SEQ ID NO :1 中 UCU 变成 CCU 的核苷酸改变，以及相当于 SEQ ID NO :54 或 90。

[0123] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 M598V，以及相当于 SEQ ID NO :24。SEQ ID NO :24 编码的高活性变异相当于 SEQ ID NO :1 中 AUG 变成 GUG 的核苷酸改变，以及相当于 SEQ ID NO :55 或 91。

[0124] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 M589V/D170D，以及相当于 SEQ ID NO :25。SEQ ID NO :25 编码的高活性变异相当于 SEQ ID NO :1 中 ATG 变成 GUG/GAC 变成 GAU 的核苷酸改变，以及相当于 SEQ ID NO :56 或 92。

[0125] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 S592G，以及相当于 SEQ ID NO :26。SEQ ID NO :26 编码的高活性变异相当于 SEQ ID NO :1 中 AGU 变成 GGU 的核苷酸改变，以及相当于 SEQ ID NO :57 或 93。

[0126] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 F594L，以及相当于 SEQ ID NO :27。SEQ ID NO :27 编码的高活性变异相当于 SEQ ID NO :1 中 UUC 变成 TTA 的核苷酸改变，以及相当于 SEQ ID NO :58 或 94。

[0127] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 Stop/WLESCN，以及相当于 SEQ ID NO :28。SEQ ID NO :28 编码的高活性变异相当于 SEQ ID NO :1 中 TGA 变成 TGG 的核苷酸改变，以及相当于 SEQ ID NO :59 或 95。

[0128] 在一些实施例中，SEQ ID NO :64 或 65 胺基酸改变包括 Stop595ELESCN/H33H，以及相当于 SEQ ID NO :29。SEQ ID NO :29 编码的高活性变异相当于 SEQ ID NO :1 中 TGA 变成 GGA/CAU 的核苷酸改变，以及相当于 SEQ ID NO :60 或 96。

[0129] 在另一较佳实施例中，编码高活性 piggyBac 转座子的核苷酸是选自上述定义编码高活性 piggyBac 转座子的核苷酸序列，以及可在严厉条件下与上述定义的互补核苷酸序列杂交。在另一实施例中，编码高活性 piggyBac 转座子的核苷酸是选自上述定义编码高活性 piggyBac 转座子的核苷酸序列，以及可在严厉条件下与上述定义的互补核苷酸序列杂交。严厉条件例如在 0.5*SSC 中 30％ (v/v) 甲醛、0.1％ (w/v)SDS 在 42C 进行 7 小时。

[0130] 测量从载体切割转座子的分析、合并转座子至细胞的基因组或染色体外 DNA 以及

转座酶结合至倒转重复的能力描述在本申请中，并且是此技艺已知的 (Ivies et al.Cell, 91,501-510(1997) ;WO 98/40510(Hackett et al.) ;WO 99/25817(Hackett et al.), WO 00/68399(Mclvor et al.) 全部内容并入本申请作为参考。为了决定本发明转座子的转作频率，基本转座子活性被正常化至 100％，决定本发明的转座子的相对活性。较佳地，本发明的转座子转座频率增加倍数较佳大于基本线转座子至少约 50％、至少约 100％、至少约 200％，更佳为至少约 300％。较佳地，两个转座子（亦即基本线转座子与测试的转座子）皆在包含转座子的载体中侧接相同的核苷酸序列。

[0131] 本发明特征也在于蛋白质序列，显示与 SEQ ID NOs 3-32 其中任何一蛋白质序列至少 80％、较佳为至少 85％、更佳为至少 90％、再更佳为至少 95％，以及最佳为至少 98％序列相同。

[0132] 本发明特征也在于蛋白质序列，显示与 SEQ ID NOs 3-32 其中任何一蛋白质序列至少 80％、较佳为至少 85％、更佳为至少 90％、再更佳为至少 95％，以及最佳为至少 98％序列相同。

[0133] “相似度 (identity)”是指两个或多个蛋白质、核苷酸之间相同的程度，可使用已知的方法例如计算器为基础的序列比对 (basic local alignment search tool, S.F.Altschul et al.,J.MoI.Biol.215(1990),403-410)，比较这些序列而决定。所述方法包含但不限于GAG编程，包含GAP(Devereux,J.,et al.,Nucleic Acids Research 12(12) : 287(1984) ;Genetics Computer Group University of Wisconsin,Madison,(WI)) ;BLASTP or BLASTN,and FASTA(Altschul,S.,et al.,J.MoI.Biol.215 :403-410)(1999))。此外，可使用 Smith Waterman 运算，决定两序列之间的相同程度。

[0134] 根据本发明，功能衍生物较佳维持哺乳动物转座酶的生物功能，亦即转座酶活性、核苷酸序列的切除以及关于切除序列插入特定目标序列的活性。根据本发明，功能衍生物可包括高活性变异的一或多个胺基酸插入、删除与/或取代，例如相当于 SEQ ID NOs 3-32。

[0135] 本申请所述的胺基酸取代较佳是保守的胺基酸取代，不会改变转座子或转座酶蛋白质的生物活性。保守的胺基酸取代特征在于具有特别大小或特征的胺基酸可取代另一胺基酸，特别是例如在本发明蛋白质的区域与催化活性或 DNA 结合活性无关。例如，其它的胺基酸序列可包含含有保守改变的胺基酸序列，不会明显改变转座酶的活性或结合特征。胺基酸序列的取代可选自其它胺基酸种类的成员。例如，非极性（疏水性）胺基酸包含丙胺酸、白胺酸、异白胺酸、缬胺酸、脯胺酸、苯丙胺酸、色胺酸与酪胺酸。极性的自然胺基酸包含甘胺酸、丝胺酸、羟丁胺酸、半胱胺酸、酪胺酸、天冬酰胺与麸胺酰胺。正电（碱性）胺基酸包含精胺酸、赖胺酸与组胺酸。负电（酸性）胺基酸包含天冬胺酸与麸胺酸。这些改变并不会实质影响由聚丙烯酰胺凝胶电泳或等电点决定的分子量。特别较佳的保守取代包含但不限于 Lys 取代 Arg，反之亦然，维持正电荷 ;Glu 取代 Asp，反之亦然，维持负电荷 ;Ser 取代 Thr，因而维持自由的 --OH ;以及 Gln 取代 Asn 维持自由的 NH_2。

[0136] 胺基酸插入与取代较佳是发生在不影响空间结构、或催化中心或 piggyBac 转座子或转座酶的结合区域的序列位置。可使用例如 CD 光谱（圆二色光谱）(Urry,1985, Absorption,circular Dichroism and ORD of Polypeptides,in :Modern Physical Methods in Biochemistry,Neuberger et al.(Ed.),Elsevier,Amsterdam) 侦测插入或删除造成的空间结构改变。产生具有不同于自然序列的取代的氨基酸序列所用的合适方法

例如美国专利号 4,737,462、4,588,585、4,959,314、5,116,943、4,879,111 与 5,017,691，这些公开内容全文并入本申请作为参考。为了避免生理降解，可将其它功能衍生物稳定化。例如可用［贝塔］胺基酸取代酰胺键，稳定蛋白质骨架而得到上述稳定化作用。

[0137] 根据本发明的一些较佳实施例，本发明的转座子可更包括标示蛋白质 (marker protein)。再者，在一些较佳实施例中，核苷酸序列可以是任何的重组蛋白质，例如此技艺中任何已知的蛋白质。例如，核苷酸序列编码的蛋白质可以是标示蛋白质，例如绿荧光蛋白质 (GFP)、蓝荧光蛋白质 (BFP)、光活化的 GFP(PA-GFP)、黄偏移绿荧光蛋白质（黄 GFP)、黄荧光蛋白质 (YFP)、强化的黄荧光蛋白质 (EYFP)、青绿色荧光蛋白质 (CFP)、强化的青绿色荧光蛋白质 (ECFP)、单体红荧光蛋白质 (mRFP1)、点火荧光蛋白质 (KFP1)、水母蛋白、自发荧光蛋白质 (AFP)、或是荧光蛋白质 JRed、TurboGFP、PhiYFP 与 PhiYFP-m、tHc-Red(HcRed-Tandem)、PS-CFP2 以及 KFP-Red（全部都是商业上可购买得到），或是其它合适的荧光蛋白质氯霉素乙酰基转移酶 (CAT)。蛋白质更可选自生长荷尔蒙，例如在转殖动物中促进生长，或是选自贝塔－半乳糖苷酶 (lacZ)、荧光酵素 (LUC) 以及类胰岛素生长因子 (IGFs)、阿法－抗－胰蛋白酶、红血球生成素 (EPO)、血液凝结系统的因子 VIII 与 XI、LDL- 受体、GATA-1 等。核苷酸序列更可以是自杀基因，编码例如细胞凋零死亡相关的酵素，以及包含 AIF、Apaf 例如 Apaf-1、Apaf-2、Apaf-3 或 APO-2(L)、APO-3(L)、Apopain、Bad、Bak、Bax、Bcl-2、Bcl-x.sub.L、Bcl-x.sub.S、bik、CAD、Calpain、Caspases 例如 Caspase-1、Caspase-2、Caspase-3、Caspase-4、Caspase-5、Caspase-6、Caspase-7、Caspase-8、Caspase-9、Caspase-10、Caspase-11、 或 Granzyme B、ced-3、ced-9、Ceramide、c-Jun、c-Myc、CPP32、crm A、Cytochrome c、D4-GDP-DI、Daxx、CdR1、DcR1、DD、DED、DISC、DNA-PK.sub.CS、DR3、DR4、DR5、FADD/MORT-1、FAK、Fas、Fas-ligand CD95/fas（受体）、FLICE/MACH、FLIP、Fodrin、fos、G-Actin、Gas-2、Gelsolin、glucocorticoid/glucocorticoid 受体、granzyme AfB、hnRNPs C1/C2、ICAD、ICE、JNK、Lamin A/B、MAP、MCL-I、Mdm-2、MEKK-I、MORT-I、NEDD、NF-.sub..kappa.B、NuMa、p53、PAK-2、PARP、Perforin、PITSLRE、PKC.delta.、pRb，Presenilin、prICE、RAIDD、Ras、RIP、Sphingomyelinase、SREBPs、单纯疱疹病毒的 thymidine kinase、TNF-alpha、TNF-alpha 受体、TRADD、TRAF2、TRAIL-R1、TRAIL-R2、TRAIL-R3、Transglutaminase、U1 70kDa snRNP、YAMA 的基因。

[0138] 本发明的 piggyBac 转座子较佳是结合 piggyBac 转座酶，相较于习知技艺的方法有一些优点，所述习知方法是关于病毒与反转录病毒的方法。例如，不同于原病毒的插入，通过 in trans 提供转座酶活性，转座子插入可以被（再）动员化 ((re)mobilized)。因此，例如除了进行耗时的微注射之外，也可根据本发明在新的位置产生转座子插入。

[0139] 本发明的 piggyBac 转座子结合上述定义的 piggyBac 转座酶蛋白质可被转染至细胞中，成为蛋白质或包含 mRNA 的核糖核酸、成为 DNA，例如成为染色体外 DNA，包含但不限于游离 DNA、成为质体 DNA 或是成为病毒核苷酸。再者，编码转座酶蛋白质的核苷酸可被转染至细胞中，成为核苷酸载体，例如质体，或是成为基因表现载体，包含病毒载体。因此，核苷酸可以是环形或是线形。本申请使用的载体是指质体、病毒载体或是黏质体 (cosmid) 可合并编码本发明的转座酶蛋白质或转座子的核苷酸。“编码序列”或是“开放读架”是指当放在适当调节序列的控制下十，可在体内被转录与 / 或转译成为多肽的核苷酸区域。

[0140] 编码转座酶蛋白质的 DNA 可稳定插入细胞的基因组中或是插入在体中，用于持续

或是可诱导型的表现。当转座酶蛋白质转染至细胞中或插入载体中成为核苷酸时，转座酶编码序列较佳是操作连结至启动子。有许多启动子可使用，包含但不限于持续启动子、组织特异性激活子、可诱导的激活子与类似物。启动子是调节信息，在细胞中结合 RNA 聚合酶，起始下游 (3' 方向) 编码序列的转录。DNA 序列操作连结至表现控制序列，例如启动子，表现控制序列控制且调节 DNA 序列的转录与转译。"操作连结"包含在要表现的 DNA 序列前具有适当的开始信息 (例如 ATG)，以及维持正确的读架，在表现控制序列的控制下，允许 DNA 序列的表现，产生所要的蛋白质产物。编码高活性 puggyBac 转座子的核苷酸序列例如 SEQ ID NO :3-SEQ ID NO :32 或本申请所述的其它高活性变异。除了上述保守改变会需要改变转座子编码核苷酸序列 (如本申请所述) 之外，还有编码高活性 piggyBac 转座子蛋白质的其它 DNA 或 RNA 序列。这些 DNA 或 RNA 序列具有与高活性 piggyBac 转座子蛋白质相同的胺基酸序列，但是利用三个字母密码子的重复性指定特定的胺基酸。例如，此技艺中已知可交替使用不同的特并 RNA 密码子 (对应于 DNA 密码子，由 U 取代 T)，编码特定的胺基酸。

[0141] DNA 与蛋白质的操作方法是此技艺中已知的，并且文献具有详细说明，例如 Sambrook et al，(1989)Molecular Cloning :A Laboratory Manual，Cold Spring Harbor Laboratory Press 或 Ausubel，R.M.，ed.(1994).Current Protocols in Molecular Biology。

[0142] 基因转移系统

[0143] 本发明特征也在于基因转移系统，包括本发明的转座子、本申请所述的合并缺陷的转座子或高活性 piggyBac 转座子，以及本申请所述的 piggyBac 转座酶。

[0144] 如上所述，piggyBac 转座酶蛋白质较佳是辨识高活性 piggyBac 转座子上的重复 (例如 IR)。因此，本发明的基因转移系统较佳是包括两个成员 :本申请描述的转座酶，以及本申请描述的高活性转座子或合并缺陷的转座子。较佳地，在一些实施例中，转座子具有至少两个重复 (例如 IRs)。当两个成员放在一起时，提供转座子活性，以及允许转座子重新定位。使用时，转座酶结合至所述重复，以及启动介于中间的 DNA 序列插入至细胞的 DNA 中，如下所述。

[0145] 在其它实施例中，所述基因转移系统包括上述定义本发明的 piggyBac 转座子结合 piggyBac 转座酶蛋白质 (或编码 piggyBac 转座酶的核苷酸，在细胞中提供活性)。这结合较佳造成插入核苷酸序列至细胞的 DNA 中。或者，通过非同质重组与许多可再现的机制，可能插入本发明的转座子至细胞的 DNA 中。使用此基因转移系统，本发明的转座子可用于基因转移。

[0146] 在一些较佳实施例中，所述基因转移系统使用 piggyBac 转座酶蛋白质，媒介插入高活性 piggyBac 转座子至不同细胞形式与不同物种的 DNA 中。较佳地，所述细胞包含适合本申请的任何细胞，包含但不限于动物细胞或来自细菌、真菌 (例如酵母菌) 或植物的细胞。较佳的动物细胞可以是脊椎动物或无脊椎动物的细胞。例如，较佳的脊椎动物细胞包含来自哺乳类动物的细胞，包含但不限于囓齿动物，例如大鼠，有蹄类哺乳动物，例如牛或山羊、绵羊、猪，或人类细胞。

[0147] 在其它实施例中，这些细胞，特别是来自上述定义哺乳动物的细胞可以是分化多能的 (pluripont)(亦即这些细胞的子代可以分化为数种有限的细胞型式，例如血液干细胞或其它干细胞)，以及分化全能 (totipotent) 细胞 (亦即这些细胞的子代可变成器官中

的任何细胞型式，例如胚胎干细胞）。使用这些细胞有利于稳定表现转座酶或得到已经转染本发明基因转移系统成员的许多细胞。此外，也可考虑卵母细胞、蛋以及一或多胚胎细胞，作为本发明基因转移系统稳定转染的目标。

[0148] 在本发明的一些实施例中，所述细胞是干细胞。

[0149] 接收本发明 piggyBac 转座子与 / 或 piggyBac 转座酶蛋白质的细胞且可插入转座子至细胞 DNA 中，所述细胞可包含但不限于淋巴细胞、肝细胞、神经细胞、肌肉细胞、许多血液细胞以及器官的不同细胞，胚胎干细胞、体干细胞例如血液前驱细胞、胚胎、受精卵、精细胞（其中有些细胞可在体外操作）。

[0150] 在一些其它的实施例中，如果本发明的 piggyBac 转座子，例如高活性 piggyBac 转座子，与细胞内的 piggyBac 转座酶蛋白质接触，则接受本申请转座子的细胞 DNA 可包含所要转染细胞（如上所述）中的任何 DNA。例如，所述 DNA 可以是细胞基因组的一部分，或是染色体外，例如游离基因、质体、环行或线形 DNA 片段。插入的典型目标例如双股 DNA。

[0151] 本申请的基因转移系统成员，亦即 piggyBac 转座酶蛋白质（蛋白质或是由本申请所述核苷酸编码的）以及本发明的 piggyBac 转座子可被转染至细胞中，较佳是转染至上述定义的细胞中，更佳是转染至相同的细胞中。

[0152] 这些成员的转染更可顺序发生或是平行发生。例如，piggyBac 转座酶蛋白质或是其编码核苷酸可同时转染至上述定义的细胞中，或是在转染哺乳动物 piggyBac 转座子之后，转染 piggyBac 转座酶蛋白质或是其编码核苷酸至上述定义的细胞中。或者，在转染 piggyBac 转座酶蛋白质或其编码核酸序列之前、同时或之后，可转染所述转座子至上述定义的细胞中。如果是平行转染，较佳是在投与之前，两个成员在个别的配方中与 / 或直接比次混合，避免转染前发生转座。此外，可重复进行投与基因转移系统的至少一成员，例如投与至少一、二或多剂量的成员。

[0153] 对于上述任何的转染反应，可用此技艺中已知的合适方式调配所述基因转移系统，或成为本申请所述的医药组合物或套组。

[0154] 在其它较佳实施例中，使用粒子轰击、电子穿孔、微注射、结合具有含脂质载体的成员，例如阳离子脂质载体、DNA 浓缩剂（例如磷酸钾、聚离胺酸或聚乙亚胺）以及插入所述成分（亦即插入核苷酸至病毒载体中，以及使病毒载体接触细胞），所述基因转移系统的成员较佳可被转染至一或多细胞中。当使用病毒载体时，所述病毒载体可包含所述技艺中已知的任何病毒载体，选自反转录病毒载体、腺病毒载体或是腺相关病毒载体。

[0155] 如上所述，编码 piggyBac 转座酶蛋白质的核苷酸可以是 RNA 或是 DNA。同样地，编码 piggyBac 转座酶蛋白质的核苷酸或本发明的转座子可转染至细胞中，成为线形片段或是环形片段，较佳是质体或是重组的病毒 DNA。

[0156] 再者，编码 piggyBac 转座酶蛋白质的核苷酸较佳是稳定的或暂时插入细胞的基因组中，促使细胞中 piggyBac 转座酶的暂时或延长表现。

[0157] 上述的基因转移系统代表修饰过的非病毒 DNA 媒介的基因转移。例如，使用病毒作为基因治疗剂会使遗传工程设计受限于病毒基因组的大小、结构与表现调节。本申请所述的非病毒载体主要是由合成的起始材料产生，因而比病毒载体更容易制造。相较于病毒试剂，非病毒试剂比较不可能成为免疫原，可重复投与。非病毒载体比病毒载体更稳定，因而比病毒载体更适合医药配方与应用。除此之外，本发明的基因转移系统是非病毒基因转

移系统，便于插入至 DNA，以及明显改善稳定基因转移的频率。

[0158] 本发明更提供有效的方法用于产生转殖动物，包含把本发明的基因转移系统用于动物。转职 DNA 并未有效率地插入染色体中。只有大约百万分之一的外来 DNA 分子会插入细胞的基因组中，通常数次切割周期进入发展。结果，大部分的转殖动物是 mosaic(Hackett et al.(1993).The molecular biology of transgenic fish.In Biochemistry and Molecular Biology of Fishes(Hochachka & Mommsen,eds)Vol.2,pp.207-240)。因此，必须培养已经传递转殖 DNA 的胚胎所培育的动物，直到分析到配子有插入的外来 DNA 存在。许多转殖动物由于位置效应而无法表现转殖基因。需要简单、可靠的程序，在一个细胞阶段，导引早期插入外来 DNA 至动物的染色体中。本发明的系统协助填补这个需要。

[0159] 在一些较佳实施例中，本发明的基因转移系统可用于产生转殖动物，在动物的一个或多个细胞中，携带特定标示或表现特定蛋白质。通常，此技艺中已知产生转殖动物的方法，以及合并本发明基因转移系统至这些技术中并不需要过度实验，例如有许多产生转殖动物的方法用于研究，或用于蛋白质生产，包含但不限于 Hackett et al.(1993,supra)。此技艺中已有产生转殖动物的其它方法的描述（例如 M.Markkula et al.Rev.Reprod.,1,97-106(1996) ;R.T.Wall et al.,J.Dairy ScL,80,2213-2224(1997))，J.C.Dalton,et al.(Adv.Exp.Med.Biol.,411,419-428(1997))and H.Lubon et al.(Transfus.Med.Rev.,10,131-143(1996))。

[0160] 在另一实施例中，本发明特征在于由本申请方法产生的转殖动物，较佳是使用本申请所述的基因转移系统。例如，转殖动物较佳可包含由所述基因转移系统插入动物基因组中的核苷酸序列，因而所述转殖动物可产生其基因产物，例如蛋白质。在转殖动物中，此蛋白质较佳是由细胞分离的产物，例如本发明的蛋白质可产生在奶、尿、血液或蛋中。可使用启动子，促进奶、尿、血液或蛋中的表现，以及这些启动子包含但不限于酪蛋白启动子、老鼠尿蛋白启动子、贝塔球蛋白启动子以及卵白蛋白启动子。已经使用在细胞中产生蛋白质的其它方法，产生重组的生长荷尔蒙、重组的胰岛素以及许多其它重组蛋白质。编码这些或其它蛋白质的核苷酸可被插入至本发明的转座子中，并且转染至细胞中。上述本发明转座子的有效转染至细胞 DNA 中发生在当有哺乳动物的 piggyBac 转座酶蛋白质存在时。当所述细胞是组织的一部分或是转殖动物的一部分时，可得到大量的重组蛋白质。

[0161] 可从脊椎动物与无脊椎动物选择转殖动物，例如鱼、鸟、哺乳动物，包含但不限于例如大鼠或老鼠的啮齿动物、例如牛或山羊、绵羊、猪的有蹄类哺乳动物，或人类。

[0162] 本发明更提供基因治疗的方法，包括导入所述基因转移系统是本申请所述细胞中。因此，本申请中所述本发明的 piggyBac 转座酶较佳包括基因，用来提供基因治疗至细胞或生物。较佳地，所述基因是在组织特异性激活子、广布启动子或在细胞中表现基因需要的一或多个其它表现控制区域的控制下。目前，测试许多基因用于许多基因治疗，包含但不限于 CFTR 基因用于囊性纤维变性、腺苷去胺酶 (ADA) 用于免疫系统疾病、因子 IX 与介白素 -2(IL-2) 用于血液细胞疾病、阿法 -1- 抗胰蛋白酶用于肺病，以及肿瘤坏死因子 (INF) 与多抗药性 (MDR) 蛋白质用于癌症治疗。可在已知的基因数据库，例如 GenBank，获得这些以及许多人类或动物特异性的基因序列，包含编码标示蛋白质与许多重组蛋白质的基因序列。

[0163] 本发明基因转移系统用于基因治疗的优点是不受限于重复之间的中间核苷酸序

列大小。使用哺乳动物 piggyBac 转座子蛋白质，可插入细胞 DNA 中的核苷酸序列大小没有已知的限制。

[0164] 在特定的较佳实施例中，对于基因至料以及本发明的其它目的，可用许多方法将所述基因转移系统转染至细胞中，例如微注射、脂质媒介的方法或是病毒媒介的方法。例如，当使用微注射时，对于本发明转座子的中间序列的大小有非常小的限制。同样地，脂质媒介的方法没有实质的大小限制。然而，导入基因转移系统至细胞中的其它方法，例如病毒媒介的方法，可限制位于重复之间核苷酸序列的长度。

[0165] 因此，在一些实施例中，本申请所述的基因转移系统可通过病毒传递至细胞，所述病毒包含反转录病毒（例如慢病毒）、腺病毒、腺关联病毒、泡疹病毒以及其它病毒。有一些传送机制的潜在组合，可用于高活性 piggyBac 转座子部分，包含由终端重复侧接的转基因以及编码转座酶的基因。例如，可在相同的重组病毒基因组中包含转座子与转座酶基因；单一感染传送基因转移系统的两部分，因而转座酶的表现而后导引从重组病毒基因组切除转座子，后续插入细胞的染色体中。在另一范例中，可通过病毒与 / 或非病毒系统的组合，例如包含脂质的试剂，分别递送转座酶与转座子。在这些例子中，转座子与 / 或转座酶基因可由重组病毒递送。在每个例子中，表显的转座酶基因导引转座子从载体 DNA（病毒基因组）释出，插入至染色体 DNA 中。

[0166] 在本发明的一些较佳实施例中，本发明的 piggyBac 转座子可用于插入突变，较佳是经过辨识突变基因。DNA 转座子特别是具有不同于习知技艺的优点，例如关于病毒与反转录病毒方法。例如，不同于原病毒插入，可通过 in trans 提供转座酶活性，再定位 (remobilize) 转座子插入。因此，取代进行耗时的微注射，根据本发明可交连 (crossing) 储存的转殖，转座子系统的上述两成员，本发明的转座子与本发明的转座酶，在新位置产生转座子插入。在较佳实施例中，为了突变生殖细胞，基因转移系统是针对实验动物的生殖系。或者，转座酶表现可使用针对特定组织或器官的许多特定启动子。此外，突变的转座子在插入位置外的再定位可用于分离反突变体 (revertants)，以及如果转座子切除与侧接 DNA 的删除相关，则本发明基因转移系统可用于产生删除突变。再者，由于转座子是由 DNA 组成，并且可维持在简单质体中，本发明的转座子与本发明基因转移系统的使用相较于高感染性的反转录病毒更为安全且更简单操作。转座酶活性提供形式可以是上述定义的 DNA、mRNA 或蛋白质。

[0167] 在另一实施例中，本发明也提供有效率的系统用于基因发现，例如基因组映像 (genome mapping)，使用本发明的基因转移系统，导入上述定义本发明的 piggyBac 转座子至基因中。在一范例中，高活性 piggyBac 转座子结合 piggyBac 转座酶蛋白质或编码 piggyBac 转座酶蛋白质的核苷酸转染至细胞中。在一些较佳实施例中，转座子较佳包括位在至少两重复之间的核苷酸序列，其中所述重复结合 piggyBac 转座酶蛋白质，以及其中所述转座子插入有 piggyBac 转座酶蛋白质的细胞的 DNA 中。在一些较佳实施例中，核苷酸序列包含标示蛋白质，例如 GFP 与限制核酸内切酵素辨识位置。在插入之后，细胞 DNA 被分离且用限制核酸内切酵素消化。例如，如果核酸内切酵素辨识位置是 6- 碱基辨识位置，以及限制核酸内切酵素是使用 6- 碱基辨识序列，则所述细胞 DNA 被切为平均约 4000-bp 片段。这些片段可被转职或是可加入连结子 (linker) 至被消化的 (digested) 片段端，提供 PCR 引子的互补序列。当加入连结子时，PCR 反应使用来自连结子的引子以及转座子中所述重

复结合的引子，用于放大片段。而后定序放大的片段，以及侧接直接重复的DNA用于搜寻计算机数据库，例如GenBank。

[0168] 使用基因转移系统用于上述方法，例如基因发现与/或基因标定，通过使用转座子作为插入突变剂，允许关于生长与发育的基因辨识、分离与定性，或是控制生长与发育的转录调节序列的辨识、分离与定性。

[0169] 在本发明的另一实施例中，本发明提供在细胞中动员(mobilizing)核苷酸序列的方法。根据此方法，所述高活性piggyBac转座子插入细胞的DNA中，如本申请所述。高活性piggyBac蛋白质或编码piggyBac转座酶蛋白质的核苷酸被转染至细胞中，以及所述蛋白质可动员（亦即移动）所述转座子从细胞的DNA内第一位置至所述细胞DNA内的第二位置。细胞的DNA较佳是基因组DNA或是外染色体DNA。本发明的方法允许转座子从基因组的一位置移动至基因组的另一位置，或是例如从细胞中的直体移动至所述细胞的基因组。

[0170] 在另一实施例中，本发明基因转移系统也可作为RNA干扰技术方法的一部分。RNA干扰(RNAi)是与mRNA或细胞的基因/基因片段互补的外来、双股RNA(dsRNA)导入细胞中，专一性结合至特定的mRNA与/或基因，因而破坏基因表现。已经证实这个技术在果蝇、秀丽隐杆线虫(Caenorhabditis elegans)、植物中有效，以及最近也正在哺乳动物细胞培养中有效。为了使用这个技术结合本发明，本发明的转座子较佳包含编码小的干扰RNA(siRNA)的短hairpin表现匣(short hairpin expression cassettes)，与细胞的mRNA与/或基因/基因片段互补。这些siRNA具有较佳长度为20至30核苷酸，更佳的长度为20至25核苷酸，以及最佳长度为21至23核苷酸。所述siRNA可针对任何mRNA与/或基因，编码上述定义的任何蛋白质，例如致癌基因(oncogene)。使用哺乳动物piggyBac转座子用于插入siRNA载体至宿主基因组中，在体外或体内提供长期表现siRNA，因而可长期消音(silencing)特定基因产物。

[0171] 诱导式分化多能的干细胞(iPS)

[0172] 在一些较佳实施例中，本发明可包含再计划的载体(reprogramming vector)，所述载体包含多作用子的(polycistronic)表现匣，所述表现匣包括转录调节因子、一或多个再计划因子以及一或多个本申请所述的高活性piggyBac转座子。较佳地，编码再计划因子是Sox、Oct、Nanog、Klf4或c-Myc。通常，干细胞是未分化的细胞，可以进行连续的成熟功能细胞。例如，血液前驱干细胞可成为任何不同形式最终分化的血液细胞。胚胎干(ES)细胞是来自于胚胎并且是分化多能的，因而具有发展成为任何器官或组织形式的能力，或是至少有潜力成为完整的胚胎。

[0173] 诱导式分化多能的干细胞通常简称为iPS细胞或是IPSCs，是分化多能的干细胞形式，是人工取得自非分化多能的细胞，通常是成人体细胞，插入一些基因。例如在表现一些干细胞基因、蛋白质与染色质甲基化图案、分裂时间、胚样体形成、畸胎瘤形成、存活的嵌合体形成以及潜力与分化能力，诱导式分化多能的干细胞被认为和例如胚胎干细胞的自然分化多能干细胞相同，但是整体内容与自然分化多能干细胞的关系仍被评估。2006年首次从老鼠细胞产生iPS(Takahashi et al.，2006，并入本申请作为参考)，以及2007年从人类细胞产生iPS细胞(Takahashi et al.，2007，并入本申请作为参考)。在干细胞研究中已经被引用成为重要的成就，使得研究人员能得到分化多能的干细胞，这对于研究是重要的，并且潜在具有治疗使用，而没有胚胎使用的争议问题。

[0174] “再计划 (reprogramming)”是指在培养或在体内，能使形成至少一种新细胞形式的能力增加的程序，所述能力增加是指相较于相同条件下没有再计划的结果。更特别地，再计划是指使体细胞具有分化多能潜力的程序。这是指如果在再计划之前没有所述子代可实质形成，而在足够的增生之后，具有新细胞型式表现型特征的子代有可测量的比例；否则，具有新细胞型式特征的比例测量上多于在再计划之前。在一些条件下，在增加偏好的程度上，具有新细胞型式特征的子代比例可至少约为 1％、5％、25％或更多。

[0175] 胚胎干 (ES) 细胞是得自于早期胚胎的分化多能干细胞。ES 细胞最先是在 1981 年建立，自 1989 年开始用于产生基因剔除鼠 (knockout mice)。在 1998 年，人类 ES 细胞建立，目前成为再生医学应用。

[0176] 不同于 ES 细胞，组织干细胞具有有限的分化潜力。组织干细胞存在组织中特定位置，并且具有未分化的细胞内结构。因此，组织干细胞的分活多能性较低。组织干细胞具有较高的核 / 质比例，以及具有一些细胞内胞器。大部分的组织干细胞具有低的分化多能性、长细胞周期以及分化能力超过个体生命。组织干细胞依取得位置分类，例如皮肤系统、消化系统、骨髓系统、神经系统等。皮肤系统的组织干细胞包含表皮干细胞、毛囊干细胞等。消化系统的组织干细胞包含胰脏（共同）干细胞、肝脏干细胞等。骨髓系统的组织干细胞包含血液前驱干细胞、间叶干细胞等。神经系统的干细胞包含神经干细胞、视网膜干细胞等。

[0177] “诱导式分化多能的干细胞”通常简称为 iPS 细胞或 iPSC，是指从非分化多能细胞人工取得的分化多能干细胞型式，典型是从成人体细胞或最终分化的细胞，例如纤维母细胞、血液前驱细胞、单核细胞、神经、表皮细胞等，插入一些基因，称为再计划因子。

[0178] iPS 细胞的产生对于用于诱导的基因是很重要的。以下的因子或其结合可用于本发明中。在一些方面中，编码 Sox 与 Oct（较佳为 Oct3/4）的核苷酸可包含在再计划因子中。例如，再计划因子可包括编码 Sox2、Oct4、Nanog 与 Lin-28 的表现匣，或是编码 Sox2、Oct4、Klf4 与 c-myc 的表现匣。编码这些再计划因子核苷酸可包括在相同的表现匣、不同表现匣、相同再计划载体或不同再计划载体中。

[0179] 已经辨识 Oct-3/4 与一些 Sox 基因家族成员 (Sox1，Sox2，Sox3 与 Sox15) 成为参与诱导过程中的重要转录调节因子，若缺乏则无法诱导。然而，已经辨识包含一些 Klf 家族 (Klf 1，Klf2，Klf4，and Klf5) 成员、Myc 家族 (C-myc，L-myc 与 N-myc)、Nanog 与 LIN28 的其它基因可增加诱导效率。Oct-3/4(Pou5f1) 是八聚体 (Oct) 转录因子家族之一，并且对于维持分化多能性具有重要作用。在 Oct-3/4+ 细胞例如胚叶细胞 (blastomeres) 与胚胎干细胞中，缺乏 Oct-3/4 会造成自发性的滋胚层 (trophoblast) 分化，以及 Oct-3/4 的存在造成胚胎干细胞的分化多能与分化潜力。“Oct”家族的许多其它基因包含 Oct-3/4 相关的 Oct1 与 Oct6，不能引发诱导，因而在诱导程序中呈现 Oct-3/4 排斥性。

[0180] 类似 Oct-3/4，基因的 Sox 家族与维持分化多能性相关，虽然相较于 Oct-3/4，也与多分化潜能与单分化潜能干细胞相关，Oct-3/4 只表现在多分化潜能干细胞中。虽然 Sox2 适用于诱导的初始基因 (Yamanaka et al. (2007)，Jaenisch et al. (1988) and Yu et al. (2007))，已经发现 Sox 家族的其它基因也在诱导程序中作用。Sox1 产生 iPS 细胞的效率如同 Sox2，以及基因 Sox3、Sox15 与 Sox18 也产生 iPS 细胞，但效率较低。

[0181] 在胚胎干细胞中，促进分化多能需要至少一 Oct 成员例如 Oct-3/4 与至少一 Sox 成员例如 Sox2。虽然 Yu et al. (2007) 已经报到 Nanog 可能是产生 iPS 细胞的因子之一，以

及 Nanog 促进再计划效率与剂量相关，但 Yamanaka et al.(2007) 报导诱导不需要 Nanog。

[0182] Klf 家族的 Klf4 最初是由 Yamanaka 等人辨识，以及由 Jaenisch 等人 (1988) 确认为产生老鼠 iPS 细胞的因子，并且由 Yamanaka 等人 (2007) 说明为产生人类 iPS 细胞的因子。然而，Thompson 等人报导人类 iPS 细胞的产生不需要 Klf4，以及 Klf4 无法产生人类 iPS 细胞。发现 Klf2 与 Klf4 可产生 iPS 细胞，以及相关基因 Klf1 与 Klf5 也可以产生 iPS 细胞，但效率较低。

[0183] Myc 家族基因是与癌症相关的前致癌基因。Yamanaka 等人与 Jaenisch 等人 (1988) 说明 c-myc 是参与产生老鼠 iPS 细胞的因子，以及 Yamanaka 等人说明它是参与产生人类 iPS 细胞的因子。然而，Thomson 等人与 Yamanaka 等人 (2007) 报导产生人类 iPS 细胞不需要 c-myc。在诱导 iPS 细胞中使用“myc”家族困扰在于 iPS 细胞最终要做为临床治疗，而 c-myc 诱导的 iPS 细胞转殖的老鼠有 25％会发展致命的畸胎瘤。已经辨识 N-myc 与 L-myc 取代 c-myc 会诱导相似的效率。

[0184] 医药组合物

[0185] 本发明更提供医药组合物，包含 piggyBac 转座酶作为蛋白质或是由核苷酸编码，与 / 或高活性 piggyBac 转座子，或本申请所述的基因转移系统，包括 piggrBac 转座酶作为蛋白质或是由核苷酸编码，结合高活性 piggyBac 转座子。所述医药组合物的提供可选择性结合医药上可接受的载体、辅剂或或运送体。在本申请中，根据本发明，医药上可接受的载体、辅剂或运送体是指无毒性载体、辅剂或运送体，不会破坏配方成分的医药活性。本发明组合物使用的医药上可接受的载体、辅剂或运送体包含但不限于离子交换剂、氧化铝、硬脂酸铝、卵磷脂、血清蛋白，例如人类血清白蛋白、缓冲物质例如磷酸盐、甘胺酸、山梨酸、己二烯酸钾、饱和蔬菜脂肪酸的部分甘油酯混合物、水、盐或电解质、氯化钠、锌盐、硅酸胶、三硅酸镁、聚乙烯四氢咯酮、纤维素为基础的物质、聚乙二醇、羧甲基纤维素钠、聚丙烯酸酯、蜡、聚乙烯 - 聚氧丙烯 - 块聚合物、聚乙烯二醇以及羊毛脂。

[0186] 本发明的医药组合物可口服、非口服投与、吸入喷洒、局部、直肠、鼻腔、脸颊、阴道投与或是经由植入的储存器。本申请所使用的非口服一词包含皮下、血管内、肌肉内、关节内、滑液内、胸骨内、脑脊髓膜内、肝内、intralesional 以及颅内注射或灌注技术。较佳地，所述医药组合物是口服、腹膜内或血管内投与。本发明医药组合物的无菌注射形式可为水溶液或油性悬浮液。可根据此技艺中已知的技术调配这些悬浮液，使用合适的分散或湿润剂以及悬浮试剂。所述无菌可注射的制备也可为无菌可注射的溶液或无毒非口无可接受的稀释剂或溶剂中的悬浮液，例如在 1,3- 丁二醇中的溶液。在这些可接受的运送体与溶剂中，可使用水、Ringer’s 溶液以及等渗透压的氯化钠溶液。除此之外，习知使用无菌、定性油作为溶剂或悬浮媒介。

[0187] 为此目的，可使用任何温和的定性油，包含合成的单或二甘油酯。例如十八烯酸与其甘油酯衍生物的脂肪酸可用于制备注射剂，可以是自然医药与液或悬浮液也可包含长链的醇类稀释剂或分散剂，例如羧甲基纤维素或类似的分散剂，通常使用在医药可接受剂量形式的配方中，包含乳剂与悬浮液。为了配方的目的，也可使用其它通常使用的接口活性剂，例如 Tweens、Spans 与其它乳化剂或生物可用性的促进剂，通常使用在医药可接受的固体、液体或其它剂量形式的制造中。

[0188] 本发明医药可接受的组合物可用任何口服可接受的剂量形式口服投与，包含但不

限于胶囊、锭剂、水溶液悬浮液或溶液。例如使用锭剂作为口服使用，通常使用的载体包含乳糖与玉米淀粉。典型添加例如硬脂酸镁的润滑剂。对于口服投与胶囊形式，有用的稀释剂包含乳糖与干燥的玉米淀粉。当需要水溶液悬浮液作为口服使用时，活性成分结合乳化与悬浮剂。若需要，也可增加一些增甜、气味或颜色剂。

[0189] 或者，本发明医药可接受的组合物可以是栓剂形式用于直肠投与。可混合本发明的基因转移系统或其成分与合适的非刺激赋形剂制备这些组合物，所述非刺激赋形剂在室温为固体而在直肠温度则为液体，因此会在直肠融化，释放所述药物。所述材料包含可可油、蜜蜡以及聚乙烯二醇。

[0190] 本发明医药可接受的组合物也可局部投与，特别是当处理目标包含可局部使用处理的区域或器官，包含眼疾、皮肤或较低肠道。合适的局部配方可制备用于这些区域或器官中。

[0191] 对于局部使用，所述医药可接受的组合物可调配在合适的药膏中，包含本发明的基因转移系统或其成分，悬浮或溶解在一或多载体中。本发明局部投予的成分载体包含但不限于矿物油、液体矿脂、白矿脂、聚乙烯二醇、聚氧乙烯、聚氧乙烯成分、乳化蜡与水。或者，所述医药可接受的组合物可调配在合适的乳液或霜中，包含所述活性成分悬浮或溶解在一或多医药可接受的载体中。合适的载体包含但不限于矿物油、去水山梨醇单硬脂酸酯、聚山梨醇酯 60、十六醇酯蜡、鲸腊醇、2- 辛基十二烷醇、苯甲醇与水。

[0192] 对于眼部使用，所述医药可接受的组合物可调配成为在等渗透压、pH 调整无菌盐水中的粒子化悬浮液或较佳为在等渗透压、pH 调整无菌盐水中的溶液，具有或没有例如苯扎氯铵的防腐剂。或者，对于眼部使用，所述医药可接受的组合物可调配在膏药中，例如矿脂。

[0193] 本发明医药可接受的组合物也可鼻腔投与或吸入。根据医药配方技艺中已知的技术，制备此组合物，以及所述组合物可制备为盐水中的溶液、使用苯甲醇或其它合适的防腐剂、吸收促进剂增进生物可用性、氟碳化物与 / 或其它习知的溶解或分散剂。

[0194] 依待处理的宿主与特定的投与形式，本发明成分的量可结合载体材料产生单剂量形式的组合物。必须注意对于任何特定病患，特定剂量与处理方法取决于不同因子，包含使用特定成分的活性、年龄、体重、一般健康、性别、饮食、投与时间、排泄率、药物结合与治疗医师的评估以及待处理疾病的严重程度。组合物中本发明成分的量取决于组合物中的特定成分。

[0195] 所述医药组合物较佳适合治疗疾病，由基因缺陷造成的特定疾病，例如囊性纤维变性、高胆固醇血症、血友病、免疫缺陷，包含 HIV、亨汀顿氏舞蹈病、阿法 - 抗 - 胰蛋白酶缺陷、直肠癌、黑色素瘤、肾癌、淋巴癌、急性骨髓白血病 (AML)、急性淋巴白血病 (ALL)、慢性骨髓性白血病 (CML)、慢性淋巴白血病 (CLL)、肠胃道肿瘤、肺癌、神经胶瘤、甲状腺癌、乳癌、前列腺癌、肝癌、病毒诱发的肿瘤例如乳突瘤病毒诱导的癌症（例如子宫颈癌）、腺病毒癌、疱疹病毒诱发的肿瘤（例如 Burkitt 淋巴癌、EBV 诱发的 B 细胞淋巴癌）、B 型肝炎诱发的肿瘤（肝细胞癌）、HTLV-1 与 HTLV-2 诱发的淋巴癌、肺癌、咽癌、肛门直肠癌、神经胶母细胞瘤、淋巴癌、直肠癌、星状细胞瘤、脑癌、胃癌、视网膜母细胞瘤、基底细胞癌、脑转移、髓母细胞瘤、阴道癌、胰脏癌、睾丸癌、黑色素瘤、膀胱癌、霍奇综合症、脑膜瘤、Schneeberger’s disease、支气管癌、脑下垂体癌、蕈状肉芽肿、咽喉癌、乳癌、神经纤维瘤、

眼睑肿瘤、Burkitt 淋巴癌、喉癌、胸腺瘤、子宫体癌、骨癌、非霍奇淋巴癌、尿道癌、CUP 综合症、寡树突神经胶质瘤、外阴癌、肠癌、食道癌、小肠肿瘤、颅咽瘤、卵巢癌、肝癌、血癌或皮肤或眼睛的癌症。

[0196] 套组

[0197] 本发明特征在于套组，包括 piggyBac 转座酶作为蛋白质或由核苷酸编码，与 / 或高活性 piggyBac 转座子；或本申请所述的基因转移系统，包括 piggyBac 转座酶作为蛋白质或由核苷酸编码，结合高活性 piggyBac 转座子；视需要结合医药可接受的载体、赋形剂或运送体，以及视需要结合使用说明。

[0198] 本发明套组的任何成分可连续或平行投与以及 / 或转染至细胞中，例如在投与且 / 或转染本发明高活性转座子之前、同时或之后，piggyBac 转座酶蛋白质或其编码核苷酸可被投与以及 / 或转染至细胞中。或者，在转染 piggyBac 转座酶蛋白质或其编码核苷酸之前、同时或之后，所述高活性 piggyBac 转座子可转染至细胞中。如果是平行转染，较佳是两种成分提供在不同的配方且 / 或在投与前直接彼此混合，避免在转染前转座。除此之外，可依时间模式投与且 / 或转染套组的至少一成分，例如投与多次剂量的此成分。

[0199] 以下范例提供完整的揭露内容给熟知此技艺的人士，描述如何制造与使用本发明的分析、筛选以及治疗方法，并且不限制本发明的范围。

[0200] 范例

[0201] 范例 1. 辨识合并缺陷的 piggyBac 变异

[0202] 本申请的实验使用产生红荧光蛋白质的 Cherry 基因，描述筛选与辨识切割 piggyBac 转座子。把 piggyBac 转座子放入所述基因中，将所述基因去活化，因而细胞不是红色。然而，piggyBac 切割回复所述基因，造成产生红色荧光蛋白质。因此，红色群落颜色增加代表突变切割更佳。

[0203] 使用突变 PCR，转殖到酵母菌的表现载体中，完成许多突变转座酶基因的收集。在洋菜盘上生长包含各个突变的群落，而后用红色荧光检视。图 4 说明已经分离的切割高活性。

[0204] 在一些较佳实施例中，合并缺陷的 piggyBac 包括野生型 piggyBac 序列中的胺基酸改变，相当于 SEQ ID NO :2。较佳地，所述胺基酸改变是 R371A、R373A 或 R371A、R373A。

[0205] 在较佳实施例中，合并缺陷的 piggyBac 相当于 SEQ ID NO :64、SEQID NO :65 或 SEQ ID NO :66 的胺基酸序列。

[0206] 范例 2. 辨识高活性变异

[0207] 使用合并缺陷的 piggyBac 突变作为起始点，本申请的发明人已经辨识高活性 piggyBac 转座子突变。酵母菌切割分析如 Mitra R. 等人所述 (piggyBac can bypass DNA synthesis during cut and paste transposition. EMBO J. Apr 9 ;27(7) :1097-109. Epub 2008 Mar 20) 用于辨识高活性突变株。使用突变 PCR 与侧接 ORF 引子作为表现建构，突变化 piggyBac ORF，而后共同转型 (co-transformation) PCR 产物与有间隙的 piggyBac 质体至包含 ura- 至 ura+ 匣的酵母菌分析株中回复转型，其中转座子切割造成 ura+ 群落形成。在 SC-Trp-His 盘上回复转型株之后，把群落悬浮在水中，点在缺少尿嘧啶的盘上，来辨识切割。在这些点测试中比较突变转型株与野生型的 ura+ 群落数目，辨识潜在的高活性变异株。而后纯化每一个高活性候选株，并且再次定量分析切割。而后悬浮包含来自确认高活

性的piggyBac基因的质体DNA，来辨识piggyBac基因突变以及造成的胺基酸改变。胺基酸改变与对应的核苷酸改变如表1所示，如下：

[0208]

L15P	CUG 变成 CCG
D19N/F395L	GAC 变成 AAC/UUU 变成 CUU
S31P/T164A	UCA 变成 CCA/ACA 变成 GCA
H33Y	CAC 变成 UAC
E44K/K334R	GAA 变成 AAA/AAG 变成 AGG
E45G	GAA 变成 GGA
C97R/T242I	UGU 变成 CGU/ACU 变成 AUU
S103P	UCC 变成 CCC
R189K/G120G	AGA 变成 AAA/GGU 变成 GGC
R189R/D450N/R526R	AGA 变成 AGG/GAC 变成 AAC
M194T	AUG 变成 ACG
M194V	AUG 变成 GUG
S213S/V436I	AGU 变成 AGC

[0209]

I221T	AUA 变成 ACA
在M6+之间的 S373P	UCA 变成 CCA
N384T	AAC 变成 ACC
C453S/N571S	UGU 变成 AGU/AAU 变成 AGU
T560A	ACU 变成 GCU
N571S	AAU 变成 AAG
S573A	UCG 变成 GCG
S584P	UCU 变成 CCU
M589V	AUG 变成 GUG
M589V/D170D	ATG 变成 GUG/GAC 变成 GAU
S592G	AGU 变成 GGU
F594L	UUC 变成 TTA
停止/WLESCN	TGA 变成 TGG
停止 595ELESCN/H33H	TGA 变成 GGA/CAC 变成 CAU

[0210] 图1中的表2显示从野生型（正常化至1）胺基酸改变与折迭增加的一些高活性突变株。

[0211] 高活性突变可发生在转座酶的任何位置，以及预期会发现许多更高活性的piggyBac变异。这些变异可能是单个胺基酸改变或是多个胺基酸改变。可使用酵母菌分析作为筛选辨识变异。可使用整个基因的PCR突变以及目标突变，使用较小的piggyBac片段或寡核苷酸的突变，针对已辨识具有高活性的突变区域。

[0212] 范例3. 使用高活性转座子的诱导型分化多能干细胞

[0213] 在一些实施例中，可使用高活性piggyBac转座子，产生使用最少基因组的诱导型分活多能干细胞。特别地，使用Oct3/4、Sox2、Klf4与c-myc作为最小的基因组。Takahashi et al.(Cell,131,861-872,November 30,2007)全文并入本申请作为参考，教导使用Oct3/4、Sox2、Klf4与c-myc，从人类皮肤纤维母细胞产生诱导型分化多能干细胞(iPS)的方法。

[0214] 其它范例

[0215] 由上述说明，显然可对本申请所述的内容进行变化与修饰而因应不同的使用与条件。所述实施例仍落于本申请权利要求的范围内。

[0216] 本申请可变因素的任何定义中组件表列包含所列示组件的单一组件或组合的定义。本申请实施例包含作为单一实施例或任何实施例的组合或任何部份的组合。

[0217] 本申请内容中所有专利与公开案皆并入本申请作为参考。

表 2

PiggyBac 高活性 -折迭增加

野生型	1
G2C, Q40R	8.7
S3N	23.7
S26P	12.3
I30V, G165S	16.4
T43A	15.9
Q55R	8.1
T57A	5.3
S61R	23.9
I82V	6.8
I90V	13.7
S103P	12.3
S103T	5.0
N113S	8.87
M185L	7.3
M194V	9.1
S230N	11.7
R281G	12.6
M282V	15.7
G316E	8.4
P410L	13.8
I426V	10.7
Q497L	8.1
K501N, K565I	59
N505D, S573L	10.9
S509G, N570S	11.1
N538K	16.4
K575R	16.9
Q591P	156
Q591R	31.8
F594L	42.2

图 1

SEQ ID NO: 1

>粉斑夜蛾 piggyBac 转座子

CCCTAGAAAGATAGTCTGCGTAAAATTGACGCATGCATTCTTGAAATATTGCTCTCTCTTTCTAAATAGCGCGAATCCGTCGCTGTGCGT

TTAGGACATCTCAGTCGCCGCTTGGAGCTCCCGTGAGGCGTGCTTGTCAATGCGGTAAGTGTCACTGATTTTGAACTATAACGACCGCGT

GAGTCAAAATGACGCATGATTATCTTTTACGTGACTTTTAAGATTTAACTCATACGATAATTATATTGTTATTTCATGTTCTACTTACGT

GATAACTTATTATATATATTTTCTTGTTATAGATATCGTGACTAAATATATAAATGGGTAGTTCTTTAGACGATGAGCATATCCT

CTCTGCTCTTCTGCAAAGCGATGACGAGCTTGTTGGTGAGGATTCTGACAGTGAAATATCAGATCACGTAAGTGAAGATGACGTCCAGAG

CGATACAGAAGAAGCGTTTATAGATGAGGTACATGAAGTGCAGCCAACGTCAAGCGGTAGTGAAATATTAGACGAACACAAATGTTATTGA

ACAACCAGGTTCTTCATTGGCTTCTAACAAAATCTTGACCTTGCCACACAGAGGACTATTAGAGGTAAGAATAAACATTGTTGGTCAACTTC

AAAGTCCACGAGGCGTAGCCGAGTCTCTGCACTGCCTCATGTCAGATCTCAAAGAGGTCCGACGCGTATGTGCCGCAATATATGACCC

ACTTTTATGCTTCAAACTATTTTTTACTGATGAGATAATTTCGGAAATTGTAAAATGGACAAATGCTGAGATATCATTGAAACGTCGGGA

ATCTATGACAGGTGCTACATTTCGTGACACGAATGAAGATGAAATCTATGCTTTCTTTGGTATTCTGGTAATGACAGCAGTGAGAAAAGA

TAACCACATGTCCACAGATGACCTCTTTGATCGATCTTTGTCAATGGTGTACGTCTCTGTAATGAGTCGTGATCGTTTTGATTTTTTGAT

ACGATGTCTTAGAATGGATGACAAAAGTATACGGCCCACACTTCGAGAAAACGATGTATTTACTCCTGTTAGAAAAATATGGGATCTCTT

TATCCATCAGTGCATACAAAATTACACTCCAGGGGCTCATTTGACCATAGATGAACAGTTACTTGGTTTTAGAGGACGGTGTCCGTTTAG

GATGTATATCCCAAACAAGCCAAGTAAGTATGGAATAAAAATCCTCATGATGTGTGACAGTGGTACAAAGTATATGATAAATGGAATGCC

TTATTTTGGGAAGAGGAACACAGACCAACGGAGTACCACTCGGTGAAGGAGTTATCAAAGCCTGTGCACGGTAGTTGTCG

TAATATTACGTGTGACAATTGGTTCACCTCAATCCCTTTGGCAAAAAACTTACTACACAAGAACCGTATAAGTTAACCATTGTGGGAACCGT

GCGATCAAACAAACGCGAGATACCGGAAGTACTGAAAAACAGTCGCTCCAGGCCAGTGGGAACATCGATGTTTTGTTTTGACGGACCCCT

TACTCTCGTCTCATATAAACCGAAGCCAGCTAAGATGGTATACTTATTATCATCTTGTGATGAGGATGCTTCTATCAACGAAAGTACCGG

TAAACCGCAAATGGTTATGTATTATAATCAAACTAAAGGCGGAGTGGACACGCTAGACCAAATGTGTTCTGTGATGACCTGCAGTAGGAA

GACGAATAGGTGGCCTATGGCATTATTGTACGGAATGATAAACATTGCCTGCATAAATTCTTTTATTATACAGCCATAATGTCAGTAG

CAAGGGAGAAAAGGTTCAAAGTCGCAAAAAATTTATGAGAAACCTTTACATGAGCCTGACGTCATCGTTTATGCGTAAGCGTTTAGAAGC

TCCTACTTTGAAGAGATATTTGCGCGATAATATCTCTAATATTTTGCCAAATGAAGTGCCTGGTACATCAGATGACAGTACTGAAGAGCC

AGTAACGAAAAAACGTACTTACTGTACTTACTGCCCCTCTAAAATAAGGCGAAAGGCAAATGCATCGTGCAAAAAAATGCAAAAAAGTTAT

TTGTCGAGAGCATAATATTGATATGTGCCAAAGTTGTTTCTGACTGACTAATAAGTATAATTTGTTTCTATTATGTATAAGTTAAGCTAA

TTACTTATTTTATAATACAACATGACTGTTTTTAAAGTACAAAATAAGTTTATTTTTGTAAAAGAGAGAATGTTTTAAAAGTTTTGTTACT

TTATAGAAGAAATTTTGAGTTTTTGTTTTTTTAATAAATAAATACATAAATAAATAGTTTGTTGAATTTATTATTAGTATGTAAGT

GTAAATATAATAAAACTTAATATGTATTCAAATTAATAAATAAACCTCGATATACAGACCGATAAAACACATGCGTCAATTTTACGCATA

TTATCTTTAACGTACGTCACAATATGATTATCTTTCTAGGG

图 2A-1

SEQ ID NO: 2

MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNVIEQPGSSLASNRILTLPQRTI
RGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISEIVKWTNAEISLKRRESMTGATFRDTNEDEIYAFF
GILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRCLRMDDKSIRPTLRENDVFTPVRKIWDLFIHQCIQNYTPGAHLTIDEQ
LLGFRGRCPFRMYIPNKPSKYGIKILMMCDSGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQ
EPYKLTIVGTVRSNKREIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLD
QMCSVMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDNISNILPNEV
PGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

图 2A-2

SEQ ID NO: 33

> 粉斑夜蛾 piggyBac 转座酶

MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNVIEQPGSSLASNKILTLPQRTI
RGKNKHCWSTSKSTRRSRVSALNHVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISEIVKWTNAEISLKRRESMTGATFRDTNEDEIYAFF
GILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRCLRMDDKSIRPTLRENDVFTPVRKIWDLFIHQCIQNYTPGAHLTIDEQ
LLGFRGRCPFRMYIPNKPSKYGIKILMMCDSGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQ
EPYKLTIVGTVRSNKREIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLD
QMCSVMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDNISNILPNEV
PGTSDDSTEEPVTKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

图 2B

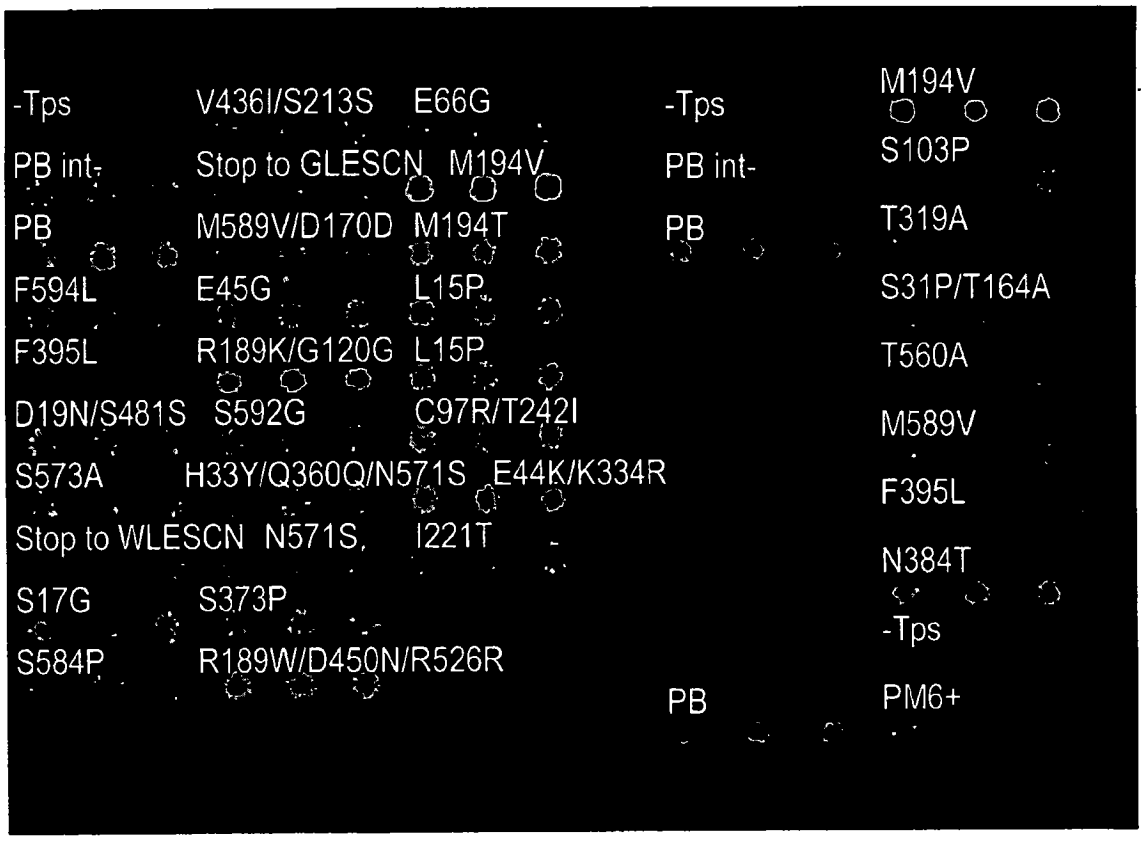

-Tps
V436I/S213S
E66G
PB int-
Stop to GLESCN
M194V
PB
M589V/D170D
M194T
F594L
E45G
L15P
F395L
R189K/G120G
L15P
D19N/S481S
S592G
C97R/T242I
S573A
H33Y/Q360Q/N571S
E44K/K334R
Stop to WLESCN
N571S,
I221T
S17G
S373P
S584P
R189W/D450N/R526R
-Tps
PB int-
PB
PB
M194V
S103P
T319A
S31P/T164A
T560A
M589V
F395L
N384T
-Tps
PM6+

图 3

SEQ ID NO: 64
MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNVIEQPGSSLASNRILTLPQRTI
RGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISEIVKWTNAEISLKRRESMTGATFRDTNEDEIYAFF
GILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRCLRMDDKSIRPTLRENDVFTPVRKIWDLFIHQCIQNYTPGAHLTIDEQ
LLGFRGRCPFRMYIPNKPSKYGIKILMMCDSGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQ
EPYKLTIVGTVASNKREIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLD
QMCSVMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDNISNILPNEV
PGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

SEQ ID NO: 65
MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNVIEQPGSSLASNRILTLPQRTI
RGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISEIVKWTNAEISLKRRESMTGATFRDTNEDEIYAFF
GILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRCLRMDDKSIRPTLRENDVFTPVRKIWDLFIHQCIQNYTPGAHLTIDEQ
LLGFRGRCPFRMYIPNKPSKYGIKILMMCDSGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQ
EPYKLTIVGTVASNKREIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLD
QMCSVMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDNISNILPNEV
PGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

SEQ ID NO: 66
MGSSLDDEHILSALLQSDDELVGEDSDSEISDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEILDEQNVIEQPGSSLASNRILTLPQRTI
RGKNKHCWSTSKSTRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISEIVKWTNAEISLKRRESMTGATFRDTNEDEIYAFF
GILVMTAVRKDNHMSTDDLFDRSLSMVYVSVMSRDRFDFLIRCLRMDDKSIRPTLRENDVFTPVRKIWDLFIHQCIQNYTPGAHLTIDEQ
LLGFRGRCPFRMYIPNKPSKYGIKILMMCDSGTKYMINGMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQ
EPYKLTIVGTVASNKREIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLD
QMCSVMTCSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMSLTSSFMRKRLEAPTLKRYLRDNISNILPNEV
PGTSDDSTEEPVMKKRTYCTYCPSKIRRKANASCKKCKKVICREHNIDMCQSCF

图 4