(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105518132 B
(45)授权公告日 2021.02.05

(21)申请号 201480042722.1
(22)申请日 2014.08.07
(65)同一申请的已公布的文献号
申请公布号 CN 105518132 A
(43)申请公布日 2016.04.20
(30)优先权数据
61/863,147 2013.08.07 US
(85)PCT国际申请进入国家阶段日
2016.01.28
(86)PCT国际申请的申请数据
PCT/US2014/050178 2014.08.07
(87)PCT国际申请的公布数据
WO2015/021298 EN 2015.02.12
(73)专利权人 瑞泽恩制药公司
地址 美国纽约州
专利权人 哈佛大学校长及研究员协会
(72)发明人 K·V·赖 G·龚 J·里恩 D·弗伦德维 D·M·瓦伦泽拉
(74)专利代理机构 北京市君合律师事务所 11517
代理人 顾云峰 吴龙瑛
(51)Int.Cl.
C12N 15/00(2006.01)
C12N 15/87(2006.01)
A61K 48/00(2006.01)
(56)对比文件
CN 101633923 A,2010.01.27
Montserrat C. Anguera等.Tsx Produces a Long Noncoding RNA and Has General Functions in the Germline, Stem Cells, and Brain.《PLoS Genetics》.2011,第7卷(第9期),1-14页.
审查员 李婧文

权利要求书5页 说明书44页
序列表1页 附图12页

(54)发明名称

缺乏lincRNA的非人类动物

(57)摘要

提供遗传修饰的非人类动物,其展现一个或多个lncRNA的功能缺乏。提供用于破坏、缺失和/或置换lncRNA编码序列的方法和组合物。提供过早衰老的遗传修饰的小鼠。此外提供被遗传修饰以包含一个或多个lncRNA的功能缺失的细胞、组织和胚胎。

CN 105518132 B

1.一种产生非人类动物的方法,包括修饰所述非人类动物使其在其基因组中包含至少一个内源性修饰的长链基因间非编码RNA (lincRNA)基因座的敲除,

其中

所述至少一个修饰的lincRNA基因座包括用插入核酸置换所述lincRNA基因座或其部分,

所述置换导致缺失全部或大部分编码所述lincRNA的序列,

所述缺失导致所述lincRNA的功能缺失型突变,

所述插入核酸包含编码报道基因的第一核苷酸序列,并且其位于与所述内源lincRNA启动子可操作地连接的位置,

所述内源lincRNA启动子驱动所述第一核苷酸序列的表达,

所述第一核酸序列的表达遵循所述lincRNA的表达模式,并且

所述非人类动物展现与所述lincRNA的功能缺失型突变相关联的表型;

其中

i)所述lincRNA是*Pint*,其中所述功能缺失型突变包括在基因坐标Chromosome 6: 31166026-31197846之间的所述*Pint*基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现:

(a) 过早衰老相关表型;

(b) 相比于野生型对照的较缓慢生长速率;

(c) 肌肉力量下降;

(d) 纤维化;

(e) 相比于野生型对照的较低体脂含量;

(f) 相比于野生型对照的较低股骨骨矿物质密度和骨质量;

(g) 相比于野生型对照降低的肌肉质量;

(h) 中值寿命降低;

(i) 脊椎弯曲;

(j) 器官萎缩;

(k) 脑发育缺陷;或

(l) (a)-(k)的其任何组合; 或者

ii)所述lincRNA是Fendrr,其中所述功能缺失型突变包括在基因坐标Chromosome 8: 121054882-121074065之间的所述Fendrr基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现围产期致死和/或肺发育缺陷;或者

iii)其中所述lincRNA是HOTTIP,其中所述功能缺失型突变包括在基因坐标Chromosome 6: 52262834-52267603或Chromosome 6: 52265374-52267603之间的所述HOTTIP基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现至少一种骨骼和/或肌肉形态或功能缺陷;或者

iv) 其中所述lincRNA是Peril,其中所述功能缺失型突变包括在基因坐标Chromosome 3: 34767849-34782292之间的所述Peril基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现脑发育缺陷、过早衰老相关表型和/或围产期致死。

2.如权利要求1所述的方法,其中所述插入核酸进一步包含编码可选择标记的第二核

酸序列。

3.如权利要求2所述的方法,其中所述第二核酸序列可操作地连接至驱动所述可选择标记的表达的启动子。

4.如权利要求2所述的方法,其中所述插入核酸包含侧接编码所述报道基因的区段和/或编码所述可选择标记的区段的位点特异性重组位点。

5.如权利要求1所述的方法,其中所述第一核苷酸序列进一步包含Kozak共有序列,所述Kozak共有序列与编码报道基因的序列可操作地连接。

6.如权利要求1所述的方法,其中所述报道基因是β-半乳糖苷酶,并且所述基因座能够报道所述lincRNA基因的转录活性。

7.如权利要求1所述的方法,其中所述非人动物针对所述敲除的修饰的lincRNA基因座是纯合子。

8.如权利要求1所述的方法,其中所述非人动物是啮齿动物。

9.如权利要求1所述的方法,其中所述非人动物是大鼠或小鼠。

10.如权利要求1-9任意一项所述的方法,其中所述非人动物是小鼠。

11.一种用于产生非人类动物的方法,所述方法包括:

(a)使多能性细胞与包含插入核酸的打靶构建体接触,所述插入核酸包含编码报道基因的第一核酸序列并且侧接5'和3'同源臂;其中所述打靶构建体与所述细胞基因组中的所述lincRNA基因座进行同源重组以形成包括修饰的lincRNA基因座的修饰的多能性细胞,并且其中与所述所关注的lincRNA基因座同源重组之后,所述插入核酸可操作地连接至内源lincRNA启动子并且置换全部或大部分编码所述lincRNA的序列,以使得

所述内源lincRNA启动子驱动所述第一核酸序列的表达,并且

所述第一核酸序列的表达遵循所述lincRNA的表达模式;

(b)将所述修饰的多能性细胞引入宿主胚胎中;和

(c)使所述宿主胚胎在代孕母体中受孕,其中所述代孕母体产生包含所述修饰的lincRNA基因座的后代,

其中所述修饰的lincRNA基因座包括缺失全部或大部分编码所述lincRNA的序列,

其中所述置换导致所述lincRNA的功能缺失型突变,并且

其中所述遗传修饰导致所述至少一个lincRNA的功能缺失型突变从而使得子代展现与所述功能缺失型突变相关联的表型;

其中

i)所述lincRNA是*Pint*,其中所述功能缺失型突变包括在基因坐标Chromosome 6:31166026-31197846之间的所述*Pint*基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现:

(a) 过早衰老相关表型;

(b) 相比于野生型对照的较缓慢生长速率;

(c) 肌肉力量下降;

(d) 纤维化;

(e) 相比于野生型对照的较低体脂含量;

(f) 相比于野生型对照的较低股骨骨矿物质密度和骨质量;

(g) 相比于野生型对照降低的肌肉质量;

(h) 中值寿命降低;

(i) 脊椎弯曲;

(j) 器官萎缩;

(k) 脑发育缺陷;或

(l) (a)-(k)的其任何组合; 或者

ii)所述lincRNA是Fendrr,其中所述功能缺失型突变包括在基因坐标Chromosome 8: 121054882-121074065之间的所述Fendrr基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现围产期致死和/或肺发育缺陷;或者

iii)其中所述lincRNA是HOTTIP,其中所述功能缺失型突变包括在基因坐标Chromosome 6: 52262834-52267603或Chromosome 6: 52265374-52267603之间的所述HOTTIP基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现至少一种骨骼和/或肌肉形态或功能缺陷;或者

iv) 其中所述lincRNA是Peril,其中所述功能缺失型突变包括在基因坐标Chromosome 3: 34767849-34782292之间的所述Peril基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现脑发育缺陷、过早衰老相关表型和/或围产期致死。

12.一种用于修饰多能性细胞中的lincRNA基因座的方法,其包括:

(a)在所述多能性细胞中引入包含插入核酸的打靶构建体,所述插入核酸包含编码报道基因的第一核酸序列并且侧接可以与所述lincRNA基因座进行同源重组的5'和3'同源臂的;和

(b)鉴定在所述lincRNA基因座包含靶向遗传修饰的修饰多能性细胞,

其中修饰的lincRNA基因座包括用插入核酸置换所述lincRNA基因座或其一部分,其导致缺失全部或大部分编码所述lincRNA的序列,

其中所述缺失导致所述lincRNA的功能缺失型突变,

所述编码报道基因的第一核苷酸序列位于与所述内源lincRNA启动子可操作地连接的位置,

所述内源lincRNA启动子驱动所述第一核苷酸序列的表达,

所述第一核酸序列的表达遵循所述lincRNA的表达模式,并且

产生自所述多能性细胞的非人动物展现与所述lincRNA的功能缺失型突变相关联的表型;

其中

i)所述lincRNA是*Pint*,其中所述功能缺失型突变包括在基因坐标Chromosome 6: 31166026-31197846之间的所述*Pint*基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现:

(a) 过早衰老相关表型;

(b) 相比于野生型对照的较缓慢生长速率;

(c) 肌肉力量下降;

(d) 纤维化;

(e) 相比于野生型对照的较低体脂含量;

(f) 相比于野生型对照的较低股骨骨矿物质密度和骨质量;

(g) 相比于野生型对照降低的肌肉质量;

(h) 中值寿命降低;

(i) 脊椎弯曲;

(j) 器官萎缩;

(k) 脑发育缺陷;或

(l) (a)-(k)的其任何组合; 或者

ii)所述lincRNA是Fendrr,其中所述功能缺失型突变包括在基因坐标Chromosome 8: 121054882-121074065之间的所述Fendrr基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现围产期致死和/或肺发育缺陷;或者

iii)其中所述lincRNA是HOTTIP,其中所述功能缺失型突变包括在基因坐标Chromosome 6: 52262834-52267603或Chromosome 6: 52265374-52267603之间的所述HOTTIP基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现至少一种骨骼和/或肌肉形态或功能缺陷;或者

iv) 其中所述lincRNA是Peril,其中所述功能缺失型突变包括在基因坐标Chromosome 3: 34767849-34782292之间的所述Peril基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现脑发育缺陷、过早衰老相关表型和/或围产期致死。

13.如权利要求12所述的方法,其中所述多能性细胞是人类iPS细胞。

14.如权利要求12所述的方法,其中所述多能性细胞是小鼠或大鼠胚胎干(ES)细胞。

15.一种获得细胞、组织或胚胎的方法,包括:

由非人动物处获取细胞、组织或胚胎,其中所述非人动物包含内源性修饰的长链基因间非编码RNA (lincRNA)基因座的敲除,其中所述至少一个修饰的lincRNA基因座包括用插入核酸置换所述lincRNA基因座或其一部分,

其中

所述置换导致缺失全部或大部分编码所述lincRNA的序列,

所述缺失导致所述lincRNA的功能缺失型突变,

所述插入核酸包含编码报道基因的第一核苷酸序列,并且其位于与所述内源lincRNA启动子可操作地连接的位置,

所述内源lincRNA启动子驱动所述第一核苷酸序列的表达,

所述第一核酸序列的表达遵循所述lincRNA的表达模式,并且

所述非人动物展现与所述lincRNA的功能缺失型突变相关联的表型;

其中

i)所述lincRNA是*Pint*,其中所述功能缺失型突变包括在基因坐标Chromosome 6: 31166026-31197846之间的所述*Pint*基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现:

(a) 过早衰老相关表型;

(b) 相比于野生型对照的较缓慢生长速率;

(c) 肌肉力量下降;

(d) 纤维化;

(e) 相比于野生型对照的较低体脂含量；

(f) 相比于野生型对照的较低股骨骨矿物质密度和骨质量；

(g) 相比于野生型对照降低的肌肉质量；

(h) 中值寿命降低；

(i) 脊椎弯曲；

(j) 器官萎缩；

(k) 脑发育缺陷;或

(l) (a)-(k)的其任何组合; 或者

ii)所述lincRNA是Fendrr,其中所述功能缺失型突变包括在基因坐标Chromosome 8: 121054882-121074065之间的所述Fendrr基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现围产期致死和/或肺发育缺陷;或者

iii)其中所述lincRNA是HOTTIP,其中所述功能缺失型突变包括在基因坐标Chromosome 6: 52262834-52267603或Chromosome 6: 52265374-52267603之间的所述HOTTIP基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现至少一种骨骼和/或肌肉形态或功能缺陷;或者

iv) 其中所述lincRNA是Peril,其中所述功能缺失型突变包括在基因坐标Chromosome 3: 34767849-34782292之间的所述Peril基因座缺失基因组序列,并且所述功能缺失型突变使得所述小鼠展现脑发育缺陷、过早衰老相关表型和/或围产期致死。

缺乏lincRNA的非人类动物

[0001] 相关申请的交叉引用

[0002] 本申请要求2013年8月7日提交的美国临时申请No.61/863,147的权益,所述临时申请特此以全文引用的方式并入本文中。

[0003] 经由EFS WEB以文本文件形式提交的序列表的引用

[0004] 序列表的正式文本以ASCII格式序列表形式经由EFS-Web以电子方式提交,文件名为448071SEQLIST.TXT,创建于2014年8月7日,并具有1千字节的大小,并且与本说明书同时提交。该ASCII格式文件中所含的序列表是本说明书的一部分并且以全文引用的方式并入本文中。

技术领域

[0005] 非人类动物、细胞和组织以及其产生方法,其包含长链非编码RNA("lncRNA")的一个或多个缺失。非人类动物和其产生方法,其包含非功能性lncRNA,或一个或多个lncRNA的敲除。遗传修饰的非人类动物展现与过早衰老一致的表型。

背景技术

[0006] 在哺乳动物中,长链非编码RNA(lncRNA)和被称为长链基因间非编码RNA(lincRNA)的子类包含约15,000种不同的转录物,其在结构、合成和其基因的染色质特性上与mRNA类似。对于绝大多数的所鉴定lncRNA,与特定lncRNA相关的功能或表型是未知的。据认为,一些lncRNA在动物中作为转录激活子或阻遏子参与转录控制;其他可在翻译后起作用或通过一些其他机制起作用。因此,取决于lncRNA的身份和功能,操纵lncRNA的能力可提供用于开发所关注表型的工具。本领域中需要用于操纵lncRNA的方法和组合物,并且需要通过lncRNA操纵产生非人类动物的表型。

发明内容

[0007] 提供含有非功能性长链非编码RNA(lncRNA)的非人类动物、细胞、组织和胚胎,包括(但不限于)一个或多个lncRNA的敲除。提供用于操纵lncRNA表达的方法和组合物。此外提供针对修饰或敲除lncRNA的打靶组合物。提供展现与一个或多个lncRNA的非功能相关的表型的非人类动物、细胞和组织。

[0008] 在一个方面,提供在基因组中包含至少一个修饰的lncRNA基因座的非人类动物,其中所述修饰的lncRNA基因座在编码lncRNA的核酸序列中包含功能缺失型突变。

[0009] 在一个实施方案中,所述lncRNA是长链基因间非编码RNA(lincRNA)。

[0010] 在一个方面,所述功能缺失型突变的特征在于至少一种lncRNA功能的破坏或敲除。

[0011] 在一个实施方案中,所述修饰的lncRNA基因座包含编码lncRNA或其部分的一个或多个外显子的缺失。在一个方面,所述破坏或敲除包括lncRNA基因座内起始于所述lncRNA基因座的第二外显子的一个或多个外显子的缺失;lncRNA基因座内起始于所述lncRNA基因

座的第一外显子的一个或多个外显子的缺失;或lncRNA基因座的整个RNA编码区的缺失。

[0012] 在一个方面,所述破坏或敲除包括用插入核酸置换lncRNA基因座或其部分。在一个实施方案中,所述插入核酸包含编码报道基因的第一核苷酸序列。在一些这样的情况下,所述第一核苷酸序列可操作地连接至驱动所述报道基因表达的启动子。在一个实施方案中,编码所述报道基因的所述第一核苷酸序列位于与内源lncRNA启动子可操作地连接的lncRNA基因座中,其中所述内源lncRNA启动子驱动所述核苷酸序列的表达。在这些情况下,核酸序列的表达遵循lncRNA的表达模式。在一个方面,所述插入核酸包含Kozak共有序列。在一个特定实施方案中,所述插入核酸的第一核苷酸序列包含Kozak共有序列。

[0013] 在一个实施方案中,所述插入核酸还包含编码可选择标记的第二核苷酸序列,其中所述第二核苷酸序列可操作地连接至启动子。

[0014] 在一个方面,所述插入核酸包含侧接编码所述报道基因的区段和/或编码所述可选择标记的区段的位点特异性重组位点。

[0015] 在多个方面和实施方案中,lncRNA基因座或其部分的置换包括用插入核酸置换lncRNA基因座内起始于所述lncRNA基因座的第二外显子的一个或多个外显子;用插入核酸置换lncRNA基因座内起始于所述lncRNA基因座的第一外显子的一个或多个外显子;或用插入核酸置换lncRNA基因座的整个RNA编码区。

[0016] 在一个实施方案中,本文提供的非人类动物的特征在于具有一种或多种以下表型:(a)过早衰老相关表型;(b)围产期致死;(c)肺发育缺陷;(d)尾部和后肢的形态畸形;(e)一个或多个组织中的肌肉质量损失;或(f)(a)-(e)的其任何组合。

[0017] 在一个方面,本文提供的非人类动物包含lncRNA Pint的破坏或敲除,并且所述非人类动物的特征在于过早衰老相关表型,其包含:(a)相比于野生型对照的较缓慢生长速率;(b)肌肉力量下降;(c)纤维化;(d)相比于野生型对照的较低体脂含量;(e)相比于野生型对照的较低股骨骨矿物质密度和骨质量;(f)相比于野生型对照降低的肌肉质量;(g)中值寿命降低;(h)脊椎弯曲(lordokyphosis);(i)器官萎缩;或(j)(a)-(i)的其任何组合。

[0018] 在一个实施方案中,本文提供的非人类动物展现大脑发育缺陷。在一些这样的情况下,所述lncRNA是Pantr2、Kantr、Peril、Celrr、Pantr1、Crnde、lincenc1、Pint、lincppara或Tug1。

[0019] 在多个方面和实施方案中,所述非人类动物是哺乳动物。在多个方面和实施方案中,所述哺乳动物是啮齿动物,例如,小鼠、大鼠或仓鼠。在多个方面和实施方案中,所述哺乳动物是羊、牛或猪物种。

[0020] 在一个方面,提供遗传修饰的非人类动物,其中所述遗传修饰导致lncRNA的功能缺失。

[0021] 在一个方面,提供遗传修饰的非人类动物,其中所述遗传修饰包含一个或多个lncRNA的破坏或敲除。

[0022] 在一个实施方案中,所述遗传修饰包含至少两个lncRNA的破坏或敲除。在一个实施方案中,所述遗传修饰包含至少三个、四个、五个或六个lncRNA的破坏或敲除。

[0023] 在一个实施方案中,所述遗传修饰包含一个或多个lncRNA的破坏或敲除并且在所述lncRNA基因座内包含编码可检测部分(报道基因)的基因,所述可检测部分(报道基因)可操作地连接至所破坏或敲除的lncRNA的启动子。在一个实施方案中,所述编码可检测部分

(报道基因)的基因选自lacZ(编码β-半乳糖苷酶)、GFP、eGFP、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强型黄色荧光蛋白(EYFP)、Emerald、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶和其组合。

[0024] 在一个实施方案中,所述遗传修饰包含一个或多个lncRNA的破坏或敲除并且在所述lncRNA基因座内包含编码可检测部分(报道基因)的基因,所述可检测部分(报道基因)可操作地连接至驱动所述可检测部分表达的启动子。

[0025] 在多个方面和实施方案中,所述可检测部分包括本领域中已知的任何报道基因。

[0026] 在一个方面,提供遗传修饰的非人类动物,其中所述遗传修饰包含编码lincRNA的lncRNA核酸序列的破坏,所述lincRNA选自HOTAIR、HOTTIP、Hoxa11os(原名为HoxA11as)、Pantr1(原名为lincRNA-Brn1-a)、Pantr2(原名为lincRNA-Brn1-b)、Ptgs2os2(原名为lincRNA-Cox2)、Eldr(原名为Fabl和lincRNA-Egfr)、Lincenc1(原名为lincRNA-Enc1)、Mannr(原名为lincRNA-Evi1)、Fendrr(原名为lincRNA-Foxf1)、Halr1(原名为Haunt和lincRNA-HoxA1)、Haglr(原名为lincRNA-HoxD3)、Celrr(原名为Celr和lincRNA-Insig2)、Crnde(原名为lincRNA-Irx5)、Kantr(原名为Spasm和lincRNA-Jarid1c)、Pint(原名为linc-Pint和lincRNA-Mkln1)、Trp53cor1(原名为lincRNA-p21)、lincppara(原名为lincRNA-Ppara)、Peril(原名为lincRNA-Sox2)、Tug1(原名为lincRNA-Tug1)和其组合。

[0027] 在一个方面,提供遗传修饰的非人类动物,其中所述遗传修饰包含Pint(原名为linc-Pint和lincRNA-Mkln1)的破坏或敲除。

[0028] 在一个方面,提供核酸基因座,其包含lncRNA的破坏。在一个实施方案中,所述破坏包含lncRNA的敲除。在一个实施方案中,所述破坏包含编码可操作地连接至lncRNA的启动子的可检测部分的基因的安置。在一个实施方案中,所述破坏包含lncRNA的敲除和编码可操作地连接lncRNA的启动子的可检测部分的基因的安置。

[0029] 在一个方面,提供核酸构建体,其包含将所述构建体靶向包含lncRNA的基因座的至少一个打靶序列,其中所述构建体能够破坏lncRNA的转录,敲除lncRNA,或置换lncRNA。

[0030] 在一个实施方案中,所述核酸构建体还包含可检测部分(具有或不具有驱动所述可检测部分的表达的外加启动子)。在一个实施方案中,所述核酸构建体还包含由启动子驱动的可选择标记基因。在一个实施方案中,所述核酸构建体包含可检测部分(具有或不具有其自身启动子)和由启动子驱动的可选择标记基因。在一个实施方案中,所述可选择标记和/或所述可检测部分侧接在引导切除所述可检测部分和/或所述可选择标记的位点特异性重组位点上游和下游。

[0031] 在一个实施方案中,提供打靶载体。在一个方面,所述打靶载体包含侧接可以与所关注的lncRNA基因座进行同源重组的5'和3'同源臂的插入核酸。在一个实施方案中,所述打靶载体的插入核酸包含编码报道基因的第一核酸序列。在一个方面,在与所关注的lncRNA基因座进行同源重组后,编码所述报道基因的所述第一核酸序列可操作地连接至驱动lncRNA在lncRNA基因座的表达的内源启动子。在一个实施方案中,所述打靶载体的插入核酸的第一核酸序列和/或第二核酸序列还包含Kozak共有序列。在一个实施方案中,所述打靶载体还包含驱动启动子表达的启动子。

[0032] 在一个方面,所述打靶载体的插入核酸还包含编码可选择标记的第二核酸序列,

其中所述第二核酸可操作地连接至启动子。在一个实施方案中，所述打靶载体还包含侧接编码报道基因的区段和/或编码可选择标记核酸的区段的位点特异性重组位点。

[0033] 在一个方面，提供展现过早衰老相关表型的非人类动物，其中所述非人类动物包含使得Pint具非功能性的修饰。在一个实施方案中，所述修饰是Pint基因座的RNA编码序列的破坏。在一个实施方案中，所述修饰是Pint基因座的整个RNA编码序列的缺失。在一个实施方案中，所述修饰包含将打靶载体插入Pint基因座中，以使得动物不再具有功能性Pint。

[0034] 在一个实施方案中，所述修饰还包含在Pint基因座的可检测部分(例如，具有或不具有其自身启动子的报道基因)和任选地可选择标记。在一个实施方案中，所述可检测部分和/或所述可选择标记侧接在引导切除所述可检测部分和/或所述可选择标记的位点特异性重组位点上游和下游。在一个实施方案中，所述非人类动物还包含可与位点特异性重组酶位点相容的诱导型位点特异性重组酶。

[0035] 在一个方面，提供非人类动物的细胞、组织或胚胎，其中所述细胞或组织缺乏编码功能性lincRNA的核酸序列，所述功能性lincRNA选自HOTAIR、HOTTIP、Hoxa11os(原名为HoxA11as)、Pantr1(原名为lincRNA-Brn1-a)、Pantr2(原名为lincRNA-Brn1-b)、Ptgs2os2(原名为lincRNA-Cox2)、Eldr(原名为Fabl和lincRNA-Egfr)、Lincenc1(原名为lincRNA-Enc1)、Mannr(原名为lincRNA-Evi1)、Fendrr(原名为lincRNA-Foxf1)、Halr1(原名为Haunt和lincRNA-HoxA1)、Haglr(原名为Mdgt和lincRNA-HoxD3)、Celrr(原名为Celr和lincRNA-Insig2)、Crnde(原名为lincRNA-Irx5)、Kantr(原名为Spasm和lincRNA-Jarid1c)、Pint(原名为linc-Pint和lincRNA-Mkln1)、Trp53cor1(原名为lincRNA-p21)、lincppara(原名为lincRNA-Ppara)、Peril(原名为lincRNA-Sox2)、Tug1(原名为lincRNA-Tug1)和其组合。

[0036] 在一个实施方案中，缺乏编码功能性lincRNA的核酸序列的细胞或组织缺乏功能性Pint(原名为lincRNA-Mlkn1)。

[0037] 在一个方面，提供核酸构建体，其包含上游同源臂和下游同源臂，其中所述上游和下游同源臂靶向lncRNA基因座，其中所述构建体能够破坏lncRNA的转录，敲除lncRNA，或置换lncRNA。

[0038] 在多个方面和实施方案中，靶向lncRNA基因座的打靶构建体包含选自Kozak序列、编码可检测部分(例如报道基因，例如如本文所述的报道基因；具有例如任选地与其可操作地连接的启动子)的序列、编码可选择标记(具有例如与其可操作地连接的启动子)的核酸序列和其组合的序列。在一个实施方案中，报道基因和/或可选择标记侧接位点特异性重组位点，所述位点特异性重组位点被布置以实现编码可选择标记基因的核酸序列和/或编码可检测部分的核酸序列的缺失。在一个实施方案中，所述构建体不包含可操作地连接至可检测部分的启动子。

[0039] 在一个方面，提供用于破坏lncRNA的方法，其包括在lncRNA基因座内插入核酸序列，其中所述插入破坏lncRNA的转录，缺失一个或多个lncRNA编码区，或缺失lncRNA的整个编码序列。

[0040] 在一个方面，提供用于产生包含lncRNA的破坏或敲除的非人类动物的方法，其包括修饰非人类动物的基因组以使得所述非人类动物不再表达lncRNA的功能型式。在一个实施方案中，所述方法包括使用打靶载体来破坏lncRNA的转录，来缺失一个或多个lncRNA编码区，或缺失所述非人类动物的基因组中的lncRNA的整个编码序列的步骤。

[0041] 在一个方面，提供用于产生包含lncRNA功能敲除的非人类动物的方法，其包括修饰多能性或全能性非人类动物细胞的基因组以破坏lncRNA的转录，缺失编码lncRNA编码区的一个或多个外显子，或缺失细胞基因组中的lncRNA的整个编码序列；使用所述细胞作为供体细胞并将所述供体细胞引入宿主胚胎中以形成供体细胞-宿主胚胎复合物；并且在合适的非人类动物中在适于受孕的条件下使所述供体细胞-宿主胚胎复合物受孕，其中在受孕后获得包含lncRNA功能敲除的后代。在一个实施方案中，所述后代被培育成关于lncRNA功能敲除的纯合性。

[0042] 在一个方面，提供用于产生包含lncRNA功能敲除的非人类动物的方法，其包括修饰非人类动物细胞的体细胞或生殖细胞的基因组以破坏lncRNA的转录，缺失一个或多个lncRNA编码区，或缺失细胞基因组中的lncRNA的整个编码序列；使用去核卵细胞中的细胞基因组来形成修饰卵细胞；使所述修饰卵细胞在合适的代孕非人类动物中在适于受孕的条件下受孕；并且获得包含lncRNA敲除的非人类动物后代。在一个实施方案中，将所述后代培育成关于lncRNA功能敲除的纯合性。

[0043] 在一个实施方案中，提供用于产生在至少一个lncRNA基因座中包含遗传修饰的非人类动物的方法。这种方法包括使多能性细胞与包含侧接5'和3'同源臂的插入核酸的打靶构建体接触；其中所述打靶构建体与所述细胞基因组中的lincRNA基因座进行同源重组以形成修饰的多能性细胞；(b)在宿主胚胎中引入所述修饰的多能性细胞；和(c)使所述宿主胚胎在代孕母体中受孕，其中所述代孕母体产生包含修饰的lncRNA基因座的后代，其中所述遗传修饰导致至少一个lncRNA的功能缺失。

[0044] 在一个方面，提供用于修饰多能性细胞中的lncRNA基因座的方法。这种方法包括在所述多能性细胞中引入包含侧接可以与所述lncRNA基因座进行同源重组的5'和3'同源臂的插入核酸的打靶构建体；并且鉴定在所述lncRNA基因座包含靶向遗传修饰的修饰多能性细胞，其中所述遗传修饰导致lncRNA功能的功能缺失。在一个实施方案中，所述多能性细胞是人类诱导型多能性(iPS)细胞。

[0045] 在多个方面和实施方案中，修饰细胞包括例如多能性细胞、诱导型多能性细胞、干细胞、胚胎干细胞等。在一个特定实施方案中，所述细胞是胚胎干(ES)细胞。在一个特定实施方案中，所述ES细胞是小鼠或大鼠ES细胞。

[0046] 在多个方面和实施方案中，所述非人类动物包括例如羊、牛、猪和鼠类动物物种。在一个特定实施方案中，所述动物是来自鼠类物种，例如，小鼠或大鼠。

[0047] 如本领域技术人员在阅读本公开后将理解，包括其他和另外的方面和实施方案。

附图说明

[0048] 本专利或申请文件含有至少一个彩色附图。在请求并支付必要费用后将由专利局提供具有彩色附图的本专利或专利申请公开的副本。

[0049] 图1示出用于靶向破坏Fendrr基因座的策略。示出包括外显子1-6的野生型小鼠Fendrr基因座的部分图谱。在同源重组后，打靶LTVEC用引入Kozak序列的LacZ-新霉素抗性表达盒置换总共19.2kb的基因组Fendrr序列。空心方框指示非编码外显子。Fendrr基因组基因座上和LacZ-新霉素抗性表达盒中的红色和绿色方框是用于打靶的同源序列。

[0050] 图2示出在妊娠中期阶段lincRNA靶向小鼠胚胎的空间和时间LacZ报道基因表达

A,将杂合e12.5胚胎固定并对于β-半乳糖苷酶染色,显示所引入LacZ报道基因在发育大脑和颅面区域(例如,Pantr1和Pantr2,Celrr和Haglr,也参见图9)、神经管(Pantr2、Halr1和lincppara)、背主动脉(Celrr)、心脏(Celrr、Haglr,也参见图9)、肺(Fendrr)、肢芽(HOTTIP、Hoxa11os和Mannr)、前肠(HOTTIP、Hoxa11os和Fendrr)、后部区域和尾部(HOTAIR、HOTTIP和Hoxa11os)中的宽泛范围的表达。类似的分析显示在Tug1中的广泛lacZ表达模式,而其他报道基因的表达可能局限于表皮(Eldr)、乳腺芽(Lincenc1,也参见图9)或须基板(Trp53cor1,也参见图9)。所示实例是每个lincRNA敲除项目的至少五种基因型确定的胚胎的代表。B,在来自指定阶段(e9.5-e12.5)的所选lincRNA(HOTTIP、Hoxa11os和Celrr)F1杂合胚胎中的表达模式显示,表达早期开始于限制位点并且然后在稍后阶段扩展超出了这个初始位点。Celrr表达在e9.5时局限于大脑并且到e12.5时发展到脊髓中。Hoxa11os表达开始于发育尾芽中并且到e12.5时发展到胚胎的整个尾区、后肢和前肢中。HOTTIP表达也开始于发育尾芽中并且然后到e11.5和e12.5时在前肢和后肢的发育远端肢体末梢中观测到。所示实例是每个lincRNA项目的至少5种基因型确定的胚胎的代表。

[0051] 图3示出在6-8周龄lincRNA F0代杂合子的大脑中的LacZ报道基因表达(蓝色)。A,Celrr,在除横向小脑和桥脑腹侧之外的灰质中广泛表达;B,Crnde,在丘中表达(背视图,箭头);C,Pantr1,在新皮层、嗅球、基底前脑和下丘脑中表达;D,Pantr2,在新皮层、嗅球、小脑、下丘脑和基底前脑中表达;E,Lincenc1,在新皮层、部分小脑和内侧下丘脑中表达,其中在颞叶皮层的嗅突起和嗅突起区域中具有特别强烈的表达模式(腹视图,红色箭头);F,Pint,在灰质中普遍表达,其中在下丘脑中具有特别强烈的表达;G,lincppara,在灰质中广泛表达,其中在下丘脑中具有特别密集的表达;H,Peril,在下丘脑中线中表达(腹视图,箭头);I,Kantr,在小脑深部层中的潜在一些表达(背视图,星号);和J,Tug1,在脊髓灰质中表达和在除新皮层之外的大多数结构中的轻微灰质表达。n=2,每个lincRNA敲除项目的基因型确定的雄性小鼠。

[0052] 图4示出Pint展现从出生后第3天至8周龄的表达增加。在F0杂合子中在第3天、第3周和第8周的LacZ报道基因表达(蓝色)显示Pint随年龄增长的表达增加。A,在第3天,仅在部分大脑、后肢的肌腱和韧带以及肺中一些细支气管中观测到β-半乳糖苷酶染色(箭头)。B,在第3周,在大脑、后肢、心脏心房、肺和肝脏中的染色增加。C,到8周龄时,全脑、后肢和胸部的骨骼肌、心房和心肌、肺和肝脏组织都展现强烈的β-半乳糖苷酶染色,这代表增加的Pint表达。所示实例是每组n>4只小鼠的代表。

[0053] 图5示出Pint敲除小鼠中的过早衰老相关表型。(A)$Pint^{-/-}$和$Pint^{+/-}$雄性小鼠相比于其野生型(WT)同窝出生仔畜展现显著较低的生长速率并且在接近6月龄时开始显示显著的体重减轻。数据以平均值+/-SEM绘制,每个组为n>9只小鼠。通过单因素ANOVA评估显著性(*,P<0.05;**,P<0.005;***,P<0.001)。(B)纯合与杂合和WT小鼠的Kaplan-Meier分析。$Pint^{-/-}$雄性小鼠相比于$Pint^{+/-}$和野生型同窝出生仔畜展现存活率显著降低。数据以在1年观测内的存活率百分比绘制。(C)$Pint^{-/-}$小鼠相比于$Pint^{+/-}$和WT同窝出生仔畜的腹侧和背侧皮肤切片。(D、E、F和G)在12、26和52周龄的身体组成的显微CT评估。(D、E)在早至26周龄时,雄性$Pint^{-/-}$和$Pint^{+/-}$小鼠展现身体脂肪显著减少。在更大年龄时,明显地在52周龄时,雌性$Pint^{-/-}$小鼠具有减少的身体脂肪(***,P<0.001,单因素ANOVA)。(F、G)在雄性和雌性$Pint^{-/-}$中,相比于其$Pint^{+/-}$和WT同窝出生仔畜,观测到股骨骨矿物质密度(BMD)显著降低

(*,P<0.05;***,P<0.001,单因素ANOVA)。(H)显微CT图像描绘出,在年龄较大的雄性和雌性$Pint^{-/-}$小鼠中,相比于WT同窝出生仔畜,可见明显的脊椎弯曲(脊柱的曲率)。(I)到12周龄时,约70%的雄性和雌性$Pint^{-/-}$小鼠具有脊椎弯曲,相比于$Pint^{+/-}$和WT同窝出生仔畜的0-20%。到26周龄时,具有脊椎弯曲的$Pint^{-/-}$小鼠的比例增加至接近90%并且出现在约60%的$Pint^{+/-}$小鼠中,相比于WT同窝出生仔畜的小于20%。对于所报道的所有观测结果,每组n>9只小鼠。

[0054] 图6示出Fendrr敲除小鼠在e13.5时展现异常的肺形态。A.在e12.5时在Fendrr KO胚胎中的LacZ报道基因表达展现在面部额鼻(FN)、主动脉性腺中肾(AGM)区域和包括肺(L)和气管(T)的呼吸道中的阳性表达。B.在e13.5时的肺解剖揭示出,相比于Het,在KO叶中的异常、杂乱、球状表型。

[0055] 图7示出在HOTAIR KO小鼠的第4节尾椎中观测到的同源转化。A.通过μCT对小鼠骨骼的骶骨和尾区的可视化揭示出,在HOTAIR KO小鼠中第4节尾椎同源转化成与第3节尾椎类似的结构。B.WT和HOTAIR KO第4节尾椎的背部、侧面和腹部比较揭示出KO中的结构异常,这指示同源转化。

[0056] 图8示出HOTTIP KO小鼠表现出异常的后肢姿势,前肢和后肢中的握力降低和肌肉萎缩表型。A.HOTTIP KO小鼠当通过尾部悬吊时展现不同寻常的后肢"紧握"姿势。WT,野生型;KO,基因敲除。B.笼耐力测试揭示出,HOTTIP KO小鼠保持从铁丝笼顶部颠倒悬吊的能力下降。每组n=5只小鼠。C.从WT、Het和KO小鼠获取右和左TA(胫前肌)、GA(腓肠肌)和Quad(四头肌)并称重。将肌肉重量针对体重标准化并计算以包括右/左肌肉重量。数据为平均值+/-SEM,每组n=6只小鼠。仅在雄性和雌性的HOTTIP KO动物的GA中观测到肌肉重量显著降低(雄性数据未示出)。星号指示KO GA肌肉重量相比于所有其他对照组的显著差异(P<0.01)。D.在WT、Het和KO中的GA肌肉纤维数的比较。在KO中观测到纤维计数显著降低。通过使用单因素ANOVA评估显著性(P<.0001)。E.肌肉纤维的平均横截面积的比较。从GA肌肉获取的横截面用针对层粘连蛋白(Sigma)的抗体染色并测量。在KO与对照骨骼肌之间没有明显的尺寸差异。对于所有肌肉分析,每组n=6只小鼠。

[0057] 图9示出展现高特异性染色模式的四种lincRNA敲除的小区域中的精确染色:Peril、Ptgs2os2、Trp53cor1和Lincenc1的特定中期妊娠lacZ表达谱。(A)Peril的LacZ报道基因谱分析显示特定神经元表达模式以及在心脏和后尾区中的强烈表达。(B)Ptgs2os2 lacZ报道基因表达局限于发育前肢和后肢的基底。(C)Trp53cor1 lacZ报道基因表达在鼻过程中对发育的须基板具特异性。从同窝仔畜采集的E12.5胚胎捕捉须基板在很短一段时间内的发育进展。(D)在$Lincenc1^{+/-}$胚胎中除去前肢和后肢以揭示乳腺芽表达(箭头)。E12.5$Lincenc1^{+/-}$胚胎的腹部图:在五对乳腺芽中检测到lacZ表达。

[0058] 图10示出以HOTTIP敲除中的跟骨损失为特征的表型:在Hottip突变小鼠中观测到骨骼畸形。除了后肢中的骨骼肌表型之外,$Hottip^{-/-}$小鼠也显示通过3D显微CT可视化的骨骼骨异常。雄性和雌性(C和F)$Hottip^{-/-}$小鼠相比于(A和D)WT和(B和E)$Hottip^{+/-}$同窝出生仔畜对照具有缩短的跟骨(箭头)。

[0059] 图11代表在胚胎和成体组织中对于研究的lncRNA敲除的报道基因表达的表格(表2)。

具体实施方式

[0060] 术语表

[0061] 术语“胚胎干细胞”或“ES细胞”包括在引入胚胎后能够促进发育胚胎的任何组织的胚胎源全能性或多能性细胞。术语“多能性细胞”包括能够发育成一种以上分化细胞类型的未分化细胞。

[0062] 术语“大打靶载体”或“LTVEC”包括源自比通常预期在真核细胞中进行同源基因打靶的其他方法所使用更大的克隆基因组DNA片段的真核细胞的大打靶载体。LTVEC的实例包括(但不限于)细菌同源染色体(BAC)和酵母人工染色体(YAC)。

[0063] 术语“重组位点”包括由位点特异性重组酶识别并且可以充当重组事件的底物的核苷酸序列。

[0064] 术语“位点特异性重组酶”包括可以促进“重组位点”之间重组的一组酶。“位点特异性重组酶”的实例包括(但不限于)Cre、Flp和Dre重组酶。

[0065] 关于核酸序列的术语“生殖系”包括可以传递至后代的核酸序列。

[0066] 短语“可操作地连接”是指以其预期方式将功能连接在一起的组分。在一种情况下,编码蛋白质的核酸序列可能可操作地连接至调控序列(例如,启动子、增强子、沉默子序列等)以保持适当的转录调控。

[0067] 如本文所用的术语“长链非编码RNA”或“lncRNA”包括长度大于200个核苷酸的非蛋白质编码转录物。

[0068] 如本文所用的术语“长链基因间非编码RNA”或“lincRNA”包括lncRNA亚群。如本文所用,lincRNA不与基因组的蛋白质编码区的外显子重叠。

[0069] 术语“基因座”被定义为基因组DNA内的DNA区段。例如,lncRNA基因座是编码lncRNA的基因组DNA内的DNA区段。

[0070] I.包含至少一个lncRNA基因座的遗传修饰的组合物

[0071] 提供包含lncRNA功能缺失的非人类动物、细胞、组织和胚胎,其包括(但不限于)一个或多个lncRNA的破坏或敲除。提供用于操纵lncRNA表达的方法和组合物。此外提供针对修饰或敲除lncRNA的打靶组合物。提供展现与一个或多个lncRNA的非功能相关表型的非人类动物、细胞和组织。虽然以下说明书是关于某些特定lncRNA的研究,但可用任何lncRNA实施所述方法和组合物。

[0072] 本文提供包含至少一个长链非编码RNA(lncRNA)基因座中的靶向遗传修饰的非人类动物、细胞、组织和胚胎。在这些情况下,修饰的lncRNA基因座包含编码lncRNA的核酸序列中的功能缺失型突变。此外提供源自包含至少一个lncRNA的功能缺失型突变的非人类动物的细胞、组织和胚胎。

[0073] 术语“功能缺失”当关于lncRNA时可以包括导致lncRNA表达的降低或缺乏和/或lncRNA活性/功能的降低或缺乏的lncRNA基因座中的任何修饰。可例如通过测定细胞或生物体中的lncRNA水平来直接测量所述lncRNA的表达水平。

[0074] 一般来说,如果lncRNA表达水平和/或lncRNA活性水平在统计学上低于($p \leq 0.05$)尚未被遗传修饰或诱变以抑制lncRNA表达和/或活性的适当的对照细胞或生物体中的lncRNA水平,则lncRNA的表达水平和/或活性降低。在特定实施方案中,相对于尚未被修饰以具有降低的lncRNA的水平和/或活性的对照细胞或生物体,lncRNA的浓度和/或活性降低

至少1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、99%或更多。

[0075] 在其他情况下，使用包括(但不限于)Southern印迹分析、DNA测序、PCR分析或表型分析的方法来选择具有降低lncRNA的表达水平和/或活性的靶向遗传修饰的细胞或生物体。然后将这些细胞或生物体用于本文所述的各种方法和组合物中。

[0076] “受试细胞”或“受试生物体”是其中已实现遗传改变如本文所公开的遗传修饰者，或者是从所改变的细胞/生物体遗传并且包含改变的细胞/生物体。“对照”或“对照细胞”或“对照生物体”提供用于测量受试细胞或生物体的表型变化的参考点。在一个实施方案中，对照细胞/生物体与在lncRNA中具有所述遗传修饰的细胞/生物体尽可能紧密地匹配，不同之处在于其缺乏所述遗传修饰或突变，从而导致表达和/或活性降低(例如，相应的细胞可源自相同的细胞系)。在其他情况下，所述对照细胞/生物体可包含例如：(a)野生型细胞/生物体，即具有与用于产生受试细胞/生物体的遗传改变的原料相同的基因型；(b)具有与原料相同的基因型但已经用空构建体(即对所关注性状没有已知作用的构建体，例如包含标记基因的构建体)遗传修饰的细胞/生物体；(c)作为受试细胞/生物体的非遗传修饰后代的细胞/生物体(即，对照细胞和受试细胞源自相同的细胞系)；(d)在遗传上与受试细胞/生物体相同但不会暴露于将诱导所关注基因表达的条件或刺激的细胞/生物体；或(e)受试细胞/生物体自身，在其中所述遗传修饰不会导致所关注多聚核苷酸表达改变的条件下。

[0077] 关于动物、细胞、组织或胚胎的术语“动物”包括哺乳动物、鱼类和鸟类。哺乳动物包括例如人类、非人类灵长类动物、猴、猿、猫、狗、马、牛、鹿、野牛、绵羊、啮齿动物(例如，小鼠、大鼠、仓鼠、豚鼠)、家畜(例如，牛物种，例如奶牛、阉牛等；羊物种，例如绵羊、山羊等；和猪物种，例如猪和野猪)。鸟类包括例如鸡、火鸡、鸵鸟、鹅、鸭等。还包括家养动物和农业动物。关于动物、细胞、组织或胚胎的短语“非人类动物”不包括人类。

[0078] 在一个实施方案中，所述动物是非人类动物。在另一个实施方案中，所述非人类动物是哺乳动物。在另一个实施方案中，所述哺乳动物是啮齿动物。在另一个实施方案中，所述啮齿动物是小鼠、大鼠或仓鼠。

[0079] 如本文所述的遗传修饰可包括从所关注lncRNA基因座的一个或多个缺失、向所关注lncRNA基因座的添加、所关注lncRNA基因座的置换和/或其任何组合。所关注基因座可包含编码区或非编码调控区。

[0080] 本文提供的遗传修饰被靶向所关注的lncRNA基因座。lncRNA的功能缺失可由lncRNA基因中的靶向遗传修饰(即，调控区、编码区、外显子和/或内含子等中的遗传修饰)产生。这类靶向修饰包括(但不限于)一个或多个核苷酸的添加、一个或多个核苷酸的缺失、一个或多个核苷酸的取代、lncRNA基因座的破坏、lncRNA基因座或其部分的敲除、lncRNA基因座或其部分的敲入、用异源核酸序列置换内源lncRNA核酸序列或其部分，或其组合。在特定实施方案中，改变至少1、2、3、4、5、7、8、9、10、50、100、400或更多个核苷酸以形成靶向基因组修饰。

[0081] 在一个实施方案中，所述功能缺失型突变的特征在于至少一个lncRNA功能的破坏或敲除。

[0082] 所述lncRNA基因座可在基因座的任何区域中被遗传修饰以使得所述修饰导致lncRNA的功能缺失。在一个实施方案中，lncRNA基因座的修饰包含整个lncRNA编码区或其部分的缺失。在一个实施方案中，修饰的lncRNA基因座包含编码lncRNA或其部分的一个或

多个外显子的缺失。在另一个实施方案中，所述缺失包含lncRNA基因座内起始于所述lncRNA基因座的第一外显子的一个或多个外显子的缺失。在其他实施方案中，所述缺失包含lncRNA基因座内起始于所述lncRNA基因座的第二外显子的一个或多个外显子的缺失。

[0083] 在一些情况下，所述lncRNA基因座或其部分被插入核酸置换。在这些情况下，所述置换可以是用插入核酸置换lncRNA基因座或其部分的整个RNA编码区，用插入核酸置换lncRNA基因座的一个或多个外显子，用插入核酸置换lncRNA基因座内起始于所述lncRNA基因座的第一外显子的一个或多个外显子，或用插入核酸置换lncRNA基因座内起始于第二外显子的一个或多个外显子。

[0084] 在一些情况下，插入核酸位于lncRNA基因座中以使得其可操作地连接内源lncRNA启动子以使得所述内源lncRNA启动子驱动所述插入核酸的表达。在这些情况下，核酸序列的表达遵循lncRNA的表达模式。

[0085] 在一个实施方案中，所述lncRNA基因座或其部分被包含编码报道基因的第一核酸序列的插入核酸置换。例如，在插入核酸包含报道基因并位于可操作地连接至lncRNA启动子的lncRNA基因座中的情况下，所述报道基因的表达受到内源lncRNA启动子驱动。可选地，插入核酸并未插入与内源lncRNA启动子的可操作连接中。在这些情况下，插入核酸可包含启动子。在一个实施方案中，插入核酸包含可操作地连接至驱动报道基因表达的启动子的报道基因。

[0086] 在一个实施方案中，所述lncRNA基因座或其部分被包含编码可选择标记的第二核酸序列的插入核酸置换。在这些情况下，所述第二核酸序列可操作地连接至驱动可选择标记表达的启动子。

[0087] 在另一个实施方案中，所述lncRNA基因座或其部分被包含报道基因和可选择标记基因的插入核酸置换。在这些情况下，所述报道基因和/或所述可选择标记基因可能会或可能不会可操作地连接至启动子。

[0088] 本文其他地方提供可用于所述方法和组合物中的各种启动子。

[0089] 这些遗传修饰(包括导致靶标lncRNA的表达和/或活性降低或调节的那些)也能够通过生殖系遗传。在特定实施方案中，所述遗传修饰导致所希望的靶标基因座的敲除。这些非人类动物例如可用于如本文其他地方所讨论的多种实验系统中。

[0090] 例如，lncRNA敲除提供动物模型以研究lncRNA功能、lncRNA在发育中的作用，和lncRNA在各种细胞途径和疾病(包括但不限于衰老(包括过早衰老)、大脑发育、胚胎发育、肺发育、骨骼发育、肌肉发育、癌症或转录调控)中的作用。

[0091] 可使用各种方法来产生靶向遗传修饰并且在本文其他地方进行描述。

[0092] A.lncRNA

[0093] 在本文提供的方法和组合物中使用的非人类动物、细胞、组织和胚胎具有导致至少一个lncRNA功能缺失的遗传修饰。lncRNA是大于200个核苷酸的长链非编码RNA。lncRNA的亚群，长链基因间非编码RNA(lincRNA)，是基因间的并且不与蛋白质编码区重叠。

[0094] 在本文提供的方法和组合物中，任何lncRNA基因座可被修饰。在一个实施方案中，遗传修饰的非人类动物、细胞、组织或胚胎包含lncRNA中的遗传修饰。在另一个实施方案中，所述lncRNA是lincRNA。

[0095] lncRNA的非限制性实例包含HOTAIR、HOTTIP、Hoxa11os(原名为HoxA11as)、Pantr1

(原名为lincRNA-Brn1-a)、Pantr2(原名为lincRNA-Brn1-b)、Ptgs2os2(原名为lincRNA-Cox2)、Eldr(原名为Fabl和lincRNA-Egfr)、lincenc1(原名为lincRNA-Enc1)、Mannr(原名为lincRNA-Evi1)、Fendrr(原名为lincRNA-Foxf1)、Halr1(原名为Haunt和lincRNA-HoxA1)、Haglr(原名为Mdgt和lincRNA-HoxD3)、Celrr(原名为Celr和lincRNA-Insig2)、Crnde(原名为lincRNA-Irx5)、Kantr(原名为Spasm和lincRNA-Jarid1c)、Pint(原名为linc-Pint和lincRNA-Mkln1)、Trp53cor1(原名为lincRNA-p21)、lincppara(原名为lincRNA-Ppara)、Peril(原名为lincRNA-Sox2)、Tug1(原名为lincRNA-Tug1)或其组合。

[0096] 现在很显然,蛋白质编码基因不是全部都属于基因组(Mattick,J.S.(2009),PLoS Genet 5:e1000459)。在哺乳动物细胞中的大规模全基因组表达研究已经揭示出,约四分之三的基因组能够以RNA形式表达(Carninci,P.等,(2005),科学(Science)309:1559-1563;Djebali,S.等,(2012),自然(Nature)489:101-108;Kapranov,P.等,(2007),科学(Science)316:1484-1488),并且大部分的转录物不编码蛋白质。在非编码转录物中包括被称为长链非编码RNA(lncRNA)的多种类别。代表来自人类细胞中接近10,000个基因组基因座的约15,000种转录物(Derrien,T.等,(2012),基因组研究(Genome Res)22:1775-1789),lncRNA和被称为长链基因间非编码RNA(lincRNA)的子类(Guttman,M.等,(2009),自然(Nature)458:223-227;Khalil等,(2009))在结构、合成和其基因的染色质特性方面与蛋白质编码mRNA类似。这种结构相似性是否延伸至与蛋白质匹配的功能多样性仍然是一个悬而未决的问题。

[0097] 关于单独的lncRNA的功能研究已经鉴定了在X染色体失活(Marahrens,Y.等,(1997),基因与发育(Genes Dev)11:156-166)、压印(Leighton,P.A.等,(1995),自然(Nature)375:34-39;Mohammad,F.等,(2010),发育(Development)137:2493-2499;Sleutels,F.等,(2002),自然(Nature)415:810-813;Takahashi,N.等,(2009),人类分子遗传学(Hum Mol Genet)18:1879-1888)、视网膜分化(Young,T.L.等,(2005),当代生物学(Curr Biol)15:501-512)以及心脏和体壁发育(Grote,P.等,(2013),发育细胞(Dev Cell)24:206-214)中的作用。关于lincRNA HOTAIR的研究首先揭示出,通过将染色质修饰复合物(在HOTAIR的情况下为多梳抑制复合物2)引导至特定基因组基因座,lincRNA可在远离其自身转录位点的位点处调控基因表达(Rinn,J.L.等,(2007),细胞(Cell)129:1311-1323)。对于Xist lncRNA在X染色体失活中(Zhao,J.等,(2008),科学(Science)322:750-756)以及对于AIR和Kcnq1ot1lncRNA在压印中已经发现类似的作用机制。这些研究结果表明lncRNA在调控基因表达中的较宽泛作用,这已经通过lincRNA和蛋白质编码基因的相关表达模式的分析得到支持,其指向lincRNA参与一组广泛范围的细胞过程和器官系统生理学(Guttman等,(2009))。许多关于lncRNA的最近研究已采用全局基因组策略,其已经建立了一类lncRNA的作用的总体情况。为了回答lncRNA对蛋白质基因表达的作用是否是广泛的、微妙的、和缓冲或特异性的、直接的和决定性的问题,需要研究其在活体动物中的单独作用。

[0098] 本文在以下说明书中提供导致小鼠基因敲除模型中的各种lncRNA敲除的遗传修饰的非限制性实例。在基因敲除小鼠中对20种lincRNA基因进行了基因表达和表型的研究,其包括LacZ谱分析,其显示组织特异性表达的不同时空模式;揭示出展现围产期致死的两种基因敲除系;并且揭示出其他表型包括过早衰老相关表型和肺、骨骼、大脑和肌肉的缺陷。

[0099] 为了开始研究lncRNA在活体动物情形中的功能，对于二十种lincRNA基因产生基因敲除小鼠系。每种突变等位基因带有lacZ报道基因，其表达谱分析揭示出在成体和胚胎中的广泛范围的时空和组织特异性转录模式。在18种纯合敲除系中，有六种(约33％)展现可辨别的突变表型，其中两种(11％)是围产期致死的，并且包括过早衰老相关表型，在大脑、肺、骨骼和肌肉中的异常形态，和基因表达模式的全局变化，其共同指向这类新型功能性RNA在胚胎发育中和在一组广泛的组织和器官的生理学中的不同作用。

[0100] B.lncRNA基因座的遗传修饰

[0101] 本文提供用于在非人类动物、细胞、组织或胚胎中的至少一个lncRNA基因座的遗传修饰的方法和组合物。

[0102] 所关注lncRNA基因座的遗传修饰可为如本文在其他地方详细描述的基因座的任何修饰(即缺失、插入、置换等)。在这些情况下，所述遗传修饰导致lncRNA的功能缺失。在一个实施方案中，所述遗传修饰包含至少一个lncRNA的破坏或敲除。

[0103] i.敲除等位基因设计和构建

[0104] 修饰的lncRNA等位基因(例如敲除)的设计和构建因若干技术问题而复杂化。例如，通常缺乏lncRNA的结构-功能关系并且lncRNA基因座不具有开放阅读框。因此，将引导等位基因设计用于修饰蛋白质编码序列如敲除的相同策略可能不会适用于lncRNA。另外，lncRNA基因的边界未被充分限定，这将使修饰lncRNA等位基因如敲除的设计进一步复杂化。下文详细地描述这些技术难题的非限制性实例和本文中用于成功地克服lncRNA敲除设计中的这些障碍的策略。

[0105] 在一个实施例中，本文提供的方法和组合物被应用于作为模型系统的小鼠。然而，即使以下说明是关于小鼠，但在本文所述的方法和组合物中可使用任何非人类动物、细胞、组织或胚胎。

[0106] 因为其发明接近二十五年前，所以用于产生遗传修饰的所谓的基因敲除小鼠的方法已经将小鼠公认为用于研究哺乳动物基因功能的首要系统(Capecchi,M.R.(2001),自然医学(Nat Med)7:1086-1090;Evans,M.J.(2001),自然医学(Nat Med)7:1081-1083;Smithies,O.(2001),自然医学(Nat Med)7:1083-1086)。除了少数例外，基因敲除小鼠技术在单独的基因研究以及大规模国际项目中的应用(Bradley,A.等,(2012),哺乳动物基因组(Mamm Genome)23:580-586)专注于蛋白质编码基因，但是最近产生用于微RNA的全局敲除小鼠资源的努力(Prosser,H.M.等,(2011),自然生物技术(Nat Biotechnol)29:840-845)(mcmanuslab.ucsf.edu/microrna_knockout)表明了将所述技术应用于非编码RNA的价值。

[0107] 然而，将敲除小鼠技术应用于lncRNA不存在一些技术问题和困难。大部分蛋白质具有已知或至少预测对于功能重要的元件或结构域。使这些基本部分的编码序列缺失通常足以产生无效等位基因。同样，可以设计条件性等位基因，其分离关键的一个或多个外显子以稍后通过组织特异性重组酶的作用缺失。因为仅对少数lncRNA建立了结构-功能关系并且不存在开放阅读框作为指导，所以可用于蛋白质编码基因的敲除策略不可用于编码lncRNA的基因组基因座。虽然lncRNA基因注释已得到改善(Derrien等,(2012))，但一些基因的精确边界可能仍然不明确，这可以使敲除等位基因设计变得复杂。应用于基因敲除小鼠的蛋白质编码基因的强大工具是用报道基因例如β-半乳糖苷酶或荧光蛋白的编码序列置换靶基因，所述β-半乳糖苷酶或荧光蛋白的表达通过所述靶基因的启动子进行控制，从

而报道其在小鼠中表达的空间和时间模式。本文其他地方提供报道基因的非限制性实例。

[0108] 报道基因置换已经成功地应用于非编码RNA如充分研究的Gt(ROSA)26Sor基因座(Zambrowicz,B.P.等,(1997),美国国家科学院院刊(Proc Natl Acad Sci USA)94:3789-3794),其编码lncRNA,和短链非编码RNA miR-155的基因(Thai,T.H.等,(2007),科学(Science)316:604-608),但可能需要开发用于产生lncRNA的这些等位基因的规则。虽然存在这些限制,但由于鉴定了数千种lncRNA,探索将基因敲除小鼠技术力量应用于这类新型基因的时机已经成熟。怀着这个目标,本文描述了二十种lincRNA的基因敲除小鼠系的产生,例如,各自带有用例如β-半乳糖苷酶报道基因置换的基因切除的缺失等位基因。

[0109] 可通过本文提供的方法和组合物修饰任何lncRNA基因座。在一个实施方案中,所述lncRNA是长链基因间非编码RNA(lincRNA)。lincRNA基因的非限制性实例列于表1中,但是本文提供的方法和组合物可用任何lncRNA实施。

[0110] 表1列出在这项研究中靶向的10种不同染色体上的20种lincRNA基因和所产生的26种敲除缺失等位基因。针对突变选择长链基因间非编码RNA类别的成员,因为根据定义,lincRNA基因是从邻近的蛋白质编码基因分离的并且其转录物不重叠(Guttman等,(2009))。这种特征允许设计缺失等位基因,其将使得与附近基因表达相互作用的可能性最小。选择靶向lincRNA基因以反映多种表达模式(Cabili,M.N.等,(2011),基因与发育(Genes Dev)25:1915-1927;Khalil,A.M.等,(2009),美国国家科学院院刊(Proc Natl Acad Sci USA)106:11667-11672),重点是神经表达,并且其可能涉及基因表达的开发和调控中。

[0111]

表 1. LincRNA 敲除缺失等位基因	
	缺失特性

[0112]

表 1. LincRNA 敲除缺失等位基因			
LincRNA	缺失起始外显子[1]	大小 (kb)	基因组坐标[2]
HOTAIR	E1	2.3	Ch15: 102945399–102947720
	E2	0.43	Ch15: 102945399–102945826
HOTTIP	E1	4.8	Ch6: 52262834–52267603
	E2	2.2	Ch6: 52265374–52267603
Hoxa11os (*HoxA11as*)	E3	3.5	Ch6: 52246320–52249795
	E4	3.1	Ch6: 52246643–52249795
	E5	0.70	Ch6: 52249094–52249795
Pantr1 (*LincRNA-Brn1-a*)	E1	47	Ch1: 42648175–42694815
Pantr2 (*LincRNA-Brn1-b*)	E1	6.5	Ch1: 42707143–42713698
Ptgs2os2 (*lincRNA-Cox2*)	E1	5.9	Ch1: 150159024–150164899
Eldr (*Fabl, LincRNA-Egfr*)	E1	17	Ch11: 16934419–16951083
Lincenc1 (*LincRNA-Enc1*)	E2	26	Ch13: 97455710–97482249
Mannr (*LincRNA-Evi1*)	E1	32	Ch3: 29891188–29923147
Fendrr (*LincRNA-Foxf1*)	E2	19	Ch8: 121054882–121074065
Halr1 (*Haunt*、*LincRNA-HoxA1*)	E2	8.6	Ch6: 52106776–52115377
Haglr (*Mdgt*、*LincRNA-HoxD3*)	E1	12	Ch2: 74750433–74762886
Celrr (*Celr*、*LincRNA-Insig2*)	E2	50	Ch1: 121087772–121137464
Crnde (*LincRNA-Irx5*)	E2	25	Ch8: 92325913–92350749
Kantr (*Spasm*、*LincRNA-Jarid1c*)	E1	29	ChX: 152298544–152327475
Pint (*Linc-Pint*、*LincRNA-Mkln1*)	E2	32	Ch6: 31166026–31197846
Trp53cor1 (*LincRNA-p21*)	E1	22	Ch17: 29057474–29079078
	E2	2.9	Ch17: 29057474–29060353
Lincppara (*LincRNA-Ppara*)	E1	29	Ch15: 85671665–85701064
Peril (*LincRNA-Sox2*)	E1	14	Ch3: 34767849–34782292
Tug1 (*LincRNA-Tug1*)	E1	9.0	Ch11: 3639794–3648758
	E2	5.7	Ch11: 3639794–3645518
[1]LincRNA 基因通常具有多种注释转录物。所有缺失都终止于最后的注释外显子。 [2]GRCm38 (GCA 000001635.3)			

[0113] 通过两个目标指导lincRNA敲除突变的设计策略。首先,产生等位基因,其将精确地报道lincRNA基因的转录活性。虽然关于组织特异性lincRNA表达存在大量证据(Cabili等,(2011)),但需要通过产生由lacZ表达谱分析提供的较高清晰度表达模式来补充这种知识基础,其可以在空间上和时间上解析组织和器官表达并揭示子结构域,以及在一些情况下通过组织解剖实验没有解析的细胞类型特异性。另外,已公布的lincRNA敲除等位基因都没有并入报道基因。其次,产生基因切除缺失,其废除了lincRNA的合成和功能以使得任何

与突变相关的表型都将指示靶向RNA的关键功能。敲除缺失的大小范围为约400bp至50kb，其中一半缺失所有的注释外显子。对于大多数剩余等位基因，缺失起始于第二外显子。将VelociGene®工程改造方法(Valenzuela,D.M.等,(2003a),自然生物技术(Nat Biotechnol)21:652-659)应用于构建和使用基于细菌人工染色体的大打靶载体(LTVEC)对于构建确保这类新型功能性大RNA的无效等位基因所需的大的基因切除缺失是至关重要的。

[0114] 关于可指导等位基因设计的lincRNA基因的结构与功能之间的关系知之甚少。关于Gt(ROSA)26Sor(Zambrowicz等,(1997))和BIC(miR-155)(Thai等,(2007))基因破坏的经验确定，在第一外显子后的缺失和插入可以产生β-半乳糖苷酶或其他报道基因的可靠的和组织特异性表达。然而，如果来自修饰等位基因的融合转录物仍然是来自在第一外显子中编码的5'部分的lincRNA的功能部分，则这种策略可能无法实现完全无效突变(Tsai,M.C.等,(2010),科学(Science)329,689-693)。因此，表1中指示的敲除等位基因设计是希望完全切除突变(其将废除lincRNA功能的可能性最高)与产生等位基因(其从β-半乳糖苷酶报道基因产生精确和翔实的基因表达谱)的目标之间的折衷。例如，对于HOTAIR基因，产生两种等位基因，一种缺失几乎整个RNA编码序列并且第二种在第二外显子中起始缺失。两种等位基因产生相同表型(描述于下)，但仅第二种充当基因表达的报道基因。

[0115] 对于处于非常接近蛋白质编码基因并且可共有相异启动子的lincRNA，缺失起点设置在第二外显子中以避免破坏邻近基因转录的可能性。图1示出Fendrr(lincRNA-Foxf1)基因的这种实例。该图示出所有等位基因共有的设计元件的实例：所有或大部分的编码lincRNA的序列的靶向缺失和被含有来自编码β-半乳糖苷酶的大肠杆菌lacZ基因的序列的表达盒和表达用于选择G418抗性ES细胞群落的新霉素磷酸转移酶的表达盒(neo^r)置换。在表型分析之前能够实现Cre介导切除的LoxP重组酶识别位点侧接药物选择盒。因为不存在与lacZ序列融合的功能性开放阅读框，所以每个等位基因可带有起始密码子和Kozak共有序列(Kozak,M.(1987),核酸研究(Nucleic Acids Res)15,8125-8148)以有效翻译β-半乳糖苷酶报道基因。Kozak共有序列的非限制性实例是A/GCCRCCATGG(SEQ ID NO:1)和GCCGCCRCCATGG(SEQ ID NO:2)，其中R是A或G。

[0116] 将LTVEC打靶载体引入ES细胞中并通过等位基因缺失法筛选正确靶向的克隆(Frendewey,D.等,(2010),酶学方法(Methods Enzymol)476,295-307)。

[0117] 将VelociMouse®法(Poueymirou,W.T.等,(2007),自然生物技术(Nat Biotechnol)25,91-99)应用于8细胞胚胎阶段注射以将靶向ES细胞转化成完全ES细胞源F0代杂合小鼠以准备用于lacZ表达谱分析或培育成纯合性。本文其他地方提供的实施例1-13中提供了用于产生lncRNA敲除动物的方法的其他细节。

[0118] ii.报道基因表达谱分析

[0119] 如本文其他地方所述，lncRNA基因座的遗传修饰可包含用插入核酸置换lncRNA基因座或其部分或者lncRNA基因座或其部分的插入/添加。在一些情况下，所述插入核酸包含报道基因。在一个实施方案中，所述报道基因位于可操作地连接内源lncRNA启动子的lncRNA基因座中。这种修饰允许由内源lncRNA启动子驱动的报道基因的表达。可选地，报道基因并非安置于与内源lncRNA启动子的可操作连接中。

[0120] 任何报道基因(或可检测部分)可用于本文提供的方法和组合物中。报道基因的非

限制性实例包括例如β-半乳糖苷酶(由lacZ基因编码)、绿色荧光蛋白(GFP)、增强型绿色荧光蛋白(eGFP)、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强型黄色荧光蛋白(EYFP)、Emerald、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶或其组合。

[0121] 以下说明是利用编码β-半乳糖苷酶的lacZ报道基因的非限制性实例。本文所述的方法和组合物可利用任何报道基因进行。

[0122] 为了研究20种靶向lincRNA基因的表达模式,将关于β-半乳糖苷酶活性的X-gal染色施用在整个胚胎或来自成年小鼠的整个固定组织和器官上。靶向lincRNA基因展现多种独特的报道基因表达模式(图11的表2),这代表了大部分的主要器官系统和组织类型。所述报道基因表达模式指示大部分的lincRNA在多个成体组织中转录,其中一种基因Pint展现在所研究的所有组织中的普遍表达。对于约三分之一的lincRNA基因,表达局限于单一器官,例如,大脑为Pantr2、Kantr和Haglr,肺为Mannr和Fendrr,泌尿生殖系统为Eldr,并且胸腔为Halr1。三种lincRNA基因(包括HOTAIR、Ptgs2os2和Haglr)不展现在任何成体组织中的表达。

[0123] 胚胎表达似乎是lincRNA的共同特征。在胚胎期12.5天(E12.5)时或附近的杂合胚胎中的β-半乳糖苷酶报道基因表达的研究揭示出所有20种靶向lincRNA基因的多种特定模式(图11的表2、图2A)。表达谱范围为普遍存在(Tug1)至高度受限,例如表皮为Eldr,须基板为Trp53cor1(图9),或乳腺芽为Lincenc1(图9)。在不同程度的HOTTIP和Hoxa11os的肢芽和尾部表达中可见的时空模式与关于HoxA簇中的相邻蛋白质编码基因所报道的那些非常类似(Hostikka,S.L.和Capecchi,M.R.(1998),发育机制(Mech Dev)70:133-145;Lu,P.等,(2008),发育(Development)135:1395-1405)。对于β-半乳糖苷酶报道基因所观测到的在后尾芽和生殖结节中的HOTAIR表达与通过原位杂交所确定的相同(Schorderet,P.和Duboule,D.(2011),PLoS Genet 7:e1002071)。在胚胎发育不同时间的β-半乳糖苷酶染色分析显示,对于一些lincRNA,表达早期开始于限制位点并且然后在稍后阶段扩展超出这个初始基因座(图2B),再次让人联想到Hox蛋白表达(Nagy,A.(2003)操纵小鼠胚胎:实验室手册(Manipulating the mouse embryo:a laboratory manual),Cold Spring Harbor,N.Y.,Cold Spring Harbor Laboratory Press)。例如,HOTTIP和Hoxa11os基因的表达开始于E9.5胚胎的极后期并且然后在稍后时间扩展至肢芽中。类似地,在E9.5时接近胚胎前端的位点处Celrr的初始表达得以维持并且在接下来两天内扩展至脊髓全长。

[0124] 与在人类组织特异性lncRNA中可见的频繁大脑表达一致(Derrien等,(2012)),二十种靶向小鼠lincRNA基因中的一半在大脑中具转录活性。关于胚胎lincRNA表达,大脑模式(图3)是独特的并且从普遍存在(lincppara和Pint)改变为高度受限的(Peril、Crnde和Kantr)。在所靶向的20种lincRNA基因中,仅Pint显示全局整体表达模式,其主要局限于出生后生活。Pint独有的是,观测到其表达随年龄增加(图4)。在3天龄新生儿中,Pint转录活性较低(大脑)或不可检测(胸腔肌肉),但是随后逐渐出现在3周龄小鼠中并且到8周龄时变得强烈和普遍存在。虽然Pint表达的强度和时机在不同器官和组织中改变,但总体趋势是在出生后表达稳步增加至成年达到平台期。这种年龄相关的动态表达模式是新颖的;本发明人没有观测到针对数百种蛋白质编码基因敲除的lacZ谱分析实验的类似概况。

[0125] iii.表型

[0126] lncRNA基因座的遗传修饰可在本文提供的非人类动物中产生多种表型。这些表型可包括例如过早衰老相关表型，包括大脑、骨骼、肌肉或肺的多种器官的发育缺陷，胚胎发育缺陷，围产期或胚胎致死，脱发，过早生长停滞，脊椎弯曲或异常姿势。

[0127] 在一个实施方案中，包含如本文所述的至少一个修饰的长链非编码RNA(lncRNA)的非人类动物的特征在于具有以下表型中的一种或多种：(a)过早衰老相关表型；(b)围产期致死；(c)肺发育缺陷；(d)尾部和后肢的形态畸形；(e)一个或多个组织中的肌肉质量损失；(f)大脑发育缺陷；或(g)(a)-(f)的其任何组合。

[0128] 在一个实施方案中，lncRNA基因座的遗传修饰导致致死。在一些情况下，lncRNA基因座的修饰是胚胎致死。在一个实施方案中，lncRNA基因座的修饰是围产期致死。在一个实施方案中，Fendrr lncRNA或Peril lncRNA的破坏或敲除导致围产期致死。在另一个实施方案中，Haglr的破坏或敲除导致致死。

[0129] 在一个实施方案中，lncRNA基因座的遗传修饰导致过早衰老相关表型。在这些动物中，过早衰老征象可包括例如体重缓慢增加，体重过早达到平台期，过早生长停滞，到约12周龄时脊椎弯曲，到约26周龄时严重脊椎弯曲，在约六个月时脱毛，在约6个月时后肢肌肉力量损失，或其组合。在一个实施方案中，导致过早衰老相关表型的遗传修饰是Pint的破坏或敲除。在一个实施方案中，所述lncRNA是Pint，并且所述非人类动物的特征在于过早衰老相关表型，其包含：(a)相比于野生型对照的较缓慢生长速率；(b)肌肉力量下降；(c)纤维化；(d)相比于野生型对照的较低体脂含量；(e)相比于野生型对照的较低股骨骨矿物质密度和骨质量；(f)相比于野生型对照降低的肌肉质量；(g)中值寿命降低；(h)脊椎弯曲；(i)器官萎缩；或(j)(a)-(i)的其任何组合。

[0130] lncRNA基因座的功能缺失型遗传修饰也可导致大脑发育缺陷。在一个实施方案中，导致过早衰老相关表型的遗传修饰是Pantr2、Kantr、Peril、Celrr、Pantr1、Crnde、lincenc1、Pint、lincppara或Tug1的破坏或敲除。在一个特定实施方案中，所述lncRNA是Pantr2。在另一个特定实施方案中，所述lncRNA是Pint。

[0131] 进行本文提供的lincRNA敲除小鼠的各种实例的表型分析并且描述于下文。

[0132] 通过lacZ谱分析揭示的全身Pint表达的引人注目的年龄相关增加(图4)表明，Pint在随着小鼠年龄维持正常健康中具有全局自我平衡作用。为了测试这种假说，将敲除等位基因培育成纯合性并且从出生起至26周龄追踪野生型(WT)、杂合(Het)和纯合敲除(KO)小鼠，并且研究生长速率和任何明显的健康状况不佳的缺陷迹象。所述Pint KO小鼠相比于WT小鼠随着年龄以较缓慢速率增加体重并且在较早期和在显著较低的体重下达到体重平台期(图5A和图5B)，这意味着过早生长停滞。雄性和雌性KO和Het小鼠展现缓慢生长表型，但是这在雄性中更明显。通过对达到年龄的单独小鼠的显微CT分析的骨骼成像揭示出，到12周龄时约70％的雄性和雌性Pint KO小鼠中出现脊椎弯曲，其中接近90％的26周龄KO小鼠显示严重的脊椎弯曲(图5C和图5D)。相比之下，仅10％至20％的26周龄WT小鼠显示轻微的年龄相关脊椎弯曲。在Pint Het小鼠中直到26周龄时仍未出现显著的脊椎弯曲，这表明Pint的不同寻常的年龄依赖性单倍剂量不足。也观测到在6月龄KO小鼠中的年龄相关脱毛，其在雌性(10只KO中的5只)中比雄性(9只中的2只)中更严重并且在同龄的Het中仅一只存在这种现象并且同龄的WT小鼠中都不存在这种现象。在约三分之二的6月龄Pint KO小鼠(雌性为60％，雄性为67％)中注意到不太严重的表型(当从尾部悬吊时的后肢紧握行为)，

相比之下同龄WT小鼠为约20％。这种表型可表明后肢肌肉力量的年龄相关损失(关于HOTTIP敲除系中的这种表型的另一个实例,参见图8)。Pint敲除小鼠中的突变相关缺陷范围表明过早衰老相关表型。

[0133] 在被培育成纯合性的19种lincRNA敲除小鼠系中,有两种(11％)(Peril和Fendrr)显示围产期致死(L.A.Goff等,未公布)。最近已报道了Fendrr基因的敲除(Grote等,(2013))。等位基因由在第一外显子中插入转录终止元件组成。在带有Fendrr的纯合突变的胚胎中,Grote等,(2013)观测到在约E13.75时的致死,其与突出的脐膨出、腹侧体壁厚度下降和造成右心房积血的心脏缺陷相关。在本文所述的Fendrr敲除系中未观测到这些表型,其从外显子2至最后的注释外显子具有26kb缺失(图1)。E12.5胚胎的X-gal染色显示在额鼻过程、上呼吸道、肺中和后大动脉-生殖腺-中肾(Aorta-Gonad-Mesonephron,AGM)区域中的lacZ表达(图6A),这在杂合(未示出)和纯合胚胎中是相同的。孤立地看在E13.5时发育的肺,揭示出敲除胚胎中的缺陷:肺叶出现坍塌、球状和杂乱(图6B)。本文所述的Fendrr基因的缺失等位基因敲除的纯合小鼠存活至出生,但在明显的呼吸困难后不久死亡。在2种不同的遗传背景上在小鼠中的Fendrr突变围产期致死表型是相同的:此处报道的C57Bl6NTac/129S6SvEvTac杂合背景和在小鼠中在单独的育种计划中进一步回交在C57BL/6背景上。

[0134] 在存活至成年期并展现突变等位基因的正常孟德尔遗传(Mendelian segregation)的lincRNA敲除小鼠中也观测到多种温和表型(L.A.Goff等,未公布)。其中,在多种lincRNA基因敲除如Pantr2、HOTAIR和HOTTIP中的lacZ表达与可检测表型之间存在强烈相关性。对于Pantr2基因切除(6.5kb完全编码序列缺失,表1)的纯合小鼠显示大脑皮层的不当发育(在CNS中强烈染色),其中降低数目的中间祖细胞和神经元迁移缺陷影响上层投射神经元的发育(L.A.Goff等,未公布)。HOTAIR和HOTTIP基因的缺失(完全或部分基因切除,表1)造成完全渗透的形态畸形。在HOTAIR KO小鼠中,观测到第4节尾椎的明显同源转化,其似乎在解剖学上与第3节尾椎类似(图7)(尾芽中的特定短暂lacZ染色)。HOTTIP KO小鼠(其在胚胎肢芽中具有阳性染色)当从其尾部悬吊时相比于野生型同窝出生仔畜显示异常的姿势(图8A)。这种行为异常伴随着握耐力损失,如通过其中将小鼠激发以保持悬吊在倒置的铁丝笼上的测试所测量。野生型和HOTTIP Het突变体抓住不放约一分钟,而其KO同窝出生仔畜在10-20秒内松开(图8B)。这种明显的握力降低与腓肠肌而非胫骨前肌或四头肌的肌肉质量损失相关(图8C)。观测到腓肠肌中纤维数目降低约40％,而非其平均尺寸降低(图8D和图8E)。除了HOTTIP敲除小鼠中的肌肉缺陷之外,还发现骨骼畸形:后肢跟骨骨骼长度缩短(图10)。

[0135] 在过去几年中已经看到,特别是在哺乳动物中,对于基因组的非蛋白质编码组分的理解呈现爆发。除了数十年来已知的非编码功能性RNA类别之外,核糖体、转移、小核、小核仁、小胞质RNA,和RNA酶P、RNA酶MRP和端粒酶的RNA组分,以及最近发现的微RNA和PIWI相关piRNA,现在可包括长链非编码RNA类别的至少10,000种成员(Carninci等,(2005);Derrien等,(2012);Djebali等,(2012);Guttman等,(2009);Kapranov等,(2007))。当我们理解了基因组存在和lncRNA基因表达时,下一个目标是发现其生物功能。作为开始应对这一挑战的第一步骤,已经应用小鼠基因敲除技术(用于测定哺乳动物基因功能的最强大工具)来产生针对主要神经系表达和预期发育功能所选择的20种lincRNA基因的敲除小鼠系的资源。

[0136] 由于lincRNA的未知结构-功能关系，在这项初始研究中关键的是产生具有缺失的敲除等位基因，其除去大部分(若非全部)的lincRNA编码潜能以确保所观测到的任何表型是完全无效等位基因的结果。许多lincRNA基因座的不明确的和复杂的注释(其中多个报道转录物或许是通过可选的剪接或转录起始位点产生)，增加了敲除等位基因设计的难度并且将使得避免亚效等位基因风险的条件性等位基因构建的实现变得困难。对于lincRNA功能重要的分子特性的新理解应向下一代lincRNA等位基因的设计通过对功能关键的更精确定向序列修饰的信息并且也容许先进和灵活的条件性策略。

[0137] 本文所述的lincRNA敲除研究的目标是产生等位基因，除了废除功能之外，也报道所述基因的时空表达模式。虽然没有蛋白质编码开放阅读框作为指导，但成功地设计了报道所有20种靶向基因的基因表达的等位基因。在成体阶段不产生lacZ表达的一种等位基因是Ptgs2os2(关于胚胎表达，参见图2A和图9)，其已知是通过炎性信号最强烈诱导的lincRNA之一(Carpenter,S.等,(2013),科学快讯(Sciencexpress)2013年8月01日；Guttman等,(2009))。在这项研究中没有进行激发实验，但Ptgs2os2敲除系应该证明lincRNA的表达如何响应于感染或其他炎性损伤以及它在过程中起到什么生物作用的有价值的研究资源。

[0138] 所观测到的基因表达模式的多样性和特异性让人联想到关于蛋白质编码基因的报道基因等位基因所见的那些。胚胎表达是由几乎所有所研究lincRNA基因共有的特征。这可能指向lincRNA在调控发育关键事件中的常见作用。观测到在胚胎发育期间的时空模式改变，很像HOX蛋白(图2B)，强烈特定表达例如对于Trp53cor1的须基板染色和对于Lincenc1的乳腺芽表达(图2A和图9)，在成体组织中的普遍表达，例如对于Pint(图4)，和表达模式的时间变化，例如在胚胎发育中关于Celrr可见的定性变化(图2B)，或对于Pint的全局表达随着年龄的新颖定量增加(图4)。因为对于这项研究所选的许多lincRNA基因已知在神经细胞谱系中表达，所以观测到大脑特异性报道基因表达(图3)，但是lacZ谱分析相比于基于细胞或解剖组织的测定法提供具有更丰富生物信息的更高分辨率。

[0139] 在培育成纯合性的19种lincRNA敲除系中所观测到的表型中，观测到两次致死(11％)，该频率可能低于对于20种随机选择的蛋白质编码基因的敲除所预期。有些低的致死率以及关于HOTAIR(图6)、HOTTIP(图7和图8)、Pantr2(L.A.Goff等,未公布)和其他(M.Sauvageau等,未公布)可见的非致死表型的细小区别和频繁的胚胎表达表明lincRNA可缓冲或调节基因表达或其他过程而非提供单一关键功能。以这种方式，lincRNA可能与其较小的非编码表亲miRNA类似，原因在于它们可能与其他功能性lncRNA共用冗余和重叠的靶标和功能。

[0140] 这项工作的一个目标是产生具有共同等位基因策略和功能性报道基因能力的lincRNA敲除小鼠系资源，其可充当更深入的表达和表型研究的主题。在每种情况下添加LacZ表达盒允许同时破坏基因功能和通过X-Gal染色研究lincRNA表达模式的调控。这些研究揭示在小鼠胚胎发生和整个成年期内lincRNA表达的动态空间和时间模式，提供对于这类新型分子在体内的调控/功能特性的显著洞察并且可寻求这些基因具功能性的精确区域。这项研究可充当用于突变lincRNA类别所有成员的大规模项目的模型，与国际蛋白质编码基因敲除小鼠联合会所实现者类似(Bradley等,(2012))。

[0141] II.用于修饰非人类动物中的lncRNA基因座的方法

[0142] 本文提供用于遗传修饰非人类动物、细胞、组织或胚胎中的lncRNA基因座的方法。

[0143] 可通过本文提供的方法修饰任何lncRNA基因座。lncRNA基因的非限制性实例包含HOTAIR、HOTTIP、Hoxa11os(原名为HoxA11as)、Pantr1(原名为lincRNA-Brn1-a)、Pantr2(原名为lincRNA-Brn1-b)、Ptgs2os2(原名为lincRNA-Cox2)、Eldr(原名为Fabl和lincRNA-Egfr)、lincenc1(原名为lincRNA-Enc1)、Mannr(原名为lincRNA-Evi1)、Fendrr(原名为lincRNA-Foxf1)、Halr1(原名为Haunt和lincRNA-HoxA1)、Haglr(原名为Mdgt和lincRNA-HoxD3)、Celrr(原名为Celr和lincRNA-Insig2)、Crnde(原名为lincRNA-Irx5)、Kantr(原名为Spasm和lincRNA-Jarid1c)、Pint(原名为linc-Pint和lincRNA-Mkln1)、Trp53cor1(原名为lincRNA-p21)、lincppara(原名为lincRNA-Ppara)、Peril(原名为lincRNA-Sox2)、Tug1(原名为lincRNA-Tug1)或其组合。

[0144] 在一个实施方案中,提供用于修饰多能性细胞中的所关注lncRNA基因座的方法。这种方法包括:(a)在多能性细胞中引入包含侧接可以与lncRNA基因座进行同源重组的5'和3'同源臂的插入核酸的打靶构建体;和(b)鉴定在所述lncRNA基因座包含靶向遗传修饰的修饰多能性细胞。在这些方法中,所述遗传修饰导致lncRNA的功能缺失。在一个实施方案中,所述多能性细胞是小鼠或大鼠胚胎干细胞。在另一个实施方案中,所述多能性细胞是人类iPS细胞。

[0145] A.打靶载体和插入核酸

[0146] 此外提供有待用于产生本文提供的遗传修饰的非人类动物、细胞、组织或胚胎的方法中的打靶载体或打靶构建体。

[0147] 在一个实施方案中,提供包含侧接可以与所关注的lncRNA基因座进行同源重组的5'和3'同源臂的插入核酸的打靶载体。

[0148] 下文详细地描述打靶载体和所述打靶载体的组分的实例(即插入核酸、所关注的多聚核苷酸、表达盒等)。

[0149] i.插入核酸

[0150] "插入核酸"或"插入多聚核苷酸"包含希望整合在靶标基因座的DNA区段。在一个实施方案中,所述插入核酸包含一个或多个所关注的多聚核苷酸。在其他实施方案中,所述插入核酸可包含一个或多个表达盒。给定表达盒可包含所关注的多聚核苷酸、编码选择标记和/或报道基因的多聚核苷酸以及影响表达的各种调控组分。

[0151] 所关注的任何多聚核苷酸可包含在各种插入多聚核苷酸中并且从而整合在靶标基因组基因座。本文公开的方法提供至少1、2、3、4、5、6个或更多个所关注的多聚核苷酸以整合至所关注的靶向lncRNA基因组基因座中。

[0152] 在一个实施方案中,包含在插入核酸中的所关注多聚核苷酸编码报道基因。在另一个实施方案中,所关注的多聚核苷酸编码可选择标记。

[0153] 在一个实施方案中,所关注的多聚核苷酸可以侧接位点特异性重组位点。在一个特定实施方案中,所述位点特异性重组位点侧接编码报道基因的区段和/或编码可选择标记的区段。

[0154] 本文其他地方详细地讨论可以包括在插入核酸内的所关注多聚核苷酸的非限制性实例,包括选择标记和报道基因。

[0155] 当整合在靶标lncRNA基因座时在插入多聚核苷酸内的所关注多聚核苷酸可以在

细胞中引入一个或多个遗传修饰。所述遗传修饰可包含内源核酸序列的缺失和/或在靶标基因组基因座中添加外源或异源或直系同源多聚核苷酸。在一个实施方案中，所述遗传修饰包含在靶标基因组基因座处用所关注的外源多聚核苷酸置换内源核酸序列。因此，本文提供的方法允许产生在靶标lncRNA基因座中包含敲除、缺失、插入、置换(“敲入”)、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换或其组合的遗传修饰。在第一、第二、第三、第四、第五、第六、第七或任何后续插入多聚核苷酸整合至靶标基因组基因座中之后可发生这些修饰。

[0156] 在插入多聚核苷酸内和/或整合在靶标基因组基因座的所关注多聚核苷酸可以包含与将其引入的细胞天然或同源的序列；所关注的多聚核苷酸可能是与将其引入的细胞异源的；所关注的多聚核苷酸可能是与将其引入的细胞外源的；所关注的多聚核苷酸可能是与将其引入的细胞直系同源的；或所关注的多聚核苷酸可能是来自与将其引入的细胞不同的物种。关于序列的术语“同源”是细胞的天然序列。关于序列的术语“异源”是来源于外来物种的序列，或者如果来自相同物种，则通过有意的人为干预基本上在组成和/或基因组基因座中从其天然形式修饰。关于序列的术语“外源”是源自外来物种的序列。术语“直系同源”是来自与另一个物种(即，物种变体)中的已知参考序列在功能上等效的一个物种的多聚核苷酸。所关注的多聚核苷酸可能是来自任何所关注的生物体，包括(但不限于)原核生物、真核生物、非人类、啮齿动物、仓鼠、小鼠、大鼠、人类、猴、禽类、农业哺乳动物或非农业哺乳动物。所关注的多聚核苷酸还可包含编码区、非编码区、调控区或基因组DNA。因此，第1个、第2个、第3个、第4个、第5个、第6个、第7个和/或任何后续插入多聚核苷酸可包含这些序列。

[0157] 在一个实施方案中，如上所述，所关注的多聚核苷酸范围可能是约500个核苷酸至约200kb。所关注的多聚核苷酸可能是约500个核苷酸至约5kb、约5kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约60kb至约70kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、或约190kb至约200kb。

[0158] 在插入多聚核苷酸内和/或插入在靶标基因组基因座的所关注多聚核苷酸可编码多肽，可编码RNA，可编码miRNA，或者它可包含所关注的任何调控区或非编码区域，包括例如调控序列、启动子序列、增强子序列、转录阻遏子结合序列、Kozak共有区段、起始密码子，或缺失非蛋白质编码序列，但不包含蛋白质编码序列的缺失。另外，在插入多聚核苷酸内和/或插入在靶标基因组基因座的所关注多聚核苷酸可编码在神经系统、骨骼系统、消化系统、循环系统、肌肉系统、呼吸系统、心血管系统、淋巴系统、内分泌系统、泌尿系统、生殖系统或其组合中表达的蛋白质。

[0159] 在一个实施方案中，所述插入核酸包含内源基因的至少一个外显子的敲入等位基因。在一个实施方案中，所述插入核酸包含整个内源基因的敲入等位基因(即，“基因交换敲入”)。

[0160] 在一个实施方案中，所述插入核酸包含调控元件，包括例如启动子、增强子或转录阻遏子结合元件。

[0161] 在其他实施方案中，所述插入核酸包含条件性等位基因。在一个实施方案中，所述

条件性等位基因是多功能等位基因，如以全文引用的方式并入的US 2011/0104799中所述。在特定实施方案中，所述条件性等位基因包含：(a)关于靶基因的转录在有义方向上的致动序列，和在有义或反义方向上的药物选择盒；(b)在反义方向上所关注的核苷酸序列(NSI)和通过倒置模块具条件性(COIN，其利用分裂外显子的内含子和可逆的基因捕捉样模块；参见例如US 2011/0104799，其以全文引用的方式并入本文中)；和(c)在暴露于第一重组酶后重组以形成条件性等位基因的可重组单元，其(i)缺乏致动序列和DSC，和(ii)含有有义方向上的NSI和反义方向上的COIN。

[0162] 在一个实施方案中，所述插入核酸包含在编码序列中的遗传修饰。在一个实施方案中，所述遗传修饰包含编码序列的缺失突变。在一个实施方案中，所述遗传修饰包含两个内源编码序列的融合。

[0163] 在一个实施方案中，所述遗传修饰包含非蛋白质编码序列的缺失，但不包含蛋白质编码序列的缺失。在一个实施方案中，所述非蛋白质编码序列的缺失包含lncRNA基因座或其部分的缺失。在一个实施方案中，所述非蛋白质编码序列的缺失包含调控元件的缺失。在一个实施方案中，所述遗传修饰包含启动子的缺失。在一个实施方案中，所述遗传修饰包含启动子或调控元件的添加。在一个实施方案中，所述遗传修饰包含启动子或调控元件的置换。

[0164] 在一个实施方案中，打靶载体的核酸序列可包含当整合至基因组中时将产生哺乳动物、非人类动物或非人类哺乳动物lncRNA基因座的区域的遗传修饰的多聚核苷酸，其中在所述lncRNA基因座的遗传修饰导致lncRNA的功能缺失。在一个实施方案中，产生lncRNA敲除(“无效等位基因”)。在另一个实施方案中，产生lncRNA基因座的破坏。

[0165] 在其他实施方案中，所述插入核酸导致哺乳动物、非人类动物或非人类哺乳动物lncRNA基因座的一部分被插入核酸序列置换。在一个实施方案中，所述插入核酸序列是报道基因核酸序列。

[0166] 给定的插入多聚核苷酸和被置换的哺乳动物、非人类或非人类哺乳动物基因座的相应区域可以是非编码区、编码区、内含子、外显子、非翻译区、调控区、启动子或增强子或其任何组合。此外，给定的插入多聚核苷酸和/或所缺失的哺乳动物、非人类或非人类哺乳动物基因座的区域可以具有任何所希望的长度，包括例如长度为10-100个核苷酸、长度为100-500个核苷酸、长度为500-1kb个核苷酸、长度为1kb至1.5kb个核苷酸、长度为1.5kb至2kb个核苷酸、长度为2kb至2.5kb个核苷酸、长度为2.5kb至3kb个核苷酸、长度为3kb至5kb个核苷酸、长度为5kb至8kb个核苷酸、长度为8kb至10kb个核苷酸或更多个核苷酸。在其他情况下，插入或置换的大小是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb。在其他实施方案中，给定的插入多聚核苷酸和/或所缺失的哺乳动物、人类细胞或非人类哺乳动物基因座的区域是至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb、17kb、18kb、19kb、20kb、25kb、30kb、35kb、40kb、45kb、50kb或更大。

[0167] 在一个实施方案中，将插入核酸插入所关注的lncRNA基因座中以使得它可操作地连接至内源lncRNA启动子。在这些情况下，lncRNA启动子驱动插入核酸序列的表达。在一个实施方案中，所述插入核酸序列是报道基因核酸序列。

[0168] 在一些情况下，所述插入核酸包含启动子。在一个实施方案中，所述插入核酸包含

可操作地连接至启动子的所关注的多聚核苷酸，所述启动子驱动所关注的多聚核苷酸的表达。在一个实施方案中，所关注的多聚核苷酸包含报道基因核酸序列。在另一个实施方案中，所关注的多聚核苷酸包含选择标记核酸序列。

[0169] 在一个实施方案中，所述启动子是组成型活性启动子。

[0170] 在一个实施方案中，所述启动子是诱导型启动子。在一个实施方案中，所述诱导型启动子是化学调控启动子。在一个实施方案中，所述化学调控启动子是醇调控启动子。在一个实施方案中，所述醇调控启动子是醇脱氢酶(alcA)基因启动子。在一个实施方案中，所述化学调控启动子是四环素调控启动子。在一个实施方案中，所述四环素调控启动子是四环素响应型启动子。在一个实施方案中，所述四环素调控启动子是四环素操作子序列(tetO)。在一个实施方案中，所述四环素调控启动子是tet-On启动子。在一个实施方案中，所述四环素调控启动子是tet-Off启动子。在一个实施方案中，所述化学调控启动子是类固醇调控启动子。在一个实施方案中，所述类固醇调控启动子是大鼠糖皮质激素受体的启动子。在一个实施方案中，所述类固醇调控启动子是雌激素受体的启动子。在一个实施方案中，所述类固醇调控启动子是蜕皮激素受体的启动子。在一个实施方案中，所述化学调控启动子是金属调控启动子。在一个实施方案中，所述金属调控启动子是金属蛋白启动子。在一个实施方案中，所述诱导型启动子是物理调控启动子。在一个实施方案中，所述物理调控启动子是温度调控启动子。在一个实施方案中，所述温度调控启动子是热休克启动子。在一个实施方案中，所述物理调控启动子是光调控启动子。在一个实施方案中，所述光调控启动子是光诱导型启动子。在一个实施方案中，所述光诱导型启动子是光阻遏启动子。

[0171] 在一个实施方案中，所述启动子是组织特异性启动子。在一个实施方案中，所述启动子是神经元特异性启动子。在一个实施方案中，所述启动子是神经胶质特异性启动子。在一个实施方案中，所述启动子是肌肉细胞特异性启动子。在一个实施方案中，所述启动子是心脏细胞特异性启动子。在一个实施方案中，所述启动子是肾细胞特异性启动子。在一个实施方案中，所述启动子是骨骼细胞特异性启动子。在一个实施方案中，所述启动子是内皮细胞特异性启动子。在一个实施方案中，所述启动子是免疫细胞特异性启动子。在一个实施方案中，所述免疫细胞启动子是B细胞启动子。在一个实施方案中，所述免疫细胞启动子是T细胞启动子。

[0172] 在一个实施方案中，所述启动子是发育调控启动子。在一个实施方案中，所述发育调控启动子仅在胚胎发育阶段具活性。在一个实施方案中，所述发育调控启动子仅在成体细胞中具活性。

[0173] 在特定实施方案中，可基于细胞类型选择启动子。因此各种启动子可用于真核细胞、哺乳动物细胞、非人类哺乳动物细胞、多能性细胞、非人类多能性细胞、人类多能性细胞、人类ES细胞、人类成体干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、纤维母细胞或CHO细胞中。

[0174] 在一些实施方案中，所述插入核酸包含侧接位点特异性重组靶序列的核酸。公认为，虽然整个插入核酸可以侧接这些位点特异性重组靶序列，但是插入核酸内的任何区域或所关注的单独的多聚核苷酸也可以侧接这些位点。可通过任何方式将位点特异性重组酶引入细胞中，包括通过将重组酶多肽引入细胞中或通过将编码位点特异性重组酶的多聚核苷酸引入宿主细胞中。编码位点特异性重组酶的多聚核苷酸可位于插入核酸内或单独的多

聚核苷酸内。所述位点特异性重组酶能够可操作地连接至在细胞中具活性的启动子，包括例如诱导型启动子、细胞内源性启动子、细胞异源性启动子、细胞特异性启动子、组织特异性启动子或发育阶段特异性启动子。可以侧接插入核酸或所述插入核酸中所关注的任何多聚核苷酸的位点特异性重组靶序列可包括(但不限于)loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attp、att、FRT、rox或其组合。

[0175] 在一些实施方案中，位点特异性重组位点侧接编码包含在插入核酸内的选择标记和/或报道基因的多聚核苷酸。在这些情况下，在靶向基因座整合插入核酸后，可以除去位点特异性重组位点之间的序列。

[0176] 在一个实施方案中，所述插入核酸包含编码选择标记的多聚核苷酸。所述选择标记可以包含在选择盒中。这些选择标记包括(但不限于)新霉素磷酸转移酶(neo^r)、潮霉素B磷酸转移酶(hyg^r)、嘌呤霉素-N-乙酰转移酶($puro^r$)、杀稻瘟菌素S脱氨酶(bsr^r)、黄嘌呤/鸟嘌呤磷酸核糖转移酶(gpt)或单纯疱疹病毒胸苷激酶(HSV-k)或其组合。在一个实施方案中，编码选择标记的多聚核苷酸可操作地连接至在细胞中具活性的启动子。在一个实施方案中，编码选择标记的多聚核苷酸侧接位点特异性重组靶序列。

[0177] 所述插入核酸还可包含可操作地连接至启动子的报道基因，其中所述报道基因编码选自以下或包含以下的报道基因蛋白：β-半乳糖苷酶(由lacZ基因编码)、GFP、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强型黄色荧光蛋白(EYFP)、Emerald、增强型绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶和/或其组合。这些报道基因能够可操作地连接至在细胞中具活性的启动子。这些启动子可以是诱导型启动子、报道基因或细胞的内源性启动子、报道基因或细胞的异源性启动子、细胞特异性启动子、组织特异性启动子或发育阶段特异性启动子。

[0178] 在一个实施方案中，所述遗传修饰包含非蛋白质编码序列的缺失，但不包含蛋白质编码序列的缺失。在一个实施方案中，所述非蛋白质编码序列的缺失包含lncRNA基因座或其部分的缺失。在一个实施方案中，所述非蛋白质编码序列的缺失包含调控元件的缺失。在一个实施方案中，所述遗传修饰包含调控元件的缺失。在一个实施方案中，所述遗传修饰包含添加启动子或调控元件。在一个实施方案中，所述遗传修饰包含置换启动子或调控元件。

[0179] ii.表达盒

[0180] 本文提供多聚核苷酸或核酸分子，其包含本文提供用于靶向lncRNA基因座的靶向基因组整合系统中使用的各种组分(即核酸酶试剂、识别位点、插入核酸、所关注的多聚核苷酸、报道基因序列、打靶载体、选择标记和其他组分中的任一种或任何组合)。

[0181] 术语“多聚核苷酸”、“多聚核苷酸序列”、“核酸序列”和“核酸片段”在本文可互换使用。这些术语涵盖核苷酸序列等。多聚核苷酸可为单链或双链的RNA或DNA的聚合物，其任选地含有合成、非天然或改变的核苷酸碱基。呈DNA聚合物形式的多聚核苷酸可包含一个或多个cDNA区段、基因组DNA、合成DNA或其混合物。多聚核苷酸可包含脱氧核糖核苷酸并且核糖核苷酸包括天然存在的分子和合成类似物，和其任何组合。本文提供的多聚核苷酸也涵盖所有形式的序列，包括(但不限于)单链形式、双链形式、发夹、茎环结构等。

[0182] 此外提供包含用于靶向lncRNA基因座的靶向基因组整合系统的各种组分的重组

多聚核苷酸。术语“重组多聚核苷酸”和“重组DNA构建体”在本文可互换使用。重组构建体包含核酸序列的人工或异源组合，例如，在自然中未发现在一起的调控和编码序列。在其他实施方案中，重组构建体可包含源自不同来源的调控序列和编码序列，或源自相同来源但以不同于自然界中所存在的方式布置的调控序列和编码序列。这种构建体可自身使用或者可与载体结合使用。如果使用载体，则载体的选择取决于如本领域技术人员众所周知的用于转化宿主细胞的方法。例如，可使用质粒载体。此外提供成功地转化、选择和增殖包含本文提供的任何分离核酸片段的宿主细胞所需的遗传元件。可通过DNA的Southern分析、mRNA表达的Northern分析、蛋白质表达的免疫印迹分析或表型分析等来实现筛选。

[0183] 在特定实施方案中，本文所述的用于靶向lncRNA基因座的靶向基因组整合系统的一种或多种组分可以提供在用于在原核细胞、真核细胞、细菌、酵母细胞或哺乳动物细胞或所关注的其他生物体或细胞类型中表达的表达盒中。所述表达盒可包括可操作地连接至本文提供的多聚核苷酸的5’和3’调控序列。“可操作地连接”包含其中可操作地连接的组分以其预期方式起作用的关系。例如，所关注的多聚核苷酸与调控序列(即启动子)之间的可操作连接是允许表达所关注的多聚核苷酸的功能连接。可操作地连接的元件可为连续的或非连续的。当用于提及两个蛋白质编码区的接合时，可操作地连接是指所述编码区在同一阅读框中。在另一种情况下，编码蛋白质的核酸序列可能可操作地连接至调控序列(例如，启动子、增强子、沉默子序列等)以保持适当的转录调控。所述表达盒可另外含有至少一个另外的所关注的多聚核苷酸以共同引入生物体中。可选地，另外的所关注的多聚核苷酸可以提供在多个表达盒上。这种表达盒具有多个限制位点和/或重组位点以在调控区的转录调控下插入重组多聚核苷酸。所述表达盒可另外含有选择标记基因。

[0184] 所述表达盒可在5’-3’方向的转录中包括转录和翻译起始区(即，启动子)、本文提供的重组多聚核苷酸，和在所关注的哺乳动物细胞或宿主细胞中具功能性的转录和翻译终止区(即终止区)。本文提供的调控区(即，启动子、转录调控区、Kozak序列和翻译终止区)和/或多聚核苷酸可为宿主细胞天然/类似的或彼此天然/类似的。可选地，本文提供的调控区和/或多聚核苷酸可为宿主细胞异源的或彼此异源的。例如，可操作地连接至异源多聚核苷酸的启动子是来自与产生所述多聚核苷酸的物种不同的物种，或者如果来自相同/类似物种，则一种或两种是基本上从其原始形式和/或基因组基因座修饰的，或所述启动子不是可操作地连接的多聚核苷酸的天然启动子。可选地，本文提供的调控区和/或重组多聚核苷酸可为完全合成的。

[0185] 终止区可能是转录起始区天然的，可能是可操作地连接的重组多聚核苷酸天然的，可能是宿主细胞天然的，或者可源自启动子、重组多聚核苷酸、宿主细胞或其任何组合的另一个来源(即外来或异源的)。

[0186] 在制备表达盒时，各种DNA片段可能是被操纵的，以提供适当方向的DNA序列。为此目的，可使用适配子或连接子来接合DNA片段或者可涉及其他操纵以提供适宜的限制位点、除去多余的DNA、除去限制位点等。为此目的，可涉及体外诱变、引物修复、限制、退火、重取代，例如转换和颠换。

[0187] 多种启动子可用于本文提供的表达盒中。所述启动子可基于所希望的结果进行选择。公认为可通过在表达盒中使用不同的启动子以调节所关注多聚核苷酸的表达的时机、位置和/或水平来增强不同的应用。如果需要的话，这些表达构建体也可含有启动子调控区

(例如,赋予诱导型、组成型、环境调控或发育调控、或细胞或组织特异性/选择性表达者)、转录起始位点、Kozak共有序列、核糖体结合位点、RNA加工信号、转录终止位点和/或多聚腺苷酸化信号。

[0188] 含有本文提供的多聚核苷酸的表达盒也可包含用于选择转化细胞的选择标记基因。将可选择标记基因用于选择转化细胞或组织。

[0189] 适当时,在所述方法和组合物中使用的序列(即,所关注的多聚核苷酸、核酸酶试剂等)可被优化以增加在细胞中的表达。也就是说,可使用在给定的所关注细胞中优选的密码子(包括例如哺乳动物优选密码子、人类优选密码子、啮齿动物优选密码子、小鼠优选密码子、大鼠优选密码子、仓鼠优选密码子等)来合成基因以改进表达。

[0190] 本文提供的各种方法和组合物可使用选择标记。各种选择标记可用于本文公开的方法和组合物中。这些选择标记可以例如赋予对抗生素如G418、潮霉素、杀稻瘟菌素、新霉素或嘌呤霉素的抗性。这些选择标记包括新霉素磷酸转移酶(neo^r)、潮霉素B磷酸转移酶(hyg^r)、嘌呤霉素-N-乙酰转移酶($puro^r$)和杀稻瘟菌素S脱氨酶(bsr^r)。在其他实施方案中,所述选择标记可操作地连接至诱导型启动子并且选择标记的表达对细胞有毒性。这些选择标记的非限制性实例包括黄嘌呤/鸟嘌呤磷酸核糖转移酶(gpt)、卤次黄嘌呤-鸟嘌呤磷酸核糖转移酶(HGPRT)或单纯疱疹病毒胸苷激酶(HSV-TK)。编码选择标记的多聚核苷酸可操作地连接至在细胞中具活性的启动子。

[0191] iii.打靶载体

[0192] 使用打靶载体来将插入核酸引入真核生物、非人类、哺乳动物、非人类哺乳动物、人类、啮齿动物、小鼠、大鼠或仓鼠核酸的所关注lncRNA基因座中。所述打靶载体包含插入核酸并且还包含侧接所述插入核酸的5'和3'同源臂。所述侧接插入核酸的同源臂对应于真核生物、非人类、哺乳动物、非人类哺乳动物、人类、啮齿动物、小鼠、大鼠或仓鼠核酸的靶标lncRNA基因座内的区域。为便于提及,靶向基因组基因座内的相应同源基因组区域被称为"靶位点"。例如,打靶载体可包含侧接与第一和第二靶位点互补的第一和第二同源臂的第一插入核酸。因而,打靶载体从而有助于通过在细胞基因组内的同源臂与互补靶位点之间发生的同源重组事件将插入核酸整合至靶基因座核酸中。

[0193] 在一个实施方案中,真核生物、哺乳动物、非人类哺乳动物、人类、啮齿动物、小鼠或仓鼠核酸的靶基因座包含与5'同源臂互补的第一核酸序列和与3'同源臂互补的第二核酸序列。在一个实施方案中,所述第一核酸序列与所述第二核酸序列间隔至少5kb。在另一个实施方案中,所述第一核酸序列与所述第二核酸序列间隔至少1kb但小于50kb。在一个实施方案中,所述第一核酸序列与所述第二核酸序列间隔至少2kb。在一个实施方案中,所述第一核酸序列与所述第二核酸序列间隔至少3kb、至少4kb、至少5kb、至少6kb、至少7b、至少8kb、至少9kb、至少10kb、至少15kb、至少20kb、至少30kb、至少40kb或至少50kb。在其他实施方案中,所述第一核酸序列与所述第二核酸序列间隔至少1kb但小于2kb、至少2kb但小于3kb、至少4kb但小于5kb、至少5kb但小于6kb、至少6kb但小于7kb、至少7kb但小于8kb、至少约8kb但小于9kb、至少9kb但小于10kb、或至少10kb但小于15kb、至少约15kb但小于约20kb、至少约20kb但小于约30kb、或至少约40kb但小于约50kb。

[0194] 打靶载体的同源臂可具有足以促进与相应靶位点的同源重组事件的任何长度,包括例如长度为至少5-10kb、5-15kb、10-20kb、20-30kb、30-40kb、40-50kb、50-60kb、60-

70kb、70-80kb、80-90kb、90-100kb、100-110kb、110-120kb、120-130kb、130-140kb、140-150kb、150-160kb、160-170kb、170-180kb、180-190kb、190-200kb或更大。如下文更详细概述，大打靶载体可使用更大长度的打靶臂。在一个特定实施方案中，5'同源臂和3'同源臂的总和为至少10kb或者5'同源臂和3'同源臂的总和为至少约16kb至约100kb或约30kb至约100kb。在其他实施方案中，LTVEC的5'和3'同源臂的合计尺寸总和为约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、或约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中，缺失的尺寸与LTVEC的5'和3'同源臂的尺寸总和相同或类似。

[0195] 当两个区域彼此共有充分水平的序列同一性以充当同源重组反应底物时，同源臂和靶位点(即同源基因组区域)“互补”或者是彼此互补的。“同源性”是指相同或共有与相应或“互补”序列的序列同一性的DNA序列。给定靶位点与在打靶载体上存在的相应同源臂之间的序列同一性可以是允许发生同源重组的任何程度的序列同一性。例如，打靶载体(或其片段)的同源臂和靶位点(或其片段)共有的序列同一性量可以是至少50％、55％、60％、65％、70％、75％、80％、81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性，以使得所述序列进行同源重组。此外，同源臂与互补靶位点之间的互补同源区可具有足以促进在裂解识别位点处同源重组的任何长度。例如，给定的同源臂和/或互补靶位点可包含长度为至少5-10kb、5-15kb、10-20kb、20-30kb、30-40kb、40-50kb、50-60kb、60-70kb、70-80kb、80-90kb、90-100kb、100-110kb、110-120kb、120-130kb、130-140kb、140-150kb、150-160kb、160-170kb、170-180kb、180-190kb、190-200kb、200kb至300kb或更大的互补同源区(例如本文其他地方所述的LTVEC载体中所述)以使得所述同源臂具有充分同源性以与细胞基因组内的相应靶位点进行同源重组。为便于提及，同源臂在本文被称为5'和3'同源臂。这种术语是指同源臂相对于打靶载体内的插入核酸的位置。

[0196] 打靶载体的同源臂因此被设计成与具有靶向基因座的靶位点互补。因此，所述同源臂可以与细胞天然的基因座互补，或者可选地它们可以与整合至细胞基因组中的异源或外源DNA区段的区域(包括但不限于转基因、表达盒，或基因组DNA的异源或外源区域)互补。可选地，打靶载体的同源臂可以与人类人工染色体的区域或适当宿主细胞中所含的任何其他工程改造基因组区域互补。此外，打靶载体的同源臂可以与BAC文库、粘性质粒文库或P1噬菌体文库的区域互补或源自BAC文库、粘性质粒文库或P1噬菌体文库的区域。因此，在特定实施方案中，打靶载体的同源臂与给定细胞的天然、异源或外源的真核生物、非人类、哺乳动物、非人类哺乳动物、人类、啮齿动物、小鼠或大鼠基因组基因座互补。在一个实施方案中，所述同源臂源自合成DNA。

[0197] 所述打靶载体(例如大打靶载体)也可包含如本文其他地方所讨论的选择盒或报道基因。所述选择盒可包含编码选择标记的核酸序列，其中所述核酸序列可操作地连接至如本文其他地方所讨论的启动子。所述打靶载体的选择标记和/或报道基因可以侧接5'和3'同源臂或发现其为同源臂的5'或3'。

[0198] 在一个实施方案中,打靶载体包含包括编码报道基因的第一核苷酸序列的插入核酸。在一些情况下,在与所关注的lncRNA基因座同源重组后,编码报道基因的第一核苷酸序列可操作地连接至驱动lncRNA在lncRNA基因座的表达的内源启动子。在另一个实施方案中,所述打靶载体的插入核酸序列包含Kozak共有序列。在插入核酸包含报道基因的这些情况下,Kozak共有序列能够可操作地连接至编码报道基因的核酸序列。

[0199] 在另一个实施方案中,所述打靶载体的插入核酸包含编码可选择标记的第二核苷酸序列。在一些情况下,所述第二核酸可操作地连接至启动子。

[0200] 在一个实施方案中,所述插入核酸的第一核苷酸序列和/或第二核苷酸序列包含Kozak共有序列。

[0201] 在一个实施方案中,所述打靶载体(例如大打靶载体)包含可操作地连接至如本文其他地方所述的启动子的报道基因和/或可选择标记基因。这些报道基因和/或可选择标记基因能够可操作地连接至在本文其他地方所述的细胞中具活性的启动子。

[0202] 在一个实施方案中,所述打靶载体包含位点特异性重组酶基因。在一个实施方案中,所述位点特异性重组酶基因编码Cre重组酶。在一个实施方案中,所述Cre重组酶基因是Crei,其中两个编码Cre重组酶的外显子被内含子隔开以防止其在原核细胞中表达。在一个实施方案中,所述位点特异性重组酶基因编码Dre重组酶。

[0203] 在一个实施方案中,所述Cre重组酶基因还包含核定位信号以促进Cre(或任何重组酶或核酸酶试剂)定位至核(例如,所述基因是NL-Cre基因)。在一个特定实施方案中,所述Cre重组酶基因还包含核定位信号和内含子(例如,NL-Crei)。

[0204] 在各种实施方案中,用于表达上文讨论的Cre或Crei重组酶的合适启动子选自或包含Prm1、Blimp1、Gata6、Gata4、Igf2、Lhx2、Lhx5和/或Pax3。在一个特定实施方案中,所述启动子是Gata6或Gata4启动子。各种启动子可来自任何生物体,包括例如啮齿动物如小鼠或大鼠、真核生物、非人类哺乳动物、哺乳动物、人类或仓鼠。在另一个特定实施方案中,所述启动子是Prm1启动子。在另一个特定实施方案中,所述启动子是小鼠Prm1启动子。在另一个特定实施方案中,所述启动子是Blimp1启动子或其片段,例如,Blimp1启动子的1kb或2kb片段。参见例如美国专利8,697,851和美国申请公开2013-0312129,其两者都以全文引用的方式并入本文中。

[0205] 在一个实施方案中,所述插入核酸包含侧接两个位点特异性重组位点的核苷酸序列。位点特异性重组位点的实例包括(但不限于)loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attp、att、FRT、rox和其组合。

[0206] iv.大打靶载体

[0207] 术语“大打靶载体”或“LTVEC”包括包含同源臂的大打靶载体,所述同源臂对应于和源自比预期在细胞中进行同源打靶的其他方法通常使用的那些更大的核酸序列和/或包含包括比预期在细胞中进行同源重组打靶的其他方法通常使用的那些更大的核酸序列的插入多聚核苷酸。在特定实施方案中,LTVEC的同源臂和/或插入多聚核苷酸包含真核细胞的基因组序列。LTVEC的尺寸过大以致于不能通过常规测定法(例如southern印迹法和长程(例如1kb-5kb)PCR)筛选打靶事件。LTVEC的实例包括(但不限于)源自细菌人工染色体(BAC)、人类人工染色体或酵母人工染色体(YAC)的载体。LTVEC的非限制性实例和其产生方法描述于例如美国专利No.6,586,251、6,596,541、7,105,348和WO 2002/036789(PCT/

US01/45375)中,所述专利各自以引用的方式并入本文中。

[0208] 所述LTVEC可具有任何长度,包括(但不限于)至少约10kb、约15kb、约20kb、约30kb、约40kb、约50kb、约60kb、约70kb、约80kb、约90kb、约100kb、约150kb、约200kb、约10kb至约15kb、约15kb至约20kb、约20kb至约30kb、约30kb至约50kb、约50kb至约300kb、约50kb至约75kb、约75kb至约100kb、约100kb to 125kb、约125kb至约150kb、约150kb至约175kb、约175kb至约200kb、约200kb至约225kb、约225kb至约250kb、约250kb至约275kb或约275kb至约300kb。

[0209] 在一个实施方案中,LTVEC的同源臂源自BAC文库、粘性质粒文库或P1噬菌体文库。在其他实施方案中,所述同源臂源自细胞的靶向lncRNA基因组基因座并且在一些情况下被设计以靶向LTVEC的靶基因组基因座不可使用常规方法靶向。在其他实施方案中,所述同源臂是源自合成DNA。

[0210] 在一个实施方案中,LTVEC中的上游同源臂和下游同源臂的总和为至少10kb。在其他实施方案中,所述上游同源臂范围为约5kb至约100kb。在一个实施方案中,所述下游同源臂范围为约5kb至约100kb。在其他实施方案中,所述上游和下游同源臂的总和为约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约70kb至约80kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约110kb至约120kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、或约190kb至约200kb。在一个实施方案中,所述缺失的尺寸与LTVEC的5'和3'同源臂的总和的尺寸相同或类似。

[0211] 在一个实施方案中,所述LTVEC包含如本文其他地方所讨论的选择盒或报道基因。

[0212] III.引入序列的方法和转基因动物的产生

[0213] 如上文所概述,本文提供方法和组合物以允许一个或多个lncRNA基因座的靶向遗传修饰。此外公认,可进行另外的靶向遗传修饰。允许这些靶向遗传修饰的这些系统可以使用多种组分,并且为便于提及,本文术语"靶向基因组整合系统"通常包括整合事件所需的所有组分(即各种核酸酶试剂、识别位点、插入DNA多聚核苷酸、打靶载体、靶基因组基因座和所关注的多聚核苷酸)。

[0214] 本文提供的方法包括在细胞中引入一个或多个包含靶向基因组整合系统的各种组分的多聚核苷酸或多肽构建体。"引入"是指以使得序列接近细胞内部的方式向细胞呈递序列(多肽或多聚核苷酸)。本文提供的方法不依赖于用于向细胞中引入靶向基因组整合系统的任何组分的特定方法,仅在于多聚核苷酸接近至少一个细胞的内部。用于向各种细胞类型中引入多聚核苷酸的方法是本领域中已知的并且包括(但不限于)稳定的转染方法、瞬时转染方法和病毒介导的方法。

[0215] 在一些实施方案中,在所述方法和组合物中使用的细胞具有稳定并入其基因组中的DNA构建体。"稳定地并入"或"稳定地引入"是指在细胞中引入多聚核苷酸以使得核苷酸序列整合至细胞基因组中并且能够被其后代继承。任何方案可用于稳定并入DNA构建体或靶向基因组整合系统的各种组分。

[0216] 转染方案以及用于向细胞中引入多肽或多聚核苷酸序列的方案可改变。非限制性转染方法包括基于化学的转染方法,包括使用脂质体;纳米粒子;磷酸钙(Graham等,

(1973).病毒学(Virology)52(2):456-67;Bacchetti等,(1977)美国国家科学院院刊(Proc Natl Acad Sci USA)74(4):1590-4;和Kriegler,M(1991).转移和表达:实验室手册(Transfer and Expression:A Laboratory Manual).New York:W.H.Freeman and Company.第96-97页);树枝状大分子;或阳离子聚合物如DEAE-葡聚糖或聚乙烯亚胺。非化学方法包括电穿孔;声致穿孔;和光学转染。基于粒子的转染包括使用基因枪、磁体辅助转染(Bertram,J.(2006)当代药物生物技术(Current Pharmaceutical Biotechnology)7,277-28)。病毒方法也可用于转染。

[0217] 可使用本文公开的各种方法来产生非人类动物。这些方法包括:(1)使用本文公开的方法在非人类动物的多能性细胞的所关注靶lncRNA基因组基因座整合一个或多个所关注的多聚核苷酸以产生在靶向lncRNA基因组基因座中包含插入多聚核苷酸的遗传修饰的多能性细胞;(2)在靶lncRNA基因组基因座选择具有一个或多个所关注的多聚核苷酸的遗传修饰的多能性细胞;(3)在例如桑葚前期的非人类动物的宿主胚胎中引入所述遗传修饰的多能性细胞;和(4)在代孕母体中植入包含所述遗传修饰的多能性细胞的宿主胚胎以产生源自遗传修饰的多能性细胞的F0代。类似方法可用于靶向激发的靶标染色体基因座。所述非人类动物可以是非人类哺乳动物、啮齿动物、小鼠、大鼠、仓鼠、猴、农业哺乳动物或家养哺乳动物、或鱼类或鸟类。

[0218] 所述多能性细胞可能是人类ES细胞、非人类ES细胞、啮齿动物ES细胞、小鼠ES细胞、大鼠ES细胞、仓鼠ES细胞、猴ES细胞、农业哺乳动物ES细胞或家养哺乳动物ES细胞。在其他实施方案中,所述多能性细胞是非人类细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、人类多能性细胞、人类ES细胞、人类成体干细胞、发育受限的人类祖细胞、人类iPS细胞、啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞。在一个实施方案中,靶向遗传修饰导致lncRNA的功能缺失。

[0219] 也可使用核转移技术来产生非人类动物。简言之,用于核转移的方法包括以下步骤:(1)将卵母细胞去核;(2)分离供体细胞或细胞核以与所述去核卵母细胞组合;(3)将所述细胞或细胞核插入去核卵母细胞中以形成重构细胞;(4)将所述重构细胞植入动物子宫中以形成胚胎;和(5)使所述胚胎发育。在这些方法中,卵母细胞通常取自死亡动物,但它们也可从活体动物的输卵管和/或卵巢分离。卵母细胞可以在去核之前在本领域普通技术人员已知的多种介质中成熟。卵母细胞的去核可以本领域普通技术人员众所周知的多种方式进行。通常通过在融合之前在透明带下显微注射供体细胞而在去核卵母细胞中插入供体细胞或细胞核以形成重构细胞。可通过在接触/融合平面(电熔)上施加DC电脉冲,通过使细胞暴露于融合促进化学物质如聚乙二醇或借助于灭活病毒如仙台病毒(Sendai virus)来诱导融合。在核供体和受体卵母细胞的融合之前、期间和/或之后,通常通过电和/或非电方式来活化重构细胞。活化方法包括电脉冲、化学诱导的休克、精子渗透、增加卵母细胞中的二价阳离子水平,和降低细胞蛋白在卵母细胞中的磷酸化(如借助于激酶抑制剂)。通常将活化的重构细胞或胚胎在本领域普通技术人员众所周知的培养基中培养并且然后转移至动物的子宫。参见例如US20080092249、WO/1999/005266A2、US20040177390、WO/2008/017234A1和美国专利No.7,612,250,其各自以引用的方式并入本文中。

[0220] 提供用于产生在生殖系中包含一种或多种如本文所述的遗传修饰的非人类动物的其他方法,其包括:(a)使用本文所述的各种方法在原核细胞中修饰非人类动物的靶向基

因组lncRNA基因座；(b)选择在所述靶向基因组基因座包含遗传修饰的修饰原核细胞；(c)从所述修饰原核细胞的基因组分离遗传修饰的打靶载体；(d)在非人类动物的多能性细胞中引入所述遗传修饰的打靶载体以产生在靶向lncRNA基因组基因座包含插入核酸的遗传修饰的多能性细胞；(e)选择所述遗传修饰的多能性细胞；(f)在桑葚前期的非人类动物的宿主胚胎中引入所述遗传修饰的多能性细胞；和(g)在代孕母体中植入包含遗传修饰的多能性细胞的宿主胚胎以产生源自所述遗传修饰的多能性细胞的F0代。在这些方法中，所述打靶载体可包含大打靶载体。所述非人类动物可能是非人类哺乳动物、啮齿动物、小鼠、大鼠、仓鼠、猴、农业哺乳动物或家养哺乳动物。所述多能性细胞可能是人类ES细胞、非人类ES细胞、啮齿动物ES细胞、小鼠ES细胞、大鼠ES细胞、仓鼠ES细胞、猴ES细胞、农业哺乳动物ES细胞或家养哺乳动物ES细胞。在其他实施方案中，所述多能性细胞是非人类细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、人类多能性细胞、人类ES细胞、人类成体干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞。在一个实施方案中，所述靶向遗传修饰导致lncRNA的功能缺失。

[0221] 在其他方法中，分离步骤(c)还包括(c1)将遗传修饰的打靶载体(即，遗传修饰的LTVEC)线性化。在其他实施方案中，引入步骤(d)还包括(d1)向多能性细胞中引入核酸酶试剂以促进同源重组。在一个实施方案中，通过向原核细胞或多能性细胞施用如本文所述的可选择剂来进行选择步骤(b)和/或(e)。在一个实施方案中，经由如本文所述的等位基因(MOA)测定法的修改来进行选择步骤(b)和/或(e)。

[0222] 在一些实施方案中，可以使用VELOCIGENE®遗传工程技术使用源自细菌人工染色体(BAC)DNA的LTVEC在细菌细胞中通过一系列同源重组反应(BHR)来进行本文所述的靶基因组基因座的各种遗传修饰(参见例如美国专利No.6,586,251和Valenzuela,D.M.等,(2003),自然生物技术(Nature Biotechnology)21(6):652-659，其以全文引用的方式并入本文中)。

[0223] 在一些实施方案中，将如本文所述的包含各种遗传修饰的lncRNA靶向的多能性和/或全能性细胞用作插入供体细胞并且经由VELOCIMOUSE®法引入来自相应生物体、例如8细胞阶段小鼠胚胎的桑葚前期胚胎中(参见例如US 7,576,259、US 7,659,442、US 7,294,754和US2008-0078000 A1，其都以全文引用的方式并入本文中)。将包含遗传修饰的多能性和/或全能性细胞的非人类动物胚胎温育至胚泡期，然后植入代孕母体中以产生F0代。在一些实施方案中，将包含如本文所述的各种遗传修饰的靶向哺乳动物ES细胞引入胚泡期胚胎中。可以经由如本文所述的等位基因(MOA)测定法的修改来鉴定带有遗传修饰的基因组基因座(即lncRNA基因座)的非人类动物。将源自遗传修饰的多能性和/或全能性细胞的所得F0代非人类动物与野生型非人类动物杂交以获得F1代后代。在用特异性引物和/或探针进行基因分型后，将对于遗传修饰的基因组基因座杂合的F1非人类动物彼此杂交以产生对于遗传修饰的基因组基因座纯合的F2代非人类动物后代。

[0224] 在一个实施方案中，提供用于产生在至少一个lncRNA基因座中包含遗传修饰的非人类动物的方法。这种方法包括：(a)使多能性细胞与包含侧接5'和3'同源臂的插入核酸的打靶构建体接触；其中所述打靶构建体与所述细胞基因组中的lncRNA基因座进行同源重组以形成修饰的多能性细胞；(b)将所述修饰的多能性细胞引入宿主胚胎中；和(c)使所述宿主胚胎在代孕母体中受孕，其中所述代孕母体产生包含修饰lncRNA基因座的后代，其中所

述遗传修饰导致至少一个lncRNA的功能缺失。

[0225] IV.细胞

[0226] 本文所述的各种方法使用用于修饰细胞中的lncRNA基因座的基因组基因座打靶系统。这些细胞包括原核细胞如细菌细胞(包括大肠杆菌),或真核细胞如酵母、昆虫、两栖动物、植物,或哺乳动物细胞(包括但不限于小鼠细胞、大鼠细胞、仓鼠细胞、兔细胞、猪细胞、牛细胞、鹿细胞、绵羊细胞、山羊细胞、鸡细胞、猫细胞、狗细胞、雪貂细胞、灵长类动物(例如,狨猴、猕猴)细胞等),和来自家养哺乳动物的细胞或来自农业哺乳动物的细胞。一些细胞是非人类、特别是非人类哺乳动物细胞。在一些实施方案中,对于不容易获得合适的可遗传修饰的多能性细胞的那些哺乳动物,使用其他方法将体细胞重编程形成多能性细胞,例如,经由在体细胞中引入包括(但不限于)Oct3/4、Sox2、KLF4、Myc、Nanog、LIN28和Glis1的多能性诱导因子的组合。在这些方法中,所述细胞也可以是哺乳动物细胞、人类细胞、非人类哺乳动物细胞、非人类细胞、来自啮齿动物、大鼠、小鼠、仓鼠的细胞、纤维母细胞或任何其他宿主细胞。在其他实施方案中,所述细胞是多能性细胞、诱导型多能性干(iPS)细胞、非人类胚胎干(ES)细胞。这些细胞包括多能性细胞,包括例如诱导型多能性干(iPS)细胞、人类iPS细胞、小鼠胚胎干(ES)细胞、大鼠胚胎干(ES)细胞、人类胚胎(ES)细胞、或发育受限的人类祖细胞、啮齿动物胚胎干(ES)细胞、小鼠胚胎干(ES)细胞或大鼠胚胎干(ES)细胞。

[0227] 非限制性实施方案包括:

[0228] 1.一种非人类动物,其在其基因组中包含至少一个修饰长链非编码RNA(lncRNA)基因座,其中所述至少一个修饰lncRNA基因座包含编码lncRNA的核酸序列的功能缺失型突变。

[0229] 2.如实施方案1所述的非人类动物,其中所述lncRNA是长链基因间非编码RNA(lincRNA)。

[0230] 3.如实施方案1或2中的任一项所述的非人类动物,其中所述功能缺失型突变的特征在于至少一种lncRNA功能的破坏或敲除。

[0231] 4.如实施方案3所述的非人类动物,其中所述修饰lncRNA基因座包含编码所述lncRNA或其部分的一个或多个外显子的缺失。

[0232] 5.如实施方案4所述的非人类动物,其中所述破坏或敲除包括:(a)lncRNA基因座内起始于所述lncRNA基因座的第二外显子的一个或多个外显子的缺失;(b)lncRNA基因座内起始于所述lncRNA基因座的第一外显子的一个或多个外显子的缺失;或(c)lncRNA基因座的整个RNA编码区的缺失。

[0233] 6.如实施方案3所述的非人类动物,其中所述破坏或敲除包括用插入核酸置换lncRNA基因座或其部分。

[0234] 7.如实施方案6所述的遗传修饰的非人类动物,其中所述插入核酸包含编码报道基因的第一核苷酸序列。

[0235] 8.如实施方案7所述的遗传修饰的非人类动物,其中所述第一核苷酸序列可操作地连接至驱动所述报道基因的表达的启动子。

[0236] 9.如实施方案7所述的遗传修饰的非人类动物,其中编码所述报道基因的所述第一核苷酸序列位于与内源lncRNA启动子可操作地连接的lncRNA基因座中,其中所述内源lncRNA启动子驱动所述核苷酸序列的表达。

[0237] 10.如实施方案9所述的非人类动物，其中所述核酸序列的表达遵循所述lncRNA的表达模式。

[0238] 11.如实施方案7所述的遗传修饰的非人类动物，其中所述第一核苷酸序列包含Kozak共有序列。

[0239] 12.如实施方案6至11中的任一项所述的非人类动物，其中所述置换包括：(a)用所述插入核酸置换lncRNA基因座内起始于所述lncRNA基因座的第二外显子的一个或多个外显子；(b)用所述插入核酸置换lncRNA基因座内起始于所述lncRNA基因座的第一外显子的一个或多个外显子；或(c)用所述插入核酸置换lncRNA基因座的整个RNA编码区。

[0240] 13.如实施方案6至12中的任一项所述的非人类动物，其中所述报道基因是β-半乳糖苷酶、绿色荧光蛋白(GFP)、增强型绿色荧光蛋白(eGFP)、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强型黄色荧光蛋白(EYFP)、Emerald、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶中的任一种或其组合。

[0241] 14.如实施方案6至13中的任一项所述的非人类动物，其中所述插入核酸还包含编码可选择标记的第二核酸序列，其中所述第二核酸序列可操作地连接至启动子。

[0242] 15.如实施方案14所述的非人类动物，其中所述插入核酸包含侧接编码所述报道基因的区段和/或编码所述可选择标记的区段的位点特异性重组位点。

[0243] 16.如实施方案1至15中的任一项所述的非人类动物，其中所述lncRNA包含Pint、Celrr、Crnde、Eldr、Fendrr、Halr1、Hotair、Hottip、Hoxa11os、Pantr1、Pantr2、Ptgs2os2、lincenc1、Trp53cor1、lincppara、Mannr、Haglr、Peril、Kantr、Tug1或其组合。

[0244] 17.如实施方案1至15中的任一项所述的非人类动物，其中所述非人类动物的特征在于具有一种或多种以下表型：(a)过早衰老相关表型；(b)围产期致死；(c)肺发育缺陷；(d)尾部和后肢的形态畸形；(e)一个或多个组织中的肌肉质量损失；或(f)(a)-(e)的其任何组合。

[0245] 18.如实施方案1所述的非人类动物，其中所述lncRNA是Pint，并且所述非人类动物的特征在于过早衰老相关表型，其包含：(a)相比于野生型对照的较缓慢生长速率；(b)肌肉力量下降；(c)纤维化；(d)相比于野生型对照的较低体脂含量；(e)相比于野生型对照的较低股骨骨矿物质密度和骨质量；(f)相比于野生型对照降低的肌肉质量；(g)中值寿命降低；(h)脊椎弯曲；(i)器官萎缩；或(j)(a)-(i)的其任何组合。

[0246] 19.如实施方案1至15中的任一项所述的非人类动物，其中所述非人类动物展现大脑发育缺陷。

[0247] 20.如实施方案19所述的非人类动物，其中所述lncRNA是Pantr2、Kantr、Peril、Celrr、Pantr1、Crnde、lincenc1、Pint、lincppara或Tug1。

[0248] 21.如实施方案1至20中的任一项所述的遗传修饰的非人类动物，其中所述非人类动物是哺乳动物。

[0249] 22.如实施方案21所述的遗传修饰的非人类动物，其中所述哺乳动物是啮齿动物。

[0250] 23.如实施方案22所述的遗传修饰的非人类动物，其中所述哺乳动物是小鼠、大鼠或仓鼠。

[0251] 24.一种细胞、组织或胚胎，其源自如实施方案1至23中的任一项所述的非人类动

物。

[0252] 25.一种打靶载体,其包含侧接可以与所关注的lncRNA基因座进行同源重组的5'和3'同源臂的插入核酸。

[0253] 26.如实施方案25所述的打靶载体,其中所述插入核酸包含编码报道基因的第一核酸序列。

[0254] 27.如实施方案26所述的打靶载体,其中在与所关注的lncRNA基因座进行同源重组后,编码所述报道基因的所述第一核酸序列可操作地连接至驱动lncRNA在所述lncRNA基因座的表达的内源启动子。

[0255] 28.如实施方案26至27中的任一项所述的打靶载体,其中所述报道基因是β-半乳糖苷酶、绿色荧光蛋白(GFP)、增强型绿色荧光蛋白(eGFP)、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强型黄色荧光蛋白(EYFP)、Emerald、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶中的任一种或其组合。

[0256] 29.如实施方案25至28中的任一项所述的打靶载体构建体,其中所述插入核酸还包含编码可选择标记的第二核酸序列,其中所述第二核酸可操作地连接至启动子。

[0257] 30.如实施方案29所述的打靶载体,其还包含侧接编码所述报道基因的区段和/或编码所述可选择标记核酸的区段的位点特异性重组位点。

[0258] 31.如实施方案26或29所述的打靶载体,其中所述第一核酸序列和/或所述第二核酸序列还包含Kozak共有序列。

[0259] 32.如实施方案25至31中的任一项所述的打靶载体,其中所述插入核酸还包含驱动所述报道基因的表达的启动子。

[0260] 33.一种产生在至少一个lncRNA基因座中包含遗传修饰的非人类动物的方法,所述方法包括:(a)使多能性细胞与包含侧接5'和3'同源臂的插入核酸的打靶构建体接触;其中所述打靶构建体与所述细胞基因组中的所述lincRNA基因座进行同源重组以形成修饰的多能性细胞;(b)将所述修饰的多能性细胞引入宿主胚胎中;和(c)使所述宿主胚胎在代孕母体中受孕,其中所述代孕母体产生包含修饰的lncRNA基因座的后代,其中所述遗传修饰导致所述至少一个lncRNA的功能缺失。

[0261] 34.如实施方案33所述的方法,其中所述lncRNA是lincRNA。

[0262] 35.如实施方案33至34中的任一项所述的方法,其中所述遗传修饰包含至少一种lncRNA功能的破坏或敲除。

[0263] 36.如实施方案33至35中的任一项所述的方法,其中所述lncRNA包含Pint、Celrr、Crnde、Eldr、Fendrr、Halr1、Hotair、Hottip、Hoxa1os、Pantr1、Pantr2、Ptgs2os2、lincenc1、Trp53cor1、lincppara、Mannr、Haglr、Peril、Kantr、Tug1或其组合。

[0264] 37.一种用于修饰多能性细胞中的lncRNA基因座的方法,其包括(a)在所述多能性细胞中引入包含侧接可以与所述lncRNA基因座进行同源重组的5'和3'同源臂的插入核酸的打靶构建体;和(b)鉴定在所述lncRNA基因座包含靶向遗传修饰的修饰多能性细胞,其中所述遗传修饰导致lncRNA功能的功能缺失。

[0265] 38.如实施方案37所述的方法,其中所述多能性细胞是人类iPS细胞。

[0266] 39.如实施方案37所述的方法,其中所述多能性细胞是小鼠或大鼠胚胎干(ES)细

胞。

[0267] 40.如实施方案37至39中的任一项所述的方法，其中所述lncRNA包含Pint、Celrr、Crnde、Eldr、Fendrr、Halr1、Hotair、Hottip、Hoxa1os、Pantr1、Pantr2、Ptgs2os2、lincenc1、Trp53cor1、lincppara、Mannr、Haglr、Peril、Kantr、Tug1或其组合。

[0268] 41.一种遗传修饰的非人类动物，其中所述遗传修饰包含Pint的敲除。

[0269] 42.如实施方案41所述的遗传修饰的非人类动物，其为啮齿动物。

[0270] 43.如实施方案42所述的遗传修饰的啮齿动物，其中所述啮齿动物选自小鼠和大鼠。

[0271] 44.一种展现过早衰老相关表型的小鼠，其中所述小鼠包含Pint的敲除。

[0272] 45.如实施方案44所述的小鼠，其中所述小鼠展现选自以下的表型：体重缓慢增加，体重过早达到平台期，过早生长停滞，到12周龄时脊椎弯曲，到26周龄时严重脊椎弯曲，在六个月时脱毛，在6个月时后肢肌肉力量损失，或其组合。

[0273] 46.一种遗传修饰的非人类动物，其包含选自HOTAIR、HOTTIP、Hoxa1os、Pantr1、Pantr2、Ptgs2os2、Eldr、Lincenc1、Mannr、Fendrr、Halr1、Haglr、Celrr、Crnde、Kantr、Pint、Trp53cor1、lincppara、Haglr、Tug1和其组合的lincRNA的敲除。

[0274] 47.如实施方案46所述的遗传修饰的小鼠，其为啮齿动物。

[0275] 48.如实施方案47所述的遗传修饰的啮齿动物，其为小鼠或大鼠。

[0276] 实施例

[0277] 实施例1:打靶载体的构建

[0278] 使用如先前所述的VelociGene®法，其允许在小鼠中快速和高通量地产生定制基因突变(Valenzuela,D.M.等,(2003b),自然生物技术(Nat Biotechnol)21:652-659)。简言之，使用来自小鼠bMQ(129S7/SvEv Brd-Hprt b-m2)或RP23BAC文库(Adams,D.J.等,(2005),基因组学(Genomics)86:753-758)的BAC克隆产生BacVec大打靶载体(LTVEC)。lacZ/neo^{r}报道基因/选择盒(图1)与用于NIH KOMP(在www.velocigene.com/komp/detail/10020可获得的序列)的ZEN-Ub1盒相同，其中例外为lacZ部分中的β-半乳糖苷酶编码序列的氨基末端被修饰以包括ATG起始密码子和Kozak共有序列(Kozak,M.(1987)核酸研究(Nucleic acids research)15:8125-8148))。

[0279] 实施例2:ES细胞打靶

[0280] 利用多孔电穿孔装置(Harvard Apparatus,Boston,MA)在体积为0.125ml的电穿孔缓冲液(Millipore)(3.3×10^{6}个细胞、0.67μg DNA)中将LTVEC引入VGF1F1杂合(129S6SvEvTac/C57BL6NTac)ES细胞(Poueymirou等,(2007);Valenzuela等,(2003a))中，接着在15cm胶凝化板上培养。在电穿孔后48小时添加含G418的选择培养基并且其后每天更换。在电穿孔后10天挑取耐药性菌落，用胰蛋白酶处理并且在胶凝化96孔板中培养至少三天，然后提前DNA并纯化。通过等位基因缺失测定法鉴定正确靶向的ES细胞克隆(Frendewey等,(2010),酶学方法(Methods in enzymology)476:295-307;Valenzuela等,(2003a),自然生物技术(Nat.Biotechnol.)21:652-659)。

[0281] 实施例3:产生LincRNA小鼠

[0282] 使用VelociMouse®法(Dechiara,T.M.,(2009),分子生物学方法(Methods Mol Biol)530:311-324;Poueymirou等,(2007),自然生物技术(Nat.Biotechnol.)25:91-99)，

其中将靶向ES细胞注入未压实的8细胞期Swiss Webster胚胎中以产生带有lincRNA敲除突变的完全ES细胞源F0代小鼠。将雄性VelociMice®直接用于lacZ表达谱分析或与C57BL/6NTac雌性交配以产生胚胎或成体以用于lacZ分析或产生F1育种并且对N2F1小鼠进行表型研究。通过指定阴道栓鉴定早晨为第0.5天(E0.5)来进行定时交配。

[0283] 实施例4:LacZ表达谱分析

[0284] 对于全胚胎染色,收集E9.5和E14.5胚胎,在PBS中洗涤并且在新鲜的0.2%戊二醛溶液中温育15至60分钟。取得胚胎卵黄囊用于基因分型。在固定后,将胚胎在洗涤缓冲液中洗涤并在37℃下在X-gal(1mg/mL)染色溶液中温育1至24小时。在染色后,将组织用洗涤缓冲液冲洗,后固定在4%多聚甲醛中,并且在70%乙醇中温育至少24小时。立即对E9.5-e11.5胚胎拍照,同时在ddH_2O中含有递增甘油和递减1%KOH的一系列溶液中清洁e12.5胚胎和更老胚胎。用Nikon SMZ800立体显微镜获取照片。将来自Fendrr e13.5胚胎的肺解剖以在清洁后用于拍照。

[0285] 对于成年小鼠研究,将6至8周龄F0代完全ES细胞源VelociMice®深度麻醉并通过使用0.2%戊二醛/4%多聚甲醛溶液进行心脏灌注来固定。解剖大脑、胸腔、心脏、肺、肝脏、脾脏、胃、肾脏、肠、泌尿生殖系统、肌肉和后肢组织,用PBS冲洗并且在0.2%戊二醛/4%多聚甲醛溶液中后固定30分钟。然后将组织洗涤并在37℃下在X-gal(1mg/mL)染色溶液中温育1至24小时。在染色后,将组织用洗涤缓冲液冲洗,后固定在4%多聚甲醛中,在一系列50%、70%和100%甘油中清洁并且关于胚胎拍照。

[0286] 实施例5:动物护理和实验程序

[0287] N2F1小鼠的表型研究开始于6-8周龄。对于定时交配,我们指定阴道栓鉴定早晨为胚胎期0.5天(E0.5)。对于LincRNA KO和野生型同窝出生仔畜当将它们在69-74F和40-60%湿度下圈养在每天12小时光照下用于研究时,从出生起观测不同发育阶段目标(呼噜、呼吸、面部和四肢异常、皮肤颜色、姿势、扶正和眼张开)直至约6-8周龄。所有实验都开始于6-8周龄并且所有动物程序的进行都符合被Regeneron制药机构动物护理和使用委员会(Regeneron Pharmaceuticals Institutional Animal Care and Use Committee)批准的方案。

[0288] 实施例6:μCT分析

[0289] 使用Quantum FX显微CT临床前体内成像系统(Perkin Elmer)对3D骨骼成像进行可视化。使用氧气/异氟烷吸入来麻醉小鼠,其中异氟烷流速为2.5L/min并且氧气流速为1.5L/min。在扫描期间,通过鼻锥在0.25L/min氧气流速下维持麻醉。在90kV和160μA下进行扫描,对于后肢为30mm视场并且对于椎骨为60mm视场。对于骨矿物质密度、骨总量、瘦肉和脂肪量分析,对于除头部之外的整个身体在60mm视场下进行两次连续扫描。手动分离右股骨用于骨矿物质密度测量。右股骨、总瘦肉和总脂肪体积都是使用Analyze 11.0软件(Mayo Clinic)测量的并且基于所确定的密度转换成质量。在扫描后,使小鼠返回其笼中并且依照Regeneron IACUC方案监测复苏。

[0290] 实施例7:悬尾试验

[0291] 当通过尾巴悬吊时,小鼠通过伸展其后肢来准备安全着陆,这通常被称为“后肢张开”。小鼠通过其尾巴悬吊10秒并且观测任何异常的紧握表型。

[0292] 实施例8:握耐力试验

[0293] 通过小鼠从铁丝网格(铁丝粗细为约2mm)倒挂的能力来评估其在5、7和10周龄时的肌肉不足迹象。将小鼠分别放置在铁丝网格上,轻轻摇晃以促使它们在网格上下颠倒时抓住不放。记录小鼠松开所花费的时间(直至最多60秒)。给与小鼠三次尝试以尽可能长时间地抓住不放并且记录最大时间以进行统计学比较。

[0294] 实施例9:肌肉组织学和组织尸检

[0295] 通过CO_2吸入、接着颈椎脱位来处死小鼠。将胫骨前肌(TA)、四头肌和腓肠肌(GA)解剖并称重。将所有采集的肌肉和器官冷冻并保持在-80℃以供未来研究用。对于组织学,将肌肉冷冻在OCT中,以12μm厚度横着冷冻切片以暴露出头部、比目鱼肌和跖肌的外侧和内侧。将相邻切片用H&E、层粘连蛋白和MHC Slow Stain染色。使用Aperio扫描显微镜对染色进行数字成像。使用Spectra软件测定纤维尺寸和计数。所有数据以平均值+/-平均值的标准误差(以误差条表示)表示。使用程序STATVIEW和/或PRISM进行方差分析(ANOVA)。统计学显著性设定为P值小于0.05。对于皮肤组织学,将背侧和腹侧皮肤区域剃毛,解剖并固定在4%多聚甲醛(PFA)中至少24小时并转移至70%乙醇。由Histoserv Labs,Inc.,Germantown,MD进行石蜡包埋、切片以及在皮肤切片上用苏木精和曙红染色。

[0296] 实施例10:Kaplan-Meier存活曲线分析

[0297] 根据Regeneron IACUC方案对动物观测52周时间并且监测发病迹象。基于发病率准则,在52周时间点之前此项研究中的小鼠无需被处死。使用Graphpad PRISM 6软件测定存活曲线和对数秩检验。

[0298] 实施例11:在二十个具有切除lincRNA的小鼠系中的不同表型和特定转录模式

[0299] 在被设计以研究长链基因间非编码RNA(lincRNA)的生物功能的20个敲除小鼠系的研究中,我们已发现多种表型,范围从围产期致死至与过早衰老以及肺、骨骼和肌肉中的形态和功能异常相关的缺陷。每个突变等位基因带有lacZ报道基因,其表达谱突出了在胚胎和成体中的广泛多种时空和组织特异性转录模式,这使我们了解表型分析并将充当这些基因的未来研究的指导。我们的研究表明,lincRNA是一类新型编码分子,与蛋白质一样,其在哺乳动物中的胚胎发育、生理学和广泛多种组织和器官的自我平衡中提供基本和重要的功能作用。

[0300] 近来明显可见,对于哺乳动物的基因型和表型之间的关系的深入了解需要我们将研究扩展超出蛋白质编码基因以包括基因组的非编码部分(Mattick JS(2009)PLoS遗传学(PLoS genetics)5:e1000459)。在哺乳动物细胞中的大规模全基因组表达研究已揭示出,约四分之三的基因组能够以RNA形式表达(Kapranov P等,(2007)科学(Science)316:1484-1488;Carninci P等,(2005)科学(Science)309:1559-1563;Djebali S等,(2012)自然(Nature)489:101-108),并且大部分转录物不编码蛋白质。在非编码转录物中包括被称为长链非编码RNA(lncRNA)的多种类别。代表来自人类细胞中接近10,000个基因组基因座的约15,000种转录物(Derrien T等,(2012)基因组研究(Genome Research)22:1775-1789),lncRNA和被称为长链基因间非编码RNA(lincRNA)的子类(Guttman M等,(2009)自然(Nature)458:223-227;Khalil AM等,(2009)美国国家科学院院刊(Proceedings of the National Academy of Sciences of the United States of America)106:11667-11672)在结构、合成和其基因的染色质特性上与蛋白质编码mRNA类似。这种结构相似性是否延伸至与蛋白质匹配的功能多样性仍然是一个悬而未决的问题。

[0301] 由于第一敲除株的产生接近二十五年前，因此小鼠已成为研究哺乳动物基因功能的首要系统(Capecchi MR(2001)自然医学(Nat Med)7:1086-1090;Evans MJ(2001)自然医学(Nat Med)7:1081-1083;Smithies O(2001)自然医学(Nat Med)7:1083-1086)。除了少数例外，敲除小鼠技术在单独的基因研究以及大规模国际项目(www.knockoutmouse.org)中的应用专注于蛋白质编码基因，但是最近产生用于微RNA的全局敲除小鼠资源的努力(Prosser,H.M.等,(2011),自然生物技术(Nat Biotechnol)29:840-845)(mcmanuslab.ucsf.edu/microrna_knockout)表明将所述技术应用于非编码RNA的价值。已存在通过在小鼠中进行基因破坏而对单独的lncRNA的一些功能研究，但约一半专注于在单一相关生物现象中所涉及的充分研究的lncRNA:X染色体失活(Marahrens Y等,(1997)基因和发育(Genes&Development)11:156-166;Sado T等,(2001)发育(Development)128:1275-1286)和身体染色体压印(Leighton PA等,(1995)自然(Nature)375:34-39;Mohammad F等,(2010)发育(Development)137:2493-2499;Sleutels F等,(2002)自然(Nature)415:810-813;Takahashi N等,(2009)人类分子遗传学(Human Molecular Genetics)18:1879-1888)。

[0302] 近来，小鼠Fendrr lncRNA的破坏导致与心脏和体壁发育缺陷相关的胚胎致死(Grote P等,(2013)发育细胞(Developmental Cell)24:206-214)。然而，lncRNA编码Gt(ROSA)26Sor(Zambrowicz BP等,(1997)美国国家科学院院刊(Proceedings of the National Academy of Sciences of the United States of America)94:3789-3794)或Malat1(Zhang B等,(2012)细胞报道(Cell Reports)2:111-123)基因中的缺失或插入突变不产生可辨识的表型。关于lncRNA基因的结构、表达和功能的新兴理解提出了使用小鼠分子遗传学来揭示与这类新型基因相关的生物功能的新的机会。

[0303] 然而，将敲除小鼠技术应用于lncRNA不存在一些技术挑战。大部分蛋白质具有已知具有功能相关性或至少预测具有功能相关性的元件或结构域。使这些基本部分缺失编码序列通常足以产生无效等位基因。同样，可以设计条件性等位基因，其分离关键的一个或多个外显子以稍后通过组织特异性重组酶的作用缺失。因为仅对少数lncRNA建立了结构-功能关系并且不存在开放阅读框作为指导，所以可用于蛋白质编码基因的敲除策略不可用于编码lncRNA的基因组基因座。虽然lncRNA基因的注释已得到改善(Derrien T等,(2012)基因组研究(Genome Research)22:1775-1789)，但一些基因的精确边界可能仍然不明确，这可能使敲除等位基因设计变得复杂。应用于蛋白质编码基因敲除小鼠的强大工具是用报道基因例如β-半乳糖苷酶或荧光蛋白的编码序列置换靶基因，所述β-半乳糖苷酶或荧光蛋白的表达通过靶基因的启动子进行控制，从而报道其在小鼠中表达的空间和时间模式。报道基因置换已成功地应用于非编码RNA如充分研究的Gt(ROSA)26Sor基因座(Zambrowicz BP等,(1997)美国国家科学院院刊(Proceedings of the National Academy of Sciences of the United States of America)94:3789-3794)，其编码lncRNA，和短链非编码RNA miR-155的基因(Thai TH等,(2007)科学(Science)316:604-608)，但可能需要开发用于产生lncRNA的这些等位基因的规则。虽然存在这些限制，但由于鉴定了数千种lncRNA，探索将基因敲除小鼠技术力量应用于这类新型基因的时机已经成熟。怀着这个目标，此处描述了通过产生敲除小鼠系阐明二十种lincRNA的功能的统一遗传学方法，各自带有用β-半乳糖苷酶报道基因置换的基因切除的缺失等位基因。

[0304] 产生具有报道基因置换的20种lincRNA缺失小鼠系

[0305] 表1列出在这项研究中靶向的10种不同染色体上的20种lincRNA基因和所产生的26种敲除缺失等位基因。我们选择突变长链基因间非编码RNA类别的成员，因为根据定义，lincRNA基因是从邻近的蛋白质编码基因分离的并且其转录物不重叠(Guttman M等，(2009)自然(Nature)458:223-227)。这种特征允许我们设计缺失等位基因，其将使得与附近基因表达相互作用的可能性最小。我们选择靶向lincRNA基因以反映多种表达模式(Khalil AM等，(2009)美国国家科学院院刊(Proceedings of the National Academy of Sciences of the United States of America)106:11667-11672;Cabili MN等，(2011)基因和发育(Genes&Development)25:1915-1927)，重点是神经表达，并且其可能涉及在基因表达的开发和调控中。

[0306] 我们关于lincRNA敲除突变的设计策略将由两个目标指导。首先，我们旨在产生将精确报道lincRNA基因的转录活性的等位基因。虽然存在关于组织特异性lincRNA表达的基于细胞和所选组织解剖研究的证据(Cabili MN等，(2011)基因和发育(Genes&Development)25:1915-1927)，但是我们希望通过产生由lacZ表达谱分析提供的更高清晰度表达模式来补充这种知识库，其可以在空间上和时间上解析组织和器官表达，从而揭示子结构域和在一些情况下通过组织解剖实验没有解析的细胞类型特异性。其次，我们努力产生废除lincRNA的合成和功能的基因切除缺失，以使得与突变相关的任何表型将指示靶向RNA的关键功能。

[0307] 敲除缺失的尺寸范围为约400bp至50kb，其中一半缺失所有的注释外显子。对于大部分其余等位基因，缺失起始于第二外显子。VelociGene®法(Valenzuela DM等，(2003)自然生物技术(Nature Biotechnology)21:652-659)用于构建和使用基于细菌人工染色体的大打靶载体(LTVEC)对于使我们能够进行确保这类新型大功能性RNA的无效等位基因所需的大的基因切除缺失是关键的。

[0308] 关于可指导等位基因设计的lincRNA基因的结构与功能之间的关系知之甚少。关于Gt(ROSA)26Sor(Zambrowicz BP等，(1997)美国国家科学院院刊(Proceedings of the National Academy of Sciences of the United States of America)94:3789-3794)和BIC(miR-155)(Thai TH等，(2007)科学(Science)316:604-608)基因破坏的经验确定，在第一外显子后的缺失和插入可以产生β-半乳糖苷酶或其他报道基因的可靠的和组织特异性表达。然而，如果来自修饰等位基因的融合转录物仍然是来自在第一外显子中编码的5'部分的lincRNA的功能部分，则这种策略可能无法实现完全无效突变(Tsai,M.C.等，(2010)，科学(Science)329,689-693)。因此，表1中指示的敲除等位基因设计是希望完全切除突变(其将废除lincRNA功能的可能性最高)与产生等位基因(其从β-半乳糖苷酶报道基因产生精确和翔实的基因表达谱)的目标之间的折衷。例如，对于HOTAIR基因，产生两种等位基因，一种缺失几乎整个RNA编码序列并且第二种在第二外显子中起始缺失。两种等位基因产生相同表型(描述于下)，但仅第二种充当基因表达的报道基因。

[0309] 对于处在非常接近蛋白质编码基因并且可共有相异启动子的lincRNA，我们将缺失起点设置在第二外显子中以避免破坏邻近基因转录的可能性。图1示出Fendrr基因的这种实例。该图示出所有等位基因共有的设计元件：所有或大部分的编码lincRNA的序列的靶向缺失和被含有来自编码β-半乳糖苷酶的大肠杆菌lacZ基因的序列的表达盒和表达用于

选择G418抗性ES细胞群落的新霉素磷酸转移酶的表达盒(neo^{r})置换。在表型分析之前能够实现Cre介导切除的LoxP重组酶识别位点侧接药物选择盒。因为不存在与lacZ序列融合的功能性开放阅读,所以每个等位基因带有起始密码子和Kozak共有序列(Kozak,M.(1987),核酸研究(Nucleic Acids Res)15,8125-8148)以有效翻译β-半乳糖苷酶报道基因。

[0310] 由LacZ报道基因谱分析揭示的特异性和多种LincRNA基因表达模式

[0311] 为了研究20种靶向lincRNA基因的表达模式,我们将关于β-半乳糖苷酶活性的X-gal染色施用在妊娠中期胚胎和成体全固定组织和器官上。靶向lincRNA基因在胚胎和成体中展现多种独特的报道基因表达模式,这代表了大部分的主要器官系统和组织类型(图11的表2)。例如,在成体组织中,Pantr2、Kantr和Peril的表达局限于大脑;Mannr和Fendrr在肺中表达;Eldr在泌尿生殖系统中表达;并且Halr1在胸腔中表达。一种lincRNA基因Pint在所有组织中都展现普遍存在的表达。我们未检测Hotair、Ptgs2os2和Haglr基因在我们所研究的任何成体组织中的表达。

[0312] 胚胎表达似乎是lincRNA的共同特征。研究在胚胎期12.5天(E12.5)时或附近的杂合胚胎中的β-半乳糖苷酶报道基因表达揭示出所有20种靶向lincRNA基因的多种特定模式(图11的表2、图2A)。表达谱范围为普遍存在(Tug1)至高度特异性,例如表皮为Eldr,须基板为Trp53cor1(图9),或乳腺芽为Lincenc1(图9)。在不同程度的Hottip和Hoxa11os的肢芽和尾部表达中可见的时空模式与关于HoxA簇中的相邻蛋白质编码基因所报道的那些非常类似(Hostikka SL,Capecchi MR(1998)发育机制(Mechanisms of Development)70:133-145;Lu P等,(2008)发育(Development)135:1395-1405)。对于β-半乳糖苷酶报道基因所观测到的在后尾芽和生殖结节中的Hotair表达与通过原位杂交所确定的相同(Schorderet P,Duboule D(2011)PLoS遗传学(PLoS genetics)7:e1002071)。在胚胎发育不同时间的X-gal染色分析显示,对于一些lincRNA基因,表达早期开始于限制位点并且然后在稍后阶段扩展超出这个初始基因座(图2B),再次让人联想到Hox蛋白表达(Nagy A(2003)操纵小鼠胚胎:实验室手册(Manipulating the mouse embryo:a laboratory manual).第3版.Cold Spring Harbor,N.Y.:Cold Spring Harbor Laboratory Press.第x,764页)。例如,Hottip和Hoxa11os基因的表达开始于E9.5胚胎的极后期并且然后在稍后时间扩展至肢芽中。类似地,在接近E9.5胚胎前端的位点处Celrr的初始表达得以维持并且在接下来两天内扩展至神经管。

[0313] 与在人类组织特异性lncRNA中可见的频繁大脑表达一致(Derrien T等,(2012)基因组研究(Genome Research)22:1775-1789),我们发现20种靶向小鼠lincRNA基因中的一半在成体大脑中具转录活性。关于胚胎lincRNA表达,大脑模式(图3)是独特的并且从普遍存在(lincppara和Pint)改变为高度受限的特定大脑结构(Peril、Crnde和Kantr)。

[0314] Pint独特的表达随年龄增加与衰老样表型相关

[0315] 在所靶向的20种lincRNA基因中,仅Pint显示全局整体表达模式,其主要局限于出生后生活(图11的表2)。Pint独有的是,我们观测到其表达随年龄增加(图4)。在3天龄新生儿中,Pint转录活性较低(大脑)或不可检测(胸腔肌肉),但是随后逐渐出现在3周龄小鼠中并且到8周龄时变得强烈和普遍存在。虽然Pint表达的强度和时机在不同器官和组织中改变,但总体趋势是在出生后表达稳步增加至成年达到平台期。据我们所知,这种年龄相关的动态表达模式是新颖的。我们没有观测到在数百种蛋白质编码基因的lacZ谱分析实验的经

验中存在类似概况(Valenzuela DM等,(2003)自然生物技术(Nature Biotechnology)21:652-659)。

[0316] 通过lacZ谱分析揭示的全身Pint表达的引人注目的年龄相关增加(图4)表明,Pint在随着小鼠年龄增长维持正常健康中具有全局自我平衡作用。为了测试这种假说,我们将Pint敲除小鼠系培育成纯合性并且进行比较纯合($Pint^{-/-}$)小鼠与野生型(WT)和杂合($Pint^{+/-}$)同窝出生仔畜对照的纵向研究。所述$Pint^{-/-}$小鼠在出生时看来是健康和正常的;然而,在3月龄时,它们开始显示出早发型衰老样表型迹象。体重测量揭示出,雄性和雌性$Pint^{-/-}$小鼠相比于其WT同窝出生仔畜展现出较缓慢的生长速率,但是这在雄性中更明显(图5A)。到一岁龄时,雄性$Pint^{-/-}$小鼠相比于其WT同窝出生仔畜轻30%以上并且$Pint^{+/-}$小鼠轻15%,而$Pint^{-/-}$雌性轻27%(数据未示出)。比较纯合与杂合和WT雄性小鼠的Kaplan-Meier分析(图5B)表明,Pint的损失与不良存活结果相关。我们发现在突变小鼠中当它们衰老时无肿瘤或病变迹象,但是一些$Pint^{-/-}$小鼠患上突出症,包括在胸部上与腹壁变薄相关的剑突过程突出(数据未示出)。当小鼠从其尾部悬吊时,存在年龄依赖性异常后肢紧握姿势(数据未示出)。这种表型的严重程度改变,但其频率随着年龄逐渐增加,这表明肌肉力量下降(关于Hottip敲除系中的另一个实例,参见图8)。我们也观测到,雄性和雌性小鼠中的脱毛(数据未示出)。从$Pint^{-/-}$小鼠的腹侧和背侧体采集的皮肤切片的组织学分析揭示出纤维化和毛囊发育的明显差异以及皮下脂肪层厚度的急剧降低(图5C)。

[0317] 通过X射线纤维断层术(显微CT)对单独的小鼠在它们衰老时的非侵入性全身分析指示,在雄性(图5D)和雌性(图5E)$Pint^{-/-}$小鼠中相比于其WT同窝出生仔畜的脂肪含量显著降低。总身体脂肪损失可能是当它们衰老时体重下降的主要因素(图5A)。$Pint^{-/-}$小鼠相比于WT也具有显著降低的股骨骨矿物质密度(图5F和图5G)。雄性小鼠在52周龄时具有显著降低的瘦体质量。雄性和雌性显示开始于26周龄的腓肠肌复合物(GA)和胫骨前肌(TA)的显著降低的肌肉质量(未示出)。骨骼成像揭示出,在雄性和雌性$Pint^{-/-}$小鼠中,相比于WT,出现严重的脊椎弯曲(图5H)。约70%的12周龄$Pint^{-/-}$小鼠显示脊椎弯曲并且到52周龄时达到100%(图5I)。相比之下,仅10%至20%的26周龄WT小鼠显示轻微的脊椎弯曲并且这种频率不随着年龄而增加。$Pint^{+/-}$小鼠直到26周龄时也没有发展成显著的脊椎弯曲,这表明Pint的不同寻常的年龄依赖性单倍剂量不足。在Pint敲除小鼠中的年龄相关病变的范围表明,Pint对于维持健康和避免在小鼠正常寿命期间的过早衰老可能是重要的。

[0318] Fendrr缺失造成围产期致死作为呼吸窘迫的结果

[0319] 在20种lincRNA敲除小鼠系中,$Peril^{-/-}$和$Fendrr^{-/-}$小鼠显示围产期致死。我们的Fendrr敲除等位基因从外显子2至最后的注释外显子具有26kb缺失(图1)。E12.5纯合胚胎的X-gal染色显示,在额鼻过程、上呼吸道、肺中和后大动脉-生殖腺-中肾(AGM)区域中的lacZ表达(图6A),这在杂合(未示出)和纯合胚胎中是相同的,表明大致正常的器官发生。孤立地看在E13.5时发育的肺,揭示出敲除胚胎中的缺陷:肺较小并且肺叶出现球状和杂乱(图6B)。Fendrr基因缺失的纯合小鼠存活至出生,但在明显的呼吸困难后不久死亡。在2种不同的遗传背景上在小鼠中的Fendrr突变围产期致死表型是相同的:此处报道的C57BL6/129杂合体和在小鼠中在单独的育种计划中进一步与C57BL/6回交(Sauvageau M等,(2013)Elife 2:e01749)。

[0320] Hotair和Hottip的缺失造成骨骼和肌肉中的形态和功能缺陷

[0321] Hotair和Hottip基因的胚胎X-gal染色显示在后部和远端肢芽中的表达受限(图2A)。与这些发育受限的表达模式一致,Hotair和Hottip基因缺失造成成年小鼠的尾部和后肢的形态畸形。在Hotair$^{-/-}$小鼠中,我们观测到第4节尾椎的明显同源转化,其变得在解剖学上与第3节尾椎类似(图7)。Hottip$^{-/-}$小鼠当从其尾部悬吊时相比于野生型同窝出生仔畜显示异常的后肢紧握姿势(图8A)。这种行为异常伴随着握耐力损失,如通过其中将小鼠激发以保持悬吊在倒置的铁丝笼上的测试所测量。野生型和Hottip$^{+/-}$突变体抓住不放约一分钟,而其纯合同窝出生仔畜在10-20秒内松开(图8B)。这种明显的握力降低与腓肠肌而非胫骨前肌或四头肌的肌肉质量损失相关(图8C)。我们观测到腓肠肌中肌肉纤维数目降低约40%,而非平均纤维尺寸降低(图8D和图8E)。除了Hottip敲除小鼠中的肌肉缺陷之外,我们还发现后肢骨骼畸形:跟骨骨骼长度缩短(图10)。

[0322] 在过去几年中,特别是在哺乳动物中,我们对于基因组的非蛋白质编码组分的理解呈现爆发。除了长期以来公认的非编码功能性RNA类别如核糖体、转移、小核、小核仁、小胞质RNA,RNA酶P、RNA酶MRP和端粒酶的RNA组分以及最近发现的微RNA和PIWI相关piRNA之外,现在可包括长链非编码RNA类别的至少10,000种成员(Kapranov P等,(2007)科学(Science)316:1484-1488;Carninci P等,(2005)科学(Science)309:1559-1563;Djebali S等,(2012)自然(Nature)489:101-108;Derrien T等,(2012)基因组研究(Genome Research)22:1775-1789;Guttman M等,(2009)自然(Nature)458:223-227)。当我们开始理解基因组存在和lncRNA基因表达时,下一个目标是发现其生物功能。作为开始应对这一挑战的第一步骤,我们应用小鼠打靶技术(用于测定哺乳动物基因功能的最强大工具)来产生针对20种lincRNA基因的敲除小鼠系的资源(Sauvageau M等,(2013)Elife 2:e01749)。

[0323] 对于lincRNA的结构-功能关系知之甚少。为此,在这项初始研究中关键的是产生具有缺失的敲除等位基因,其除去大部分(若非全部)的lincRNA编码潜能以使得产生功能缺失型突变的可能性最高。许多lincRNA基因座的不明确的和不完全的注释(其中多个报道转录物或许是通过可选的剪接或转录起始位点产生),增加了敲除等位基因设计的难度。对于lincRNA功能重要的分子特性的新理解应向下一代lincRNA等位基因的设计提供对功能关键的更精确定向序列修饰的信息并且也容许先进和灵活的条件性策略。

[0324] 我们的lincRNA敲除研究的关键目标是产生等位基因,除了废除功能之外,也报道所述基因的时空表达模式。虽然没有蛋白质编码开放阅读框作为指导,但我们成功地设计了报道所有20种靶向基因的基因表达的等位基因。在成体阶段不产生lacZ表达的一种等位基因是Ptgs2os2(关于胚胎表达,参见图2A和图9),其已知是通过炎性信号最强烈诱导的lincRNA之一(Guttman M等,(2009)自然(Nature)458:223-227;Carpenter S等,(2013)科学(Science)341(6147):789-92)。Ptgs2os2敲除系应该证明lincRNA的表达如何响应于感染或其他炎性损伤以及它在过程中起到什么生物作用的有价值的研究资源。

[0325] 我们在选择这项研究靶向哪些lincRNA基因时所应用的标准之一是预期在神经组织中表达。十种靶向基因显示在成体大脑中的lacZ报道基因表达并且各自展现独特的模式(图3),范围从强烈全脑表达(Pint)至在大部分结构中的轻微灰质表达(Tug1)至丘核(Crnde)或下丘脑中线(Peril)所独有的高度受限的表达。大脑中的基因表达模式的多样性和特异性在其他组织中也是明显的并且与我们关于蛋白质编码基因的报道基因等位基因所见的类似。我们的lincRNA基因lacZ表达谱分析模式与在野生型小鼠组织中通过RNA定量

实验所见的组织特异性表达一致(Sauvageau M等,(2013)Elife 2:e01749)。然而,在这项研究之前,没有理解lincRNA基因表达的强烈组织和细胞类型特异性,因为先前的定量方法不能提供lacZ报道基因谱分析的高清晰度和细胞类型分辨率(图2A)。

[0326] 胚胎表达是我们所研究的所有lincRNA基因共有的特征。LacZ谱分析提供完整胚胎的高清晰度视图,其揭示出每种lncRNA所独有的广泛范围的特定模式。实例包括在须基板中关于Trp53cor1和在乳腺芽中关于lincenc1观测到的强烈特异性表达,Eldr的表皮表达,Hottip和Hoxa11os的肢芽表达,和Tug1的普遍存在的表达(图2A和图2B和图9)。这些改变的模式可能指向lincRNA在发育关键事件调控中的常见作用。

[0327] lacZ谱分析的另一种价值在于它可以指导和聚焦表型研究的设计。例如,杂合Hotair和Hottip胚胎的高度受限的后部表达模式表明,我们可能找出在后部身体部分中的敲除表型。与这种预期一致,我们观测到$Hotair^{-/-}$小鼠中第4节尾椎的明显同源转化(图6),并且我们发现$Hottip^{-/-}$小鼠中的后肢异常,其包括肌肉虚弱和骨骼畸形(图8和图10)。在具有不同Hotair敲除等位基因的小鼠中也观测到Hotair同源表型(H.Chang,个人通讯)。我们发现,杂合子中的Fendrr表达局限于成年小鼠中的肺(图11的表2)并且在胚胎中的呼吸道发育中是显著的(图2A)。或许并不奇怪,Fendrr纯合子展现呼吸窘迫并且随后由于肺的缺陷结构成熟而出现围产期死亡。我们的Fendrr敲除表型类似于罕见的人类致死肺发育障碍肺泡毛细血管发育不良与肺静脉错位(ACD/MPV),其中患者表现出肺叶发育不足并且在出生后数分钟至数小时内遭受出生后呼吸窘迫(Bishop NB等,(2011)美国呼吸与重症监护医学杂志(American Journal of Respiratory and Critical Care Medicine)184:172-179)。至少一名ACD/MPV患者报道在FOXF1-AS1基因内具有11kb缺失,所述FOXF1-AS1基因是小鼠Fendrr的人类同源物,表达于正常新生儿肺中(Szafranski P等,(2013)基因组研究(Genome Research)23:23-33)。Grote等(Grote P等,(2013),发育细胞(Developmental Cell)24:206-214)报道了具有Fendrr基因修饰的突变小鼠,其在约E13.75时产生与突出的脐膨出、腹侧体壁厚度降低和造成右心房积血的心脏缺陷相关的致死。我们没有观测到这些表型中的任一种。表型之间的差异可通过不同的等位基因设计来解释。我们的等位基因最后缺失Fendrr外显子2,被设计以避免可与相邻Foxf1蛋白质编码基因共用的启动子的破坏。Grote等(Grote P等,(2013),发育细胞(Developmental Cell)24:206-214)的Fendrr等位基因由在第一外显子中插入转录终止元件组成并且不包括报道基因。

[0328] 我们观测到的最明显的成体表达模式是对于Pint,其展现当小鼠从新生儿生长至成熟成体时X-gal染色的程度和强度增加(图4)。这种惊人的年龄相关模式促使我们对于生长速率和明显的异常健康迹象进行纵向分析。与WT小鼠相比,我们发现,随着$Pint^{-/-}$小鼠衰老,它们表现出渐进式脱发和肌肉减弱迹象、严重的脊椎弯曲、降低的身体脂肪和骨矿物质密度、较缓慢的生长速率和降低的存活率。令人惊讶的是,这些结果在杂合小鼠中被复制,但程度较轻。这种范围的年龄相关表型以及不同寻常的基因表达随年龄增加暗示小鼠可能需要临界剂量的Pint用于大体维持在正常寿命期间的健康和组织功能,并且首次指向LincRNA在生理衰老中的潜在作用。最近的研究显示,Pint是p53的直接靶标,通过多梳抑制复合物2(PRC2)提供p53途径与表观遗传沉默之间的联系(Marin-Bejar等,2013)。越来越多的身体证据表明,p53在细胞衰老和衰老控制中的关键作用。研究Pint的调控和其在p53依赖性细胞衰老和生物体衰老中的潜在参与将引起极大关注。这可能揭示在哺乳动物中的生

理衰老过程中的关键机制,其在人类疾病(包括与衰老和癌症相关的那些)中具有潜在临床意义。

[0329] 我们开始这项研究的目的不仅在于要阐明我们选择突变基因的20种特定lincRNA的功能,而且要获得对于一类lncRNA的一般性质的更好理解。这种集合可充当更大规模努力的种子来突变lincRNA基因家族的更多成员。许多lincRNA已显示与参与在染色质水平调控转录的蛋白质相关。这可能表明基因表达中的广泛、一般和冗余功能很像miRNA在转录后水平上维持组织特异性基因表达谱的相互作用。然而,我们的结果似乎指向不同的方向。此处和在Sauvageau等(Sauvageau M等,(2013)Elife 2:e01749)中描述的独特表型和强烈特异性表达模式主张lincRNA的特异性、直接和决定性功能。虽然这项研究仅仅是这种基因敲除小鼠集合分析的开始,但它揭示lincRNA作为一类新型功能性编码分子,与蛋白质一样,其在哺乳动物中的胚胎发育、生理学和一组广泛组织和器官的自我平衡中提供多种作用。

序列表

<110> K·V·赖
G·龚
J·里恩
D·弗伦德维
D·M·瓦伦泽拉

<120> 缺乏lincRNA的非人类动物

<130> 57766-448071

<150> 61/863,147
<151> 2013-08-07

<160> 2

<170> FastSEQ，用于Windows 4.0版

<210> 1
<211> 10
<212> DNA
<213> 人工序列

[0001] <220>
<223> Kozak共有序列

<220>
<221> misc_feature
<222> 1, 4
<223> R为A或G

<400> 1
rccrccatgg 10

<210> 2
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> Kozak共有序列

<220>
<221> misc_feature
<222> 7
<223> R为A或G

<400> 2
gccgccrcca tgg 13

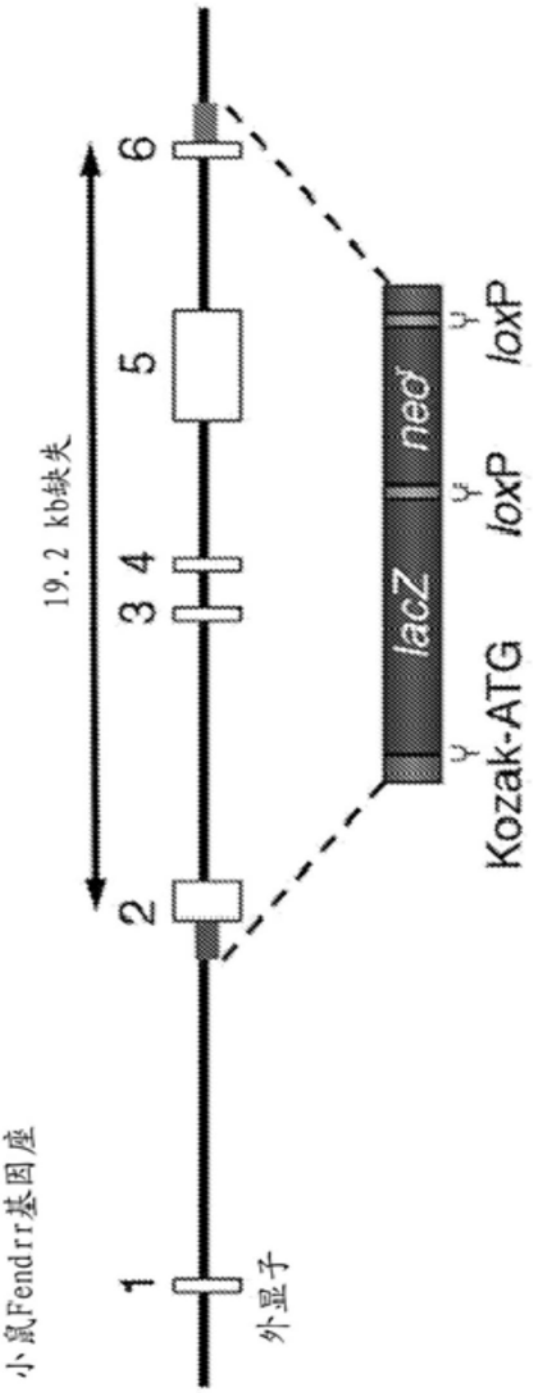

小鼠Fendrr基因座
外显子
1
2
3 4
5
6
19.2 kb缺失
Kozak-ATG
lacZ
neor
loxP
loxP

图1

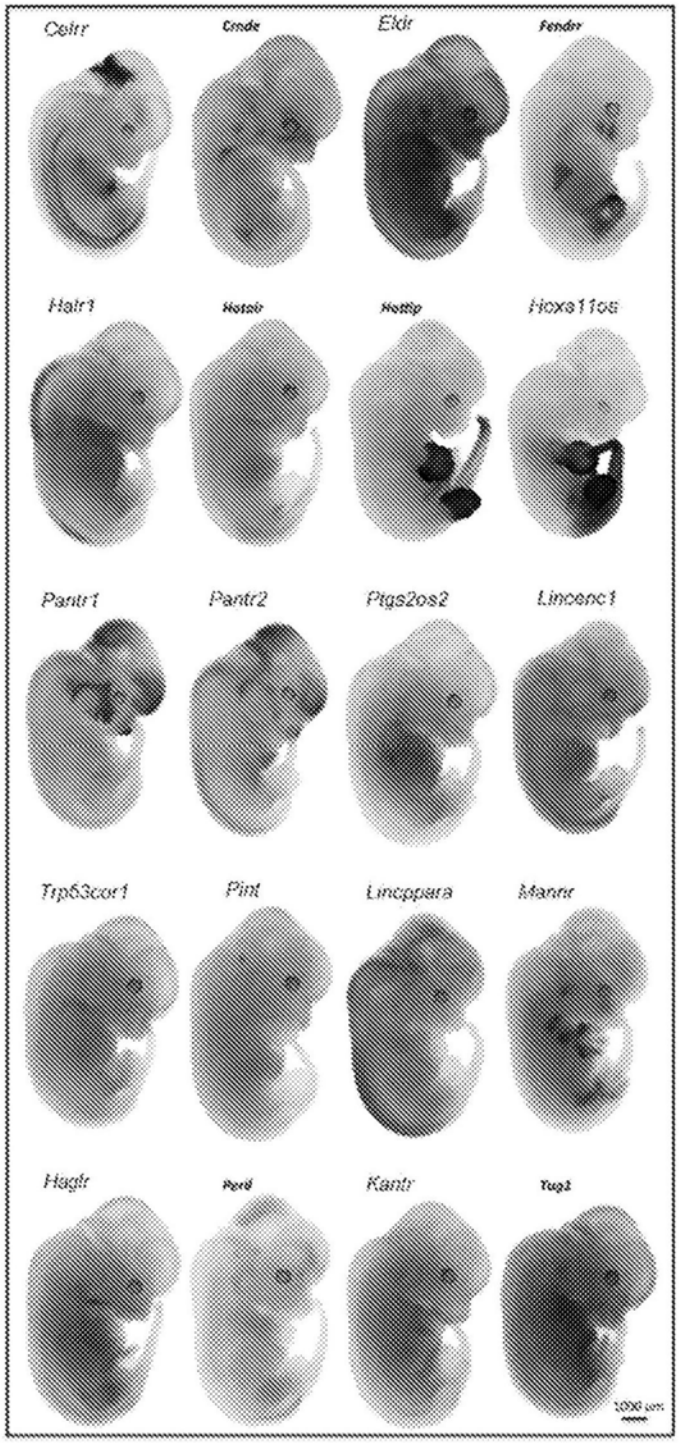

Celrr
Cmde
Eldr
Fendrr
Halr1
Hotair
Hottip
Hoxa11os
Pantr1
Pantr2
Ptgs2os2
Lincenc1
Trp53cor1
Pint
Lincppara
Mannr
Hagfr
Kantr
Tug1

图2A

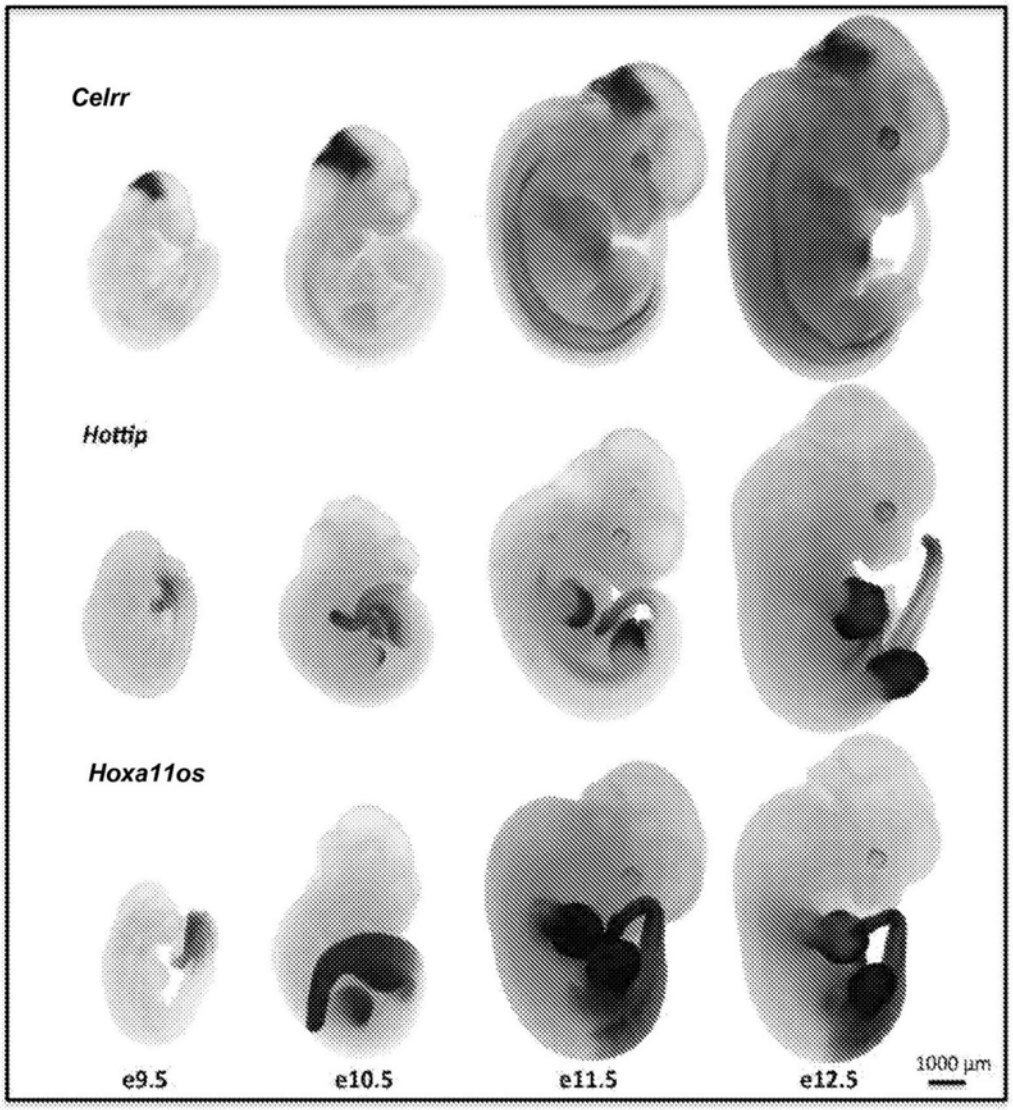


图2B

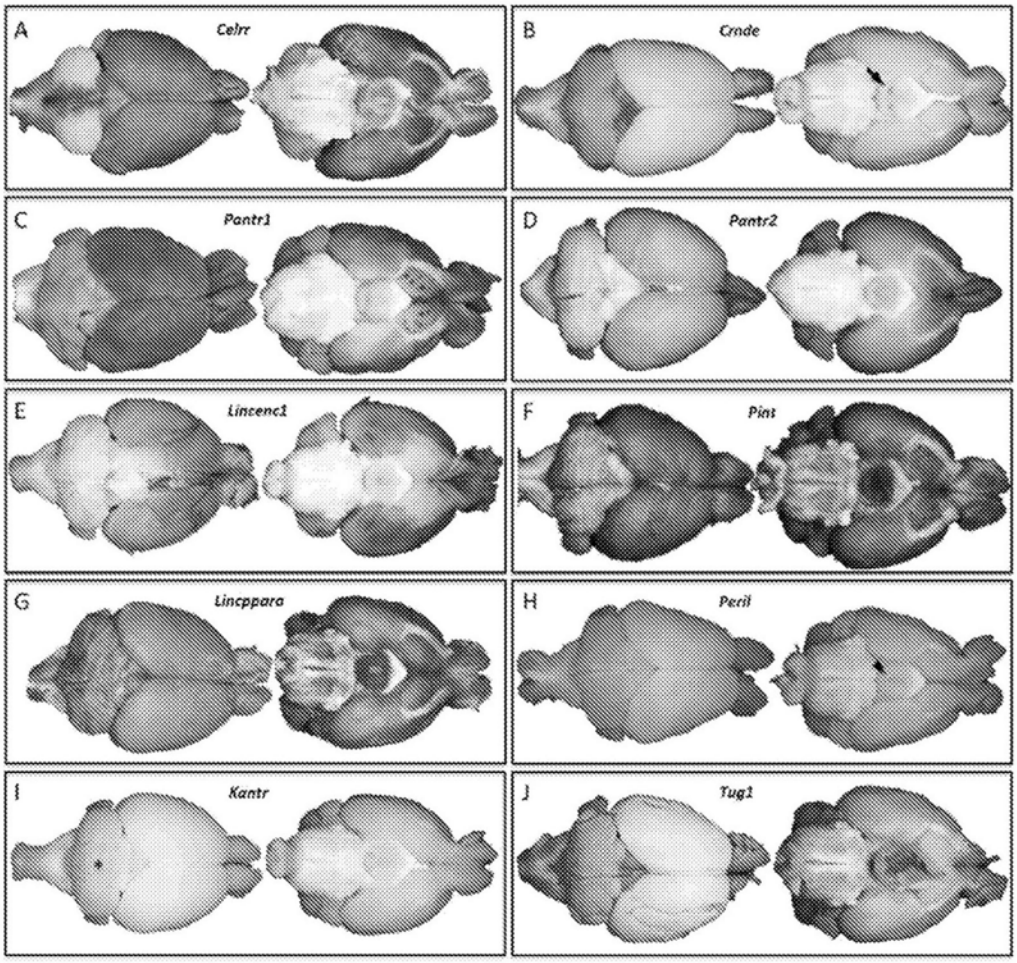

A
Celrr
B
Crnde
C
Pantr1
D
Pantr2
E
Lincenc1
F
Pint
G
Lincppara
H
Peril
I
Kantr
J
Tug1

图3

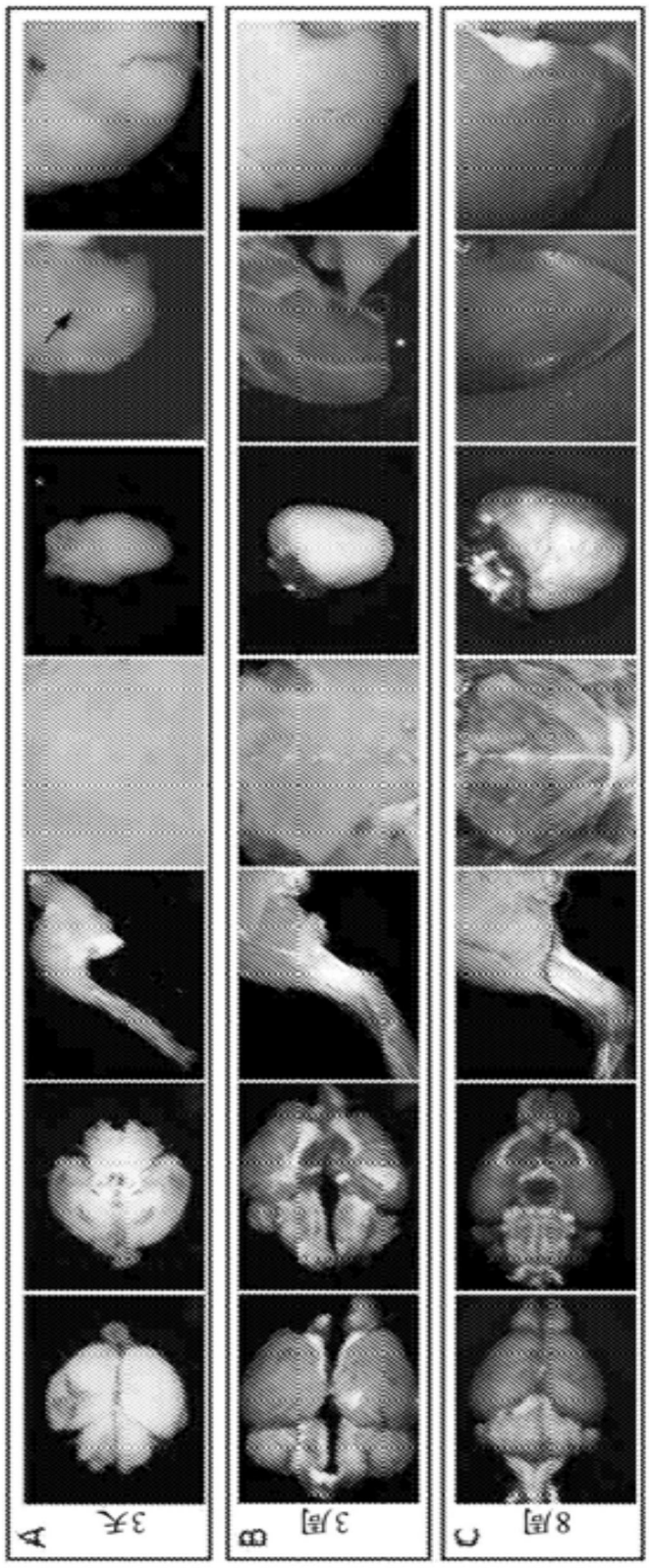


图4

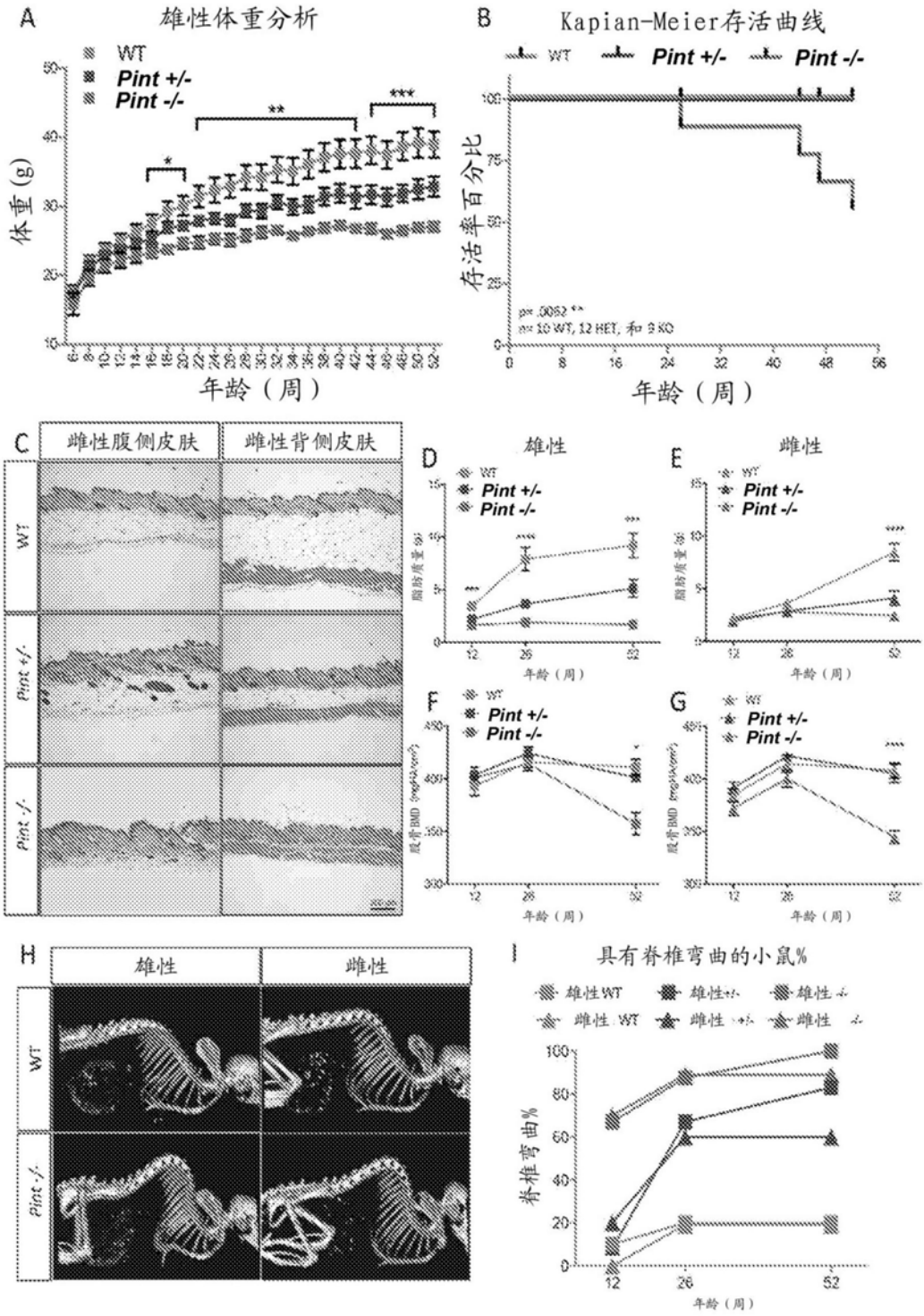

A
雄性体重分析
WT
Pint +/-
Pint -/-
体重(g)
年龄（周）
B
Kapian-Meier存活曲线
存活率百分比
年龄（周）
C
雌性腹侧皮肤
雌性背侧皮肤
D
雄性
E
雌性
脂肪质量 (g)
F
G
骨BMD
H
雄性
雌性
I
具有脊椎弯曲的小鼠%
脊椎弯曲%

图5

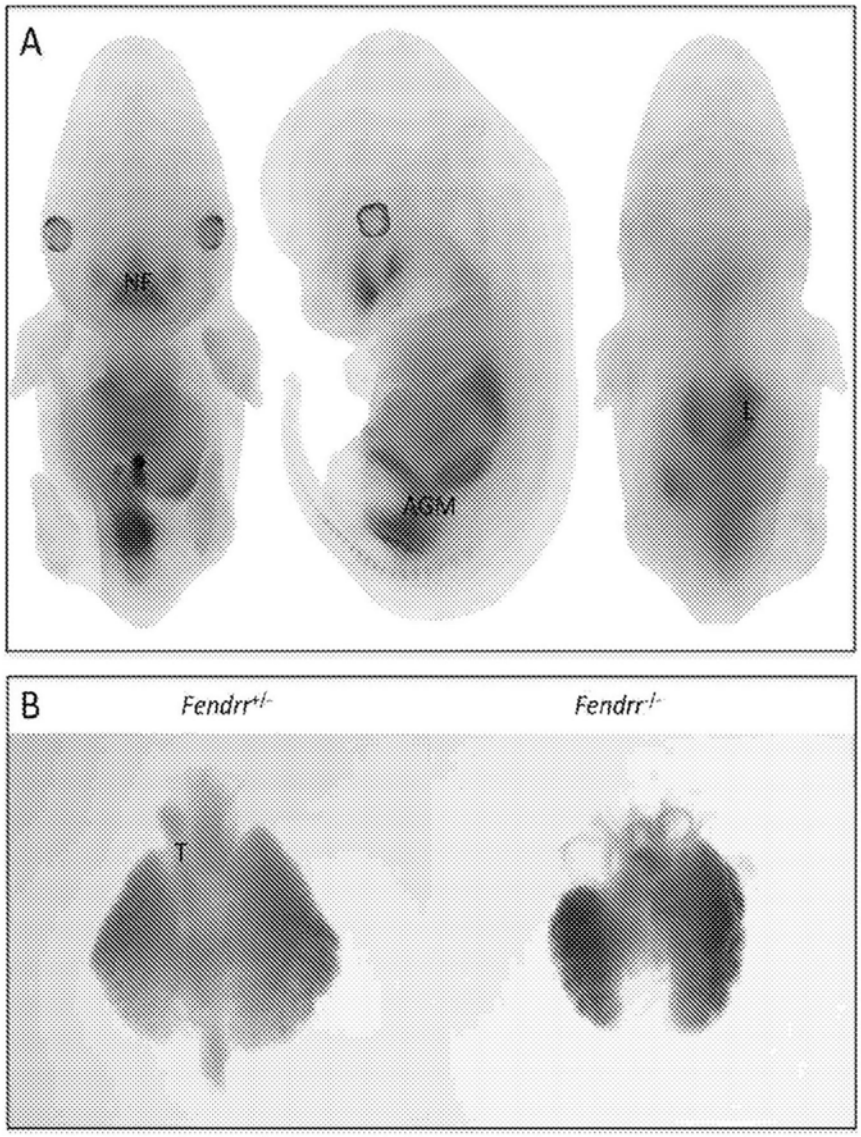


图6

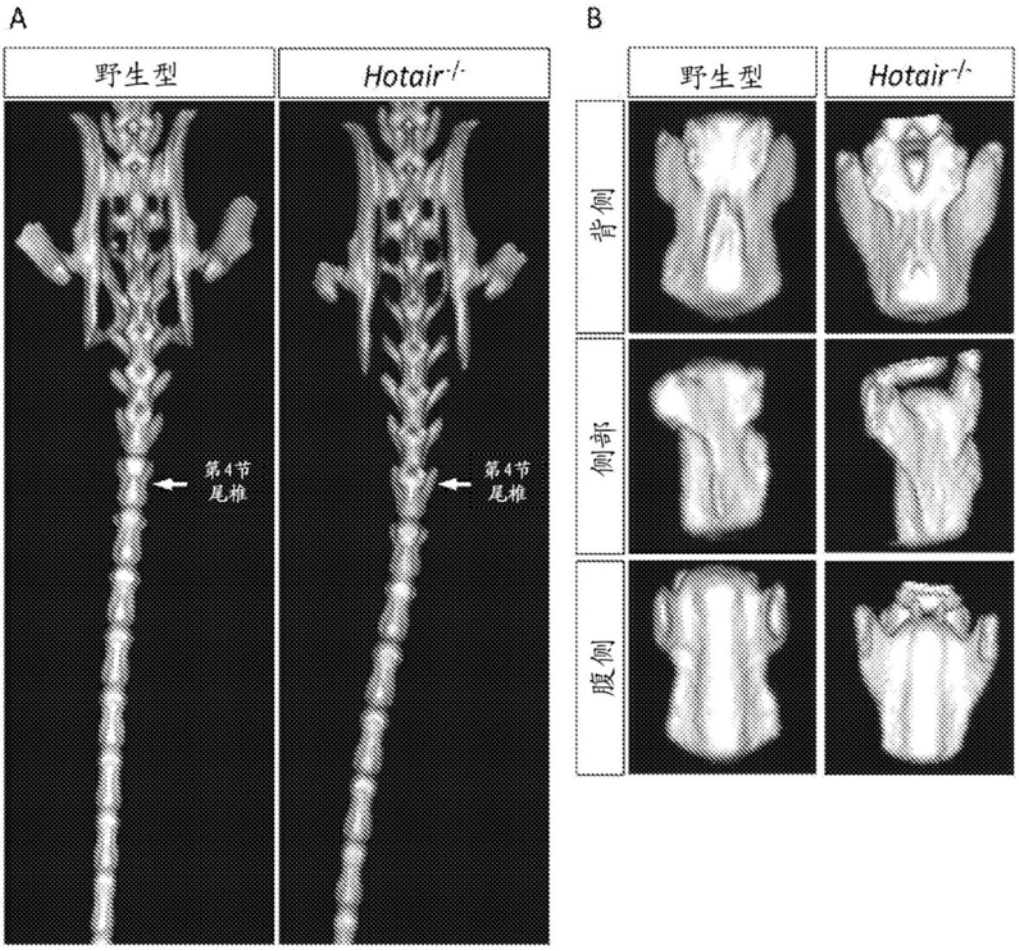


图7

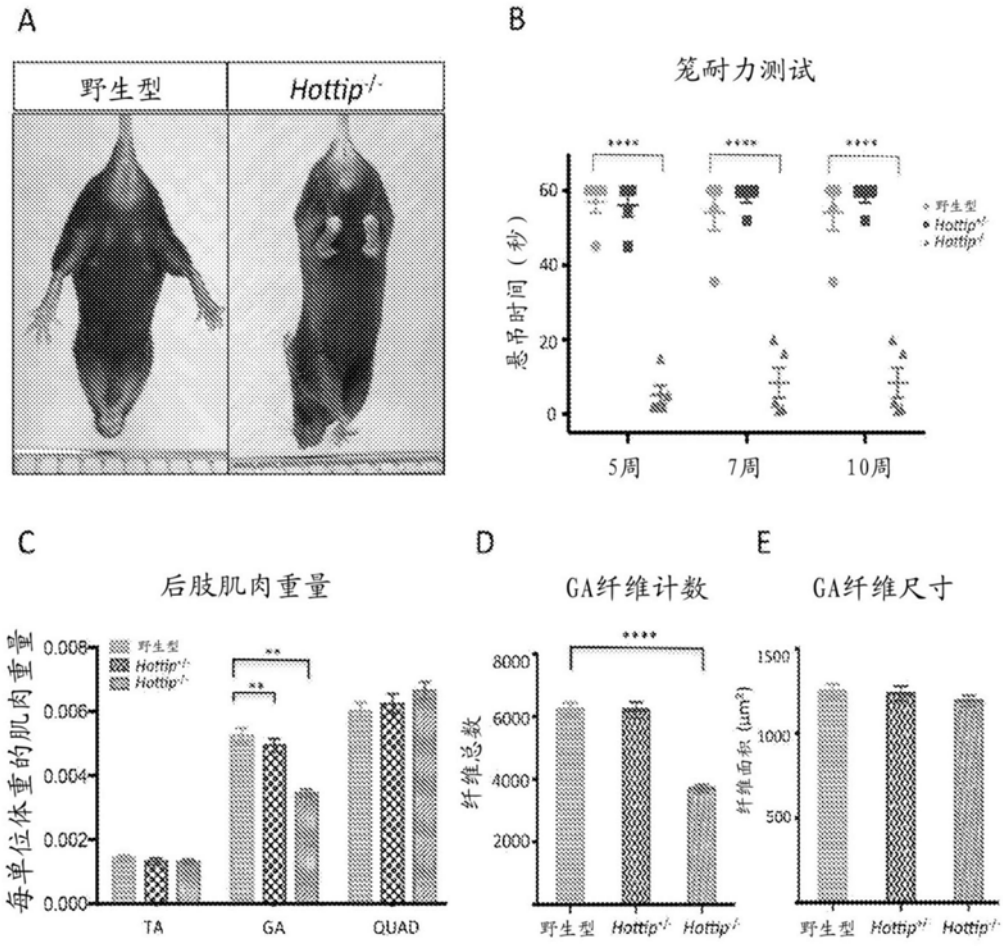

A
野生型
Hottip-/-
B
笼耐力测试
悬吊时间（秒）
5周
7周
10周
C
后肢肌肉重量
每单位体重的肌肉重量
TA
GA
QUAD
D
GA纤维计数
纤维总数
E
GA纤维尺寸
纤维面积 (μm²)

图8

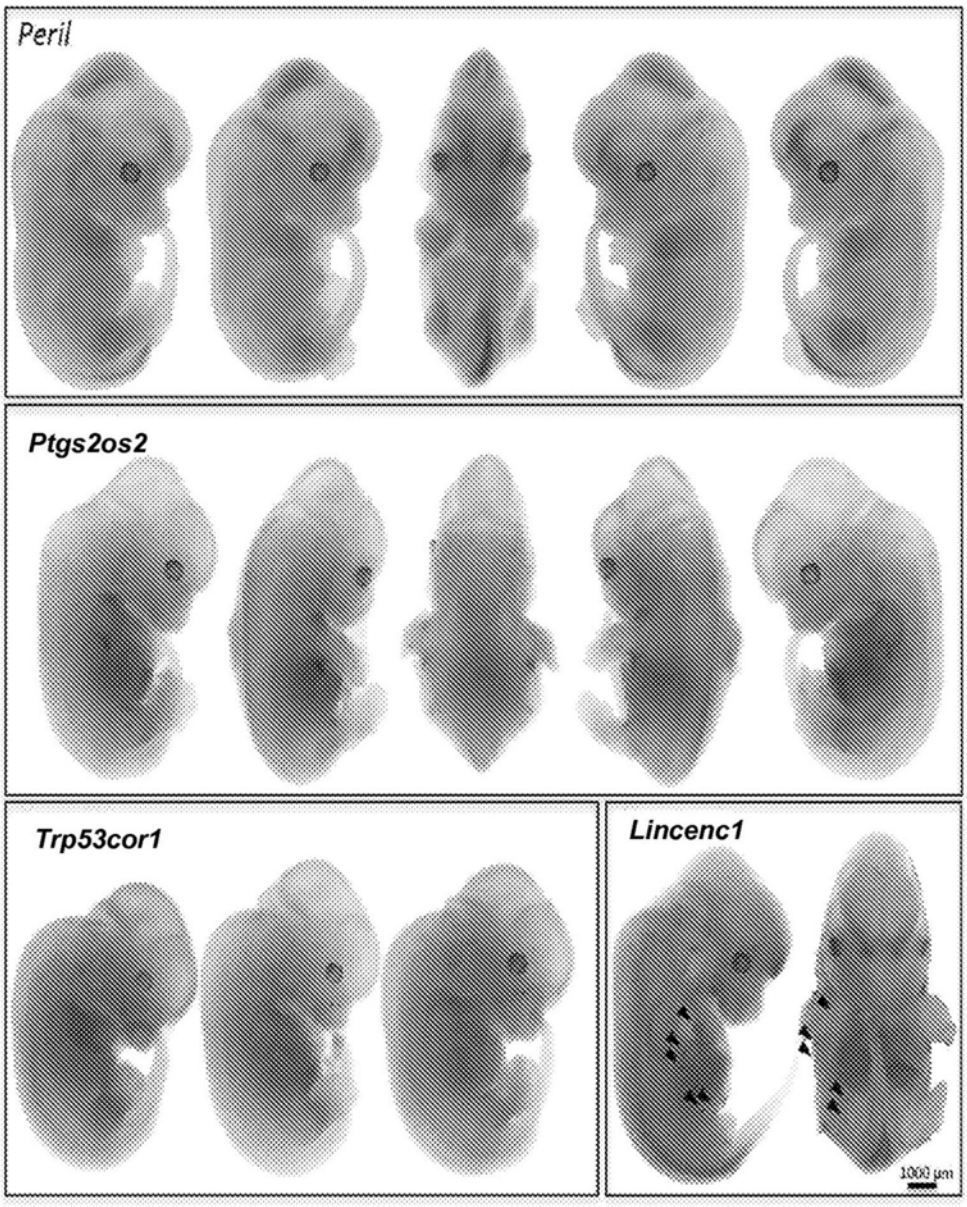


图9

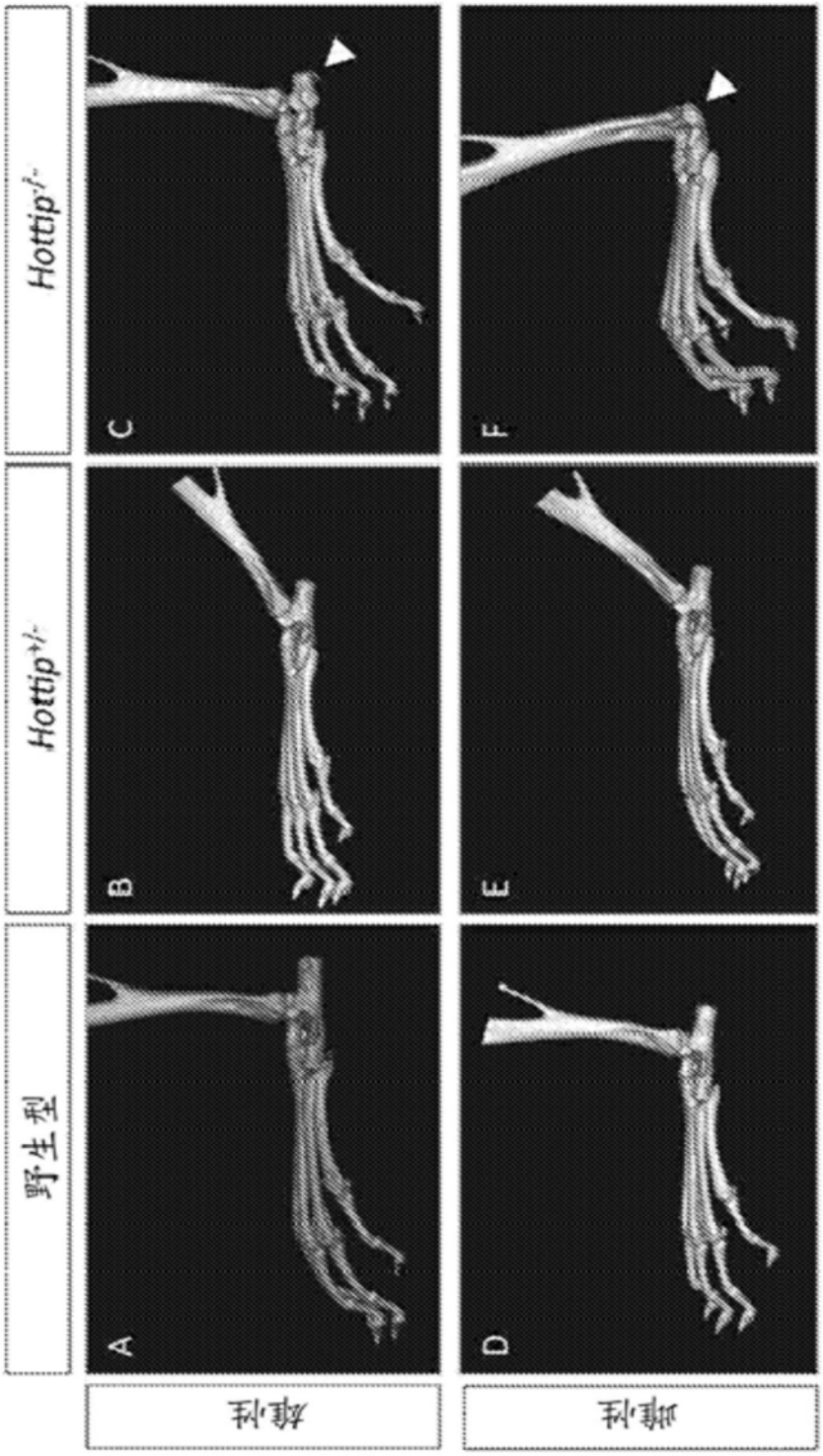


图10

表2. lacZ报道基因在胚胎和成体组织中的表达的汇总

LincRNA	胚胎	大脑	心脏	肺	肝脏	肋骨	脾脏	肠	胃	肾脏	泌尿生殖系统	后肢
Celrr	+	+				+						+
Crnde	+	+		+		+					+	+
Eldr	+										+	
Fendrr	+			+								
Halr1	+					+						
Hotair	+											
Hotlip	+							+			+	+
Hoxa11os	+					+		+		+	+	+
Pantr1	+	+				+				+		
Pantr2	+	+										
Ptgs2os2	+											
lincenc1	+	+										+
Trp53cor1	+									+	+	
Pint	+	+	+	+	+	+	+	+	+	+	+	+
lincppara	+	+		+		+			+		+	+
Mannr	+			+								
Haglr	+											
Penl	+	+										
Kantr	+	+										
Tug1	+	+	+			+				+	+	

图11