



(12) 发明专利申请

(10) 申请公布号 CN 116670769 A

(43) 申请公布日 2023. 08. 29

(21) 申请号 202180079761.9

(22) 申请日 2021.09.30

(30) 优先权数据

63/085,878 2020.09.30 US

(85) PCT国际申请进入国家阶段日

2023.05.26

(86) PCT国际申请的申请数据

PCT/US2021/052987 2021.09.30

(87) PCT国际申请的公布数据

W02022/072709 EN 2022.04.07

(71) 申请人 加利福尼亚大学董事会

地址 美国加利福尼亚州

(72) 发明人 A·R·穆奥特里 F·帕佩斯

(74) 专利代理机构 上海弼兴律师事务所 31283

专利代理师 袁文正 徐婕超

(51) Int.Cl.

G16B 20/20 (2006.01)

权利要求书4页 说明书48页

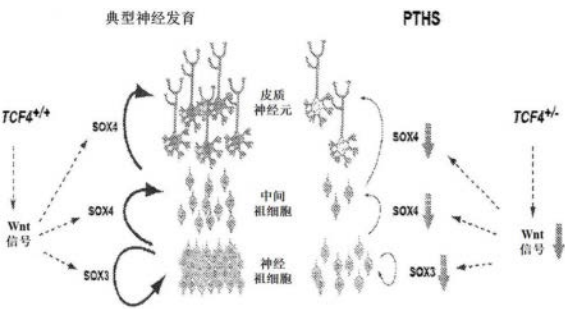
序列表25页 附图93页

(54) 发明名称

调节TCF4基因表达和治疗皮特霍普金斯综合征的组合物和方法

(57) 摘要

本公开提供了编码TCF4多肽的重组盒和载体,及其在治疗神经或神经发育疾病和病症中的应用。



1. 一种重组核酸构建体,其包含微型启动子,所述微型启动子可操作地连接到TCF4多肽的编码序列。

2. 如权利要求1所述的重组核酸构建体,其中,所述重组核酸构建体还包含一个或多个转录因子结合基序。

3. 如权利要求2所述的重组核酸构建体,其中,所述一个或多个转录因子结合基序为microE5基序。

4. 如权利要求3所述的重组核酸构建体,其中,所述重组核酸构建体包含1至15个microE5基序。

5. 如权利要求4所述的重组核酸构建体,其中,所述重组核酸构建体包含至少5个microE5基序。

6. 如权利要求4所述的重组核酸构建体,其中,所述重组核酸构建体包含至少10个microE5基序。

7. 如权利要求4所述的重组核酸构建体,其中,所述重组核酸构建体包含12个microE5基序。

8. 如权利要求4所述的重组核酸构建体,其中,所述重组核酸构建体具有以下一般结构: microE5_n-微型启动子-TCF4编码序列,其中n为5至15范围内的整数。

9. 如权利要求4至8中任一项所述的重组核酸构建体,其中,所述microE5基序包含如SEQ ID NO:10所示的核苷酸序列。

10. 如权利要求1至9中任一项所述的重组核酸构建体,其中,所述TCF4多肽为TCF4-B。

11. 如权利要求10所述的重组核酸构建体,其中,所述TCF4多肽包含与SEQ ID NO:2具有至少85%序列同一性的氨基酸序列。

12. 如权利要求11所述的重组核酸构建体,其中,所述TCF4多肽包含与SEQ ID NO:2具有至少90%序列同一性的氨基酸序列。

13. 如权利要求11所述的重组核酸构建体,其中,所述TCF4多肽包含与SEQ ID NO:2具有至少95%序列同一性的氨基酸序列。

14. 如权利要求11所述的重组核酸构建体,其中,所述TCF4多肽包含与SEQ ID NO:2具有至少98%序列同一性的氨基酸序列。

15. 如权利要求1至14中任一项所述的重组核酸构建体,其中,所述TCF4编码序列包含与SEQ ID NO:1至少80%相同的核苷酸序列。

16. 如权利要求15所述的重组核酸构建体,其中,所述TCF4编码序列包含与SEQ ID NO:1至少85%相同的核苷酸序列。

17. 如权利要求16所述的重组核酸构建体,其中,所述TCF4编码序列包含与SEQ ID NO:1至少90%相同的核苷酸序列。

18. 如权利要求16所述的重组核酸构建体,其中,所述TCF4编码序列包含与SEQ ID NO:1至少95%相同的核苷酸序列。

19. 如权利要求1至18中任一项所述的重组核酸构建体,其中,所述微型启动子包含核心启动子。

20. 如权利要求19所述的重组核酸构建体,其中,所述微型启动子包含与SEQ ID NO:3具有至少70%序列同一性的核苷酸序列。

21. 如权利要求19所述的重组核酸构建体,其中,所述微型启动子包含与SEQ ID NO:3具有至少80%序列同一性的核苷酸序列。

22. 如权利要求19所述的重组核酸构建体,其中,所述微型启动子包含与SEQ ID NO:3具有至少90%序列同一性的核苷酸序列。

23. 如权利要求19所述的重组核酸构建体,其中,所述微型启动子包含如SEQ ID NO:3所示的核苷酸序列,任选地具有独立选自缺失、插入和替换的1至5个核苷酸修饰。

24. 如权利要求1至23中任一项所述的重组核酸构建体,其中,所述核酸构建体包含与SEQ ID NO:4、5、6、7或8中的任何一个至少80%相同的核苷酸序列。

25. 一种包含如权利要求1至24中任一项所述的重组核酸构建体的载体。

26. 如权利要求25所述的载体,其中,所述载体为病毒载体。

27. 如权利要求26所述的载体,其中,所述病毒载体为逆转录病毒载体。

28. 如权利要求25所述的载体,其中,所述载体为腺相关病毒(AAV)载体、慢病毒载体或gamma-逆转录病毒载体。

29. 如权利要求28所述的载体,其中,所述载体为AAV9载体。

30. 一种重组细胞,其包含如权利要求1至24中任一项所述的重组核酸构建体或如权利要求25至28中任一项所述的载体。

31. 一种包含如权利要求25至29中任一项所述的载体的药物组合物。

32. 一种治疗受试者相关的神经或神经发育的疾病或病症的方法,包含用如权利要求1至24中任一项所述的重组核酸构建体转化所述受试者的神经元,或施用如权利要求25至39中任一项所述的载体,或向所述受试者施用如权利要求31所述的药物组合物。

33. 如权利要求32所述的方法,其中,所述神经或神经发育的疾病或病症为皮特-霍普金斯综合征、精神分裂症、自闭症、自闭症谱系障碍或18q综合征。

34. 如权利要求32所述的方法,其中,所述神经或神经发育的疾病或病症为皮特-霍普金斯综合征,并且与TCF4单倍体剂量不足有关。

35. 如权利要求32至34中任一项所述的方法,其中,所述受试者在TCF4基因中具有一个或多个单核苷酸多态性。

36. 如权利要求32至34中任一项所述的方法,其中,所述受试者具有包含至少部分TCF4基因的染色体缺失。

37. 如权利要求36所述的方法,其中,所述受试者具有TCF4基因的完全缺失。

38. 如权利要求32至34中任一项所述的方法,其中,所述受试者具有包含至少部分TCF4基因的染色体易位。

39. 如权利要求32至34中任一项所述的方法,其中,所述受试者在TCF4基因中具有易位、移码或无义突变。

40. 如权利要求32至39中任一项所述的方法,其中,所述受试者为婴儿受试者或儿科受试者。

41. 如权利要求40所述的方法,其中,所述受试者约为小于或等于16岁。

41. 如权利要求40所述的方法,其中,所述受试者约为小于或等于12岁。

42. 如权利要求40所述的方法,其中,所述受试者约为小于或等于8岁。

43. 如权利要求40所述的方法,其中,所述受试者约为小于或等于5岁。

44. 如权利要求40所述的方法,其中,所述受试者约为小于或等于2岁。
45. 如权利要求32至39中任一项所述的方法,其中,所述受试者为成人受试者。
46. 一种治疗受试者中TCF4单倍体剂量不足相关的神经或神经发育的疾病或病症的方法,其包含在所述受试者的神经元中增加一种或多种SOX3和SOX4的表达。
47. 如权利要求46所述的方法,其中,通过向所述神经元引入表达TCF4-B多肽的重组核酸构建体来增加SOX3和/或SOX4的表达。
48. 如权利要求47所述的方法,其中,所述重组核酸构建体包含微型启动子,所述微型启动子可操作地连接到TCF4-B多肽的编码序列。
49. 如权利要求48所述的方法,其中,所述重组核酸构建体还包含一个或多个转录因子结合基序。
50. 如权利要求49所述的方法,其中,所述一个或多个转录因子结合基序为microE5基序。
51. 如权利要求50所述的方法,其中,所述重组核酸构建体包含1至15个microE5基序。
52. 如权利要求51所述的方法,其中,所述重组核酸构建体包含至少5个microE5基序。
53. 如权利要求51所述的方法,其中,所述重组核酸构建体包含至少10个microE5基序。
54. 如权利要求51所述的方法,其中,所述重组核酸构建体包含12个microE5基序。
55. 如权利要求51所述的方法,其中,所述重组核酸构建体具有以下一般结构: microE5_n-微型启动子-TCF4编码序列,其中n为5至15范围内的整数。
56. 如权利要求50至55中任一项所述的方法,其中,所述microE5基序包含如SEQ ID NO:10所示的核苷酸序列。
57. 如权利要求47至56中任一项所述的方法,其中,用病毒载体将所述表达TCF4-B多肽的重组核酸构建体递送到所述受试者。
58. 如权利要求57所述的方法,其中,所述病毒载体为逆转录病毒载体。
59. 如权利要求57所述的方法,其中,所述载体为腺相关病毒(AAV)载体、慢病毒载体或gamma-逆转录病毒载体。
60. 如权利要求59所述的方法,其中,所述载体为AAV9载体。
61. 如权利要求46至60中任一项所述的方法,其中,所述神经或神经发育的疾病或病症为皮特-霍普金斯综合征、精神分裂症、自闭症、自闭症谱系障碍或18q综合征。
62. 如权利要求61所述的方法,其中,所述神经或神经发育的疾病或病症为皮特-霍普金斯综合征。
63. 如权利要求46至62中任一项所述的方法,其中,所述受试者在TCF4基因中具有一个或多个单核苷酸多态性。
64. 如权利要求46至62中任一项所述的方法,其中,所述受试者具有包含至少部分TCF4基因的染色体缺失。
65. 如权利要求46至62中任一项所述的方法,其中,所述受试者具有TCF4基因的完全缺失。
66. 如权利要求46至62中任一项所述的方法,其中,所述受试者具有包含至少部分TCF4基因的染色体易位。
67. 如权利要求46至62中任一项所述的方法,其中,所述受试者在TCF4基因中具有易

位、移码或无义突变。

68. 如权利要求46至67中任一项所述的方法, 其中, 所述受试者为婴儿受试者或儿科受试者。

69. 如权利要求68所述的方法, 其中, 所述受试者约为小于或等于16岁。

70. 如权利要求68所述的方法, 其中, 所述受试者约为小于或等于12岁。

71. 如权利要求68所述的方法, 其中, 所述受试者约为小于或等于8岁。

72. 如权利要求68所述的方法, 其中, 所述受试者约为小于或等于5岁。

73. 如权利要求68所述的方法, 其中, 所述受试者约为小于或等于2岁。

74. 如权利要求47至67中任一项所述的方法, 其中, 所述受试者为成人受试者。

调节TCF4基因表达和治疗皮特霍普金斯综合征的组合物和方法

[0001] 相关申请的交叉引用

[0002] 本申请根据35 U.S.C. §119要求2020年9月30日提交的序列号63/085,878的美国临时申请的优先权,其公开内容通过引用并入本文。

技术领域

[0003] 本公开涉及用于治疗神经或神经发育疾病和病症的方法和组合物。

[0004] 通过引用并入序列表

[0005] 随附本申请文件的标题为“Sequence-Listing_ST25”的序列表,创建于2021年9月30日,具有40,086字节的数据,在IBM-PC、MS-Windows操作系统上机器格式化。出于所有目的,特此通过引用并入整个序列表。

背景技术

[0006] 神经和神经发育的疾病和病症是慢性的和令人衰弱的,其包括精神分裂症、自闭症、自闭症谱系障碍。皮特-霍普金斯综合征(PTHS)和18q综合征是罕见的神经发育障碍,其症状包括智力障碍、语言习得失败、运动学习缺陷、过度换气、癫痫、自闭症行为和胃肠道异常。在精神分裂症的临床全基因组关联分析(GWAS)中,包含TCF4的基因座中的特定单核苷酸多态性(SNPs)是最快达到全基因组显著的多态性之一。这些神经精神障碍的特点是明显的认知缺陷,这表明这些障碍之间不仅存在遗传重叠,还可能存在病理生理学重叠。

[0007] TCF4是一种碱性的螺旋-环-螺旋(bHLH)转录因子(TF),其与自身或其他bHLH TFs形成同源或异源二聚体。TCF4的二聚化允许其识别E-box结合位点(基序:CANNTG),并且指导DNA结合可能导致转录的抑制或激活,这取决于与TCF4结合的蛋白质复合物。在人类发育过程中,TCF4基因在整个中枢神经系统(CNS)中高度表达,但其表达和剪接的调节是复杂的,因为已经鉴定出多个包含不同5'外显子和内部剪接的选择性转录本。TCF4下游调控的基因尚不清楚;由于E-box序列的特异性有限,以及由于异二聚化、发育表达和细胞类型特异性而导致的TCF4的环境依赖性调节,使得情况变得复杂。

发明内容

[0008] 本公开提供了用于将分子递送到细胞中的方法和组合物。

[0009] 本公开提供了一种重组核酸构建体,其包括可操作地连接到TCF4多肽的编码序列的微型启动子。在一个实施方案中,所述核酸还包括一个或多个转录因子结合基序。在进一步的实施方案中,所述一个或多个转录因子结合基序是microE5基序。在另一实施方案中,所述重组核酸包含1至15个microE5基序。在又一实施方案中,所述重组核酸包含至少5个microE5基序、至少10个microE5基序或至少12个microE5基序。在另一或进一步的实施方案中,所述核酸具有以下一般结构: microE5_n-微型启动子-TCF4编码序列,其中n是5至15范围内的整数。在再一个或进一步的实施方案中,所述microE5基序包含如SEQ ID NO:10所示的

核苷酸序列。在前述任何实施方案的又一或进一步实施方案中,所述TCF4多肽是TCF4-B。在进一步的实施方案中,所述TCF4多肽包含与SEQ ID NO:2具有至少85%、90%、95%、98%或更高序列同一性的氨基酸序列。在前述任何实施方案的又一个或进一步实施方案中,所述TCF4编码序列包括与SEQ ID NO:1具有至少80%、85%、90%、95%或更高同一性的核苷酸序列。在另一个实施方案中,所述TCF4编码序列在严格的条件下与SEQ ID NO:1组成的序列杂交。在另一实施方案中,上述任何实施方案中的微型启动子包括核心启动子。在进一步的实施方案中,所述微型启动子包含与SEQ ID NO:3具有至少70%、80%、90%或更高序列同一性的核苷酸序列。在另一个实施方案中,所述微型启动子包含如SEQ ID NO:3所示的核苷酸序列,任选地其具有独立选自缺失、插入和替换的1至5个核苷酸修饰。在上述任何实施方案的另一或进一步实施方案中,所述核酸构建体包含与SEQ ID NO:4、5、6、7或8中任何一个至少80%相同的核苷酸序列。

[0010] 本公开还提供了一种包含上述任何一种重组核酸的载体。在进一步的实施方案中,所述载体是病毒载体。在另一个实施方案中,所述病毒载体是逆转录病毒载体。在又一个实施方案中,所述载体是腺相关病毒(AAV)载体、慢病毒载体或gamma-逆转录病毒载体。在另一实施方案中,所述载体是AAV9载体。

[0011] 本公开还提供了一种重组细胞,其包括本公开的所述重组核酸构建体或本公开的所述载体。

[0012] 本公开还提供了一种包含本公开所述载体的药物组合物。

[0013] 本公开还提供了一种治疗受试者的神经或神经发育的疾病或病症的方法,包括用本公开所述的重组核酸构建体转化受试者的神经元或施用本公开所述的载体,或将本公开所述的药物组合物施用给所述受试者。在另一个实施方案中,所述神经或神经发育疾病或病症是皮特-霍普金斯综合征、精神分裂症、自闭症、自闭症谱系障碍或18q综合征。在另一个实施方案中,所述神经或神经发育的疾病或病症是皮特-霍普金斯综合征,并且与TCF4单倍体剂量不足有关。在另一个或进一步的实施方案中,所述受试者在TCF4基因中具有一个或多个单核苷酸多态性。在另一个或进一步的实施方案中,所述受试者具有包含至少部分TCF4基因的染色体缺失。在进一步的实施方案中,所述受试者具有TCF4基因的完全缺失。在另一个实施方案中,所述受试者具有包含至少部分TCF4基因的染色体易位。在另一个实施方案中,所述受试者在TCF4基因中具有易位、移码或无义突变。在另一个或进一步实施方案中,所述受试者是婴儿或儿科受试者。在进一步的实施方案中,所述受试者约为小于或等于16岁。在又一个实施方案中,所述受试者约为小于或等于12岁。在又一个实施方案中,所述受试者约为小于或等于8岁、约小于或等于5岁、或约小于或等于2岁。在另一个实施方案中,所述受试者是成人受试者。

[0014] 本公开还提供了一种治疗与受试者中TCF4单倍体剂量不足有关的神经或神经发育的疾病或病症的方法,包括在所述受试者的神经元中增加一种或多种SOX3和SOX4的表达。在另一个实施方案中,通过向所述神经元引入表达TCF4-B多肽的重组核酸构建体来增加SOX3和/或SOX4的表达。在进一步的实施方案中,所述重组核酸构建体包含可操作地连接到TCF4-B多肽的编码序列的微型启动子。在更进一步的实施方案中,所述重组核酸构建体还包含一个或多个转录因子结合基序。在进一步的实施方案中,所述一个或多个转录因子结合基序是microE5基序。在另一实施方案中,所述重组核酸构建体包含1至15个microE5基

序。在另一个或进一步的实施方案中,所述重组核酸包含至少5个、至少10个或至少12个microE5基序。在另一个实施方案中,所述核酸具有以下一般结构: microE5_n-微型启动子-TCF4编码序列,其中n是5至15范围内的整数。在另一或进一步的实施方案中,所述microE5基序包含如SEQ ID NO:10所示的核苷酸序列。在又一个或进一步的实施方案中,用病毒载体将所述表达TCF4-B多肽的重组核酸构建体递送到所述受试者。在进一步的实施方案中,所述病毒载体为逆转录病毒载体。在另一实施方案中,所述载体是腺相关病毒(AAV)载体、慢病毒载体或gamma-逆转录病毒载体。在进一步的实施方案中,所述载体是AAV9载体。在上述任何实施方案的另一个或进一步实施方案中,所述神经或神经发育的疾病或病症是皮特-霍普金斯综合征、精神分裂症、自闭症、自闭症谱系障碍或18q综合征。在进一步的实施方案中,所述神经或神经发育的疾病或病症是皮特-霍普金斯综合征。在前述任何实施方案的又一或进一步实施方案中,所述受试者在TCF4基因中具有一个或多个单核苷酸多态性。在前述任何实施方案的又一或进一步实施方案中,所述受试者具有包含至少部分TCF4基因的染色体缺失。在前述任何实施方案的另一或进一步实施方案中,所述受试者具有TCF4基因的完全缺失。在前述任何实施方案的又一或进一步实施方案中,所述受试者具有包含至少部分TCF4基因的染色体易位。在前述任何实施方案的又一或进一步实施方案中,所述受试者在TCF4基因中具有易位、移码或无义突变。在前述任何实施方案的又一或进一步实施方案中,所述受试者是婴儿或儿科受试者。在进一步的实施方案中,所述受试者约为小于或等于16岁、约小于或等于12岁、约小于或等于8岁、约小于或等于5岁或约小于或等于2岁。在另一个实施方案中,所述受试者是成人受试者。

[0015] 一个或多个实施方案的细节在附图和下面的描述中阐述。从说明书和附图以及权利要求书中,其他特征、目的和优点将变得显而易见。

附图说明

[0016] 图1A-B显示了神经祖细胞中TCF4的表达水平。(A) 每个条形代表PTHS个体(红色条形)和健康对照(患病子代的亲本,黑色条形)产生的RNA测序文库中的TPM表达丰度(每百万转录)。横轴上的转录本异构体根据Ensembl数据库记录的人类TCF4基因命名。(B) 每个转录本异构体与相应蛋白质异构体之间的对应关系,根据Sepp等人,2011年命名(Functional Diversity of Human Basic Helix-Loop-Helix Transcription Factor TCF4 Isoforms Generated by Alternative 59Exon Usage and Splicing.PLoS ONE 6(7):e22138.doi:10.1371/journal.pone.0022138)。

[0017] 图2A-C显示了TCF4过表达的DNA盒的构建和验证。(A) 表达构建体的示意图,其中TCF4-B cDNA序列受微型启动子(minP)调控,微型启动子前有不同数量的 μ E5盒(μ E5盒)。(B) 在HEK293细胞中表达A中所述构建体、空载和“无DNA”对照,对其TCF4蛋白进行蛋白质印迹分析。代表性印迹n=3个重复。 β -肌动蛋白用作内参对照。(C) 在本研究合成的DNA构建体的一个示例中,TCF4-B编码序列前面的调控元件的DNA序列包括12个 μ E5盒和微型启动子minP。

[0018] 图3A-B显示了将DNA构建体引入患者来源的神经祖细胞后,TCF4及其靶基因之一的表达增加。(A) 根据图2C中概述的细节,用含有12个或6个 μ E5盒的慢病毒颗粒(病毒载体)转染对照细胞系(橙色条形)和PTHS来源细胞系(蓝色条形)后,TCF4基因的相对表达水平。

实验使用了两个患者系和两个对照。 n = 每个处理有3个生物学重复;每个生物学重复有3个技术重复。(B)与A相似,但是对于GADD45G基因,其为人类祖细胞中已知的TCF4靶标之一。用灰色条形代表用病毒颗粒转染的对照系,以突出显示橙色条形(正常水平的GADD45G表达)和蓝色条形(遗传操作前后患病细胞中的表达)之间的比较。

[0019] 图4A-E提供了本公开的示例性TCF4盒(SEQ ID NO:4-8)。

[0020] 图5A-E显示了PTHS类器官的神经祖细胞和皮质神经元表现出异常发育和含量改变。(A)体外培养超过4周的来自对照组(亲本)和PTHS个体的皮质类器官(Ct0)的亮视野显微镜图像。(B)左:在体外培养4周时,4对亲子的Ct0大小分布(有关本研究涉及的受试者的描述,请参见表1)。右:4周时的平均Ct0大小。 N = 每组4名受试者(根据表1中用不同符号表示的图例),每个受试者12-30个类器官。(C)体外培养超过4周的对照和PTHS的皮质下类器官(subpallial organoids, sP0)的显微镜图像。由于类器官的尺寸很大,4周的图像直接在直径为3.5cm的平板中拍摄。(D)在两个发育阶段对SOX2+细胞含量进行定量。 N = 4个受试者(符号),每个受试者3个批次,每批6个类器官,每个类器官4个随机 $100 \times 100 \mu\text{m}$ 关注区域(ROI)。图12F为sP0中SOX2+细胞的定量。(E)在Ct0发育的两个阶段表达CTIP2的皮质神经元含量的定量。 N = 4个受试者(符号),每个受试者3个批次,每批6个类器官,每个类器官4个随机ROI。图12G为Ct0中SATB2+细胞的定量。

[0021] 图6A-J显示了PTHS类器官中神经祖细胞的百分比增加,兴奋性皮质神经元和抑制性中间神经元的百分比降低。(A)统一流形逼近和投影(Uniform Manifold Approximation and Projection, UMAP)二维还原亲本对照和PTHS类器官的Ct0和sP0的单细胞RNA-Seq转录组学分析。颜色代码代表6个带注释的亚群:Pr-Glut,谷氨酸能谱系中的神经祖细胞;IP-Glut,谷氨酸能谱系中的中间祖细胞;N-Glut,谷氨酸能神经元;Pr-GABA,抑制性谱系中的神经祖细胞;IP-GABA,抑制性谱系中的中间祖细胞;N-GABA,GABA能中间神经元(另见图13A)。未显示其他细胞类型。(B)轨迹分析表明Ct0和sP0中存在单独的细胞分化谱系。颜色图表示每个谱系(谷氨酸能或GABA能)沿拟时间的进展。(C)亲本和PTHS Ct0之间不同细胞类型含量的比较。颜色代码与A中相同。黑点代表A中未描述的其他群体中的细胞。(D)Ct0中每个亚群中细胞类型百分比的定量(颜色代码与A中相同)。(E)左:Ct0中Pr-Glut亚群的SOX2表达水平;每个点代表一个细胞。右:表达SOX2的细胞百分比(高于阈值等于平均值的40%)。(F)亲本和PTHS sP0之间的细胞群体比较。(G)sP0中每个亚群中细胞类型的百分比的定量。(H)左:sP0中Pr-GABA亚群的SOX2表达水平。右:SOX2+细胞的百分比。(I,J)左:Ct0的N-Glut亚群(I)或sP0的N-GABA亚群(J)中CTIP2和SATB2(I)或GAD2(J)的表达。右:PTHS Ct0和sP0中神经元亚型百分比严重降低(表达高于阈值,相当于平均值的40%)。

[0022] 图7A-H显示了PTHS神经元表现出异常的电生理特性和基因表达程序。(A)左:Ct0接种在多电极阵列(MEA)板上。右:Ct0的平均放电率。 N = 每组4个受试者(符号),每个受试者3个独立重复。(B)对二维培养条件下的iPSC来源神经元进行MAP2免疫染色(白色)。(C)亲本和PTHS来源神经元之间轴突长度和胞体面积的比较。中位数由彩色线表示。 N = 每行30-80个神经元(点),来自亲子对#1和#4的受试者。(D)膜片钳电生理探究iPSC来源神经元,显示了PTHS神经元的脉冲率(顶部)和内在兴奋性(底部)降低。 N = 10(亲本)或9(PTHS)个神经元。(E)PTHS(蓝线)和亲本对照(橙线)之间的钠和钾电流测量比较。 N = 10(亲本)或9(PTHS)个神经元。(F)左:Ct0的N-Glut神经元中FOS的表达;小提琴图表示每个群体的基因表达分

布;N=1401(亲本)和380(PTHS)个细胞。右:FOS+细胞的百分比。(G) iPSC来源神经元中特定神经元基因的相对表达(RT-qPCR)。N=每组4个受试者(符号),每个受试者3个独立重复,每个样本2个技术重复。在对照组中,平均基因表达标准化为1。STMN2, stathmin 2;TAC1,速激肽前体1(tachykinin precursor);CNTN2,接触蛋白1;INA,连接蛋白神经元中间丝蛋白 α ;ADCYAP1,腺苷酸环化酶激活多肽1;SYT13,突触结合蛋白13;和SLC17A6,囊泡型谷氨酸转运蛋白2(VGLUT2)。(H) sP0的N-GABA神经元的单细胞转录组学数据中与G中相同基因的表达。样本数量与F中的样本数量相同。条形图表示平均值+SEM。n.s.为无统计学显著性差异,* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$;Kruskal-Wallis H检验(F和H)、Welch's t检验(A)、重复测量使用带有Greenhouse-Geisser校正的方差分析后进行LSD事后检验(C),或方差分析后进行HSD事后检验(D和E)。比例尺为100 μ m。在F和H中,统计比较是每个基因的基因表达平均值之间的比较。

[0023] 图8A-J显示了PTHS的神经祖细胞以较低的速率增殖。(A)左,第2周的Ct0的SOX2和MAP2的免疫染色。箭头标记玫瑰结示例。中间,图表显示第2周亲本和PTHS类器官中的玫瑰结数量。右,类器官中SOX2+细胞的密度。N=每组4个受试者(符号),3个技术重复。(B)左: NPC的增长曲线;线条表示平均细胞数;N=每个时间点(圆圈)3个独立重复,3个技术重复。右:培养4天后神经祖细胞的相对活细胞计数(起始细胞数=100,000)。N=每组4名受试者(符号),3个独立的生物学重复,3个技术重复。(C) NPC中膜联蛋白V阳性(凋亡)细胞的定量。N=每组4个受试者(符号),每个受试者3个独立重复。(D) NPC增殖测定中的活细胞计数。N=每组4个受试者(符号),每个受试者3个独立重复。(E)左:EdU阳性(分裂)NPC的流式细胞术评估。右:EdU+细胞的百分比;N=每组3个受试者(符号),每个受试者3个独立重复,6个技术重复。(F) PTHS NPC的形态异常,显示出扁平增大的细胞(箭头)。(G)左: NPC中衰老相关 β -半乳糖苷酶(SA- β -gal)活性(绿色荧光)的染色。定量显示在右侧。N=每组4个受试者(符号),每个受试者3个独立重复。(H) NPC中CDKN2A(细胞周期依赖性激酶抑制剂2A;左)和LMNB1(核纤层蛋白B1;右)的相对表达。N=每组4个受试者(符号),每个受试者3个独立重复,2个技术重复。(I) CDKN2A在死后PTHS皮质样品中相对表达。(J) Ct0中p16^{INK4a}(左)和凋亡(裂解半胱天冬酶3,CC3+;右)细胞的定量。N=每组4个受试者(符号),2个批次,每批6个类器官,每个类器官4个100 $\times$ 100 μ m ROI。所有条形图表示平均值+SEM。n.s.=无统计学意义差异,* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$;方差分析(在C的左侧)或其余比较为Welch's t检验。H中,对照的平均基因表达标准化为1。DAPI将核染为蓝色。比例尺为100 μ m。

[0024] 图9A-K示出了Wnt信号通路的调控挽救PTHS神经祖细胞的异常增殖。(A) 亲本和PTHS NPC间Wnt信号通路基因表达丰度(每百万分之一的转录本,TPM)的比率。N=4对亲子(符号)。(B) NPC中特定的Wnt信号基因的相对表达。N=每组4个受试者(符号),每个受试者3个独立重复,2个技术重复。(C) PTHS NPC中的Wnt信号活性降低(TOP-Flash测定)。N=每组4个受试者(符号),每个受试者3个独立重复。“亲本”组中平均活性(任意单位)标准化为1。(D) 死后PTHS皮质样本中特定Wnt基因的相对表达。(E) 向对照NPC施用Wnt通路拮抗剂DKK-1和ICG-001(黄色条型),PTHS祖细胞表型增殖缺陷。N=每组3个重复(点);表1中细胞取自亲子对#4。(F) ICG-001处理后,PTHS类器官显示低神经祖细胞含量(SOX2)的表型。N=3个独立的实验(点);每组每个实验评估12个类器官;每个类器官4个随机100 $\times$ 100 μ m ROI。(G) 显示用Wnt通路激动剂CHIR99021处理NPC的活细胞计数。N=每组3个受试者(符号),每个受试者

3个独立重复,3个技术重复。(H)用CHIR99021处理的NPC的EdU增殖测定。左图表示亲子对#4的数据,右图显示亲子对#1的数据。 $N=3$ 次技术重复。(I)用CHIR99021处理的NPC中 $p16^{INK4a+}$ (衰老)细胞的定量。数据针对亲子对#4(另请参阅图16E中对#1)。(J)用CHIR99021处理PTHS NPC挽救了几个祖细胞基因的表达。 $N=$ 每组4个受试者(符号),每个受试者3个独立重复,2个技术重复。(K)用CHIR99021处理PTHS Ct0后SOX2+细胞的定量。 $N=3$ 个实验(点);每组实验6个类器官;每个类器官4个随机 $100 \times 100 \mu\text{m}$ ROI。条形图表示平均值+SEM。*n.s.*,无统计学意义差异;** $p<0.01$,*** $p<0.001$;对其余组进行Welch's *t*检验(B,C和D)或方差分析。在B、D和J中,对照基因的平均表达为1。在K中,统计比较PTHS+CHIR组和PTHS+DMSO(对照)组之间的平均值。比例尺为 $100 \mu\text{m}$ 。

[0025] 图10A-L显示了SOX基因在PTHS细胞病理生理学中的参与机制。(A) NPC中几个SOX基因的表达丰度(TPM)比。 $N=4$ 个亲子对(符号)。条形上方的括号显示SOX基因类别。(B)顶部: NPC中SOX3的表达丰度(TPM);底部:通过qPCR检测SOX3的相对表达。 $N=$ 每组4个受试者(符号),每个受试者3个独立重复,2个技术重复。(C) NPC(表1中来自亲本#4的细胞)中shRNA介导的TCF4敲低后,SOX3下调。 $N=$ 每组3个独立重复(点),2个技术重复。(D)死后PTHS皮质样品中SOX3的相对表达。(E)死后PTHS样本中SOX3的免疫染色(每个基因型显示两个ROI)。(F)用CHIR99021处理PTHS NPC挽救异常SOX3表达。 $N=$ 每组4个受试者(符号),每个受试者3个独立重复,2个技术重复。(G) shRNA介导的SOX3敲低降低祖细胞增殖。 $N=$ 每组3个独立重复(点),3个技术重复。(H)PTHS NPC中的SOX4表达降低。顶部:表达丰度(TPM)。底部:相对表达。 $N=$ 每组4个受试者(符号),每个受试者3个独立重复,2个技术重复。(I)PTHS Ct0的中间祖细胞(IP-Glut)和神经元(N-Glut)中SOX4的表达降低。 $N=717$ 个(亲本)和382个(PTHS) IP-Glut细胞,或1401个(亲本)和380个(PTHS) N-Glut神经元。(J)神经元(MAP2+)和NPC(SOX2+)之间的比率代表神经元分化率。 $N=$ 每组4个受试者(符号)。(K)SOX4敲低后MAP2+和SOX2+之间的比率。 $N=$ 每个基因型3-4个ROI;细胞来自亲子对#1和#4。(L)SOX4敲低后MAP2和SOX2基因表达水平之间的比率。 $N=4$ 个生物学重复;细胞来自亲子对#4。条形图表示平均值+SEM。 $*p<0.05$;** $p<0.01$;*** $p<0.001$;Kruskal-Wallis H检验(I)或方差分析(F和G)和其余组中Welch's *t*检验。比例尺为 $100 \mu\text{m}$ 。用DAPI将核染成蓝色。

[0026] 图11A-H显示了对PTHS类器官进行TCF4表达的遗传矫正,逆转了异常表型。(A)基于CRISPR对TCF4表达进行反-表观遗传(trans-epigenetic)矫正的示意图,使用包含向导RNA(gRNA)、转录激活模块MPH和死亡Cas9的构建体。(B)顶部:病毒应用方案。底部:对PTHS脑类器官进行TCF4表达校正(PTHS+TCF4 gRNA)的亮视野图像,与转导了干扰gRNA(scr gRNA)的对照对比。(C)TCF4免疫染色后转导的类器官的荧光显微镜图像;C':簇状TCF4+细胞(箭头)异常生长。(D)在两个发育时间点对转导的类器官中MAP2和SOX2染色。箭头:scr gRNA PTHS中的异常生长。高放大倍率插图(inset):类器官生长中簇状异常的MAP2+细胞。(E)TCF4-B编码序列置于由微型启动子minP和不同数量的TCF4结合位点(μE5 盒)组成的合成启动子控制下的过表达构建体。(F)顶部:病毒应用方案。底部:显微镜图像显示了转导TCF4过表达载体或空载体(ctrl) (顶部线)的6周龄类器官的一般形态,对SOX2和MAP2(中间线),以及CTIP2(底部线)进行免疫染色。箭头,神经玫瑰结。(G)左,光栅图显示通过多电极阵列(MEA)分析转导的2至3月龄的类器官的电活动。每行代表一个电极。垂直红色矩形表示在网络级别发生的电活动放电事件(网络放电)。右,随着时间的推移,转导类器官的平均放电

率(右上)的量化(光栅图另见图18L)和网络放电次数(下图)。(H) 上图:病毒应用方案。下图:TCF4(顶部线)、SOX2和MAP2(中间)以及CTIP2(底部)的免疫染色。箭头,神经玫瑰结。条形图表示平均值+SEM。n.s.,无统计学意义差异,* $p<0.05$;** $p<0.01$;*** $p<0.001$;G中每个时间点使用Welch's t检验。比例尺为100 μ m。用DAPI将核染为蓝色。

[0027] 图12A-M显示了PTHS iPSC表现出正常的生长速率,并且可以分化成神经元。(A) 不同患者TCF4外显子(每个矩形顶部的数字)的结构。白色矩形代表由于部分或整个基因删除而外显子缺失。带有粗边框的矩形表示每种情况的编码序列。AD1至AD3,转录激活结构域;bHLH,碱性螺旋-环-螺旋DNA结合域。显示了外显子1和2,但它们不是TCF4基因的主要转录本(称为TCF4-B)的一部分。每个患者带有的突变类型的详细信息在右侧给出(更多信息另见表1)。(B) 通过免疫共沉淀杂交进行SNP图谱的数字核型分析示例,样本为患者PTHS#2(表1),显示了18号染色体上的大量缺失(星号)。左边的数字代表染色体。每个染色体图中的y轴代表单个杂交芯片探针(点)的logR比值。(C) 比较亲本(对照组)、PTHS和对照iPSC系(取自本研究之外的受试者)(WT83)之间的iPSC克隆生长速率(达到传代所需的大小的天数,即2mm)。N=5个受试者(符号)。每次测量是5个独立平板的中位数。(D) 显示培养iPSC来源的神经元的代表性亮视野显微镜图像。注意亲本和PTHS组的祖细胞和神经元细胞体中束的形成。(E) 对D的培养物中MAP2(红色)和SOX2(绿色)染色的荧光显微镜图像,显示了两组中iPSC分化成NPC(SOX2+)和神经元(MAP2+)的能力。(F) 左,来源于iPSC的PTHS和亲本对照Ct0的两个不同批次(克隆)的亮视野图像。中,类器官尺寸分布。右,SOX2和MAP2的免疫染色。箭头指向患者系中的小玫瑰结。(G) 亲本和PTHS的4周和10周Ct0中SOX2+细胞的基因型的百分比。N=4个受试者(符号)。(H) 体外培养6周后,对亲本和PTHS sP0中NPC标志物SOX2(绿色)和神经元标志物MAP2(红色)进行免疫染色的荧光显微镜图像。(I) 体外培养6周时,亲本和PTHS sP0中SOX2+细胞的密度定量。N=4个受试者(符号),每个受试者2个批次,每批6个类器官,每个类器官随机4个100 $\times$ 100 μ m关注区域(ROI)。(J) 在体外发育的两个阶段,对亲本和PTHS Ct0中表达SATB2的皮质神经元的密度定量。N=4个受试者(符号),每个受试者3个批次,每批6个类器官,每个类器官随机4个ROI。(K) 死后PTHS皮质样本中神经标志物的相对表达。(L) 死后样本中CTIP2+细胞百分比的定量。(M) vGLUT1和GAD65/67表达的定量,根据PTHS和对照Ct0和sP0中每单位面积阈值强度以上的像素数判断。N=每个条件(圆圈)6个部分,每个部分4个ROI。条形图表示平均值+SEM。n.s.,无统计学意义差异,** $p<0.01$,*** $p<0.001$;双样本Welch's t检验(D、G、I和K)、单因素方差分析后针对各自亲本组进行Tukey-Kramer's HSD事后检验(G),或Wilcoxon-Mann-Whitney U检验(M)。比例尺为100 μ m。

[0028] 图13A-I显示了亚群和相应对照的单细胞RNA-Seq实验的注释。(A) 点图显示了图6A中所示六个细胞亚群中特定标记基因的表达。Pr-Glut,谷氨酸能谱系中的神经祖细胞;IP-Glut,谷氨酸能谱系中的中间祖细胞;N-Glut,谷氨酸能神经元;Pr-GABA,抑制性谱系中的神经祖细胞;IP-GABA,抑制性谱系中的中间祖细胞;N-GABA,GABA能中间神经元。“其他”代表未包含在前六类中的异质细胞组。点大小是每个亚群中具有可检测表达相应基因的细胞的百分比。(B) A中显示的标记物基因的小提琴图,显示了六个分析的细胞亚群(和“其他”)的表达范围。颜色代码与图2A中相同。GRIN2、GAD1和GAD2的中位数较低,因此每个细胞中的表达表示为点。(C) 单细胞RNA-Seq质量控制数据。小提琴图表示A中列出的亚群的读数计数、检测到的基因数(特征)和线粒体基因(mtRNA)的百分比。颜色代码与图6A中相同,“其

他”以黑色显示。(D) UMAP显示了TCF4、神经谱系标志物SOX2和MAP2、中胚层标志物MIXL1和TBXT (Brachyury) 以及内胚层标志物CFTR和SOX17的表达。紫色的强度表示相对表达水平。(E) “其他”亚群中未分配细胞的特征。左上, 亲本和PTHS的Ct0和sP0亚群中“其他”的细胞和线粒体RNA含量的百分比。右上, 亲本(橙色) 和PTHS(蓝色) 类器官的Ct0和sP0中神经谱系标志物SOX2和MAP2的表达水平。底部, 亲本(每组左) 和PTHS(每组右) 基因型的Ct0和sP0中, D中所示中胚层和内胚层标志物的表达水平。(F) 左, UMAP显示了亲本和PTHS Ct0中星形胶质细胞标志物S100B和ALDH1L1的表达。右, 亲本和PTHS的Ct0中相同基因表达水平的比较。(G) 对照显示了独立批次中Ct0形成的可重复性(重复#1和#2; 左) 和单细胞RNA-Seq分析的稳定性, 这些分析与在独立批次的亲本来源类器官(右) 中检测到的每种细胞类型的百分比相似。颜色代码与图6A相同。(H, I) 左: 比较亲本和PTHS的Ct0的N-Glut亚群中编码皮质神经元亚型TBR1 (H) 和CUX (I) 标志物基因的表达。右: PTHS Ct0中神经元亚型的百分比严重降低(表达每个标记基因的细胞超过阈值, 相当于相应平均值的40%)。n.s., 无统计学意义差异, Wilcoxon-Mann-Whitney U检验(F)。

[0029] 图14A-H显示了用于研究2D培养物的神经元的支持数据。(A) 进行多电极阵列(MEA) 测定的Ct0中平均放电速率的时间过程, 显示了亲本(橙色) 和PTHS(蓝色) 类器官之间的比较。每个患者用不同的符号表示, 如表1所示。红线和误差线代表一段时间内所有受试者的平均值。N=4个受试者(符号), 每个受试者3个独立重复。(B) 亲本来源的iPSC经2D培养分化, 其神经元(MAP2标记以红色显示) 中TCF4蛋白(绿色) 表达的代表性荧光显微镜图像。(C) iPSC来源的PTHS与亲本神经元培养之间的TCF4的相对表达水平(RT-qPCR)。N=4个受试者(符号), 每个受试者3个独立重复, 每个样本2个技术重复。将每个PTHS样本与相应的亲本进行比较(表达标准化为1)。(D) 通过膜片钳电生理分析比较2D培养中PTHS(蓝色圆圈) 和亲本对照神经元(圆圈) 之间的膜电容。(E) 探究钠(顶部) 和钾(底部) 电流密度, 比较PTHS(蓝色线) 和对照(橙色线) 神经元。N=10(亲本) 或9(PTHS) 个神经元。(F) 热图显示了来自一个亲本和对应PTHS子代(表1中的PTHS#4) 的神经元的RNA-Seq文库中20,000个表达最高的基因的表达水平。亲子比较中差异表达(DE) 基因的数量如图上方所示。(G) D中下调的DE基因的基因本体论-生物过程(顶部) 和通路分析(底部) 的点图结果。对于每个分析, 显示了调整后p值的前10个类别。点大小代表属于每个分类类别的DE基因的数量, 点的颜色是调整后的p值, x轴表示RNA-Seq文库中每个类别中DE表达基因的基因百分比。注意存在参与谷氨酸能和GABA能传递的基因。(H) 对照和PTHS神经元中表达的基因的MA图。灰点是对照组和PTHS之间没有统计学上显著差异表达的基因。蓝点和红点是具有统计学意义的DE基因。红点代表编码钠或钾通道的基因, 其 $\log_2$ 倍数变化高于4(虚线)。条形图表示平均值+SEM。* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Welch's t检验(A和D) 或方差分析后进行HSD事后检验(E)。比例尺为100 μm 。

[0030] 图15A-N显示了神经祖细胞中的表达分析。(A) 通过RNA-Seq文库中TPM表达丰度判断亲本和PTHS受试者的iPSC来源的神经祖细胞中几种NPC标记物表达的验证。N=每组4个受试者(符号), 每个受试者3个独立重复。(B) iPSC和iPSC来源的NPC和神经元培养物中的TCF4的相对表达水平(RT-qPCR), 以及其在非神经细胞系(HEK293T) 中的表达。N=每组4个独立重复, 每个样品2个技术重复。神经祖细胞组平均表达标准化为1。(C) 2D培养下NPC中TCF4(红色) 和NPC标志物巢蛋白(绿色) 免疫染色后的代表性荧光显微镜图像。插图中的放

大倍率更高。(D) 对照(亲本) Ct0的玫瑰结的NPC中TCF4(红色) 丰富表达的代表性荧光显微镜图像。(E) 亲本来源和相应PTHS子代的NPC中TCF4的相对表达(RT-qPCR)。N=每组5名受试者(符号;表1中的受试者PTHS#1至#5), 每个受试者3个独立重复, 每个样本2个技术重复。注意, 一个PTHS受试者(圆圈符号) 的表达没有显著减少, 这是意料之外的, 因为该受试者(PTHS#4) 携带的点突变预计不会影响转录本丰度。所有其他受试者都有预期会降低转录本含量的突变(无义突变、移码突变、全基因缺失和易位)。每个PTHS受试者的平均表达量显示为相对于其各自的亲本(所有亲本均值标准化为1)。(F) 亲本来源和对应PTHS子代的NPC中GADD45G的相对表达。N=每组5个受试者(符号), 每个受试者3个独立重复, 每个样本2个技术重复。(G) 更高放大倍率的荧光图像显示PTHS中TCF4的表达减少(受试者PTHS#2) 以及TCF4蛋白在细胞核(箭头) 外的可能错误定位。(H) PTHS和对照样品之间复制衰老的转录组标志物之间表达的比率。细胞来自亲子对#4(表1)。对于每个基因, 每次传代的平均值由3个独立的生物样品确定。每条线连接早期和晚期传代条件下某个基因的表达。标记基因根据类别(在衰老细胞中下调或上调) 分开。注意, 晚期传代情况下错误调控更明显。亲子对#1也得到了类似的结果。(I) 表达衰老标志物p16^{INK4a}的细胞的百分比定量, 这些细胞也表达神经谱系标志物巢蛋白、祖细胞标志物SOX2、中胚层标志物Brachyury和内胚层标志物SOX17。对于SOX2, 一些细胞表现出强染色, 而另一些则表现出弱染色。(J) 左, Ct0中对SOX2、MAP2和p16^{INK4a}染色的高放大倍率图像, 然后定量共表达p16^{INK4a}与SOX2或MAP2的细胞(右)。(K) 左: shRNA介导的TCF4敲低降低了NPC增殖。N=每组3个独立重复(点), 每个样品3个技术重复。初始接种密度为 1×10^5 个细胞。右: 同一组中EdU-阳性NPC的百分比量化; N=每组3个独立重复(点)。(L) shRNA介导的TCF4敲低导致TCF4和TCF4下游靶基因GADD45G的表达降低, 以及衰老标志物CDKN2A的表达增加。N=每组3个独立重复(点), 每个样品2个技术重复。(M) 热图显示了源自2个亲本和相应PTHS子代的NPC的RNA-Seq文库中20,000个最高表达的基因的表达水平。所有四个亲子比较中差异表达(DE) 基因的交集数量如图的上方所示。(N) M中列出的DE下调基因的基因本体论-生物过程(顶部) 和通路分析(底部) 的点图结果, 所有4个亲子对之间交集的结果。对于每个分析, 将显示调整后p值前10的类别。点的大小表示属于每个分类类别DE基因的数量, 点的颜色是调整后的p值, x轴表示RNA-Seq文库中每个类别中是DE表达基因的基因的百分比。注意Wnt信号通路中存在下调基因。条形图表示平均值+SEM。* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; 假设方差不齐的双样本Welch' st检验(E、F和K) 或方差分析后进行Tukey-Kramer HSD事后检验(K)。在L中, 每个亲本+对照shRNA组的平均基因表达标准化为1。比例尺为100 μ m。

[0031] 图16A-N显示了Wnt信号调控类器官的其他控制。(A) 比较亲本(橙色) 和PTHS(蓝色) NPC中Wnt信号通路中特定基因的表达丰度。N=每组4个受试者(符号), 每个受试者3个独立重复。(B) 用DKK-1处理亲本来源的NPC增加衰老标志物CDKN2A的表达。N=4个生物学重复(符号)。(C) 用Wnt通路拮抗剂ICG-001(黄色条形) 处理对照Ct0, 表型模拟了PTHS类器官中的小类器官尺寸。N=3个独立重复(点), 每组每个实验12-30个测量的类器官。(D) 通过TOP-Flash功能报告基因测定测量, 确定用激动剂CHIR99021处理NPC后Wnt信号活性增加。N=每组4个受试者(符号), 每个受试者3个独立重复。在“亲本+DMSO”组中, 平均活性(任意单位) 标准化为1。(E) 用中胚层标记物Brachyury和衰老标记物p16^{INK4a}(顶行)、内胚层标记物SOX17或TCF4(中行) 以及神经谱系标记物巢蛋白和SOX2(底行) 染色亲本和PTHS受试者的

NPC后,用CHIR99021(或DMSO作为对照)处理的荧光显微镜图像。高放大率插图中的箭头表示共定位。(F)用Wnt通路激动剂CHIR99021(浅蓝色条形)处理PTHS NPC,挽救了类器官尺寸。 $N=3$ 个独立重复(点),每组每个实验15-20个测量的类器官。(G和H)单细胞RNA-Seq揭示了用Wnt激动剂CHIR99021处理sP0(G)和Ct0(H)会增加NPC的量。左:UMAP表示细胞多样性的比较,如图6A显示6个细胞亚群。Pr-Glut,谷氨酸能谱系中的神经祖细胞;IP-Glut,谷氨酸能谱系中的中间祖细胞;N-Glut,谷氨酸能神经元;Pr-GABA,抑制性谱系中的神经祖细胞;IP-GABA,抑制性谱系中的中间祖细胞;N-GABA,GABA能中间神经元。未显示其他细胞类型。右:神经祖细胞和神经元百分比的量化。(I)用CHIR99021处理的Ct0类器官中GAD1和GAD2的相对表达。 $N=$ 亲子对#4的每个条件的3个生物学重复。(J)上:UMAP显示用CHIR99021处理的PTHS sP0(右)与未处理的PTHS sP0(左)中TCF4的表达的比较。紫色的强度表示相对表达水平。下:小提琴图显示了在未经处理和CHIR处理的PTHS sP0的单细胞(点)中TCF4的表达水平。(K)用CHIR99021处理PTHS NPC增加TCF4和TCF4下游靶基因GADD45G的表达。 $N=$ 每组4个受试者(符号),每个受试者3个独立重复。“亲本+DMSO”组中每个基因的平均表达水平标准化为1。(L)对4周龄的Ct0进行 β -连环蛋白的免疫染色,对照类器官中定位于玫瑰结中心,但在PTHS Ct0(箭头)中染色紊乱。(M)CTNNB1(β -连环蛋白)在NPC(左上)或Ct0中兴奋性谱系祖细胞(右上)中的表达水平。(N)编码钙粘蛋白23(CDH23)和原钙粘蛋白15(PCDH15)的基因的表达,它们是亲本和PTHS的NPC之间的DE基因(左)。右图表示用Wnt激动剂CHIR99021处理的NPC的表达水平。条形图表示平均值+SEM。 $n.s.$,无统计学意义差异, $*p<0.05$, $**p<0.01$, $***p<0.001$;Welch's t 检验M和N(左图)和方差分析后进行HSD事后检验。在B、I、K和N(右图)中,每个对照组的平均基因表达标准化为1。在D中,亲本+DMSO组中的Wnt信号活性设置为1。在K中,亲本+DMSO组中每个基因的平均表达量设置为1。在K中,统计比较是PTHS+CHIR和PTHS+DMSO(对照)组之间的平均值。

[0032] 图17A-L显示了中间祖细胞在PTHS类器官中的丰度较低。(A)小提琴图显示Ct0和sP0细胞亚群中SOX1、SOX3、SOX4和SOX11的表达。SOX2表达见图13B。Pr-Glut,谷氨酸能谱系中的神经祖细胞;IP-Glut,谷氨酸能谱系中的中间祖细胞;N-Glut,谷氨酸能神经元;Pr-GABA,抑制性谱系中的神经祖细胞;IP-GABA,抑制性谱系中的中间祖细胞;N-GABA,GABA能中间神经元。(B和C)SOX1(B)和SOX3(C)在Ct0中的祖细胞(Pr-Glut)和中间祖细胞(IP-Glut)以及sP0中的祖细胞(Pr-GABA)和中间祖细胞(IP-GABA)中的表达。每个点代表一个细胞;小提琴图表示每个群体中基因表达的分布; $N=$ 亲本和PTHS Ct0组中分别959和1230个Pr-Glut细胞; $N=$ 亲本和PTHS Ct0组中分别717和382个IP-Glut细胞; $N=$ 亲本和PTHS sP0组中分别346和1376个Pr-GABA细胞; $N=$ 亲本和PTHS sP0组中分别2737和105个IP-GABA细胞。(D)shRNA介导的SOX3敲低后亲本对照NPC的2D培养物中,SOX3(RT-qPCR)的相对表达。 $N=3$ 个独立重复(点),每个样品2个技术重复。存在表达较低的趋势,低于显著性($P\approx 0.50$),但处于预期方向。亲本+对照shRNA组的平均表达标准化为1。使用的NPC来自亲本系#4(表1)。(E)shRNA介导的SOX3敲低后亲本对照NPC的2D培养物中,CDKN2A、ASCL1和HES1(RT-qPCR)的相对表达。 $N=3$ 个独立重复(点),每个样品2个技术重复。使用的NPC来自亲本系#4(表1)。(F)SOX3过表达后亲本和PTHS的NPC的活细胞计数。 $N=3$ 个生物学重复,来自亲子对#4的细胞。(G)SOX3过表达后SOX3的相对表达。 $N=3$ 个生物学重复,来自亲子对#4的细胞。(H)SOX4在PTHS sP0的中间祖细胞(IP-GABA)和神经元(N-GABA)中表达正常。每个点代表一个

细胞;小提琴图表示每个群体中基因表达的分布;N=亲本和PTHS组分别2737和105个IP-GABA细胞;N=亲本和PTHS组分别2661和988个N-GABA神经元。尽管如Ct0所示,每个细胞的表达没有改变(图10G),但PTHS类器官中中间祖细胞和神经元的数量明显减少。小提琴图的颜色代码与B相同。(I) PTHS Ct0和sP0中的中间祖细胞(IP-Glut和IP-GABA)和神经元(N-Glut和N-GABA)中SOX11表达正常。每个点代表一个细胞;小提琴图表示每个群体中基因表达的分布;N=亲本和PTHS Ct0组分别717和382个IP-Glut细胞;N=亲本和PTHS Ct0组分别1401和380个N-Glut神经元;N=亲本和PTHS sP0组分别2737和105个IP-GABA细胞;N=亲本和PTHS sP0组分别2661和988个N-GABA神经元;N=亲本和PTHS Ct0组分别717和382个IP-Glut细胞。小提琴图的颜色代码与B中的相同。(J) 量化来自亲本对照和PTHS受试者的分化神经元的2D培养物中MAP2+细胞的百分比。N=每组4个受试者(符号),2个独立分化实验,每个受试者3个独立重复,每个独立重复4个随机选择的视野。(K) 单细胞RNA-Seq的UMAP表示PTHS和亲本对照Ct0和sP0的结果,在每个图中以红色突出显示中间祖细胞。IPs的百分比显示在每个象限的左下角。(L) 左:小提琴图显示Ct0和sP0的IPs和神经元中POU3F2(编码BRN2,以中间祖细胞表达)的表达。每个点代表一个细胞。右:从单细胞RNA-Seq数据中的细胞群体定量判断,PTHS Ct0和sP0中中间祖细胞的百分比严重降低。小提琴图的颜色代码与B中的相同。条形图表示平均值+SEM.n.s.,无统计学意义差异,* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$;假设方差不齐(E和J)的Welch's t检验,Kruskal-Wallis H检验用于比较PTHS与对应亲本中的基因表达(B、C、H、I和L中的左图)。在D中,比较亲本组 and 对应PTHS组之间,在极限p值为0.5时具有统计学意义。†表示PTHS均值在统计意义上与亲本均值显著不同,但倍数变化小于10%。

[0033] 图18A-0提供了TCF4表达的遗传矫正的详细信息。(A) 亲本和PTHS样品(第一行)中TCF4位点(红色条形)替代启动子的标准化转录活性。第二行描绘了TCF4位点的示意图,显示了其外显子的位置。显示了三个选定的TCF4替代启动子的设计的gRNA(蓝色箭头)的位置,位于外显子3b、8a和10a的上游。其余行显示了由外显子3b、8a和10a启动的转录形成的转录本,它们分别产生TCF4蛋白异构体TCF4-B、TCF4-D和TCF4-A。(B) 测试五种TCF4 gRNA对SH-S5Y5细胞中TCF4表达的反式激活效率。scr gRNA,对照干扰gRNA;无gRNA,空表达构建体。N=4个独立重复,3个技术重复。(C) 左:2D神经元培养物中CNTNAP2的相对表达。N=3(对照)或4(PTHS)受试者(符号),每个受试者3个独立重复,2个技术重复。下:反表观遗传学TCF4表达校正增加SH-S5Y5细胞中的CNTNAP2。N=4个独立重复,3个技术重复。右:TCF4靶基因KCNQ1在PTHS神经元中的相对表达(RT-qPCR)大大增加。N=每组4个受试者(符号),每个受试者3个独立重复,每个样本2个技术重复。下:TCF4表达校正降低转染SH-S5Y5细胞中的KCNQ1表达。(D) 顶部,TCF4校正后TCF4表达水平升高。底部,正常(编码序列第959位为C)和突变(该位为T)TCF4等位基因的表达丰度之比。N=每组3个独立重复(点),每个样品10个混合类器官。(E) TCF4校正后GADD45G(TCF4下游靶基因)、CDKN2A、SOX3和MAP2的表达水平。N=每组3个独立重复(点),每个样品10个混合类器官。(F) TCF4校正后DCX表达水平的校正。N=每组3个独立重复(点),每个样品10个混合类器官。(G) DCX在PTHS死后皮质组织中的相对表达。(H) 在PTHS Ct0和sP0中的中间祖细胞(IP-Glut和IP-GABA)和神经元(N-Glut和N-GABA)中DCX表达降低。N=717(亲本)和382(PTHS)个IP-Glut细胞、1401(亲本)和380(PTHS)个N-Glut神经元、2737(亲本)和105(PTHS)个IP-GABA细胞或2661(亲本)和988(PTHS)个N-GABA

神经元。(I)根据RNA-Seq数据判断,与来自亲本对照的神经元相比,PTHS神经元中DCX的表达丰度(以百万分之一的转录本或TPM为单位)显著降低。 $N=$ 每组3个独立重复(点)。(J)使用含有编码TCF4-B亚型的cDNA的过表达盒(顶部)验证TCF4表达的校正,前面是含有微E5(μ E5)调节结合位点的合成启动子,该启动子允许其在表达TCF4的细胞中过表达,从而防止异位表达。将该构建体的多个版本(具有6或12个 μ E5盒)转染到2D培养的PTHS NPC中,然后用RT-qPCR评估TCF4和TCF4靶基因GADD45G表达的增加。 $N=$ 每组3个独立重复。转染含有微型启动子(minP;预计不会增加TCF4水平)的过表达盒(具有6或12个 μ E5盒)和对照的细胞,对比其相对表达水平的平均值。亲本组中每个基因的表达标准化为1。(K)顶部行,在类器官衍生方案开始时用慢病毒载体载入TCF4OE的Ct0,其中TCF4和CDKN2A的相对表达。底部行,载入TCF4 OE的Ct0中SOX2和CTIP2-阳性细胞的密度。 $N=$ 每个受试者3个生物学重复,用来自亲子对#1和#4的类器官。(L)代表性光栅图,显示MEA测定亲本(左)和PTHS(右上)Ct0之间的放电活性差异,但用TCF4 OE(右下)慢病毒载体处理PTHS类器官,活性被部分挽救。(M)转导含有12个 μ E5盒(TCF4 OE)的OE AAV载体的8周龄的Ct0的低放大倍率图像。(N)顶部行,在神经诱导阶段结束后将带有AAV载体的TCF4 OE载入Ct0中,TCF4和CDKN2A的相对表达。底部行,载入TCF4 OE的Ct0中SOX2和CTIP2-阳性细胞的密度。 $N=$ 每个受试者3个生物学重复,用来自亲子对#1和#4的类器官。(O)解释PTHS神经结构中异常细胞表型的现有机制模型。由于PTHS中的TCF4单倍体不足,Wnt信号活性降低,进而导致NPC中SOX3表达降低,损害增殖。此外,SOX4在PTHS细胞中也被下调,这表明它损害了PTHS神经组织中的神经元分化和含量。scr gRNA,干扰(对照)向导RNA。条形图表示平均值+SEM。 $*p<0.05$, $**p<0.01$, $***p<0.001$; Welch's t检验(D)或方差分析后进行Tukey-Kramer's HSD事后检验(C)。

[0034] 详细说明

[0035] 如本文和随附权利要求书中使用的,除非上下文另有明确规定,单数形式“(a)”、“一个(an)”和“所述(the)”包括复数指代物。因此,例如,提及“一个启动子”包括多个这样的启动子,而提及“该构建体”包括提及一个或多个构建体,等等。

[0036] 除非另有定义,本文中使用的所有技术和科学术语与本公开所属领域的普通技术人员通常理解的含义相同。尽管与本文描述的方法和材料相似或等效的方法和材料可用于本公开的方法和组合物的实践中,但本文描述了示例性的方法、设备和材料。

[0037] 同样,除非另有说明,“或”是指“和/或”。同样,“包含(comprise、comprises、comprising)”和“包括(include、includes、including)”是可以互换的,无限制意义。

[0038] 需要进一步理解的是,在描述各种实施方案时使用术语“包含”的地方,本领域技术人员应理解,在某些特定情况下,可以替代地使用“基本上由...组成”或“由...组成”的语言来描述实施方案。

[0039] 上文和全文中讨论的公开仅供它们在本申请的申请日之前公开。本文中的任何内容均不被解释为承认任何公开是现有技术。此外,对于一个或多个公开中提出的与本公开中明确定义的术语相似或相同的任何术语,本公开中明确规定的术语定义将在所有方面具有控制权。

[0040] 转录因子4(TCF4;OMIM 602272)编码螺旋-环-螺旋转录因子,其与神经发育的若干个方面有关,包括神经发生、细胞存活、细胞周期调节、神经元分化、神经谱系定向生长和神经元兴奋性。许多替代转录本是从TCF4位点转录而来的,其中一些在大脑发育过程中高

度表达。

[0041] 许多研究发现TCF4基因变异在遗传上与一系列神经精神疾病相关,即精神分裂症、双相情感障碍、创伤后应激障碍和重度抑郁症。重要的是,TCF4的新发(*de novo*)杂合突变导致一种称为皮特-霍普金斯综合征(MIM#610954)的自闭症谱系障碍,并且类似的综合征已被证明是由TCF4下游靶基因NRXN1 β 和CNTNAP2的突变引起的。尽管有这些知识,但对TCF4突变导致神经发育和功能改变的分子和细胞机制知之甚少。

[0042] PTHS包括严重衰弱的临床症状,例如深度认知障碍、发育迟缓、全身性肌张力低下(*generalized hypotonia*)、呼吸异常、癫痫发作、言语缺乏、典型的自闭症行为、慢性便秘和独特的面部完形(*gestalt*)。大多数PTHS患者表现出独有的TCF4突变,其可能是跨越整个基因的大染色体缺失、部分基因缺失、易位、移码、无义、剪接位点或错义突变,其中大多数被认为是损害TCF4转录活性的功能缺失突变。

[0043] 已经产生了若干种携带TCF4突变的转基因小鼠系作为PTHS动物模型。其中一些小鼠系显示PTHS样表型,包括社交互动、联想记忆,感觉运动门控和胃肠道运输改变的缺陷。对这些动物脑组织的检查显示了皮质发育异常、海马和脑桥核发育过程中神经元迁移改变以及少突胶质细胞分化受损。然而,这些小鼠并没有表现出完整的临床相关症状,其包括许多最衰弱的症状,如严重的运动延迟和肌张力低下。此外,这些小鼠系中只有少数携带如在PTHS患者中发现的杂合突变。

[0044] 为了研究TCF突变,本公开使用神经祖细胞(NPC)和从患者来源的诱导多能干细胞(iPSC)体外分化的神经元,这允许在相关基因组背景下分析疾病对单个细胞类型的影响。此外,还产生了模式化的皮质类器官——包括含有兴奋性和抑制性神经元谱系的皮质和皮质下类器官。这些三维(3D)神经结构始终显示一系列不同的细胞群,并已成功用于模拟若干种疾病早期神经发育期间的细胞病理学。

[0045] 本公开表明PTHS皮质类器官在尺寸和结构上异常,含有比对照类器官更高百分比的NPC和更少的神经元。PTHS来源的NPC表现出增殖降低和分化成神经元的能力受损。值得注意的是,这些神经系统的分子探测揭示了一种病理机制,通过这种机制,TCF4中的突变导致典型Wnt/ β -catenin信号传导减少,从而导致SOX转录因子表达减少,导致细胞异常。本公开还证明了Wnt信号传导的药理学操作从基因上校正了TCF4本身的表达,导致细胞水平上神经特征的恢复。综上所述,本公开的数据揭示了相关人类细胞类型中新的细胞和分子PTHS表型,并表明这些表型是可逆的,为PTHS或其他与TCF4相关的遗传疾病患者的治疗干预提供了途径。

[0046] 与神经2D培养系统相结合的两种类型(皮质和皮质下)的患者来源的脑类器官用于研究与TCF4临床相关突变相关的病理生理学和异常分子机制。这些细胞来源于患有PTHS的儿科患者,PTHS是一种仅由TCF4突变引起的毁灭性自闭症谱系疾病。本公开表明,PTHS NPC增殖速度较慢并显示出神经元分化受损。此外,PTHS类器官表现出异常的电特性,并且含有较少的皮质神经元(图5-7)。

[0047] 本公开展示了一个模型(图180),根据该模型,低增殖性PTHS NPC的病理分子机制包括一系列分子事件,导致从TCF4功能缺失突变到Wnt信号活性降低(图8)。数据显示,在机制上,Wnt信号位于TCF4下游,其在患者细胞中以明显的级联方式失调,这一结果为疾病病理学提供了治疗干预和更好理解。PTHS样品中Wnt信号传导的药理学激活可以完全校正异

常的NPC增殖表型、类器官的形态和衰老标志物以及下游参与分子的表达(图9),其为药物治疗提供条件。

[0048] 本公开还提供了Wnt信号控制两种SOX转录因子SOX3和SOX4表达的机制证据(图10)。SOX3在PTHS NPC、类器官和死后皮质样品中表达减少,由此发现SOX3下调导致NPC增殖降低(图10)。有趣的是,SOX3的突变与另一种神经发育障碍(X-连锁智力障碍)有关,这表明这种情况与PTHS之间存在重叠的分子机制。

[0049] PTHS NPC还表现出神经元分化受损,这与螺旋-环-螺旋转录因子NEUROG1、NEUROG2和ASCL1的前神经作用一致,已知它们与TCF4相互作用。有趣的是,发现了在PTHS NPC和PTHS类器官的中间祖细胞和神经元中SOX4表达降低(图10)。已知SOX4转录因子参与神经元分化,这与PTHS类器官含有较少神经元,并且患者来源的NPC的分化速率比对照细胞慢的发现一致,在SOX4的表达被敲低的分化神经元培养物中也观察到了这种表型(图10)。据推测,TCF4单倍体剂量不足导致SOX4下调,导致神经元分化降低(图180)。

[0050] 细胞增殖和分化的缺陷是导致在PTHS类器官和在来自PTHS个体的死后脑组织中观察到皮质神经元含量较低的因素。应该注意的是,类器官和死后样本中神经元含量的降低与通过MRI检测的一些PTHS儿童的胼胝体小或缺失一致。值得注意的是,PTHS神经组织表现出如此程度的紊乱和皮质神经元含量的减少,将有益于确定哪些临床症状是由这些异常引起的,或者这种影响是否在神经发育期间或完全形成的神经系统中表现出来。

[0051] 携带纯合功能丧失突变的Tcf4完全敲除小鼠表现出皮质神经元群体(包括SATB2-表达细胞和BRN2-表达细胞)的实质性改变,但是这些改变在Tcf4^{+/-}小鼠中明显较温和,其表现出的表型肯定没有死后PTHS皮质样品中组织紊乱和基因表达改变的水平突出(图5和12)。这表明小鼠模型不是研究TCF4杂合突变的理想模型,就像在PTHS患者中看到的那样。相比之下,本文提供的数据显示PTHS类器官中皮质神经元分化的严重受损,这与死后脑组织中的观察结果一致,这表明脑类器官模型为观察与PTHS相关的神经发育异常提供了新的机会。Tcf4^{+/-}小鼠的大脑与PTHS人类器官或死后样本之间表型严重程度差异可能反映了人类和啮齿动物之间神经发育的重要进化区别,因此,在上述和其他神经发育状况的背景下,证明使用患者来源的系统能够更好地了解病理生理学。

[0052] 本公开还展示了一系列遗传操作实验,以增强TCF4的表达并因此在体外校正其在PTHS神经组织中的表达(图11)。这些方法包括:过量表达额外的TCF4基因拷贝和CRISPR介导来自内源性TCF4位点的反-表观遗传学增强表达,这导致异常细胞表型的逆转,这一重要发现可能指导治疗PTHS的治疗工作。此外,由于CRISPR介导的TCF4表达的校正增强了突变和正常内源性等位基因的转录,因此该实验明确证明此处观察到的PTHS表型是由单倍体剂量不足引起的,而不是由显性负效应引起的。本公开的方法和组合物还有益于理解其他遗传疾病,包括归类为皮特-霍普金斯样综合征的常染色体隐性智力残疾病症(其由TCF4下游靶基因NRXN1B和CNTNAP2中的突变引起),以及精神分裂症和其他可能以TCF4为遗传成分的疾病。

[0053] 除非另有说明,本文所描述的技术的实践将采用本领域技术范围内的组织培养、免疫学、分子生物学、微生物学、细胞生物学和重组DNA的常规技术。例如,参见Green和Sambrook编辑(2012)Molecular Cloning:A Laboratory Manual,第4版;Ausubel等人编辑的系列(2015)Current Protocols in Molecular Biology;the series Methods in

Enzymology(学术出版社,纽约);MacPherson等人(2015)PCR 1:A Practical Approach(牛津大学出版社的IRL出版社);MacPherson等人(1995)PCR 2:A Practical Approach;McPherson等人(2006)PCR:The Basics(Garland Science);Harlow和Lane编辑(1999)Antibodies,A Laboratory Manual;Greenfield编辑(2014)Antibodies,A Laboratory Manual;Freshney(2010)Culture of Animal Cells:A Manual of Basic Technique,第6版;Gait编辑(1984)Oligonucleotide Synthesis;美国专利号4,683,195;Hames和Higgins编辑(1984)核酸杂交;Anderson(1999)Nucleic Acid Hybridization;Herdewijn编辑(2005)Oligonucleotide Synthesis:Methods and Applications;Hames和Higgins编辑(1984)Transcription and Translation;Buzdin和Lukyanov编辑(2007)Nucleic Acids Hybridization:Modern Applications;Immobilized Cells and Enzymes(IRL出版社(1986));Grandi编辑(2007)In vitro Transcription and Translation Protocols,第2版;Guisan编辑(2006)Immobilization of Enzymes and Cells;Perbal(1988)A Practical Guide to Molecular Cloning,第2版;Miller和Calos编辑,(1987)Gene Transfer Vectors for Mammalian Cells(冷泉港实验室);Makrides编辑(2003)Gene Transfer and Expression in Mammalian Cells;Mayer和Walker编辑(1987)Immunochemical Methods in Cell and Molecular Biology(学术出版社,伦敦);Lundblad和Macdonald编辑(2010)Handbook of Biochemistry and Molecular Biology,第4版;Herzenberg等人编辑(1996)Weir's Handbook of Experimental Immunology,第5版;和/或其更新版本。

[0054] 本说明书中使用的术语仅用于描述特定实施方案,而不适用于限定本申请。

[0055] 所有数值名称(例如pH、温度、时间、浓度和分子量),包括范围,都为以1.0或0.1递增(+)或(-)变化的近似值,视情况而定或者可选地变化 $\pm 15\%$ 、或者可选地 10% 、或者可选地 5% 、或者可选地 2% 。应当理解,尽管并非总是明确说明,所有数字名称前面都带有“大约”一词。还应理解,尽管并非总是明确说明,本文中描述的试剂仅仅是示例性的,并且其等同物在本领域是已知的。

[0056] 除非上下文另有指示,否则特别意在,本公开描述的各种特征可以任意组合使用。此外,本公开还考虑到,在一些实施方案中,可以排除或省略本文中阐述的任何特征或特征的组合。为了说明,如果说明书规定复合物包括组分A、B和C,则具体意图是A、B或C中的任何一种,或其组合,可以单独或以任何组合的方式省略和否认。

[0057] 除非另有明确说明,所有指定的实施方案、特征和术语意在包含所列举的实施方案、特征或术语及其生物学等同物。

[0058] “蛋白质”或“多肽”术语在本文中可互换使用,其包括一个或多个称为氨基酸的化学构件链,其通过称为肽键的化学键连接在一起。

[0059] 本文中使用的术语“大约”可以表示由本领域普通技术人员确定的给定值,其可接受的误差范围内,这可以部分地取决于该值是如何测量或确定的,例如,测量系统的局限性。例如,根据本领域的实践,“大约”可以表示给定值的正或负 10% 。或者,“大约”可以表示给定值的正或负 20% 、正或负 10% 、正或负 5% 或者正或负 1% 的范围。或者,特别是就生物系统或过程而言,该术语可以表示在一个数量级内,在5倍或2倍以内。在本申请和权利要求书中描述特定值的情况下,除非另有说明,术语“大约”的含义可以假定给定值在可接受的

误差范围内。此外,在给定值的范围和/或子范围的情况下,范围和/或子范围可以包括范围和/或子范围的端点。在某些情况下,变化可以包括指定量或浓度的20%、10%、5%、1%、0.5%,甚至0.1%的量。

[0060] 对于本文中数值范围的复述,明确地考虑了其具有相同精确度的每个中间数字。例如,对于6-9的范围,除了6和9之外,还考虑数字7和8,对于范围6.0-7.0,明确考虑数字6.0、6.1、6.2、6.3、6.4、6.5、6.6、6.7、6.8、6.9和7.0。

[0061] 本文中使用的术语“腺相关病毒”或“AAV”是指与该名称相关联的病毒类的成员,并且属于依赖性细小病毒属、细小病毒科。已知该病毒的多种血清型适用于基因转导;所有已知的血清型都可以感染各种组织类型的细胞。现有技术中公开了至少11种顺序编号的血清型。可用于本公开目的的非限制性示例性血清型包括11种血清型中的任何一种,例如AAV2和AAV9。本文中使用的术语“慢病毒”是指与该名称相关联的病毒类的成员,并且属于慢病毒属、逆转录病毒科。虽然已知一些慢病毒会引起疾病,但已知其他慢病毒适用于基因转导。例如,参见Tomás等人(2013) *Biochemistry, Genetics and Molecular Biology: “Gene Therapy-Tools and Potential Applications”*, ISBN 978-953-51-1014-9, DOI: 10.5772/52534。

[0062] 术语“Cas9”可以指该名称所指的CRISPR相关核酸内切酶。非限制性示例性Cas9包括金黄色葡萄球菌(*Staphylococcus aureus*) Cas9、核酸酶死亡Cas9,以及它们各自的直系同源物和生物学等同物。直系同源物包括但不限于化脓性链球菌(*Streptococcus pyogenes*) Cas9(“spCas9”)、来源于嗜热链球菌(*Streptococcus thermophiles*)、嗜肺军团菌(*Legionella pneumophila*)、乳糖奈瑟菌(*Neisseria lactamica*)、脑膜炎奈瑟菌(*Neisseria meningitidis*)和新凶手弗朗西斯菌(*Francisella novicida*)的Cas9;和来自多种细菌种属的Cpf1(执行类似于Cas9的切割功能),包括氨基酸球菌属(*Acidaminococcus* spp.)和新凶手弗朗西斯菌U112。例如,可以使用UniProtKB G3ECR1(CAS9_STRTR)以及缺少核酸内切酶活性的死亡Cas9或dCas9(例如,在RuvC和HNH结构域中都有突变)。术语“Cas9”可以进一步指与所指Cas9具有至少约60%、65%、70%、75%、80%、85%、90%、95%或99%的同一性的等同物,包括但不限于其它大的Cas9蛋白。在一些实施方案中,Cas9来源于空肠弯曲杆菌(*Campylobacter jejuni*)或另一种长度为1000个或小于1000个的氨基酸的Cas9直系同源物。

[0063] 本文所用的术语“盒”或“表达盒”是指模块化多核苷酸构建体,其可包含一个或多个结构域,使得所述表达盒能够在不同载体系统之间有效转移,并且在表达时提供基本相似的编码构造或表达谱。

[0064] “TCF4盒”是指包含TCF4编码序列的盒。在一个实施方案中,TCF4盒包括至少一个微型启动子盒或可操作地连接到编码TCF4多肽(例如TCF4-B多肽)的多核苷酸的核心启动子盒。在一些实施方案中,TCF4盒可以包含一个或多个μE5盒。因此,“TCF4盒”可以包含可操作地链接到TCF4多肽(例如TCF4-B多肽)的编码序列的单个微型启动子,并且可以包含一个或多个可操作地与微型启动子相关联的μE5盒。表达盒的示例如图4A-E所示(分别为SEQ ID NO:4、5、6、7和8)。应知道,图4中提供的序列可以变化1%至15%(例如,与图4A-D或E中的序列85%-99%相同),只要这些变体仍然可以驱动功能性TCF4多肽的转录。

[0065] 如本文所用,术语“CRISPR”可以指依赖于成簇规则间隔短回文重复途径的序列特

异性遗传操作技术。CRISPR可用于进行基因编辑和/或基因调控,以及简单地将蛋白质靶向特定的基因组位置。“基因编辑”可以指一种通过引入多核苷酸序列的缺失、插入、单链或双链断裂或碱基替换来改变靶向多核苷酸的核苷酸序列的基因工程。在一些方面,CRISPR介导的基因编辑利用非同源末端连接(NHEJ)或同源重组的途径来进行编辑。基因调控可以指增加或减少特定基因产物(如蛋白质或RNA)的产生。

[0066] 本文中使用的术语“缺乏”可以指特定药剂低于正常(生理上可接受的)水平。在蛋白质的背景下,缺乏可以指全长蛋白质低于正常水平。

[0067] 如本文所用,术语“结构域”可以指多肽或多核苷酸的特定区域并且与特定功能相关。例如,“结合RNA结合蛋白的结构域”可以指结合一种或多种控制表达的多肽的多核苷酸的结构域。

[0068] 术语“编码”当它应用于多核苷酸时可以指多核苷酸,如果在其天然状态下或当通过本领域技术人员公知的方法操纵时,它可以被转录和/或翻译以产生多肽和/或其片段的mRNA,其被称为“编码”多肽。反义链与上述核酸互补,编码序列可以由此推导出来。

[0069] 术语“等同物”或“生物等同物”在指代特定分子、生物或细胞材料时可互换使用,并且意指具有最小同源性而仍保持所需结构或功能的那些。

[0070] 本文中使用的术语“gRNA”或“向导RNA”可以指用于靶向特定多核苷酸序列的向导RNA序列,为采用CRISPR技术进行基因编辑。设计gRNA和供体治疗性多核苷酸以实现靶标特异性的技术是本领域公知的。例如,Doench,J.等人,Nature biotechnology 2014;32(12):1262-7、Mohr,S.等人(2016)FEBS Journal 283:3232-38和Graham,D.等人Genome Biol.2015;16:260。gRNA包含或可选地基本上由或进一步由包含CRISPR RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)的融合多核苷酸组成;或包含一个含有CRISPR RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)的多核苷酸。在某些方面,gRNA是合成的(Kelley,M.等人(2016)J of Biotechnology 233(2016)74-83)。

[0071] “同源性”或“同一性”或“相似性”可以指两个肽之间或两个核酸分子之间序列的相似性。同源性可以通过比较每个序列中的位置来确定,可以比对以进行比较。当比较序列中的某个位置被相同的碱基或氨基酸占据时,该分子在该位置是同源的。序列之间的同源性是序列间共有的匹配数或同源位置数的函数。“不相关”或“非同源”序列与本公开的序列之一共有小于40%的同一性,或者选择性地小于25%的同一性。

[0072] 从实际来看,特定序列可以与本文描述的任何序列(其可以对应于本文描述的特定核酸序列)至少50%、60%、70%、80%、85%、90%、92%、95%、96%、97%、98%或99%相同,常规地,上述特定序列可以使用已知的计算机程序来确定,例如Bestfit程序(威斯康星序列分析软件包,Unix版本8,遗传学计算机课题组,大学研究园,575科学驱动,麦迪逊,Wis.53711)。当使用Bestfit或任何其他序列比对程序来确定特定序列是否与参考序列如95%相同时,可以设置参数,以便在参考序列的整个长度上计算同一性百分比,并且允许高达总参考序列5%的同源性缺口。

[0073] 例如,在具体实施方案中,参考序列(查询序列,即本公开的序列)和受试者序列之间的同一性也称为全局序列比对,其可以使用基于Brutlag等人算法(Comp.App.Biosci.6:237-245(1990))的FASTDB计算机程序来确定。在某些情况下,狭义地解释同一性,用FASTDB氨基酸比对特定实施方案的参数可以包括:评分方案=PAM(接受突变的百分比)0、k-元组

=2、不匹配罚分=1、插入罚分=20、随机组长度=0、截断分数=1、窗口大小=序列长度、缺口罚分=5、缺口大小罚分=0.05、窗口大小=500或受试者序列的长度,以较短者为准。根据该实施方案,如果由于N端或C端缺失而不是由于内部缺失而导致受试者序列短于查询序列,因考虑到FASTDB程序在计算全局同一性百分比时不考虑受试者序列的N端和C端截断的事实,可以对结果进行手动更正。对于在N端和C端截断的受试者序列,相对于查询序列,可以通过计算查询序列中与受试者序列的N端和C端侧向的残基数(与相应的主体残基不匹配/比对)占查询序列总基数的百分比来校正同一性百分比。残基是否匹配/比对的确定可以通过FASTDB序列比对的结果来确定。然后,可以从由FASTDB程序使用指定的参数计算的同一性百分比中减去此百分比,以得出最终的同一性百分比分数。该最终同一性百分比分数可用于本实施方案的目的。在某些情况下,仅考虑受试者序列的N端和C端的残基,这些残基与查询序列不匹配/对齐,以便手动调整同一性百分比分数。也就是说,此手动校正仅考虑在受试者序列最远的N端和C端残基之外的查询残基位置。例如,可以将90个残基的受试者序列与100个残基的查询序列进行对齐,以确定同一性百分比。缺失发生在受试者序列的N末端,因此,FASTDB比对不显示N端前10个残基的匹配/对齐。10个未配对的残基代表序列的10%(N端和C端的残基数不匹配/查询序列中的残基总数),因此从FASTDB程序计算的同一性百分比分数中减去10%。如果剩余的90个残基完全匹配,则最终的同一性百分比可以是90%。在另一个示例中,将90个残基的受试者序列与100个残基的查询序列进行比较。这次的删除是内部删除,因此在受试者序列的N端或C端没有与查询序列不匹配/对齐的残留物。在这种情况下,不会手动更正FASTDB计算的同一性百分比。同样,仅手动更正与查询序列不匹配/对齐的受试者序列的N端和C端之外的残基位置(如FASTDB对齐所示)。两个多核苷酸可以使用类似的技术进行比对。

[0074] “杂交”可以指其中一个或多个多核苷酸反应形成复合物的反应,该复合物通过核苷酸残基的碱基之间的氢键而稳定。氢键可以通过沃森-克里克碱基配对、霍格斯坦结合或任何其他序列特异性方式发生。该复合物可以包括形成双链结构的两条链、形成多链复合物的三条或多条链、单个自杂交链或上述的任意组合。杂交反应可以构成更广泛过程中的步骤,例如PC反应的启动,或核酶对多核苷酸的酶切。

[0075] 举例来说,严格的杂交条件包括:孵育温度为约25℃至约37℃;杂交缓冲液浓度为约6×SSC至约10×SSC;甲酰胺浓度为约0%至约25%;且洗涤液为约4×SSC至约8×SSC。举例来说,温和的杂交条件包括:孵育温度为约40℃至约50℃;缓冲液浓度为约9×SSC至约2×SSC;甲酰胺浓度为约30%至约50%;且洗涤液为约5×SSC至约2×SSC。举例来说,高严格的条件包括:孵育温度为约55℃至约68℃;缓冲液浓度为约1×SSC至约0.1×SSC;甲酰胺浓度为约55%至约75%;且洗涤液为约1×SSC、0.1×SSC或去离子水。一般来说,杂交孵育时间为5分钟至24小时,有1、2或更多个洗涤步骤,洗涤孵育时间约为1、2或15分钟。SSC为0.15M NaCl和15mM柠檬酸盐缓冲液。此处可理解为,可以采用使用其他缓冲液体系的SSC的等同物。

[0076] 本文所用的术语“分离的”可以指基本上不含其它材料的分子、生物制剂或细胞材料。在一个方面,术语“分离的”可以指核酸(例如DNA或RNA)、或蛋白质或多肽(例如,抗体或其衍生物)、或细胞或细胞器、或组织或器官分别与存在于天然来源中的其它DNA或RNA、或蛋白质或多肽、或细胞或细胞器、或组织或器官分离。术语“分离的”也可以指通过重组DNA

技术生产时基本上不含细胞物质、病毒物质或培养基的核酸或肽,或化学合成时基本上不含化学前体或其他化学物质的核酸或肽。此外,“分离的核酸”是指包括自然状态下不存在的片段,并且包括可能自然状态下没有发现的核苷酸片段。术语“分离的”在本文中也用于指从其它细胞蛋白中分离的多肽,并且意在包括纯化和重组的多肽。术语“分离的”在本文中也用于指从其它细胞或组织中分离的细胞或组织,并且旨在包括培养的和工程化的细胞或组织。

[0077] 术语“蛋白质”、“肽”和“多肽”可互换使用,并且其最广义的意义是指两个或多个亚氨基酸、氨基酸类似物或拟肽的化合物。亚基可以通过肽键连接。在另一个实施方案中,亚基可以通过其它键连接,例如酯、醚等。蛋白质或肽可以包含至少两个氨基酸,并且对可以包含蛋白质或肽序列的最大氨基酸数量没有限制。本文所用的术语“氨基酸”可以指天然和/或非天然或合成氨基酸,包括甘氨酸和D和L两种光学异构体、氨基酸类似物和拟肽。如本文所用,术语“融合蛋白”可以指由来自一个以上天然存在或重组产生的蛋白质的结构域组成的蛋白质,通常其中每个结构域具有不同的功能。在这方面,术语“接头”可以指用于将这些结构域连接在一起的蛋白质片段——可选地保持融合蛋白质结构域的构象和/或防止融合蛋白质结构域之间的不利相互作用,这种不利相互作用可能损害它们各自的功能。

[0078] 术语“多核苷酸”和“寡核苷酸”可互换使用,并指任何长度的核苷酸的聚合形式,脱氧核糖核苷酸或核糖核苷酸或其类似物。多核苷酸可以具有任何三维结构,并且可以执行任何已知或未知的功能。以下是多核苷酸的非限制性示例:基因或基因片段(例如,探针、引物、EST或SAGE标签)、外显子、内含子、信使RNA(mRNA)、转运RNA、核糖体RNA、RNAi、核酶、cDNA、重组多核苷酸、支链多核苷酸、质粒、载体、任意序列的分离DNA、任意序列的分离RNA、核酸探针和引物。多核苷酸可以包含修饰的核苷酸,例如甲基化的核苷酸和核苷酸类似物。如果存在,可以在多核苷酸组装之前或之后对核苷酸结构进行修饰。核苷酸序列可以被非核苷酸组分中断。聚合后可以进一步修饰多核苷酸,例如通过与标记元件偶联。上述术语也可以指双链和单链分子。除非另有说明或要求,本公开的任何多核苷酸的实施方案既包括双链形式,也包括已知或预测构成双链形式的两种互补单链形式中的每一种。

[0079] 可以理解,本文中描述的多核苷酸包括“基因”,并且本文中描述的核酸分子包括“载体”或“质粒”。例如,编码TCF4的多核苷酸可以由TCF4基因或其同系物编码。因此,术语“基因”,也称为“结构基因”是指编码特定氨基酸序列的多核苷酸,其包含一种或多种蛋白质或酶的全部或部分,并且可以包括调节性(非转录的)DNA序列,例如启动子序列,其确定例如基因表达的条件。基因的转录区域可能包括非翻译区域,其包括内含子、5'-非翻译区(UTR)和3'-UTR,以及编码序列。术语“核酸”或“重组核酸”是指多核苷酸,例如脱氧核糖核酸(DNA),并在适当情况下指核糖核酸(RNA)。本文提供的任何包含胸腺嘧啶(T)的序列都可以通过用“U”(尿嘧啶)替换“T”来转换为RNA序列。因此,本文考虑了DNA和RNA序列。

[0080] 本领域技术人员将认识到,由于遗传密码的简并性质,核苷酸序列不同的各种DNA或RNA化合物的可用于编码本公开的给定氨基酸序列。编码TCF4的天然DNA或RNA序列只是本公开的说明性实施方案,并且本公开包括编码本公开方法中使用的多肽和蛋白质氨基酸序列的任何序列的DNA化合物。以类似的方式,多肽通常可以耐受其氨基酸序列中的一个或多个氨基酸替换、缺失和插入,而不会损失或显着损失所需的活性。本公开包括具有替代的氨基酸序列的此类多肽,并且由本文所示的DNA序列编码的氨基酸序列仅说明本公开的实

施方案。

[0081] 本公开的核酸可以使用cDNA、mRNA或替代的基因组DNA作为模板和适当的寡核苷酸引物根据标准的PCR扩增技术进行扩增。如此扩增的核酸可以克隆到适当的载体中,并通过DNA序列分析进行表征。此外,对应于核苷酸序列的寡核苷酸可以通过标准合成技术制备,例如,使用自动DNA合成器。

[0082] 还可以理解,编码与本文所述的TCF4多肽的同源多肽的分离的核酸分子,可以通过在编码特定多肽的核苷酸序列中引入一个或多个核苷酸替换、插入或缺失来创建,使得一个或多个氨基酸替换、插入或缺失被引入到编码的蛋白质中。可以通过标准技术将突变引入多核苷酸中,例如定点诱变和PCR介导的诱变。与那些可能需要进行非保守氨基酸替换的位置相反,在某些位置优选进行保守的氨基酸替换。“保守氨基酸替换”是指氨基酸残基被具有相似侧链的氨基酸残基替换。本领域已定义了具有相似侧链的氨基酸残基家族。这些家族包括具有碱性侧链(例如,赖氨酸、精氨酸、组氨酸)、酸性侧链(例如,天冬氨酸、谷氨酸)、不带电荷的极性侧链(例如,甘氨酸、天冬酰胺、谷氨酰胺、丝氨酸、苏氨酸、酪氨酸、半胱氨酸)、非极性侧链(例如,丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、蛋氨酸、色氨酸)、 β -支链侧链(例如,苏氨酸、缬氨酸、异亮氨酸)和芳香族侧链(例如,酪氨酸、苯丙氨酸、色氨酸、组氨酸)。

[0083] 术语“多核苷酸序列”是多核苷酸分子的字母表示。这种字母表示可以输入到具有中央处理器的计算机的数据库中,并用于生物信息学应用,例如功能基因组学和同源性搜索。

[0084] 本文所用的术语“重组表达系统”是指通过重组形成的用于表达某些遗传物质的一个或多个遗传构建体;在这方面,术语“构建体”可与本文定义的术语“载体”互换。

[0085] 如本文所用,与蛋白质表达相关的术语“恢复”可以指在先前蛋白质表达由于突变而被截断的情况下,建立全长蛋白质表达的能力。在“恢复活性”的内容中,该术语包括将蛋白质的表达影响到其正常的非突变水平,其中突变导致异常表达(例如,太低或太高)。

[0086] “转化”是指将载体引入宿主细胞的过程。转化(或转导,或转染)可以通过多种方式中的任何一种来实现,包括病毒递送、电穿孔、显微注射、生物制剂(或粒子轰击介导的递送)等。

[0087] 本文所用的术语“治愈”、“治疗”等用于表示获得期望的药理学和/或生理学效果。效果可以是预防(完全或部分预防疾病、病症或其状况或体征或症状),和/或治疗(可以部分或完全治愈病症和/或该病症的不良反应)。

[0088] 本文所用的术语“载体”可以指用于在不同宿主之间转移的核酸构建体,包括但不限于质粒、病毒、粘粒、噬菌体、BAC、YAC等。“病毒载体”被定义为重组产生的病毒或病毒颗粒,其包含可在体内、离体或体外递送到宿主细胞中的多核苷酸。在一些实施方案中,质粒载体可以从市售载体制备。在其他实施方案中,根据本领域已知的技术,病毒载体可以由杆状病毒、逆转录病毒、腺病毒、AAV等产生。在一个实施方案中,病毒载体是慢病毒载体。病毒载体的例子包括逆转录病毒载体、腺病毒载体、腺相关病毒载体、甲病毒载体等。基于传染性烟草花叶病毒(TMV)的载体可用于制造蛋白质,并且据报道在烟叶中表达Griffithsin(O'Keefe等人(2009)Proc.Nat.Acad.Sci.USA 106(15):6099-6104)。还开发了甲病毒载体(例如基于Semliki Forest病毒的载体和基于Sindbis病毒的载体)用于基因治疗和免疫治

疗。参见Schlesinger&Dubensky (1999) Curr. Opin. Biotechnol. 5:434-439和Ying等人 (1999) Nat. Med. 5(7):823-827。在逆转录病毒载体介导基因转移的方面,载体构建体可以指包含逆转录病毒基因组或其部分的多核苷酸和目的基因。有关用于基因转移的载体的现代方法的更多详细信息可以在,例如,Kotterman等人 (2015) Viral Vectors for Gene Therapy: Translational and Clinical Outlook Annual Review of Biomedical Engineering 17中找到。包含启动子和多核苷酸的载体可以操作地连接到其中的克隆位点是本领域公知的。上述载体能够在体外或体内转录RNA,并且可以从安捷伦科技 (Santa Clara, Calif.) 和Promega Biotech (Madison, Wis.) 等来源商购获得。

[0089] PTHS可由TCF4基因中的杂合突变引起,其编码碱性螺旋-环-螺旋 (bHLH) 转录因子。PTHS患者表现出严重的智力障碍和认知障碍,明显的发育迟缓,完全丧失口语,以及特征性的面部完形。大多数患者表现为肌张力低下、运动迟缓和/或协调性受损。常见的表现是便秘,可能是由于肠道神经系统异常。呼吸异常和癫痫发作是一种可变的临床表现,有时是晚发的。自闭症行为包括缺乏语言交流、智力障碍和重复的以自我为中心的行为。

[0090] TCF4基因位于18号染色体 (18q21.2) 上,并且包含18个编码外显子。其最长和最广泛研究的选择性剪接转录本编码TCF4-B蛋白亚型,这是一种在发育过程中在整个大脑中高度表达的bHLH转录因子。TCF4蛋白与E-box调节序列 (共有CANNTG) 结合,并与免疫系统、上皮-间充质转化和神经系统的许多发育过程有关。大多数PTHS患者在TCF4基因中表现出独有突变,这可能是跨越整个基因的大染色体缺失、部分基因缺失、易位或点突变。

[0091] 本公开提供了用于改变人TCF4基因 (OMIM 602272; 又名E2-2、ITF2、PTHS、SEF2和bHLHb19) 表达的DNA构建体和方法,因此可用于开发在患病细胞和组织或在具有TCF4表达降低的个体中增加TCF4基因表达的方法,所述个体例如,但不限于,患有称为皮特-霍普金斯综合征 (PTHS; MIM#610954) 遗传疾病的人类受试者。值得注意的是,根据WGAS研究,TCF4也是精神分裂症的最高风险基因。

[0092] 本公开提供了多个表达盒,其可与合适的DNA构建体和载体一起使用,以递送和/或增加TCF4或TCF4的表达。例如,在一个实施方案中,将TCF4编码序列 (例如,TCF4基因) 的额外拷贝插入到靶细胞或组织中。DNA构建体包括TCF4-B转录本或变体的编码序列,其通常以DNA调控元件为前导,一旦将构建体转入靶细胞中,这些元件就可以调控表达速率。TCF4-B基于其在神经祖细胞和神经元中的高表达水平作为示例 (图1)。在一个实施方案中,表达盒TCF4-B cDNA序列可以置于合成微型启动子 (minP) 的控制之下,前面是由TCF4蛋白本身识别的不同数量 (0、1、2、3、4、5、6、7、8、9、10、11、12、3、14、15) 的调节序列,称为 μ E5盒 (图2)。这种方法仅为通常发现TCF4基因的细胞类型中提供TCF4基因表达。

[0093] 结果表明,TCF4表达的水平可以通过改变 μ E5盒的数量来操纵,提供靶细胞中TCF4基因的可控过表达 (图2)。此外,这些DNA构建体已被用于慢病毒颗粒中,以转染 (转导) 源自PTHS患者的患病靶细胞。特别是,这些构建体已被引入神经祖细胞 (NPC) 中。实验验证了它们可以增强这些细胞中的TCF4表达 (图3A),将其表达水平提高2倍至5倍,这具体取决于测试的特定构建体中 μ E5盒的数量。此外,这种遗传操作校正了TCF4靶基因 (如GADD45G (图3B)) 的表达,使其水平恢复到对照细胞系中的正常水平 (图3B)。

[0094] 如本文所述,本公开的盒可包含可操作地与多核苷酸连接的微型启动子,所述多核苷酸与TCF4 (TCF4-B) cDNA具有至少85%、90%、92%、95%、98%、99%或100%相同:

[0095] ATGCATCACCAACAGCGAATGGCTGCCTTAGGGACGGACAAAGAGCTGAGTGATTTACTGGATTTCAGTGCGATGTTTTACCTCCTGTGAGCAGTGGGAAAAATGGACCAACTTCTTTGGCAAGTGGACATTTTACTGGCTCA AATGTAGAAGACAGAAGTAGCTCAGGGTCTGGGGGAATGGAGGACATCCAAGCCCGTCCAGGA ACTATGGAGATG GGA CTCCCTATGACCACATGACCAGCAGGGACCTTGGGTACATGACAATCTCTCTCCACCTTTTGTCAATTCCAG AATACAAAGTAAAAACAGAAAGGGGCTCATACTCATCTTATGGGAGAGAATCAA ACTTACAGGGTTGCCACCAGCAG AGTCTCCTTGGAGGTGACATGGATATGGGCAACCCAGGAACCCCTTTCGCCCACCAAACCTGGTTCCCAGTACTATC AGTATTCTAGCAATAATCCCCGAAGGAGGCCTCTTCACAGTAGTGCCATGGAGGTGCAGACAAAGAAAGTTCGAAA AGTTCTCCAGGTTTGCCATCTTCAGTCTATGCTCCATCAGCAAGCACTGCCGACTACAATAGGGACTCGCCAGGC TATCCTTCTCCAAACCAGCAACCAGCACTTTCCTAGCTCCTTCTTCATGCAAGATGGCCATCACAGCAGTGACC CTTGGAGCTCCTCCAGTGGGATGAATCAGCCTGGCTATGCAGGAATGTTGGGCAACTCTTCTCATATTCCACAGTC CAGCAGCTACTGTAGCCTGCATCCACATGAACGTTTGGAGCTATCCATCACACTCCTCAGCAGACATCAATTCCAGT CTTCTCTCCGATGTCCACTTTCCATCGTAGTGGTACAAACCATTACAGCACCTCTTCCTGTACGCCTCCTGCCAACG GGACAGACAGTATAATGGCAAATAGAGGAAGCGGGGAGCCGGCAGCTCCCAGACTGGAGATGCTCTGGGGAAAGC ACTTGCTTCGATCTATTCTCCAGATCACACTAACAACAGCTTTTCATCAAACCCCTCAACTCCTGTTGGCTCTCCT CCATCTCTCTCAGCAGGCACAGCTGTTTGGTCTAGAAATGGAGGACAGGCCTCATCGTCTCCTAATTATGAAGGAC CTTTACACTCTTTGCAAAGCCGAATTGAAGATCGTTTAGAAAGACTGGATGATGCTATTCATGTTCTCCGGAACCA TGCAGTGGGCCCATCCACAGCTATGCCTGGTGGTCATGGGGACATGCATGGAATCATTGGACCTTCTCATAATGGA GCCATGGGTGGTCTGGGCTCAGGGTATGGAACCGGCCTTCTTTCAGCCAACAGACATTCACTCATGGTGGGGACCC ATCGTGAAGATGGCGTGGCCCTGAGAGGCAGCCATTCTCTTCTGCCAAACCAGGTTCCGGTTCACAGCTTCTGT CCAGTCTGCGACTTCCCCTGACCTGAACCCACCCAGGACCTTACAGAGGCATGCCACCAGGACTACAGGGGCAG AGTGTCTCCTCTGGCAGCTCTGAGATCAAATCCGATGACGAGGGTGATGAGAACCTGCAAGACACGAAATCTTCGG AGGACAAGAAATTAGATGACGACAAGAAGGATATCAAATCAATTACTAGGTCAAGATCTAGCAATAATGACGATGA GGACCTGACACCAGAGCAGAAGGCAGAGCGTGAGAAGGAGCGGAGGATGGCCAACAATGCCCAGAGCGTCTGCGG GTCCGTGACATCAACGAGGCTTTCAAAGAGCTCGGCCGCATGGTGCAGCTCCACCTCAAGAGTGACAAGCCCCAGA CCAAGCTCCTGATCCTCCACCAGGCGTGGCCGTCATCCTCAGTCTGGAGCAGCAAGTCCGAGAAAGGAATCTGAA TCCGAAAGCTGCGTGTCTGAAAAGAAGGGAGGAAGAGAAGGTGTCTCAGAGCTCCCCCTCTCTCCTTGGCCGGC CCACACCCTGGAATGGGAGACGCATCGAATCACATGGGACAGATGTAA (SEQ ID NO:1)

[0096] 并编码多肽 (SEQ ID NO:2) :

[0097] MHHQQRMAALGTDKELSDLDFSAMFSPVSSGKNGPTSLASGHFTGSNVEDRSSSGSWGNGHPSPSR NYGDGTPYDHMTSRDLGSHDNLSPFVNSRIQSKTERGSYSSYGRESNLQGCHQQSLLGGDMDMGNPGTLSPTKPGS QYYQYSSNNPRRRPLHSSAMEVQTKKVRKVPPGLPSSVYAPSASTADYNRDSPGYPSSKPATSTFPSSFFMQDGHHS SDPWSSSSGMNQPGYAGMLGNSSHIPQSSSYCSLHPHERLSYPSSHSSADINSSLPPMSTFHRSGTNHYSTSSCTPPA NGTDSIMANRSGAAGSSQTGDALGKALASIYSPDHTNNSFSSNPSTPVGSPPSLSAGTAVWSRNGQASSSPNYEG PLHSLQSRIEDRLERLDDAIHVLNRHAVGPSTAMPGGHGDHGIIGPSHNGAMGGLSGYGTGLLSANRHSLMVGTH REDGVALRGSHSLLPNQVPVQLPVQSATSPDLNPPQDPYRGMPPGLQGQSVSSGSSEIKSDDEGDENLQDTKSSD KKLDDDKKDIKSITRSRSSNDDDLTPEQKAEREKERRMANNARERLRVRDINEAFKELGRMVQLHLKSDKPQTKL LILHQA VAVILSLEQQVRERNLNPKAACLRREEEEKVSSEPPPLSLAGPHPGMGDASNHMGQM

[0098] 上述表达盒可以包含微型启动子和一个或多个 (例如, 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15或更多) 的蛋白质结合结构域, 例如一个或多个可操作地连接到上述微型

启动子的 μ E5盒结构域。 μ E5盒具有以下序列:cacctg;并且在一些实施方案中, μ E5与下一个相邻 μ E5间隔约3至约10个核苷酸,例如约4至约8个核苷酸(例如,约6个核苷酸),使得其可以包含如SEQ ID NO:10所示的序列)。示例性 μ E5如SEQ ID NO:10所示。示例性微型启动子包括agagggtatataatggaagctcgacttccag (SEQ ID NO:3)。本文描述了其它微型启动子或核心启动子。在一些实施方案中,如SEQ ID NO:3所示的微型启动子通过间隔区(例如caagaa)与 μ E5结构域分离。

[0099] 表达盒可以通过重组分子生物学技术定位到合适的载体中,以促进递送到细胞。如下所述,合适的载体可以是DNA构建体或病毒载体(例如,腺病毒载体、逆转录病毒载体,例如慢病毒载体和gamma病毒载体)。

[0100] 大多数启动子相当大,通常超过600bp;全长的启动子可以是上千碱基。可以产生更小的启动子,其允许从逆转录病毒载体(包括复制和非复制病毒载体)的载体中可靠地表达哺乳动物细胞中的转基因。如上所述,合适的微型启动子可以包含SEQ ID NO:3。其他合适的微型启动子可以从Kadanaga和合作者(Juven-Gershon等人,Nature Methods,11:917-922,2006)描述的“核心”启动子中获得。这些核心启动子基于腺病毒主要晚期(AdML)和巨细胞病毒(CMV)主要直接早期基因,以及合成的“超级核心启动子”-1(SCP1)。其它细胞核心启动子包括但不限于:人血红蛋白加氧酶近端启动子(121bp;Tyrrell等人,Carcinogenesis,14:761-765,1993)、CTP:磷酸胆碱胞苷转移酶(CCT)启动子(240bp;Zhou等人,Am.J.Respir.Cell Mol.Biol.,30:61-68,2004);人ASK(S期激酶的激活剂,也称为HsDbf4基因,63bp;Yamada等人,J.Biol.Chem.,277:27668-27681,2002);和HSVTK基因内核(A1-Shawi等人,Mol.Cell.Biol.,11:4207,1991;Salamon等人,Mol.Cell.Biol.,15:5322,1995)。此外,这些“核心”启动子可用作进一步修饰以改善启动子活性的起点。例如,这种修饰包括向“核心”启动子添加其它结构域和序列以改进功能(例如,增强子、Kozak序列等)。在一个实施方案中,这种进一步的修饰可以包括添加增强子或转录结合蛋白序列。

[0101] 这些核心启动子的长度约为30-80bp,因此,当用于病毒载体时,其为转基因序列提供了充足的额外容量。使用这种启动子可以为例如TCF4基因或编码序列(例如,SEQ ID NO:1)提供有用的基因表达。

[0102] 此外,使用合理的设计技术可以使用各种启动子组分来优化载体和表达盒的表达和稳定性。这种优化的核心启动子为病毒多核苷酸提供了更有效的表达和稳定性。例如,“设计者”启动子可以包括核心启动子,该核心启动子已被进一步修饰以包括一个或多个适合于稳定性和表达的附加元件。

[0103] 本文所用的“核心启动子”是指包含约30-100bp且缺乏增强子元件的微型启动子。这些核心启动子包括但不限于SCP1、AdML和CMV核心启动子以及SEQ ID NO:3的启动子。示例性启动子可以包括SEQ ID NO:3。

[0104] 核心启动子包括某些病毒启动子。本文所用的病毒启动子是指具有核心序列但通常还有一些进一步的辅助元件的启动子。例如,SV40的早期启动子包含三类元件:TATA盒、起始位点和GC重复(Barrera-Saldana等人,EMBO J,4:3839-3849,1985;Yaniv,病毒学,384:369-374,2009)。TATA盒位于转录起始位点上游约20个碱基对的位置。GC重复区域是包含六个GC盒的21个碱基对重复序列,是决定转录方向的位点。上述核心启动子序列约为100bp。添加额外的72个碱基对重复序列,从而使其成为“小启动子”,作为转录增强子,可将

启动子的功能增加约10倍。当SP1蛋白与21bp重复序列相互作用时,它会结合最前面或最后的三个GC盒。前三个的结合启动早期表达,后三个的结合启动后期表达。72bp重复序列的功能是增强稳定RNA的量并提高合成速率。这是通过与AP1(激活蛋白1)结合(二聚化),以产生3'多腺苷酸化和5'帽的主要转录本来完成的。其它病毒启动子,如Rous Sarcom病毒(RSV)、HBV X基因启动子和疱疹胸苷激酶核心启动子也可以用作选择所需功能的基础。

[0105] 核心启动子通常包含相对于转录起始位点+1的-40至+40 (Juven-Gershon和Kadonaga, Dev. Biol., 339:225-229, 2010), 其定义了RNA聚合酶II机械上启动转录的位置。通常, RNA聚合酶II与许多转录因子相互作用, 这些转录因子与启动子中的DNA基序结合。这些因子通常被称为“一般”或“基础”转录因子, 包括但不限于TFIIA (RNA聚合酶IIA的转录因子)、TFIIB、TFIID、TFIIE、TFIIF和TFIIH。这些因子以“一般”的方式与所有核心启动子一起发挥作用; 因此, 它们通常被称为“基础”转录因子。

[0106] Juven-Gershon等人, 2006 (同上), 描述了核心启动子的元件。例如, pRC/CMV核心启动子由长度为81bp的TATA盒组成; CMV核心启动子由TATA盒和起始位点组成; 而SCP合成核心启动子 (SCP1和SCP2) 由TATA盒、Inr (起始子)、MTE位点 (十基序元件) 和DPE位点 (下游启动子元件) 组成, 长度约为81bp。与简单的pRC/CMV核心启动子相比, SCP合成启动子的表达改善。

[0107] 本文所用的“微型启动子”或“小启动子”是指促进可操作连接的基因或编码核酸序列的转录的调控结构域。顾名思义, 微型启动子包括有效转录和/或翻译可操作连接的编码序列所需的最少量的元件。微型启动子可以包括结合附加调控元件的“核心启动子”或“修饰的核心启动子”。通常, 微型启动子或修饰的核心启动子的长度约为30-600bp, 而核心启动子通常小于约100bp (例如, 约30-80bp)。在其它存在核心启动子的实施方案中, 盒可任选地包含增强子元件或核心启动子序列上游或下游的另一元件, 其促进可操作连接的编码序列的表达高于单独核心启动子的表达水平。。

[0108] 因此, 本公开提供了源自细胞元件的微型启动子 (例如, 修饰的核心启动子), 如针对“核心启动子”元件所确定的, 所述“核心启动子”元件允许在靶细胞中以显著水平普遍表达, 且通常可用于稳定掺入到载体中, 且尤其可用于稳定掺入到病毒载体中, 以允许有效表达转基因。还提供了由核心启动子加上最小增强子序列和/或Kozak序列组成的微型启动子, 其仍低于200、400或600bp, 且与缺少上述序列的核心启动子相比, 基因表达更好。上述微型启动子包括修饰的核心启动子和天然存在的组织特异性启动子, 例如弹性蛋白启动子 (对胰腺腺泡细胞具有特异性, (204bp; Hammer等人, Mol Cell Biol., 7:2956-2967, 1987) 和来自小鼠和人的细胞周期依赖性ASK基因的启动子 (63-380bp; Yamada等, J. Biol. Chem., 277:27668-27681, 2002)。普遍表达的微型启动子还包括病毒启动子, 例如SV40早期和晚期启动子 (约340bp)、RSV LTR启动子 (约270bp) 和HBV X基因启动子 (约180bp) (例如, R Anish等人, PLoS One, 4:5103, 2009), 其没有规范的“TATTAA盒”, 并且具有5'-CCCCGTTGCCCG-3'的13bp核心序列。

[0109] 如本文所述, 可以在各种表达盒和载体 (包括有复制能力和无复制能力的病毒载体) 使用单独或包括用于表达的附加元件的上述微型启动子, 以表达TCF4编码序列 (例如, SEQ ID NO:1)。例如, 本公开提供了可以并入表达载体或病毒载体的盒。已知多种载体, 并且可以将盒克隆到此类表达载体或病毒载体中。例如, 一些病毒载体可以将盒克隆到长

末端重复序列(LTR)中。其他载体可以将表达盒克隆到包膜基因的下游和3'LTR的上游。然而,其他非复制载体由于已经去除了关键基因(例如,gag和pol),其具有更大的盒容量。

[0110] 另一种适合用于CNS的递送载体包括纳米颗粒,其通常具有小于200nm、或小于约150nm、或小于约100nm的尺寸。其可能包括基于脂质的纳米颗粒、聚合物纳米颗粒,树枝状聚合物和无机纳米颗粒,其中一些可以定制为通过血脑屏障(BBB)。在一些实施方案中,递送系统通过使用转运蛋白或受体的配体来增强穿过BBB的纳米颗粒摄取,从而主动靶向递送。这种方法的优选途径是受体(或转运蛋白)介导的转胞吞作用,通过该转胞吞作用,货物(例如纳米颗粒)在大脑内皮细胞(EC)的顶端和基底外侧表面之间运输。例如,低密度脂蛋白通过受体介导过程穿过EC进行转胞吞作用,绕过溶酶体区室并在脑侧的基底外侧表面释放。此外,由于BBB含有氨基酸的转运蛋白,因此使用天然存在的精氨酸转运蛋白进行递送是递送到大脑的一种方法。另一种大脑递送的载体是外泌体,外泌体是由细胞分泌的小细胞外囊泡。与其他合成纳米颗粒相比,外泌体的主要优点是它们的非免疫原性,导致长而稳定的循环。

[0111] 本公开提供了用于治疗 and/或减轻与中枢神经系统(CNS)或周围神经系统(PNS)的神经元细胞中TCF4的异常表达相关的神经或神经发育疾病和病症的症状的方法和组合物,通过施用有效量的包含本公开的TCF4盒的构建体,使得该盒由神经元细胞表达。神经或神经发育疾病或病症也可能与神经元细胞中TCF4转录因子基因表达和/或蛋白质功能的缺陷或异常有关,例如,通过突变或单倍体剂量不足。这些神经或神经发育疾病和病症包括,例如,皮特-霍普金斯综合症(PTHS)、精神分裂症、自闭症、自闭症谱系障碍等。可以设计TCF4盒,使得包含该盒的构建体在神经元细胞中异位表达。如本文所用,异位表达是指蛋白质在不正常表达的细胞和/或组织中表达和/或活性。在本例中,TCF4异常、非正常或非典型表达或活性。

[0112] 本公开提供了一种治疗和/或减轻神经或神经发育疾病或病症的症状的方法,包括递送TCF4盒和表达该盒以治疗和/或减轻所述神经或神经发育疾病或病症的症状。在一个实施方案中,包含TCF4盒的载体是具有如SEQ ID NO:9所示序列的AAV9载体或与如SEQ ID NO:9所示序列具有至少80%、85%、90%、92%、95%、97%、98%或99%相同的序列。在一个实施方案中,神经或神经发育疾病或病症与神经元细胞中TCF4转录因子基因表达和/或蛋白质功能缺陷或异常有关。在一个实施方案中,有需要的受试者具有、被怀疑具有或处于具有(例如,已被鉴定为具有TCF4突变)上述神经或神经发育疾病或病症的风险。在实施方案中,所述神经或神经发育疾病或病症是皮特-霍普金斯综合征、精神分裂症、自闭症、自闭症谱系障碍或18q综合征等。

[0113] 本公开还提供了一种在有需要的受试者中治疗和/或减轻与异常或缺陷神经元TCF4表达和/或功能相关的神经或神经发育疾病或病症的症状的方法,通过向受试者施用治疗有效量的本公开的TCF4构建体(例如,包含TCF4盒的载体)。

[0114] 本公开还提供了施用本公开所述载体和/或表达盒的药物组合物,其可以方便地以剂量单位形式呈现并且可以通过药学领域公知的任何方法制备。上述药物组合物可以例如通过将本文提供的载体和/或含盒的组合物与液体载体、细碎的固体载体或两者均匀紧密地结合来制备。在药物组合物中包括足以产生所需治疗效果的量的本文提供的化合物。

[0115] 系统制剂包括那些设计通过注射(例如,皮下、静脉内、输液、肌肉内、颅内、脊管

内、鞘内或腹腔内注射)施用的制剂以及那些设计用于透皮、经粘膜、口服或肺部施用的制剂。

[0116] 有用的可注射制剂包括本文提供的化合物在水性或油性载体中形成的无菌悬浮液、溶液或乳液。组合物还可以含有配制剂,例如悬浮剂、稳定剂和/或分散剂。注射制剂可以单位剂型呈现,例如,在安瓿或多剂量容器中,并且可以含有添加的防腐剂。

[0117] 或者,可注射制剂可以在使用前以粉末形式提供,用合适的载体复溶,包括但不限于无菌的无热原水、缓冲液和葡萄糖溶液。为此,本文提供的组合物可以通过任何公知的技术进行干燥,例如冻干,并在使用前进行复溶。

[0118] “施用”可以在整个治疗过程中以连续或间歇的方式一次施用。确定最有效的施用手段和剂量的方法为本领域技术人员公知,并且可以随用于治疗的组合、治疗目的、被治疗的靶细胞和被治疗的受试者而变化。单次或多次施用可以在治疗医生选择剂量水平和模式的情况下进行。合适的剂量制剂和施用药剂的方法是本领域已知的。施用途径和确定最有效的施用途径的方法也是本领域技术人员已知的,并且可以随用于治疗的组合、治疗目的、被治疗对象的健康状况或疾病阶段以及靶细胞或组织而变化。

[0119] 施用可以指可用于将化合物或组合物递送到所需的生物作用位点(诸如DNA构建体、病毒载体或其它)的方法。这些方法可以包括肠胃外施用(包括静脉内、皮下、鞘内、腹腔内、肌肉内、血管内或输液)、脑内和脊髓内。在某些情况下,受试者可以在没有监督的情况下管理作文。在某些情况下,受试者可以在医疗专业人员(例如,医生、护士、医师助理、秩序员,临终关怀工作者等)的监督下施用组合物。在某些情况下,医疗专业人员可以施用组合物。在某些情况下,美容专业人员可以施用组合物。

[0120] 本文公开的组合物的施用或应用可以连续或非连续进行至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100天的治疗持续时间。在某些情况下,治疗持续时间可以是约1至约30天、约2至约30天、约3至约30天、约4至约30天、约5至约30天、约6至约30天、约7至约30天、约8至约30天、约9至约30天、约10至约30天、约11至约30天、约12至约30天、约13至约30天、约14至约30天、约15至约30天、约16至约30天、约17至约30天、约18至约30天、约19至约30天、约20至约30天、约21至约30天、约22至约30天、约23至约30天、约24至约30天、约25至约30天、约26至约30天、约27至约30天、约28至约30天或约29至约30天。

[0121] 本文公开的组合物的施用或应用可以每天至少进行1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23或24次。在某些情况下,施用或应用本文公开的组合物可以每周至少进行1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或21次。在一些情况下,施用或应用本文公开的组合物可以每月至少进行1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89或90次。

[0122] 在一些情况下,组合物可以作为单剂量或分剂量施用/施用。在某些情况下,本文中描述的组合物可以在第一时间点和第二时间点施用。在某些情况下,可以施用组合物使得第一次施用先于另一个施用,施用时间相差1小时、2小时、4小时、8小时、12小时、16小时、20小时、1天、2天、2天、7天、2周、4周、2个月、3个月、4个月、5个月、6个月、7个月、8个月、9个月、10个月、11个月、1年或以上。

[0123] “组合物”通常指活性剂的组合,例如本公开的TCF4盒,通常在诸如病毒载体(例如,和AAV9载体)的载体和天然存在或非天然存在的惰性或活性载体中,例如佐剂、稀释剂、粘合剂、稳定剂、缓冲剂、盐、亲脂性溶剂、防腐剂、佐剂等以及包括药学上可接受的载体。在一个实施方案中,组合物包含与SEQ ID NO:9至少80%-100%相同的序列。载体还包括药用赋形剂和添加剂蛋白质、肽、氨基酸、脂质和碳水化合物(例如,糖,包括单糖、二寡糖、三低聚糖、四低聚糖和低聚糖;衍生化糖如醛糖醇、醛糖酸、酯化糖等;和多糖或糖聚合物),它们可以单独或组合存在,包含占其重量或体积1-99.99%的单独或组合。示例性蛋白质赋形剂包括血清白蛋白,例如人血清白蛋白(HSA)、重组人白蛋白(rHA)、明胶、酪蛋白等。具有代表性的氨基酸组分也可起到缓冲能力,包括丙氨酸、精氨酸、甘氨酸、精氨酸、甜菜碱、组氨酸、谷氨酸、天冬氨酸、半胱氨酸、赖氨酸、亮氨酸、异亮氨酸、缬氨酸、蛋氨酸、苯丙氨酸、阿斯巴甜等。碳水化合物赋形剂也意在该技术的范围内,其实例包括但不限于单糖如果糖、麦芽糖、半乳糖、葡萄糖、D-甘露糖、山梨糖等;双糖,如乳糖、蔗糖、海藻糖、纤维二糖等;多糖,如棉子糖、蜜糖、麦芽糊精、葡聚糖、淀粉等;和醛糖醇,如甘露醇、木糖醇、麦芽糖醇、乳糖醇、木糖醇山梨醇(glucitol)和肌醇。

[0124] 根据本公开使用的组合物和药物制剂可以包装成剂型单位形式,以便于施用和剂量的均匀性。术语“单位剂量”或“剂量”可以指适合在受试者中使用的物理离散单位,每个单位含有预定量的组合物,计算成产生与其施用相关的所需反应,即适当的途径和方案。根据治疗次数和单位剂量,施用量取决于期望的结果和/或保护。组合物的精确数量也取决于从业者的判断,并且因人而异。影响剂量的因素包括受试者的身体和临床状态、施用途径、预期的治疗目标(缓解症状与治愈)以及特定组合物的效力、稳定性和毒性。在配制时,溶液可以以与制剂剂量相容的方式施用,并且以治疗或预防有效的量施用。制剂容易以各种剂型施用,例如本文所述的可注射溶液的类型。

实施例

[0125] 人类受试者。受试者是通过皮特霍普金斯研究基金会招募的志愿家庭的成员。PTHS受试者(表1)是根据详细的临床和分子诊断信息的可用性选择的,包括他们携带的TCF4突变的类型。对于携带点突变、小插入缺失或易位的患者,通过对TCF4位点进行重测序来确认每个TCF4突变的细节。所有参与的家庭都回答了一份详细和个性化的问卷,以收集与患者PTHS临床症状相关的信息,包括有关神经病学结果、认知、行为和胃肠病学临床表现、诊断年龄、一般生活质量、动作发展指标的时间演变、沟通水平、畸形面部特征、泌尿系统症状、视力问题、感觉反应、睡眠障碍、呼吸异常(例如呼吸暂停和过度换气)、饮食习惯和肠道症状、癫痫发作史以及MRI结果。这些数据如表1所示。为了最大限度地提高可比性,本研究仅选择男性受试者。对照对象是相应患者的父亲,他们没有精神或遗传病史。所有受试者的参与都得到了进行该研究的机构的人类受试者伦理委员会的批准。在收到对研究的

全面描述后,获得了所有参与家庭的书面知情同意。重要的是要注意,TCF4 (转录因子4) 不应与TCF-4 (T细胞因子4) 混淆,TCF-4是TCFL7的旧的而过时的名称,TCFL7是TCF-LEF蛋白之一,是Wnt信号通路的终点,与PTHS患者中突变的TCF4完全无关。

[0126] 表1:参与受试者和临床特征总结。与所有附图相关。

[0127]

患者	附图中符号	性别	年龄	TCF4 突变类型	PTHS 临床特征	abMRI
PTHS #1	◆	M	14	小插入 (c.1067_1068insTC)	FG MM CT SZ BA VA	是
PTHS #2	■	M	8	全基因缺失 (del 6.6 q21.2-q21.33)	FG MM AS CT VA RB	是
PTHS #3	▲	M	10	易位 [t(2;18)(chr2:197,989,212-197,989,746_chrl8:52,895,297-52,895,597)]	FG SM AS CT VA RB	否
PTHS #4	●	M	4	点突变 (c.959C>T; Thr320Ile)	FG SM AS CT VA	否
PTHS #5	+	M	7	部分基因缺失 (c.454_1072del)	FG SM AS SZ RB	是
PTHS #6		F	7			
亲本#1	◆	M	40	-	-	-
亲本#2	■	M	30	-	-	-
亲本#3	▲	M	50	-	-	-
亲本#4	●	M	44	-	-	-
亲本#5	+	M	45	-	-	-

[0128] 对于所有患者,对照组是性别匹配的亲本。

[0129] 缩写:FG:畸形面部(dysmorphic facial)特征(gestalt);SM/MM:3岁时重度(SM, severe)或轻度(MM, mild)运动延迟(motor delay);AS:言语缺陷(absent speech);CT:便秘(constipation);SZ:癫痫发作(seizures);BA:呼吸问题(breathing problems)(过度换气或呼吸暂停);UA:泌尿异常(urinary abnormalities)(尿滞留或尿失禁);VA:视觉异常(visual abnormalities)(眼部异常);RB:重复行为(repetitive behaviors);abMRI: MRI检测到的脑异常(brain anomalies)(胼胝体变薄)。

[0130] 将皮肤成纤维细胞重编程为诱导多能干细胞(iPSCs)。从PTHS和对照受试者的活检中获得皮肤成纤维细胞,然后在含有10%胎牛血清和青霉素/链霉素的DMEM/F12培养基中培养。如(Marchetto等人,2017)所述,iPSCs通过细胞重编程从成纤维细胞中获得。简而言之,用含有OCT4、SOX2、KLF4和MYC过表达盒的仙台病毒转导成纤维细胞培养物(Cytotune iPS 2.0仙台重编程试剂盒;Thermo Fisher Scientific)。转导七天后,将细胞重新接种到由DMEM/F12中的小鼠胚胎成纤维细胞(mEF)组成的饲养层上,该饲养层含有20% Knockout血清替代物(Thermo Fisher Scientific)、1%非必需氨基酸(NEAA)和100μMβ-巯基乙醇。2周后鉴定iPSC克隆,并将其转移到涂有Matrigel® (BD Biosciences)的6cm平板中,之后将其保持在mTeSR1培养基(StemCell Technologies)中,并借助移液器吸头通过手动挑取进行传代。研究中的每个受试者共产生了20个iPSC系,所有这些系都通过免疫染色和SNP定位的组合进行分析,以排除不需要的染色体异常和突变的存在(图12B中的示例)。所有iPSC克隆传代直至P10,选择此次传代后的2个克隆进行进一步NPC和类器官衍生。本研究中的大多

数实验都是每个受试者使用一个P15 iPSC克隆进行的。每两周对培养物进行一次支原体检测,在任何阶段从未发现污染。

[0131] 通过SOX2、OCT4、NANOG和LIN28的免疫染色对iPSC进行验证。简而言之,在LabTek II 8孔腔室载玻片(Thermo Fisher Scientific)的孔内培养总共20个克隆,直到它们尺寸达到2mm。然后用4%多聚甲醛溶液固定克隆10分钟,用1×磷酸盐缓冲盐水(PBS)洗涤一次,用1% Triton X-100透化5分钟,在1×PBS中再次洗涤,并用10%牛血清白蛋白(BSA)/1% Triton X-100/1×PBS封闭。在同一封闭液中与一抗在4℃孵育16小时。使用的一抗是抗-SOX2(Abcam;ab97959)、抗-OCT4(Abcam;ab19857)、抗-NANOG(GeneTex;GTX100863)和抗-LIN28(Cell Signaling;3978)。在1×PBS中洗涤3次后,将克隆与荧光标记的二抗孵育3小时,并将细胞核用1μg/mL DAPI(Thermo Fisher Scientific)复染30分钟。载玻片用ProLong Gold抗猝灭封片剂(Thermo Fisher Scientific)封装。

[0132] 为了鉴定不需要的染色体结构改变,使用iScan系统(Illumina)和Infinium HumanCytoSNP-12BeadChip(Illumina;299,140个遗传标记物)对从iPSC系中提取的基因组DNA进行扩增、缺失、拷贝数变异和重排的全基因组分析。未发现包含明显较大缺失和重复的克隆。图12B给出了使用该技术为PTHS患者#2进行的核型分析的示例,显示该患者系的18号染色体长臂上预期的大部分缺失。

[0133] 皮质(pallial)和皮质下(subpallial)类器官的生成。为了产生皮质(皮质(cortical))脑类器官(Ct0),使用Accutase(Thermo Fisher Scientific;用等体积1×PBS稀释)在37℃下分离iPSC克隆12分钟。在以150×g离心3分钟后,将单细胞重悬于补充有10mM SB431542(Stemgent)和1mM dorsomorphin(R&D Systems)的mTeSR1培养基(StemCell Technologies)中。将大约3-4百万个细胞接种到低结合6孔板的每个孔中,并放置在CO₂培养箱中95rpm速度的摇床上。在前24小时内,培养基补充5mM Rho激酶抑制剂(Y-27632;Calbiochem,Sigma-Aldrich)。在三天内,细胞聚集形成球形拟胚体,之后mTeSR1被替换为含有神经基础培养基(Thermo Fisher Scientific)的神经诱导培养基,神经基础培养基含有GlutaMAX、1% Gem21 NeuroPlex补充剂(Gemini Bio-Products)、1% N2NeuroPlex(Gemini Bio-Products)、1% NEAA(Thermo Fisher Scientific)、1% 青霉素/链霉素(Thermo Fisher Scientific)、10mM SB431542和1mM dorsomorphin,培养7天。接下来,将培养基替换为NPC增殖培养基,该培养基由含有GlutaMAX、1% Gem21、1% NEAA、20ng/mL FGF-2(Thermo Fisher Scientific)的神经基础培养基组成,培养7天,然后在同一培养基中再补充20ng/mL EGF(PeproTech)后再培养7天。通过更换到含有1% GlutaMAX、1% Gem21、1% NEAA、10ng/mL BDNF、10ng/mL GDNF、10ng/mL NT-3(均来自PeproTech)、200mM L-抗坏血酸和1mM二丁酰-cAMP(Sigma-Aldrich)的神经基础培养基,培养7天,实现了神经元分化和类器官成熟。在此期间之后,根据需要将Ct0维持在含有GlutaMAX、1% Gem21、1% NEAA的神经基础培养基中,每3-4天更换一次培养基。对于每个受试者,大部分实验至少使用3个独立批次(通常超过10个批次)进行,在整个研究的数据中被认为是独立的生物学重复,每个批次至少3个技术重复(类器官的孔)。对于在4个或更多个独立批次上进行的表型评估,使用两个或多个独立的iPSC克隆来产生类器官(和NPC)并确认基因型的影响,如图12F所示。

[0134] 对于皮质下类器官(sP0)的生成,使用了先前发表在(Birey等人,2017)中的方案,

并进行了一些修改。在mTeSR1中培养拟胚体三天后,从第4天到第10天,将它们转移到含有5 μ M Wnt通路抑制剂IWP-2 (SelleckChem)的神经诱导培养基中(补充有1% GlutaMAX、1% Gem21、1% N2、1% NEAA、1% 青霉素/链霉素、10mM SB431542和1mM dorsomorphin的神经基础培养基)。接下来,用NPC增殖培养基替换上述培养基,该培养基由含有1% GlutaMAX、1% Gem21、1% NEAA、20ng/mL FGF-2和100nM SHH通路激动剂SAG (SelleckChem)的神经基础培养基组成,培养7天,然后在补充有20ng/mL EGF (PeproTech)的同一培养基中再培养2天。在没有SAG的同一培养基中再培养5天,完成了NPC增殖阶段。随后是神经元分化和类器官成熟阶段,其使用与Ct0衍生方案中相同类型的培养基和持续时间进行。

[0135] 免疫荧光染色。在体外培养所需时间后,将Ct0和sP0用4%多聚甲醛在4℃下固定4-8小时,并在30%蔗糖中冷冻保护12小时。然后将类器官嵌入TissueTek (Leica Microsystems)中,并在Leica VT1000S低温恒温器上切片以产生20 μ m切片。为进行染色,将载玻片风干10分钟,在1% Triton X-100/1 $\times$ PBS中透化2分钟,并用0.1% Triton X-100/3% BSA/1 $\times$ PBS在25℃下封闭1小时,然后在同一溶液中与一抗在4℃下孵育16小时。使用的一抗是:大鼠抗-CTIP2 (Abcam; ab18465; 1:500)、兔抗-SATB2 (Abcam; ab34735; 1:200)、鸡抗-MAP2 (Abcam; ab5392; 1:1000)、兔抗-SOX2 (Cell Signaling Technology; 2748; 1:500)、兔抗-GAD65/67 (Abcam; ab11070; 1:200)、兔抗-CUX1 (CUTL1或CASP) (Abcam; ab54583; 1:200)、兔抗-TCF4 (Abcam; ab217668; 1:1000)、兔抗-vGLUT1 (Synaptic Systems; 135311; 1:500)、兔抗-CC3(裂解半胱天冬酶3) (Cell Signaling; 9664S; 1:500)、兔抗-双皮质素(DCX) (Abcam; ab18723; 1:200)、小鼠抗-Cas9 (Abcam; ab210571; 1:200)、小鼠抗-p16^{INK4a} (CDKN2A) (Abcam; ab54210; 1:1000)、兔抗-SOX3 (Abcam; ab183606; 1:200)、小鼠抗-巢蛋白 (Abcam; ab22035; 1:1000)、山羊抗-SOX17 (R&D Systems; AF1924; 1:200)、兔抗-Brachyury (Sigma; B8436; 1:200)、或兔anti--catenin (Cell Signaling; 9582S; 1:100)。在含有一抗的溶液中孵育后,将载玻片在1 $\times$ PBS中洗涤3次,每次5分钟,并在与一抗相同类型的溶液中与荧光标记的二抗(Alexa Fluor 488-或555-偶联抗体; 1:500稀释; Thermo Fisher Scientific)在25℃下孵育3小时。在1 $\times$ PBS中进一步洗涤后,将载玻片用DAPI溶液(1 μ g/mL)复染45分钟并用ProLong Gold固定。所有图像均使用配备Apotome的蔡司(Zeiss)荧光显微镜(Axio Observer Apotome, 蔡司)拍摄。在每系列收集10个光学切片后,使用ZEN软件(蔡司)的最大密度特征预测DCX-染色类器官的z系列堆叠图像。对于p16^{INK4a}染色,将载玻片在1 $\times$ Universal HIER抗原修复试剂 (Abcam; ab208572) 中在60℃孵育10分钟来进行抗原修复,然后进行常规免疫染色。为了在p16^{INK4a}共染色后计数SOX2+细胞(图15J),使用原始未处理的图像,并将强染色细胞定义为每个图像中平均像素强度介于上第三个四分位数和最大像素强度之间的细胞。剩余的SOX2+细胞被认为弱染色。

[0136] 为了定量类器官切片中的细胞类型,在每个成像切片中随机采样4个100 $\times$ 100 μ m关注区域(regions of interest, ROI)。每个样品的标记细胞的平均数量首先通过平均每个ROI中的标记细胞数量以产生每个切片的标记细胞的平均值,然后平均每个受试者的所有切片中的这些平均值来计算的。受试者和切片数量的量化在图例中表示。由于vGLUT1主要存在于细胞体外,因此对vGLUT1和GAD65/67的量化(图12M)通过使用ImageJ软件上的彩色像素计数器插件(Color Pixel Counter plugin)对原始未处理荧光显微镜图像中的像素进行计数,计数大小为1像素且颜色密度高于50(范围为0至255)的颗粒。根据这些规则,

像素的平均百分比是通过每个切片大于四个 $100 \times 100 \mu\text{m}$ ROI和每个受试者6个切片计算的。

[0137] 对于NPC的免疫荧光标记,将这些细胞以每孔50,000个细胞的密度接种在LabTek II 8孔腔室载玻片上。当细胞聚集至50%时,用与iPSC克隆相同的方式固定并处理它们以进行免疫染色,使用以下一抗:兔抗-TCF4 (Abcam;ab217668;1:1000) 和鸡抗-波形蛋白(VIM) (Abcam;ab22651;1:2000)。还对NPC进行染色,在如上对抗原修复后,使用CellEvent™衰老绿色检测试剂盒(Thermo Fisher Scientific;C10850)检测 β -半乳糖苷酶(SA- β -gal)相关的衰老。将上述用于计数p16^{INK4a}共染色后弱染色和强染色SOX2+细胞的相同方法应用于NPC。

[0138] 死后脑样本采集与分析。患者#6(表1)在7岁时死于矫正脊柱侧弯的外科手术,原因是与PTHS神经系统症状无关的并发症。医院病理学家立即解剖了大脑并获取皮质组织,其包括前运动和前额叶区域交界处的整个皮质宽度。海马组织也被获取,但在本研究中没有描述。在用20%蔗糖冷冻保护并在振动切片机下切片,然后如上所述进行免疫染色之前,脑组织在福尔马林下固定24小时,然后在4%多聚甲醛中固定6h。将PTHS图像与使用来自商业来源(NOVUS;NBP2-77523)的正常脑组织的平行染色的切片(图10E)进行比较,其取自一名没有疾病或神经病理学体征的12岁男性。由于PTHS特征性紊乱的解剖结构,PTHS脑组织无法使用分层结构作为组织内部定位的指标,通过与从皮质表面等效深度(以mm为单位)的关注区域(ROI)收集的匹配图像进行比较。病理学家在解剖前在皮质组织的脑回和宽度的总体外观方面没有观察到显著差异。

[0139] 类器官单细胞RNA测序分析。通过结合镊子的机械分离和Accutase的酶消化,将Ct0和sP0分离10分钟产生单细胞悬液。对于每个文库,总共分离15个类器官,并将所得细胞汇集并随后过滤以分离单个细胞,以便在同一天进行RNA测序分析。将分离的细胞沉淀(3分钟, $100 \times g$,4℃)并重悬于10mL神经基础培养基中。使用Chemometec自动细胞计数器测定每个文库中单个细胞的浓度,发现所有文库中的最小群体活力为85%。根据制造商的方案,使用Chromium Single Cell 3'v3文库试剂盒($10 \times$ Genomics)制备单细胞RNA-seq文库。每个样品在Chromium芯片上上样了大约20,000个细胞。所有步骤,包括GEM(Gel beads in emulsion,乳液中凝胶珠)制备、逆转录、PCR扩增和Illumina文库构建均在T100热循环仪(Bio-Rad)上进行。从GEM中提取的cDNA使用MyOne硅烷珠(Thermo Fisher Scientific)纯化,PCR扩增共10个循环,然后使用SPRIselect试剂盒(B23317,Beckman Coulter)纯化。接下来,对每个文库的cDNA库进行酶切,并使用SPRIselect试剂盒进行两次片段大小选择。最后,连接Illumina接头以制备用于测序的文库,然后使用SPRIselect试剂盒进行另一轮两次片段大小选择。最终文库片段大小范围为300-700bp,平均片段大小约为450bp。使用Qubit dsDNA HS检测试剂盒(Thermo Fisher Scientific)对Illumina文库进行定量,并在高灵敏度D1000 tapestation(Agilent)上进行片段大小质量控制。文库在NovaSeq 6000S4测序仪(Illumina)上测序,每个细胞产生20,000个读数(reads),或每个文库4亿个读数,其中read_1进行26个循环,索引(index)进行8个循环,read_2进行98个循环,其包含基因序列。

[0140] 使用Cell Ranger(版本4.0.0)软件中的“cellranger count”命令和人类转录本的GRCh382020-A参考数据集分别生成每个单细胞RNA-Seq文库的特征计数矩阵。然后将独

立文库标准化为相同的测序深度,并使用“cellranger aggr”命令聚合到单个特征-条形码(feature-barcode)矩阵中。通过自动注释和精选手动检查的组合来描述细胞类型亚群。首先,将处理后的数据传输到Cell Loupe软件(10×Genomics),并使用k均值(k-means)聚类分析到8组的单个细胞分区组。然后在Cell Loupe分配的每个亚群中目视检查标记基因的表达。接下来,分析图13B中各标记基因的表达情况。并根据这些基因的表达模式手动调整为k均值分配的亚群。首先对亚群进行无偏定量,然后手动管理的组合方法可以最大程度地鉴定生物学相关的细胞组。很明显,这些亚群可以进一步细分为其他细胞组,但决定将重点放在单细胞数据所示的兴奋性和抑制性谱系中含有祖细胞、中间祖细胞和神经元的组(图6A-J)。存在其他次要亚群,但未显示在大多数面板中,因为重点是上述6个亚群。其他群体在图13A-B中被统称为“其他”,并进一步分析得出的结论是,该类别中的细胞数量在亲本和PTHS之间没有差异,并且它们不是非神经起源的细胞,它们在Ct0或sP0中不以显著的数量存在(图13E)。线粒体基因被用作识别凋亡细胞的指标,凋亡细胞在所有文库中通常很少见(小于5%)。

[0141] 接着,将Seurat文库(版本3.2.2)(Butler等人,2018)用于特征-条形码矩阵的下游处理和分析。首先,将Cell Ranger生成的聚合矩阵导入Seurat,并通过将每个细胞的特征计数除以该细胞的总数进行归一化,然后在执行对数转换之前将数据缩放到每个单元格10,000个计数(“NormalizeData”函数)。接下来,使用“FindVariableFunctions”函数识别可变特征,该函数将多项式曲线拟合到均值-方差关系,根据给定表达式的预期方差标准化特征计数,并选择方差最高的3,000个特征。然后将高度可变的特征缩放到细胞间的平均表达为0和方差为1的分布(“ScaleData”函数),随后用于执行线性降维(PCA,“RunPCA”函数)。前15个PCA维度用于使用UMAP算法(“RunUMAP”函数)将细胞嵌入到非线性降维空间中。

[0142] 使用Monocle 3(版本0.2.2)(Cao等人,2019)对兴奋性和抑制性谱系独立执行无监督轨迹(伪时间)推断(图6B)。具体来说,使用Leiden方法在UMAP嵌入(“cluster_cells”函数)中进行细胞聚类,然后将代表分化轨迹的未分区主图拟合到数据(“learn_graph”函数)。最后,通过将轨迹根源到手动注释的祖细胞亚群(“order_cells”函数)来对细胞进行排序。伪时间是细胞与轨迹起点之间的转录距离(抽象单位),沿最短路径测量。

[0143] Cell Loupe软件用于量化每个亚群和文库中细胞的百分比(图6D、6G、13F、13G、16H和17K)。为了计算表达某种基因的细胞的百分比,使用Seurat和R包来计数表达该基因高于对应于被比较的每组中基因表达平均值的40%的阈值水平的细胞数,(图6E、6H-J、7F、13H、13I和17L)。对于特定亚群之间基因表达水平的统计比较,使用Mann-Whitney U检验(2组)或Kruskal-Wallis检验,然后使用Dunn事后(Dunn's post-hoc)检验(超过2组和成对比较)。

[0144] NPC衍生和神经元分化。将维持在mTeSR1培养基中的iPSC克隆转移到含有N2和GEM21补充剂的DMEM/F12培养基(StemCell Technologies)中。2天后,用Accutase从平板上提取菌落,并在相同培养基中培养,加入10mM SB431542和1mM dorsomorphin,在平台摇床上悬浮培养,直到形成胚状体。以这种方式培养2周后,将胚状体直接接种到Matrigel包被的培养皿上,并保持在含有N2和SM1补充剂(StemCell Technologies)、20ng/mL FGF-2和1%青霉素/链霉素的DMEM/F12培养基中。3至5天后,出现玫瑰结(rosette),7天后,玫瑰结被手动挑取并重新铺到Matrigel涂层的培养皿上。NPC在玫瑰结周围生长,与Accutase分离

5分钟,然后重新接种到涂有10 μ g/mL聚鸟氨酸(Sigma-Aldrich)和5 μ g/mL层粘连蛋白(Thermo Fisher Scientific)的平板上以产生传代1(P1)。将NPC维持在含有N2和SM1补充剂、20ng/mL FGF-2和1%青霉素/链霉素的DMEM/F12培养基中培养到20代。培养物不是在含有Wnt或Shh激动剂/拮抗剂(如环巴胺)的培养基中衍生的,因为用人工高浓度的上述物质处理祖细胞可能会影响细胞的增殖速率,从而可能在评估NPC增殖时增加混淆因素。

[0145] 对于神经元分化,将NPC接种到涂有聚鸟氨酸和层粘连蛋白的平板上并在NPC培养基中培养至聚集至90%,此时将培养基改为含有N2和SM1补充剂以及1%青霉素/链霉素的DMEM/F12,培养基每3至4天更换一次。当神经元突起在一周后开始生长时,将培养基更改为BrainPhys神经元培养基(StemCell Technologies),并且将细胞在上述条件下保持长达4个月,每3至4天更换一次培养基。如图7D-E,在BrainPhys培养基中3或4个月后对神经元培养物进行电生理测量,此时培养物中的绝大多数细胞为MAP2+(亲本为95.4 $\pm$ 2.4%,PTHS为93.2 $\pm$ 1.4%; $P=2.4$;未配对的Welch's t -检验)。

[0146] 神经元分化率的定量(图10J、10K和17J)是通过在BrainPhys培养基中分化2个月后接种到LabTek II腔室载玻片上的分化神经元培养物中计数MAP2+和SOX2+细胞来完成的,然后如前所述进行免疫荧光染色。

[0147] 对NPCs和神经元培养物进行RNA测序。使用RNeasy Mini Plus试剂盒(Qiagen),从4名受试者和4个各自的亲本对照的NPC的第15代中分离出RNA,用于大多数分析,从2名受试者和2个各自的对照的NPC的第5代中分离RNA,用于图15H中的分析,从BrainPhys培养基培养2个月后分化的神经元培养物中(1名受试者和各自的亲本对照)进行FACS分选以纯化CD184+/CD44-/CD24+群体(图7G、14C、14F、14G和18L)。对于每个受试者,从3个独立制备的生物学重复中提取RNA。使用链式TruSeq试剂盒(Illumina)从每个样品中总共使用1 μ g RNA进行Illumina文库制备。RNA在Illumina NovaSeq 6000S4仪器上测序,其具有150bp配对末端读数,每个文库产生约4000万个测序片段。

[0148] 为了从大量RNA-Seq数据估计转录水平的表达,使用了Salmon(版本0.14.1)软件(Patro等人,2017),其具有选择性映射('--validateMappings')和序列特异性偏差校正('--seqBias')、GC内容偏差('--seqBias')和片段位置偏差('--posBias')。读取映射的参考转录本是从GENCODE 32基本注释中获得的(Frankish等人,2019年)。对于每个样品,异常值由高间-重复欧氏距离(between-replicate Euclidean distances)定义(在转换后达到同方差性,如本文其他地方所述),这导致从后续表达分析中仅排除了PTHS患者#3的一个文库重复。其余41个文库全部通过质量控制阶段并被保留。

[0149] 使用DESeq2(版本1.22.1)对源自PTHS患者的细胞及其各自亲本对照之间进行成对差异表达(DE)测试(Love等人,2014)。tximport(版本1.10.1)(Soneson等人,2015)用于将转录本丰度聚合为基因水平计数。接下来,使用量化因子方法(Anders和Huber,2010)进行样品间归一化,并将局部离散模型拟合到归一化计数。最后,将负二项式广义线性模型拟合到数据中,使用apeglm算法缩小效应量(log2FoldChange)(Zhu等人,2019),并使用基于阈值的Wald检验("lfcThreshold=0.5")完成严格的统计测试。DE转录本根据其s值(<0.005)确定。为了聚类 and 可视化目的,使用设置为"TRUE"的"盲"参数的"方差稳定变换"函数,将计数数据转换为近似同方差的矩阵(图14F和15M)。

[0150] 为了获得所有受试者的DE基因列表,首先推导了每个PTHS受试者与其各自亲本之

间的DE基因列表(表3),然后对4个列表进行交叉检查并选择所有4对亲子对共有的DE基因。然后将最终列表用于基因集富集评估,然后使用基于Web的WebGestalt工具(Wang等人,2017)使用默认参数进行基因本体(Gene Ontology,GO)和通路分析。WebGestalt进行排列以获得每个GO项的过度表示Z分数和富集p值。对于通路分析,使用了带有默认参数的KEGG选项。对于所有分析,每个类别至少使用5个基因,BH多重检验校正,并选择错误发现率的显著性水平为0.05。

[0151] 实时荧光定量PCR。根据制造商的建议,使用RNeasy Mini Plus试剂盒(Qiagen)提取总RNA,然后在色谱柱上进行DNase I处理。使用Superscript III第一链逆转录系统(Thermo Fisher Scientific)将2.5 μ g的总RNA逆转录成cDNA。使用预先验证的FAM-MGB TaqMan探针(Thermo Fisher Scientific)和不含UNG的TaqMan通用预混液II(Thermo Fisher Scientific)在CFX Connect实时荧光定量PCR检测系统(Bio-Rad)上进行实时定量PCR(RT-qPCR),其具有以下循环参数:94 $^{\circ}$ C 3min,然后94 $^{\circ}$ C 30s和68 $^{\circ}$ C 1min,40个循环。分析了所有探针的扩增和变性曲线,以验证仅一个扩增子的扩增。所有RT-qPCR分析均使用从每个受试者/条件至少3个独立生物样本中提取的RNA,并标准化为以下内源性对照基因(TBP、ACTB和GAPDH)。使用传统的 $\Delta\Delta C_t$ 方法计算相对表达量。

[0152] 分别使用了以下TaqMan探针:STMN2(Hs00199796_m1)、TAC1(Hs00243225_m1)、INA(Hs00190771_m1)、SLC17A6(Hs00220439_m1)、CDKN2A(Hs00923894_m1)、LMNB1(Hs01059210_m1)、WNT2B(Hs00921615_m1)、WNT3(Hs00902257_m1)、WNT5A(Hs00180103_m1)、SFRP2(Hs00293258_m1)、ASCL1(Hs00269932_m1)、NEUROD1(Hs00159598_m1)、HES1(Hs00172878_m1)、SOX2(Hs04234836_s1)、SOX3(Hs00271627_s1)、SOX4(Hs00268388_s1)、TCF4(Hs00972432_m1)、CNTNAP2(Hs01034283_m1)、GADD45G(Hs00198672_m1)、MAP2(Hs01103234_g1)、VIM(Hs00185584_m1)、NES(Hs04187831_g1)、ID3(Hs00171409_m1)、KCNQ1(Hs00165003_m1)、BCL11B(CTIP2)(Hs00256257_m1)、SATB2(Hs00392652_m1)、CUX1(Hs00738851_m1)、TBR1(Hs00232429_m1)、CDH23(Hs00254446_m1)和PCDH15(Hs00263709_m1)。

[0153] 神经元形态测量。使用Neurolucida神经元追踪软件(MBF Bioscience)对神经元进行形态学分析(图7C)。从共聚焦图像中鉴定出单个MAP2+神经元,这些图像清楚地显示了从细胞体分支的过程数量,完全根到尖端长度的过程或完整的细胞体。仅计算最短树突至少比细胞体直径长3倍的神经元。评估来自每个细胞系的至少2个克隆的随机图像。“轮廓(contour)”函数用于沿着完整过程的最长路径跟踪和求和每条曲线的增量长度,以得出其总长度。还使用“轮廓”函数追踪细胞体的轮廓,并由软件自动计算生成的表面积。

[0154] 多电极阵列分析。Axion Biosystems的12孔多电极阵列板用于从类器官获取电活动读数。使用含有GlutaMAX、1% Gem21、1% NEAA、10ng/mL BDNF、10ng/mL GDNF、10ng/mL NT-3、200mM L-抗坏血酸和1mM二丁酰基-cAMP的神经基础培养基,根据本文所述的类器官衍生方案,在20天时将6个类器官接种到每个孔中。将其在该培养基中维持7天,然后转移到含有1%谷氨酸MAX、1% Gem21、1% NEAA和0.5%青霉素/链霉素的神经基础培养基另外培养7天。在此时间范围之后,将种子类器官保存在BrainPhys培养基中,直到测量时间。每个受试者至少进行2次独立实验,每个实验中每个受试者进行3次独立重复。从转移到BrainPhys培养基后7天开始评估类器官的电生理参数。图7A中报告的数据来自在

BrainPhys培养基中培养30天的类器官。图11G中报告的数据来自在BrainPhys培养基中培养长达90天的类器官。

[0155] 使用Maestro系统和AxIS软件(Axion Biosystems)进行记录,其具有从10Hz到2.5kHz的带宽滤波器。使用自适应阈值计算脉冲检测,该阈值是每个电极估计噪声标准偏差的5.5倍。在记录之前,将板在Maestro仪器中保持3分钟不动,再记录3分钟。使用Axion Biosystems神经计量工具分析数据,条件是如果电极在1分钟内至少出现5个脉冲(最少5个脉冲/分钟),则电极被视为是活跃的。在受试者的所有孔中的活跃电极上计算受试者的平均放电速率。网络放电(network burst)被定义为在孔中超过25%的活跃电极中发生的超过10个脉冲的放电,最大脉冲间隔为100毫秒。

[0156] 膜片钳电生理分析。在二维(单层)培养物中对神经元进行全细胞膜片钳记录,神经元为在FGF-2停用后,在涂有聚鸟氨酸和层粘连蛋白的35mm培养皿上从NPC分化4个月而来。在所有平板中都实现了相似的神元密度。细胞外溶液为130mM NaCl、3mM KCl、1mM CaCl₂、1mM MgCl₂、10mM HEPES和10mM葡萄糖,用1M NaOH调节pH为7.4(添加约4mM Na⁺)。玻璃电极的内部溶液为138mM K-葡萄糖酸盐、4mM KCl、10mM Na₂-磷酸肌酸、0.2mM CaCl₂、10mM HEPES(Na⁺盐)、1mM EGTA、4mM Mg-ATP、0.3mM Na-GTP,用1M KOH调节pH为7.4(添加~3mM K⁺)。将所有溶液的渗透压调节至290mOsm。将长丝硼硅酸盐玻璃毛细管(1.2mm OD, 0.69mm ID, World Precision Instruments)拉在Flaming/Brown拉针仪(micropipette puller)(型号P-87, Sutter Instrument)上。全细胞记录的电极电阻为4-6MΩ。Axon CV-4探头和Axopatch 200A放大器(Molecular Devices)用于室温下的电生理记录。对于诱发AP记录,采用电流钳配置,注入小电流以将膜电位保持在-70mV。然后,使用电压钳配置记录电压依赖性神经Na⁺和K⁺电流。记录以1kHz进行低通滤波,并使用DigiData 1322A(Molecular Devices)以10kHz进行数字化。液接电位为零。使用pCLAMP 10软件(Molecular Devices)离线分析电生理学数据。使用双尾Welch's t检验进行统计比较,每组10个(亲本)或9个(PTHS)神经元,显著性阈值为p=0.05。

[0157] 增殖和凋亡测定。为了通过细胞计数定量细胞增殖,在12孔聚鸟氨酸/层粘连蛋白包被的平板上每孔接种100,000个细胞。每个受试者至少进行2次实验,每个实验每个受试者进行3次技术重复。在指定的天数后,取出细胞通过Accutase处理5分钟,重悬于等体积的DMEM/F12中,并使用Chemometec Via-1盒计数,其还计数总活细胞。

[0158] 对于EdU细胞周期测定,使用Click-iT EdU流式细胞术测定试剂盒(Thermo Fisher Scientific),遵循制造商的方案。简而言之,将10cm培养皿中聚集70%的NPC用Accutase分离,重悬于StemDiff神经祖细胞培养基(StemCell Technologies)中,并以 0.2×10^6 个细胞/孔的密度接种到基质胶包被的6孔板上。将细胞在37℃和5% CO₂下孵育12h,然后将EdU以终浓度为10μM加入培养基中。将细胞再孵育2.5h以使EdU掺入,随后通过Accutase介导的分离收获,重悬于3mL含1%BSA的1×PBS中,并在500×g下沉淀5分钟。将沉淀重悬于试剂盒的固定溶液中并在25℃避光孵育15min,然后加入3mL含1%BSA的1×PBS以停止固定。接下来,将NPC在500×g下沉淀5min,除去上清液,并将沉淀的细胞在1×基于Click-iT皂苷的透化和洗涤试剂中孵育15min。在孵育期间,根据制造商的方案制备Click-iT反应混合物,然后加入样品中,然后匀浆并孵育30min,避光。细胞每5min重新匀浆一次,然后在3mL的1×Click-iT透化和洗涤试剂中洗涤,沉淀并重悬于同一溶液中,然后在含有

20 μ g/mL碘化丙啶的1 $\times$ PBS/0.1% Triton X-100/100 μ g/mL RNase A溶液中进行核染色。核染色后,立即将细胞转移到冰上并保持在4 $^{\circ}$ C,避光,直到在LSR Fortessa X-20细胞细胞仪(BD Biosciences)中进行分析。

[0159] NPC的凋亡测定使用带有膜联蛋白V FITC和PI的死细胞凋亡试剂盒(V13242, Thermo Fisher Scientific)进行,遵循制造商的方案,并在上述同一仪器中通过流式细胞术进行分析。

[0160] TOP-Flash荧光素酶报告基因Wnt功能测定。为了评估Wnt信号水平,用M50 Super8 $\times$ TOPFlash质粒(Addgene#12456;[<http://n2t.net/addgene:12456>;RRID:Addgene_12456;称为TOP-Flash荧光素酶报告质粒)转染24孔板中聚集70%的NPC的培养物(使用来自StemCell Technologies的神经祖细胞培养基),用于评估 β 连环蛋白介导的转录激活。该质粒含有最小的TA病毒启动子,驱动萤火虫荧光素酶基因的表达,该基因前面有七个TCF/LEF结合位点(AGATCAAAGG;SEQ ID NO:1)(Veeman等人,2003),不要与TCF4混淆。对照NPC转染M51 Super8 $\times$ FOPflash质粒,该质粒具有突变的TCF/LEF结合位点(Addgene质粒#12457;[<http://n2t.net/addgene:12457>;RRID:Addgene_12457)。

[0161] 使用Amaxa核转染小鼠神经干细胞核转染试剂盒(Lonza),根据制造商的建议对NPCs进行转染。24h后,补充培养基,并使用Pierce Firefly荧光素酶快速测定试剂盒(Thermo Fisher Scientific)对50,000个细胞的样品使用Synergy酶标仪(BioTek Instruments)进行荧光素酶测定。所有测定均在每个NPC系(每个受试者)进行3个独立重复和3个技术重复。活动水平表示为根据各自对照中的平均活动归一化的任意单位。

[0162] Wnt信号操纵。为了操纵NPC中的Wnt/ β -连环蛋白信号通路,将200,000个细胞接种在6孔板上,然后用特异性激动剂CHIR99021(1 μ M)处理4天。对照用相同浓度的DMSO(CHIR稀释剂)处理相同持续时间。在单独的实验中,用Wnt信号拮抗剂DKK-1(25 μ M)或ICG-001(1 μ M)处理细胞3至5天。在所有情况下,通过转染本文所述的TOP-Flash质粒测定处理过的细胞以测量Wnt通路的活性。对于所有实验,每个主题行使用3个生物学重复,并且在至少3个独立实验中获得了相似的结果。

[0163] 在Ct0或sP0祖细胞增殖阶段的第一天(当FGF-2首次添加到生长的类器官中时)用1 μ M CHIR99021(或DMSO,作为对照)处理,培养基类型与未处理的类器官相同。同样,在祖细胞增殖阶段的第一天用Wnt拮抗剂ICG-001(1 μ M)进行处理。在所有情况下,对每个类器官系的至少6个独立重复进行上述处理。

[0164] 敲低TCF4、SOX3和SOX4。对于NPC中的TCF4和SOX3敲低,根据制造商的建议使用Amaxa核转染小鼠神经干细胞核转染试剂盒(Lonza)将shRNA Mission质粒(Sigma Millipore)转染至100,000个细胞。使用了在pLKO.1质粒骨架(TRC2系列)中制备的SHCLND-NM-005834(SOX3)和SHCLND-NM_003199(TCF4)预验证的MissionshRNA载体(Sigma Millipore)。以SHC201空TRC2载体为对照。转染后4天,对细胞进行计数,并使用RNeasy Mini Plus试剂盒(Qiagen)提取RNA,然后通过RT-qPCR分析所选基因的基因表达。每个实验中每个受试者/条件3个重复进行细胞计数,或3个独立重复进行RNA提取和基因表达评估。由于转染后未进行选择,因此观察到的SOX3或TCF4敲低对其他基因表达的影响应解释为转染群体中所有细胞的平均变异。这可以解释例如NPC中的TCF4敲低导致SOX3表达降低(图10C),其效应量小于对照和PTHS NPC样品比较中观察到的效应量(图10B)的原因。

[0165] 对于神经元中的SOX4敲低(图8I-J),避免了需要转染正在分化的神经元培养物的方法,因为这会导致表型改变、细胞死亡和细胞密度变化。因此,使用反义寡核苷酸(ASOs)方法,其中两种在所有实验中组合使用。ASO是使用制造商的设计工具设计的,作为反义16核苷酸长锁核酸(locked nucleic acid,LNA)寡核苷酸(Qiagen),其在侧翼区域富含LNA,但在无LNA的中心间隙(GapmeRs)中含有规则的DNA核苷酸。将每种ASO重悬于10mM Tris pH 7.5/0.1mM EDTA中,并以1μM终浓度使用。停用FGF-2两周后,在FGF-2停用后第15、20和25天用ASO处理正在分化的神经元培养物,通过直接将其施加到培养基进行无辅助摄取(gymnosis),。在最后一次用ASO处理后三天固定或收获培养物以进行RNA提取。

[0166] SOX3过表达。对于NPC中的SOX3过表达(图17F-G),使用Amaxa核转染小鼠神经干细胞核转染试剂盒(Lonza)和1.5μg pENTER-CMV-SOX3质粒转染100,000个细胞(Vigene Biosciences;CH850241),其中SOX3编码序列由巨细胞病毒(CMV)启动子控制。转染后4天,计数细胞并使用RNeasy Mini Plus试剂盒(Qiagen)提取RNA,然后如上所述通过RT-qPCR分析选定基因的基因表达。每个实验中每个受试者/条件3个重复进行细胞计数,或3次独立重复进行RNA提取。

[0167] TCF4过表达。在通过内源性位点的反表观遗传激活测试CRISPR介导的TCF4表达增强之前,通过在对照组和PTHS NPC中转染置于人工启动子控制之下的TCF4-B转录本变体编码序列的表达盒来测试过表达TCF4的效果(图11E和18J)。对于对照条件,编码序列由人工minP启动子(AGAGGGTATATAATGGAAGCTCGACTTCCAG;SEQ ID NO:2)控制。其他构建体包含TCF4-B编码序列,该序列前有minP启动子和不同数量(6或12)的μE5 TCF4调节性DNA结合位点(CACCTG),这些位点由CAAGAA组成的间隔序列分离。这些构建体是通过基于PCR的反应制备的,以将包含人工启动子的超聚寡核苷酸(minP_TCF4,E-box-x6-minP_TCF4或E-box-x12-minP_TCF4;Integrated DNA technologies)连接至TCF4-B编码序列,TCF4-B编码序列分别从人脑cDNA(Promega)中使用引物TCF4B_CDNA正向和TCF4B_cDNA反向通过RT-PCR扩增。将所得PCR片段克隆到pLenti-III无启动子载体(Applied Biological Materials)的EcoRI和XhoI限制性位点中。使用本文所述的方案用这些质粒转染NPC,然后如前文所述使用RNeasy Mini Plus试剂盒(Qiagen)和RT-qPCR提取总RNA。

[0168] 表2:试剂

[0169]	寡核苷酸	SEQ ID NO:
	gRNA 序列: 3bS1: CTTTATAAGCCCGCAGTTCC	11
	gRNA 序列: 3bS2: CCGCAGTTCCCGGATGTGAA	12
	gRNA 序列: 3bS3: GTCGACCAGCACCGCCATCT	13
	gRNA 序列: 3bA1: GGTAACAGAGCGCCTAGAG	14

[0170]

gRNA 序列: 3bA2: CATTACATCCGGGAACTGC	15
gRNA 序列: 8aS1: ACATAGGAAGGTACGACTTC	16
gRNA 序列: 8aS2: TTTACGTACCAGACATAGGA	17
gRNA 序列: 8aS3: TTCCTTTACGTACCAGACAT	18
gRNA 序列: 8aA1: TTCCTATGTCTGGTACGTAA	19
gRNA 序列: 8aA2: AAGTCGTACCTTCCTATGTC	20
gRNA 序列: 10aS1: CACGCTTGCCCGGCCATAT	21
gRNA 序列: 10aS2: CTTTGCATATTCACCACGCT	22
gRNA 序列: 10aS3: AGCGTCTGACAGCAGCGCCG	23
gRNA 序列: 10aA1: TCAACTTTGCGCAGCGGAGC	24
gRNA 序列: 10aA2: AGACGCTCAACTTTGCGCAG	25
gRNA 序列: 非靶向干扰: AAATGTGAGATCAGAGTAAT	26
引物: <i>TCF4</i> 3b 正向: GTGCCGAAACTACACTTTTGTG	27
引物: <i>TCF4</i> 3b 反向: ACAAGATTATGCACCTGGCT	28
引物: <i>TCF4</i> 8a 正向: TGGAAAGGGGATGCTTACTCTC	29
引物: <i>TCF4</i> 8a 反向: AACTCTAATGACCTCCGCCT	30
引物: <i>TCF4</i> 10a 正向: TTACCCTAAGGCAGTTGAGTGGA	31
引物: <i>TCF4</i> 10a 反向: TTGTCAAAAATCCCCCTCGCA	32
引物: <i>TCF4</i> 4/5 正向(TCF4-B 亚型): CCTCCTGTGAGCAGTGGA	33
引物: <i>TCF4</i> 4/5 反向(TCF4-B 亚型): CTGGACGGGCTTGGATGTC	34
引物: <i>TCF4</i> 9/10 正向(TCF4-D 亚型): GCCATCTTCAGTCTATGCTCCATC	35
引物: <i>TCF4</i> 9/10 反向(TCF4-D 亚型): TAGGGAAAGTGCTGGTTGCTGG	36
引物: <i>TCF4</i> -A 正向(TCF4-A 亚型): GGAAAGCGGTCTATGCTCCAT	37
引物: <i>TCF4</i> -A(9/10)反向(TCF4-A 亚型): TAGGGAAAGTGCTGGTTGCTGG	38
引物: <i>TCF4</i> 12/13 正向(所有亚型): CTTCTCCGATGTCCACTTTCCA	39
引物: <i>TCF4</i> 12/13 反向(所有亚型): CCCGCTTCCTCTATTTGCCAT	40
引物: <i>MPH</i> 正向: ATCAGGGCGTGTCCATGTCTCA	41
引物: <i>MPH</i> 反向: CGGGTAATGGCTTCGGGGTA	42
引物: <i>dCas9</i> 正向: GCCGACGCTAATCTGGACAAAGT	43
引物: <i>dCas9</i> 反向: TGGTCAGGGTAAACAGGTGGATG	44
引物: <i>TBP</i> 正向: TGTATCCACAGTGAATCTTGGTTG	45
引物: <i>TBP</i> 反向: GGTTCGTGGCTCTCTTATCCTC	46
引物: <i>CNTNAP2</i> 正向: TCACACAGACCAAGATGAGCCAA	47
引物: <i>CNTNAP2</i> 反向: TAGGAAGCGAACCTCGTGCCA	48
引物: <i>KCNQ1</i> 正向: GGACAAAGACAATGGGGTGACT	49
引物: <i>KCNQ1</i> 反向: GTGTTGGGCTCTTCCTTACAGAA	50
引物: <i>TCF4B</i> cDNA 正向: ATGCATCACCAACAGCGAATGG	51
引物: <i>TCF4B</i> cDNA 反向: ATCCTCGAGTTACATCTGTCCCATGTGATTCG	52
引物: minP_TCF4: GAATTCAGAGGGTATATAATGGAAGCTCGACTTCCAGATGCATCACCA ACAGCGAATGG	53
引物: E-box-x6-minP_TCF4: GAATTC AACACCTGCAAGAACACCTGCAAGAACACCTGCAAGAACCA CCTGCAAGAACACCTGCAAGAACACCTGCAAGAGAGGGTATATAATG GAAGCTCGACTTCCAGATGCATCACCAACAGCGAATGG	54

[0171]	引物: E-box-x12-minP_TCF4: GAATTCACCTGCAAGAACACCTGCAAGAACACCTGCAAGAACACCT GCAAGAACACCTGCAAGAACACCTGCAAGAACACCTGCAAGAACAC CTGCAAGAACACCTGCAAGAACACCTGCAAGAACACCTGCAAGAAC ACCTGCAAGAGAGGGTATATAATGGAAGCTCGACTTCCAGATGCATC ACCAACAGCGAATGG	55
--------	--	----

[0172] 对于类器官转导实验(图11F-H和19M-S),使用第二代慢病毒生产质粒psPAX2(Addgene#12260)和pMD2.G(Addgene#12259)制备慢病毒颗粒。将30个聚集80%的HEK293T细胞的10cm平板每板转染10 μ g质粒,根据制造商的建议使用2nd Generation Packaging mix(ABM;LV003)和Lentifectin转染试剂(ABM;G074)。转染两天后,从所有平板中收获上清液,并使用PEG-it病毒沉淀溶液(Systems Biosciences;LV810A-1)利用PEG沉淀纯化病毒。使用qPCR慢病毒滴定试剂盒(ABM;LV900)实现滴度的测定。所有滴度均高于每 μ L10⁹ IU(颗粒)。从Applied Biological Materials(ABM)订购等效的AAV9血清型的AAV载体,每种构建体中含有人类生长激素(hGH)终止子。所有滴度均在商业上测定为每 μ L>10⁹ IU(颗粒)。通过在类器官衍生的第一天混合350万个分离的iPSC(慢病毒;图11F)或通过在神经诱导期的最后一天之后将AAV病毒直接添加到培养基中(图11H)实现类器官(Ct0)的转导,以适当的病毒量获得每种类型的载体的感染复数(multiplicity of infection,MOI)为5。

[0173] CRISPR介导的TCF4表达的反表观遗传校正。首先,从PTHS和对照NPC分析RNA测序文库,以确定来自人TCF4基因的众多替代启动子的转录活性(Sepp等人,2011)。使用(Demircioğlu等人,2019)中所述的连接读数计数方法进行启动子活性估计。简而言之,通过使用STAR aligner(版本STAR 2.7.6a)将RNA-seq读数映射到GRCh38.p13基因组组装上,使用GENCODE 34原始(primary)注释作为参考来确定外显子坐标,从而获得外显子连接计数。接下来,使用proActiv R软件包(版本0.99.0)通过计数映射到每个TCF4转录本的第一组内含子的连接读数来估计启动子活性,然后使用大小因子方法和数据对数转换对计数进行归一化。该方法将外显子3b、8a和10a上游的启动子鉴定为亲本和PTHS样品中最活跃的启动子(图18A),因此被选用于CRISPR介导的TCF4转录活性的反表观遗传操纵。

[0174] 在这三个启动子中,gRNA基于位于相应转录起始位点(transcriptional start sites,TSS)的-100和+50之间的序列而设计(Liao等人,2017)(图18A)。对于每个启动子,根据设计的计算工具生成的分数选择3个正义和2个反义gRNA(Hsu等人,2013)。选择非靶向干扰序列(scrambled sequence)(gRNA序列见表2)作为对照gRNA序列。通过首先将相应的序列插入传统的CRISPR pSpCas9(BB)-2A-Puro质粒(Addgene#48139;[http://n2t.net/addgene:48139;RRID:Addgene_48139]),然后在预实验中测试每个gRNA产生插入缺失的效率。为此,pSpCas9(BB)-2A-Puro用BpiI(Thermo Fisher Scientific)消化。每对合成的gRNA寡核苷酸均用T4多核苷酸激酶(Promega)磷酸化,并在以下条件下通过在热循环仪中孵育退火:在37℃下孵育30min,在95℃下5min,并以5℃min⁻¹降温至25℃。然后将每个gRNA的磷酸化寡核苷酸双链连接到消化的质粒上,方法是在25℃下与T4 DNA连接酶(Promega)孵育1小时。感受态细胞(Stbl3大肠杆菌菌株;Thermo Fisher Scientific)用每种连接产物进行转化,并使用PureYield质粒小量制备系统(Promega)从每个克隆中提取质粒DNA,然后使用hU6-F通用引物通过Sanger测序进行验证。

[0175] 接着,将HEK293T细胞(ATCC)在含有10% FBS和1%青霉素/链霉素的DMEM中培养

至聚集70%，并用聚乙亚胺(PEI;Sigma-Aldrich)以3:1的PEI/DNA(w/w)的比例转染每种gRNA质粒，每mL培养基含有1 μ g DNA。将PEI和DNA在Opti-MEM(Gibco)中以培养基总体积的1/20稀释，然后孵育30分钟，然后直接施用于细胞顶部。转染后16小时用培养基替换转染培养基。为了确认所选的TCF4 gRNA序列确实靶向人TCF4基因的TSS 3b、8a和10a，进行了T7核酸内切酶I测定。通过用补充有1 μ g mL⁻¹嘌呤霉素的培养基替换转染培养基来选择转化的细胞，直到阴性对照中的所有细胞死亡(~72小时)。然后根据制造商的说明，使用Illustra血液基因组制备Minispin试剂盒(GE)提取基因组DNA。在每个gRNA的靶位点两侧设计一对引物，并使用具有Q5高保真DNA聚合酶(NEB)进行终点PCR。扩增子使用向导SV凝胶和PCR清洁系统(Promega)纯化，并使用Qubit DNA BR检测试剂盒(Thermo Fisher Scientific)进行定量。对于T7核酸内切酶I测定，将每个样品的300ng扩增子与2 μ L NEBuffer 2和H₂O(最终体积为19.5 μ L)在热循环仪中孵育，循环参数如下：在95℃下5分钟，以-2℃ min⁻¹降温至85℃，以-0.1℃ min⁻¹降温至25℃。变性并逐渐重新退火以形成DNA异质双链体后，将5U的T7核酸内切酶I(NEB)加入样品中，并在37℃下孵育30分钟。产物在1.5%琼脂糖凝胶上运行，并根据未消化条带和消化片段质量数之间的比率估计每个gRNA的CRISPR介导的产生插入缺失的效率。此外，对扩增子进行深度测序，并计算带有插入缺失的克隆的百分比。

[0176] 所有gRNA也克隆到pLentiSAMv2质粒(Addgene质粒#75112;[http://n2t.net/addgene:75112;RRID:Addgene_75112])中，其承载了U6启动子控制的在四环和茎环2上都含有MS2环的gRNA序列，以及EF1 α 启动子的控制的与VP64基因融合的死亡Cas9(dead Cas9, dCas9)基因。通过用Esp3I(Thermo Fisher Scientific)消化pLentiSAMv2进行克隆。接下来，通过在HEK293T细胞(ATCC)中转染携带测试gRNA或干扰gRNA(对照)的合适pLentiSAMv2载体来制备慢病毒颗粒。

[0177] 为了评估设计的TCF4 gRNA序列在通过反表观遗传激活增加TCF4基因内源表达方面的效率，在SH-SY5Y细胞中转染pLentiSAMv2和pLentiMPHv2(Addgene#89308;[http://n2t.net/addgene:89308;RRID:Addgene_89308])质粒，然后进行RT-qPCR以验证TCF4转录本的水平。pLentiMPHv2载体含有EF1 α 启动子控制的MS2-P65-HSF1激活辅助(MPH)复合基因(Liao等人,2017)，与gRNA和死亡Cas9结合，用于TCF4位点的反表观遗传激活。

[0178] 在含有10% FBS和1%青霉素/链霉素的DMEM/F12中培养SH-SY5Y细胞。SH-SY5Y细胞用FuGENE HD转染试剂(Promega)以4:1FuGENE/DNA(v/w)的比例进行转染，每mL培养基含有2 μ g DNA。FuGENE和DNA均在Opti-MEM(Gibco)中以培养基总体积的1/10稀释，在应用于细胞之前不进行孵育。转染16小时后，用补充有10 μ g mL⁻¹杀稻瘟菌素S(Sigma-Aldrich)的培养基替换转染培养基，以筛选转染的细胞。对照细胞死亡(~72小时)后，用培养基替换选择培养基以使细胞扩增。对于每个gRNA，转染3次。根据制造商的说明，用TRIzol试剂(Thermo Fisher Scientific)纯化来自选定细胞的RNA。cDNA由ImProm-II逆转录系统(Promega)合成。对于RT-qPCR反应，引物对被设计为能够检测(I)编码TCF4-B,TCF4-D或TCF4-A的转录本(取决于每个gRNA靶向的相应启动子)，(II)内源性TBP基因的转录本，以及外源性dCas9(由lentiSAMv2编码)和MPH(由lentiMPHv2编码)基因，以及(III)来自受TCF4转录调控基因的转录本。所有反应均进行技术重复，在QuantStudio 6Flex实时荧光定量PCR系统(Applied Biosystems)上使用PowerUp SYBR绿色预混液(Applied Biosystems)。每次运行结束时始终包括熔融曲线步骤。使用空的lentiSAMv2质粒或含有干扰gRNA的质粒转染的样品作为通

过传统 $\Delta \Delta Ct$ 方法定量 TCF4 和 TCF4 靶基因转录本相对水平的参考, TBP、dCas9 和 MPH 的转录水平用于样品之间的标准化。

[0179] 对于类器官转导实验(图11B-D和18E-F), 由 pLentiMPHv2 载体和含有不同 gRNA 的几种 pLentiSAMv2 版本的载体制备慢病毒颗粒(Liao 等人, 2017)。对于病毒制备, 使用第二代慢病毒生产质粒 psPAX2 (Addgene#12260) 和 pMD2.G (Addgene#12259)。根据制造商的建议, 将 20 个 10cm 平板上的聚集 80% 的 HEK293T 细胞用 10 μ g/板的质粒, 2nd Generation Packaging mix (ABM; LV003) 和 Lentifectin 转染试剂 (ABM; G074) 转染。转染两天后, 从所有平板中收获上清液, 并使用 PEG-it 病毒沉淀溶液 (Systems Biosciences; LV810A-1) 利用 PEG 沉淀纯化病毒。使用 qPCR 慢病毒滴定试剂盒 (ABM; LV900) 进行滴度测定。所有滴度均高于每毫升 10⁹ IU (颗粒)。

[0180] 类器官 (sP0) 的转导是通过在类器官衍生的第一天将 250 万个分离的 iPSC 与适当的病毒量混合来实现的, 以获得每种类型的慢病毒的感染复数 (MOI) 为 5。处于“干扰 gRNA”条件 (对照) 的类器官由含有干扰 gRNA 的 pLentiMPHv2 和 pLentiSAMv2 质粒产生的慢病毒共转导。“TCF4 gRNA”组中的类器官与含有 gRNA 版本 3bS3 的 pLentiMPHv2 和 pLentiSAMv2 质粒产生的慢病毒共转导。在类器官衍生的第二天和第三天, 更换培养基并再次加入慢病毒。在这 3 天内, 在含有 SB431542 和 dorsomorphin 的 mTeSR1 培养基存在下形成拟胚体。从第四天开始, 根据常规方案更换培养基, 不添加病毒。通过使用本文和上述方案通过免疫染色评估 Cas9 表达来确认转导。

[0181] 统计分析。除非另有说明, 否则数据以平均值+SEM 表示。由于可用的 PTHS 样本有限, 因此没有进行诸如功效分析之类的统计方法来确定样本量, 这些样本是根据有关每个患者携带的 TCF4 突变类型的详细信息的可用性选择的。然而, 基于在整个研究中观察到的强烈而一致的效应量以及所有受试者 (NPC 和类器官) 细胞系的差异水平, 增加样本量预计不会改变结果的统计学显著性。

[0182] 在整个研究中分别使用了不同类型的统计检验, 如相应的附图描述所示。通常, 在测量类器官尺寸、相对表达水平或表达丰度的实验中, 两组之间的均值 (PTHS 与亲本) 的比较使用双样本 Welch's t 检验, 假设方差不等和异方差。当在两个以上的组中比较这些类型的均值时, 使用单因素方差分析 (ANOVA), 然后使用 Tukey 的诚实显著差异 (Honestly Significantly Different, HSD) 事后检验。为了比较两个样品之间单细胞 RNA-Seq 数据中的平均基因表达, 使用了非参数的 Mann-Whitney U 检验。对于两个以上样本之间的相同类型的比较, 使用 Kruskal-Wallis 检验, 然后使用 Dunn 事后检验。对于单细胞转录组学数据中平均基因表达值的比较, 计算出的 p 值在图中显示为星号, 但在 PTHS 和亲本之间的倍数变化小于亲本组平均值的任意值的比较中添加了一个十字符号 (†)。为了比较 PTHS 和对照神经元之间的轴突长度和胞体面积, 使用 ANOVA 与 Geisser-Greenhouse 校正进行重复测量, 然后进行 Fisher's 最小显著差异 (Least Significantly Different, LSD) 事后检验。

[0183] 样品大小在图例中表示。P 值在图中报告为星号, 显著性水平定义为 $p < 0.05$ (*)、 $p < 0.01$ (**) 或 $p < 0.001$ (***)。当实验涉及每个受试者细胞系多个独立重复, 或每个独立重复涉及多个技术重复时, 重复次数也显示在图例中, 即使每个统计测试仅基于不同受试者均值之间的比较计算。盲法用于比较患者和对照样本的大多数分析。使用 Prism 软件 (GraphPad)、Rstudio、G*Power 和 WebPower 进行统计分析。

[0184] 来源于皮特-霍普金斯综合征患者的脑皮质类器官表现出异常的大小和形态。为了深入了解由TCF4突变引起的在很大程度上未知的细胞病理生理学,通过对来自五名PTHS患者和相应性别匹配亲本的皮肤成纤维细胞进行细胞重编程来产生iPSC系(表1)。这些个体携带TCF4基因部分或全部删除的突变,包括去除其必需的bHLH DNA结合域,或影响其转录激活域之一(图12A)。检查所有iPSC克隆的干细胞标志物表达,基于SNP图谱的核型分析显示其没有不必要的染色体异常(图12B)。在生长速率(图12C)或体外衍生NPC和神经元的一般能力(图12D-E)方面,PTHS和对照iPSC系之间没有观察到差异。

[0185] 接着,将iPSC系用于生成脑皮质类器官(Ct0)(图5A),随后在细胞和分子水平上评估异常表型。Ct0在早期发育过程中的转录类似于人类皮质,包括功能性谷氨酸能和GABA能神经元,并表现出功能类似于人类神经发育期间观察到的特征的细胞群。对照(亲本)Ct0如预期表现出三维组织成球状体,其尺寸不断增大并形成清晰可见的玫瑰结状细胞聚集体(图5A中顶行的箭头)。与此形成鲜明对比的是,PTHS Ct0更小(图5A-B),并且表现出明显较少的可辨别玫瑰结和更小的尺寸(图5A)。一些PTHS类器官表现出极化结构(图5A底行中的箭头)。这些表型在使用来自同一患者的不同克隆进行的批次中是一致的(图12F)。

[0186] 虽然Ct0概括了皮质发育的几个方面,但它们大多缺乏皮质下起源的GABA能细胞(图6D)。因此,还衍生出了皮质下类器官(sP0,图5C),其包含神经祖细胞和GABA能抑制神经元。与在Ct0中观察到的情况类似,PTHS sP0也显示出较小的尺寸(图5C)和异常的内部组织,很少或没有玫瑰结(图12H)。

[0187] 总之,这些结果表明PTHS脑类器官表现出异常形态,表明PTHS患者潜在的神经发育过程可能被改变。此外,对照组和患者来源的脑类器官之间的表型差异程度证实了这种人细胞模型可用于研究PTHS病理生理学和分子机制。

[0188] PTHS类器官中祖细胞和神经元的含量异常。较小的类器官可能是由一系列改变的细胞过程引起的,例如细胞分裂减少或细胞凋亡增加、异常迁移或衰老。为了确定PTHS类器官中哪些过程存在缺陷,分析了几种关键细胞亚型的组织和含量。首先,对患者和对照Ct0和sP0的组织学切片进行神经祖细胞标志物SOX2的免疫染色。在体外4周时,对照Ct0包含大量由围绕脑室样管腔(ventricle-like lumen)的神经祖细胞组成的玫瑰结,类似于发育中人脑室和脑室下区祖细胞的分布。当这些富含祖细胞的结构分化成几种神经元亚型时,玫瑰结的大小减小。与对照Ct0相比,PTHS类器官显示出很少的玫瑰结状结构,并且神经祖细胞分散在整个类器官中,没有任何明显的组织聚类(参见图12F中的支持对照)。此外,大多数PTHS Ct0是极化的,SOX2阳性细胞集中在一侧。研究发现,与对照类器官相比,PTHS Ct0的密度显著降低,但神经祖细胞的百分比更高(图5D和12F),与祖细胞倾向于定位的较少的玫瑰结结构保持一致。同样,PTHS sP0显示SOX2+祖细胞的含量降低(图12H-I)。

[0189] 用于神经元标记物MAP2的免疫染色显示,对照Ct0具有分散在整个球状体中的神经元,特别是在玫瑰结周围和之间,并且神经元含量随着亲本Ct0发育的进行而增加。不同的是,即使在类器官发育的后期阶段,PTHS Ct0和sP0也具有不太明显的MAP2标记(图12H)。亲本Ct0表现出典型的皮质发育模式,概括了人类皮质中神经元分化的时间进展,其中首先形成深层神经元(即CTIP2+细胞),然后是浅层神经元(SATB2+和CUX1+细胞)的分化(图12J)。相比之下,PTHS Ct0表现出皮质神经元亚型的含量严重降低(图5E和12J)。此外,成熟的PTHS Ct0显示囊泡型谷氨酸转运蛋白家族成员1(vesicular glutamate transporter

family member 1, vGLUT1) 的染色减少 (图12M), vGLUT1是兴奋性神经元的标志物, 之前已证明Ct0中大量存在。同样, PTHS sP0减少了GABA能神经元标记物GAD65/67的染色 (图12H)。重要的是, 在死后PTHS脑样本中观察到MAP2和皮质神经元标记物的表达同样降低 (图12K) 以及皮质型神经元数量减少 (图12L)。

[0190] 总之, 这些数据表明, 在细胞水平上患者来源的类器官与在患者中观察到的神经表型非常匹配, 并证实PTHS的特征在于皮质神经元含量和组织的严重缺陷。

[0191] PTHS类器官表现出较低的神经元百分比和较高的祖细胞百分比。为了更好地量化脑类器官的细胞多样性, 对来自Ct0和sP0的游离细胞进行了单细胞RNA测序 (scRNA-Seq)。分析了六个带注释的细胞亚群——Ct0和sP0中的神经祖细胞、中间祖细胞和成熟神经元 (图6A, 13A-C)。进一步分析了这些细胞组, 因为分化轨迹分析表明它们组成了两个不同的分化谱系, 其中神经祖细胞通过中间祖细胞阶段进展为谷氨酸能神经元 (兴奋谱系) 或GABA能神经元 (抑制谱系) (图6B)。因此, 研究这些群体适用于评估皮质细胞类型谱系的神经发育缺陷。类器官不包含表达中胚层和内胚层标志物的细胞 (图12C), 并且未研究其他较小的群体 (图13A中的“其他”), 因为即使它们是神经来源的 (图13E), 它们也不能明确地分配到上述选择进行分析的六个群体。

[0192] PTHS类器官具有降低的单位面积祖细胞密度 (图5D), 但scRNA-Seq数据和免疫染色显示, 相对于对照类器官, PTHS Ct0中祖细胞的百分比更高 (图6C、6D、6E和12G)。同样, PTHS sP0具有比对照sP0更高的皮质下祖细胞百分比 (图6F、6G和6H)。此外, PTHS和对照类器官中的星形胶质细胞含量很小且相似 (图13F), 排除了星形胶质细胞含量是PTHS类器官表型异常的潜在原因。

[0193] 与PTHS类器官具有较低神经元含量的发现一致 (图5D和5E), scRNA-Seq分析显示, 与对照类器官相比, PTHS Ct0和sP0中兴奋性和抑制性神经元的百分比分别降低 (图6C、6D、6F和6G)。此外, 在PTHS Ct0中表达BCL11B (编码CTIP2)、SATB2、TBR1和CUX1的神经元百分比降低 (图6I、13H和13I), PTHS sP0中表达GAD2 (编码GAD67) 的神经元数量也较低 (图6J)。

[0194] 总之, 这些数据表明PTHS类器官具有比例更多的祖细胞和更少的神经元, 表明该疾病的病理生理学包括祖细胞增殖和/或分化为神经元的缺陷。

[0195] PTHS神经元表现出异常的放电特性。PTHS类器官中神经元含量的减少表明神经元回路的形成可能在患者的神经组织中受损。此外, PTHS Ct0中vGLUT1染色的减少可能表明突变神经元建立较少的突触并表现出电活动受损。为了研究这些问题, 使用了2D培养物和类器官中的PTHS神经元。首先, 使用多电极阵列 (MEA) 测定来分析神经元活动, 发现与对照Ct0相比, PTHS的平均神经元放电率要低得多 (图7A和14A)。PTHS Ct0含有大量的神经元 (图5D和5E), 因此电活动的这种减少可能反映了细胞水平的电生理缺陷。然而, PTHS Ct0中的神经元含量受损 (图5E、6I、12J、13H和13I), 因此PTHS电活动缺陷在形式上可能是类器官中连接不良或神经元密度差的结果。

[0196] 为了评估TCF4单倍体剂量不足对单个神经元的影响并确定类器官中活性的降低是否可能源于异常的电神经元特性, 研究了2D神经元培养物。首先, 进行实验以确认TCF4在对照神经元中表达 (图14B), 并且与各自的亲本对照相比, PTHS神经元的TCF4表达减少 (图14C)。通过分析神经元树状结构 (图7B), 得出结论, PTHS神经元的胞体区域与亲本对照的胞体区域没有区别, 而PTHS神经元的神经元突起更长 (图7C)。接下来, 进行膜片钳分析以评估

来自最显著受损的患者系的2D培养中的神经元(图7A)。结果发现,PTHS神经元表现出严重的内在兴奋性(图7D)、膜电容以及钠和钾电流的下降(图7E、14D和14E)。此外,与对照Ct0相比,在PTHS Ct0的神经元中观察到神经元活动的替代标志物FOS表达较低(图7F)。综合起来,这些数据表明PTHS神经元在网络和细胞水平上表现出严重的电特性缺陷。

[0197] 这种神经元功能障碍可能由PTHS神经元中的异常基因表达引起,因此用RNA测序探究这些细胞中的转录组改变,比较在2D培养条件下从iPSC衍生的患者和对照神经祖细胞分化的神经元。差异表达(DE)分析比较了来自2月龄的FACS分选培养物的PTHS和对照神经元,揭示了一系列错误调节的基因(图14F),其中一些基因参与神经发生、神经元同一性、分化和神经元兴奋性的调节(图14G)。在具有较高倍数变化(高于4倍)的基因中,有几个基因在神经元功能中很重要(图7G),这些基因在PTHS类器官的神经元中也下调(图7H)。

[0198] 总之,这些数据表明来源于PTHS患者的神经元在形态学、生理学和转录组方面是异常的。重要的是,DE基因列表包括钾电压门控通道亚家族Q成员1(potassium voltage-gated channel subfamily Q member 1, KCNQ1),先前显示其在Tcf4敲低后使小鼠神经元的内在兴奋性失调(Rannals等人,2016),以及许多其他离子通道(图14H),其不仅提供了对PTHS神经元内在兴奋性缺陷机制的见解,而且还为药物治疗干预提供了新的机会。

[0199] PTHS神经祖细胞显示出较低的增殖速率和复制性衰老。PTHS Ct0和sP0具有较少的神经玫瑰结和较低的NPC密度(但较高百分比)的发现,提出了是否这些表型是异常神经诱导、祖细胞增殖减少或分化受损的结果。为了评估这些可能性,在神经诱导阶段(第2周)后计数玫瑰结的数量,并显示它在亲本对照和PTHS类器官中相似,SOX2+细胞的密度也相似(图8A)。这些结果,加上类器官中缺乏表达非神经标志物的细胞(图13D和13E),强烈表明PTHS类器官中的神经诱导是正常的,并且在类器官成熟后期由于祖细胞增殖不良或分化受损,玫瑰结在器官成熟后期减少。为了解析这两种可能性,iPSC衍生的NPC是由PTHS个体和亲本对照产生的(参见图15A中NPC标志物的表达)。这些细胞确实表达TCF4(图15B、15C和15D),并且在PTHS NPC中表达降低(图15E和15G)。重要的是,在PTHS NPC中,TCF4的直接转录靶标GADD45G的表达强烈降低(图15F),证实了TCF4功能在所有患者系中都显著受损。观察到PTHS NPC在2D培养中的生长速度明显慢于对照系(图8B和8D)。由于这种差异可能来自增殖减少或凋亡增加,因此使用膜联蛋白V介导的流式细胞术来评估凋亡率,并得出结论,PTHS和亲本NPC在凋亡细胞的百分比上没有显著差异,通常小于5%(图8C)。

[0200] 接下来,通过用5-乙炔基-2'-脱氧尿苷(EdU)孵育细胞培养物来评估NPC的增殖能力,EdU是一种胸腺嘧啶核苷类似物,在合成过程中被掺入DNA中,然后基于流式细胞术测定进行分裂的细胞百分比。PTHS系的增殖细胞百分比约为对照NPC的一半(图8D)。还观察到PTHS NPC经常呈现出非典型的扩大、扁平的形态(图8G中的箭头),这在对照系中没有观察到。异常形态和增殖活性降低的结合导致了PTHS神经祖细胞正在经历早熟复制衰老的假设。这是一个定义明确的细胞过程,其特征为细胞周期停滞和随后的增殖停止,并显示参与各种生理和病理缺陷。事实上,除了较大的细胞大小外,在PTHS NPC中还观察到三个表明复制衰老的特征: β -半乳糖苷酶活性增强(SA- β -gal;图8G),核纤层蛋白B的表达降低(LMNB1;图8H)、细胞周期蛋白依赖性激酶抑制剂基因CDKN2A(图8H)和CDKN1A的表达显著增加——其表达导致细胞分裂停滞并作为复制衰老的标志物。有趣的是,这些特征随着传代次数的增加而增强(图15H)。而且,衰老的NPC是巢蛋白+,其大部分是SOX2+;SOX2在一些NPC中表达

降低,并且对Brachyury和SOX17呈阴性(图16E),表明它们是衰老的NPC,而不是错误分化的细胞。

[0201] 有趣的是,衰老标志物的表达在PTHS死后样本中也强烈上调(图8I)。PTHS Ct0含有许多表达CDKN2A基因产物p16^{INK4A}的神经谱系细胞(图8J和15J),这些细胞不是在PTHS和对照类器官中同样不常见的凋亡细胞(图8J)。重要的是,shRNA介导的TCF4敲低在对照NPC中导致增殖减少(图15K)和CDKN2A更高的表达(图15L),进一步加强了患者来源NPC中TCF4表达减少与衰老增加和增殖减少之间的联系。

[0202] 总之,这些数据表明PTHS在细胞水平上的病理机制涉及NPCs的增殖减少和衰老增强。

[0203] PTHS神经祖细胞和类器官中Wnt信号的校正拯救异常表型。为了获得对PTHS NPC异常增殖活性的机制见解,对来自4对亲子的PTHS和对照细胞之间的差异表达(DE)基因进行了无偏研究(图15M),然后进行了基因集富集分析(图15N)。这种方法主要揭示了Wnt信号通路中基因表达的改变(图9A和16A)。由于Wnt通路与许多组织中的NPC增殖有关,因此提出了异常Wnt活性可能与PTHS细胞中观察到的较低NPC增殖率有因果关系的假设。为了验证这一假设,分析了PTHS NPC中Wnt成分的表达,以确认WNT2B、WNT3、WNT5A和SFRP2在患者系中的表达较低(图9A-B),并且使用荧光素酶报告基因评估信号传导功能,表明典型Wnt/ β -连环蛋白信号活性显著降低(图9C)。重要的是,几种Wnt信号成分的表达在死后PTHS脑样本中显著下调(图9D)。

[0204] 用Wnt信号拮抗剂DKK-1和ICG-001处理对照NPC,表型模拟了PTHS祖细胞增殖率降低(图9E)和CDKN2A表达增加(图16B)。用ICG-001处理对照Ct0,ICG-001是一种易于穿透三维类器官结构的扩散小分子,其导致类器官尺寸(图16C)和SOX2+细胞含量显著降低(图9F)。作为反向方法,用Wnt信号激动剂CHIR99021处理PTHS NPC。首先,确认Wnt信号在处理的细胞中增加(图16D)。用CHIR99021处理恢复了PTHS NPC的增殖速率(图9G-H),降低了p16^{INK4a}衰老细胞的百分比(图16E),并增加了促增殖基因HES1以及前神经基因ASCL1和NEUROD1的表达(图9J)。PTHS Ct0的治疗导致类器官尺寸(图16F)和NPC含量(图9K)显著增加,并重新出现明显的神经玫瑰结。对CHIR99021处理的PTHS Ct0和sP0中的细胞多样性的分析证实了祖细胞数量的增加(图16G-H)。

[0205] 处理过的Ct0具有增加的皮质下型神经祖细胞的百分比(图16H),并且在CHIR99021处理的亲本和PTHS类器官中皮质下标志物的表达更高(图16I),增加了这部分细胞命运变化的可能性,从皮质-皮质(cortico-pallial)到皮质下(subpallial)轨迹,可能导致Wnt激活后在PTHS细胞中观察到的细胞表型被逆转。然而,CHIR处理不会导致亲本类器官中大小(图16F)或祖细胞含量(图9K)的增加,并且CHIR处理不会导致2D培养中亲本NPC的增殖增加(图9G-H)或衰老细胞的百分比降低(图9I)。这些结果排除了命运限制导致PTHS细胞Wnt信号激动激活后表型校正的原因。

[0206] CHIR处理增加了2D培养的PTHS NPC中TCF4和TCF4下游靶标的表达(图16K),这是先前报道的其他细胞类型暴露于高CHIR浓度时的效果,其增加了Wnt激活后的表型校正。然而,3D结构的类器官中的神经祖细胞在CHIR处理后没有表现出TCF4水平的增加(图16J),这使我们能够得出结论,Wnt信号激活后PTHS类器官增殖缺陷的挽救不是由于TCF4本身表达增加。

[0207] PTHS类器官中神经玫瑰结数量较少(图9D),这表明祖细胞的神经上皮结构是有缺陷的。由于 β -连环蛋白是Wnt信号通路的关键组成部分,也是上皮细胞粘附和完整性的重要调节因子,因此一个合理的假设是PTHS NPC中Wnt信号传导的减弱导致 β -连环蛋白表达失调,导致玫瑰结的解体和无法组织神经祖细胞。事实上,即使PTHS NPC和PTHS Ct0祖细胞中 β -连环蛋白的表达水平保持不变(图16M),PTHS类器官中 β -连环蛋白的表达紊乱(图16L),这加强了Wnt信号下调导致PTHS神经发育期间神经上皮完整性缺陷的可能性。

[0208] PTHS NPC中另一G0分类的DE基因是“钙粘蛋白”(图15N),由此进行实验以确定钙粘蛋白或原钙粘蛋白在PTHS细胞中的表达是否改变。这一分类中的大多数DE基因实际上是Wnt通路的组成部分,除了CDH23和PCDH15。CDH23的表达可以忽略不计(图16N),因此排除该基因作为潜在的机制候选者。PCDH15在PTHS NPC中显著下调,但是在细胞表型被校正的相同条件下(图9G-I),NPC中的CHIR99021处理进一步降低了其表达(图16N),排除了PCDH15作为PTHS NPC中异常表型的合理原因。

[0209] 综上,实验证实了Wnt信号在PTHS NPC增殖缺陷中的参与机制。更重要的是,这些结果表明,PTHS NPC和类器官中描述的异常表型可以在药理学上得到校正,这一观察结果可能为指导未来治疗TCF4单倍体剂量不足引起的疾病做出努力。

[0210] SOX基因在改变PTHS NPC增殖和分化中的参与机制。由于与亲本类器官相比,PTHS类器官含有更高百分比的NPC、更少的神经元和改变的Wnt信号传导(图5和9),因此进行了实验以确定TCF4下游的机制参与者,以及可以控制NPC增殖和分化的Wnt通路。由于已经描述了SRY相关HMG-box(SOX)蛋白与Wnt信号传导之间的相互作用,并且由于它们在细胞增殖/分化中的已知作用,因此研究了PTHS NPC中所有SOX基因的表达。结果显示,SOX1、SOX2、SOX3和SOX4在患者来源的细胞中显著下调(图10A)。SOXB亚家族的成员(SOX1、SOX2和SOX3)传统上被认为是细胞增殖的调节因子。事实上,发现这些基因主要在Ct0和sP0的祖细胞和中间祖细胞中表达(图13B和图17A-C)。SOX1在类器官中没有明显表达(图17A),因此将实验工作集中在SOX3上;此外,所有PTHS系都表现出该基因的表达降低,在某些患者中相当显著(图10A-B)。

[0211] 首先,进行实验以确定SOX3是否在功能上是TCF4的下游,因为shRNA介导的TCF4敲低在对照NPC中导致SOX3表达降低(图10C)。对死后PTHS皮质样本的研究表明,SOX3的表达(图10D)和SOX3+细胞的数量(图10E)都严重降低。还验证了SOX3位于Wnt信号通路的下游,因为用Wnt激动剂CHIR99021处理PTHS祖细胞增加了SOX3表达(图10F)。重要的是,shRNA介导的SOX3敲低导致对照NPC的细胞计数减少(图10G和17D),细胞周期阻滞基因CDKN2A(衰老标志物)的表达增加(图17E),以及促增殖基因HES1和前神经基因ASCL1的表达降低(图17E),这与在PTHS NPC中发现的表型相匹配。有趣的是,转染介导的瞬时SOX3过表达并没有挽救PTHS NPC的增殖缺陷(图17F-G)。这可能是由于在NPC增殖测定的许多天中缺乏持续的SOX3过表达,或者存在其他平行失调的通路。

[0212] PTHS中的NPC分化率较低,这是通过分化2D培养物中的神经元与祖细胞比率来判断的(图10J和图17J)。此外,与对照Ct0和sP0相比,PTHS中的中间祖细胞更稀少(图17K),表达中间祖细胞标记物POU3F2(编码BRN2)的细胞在PTHS类器官中数量较少(图17L)。综合起来,这些结果表明PTHS神经组织中祖细胞异常分化为神经元。鉴于SOX转录因子中SOXC亚家族成员(SOX4和SOX11)在神经发生过程中作为促分化因子的已知作用,PTHS分化异常的一

个合理的假设是由于SOXC表达较低。事实上,这些转录因子基因在Ct0和sP0的中间祖细胞和神经元中表达(图10I和图17H-I)。SOX4参与中间祖细胞的产生及其分化为早产(CTIP2-和TBR1-阳性)和晚产(BRN2-、SATB2和CUX1-阳性)皮质神经元。已证实,SOX4在三个PTHS祖细胞系(图10H)、PTHS Ct0中间祖细胞和兴奋性谱系神经元(图10I)和PTHS sP0的GABA能神经元(图17H)中表达较低。为了测试SOX4在细胞分化病理学中的参与,在从两个亲本系分化神经元2D培养物中,进行了锁核酸反义寡核苷酸(LNA ASO)介导的SOX4敲低。细胞计数和转录组学分析均显示,SOX4敲低后分化率(MAP2与SOX2比例)降低(图10L),模拟了PTHS神经元培养物的异常表型(图10J)。

[0213] 一种模型是TCF4功能丧失导致Wnt下调并因此降低SOX3表达,导致增殖减少和细胞衰老增加。同时,SOX4表达的降低会导致PTHS神经组织中的分化受损,从而导致患者来源的细胞中观察到病理表型。

[0214] 通过TCF4表达的遗传校正逆转异常的PTHS表型。进行实验以对TCF4进行遗传操控。首先,使用基于CRISPR的反表观遗传策略在PTHS类器官中校正TCF4表达(Liao等人,2017)。在该方法中,用两种病毒载体将三种表达盒递送至靶细胞:一种编码与工程RNA发夹适配体偶联的短向导RNA(gRNA);第二种编码结合上述适配体的转录激活复合物(MPH);第三种编码死亡Cas9。期望靶细胞中上述表达盒的表达能够表观遗传地反激活内源性TCF4位点,因为Cas9介导的gRNA与TCF4启动子结合聚集了MPH,增强了下游基因的转录(图11A)。

[0215] 建立了包含15种不同gRNA的表达盒的集合,这些表达盒靶向TCF4基因的3个替代启动子(外显子3b、8a和10a的上游)。这些启动子分别产生编码TCF4蛋白亚型B、D和A的转录本,在PTHS和亲本对照样品中最活跃(图18A)。发现一些gRNA可以有效地反激活神经元细胞系SH-SY5Y中的TCF4及其靶基因,并且一些gRNA理想地提供2倍的TCF4表达增量(图18B-C)。接下来,将含有最有效gRNA的表达盒转导到源自患者#4及其亲本对照的PTHS类器官中(图11B)。之所以选择该患者系,是因为与相应的对照组相比,它在类器官尺寸和细胞含量方面显示出最大的差异(图5B和5D),从而可以更轻松地确定TCF4校正的益处。TCF4校正通过TCF4免疫标记强度(图11C)和TCF4 mRNA水平(图18D)的增加来验证。PTHS系#4中的TCF4表达与亲本对照系中的表达相同(图15E和图18D中的上图),因为患者#4具有点突变,预计不会降低转录水平(图12A)。CRISPR介导的校正策略增强了内源性和突变等位基因(图18D,下图),并且全局增加的TCF4水平伴随着TCF4下游靶标GADD45G(图18E)的校正,揭示了TCF4位点的功能性校正。

[0216] 用TCF4gRNA载体转导的类器官显示衰老基因CDKN2A的表达降低,以及神经元标记物MAP2表达的校正(图18E)。此外,这些类器官中的SOX3表达增加(图18E),即使是部分校正,其可能是因为并非类器官中的所有细胞都表达SOX3。接受TCF4校正的类器官的PTHS组织学表型异常被挽救(图11D),其产生没有异常生长的正常球状体(中间图中的箭头)。转导类器官的SOX2和MAP2染色表明,形态学的校正伴随着类器官内部结构的重建,祖细胞在管腔周围形成神经玫瑰结(右图中的箭头;图11D)。PTHS类器官中的生长物通常含有MAP2+细胞的聚集体(图11J中的插图),这是一种异常特征,在用TCF4 gRNA转导的TCF4校正后在类器官中消失。

[0217] 此外,还研究了PTHS类器官中存在可能表明皮质神经元形成改变的未成熟神经元。DCX(双皮质素)被用作未成熟神经元的标志物,用其验证了尽管在亲本和PTHS类器官中

都发现DCX+细胞,但这些细胞有时会在PTHS类器官的生长物中形成异常形状的高口径DCX纤维束,这一特征在校正TCF4表达后被修正。有趣的是,在PTHS死后皮质组织样本(图18G)、PTHS Ct0和sP0的中间祖细胞和神经元(图18H)以及2D培养物中的PTHS神经元(图18I)中的DCX表达较低。值得注意的是,用TCF4 gRNA转导的类器官中DCX表达得到校正(图18F),进一步表明在CRISPR介导的TCF4水平校正后,未成熟神经元的数量恢复正常。

[0218] 上述CRISPR策略需要使用两种病毒载体,它们必须以最佳水平表达以促进TCF4表达的校正。作为替代方案,采用了一种更简单的校正TCF4水平的程序,其中细胞和类器官转导了对TCF4基因的额外拷贝进行过表达(OE)的慢病毒或AAV。在这些载体中,TCF4-B编码序列被置于TCF4结合基序(μ E5盒)的控制之下(图11E和18J),这种方法目的在于防止异位TCF4表达。首先,转导的PTHS NPC中TCF4和GADD45G表达在被校正(图18J),这表明该策略可用于TCF4遗传校正。接下来,在类器官衍生方案开始时观察到慢病毒TCF4 OE构建体的转导,并导致TCF4标记强度增加,并校正类器官中的TCF4和CDKN2A水平(图18K)。在TCF4 OE之后,成熟的PTHS类器官表现出丰富的神经玫瑰结(图11E),以及一般形态的校正和SOX2+祖细胞和CTIP2+皮质神经元数量的挽救(图11E和18K)。TCF4 OE的PTHS类器官的上述效应伴随着两个关键电生理参数——平均放电率和网络电脉冲次数的显著改善(图11G),这清楚地表明校正后的类器官的功能挽救。

[0219] 当使用AAV载体在神经诱导阶段后进行TCF4 OE时(图11H),实现了TCF4标记强度的明显增加(图11H),以及TCF4和CDKN2A表达、SOX2+和CTIP2+细胞的数量(图18N)以及玫瑰结丰度的重新出现(图11H)的校正。该实验不仅表明PTHS细胞病理学可以逆转,而且还表明TCF4单倍体剂量不足不会导致神经诱导受损,这与PTHS类器官发育早期阶段玫瑰结的存在一致(图8A),加强了细胞病理生理学涉及祖细胞增殖缺陷的假设。

[0220] 这些实验提出了明确的证据,证明了在PTHS类器官中观察到的病理学是TCF4表达减少的结果。重要的是,这些数据提供了概念验证,即TCF4单倍体剂量不足引起的病理生理学,包括祖细胞增殖和神经元分化受损以及细胞衰老和SOX基因表达失调,其可以在细胞和组织水平上得到纠正,为这种情况急需的治疗铺平了道路。

[0221] 本文已经描述了多个实施例。然而,可以理解,可以在不偏离说明书的精神和范围的情况下进行各种修改。因此,其它实施例也均在以下权利要求的范围之内。

SEQUENCE LISTING

<110> 加利福尼亚大学董事会
 <120> 调节TCF4基因表达和治疗皮特霍普金斯综合征的组合物和方法
 <130> P23111922WP
 <150> US 63/085,878
 <151> 2020-09-30
 <160> 55
 <170> PatentIn version 3.5
 <210> 1
 <211> 2016
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> CDS
 <222> (1) .. (2016)
 <400> 1
 atg cat cac caa cag cga atg gct gcc tta ggg acg gac aaa gag ctg 48
 Met His His Gln Gln Arg Met Ala Ala Leu Gly Thr Asp Lys Glu Leu
 1 5 10 15
 agt gat tta ctg gat ttc agt gcg atg ttt tca cct cct gtg agc agt 96
 Ser Asp Leu Leu Asp Phe Ser Ala Met Phe Ser Pro Pro Val Ser Ser
 20 25 30
 ggg aaa aat gga cca act tct ttg gca agt gga cat ttt act ggc tca 144
 Gly Lys Asn Gly Pro Thr Ser Leu Ala Ser Gly His Phe Thr Gly Ser
 35 40 45
 aat gta gaa gac aga agt agc tca ggg tcc tgg ggg aat gga gga cat 192
 Asn Val Glu Asp Arg Ser Ser Ser Gly Ser Trp Gly Asn Gly Gly His
 50 55 60
 cca agc ccg tcc agg aac tat gga gat ggg act ccc tat gac cac atg 240
 Pro Ser Pro Ser Arg Asn Tyr Gly Asp Gly Thr Pro Tyr Asp His Met
 65 70 75 80
 acc agc agg gac ctt ggg tca cat gac aat ctc tct cca cct ttt gtc 288
 Thr Ser Arg Asp Leu Gly Ser His Asp Asn Leu Ser Pro Pro Phe Val
 85 90 95
 aat tcc aga ata caa agt aaa aca gaa agg ggc tca tac tca tct tat 336
 Asn Ser Arg Ile Gln Ser Lys Thr Glu Arg Gly Ser Tyr Ser Ser Tyr
 100 105 110
 ggg aga gaa tca aac tta cag ggt tgc cac cag cag agt ctc ctt gga 384

Gly	Arg	Glu	Ser	Asn	Leu	Gln	Gly	Cys	His	Gln	Gln	Ser	Leu	Leu	Gly	
115				120				125								
ggt	gac	atg	gat	atg	ggc	aac	cca	gga	acc	ctt	tcg	ccc	acc	aaa	cct	432
Gly	Asp	Met	Asp	Met	Gly	Asn	Pro	Gly	Thr	Leu	Ser	Pro	Thr	Lys	Pro	
130				135				140								
ggt	tcc	cag	tac	tat	cag	tat	tct	agc	aat	aat	ccc	cga	agg	agg	cct	480
Gly	Ser	Gln	Tyr	Tyr	Gln	Tyr	Ser	Ser	Asn	Asn	Pro	Arg	Arg	Arg	Pro	
145				150				155				160				
ctt	cac	agt	agt	gcc	atg	gag	gtg	cag	aca	aag	aaa	gtt	cga	aaa	gtt	528
Leu	His	Ser	Ser	Ala	Met	Glu	Val	Gln	Thr	Lys	Lys	Val	Arg	Lys	Val	
165				170				175								
cct	cca	ggt	ttg	cca	tct	tca	gtc	tat	gct	cca	tca	gca	agc	act	gcc	576
Pro	Pro	Gly	Leu	Pro	Ser	Ser	Val	Tyr	Ala	Pro	Ser	Ala	Ser	Thr	Ala	
180				185				190								
gac	tac	aat	agg	gac	tcg	cca	ggc	tat	cct	tcc	tcc	aaa	cca	gca	acc	624
Asp	Tyr	Asn	Arg	Asp	Ser	Pro	Gly	Tyr	Pro	Ser	Ser	Lys	Pro	Ala	Thr	
195				200				205								
agc	act	ttc	cct	agc	tcc	ttc	ttc	atg	caa	gat	ggc	cat	cac	agc	agt	672
Ser	Thr	Phe	Pro	Ser	Ser	Phe	Phe	Met	Gln	Asp	Gly	His	His	Ser	Ser	
210				215				220								
gac	cct	tgg	agc	tcc	tcc	agt	ggg	atg	aat	cag	cct	ggc	tat	gca	gga	720
Asp	Pro	Trp	Ser	Ser	Ser	Ser	Gly	Met	Asn	Gln	Pro	Gly	Tyr	Ala	Gly	
225				230				235				240				
atg	ttg	ggc	aac	tct	tct	cat	att	cca	cag	tcc	agc	agc	tac	tgt	agc	768
Met	Leu	Gly	Asn	Ser	Ser	His	Ile	Pro	Gln	Ser	Ser	Ser	Tyr	Cys	Ser	
245				250				255								
ctg	cat	cca	cat	gaa	cgt	ttg	agc	tat	cca	tca	cac	tcc	tca	gca	gac	816
Leu	His	Pro	His	Glu	Arg	Leu	Ser	Tyr	Pro	Ser	His	Ser	Ser	Ala	Asp	
260				265				270								
atc	aat	tcc	agt	ctt	cct	ccg	atg	tcc	act	ttc	cat	cgt	agt	ggt	aca	864
Ile	Asn	Ser	Ser	Leu	Pro	Pro	Met	Ser	Thr	Phe	His	Arg	Ser	Gly	Thr	
275				280				285								
aac	cat	tac	agc	acc	tct	tcc	tgt	acg	cct	cct	gcc	aac	ggg	aca	gac	912
Asn	His	Tyr	Ser	Thr	Ser	Ser	Cys	Thr	Pro	Pro	Ala	Asn	Gly	Thr	Asp	
290				295				300								
agt	ata	atg	gca	aat	aga	gga	agc	ggg	gca	gcc	ggc	agc	tcc	cag	act	960
Ser	Ile	Met	Ala	Asn	Arg	Gly	Ser	Gly	Ala	Ala	Gly	Ser	Ser	Gln	Thr	
305				310				315				320				
gga	gat	gct	ctg	ggg	aaa	gca	ctt	gct	tcg	atc	tat	tct	cca	gat	cac	1008

Gly	Asp	Ala	Leu	Gly	Lys	Ala	Leu	Ala	Ser	Ile	Tyr	Ser	Pro	Asp	His	
325				330				335								
act	aac	aac	agc	ttt	tca	tca	aac	cct	tca	act	cct	gtt	ggc	tct	cct	1056
Thr	Asn	Asn	Ser	Phe	Ser	Ser	Asn	Pro	Ser	Thr	Pro	Val	Gly	Ser	Pro	
340				345				350								
cca	tct	ctc	tca	gca	ggc	aca	gct	gtt	tgg	tct	aga	aat	gga	gga	cag	1104
Pro	Ser	Leu	Ser	Ala	Gly	Thr	Ala	Val	Trp	Ser	Arg	Asn	Gly	Gly	Gln	
355				360				365								
gcc	tca	tcg	tct	cct	aat	tat	gaa	gga	ccc	tta	cac	tct	ttg	caa	agc	1152
Ala	Ser	Ser	Ser	Pro	Asn	Tyr	Glu	Gly	Pro	Leu	His	Ser	Leu	Gln	Ser	
370				375				380								
cga	att	gaa	gat	cgt	tta	gaa	aga	ctg	gat	gat	gct	att	cat	gtt	ctc	1200
Arg	Ile	Glu	Asp	Arg	Leu	Glu	Arg	Leu	Asp	Asp	Ala	Ile	His	Val	Leu	
385				390				395				400				
cgg	aac	cat	gca	gtg	ggc	cca	tcc	aca	gct	atg	cct	ggt	ggt	cat	ggg	1248
Arg	Asn	His	Ala	Val	Gly	Pro	Ser	Thr	Ala	Met	Pro	Gly	Gly	His	Gly	
405				410				415								
gac	atg	cat	gga	atc	att	gga	cct	tct	cat	aat	gga	gcc	atg	ggt	ggt	1296
Asp	Met	His	Gly	Ile	Ile	Gly	Pro	Ser	His	Asn	Gly	Ala	Met	Gly	Gly	
420				425				430								
ctg	ggc	tca	ggg	tat	gga	acc	ggc	ctt	ctt	tca	gcc	aac	aga	cat	tca	1344
Leu	Gly	Ser	Gly	Tyr	Gly	Thr	Gly	Leu	Leu	Ser	Ala	Asn	Arg	His	Ser	
435				440				445								
ctc	atg	gtg	ggg	acc	cat	cgt	gaa	gat	ggc	gtg	gcc	ctg	aga	ggc	agc	1392
Leu	Met	Val	Gly	Thr	His	Arg	Glu	Asp	Gly	Val	Ala	Leu	Arg	Gly	Ser	
450				455				460								
cat	tct	ctt	ctg	cca	aac	cag	gtt	ccg	gtt	cca	cag	ctt	cct	gtc	cag	1440
His	Ser	Leu	Leu	Pro	Asn	Gln	Val	Pro	Val	Pro	Gln	Leu	Pro	Val	Gln	
465				470				475				480				
tct	gcg	act	tcc	cct	gac	ctg	aac	cca	ccc	cag	gac	cct	tac	aga	ggc	1488
Ser	Ala	Thr	Ser	Pro	Asp	Leu	Asn	Pro	Pro	Gln	Asp	Pro	Tyr	Arg	Gly	
485				490				495								
atg	cca	cca	gga	cta	cag	ggg	cag	agt	gtc	tcc	tct	ggc	agc	tct	gag	1536
Met	Pro	Pro	Gly	Leu	Gln	Gly	Gln	Ser	Val	Ser	Ser	Gly	Ser	Ser	Glu	
500				505				510								
atc	aaa	tcc	gat	gac	gag	ggt	gat	gag	aac	ctg	caa	gac	acg	aaa	tct	1584
Ile	Lys	Ser	Asp	Asp	Glu	Gly	Asp	Glu	Asn	Leu	Gln	Asp	Thr	Lys	Ser	
515				520				525								
tcg	gag	gac	aag	aaa	tta	gat	gac	gac	aag	aag	gat	atc	aaa	tca	att	1632

Ser Glu Asp Lys Lys Leu Asp Asp Asp Lys Lys Asp Ile Lys Ser Ile	
530	535
act agg tca aga tct agc aat aat gac gat gag gac ctg aca cca gag	1680
Thr Arg Ser Arg Ser Ser Asn Asn Asp Asp Glu Asp Leu Thr Pro Glu	
545	550
cag aag gca gag cgt gag aag gag cgg agg atg gcc aac aat gcc cga	1728
Gln Lys Ala Glu Arg Glu Lys Glu Arg Arg Met Ala Asn Asn Ala Arg	
565	570
gag cgt ctg cgg gtc cgt gac atc aac gag gct ttc aaa gag ctc ggc	1776
Glu Arg Leu Arg Val Arg Asp Ile Asn Glu Ala Phe Lys Glu Leu Gly	
580	585
cgc atg gtg cag ctc cac ctc aag agt gac aag ccc cag acc aag ctc	1824
Arg Met Val Gln Leu His Leu Lys Ser Asp Lys Pro Gln Thr Lys Leu	
595	600
ctg atc ctc cac cag gcg gtg gcc gtc atc ctc agt ctg gag cag caa	1872
Leu Ile Leu His Gln Ala Val Ala Val Ile Leu Ser Leu Glu Gln Gln	
610	615
gtc cga gaa agg aat ctg aat ccg aaa gct gcg tgt ctg aaa aga agg	1920
Val Arg Glu Arg Asn Leu Asn Pro Lys Ala Ala Cys Leu Lys Arg Arg	
625	630
gag gaa gag aag gtg tcc tca gag cct ccc cct ctc tcc ttg gcc ggc	1968
Glu Glu Glu Lys Val Ser Ser Glu Pro Pro Pro Leu Ser Leu Ala Gly	
645	650
cca cac cct gga atg gga gac gca tcg aat cac atg gga cag atg taa	2016
Pro His Pro Gly Met Gly Asp Ala Ser Asn His Met Gly Gln Met	
660	665
<210> 2	
<211> 671	
<212> PRT	
<213> Homo sapiens	
<400> 2	
Met His His Gln Gln Arg Met Ala Ala Leu Gly Thr Asp Lys Glu Leu	
1	5
Ser Asp Leu Leu Asp Phe Ser Ala Met Phe Ser Pro Pro Val Ser Ser	
20	25
Gly Lys Asn Gly Pro Thr Ser Leu Ala Ser Gly His Phe Thr Gly Ser	
35	40
Asn Val Glu Asp Arg Ser Ser Ser Gly Ser Trp Gly Asn Gly Gly His	
50	55
	60

Pro	Ser	Pro	Ser	Arg	Asn	Tyr	Gly	Asp	Gly	Thr	Pro	Tyr	Asp	His	Met
65					70					75					80
Thr	Ser	Arg	Asp	Leu	Gly	Ser	His	Asp	Asn	Leu	Ser	Pro	Pro	Phe	Val
				85					90					95	
Asn	Ser	Arg	Ile	Gln	Ser	Lys	Thr	Glu	Arg	Gly	Ser	Tyr	Ser	Ser	Tyr
			100					105					110		
Gly	Arg	Glu	Ser	Asn	Leu	Gln	Gly	Cys	His	Gln	Gln	Ser	Leu	Leu	Gly
		115					120					125			
Gly	Asp	Met	Asp	Met	Gly	Asn	Pro	Gly	Thr	Leu	Ser	Pro	Thr	Lys	Pro
	130					135					140				
Gly	Ser	Gln	Tyr	Tyr	Gln	Tyr	Ser	Ser	Asn	Asn	Pro	Arg	Arg	Arg	Pro
145					150					155					160
Leu	His	Ser	Ser	Ala	Met	Glu	Val	Gln	Thr	Lys	Lys	Val	Arg	Lys	Val
				165					170					175	
Pro	Pro	Gly	Leu	Pro	Ser	Ser	Val	Tyr	Ala	Pro	Ser	Ala	Ser	Thr	Ala
			180					185					190		
Asp	Tyr	Asn	Arg	Asp	Ser	Pro	Gly	Tyr	Pro	Ser	Ser	Lys	Pro	Ala	Thr
	195						200					205			
Ser	Thr	Phe	Pro	Ser	Ser	Phe	Phe	Met	Gln	Asp	Gly	His	His	Ser	Ser
	210					215					220				
Asp	Pro	Trp	Ser	Ser	Ser	Ser	Gly	Met	Asn	Gln	Pro	Gly	Tyr	Ala	Gly
225					230					235					240
Met	Leu	Gly	Asn	Ser	Ser	His	Ile	Pro	Gln	Ser	Ser	Ser	Tyr	Cys	Ser
			245						250				255		
Leu	His	Pro	His	Glu	Arg	Leu	Ser	Tyr	Pro	Ser	His	Ser	Ser	Ala	Asp
			260					265					270		
Ile	Asn	Ser	Ser	Leu	Pro	Pro	Met	Ser	Thr	Phe	His	Arg	Ser	Gly	Thr
	275						280					285			
Asn	His	Tyr	Ser	Thr	Ser	Ser	Cys	Thr	Pro	Pro	Ala	Asn	Gly	Thr	Asp
	290					295					300				
Ser	Ile	Met	Ala	Asn	Arg	Gly	Ser	Gly	Ala	Ala	Gly	Ser	Ser	Gln	Thr
305				310					315						320
Gly	Asp	Ala	Leu	Gly	Lys	Ala	Leu	Ala	Ser	Ile	Tyr	Ser	Pro	Asp	His
			325					330					335		
Thr	Asn	Asn	Ser	Phe	Ser	Ser	Asn	Pro	Ser	Thr	Pro	Val	Gly	Ser	Pro
			340					345					350		
Pro	Ser	Leu	Ser	Ala	Gly	Thr	Ala	Val	Trp	Ser	Arg	Asn	Gly	Gly	Gln
	355						360				365				
Ala	Ser	Ser	Ser	Pro	Asn	Tyr	Glu	Gly	Pro	Leu	His	Ser	Leu	Gln	Ser

370	375	380
Arg Ile Glu Asp Arg Leu Glu Arg Leu Asp Asp Ala Ile His Val Leu		
385	390	395
Arg Asn His Ala Val Gly Pro Ser Thr Ala Met Pro Gly Gly His Gly		400
405	410	415
Asp Met His Gly Ile Ile Gly Pro Ser His Asn Gly Ala Met Gly Gly		
420	425	430
Leu Gly Ser Gly Tyr Gly Thr Gly Leu Leu Ser Ala Asn Arg His Ser		
435	440	445
Leu Met Val Gly Thr His Arg Glu Asp Gly Val Ala Leu Arg Gly Ser		
450	455	460
His Ser Leu Leu Pro Asn Gln Val Pro Val Pro Gln Leu Pro Val Gln		
465	470	475
Ser Ala Thr Ser Pro Asp Leu Asn Pro Pro Gln Asp Pro Tyr Arg Gly		
485	490	495
Met Pro Pro Gly Leu Gln Gly Gln Ser Val Ser Ser Gly Ser Ser Glu		
500	505	510
Ile Lys Ser Asp Asp Glu Gly Asp Glu Asn Leu Gln Asp Thr Lys Ser		
515	520	525
Ser Glu Asp Lys Lys Leu Asp Asp Asp Lys Lys Asp Ile Lys Ser Ile		
530	535	540
Thr Arg Ser Arg Ser Ser Asn Asn Asp Asp Glu Asp Leu Thr Pro Glu		
545	550	555
Gln Lys Ala Glu Arg Glu Lys Glu Arg Arg Met Ala Asn Asn Ala Arg		
565	570	575
Glu Arg Leu Arg Val Arg Asp Ile Asn Glu Ala Phe Lys Glu Leu Gly		
580	585	590
Arg Met Val Gln Leu His Leu Lys Ser Asp Lys Pro Gln Thr Lys Leu		
595	600	605
Leu Ile Leu His Gln Ala Val Ala Val Ile Leu Ser Leu Glu Gln Gln		
610	615	620
Val Arg Glu Arg Asn Leu Asn Pro Lys Ala Ala Cys Leu Lys Arg Arg		
625	630	635
Glu Glu Glu Lys Val Ser Ser Glu Pro Pro Pro Leu Ser Leu Ala Gly		
645	650	655
Pro His Pro Gly Met Gly Asp Ala Ser Asn His Met Gly Gln Met		
660	665	670

<210> 3

<211> 31

<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	微型启动子	
<400>	3	
	agagggtata taatggaagc tcgacttcca g	31
<210>	4	
<211>	2204	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	ARN65表达盒	
<400>	4	
	gaattcacct gcaagaacac ctgcaagaac acctgcaaga acacctgcaa gaacacctgc	60
	aagaacacct gcaagaacac ctgcaagaac acctgcaaga acacctgcaa gaacacctgc	120
	aagaacacct gcaagaacac ctgcaagaga gggtatataa tggaagctcg atttccagat	180
	gcatacccaa cagcgaatgg ctgccttagg gacggacaaa gagctgagtg atttactgga	240
	tttcagtgcg atgtttttcac ctctgtgag cagtgggaaa aatggaccaa cttctttggc	300
	aagtggacat ttactggct caaatgtaga agacagaagt agctcagggt cctgggggaa	360
	tggaggacat ccaagcccggt ccaggaacta tggagatggg actccctatg accacatgac	420
	cagcagggac ctgggtcac atgacaatct ctctccacct ttgtcaatt ccagaatata	480
	aagtaaaaca gaaaggggct catactcatc ttatgggaga gaatcaaact tacagggttg	540
	ccaccagcag agtctccttg gaggtgacat ggatatgggc aaccaggaa ccctttcgcc	600
	caccaaacct ggttcccagt actatcagta ttctagcaat aatccccgaa ggaggcctct	660
	tcacagtagt gccatggagg tgcagacaaa gaaagttcga aaagttcctc caggtttgcc	720
	atcttcagtc tatgctccat cagcaagcac tgccgactac aatagggact cgccaggcta	780
	tccttcctcc aaaccagcaa ccagcacttt ccctagctcc ttcttcatgc aagatggcca	840
	tcacagcagt gacccttga gctcctccag tgggatgaat cagcctggct atgcaggaat	900
	gttgggcaac tcttctcata ttccacagtc cagcagctac tgtagcctgc atccacatga	960
	acgtttgagc tatccatcac actcctcagc agacatcaat tccagtcttc ctccgatgtc	1020
	cactttccat cgtagtggtta caaaccatta cagcacctct tcctgtacgc ctctgcca	1080
	cgggacagac agtataatgg caaatagagg aagcggggca gccggcagct ccagactgg	1140
	agatgctctg gggaaagcac ttgcttcgat ctattctcca gatcacacta acaacagctt	1200
	ttcatcaaac cttcaactc ctgttggtc tcctccatct ctctcagcag gcacagctgt	1260
	ttggtctaga aatggaggac aggcctcatc gtctcctaata tatgaaggac cttacactc	1320
	tttgcaaagc cgaattgaag atcgtttaga aagactggat gatgctatc atgttctccg	1380
	gaaccatgca gtgggcccac ccacagctat gcctggtggt catggggaca tgcattggaat	1440
	cattggacct tctcataatg gagccatggg tggtctgggc tcagggtatg gaaccggcct	1500
	tctttcagcc aacagacatt cactcatggt ggggacccat cgtgaagatg gcgtggccct	1560

gagaggcagc cattctcttc tgccaaacca ggttccggtt ccacagcttc ctgtccagtc	1620
tgcgacttcc cctgacctga acccacccca ggacccttac agaggcatgc caccaggact	1680
acaggggcag agtgtctcct ctggcagctc tgagatcaaa tccgatgacg agggatgatga	1740
gaacctgcaa gacacgaaat cttcggagga caagaaatta gatgacgaca agaaggatat	1800
caaatcaatt actagggtcaa gatctagcaa taatgacgat gaggacctga caccagagca	1860
gaaggcagag cgtgagaagg agcggaggat ggccaacaat gcccagagagc gtctgcgggt	1920
ccgtgacatc aacgaggctt tcaaagagct cggccgcatg gtgcagctcc acctcaagag	1980
tgacaagccc cagaccaagc tctgatacct ccaccaggcg gtggccgtca tctcagttct	2040
ggagcagcaa gtccgagaaa ggaatctgaa tccgaaagct gcgtgtctga aaagaaggga	2100
ggaagagaag gtgtcctcag agcctcccc tctctccttg gccggccac acctggaat	2160
gggagacgca tcgaatcaca tgggacagat gtaactcgag gata	2204
<210> 5	
<211> 2063	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> ARN60表达盒	
<400> 5	
gaattcagag ggtatataat ggaagctcga cttccagatg catcaccaac agcgaatggc	60
tgccttaggg acggacaaag agctgagtga tttactggat ttcagtgcga tgttttcacc	120
tctgtgagc agtgggaaaa atggaccaac ttctttggca agtggacatt ttactggctc	180
aaatgtagaa gacagaagta gctcagggtc ctgggggaat ggaggacatc caagcccgtc	240
caggaactat ggagatggga ctccctatga ccacatgacc agcagggacc ttgggtcaca	300
tgacaatctc tctccacctt ttgtcaattc cagaatacaa agtaaaacag aaaggggctc	360
atactcatct tatgggagag aatcaaactt acagggttgc caccagcaga gtctccttgg	420
aggtgacatg gatatgggca acccaggaac cttttgccc accaaacctg gttcccagta	480
ctatcagtat tctagcaata atccccgaag gaggcctctt cacagtagtg ccatggaggt	540
gcagacaaaag aaagttcgaa aagttcctcc aggtttgcca tcttcagtct atgtccatc	600
agcaagcact gccgactaca atagggactc gccaggetat ctttctcca aaccagcaac	660
cagcactttc ctagctcctt tcttcatgca agatggccat cacagcagtg acccttggag	720
ctcctccagt gggatgaatc agcctggcta tgcaggaatg ttgggcaact cttctcatat	780
tccacagtcc agcagctact gtagcctgca tccacatgaa cgtttgagct atccatcaca	840
ctcctcagca gacatcaatt ccagtcttcc tccgatgtcc actttccatc gtagtggtac	900
aaaccattac agcacctctt cctgtacgcc tctgccaac gggacagaca gtataatggc	960
aaatagagga agcggggcag ccggcagctc ccagactgga gatgctctgg ggaaagcact	1020
tgttcgatc tattctccag atcacactaa caacagcttt tcatcaaacc cttcaactcc	1080
tgttggtctt cctccatctc tctcagcagg cacagctgtt tggcttagaa atggaggaca	1140
ggcctcatcg tctcctaatt atgaaggacc cttacactct ttgcaaagcc gaattgaaga	1200
tcgttttagaa agactggatg atgctattca tgttctccgg aacctgcag tgggcccac	1260

cacagctatg cctgggtggc atggggacat gcatggaatc attggacctt ctcataatgg	1320
agccatgggt ggtctgggct cagggtatgg aaccggcctt ctttcagcca acagacattc	1380
actcatgggtg gggacccatc gtgaagatgg cgtggccctg agaggcagcc attctcttct	1440
gccaaaccag gttccggttc cacagcttcc tgtccagtct gcgacttccc ctgacctgaa	1500
cccaccccag gacccttaca gaggcattgcc accaggacta caggggcaga gtgtctcctc	1560
tggcagctct gagatcaaat ccgatgacga gggatgatgag aacctgcaag acacgaaatc	1620
ttcggaggac aagaaaattag atgacgacaa gaaggatata aatcaatta ctaggtcaag	1680
atctagcaat aatgacgatg aggacctgac accagagcag aaggcagagc gtgagaagga	1740
gcggaggatg gccaaacaatg cccgagagcg tctgcgggtc cgtgacatca acgaggcttt	1800
caaagagctc ggccgcatgg tgcagctcca cctcaagagt gacaagcccc agaccaagct	1860
cctgatcctc caccaggcgg tggccgtcat cctcagtctg gagcagcaag tccgagaaag	1920
gaatctgaat ccgaaagctg cgtgtctgaa aagaaggag gaagagaagg tgtctcaga	1980
gcctccccct ctctccttgg ccggcccaca ccctggaatg ggagacgcat cgaatcacat	2040
gggacagatg taactcgagg ata	2063

<210> 6

<211> 2075

<212> DNA

<213> Artificial Sequence

<220>

<223> ARN61表达盒

<400> 6

gaattcaaca cctgcaagag aggttatata atggaagctc gacttccaga tgcattacca	60
acagcgaatg gctgcccttag ggacggacaa agagctgagt gatttactgg atttcagtgc	120
gatgttttca cctcctgtga gcagtgggaa aaatggacca acttctttgg caagtggaca	180
ttttactggc tcaaattgtag aagacagaag tagctcaggg tcctggggga atggaggaca	240
tccaagccccg tccaggaact atggagatgg gactccctat gaccacatga ccagcaggga	300
ccttgggtca catgacaatc tctctccacc tttgtcaat tccagaatac aaagtaaaac	360
agaaaggggc tcatactcat cttatgggag agaataaac ttacagggtt gccaccagca	420
gagtctcctt ggaggtgaca tggatatggg caaccaggga accctttcgc ccaccaaac	480
tggttcccag tactatcagt attctagcaa taatccccga aggaggctc ttcacagtag	540
tgccatggag gtgcagacaa agaaagtctg aaaagtctct ccaggtttgc catcttcagt	600
ctatgctcca tcagcaagca ctgccgacta caatagggac tcgccaggct atccttcctc	660
caaaccagca accagcaatt tccctagctc cttcttcatg caagatggcc atcacagcag	720
tgacccttgg agctcctcca gtgggatgaa tcagcctggc tatgcaggaa tgttgggcaa	780
ctcttctcat attccacagt ccagcagcta ctgtagcctg catccacatg aacgtttgag	840
ctatccatca cactcctcag cagacatcaa ttccagtctt cctccgatgt ccactttcca	900
tcgtagtgg acaaaccatt acagcacctc ttctgtacg cctcctgcca acgggacaga	960
cagtataatg gcaaataagag gaagcggggc agccggcagc tcccagactg gagatgctct	1020
ggggaaagca cttgcttcca tctattctcc agatcacact aacaacagct tttcatcaaa	1080

```

cccttcaact cctgttggct ctctccatc tctctcagca ggcacagctg tttggtctag 1140
aaatggagga caggcctcat cgtctcctaa ttatgaagga cccttacact ctttgcaaag 1200
ccgaattgaa gatcgttttag aaagactgga tgatgctatt catgttctcc ggaaccatgc 1260
agtgggcccc tccacagcta tgcctggtgg tcatggggac atgcatggaa tcattggacc 1320
ttctcataat ggagccatgg gtggtctggg ctcagggtat ggaaccggcc ttctttcagc 1380
caacagacat tcaactcatg tggggaccca tcgtgaagat ggctggccc tgagaggcag 1440
ccattctctt ctgcccacc aggttccggt tccacagctt cctgtccagt ctgcgacttc 1500
ccctgacctg aaccaccccc aggacctta cagaggcatg ccaccaggac tacaggggca 1560
gagtgtctcc tctggcagct ctgagatcaa atccgatgac gagggtgatg agaacctgca 1620
agacacgaaa tcttcggagg acaagaaatt agatgacgac aagaaggata tcaaatcaat 1680
tactaggtca agatctagca ataatgacga tgaggacctg acaccagagc agaaggcaga 1740
gcgtgagaag gagcggagga tggccaacaa tgcccagag cgtctgcggg tccgtgacat 1800
caacgaggct ttcaaagagc tcggccgcat ggtgcagctc cacctcaaga gtgacaagcc 1860
ccagaccaag ctctgatcc tccaccaggc ggtggccgctc atctcagtc tggagcagca 1920
agtccgagaa aggaatctga atccgaaagc tgcgtgtctg aaaagaagg aggaagagaa 1980
ggtgtcctca gagcctcccc ctctctcctt ggccggcccc caccctggaa tgggagacgc 2040
atcgaatcac atgggacaga tgtaactcga ggata 2075

```

<210> 7

<211> 2099

<212> DNA

<213> Artificial Sequence

<220>

<223> ARN62表达盒

<400> 7

```

gaattcaaca cctgcaagaa cacctgcaag aacacctgca agagagggtata tataatggaa 60
gctcgacttc cagatgcatc accaacagcg aatggctgcc ttagggacgg acaaagagct 120
gagtgatttta ctggatttca gtgcgatgtt ttacacctct gtgagcagtg ggaaaaatgg 180
accaaacttct ttggcaagtg gacattttac tggtcctaat gtagaagaca gaagtagctc 240
agggtccttg gggaatggag gacatccaag cccgtccagg aactatggag atgggactcc 300
ctatgaccac atgaccagca gggaccttgg gtcacatgac aatctctctc caccttttgt 360
caattccaga atacaaaagta aaacagaaaag gggctcatac tcattcttatg ggagagaatc 420
aaacttacag ggttgccacc agcagagtct ccttgagggt gacatggata tgggcaaccc 480
aggaaccctt tcgcccacca aacctgggtc ccagtactat cagtattcta gcaataatcc 540
ccgaaggagg cctcttcaca gtagtgccat ggaggtgcag acaaagaaag ttcgaaaagt 600
tcctccaggt ttgccatctt cagtctatgc tccatcagca agcactgccg actacaatag 660
ggactcgcca ggctatcctt cctccaaacc agcaaccagc actttcccta gctccttctt 720
catgcaagat ggccatcaca gcagtgacct ttggagctcc tccagtggga tgaatcagcc 780
tggtatgca ggaatgttgg gcaactcttc tcatattcca cagtccagca gctactgtag 840
cctgcatcca catgaacgtt tgagctatcc atcacactcc tcagcagaca tcaattccag 900

```

tcttcctccg atgtccactt tccatcgtag tggtaaaaac cattacagca cctcttcctg	960
tacgcctcct gccaacggga cagacagtat aatggcaaat agaggaagcg gggcagccgg	1020
cagctcccag actggagatg ctctggggaa agcacttgct tcgatctatt ctccagatca	1080
cactaacaac agcttttcat caaaccttc aactcctggt ggctctcctc catctctctc	1140
agcaggcaca gctgtttggt ctagaaatgg aggacaggcc tcatcgtctc ctaattatga	1200
aggaccctta cactctttgc aaagccgaat tgaagatcgt ttagaaagac tggatgatgc	1260
tattcatggt ctccggaacc atgcagtggg cccatccaca gctatgcctg gtggatcatgg	1320
ggacatgcat ggaatcattg gaccttctca taatggagcc atgggtgggc tgggctcagg	1380
gtatggaacc ggcttcttt cagccaacag acattcactc atgggtgggga cccatcgtga	1440
agatggcgtg gccctgagag gcagccattc tcttctgcca aaccagggtc cggttccaca	1500
gcttcctgtc cagtctgcga cttccctga cctgaacca ccccaggacc cttacagagg	1560
catgccacca ggactacagg ggacagagtgt ctctctggc agctctgaga tcaaatccga	1620
tgacgagggt gatgagaacc tgcaagacac gaaatcttcg gaggacaaga aattagatga	1680
cgacaagaag gatatacaat caattactag gtcaagatct agcaataatg acgatgagga	1740
cctgacacca gagcagaagg cagagcgtga gaaggagcgg aggatggcca acaatgcccg	1800
agagcgtctg cgggtccgtg acatcaacga ggctttcaaa gagctcgcc gcattggtgca	1860
gctccacctc aagagtgaca agccccagac caagctcctg atcctccacc aggcggtggc	1920
cgatcatctc agtctggagc agcaagtccg agaaaggaat ctgaatccga aagctgcgtg	1980
tctgaaaaga agggaggaag agaaggtgtc ctgagagcct cccctctctc cttggccgg	2040
cccacaccct ggaatgggag acgcatcgaa tcacatggga cagatgtaac tcgaggata	2099

<210> 8

<211> 2135

<212> DNA

<213> Artificial Sequence

<220>

<223> ARN63表达盒

<400> 8

gaattcaaca cctgcaagaa cacctgcaag aacacctgca agaacctg caagaacacc	60
tgcaagaaca cctgcaagag aggttatata atggaagtc gacttccaga tgcattacca	120
acagcgaatg gctgccttag ggacggacaa agagctgagt gatttactgg atttcagtgc	180
gatgttttca cctcctgtga gcagtgggaa aatggacca acttctttgg caagtggaca	240
ttttactggc tcaaatgtag aagacagaag tagctcaggg tcctggggga atggaggaca	300
tccaagcccc tccaggaaact atggagatgg gactccctat gaccacatga ccagcaggga	360
ccttgggtca catgacaatc tctctccacc tttgtcaat tccagaatac aaagtaaaac	420
agaaaggggc tcatactcat cttatgggag agaataaac ttacagggtt gccaccagca	480
gagtctcctt ggaggtgaca tggatatggg caaccagga accctttcgc ccaccaaac	540
tggttcccag tactatcagt attctagcaa taatccccga aggagcctc ttcacagtag	600
tgccatggag gtgcagacaa agaaagtctg aaaagtctc ccaggtttgc catcttcagt	660
ctatgctcca tcagcaagca ctgccgacta caataggac tcgccaggct atccttctc	720

caaaccagca accagcactt tccctagctc cttcttcatg caagatggcc atcacagcag	780
tgacccttgg agctcctcca gtgggatgaa tcagcctggc tatgcaggaa tggtgggcaa	840
ctcttctcat attccacagt ccagcagcta ctgtagcctg catccacatg aacgtttgag	900
ctatccatca cactcctcag cagacatcaa ttccagtctt cctccgatgt ccactttcca	960
tcgtagtggg acaaaccatt acagcacctc ttctgtacg cctcctgcca acgggacaga	1020
cagtataatg gcaaataagag gaagcggggc agccggcagc tcccagactg gagatgctct	1080
ggggaaaagca cttgcttcga tctattctcc agatcacact aacaacagct tttcatcaaa	1140
cccttcaact cctgttggct ctctccatc tctctcagca ggcacagctg tttggtctag	1200
aaatggagga caggcctcat cgtctcctaa ttatgaagga cccttacact ctttgcaaag	1260
ccgaattgaa gatcgttttag aaagactgga tgatgctatt catgttctcc ggaaccatgc	1320
agtgggcca tccacagcta tgctgtggg tcatggggac atgcatggaa tcattggacc	1380
ttctcataat ggagccatgg gtggtctggg ctcagggtat ggaaccggcc ttctttcagc	1440
caacagacat tcactcatgg tggggacca tcgtgaagat ggcgtggccc tgagaggcag	1500
ccattctctt ctgccaacc aggttccggt tccacagctt cctgtccagt ctgcgacttc	1560
ccctgacctg aaccacccc aggaccctta cagaggcatg ccaccaggac tacaggggca	1620
gagtgtctcc tctggcagct ctgagatcaa atccgatgac gagggtgatg agaacctgca	1680
agacacgaaa tcttcggagg acaagaaatt agatgacgac aagaaggata tcaaatcaat	1740
tactaggtca agatctagca ataatgacga tgaggacctg acaccagagc agaaggcaga	1800
gcgtgagaag gagcggagga tggccaacaa tgcccagag cgtctgcggg tccgtgacat	1860
caacgaggct ttcaaagagc tcggccgcat ggtgcagctc cacctcaaga gtgacaagcc	1920
ccagaccaag ctctgatcc tccaccaggc ggtggccgctc atctcagtc tggagcagca	1980
agtccgagaa aggaatctga atccgaaagc tgctgtctg aaaagaaggg aggaagagaa	2040
ggtgtcctca gagcctcccc ctctctcctt ggccggccca caccctggaa tgggagacgc	2100
atcgaatcac atgggacaga tgtaactcga ggata	2135

<210> 9

<211> 6294

<212> DNA

<213> Artificial Sequence

<220>

<223> PAAV(exp) -SYN1>hTCF4

<400> 9

cctgcaggca gctgcgcgt cgtcgtctca ctgaggccgc ccgggcaaag cccgggcgtc	60
gggcgacctt tggtcgccc gctcagtgat gcgagcgagc gcgcagagag ggagtggcca	120
actccatcac taggggttcc ttctagacaa ctttgtatag aaaagttgct gcagagggcc	180
ctgcgtatga gtgcaagtgg gttttaggac caggatgagg cggggtgggg gtgcctacct	240
gacgaccgac cccgaccac tggacaagca cccaaccccc attccccaaa ttgcgcatcc	300
cctatcagag agggggaggg gaaacaggat gcggcgaggc gcgtgcgcac tgccagcttc	360
agcaccgcgg acagtgcctt cgccccgcc tggcggcgcg cgccaccgcc gcctcagcac	420
tgaaggcgcg ctgacgtcac tcgccgtcc cccgcaaact ccccttccc gccaccttgg	480

tcgcgtccgc	gccgccgccg	gcccagccgg	accgcaccac	gcgaggcgcg	agataggggg	540
gcacggggcg	gaccatctgc	gctgcggcgc	cggcgactca	gcgctgcctc	agtctgcggt	600
gggcagcgga	ggagtcgtgt	cgtgcctgag	agcgcagcaa	gtttgtacaa	aaaagcaggc	660
tgccaccatg	catcaccaac	agcgaatggc	tgccttaggg	acggacaaaag	agctgagtga	720
tttactggat	ttcagtgcga	tgttttcacc	tcctgtgagc	agtgggaaaa	atggaccaac	780
ttctttggca	agtggacatt	ttactggctc	aaatgtagaa	gacagaagta	gctcagggtc	840
ctgggggaat	ggaggacatc	caagcccgtc	caggaactat	ggagatggga	ctccctatga	900
ccacatgacc	agcagggacc	ttgggtcaca	tgacaatctc	tctccacctt	ttgtcaattc	960
cagaatacaa	agtaaaacag	aaaggggctc	atactcatct	tatgggagag	aatcaaactt	1020
acagggttgc	caccagcaga	gtctccttgg	aggtgacatg	gatatgggca	accaggaac	1080
cctttcgccc	accaaacctg	gttcccagta	ctatcagtat	tctagcaata	atccccgaag	1140
gaggcctctt	cacagtagtg	ccatggaggt	acagacaaag	aaagttcgaa	aagtctctcc	1200
aggtttgcca	tcttcagtct	atgtctccatc	agcaagcact	gccgactaca	atagggactc	1260
gccaggetat	ccttcctcca	aaccagcaac	cagcactttc	cctagctcct	tcttcatgca	1320
agatggccat	cacagcagtg	acccttggag	ctcctccagt	gggatgaatc	agcctggcta	1380
tgcaggaatg	ttgggcaact	cttctcatat	tccacagtcc	agcagctact	gtagcctgca	1440
tccacatgaa	cgtttgagct	atccatcaca	ctcctcagca	gacatcaatt	ccagtcttcc	1500
tccgatgtcc	actttccatc	gtagtggtag	aaaccattac	agcacctctt	cctgtacgcc	1560
tcctgccaac	gggacagaca	gtataatggc	aaatagagga	agcggggcag	ccggcagctc	1620
ccagactgga	gatgctctgg	ggaaagcact	tgcttcgatc	tattctccag	atcacactaa	1680
caacagcttt	tcatcaaacc	cttcaactcc	tgttggtctt	cctccatctc	tctcagcagg	1740
cacagctggt	tgggtctagaa	atggaggaca	ggcctcatcg	tctcctaatt	atgaaggacc	1800
cttacactct	ttgcaaagcc	gaattgaaga	tcgttttagaa	agactggatg	atgctattca	1860
tgttctccgg	aaccatgcag	tggggccatc	cacagctatg	cctggtgggc	atggggacat	1920
gcatggaatc	attggacctt	ctcataatgg	agccatgggt	ggtctgggct	cagggtatgg	1980
aaccggcctt	ctttcagcca	acagacattc	actcatgggt	gggacccatc	gtgaagatgg	2040
cgtggccctg	agaggcagcc	attctcttct	gccaaaccag	gttccgggtc	cacagcttcc	2100
tgtccagtct	gcgacttccc	ctgacctgaa	cccaccccag	gacccttaca	gaggcatgcc	2160
accaggacta	caggggcaga	gtgtctctc	tggcagctct	gagatcaa	ccgatgacga	2220
gggtgatgag	aacctgcaag	acacgaaatc	ttcggaggac	aagaaattag	atgacgacaa	2280
gaaggatatc	aaatcaatta	ctaggtcaag	atctagcaat	aatgacgatg	aggacctgac	2340
accagagcag	aaggcagagc	gtgagaagga	gcggaggatg	gccaacaatg	cccagagcgc	2400
tctgcgggtc	cgtgacatca	acgaggcttt	caaagagctc	ggccgcatgg	tgcagctcca	2460
cctcaagagt	gacaagcccc	agaccaagct	cctgatectc	caccaggcgc	tggccgtcat	2520
cctcagtctg	gagcagcaag	tccgagaaaag	gaatctgaat	ccgaaagctg	cgtgtctgaa	2580
aagaaggagg	gaagagaagg	tgtcctcaga	gcctccccct	ctctccttgg	ccggcccaca	2640
ccctggaatg	ggagacgcat	cgaatcacat	gggacagatg	taaaccagc	tttcttgtac	2700
aaagtgggaa	ttccgataat	caacctctgg	attacaaaat	ttgtgaaaga	ttgactggta	2760
ttcttaacta	tgttgctcct	tttacgctat	gtggatacgc	tgctttaatg	cctttgtatc	2820

atgctattgc	ttcccgtatg	gctttcattt	tctcctcctt	gtataaatcc	tggttgctgt	2880
ctcttttatga	ggagtgtgtg	cccgttgtca	ggcaacgtgg	cgtggtgtgc	actgtgtttg	2940
ctgacgcaac	ccccactggg	tggggcattg	ccaccacctg	tcagctcctt	tccgggactt	3000
tcgctttccc	cctccctatt	gccacggcgg	aactcatcgc	cgcctgcctt	gcccgtgct	3060
ggacaggggc	tcggctgttg	ggcactgaca	attccgtggg	gttgctgggg	aagctgacgt	3120
cctttccatg	gctgctcgcc	tgtgttgcca	cctggattct	gcgcgggacg	tccttctgct	3180
acgtcccttc	ggccctcaat	ccagcggacc	ttccttcccg	cggcctgctg	ccggctctgc	3240
ggcctcttcc	gcgtcttcgc	cttcgccttc	agacgagtcg	gatctccctt	tgggccgcct	3300
ccccgcacg	ggaattccta	gagctcgctg	atcagcctcg	actgtgcctt	ctagttgccca	3360
gccatctgtt	gtttgcccct	ccccgtgcc	ttccttgacc	ctggaagggtg	ccactcccac	3420
tgtcctttcc	taataaaaatg	aggaaattgc	atcgcatgtt	ctgagtaggt	gtcattctat	3480
tctggggggg	gggggtgggg	aggacagcaa	gggggaggat	tggaagaga	atagcaggca	3540
tgtgggggag	ggccgcagga	acccttagtg	atggagttag	ccactccctc	tctgcgcgct	3600
cgtcgcctca	ctgaggccgg	gcgaccaaag	gtcgcccgac	gcccgggctt	tgcccgggcg	3660
gcctcagtga	gcgagcgagc	gcgcagctgc	ctgcaggggc	gcctgatgcg	gtattttctc	3720
cttacgcac	tgtgcggtat	ttcacaccgc	atacgtcaaa	gcaaccatag	tacgcgcctt	3780
gtagcggcgc	attaagcgcg	gcgggggttg	tggttacgcg	cagcgtgacc	gctacacttg	3840
ccagcgcctt	agcgcggcgt	cctttcgctt	tcttcccttc	ctttctcgcc	acgttcgccg	3900
gctttccccg	tcaagctcta	aatcgggggc	tccctttagg	gttccgattt	agtgttttac	3960
ggcacctcga	ccccaaaaaa	cttgatttgg	gtgatggttc	acgtagtggg	ccatcgccct	4020
gatagacggg	ttttcgccct	ttgacgttgg	agtccacgtt	ctttaatagt	ggactcttgt	4080
tccaaactgg	aacaacactc	aactctatct	cgggctattc	ttttgattta	taagggattt	4140
tgccgatttc	ggtctatttg	ttaaaaaatg	agctgattta	acaaaaattt	aacgcgaatt	4200
ttaacaaaat	attaacgttt	acaattttat	ggtgcactct	cagtacaatc	tgctctgatg	4260
ccgcatagtt	aagccagccc	cgacaccgcg	caacaccgcg	tgacgcgccc	tgacgggctt	4320
gtctgctccc	ggcatccgct	tacagacaag	ctgtgaccgt	ctccgggagc	tgcatgtgtc	4380
agagggtttt	accgtcatca	ccgaaacgcg	cgagacgaaa	gggcctcgtg	atagcctat	4440
ttttataggt	taatgtcatg	ataataatgg	tttcttagac	gtcaggtggc	acttttcggg	4500
gaaatgtgcg	cggaaccctt	atttgtttat	ttttctaaat	acattcaaat	atgtatccgc	4560
tcatgagaca	ataaccctga	taaatgcttc	aataatattg	aaaaaggaag	agtatgagta	4620
ttcaacattt	ccgtgtcgcc	cttatccctt	tttttgccgc	atgttgccct	cctgtttttg	4680
ctcaccacga	aacgctgggt	aaagtaaaag	atgctgaaga	tcagttgggt	gcacgagtgg	4740
gttacatcga	actggatctc	aacagcggta	agatccttga	gagttttcgc	cccgaagaac	4800
gttttccaat	gatgagcact	tttaaagtgc	tgctatgtgg	cgcggtatta	tcccgatttg	4860
acgccgggca	agagcaactc	ggtcgccgca	tacactattc	tcagaatgac	ttggttgagt	4920
actcaccagt	cacagaaaag	catcttacgg	atggcatgac	agtaagagaa	ttatgcagtg	4980
ctgccataac	catgagtgat	aacactgcgg	ccaacttact	tctgacaacg	atcgaggagc	5040
cgaaggagct	aaccgctttt	ttgcacaaca	tgggggatca	tgtaactcgc	cttgatcggt	5100
gggaaccgga	gctgaatgaa	gccataccaa	acgacgagcg	tgacaccacg	atgcctgtag	5160

```

caatggcaac aacgttgccg aaactattaa ctggcgaact acttactcta gcttcccggc 5220
aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 5280
ttccggctgg ctggtttatt gctgataaat ctggagccgg tgagcgtgga agccgcggta 5340
tcattgcagc actggggcca gatggttaagc cctcccgtat cgtagttatc tacacgacgg 5400
ggagtcaggc aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga 5460
ttaagcattg gtaactgtca gaccaagttt actcatatat acttttagatt gattttaaac 5520
ttcatttttta atttaaaagg atctaggtga agatcctttt tgataatctc atgacaaaaa 5580
tcccttaacg tgagttttcg ttccactgag cgtcagaccc cgtagaaaag atcaaaggat 5640
cttcttgaga tccttttttt ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc 5700
taccagcggg ggtttgtttg ccggatcaag agtaccacac tctttttccg aaggtaactg 5760
gcttcagcag agcgcagata ccaaatactg ttcttctagt gtagccgtag ttaggccacc 5820
acttcaagaa ctctgtagca ccgcctacat acctcgtctt gctaactctg ttaccagtgg 5880
ctgctgccag tggcgataag tcgtgtctta ccgggttgga ctcaagacga tagttaccgg 5940
ataaggcgca gcggtcgggc tgaacggggg gtctgtgcac acagcccagc ttggagcgaa 6000
cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc acgcttcccg 6060
aagggagaaa ggcgacagg tatccggtta gcggcagggt cggaacagga gagcgcacga 6120
gggagcttcc agggggaaac gcctggtatc tttatagtcc tgtcgggttt cgccacctct 6180
gacttgagcg tcgatttttg tgatgctcgt caggggggcg gaggcctatgg aaaaacgcca 6240
gcaacgcggc ctttttacgg ttcttggcct tttgctggcc ttttgctcac atgt 6294

```

<210> 10

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> microE5盒

<400> 10

cacctgcaag aa

12

<210> 11

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> gRNA序列

<400> 11

ctttataagc ccgcagttcc

20

<210> 12

<211> 20

<212> DNA

<213> Artificial Sequence

<220>		
<223>	gRNA序列	
<400>	12	
ccgcagttcc	cggatgtgaa	20
<210>	13	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	13	
gtcgaccagc	accgccatct	20
<210>	14	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	14	
ggtaaacaga	gcgccctagag	20
<210>	15	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	15	
cattcacatc	cgggaactgc	20
<210>	16	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	16	
acataggaag	gtacgacttc	20
<210>	17	
<211>	20	
<212>	DNA	

<213> Artificial Sequence	
<220>	
<223> gRNA序列	
<400> 17	
tttacgtacc agacatagga	20
<210> 18	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> gRNA序列	
<400> 18	
ttcctttacg taccagacat	20
<210> 19	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> gRNA序列	
<400> 19	
ttcctatgtc tggtagctaa	20
<210> 20	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> gRNA序列	
<400> 20	
aagtcgtacc ttcctatgtc	20
<210> 21	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> gRNA序列	
<400> 21	
cacgcttggc ccggccatat	20
<210> 22	
<211> 20	

<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	22	
	ctttgcatat tcaccacgct	20
<210>	23	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	23	
	agcgtctgac agcagcgccg	20
<210>	24	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	24	
	tcaactttgc gcagcggagc	20
<210>	25	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列	
<400>	25	
	agacgctcaa ctttgcgcag	20
<210>	26	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	gRNA序列-干扰	
<400>	26	
	aaatgtgaga tcagagtaat	20
<210>	27	

<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 3b正向引物	
<400>	27	
	gtgccgaaac tacacttttg tg	22
<210>	28	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 3b反向引物	
<400>	28	
	acaagattat gcacctggct	20
<210>	29	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 8a正向引物	
<400>	29	
	tggaaagggg atgcttactc tc	22
<210>	30	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 8a反向引物	
<400>	30	
	aactctaattg acctccgcct	20
<210>	31	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 10a正向引物	
<400>	31	
	ttaccctaag gcagttgagt gga	23

<210>	32	
<211>	21	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 10a反向引物	
<400>	32	
	ttgtcaaaaa tccccctcgc a	21
<210>	33	
<211>	19	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 4/5正向引物	
<400>	33	
	cctcctgtga gcagtggga	19
<210>	34	
<211>	19	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 4/5反向引物	
<400>	34	
	ctggacgggc ttggatgtc	19
<210>	35	
<211>	24	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 9/10正向引物	
<400>	35	
	gccatcttca gtctatgctc catc	24
<210>	36	
<211>	22	
<212>	DNA	
<213>	Artificial Sequence	
<220>		
<223>	TCF4 9/10反向引物	
<400>	36	

tagggaaagt gctggttgct gg	22
<210> 37	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> TCF4-A正向引物	
<400> 37	
ggaaagcgggt ctatgctcca t	21
<210> 38	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> TCF4-A反向引物	
<400> 38	
tagggaaagt gctggttgct gg	22
<210> 39	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> TCF4 12/13正向引物	
<400> 39	
cttcctccga tgtccacttt cca	23
<210> 40	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> TCF4 12/13反向引物	
<400> 40	
cccgttcct ctatttgcca t	21
<210> 41	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> MPH正向引物	

<400> 41	
atcagggcgt gtccatgtct ca	22
<210> 42	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> MPH反向引物	
<400> 42	
cgggtaatgg cttcggggta	20
<210> 43	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> dCAS9正向引物	
<400> 43	
gccgacgcta atctggacaa agt	23
<210> 44	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> dCas9反向引物	
<400> 44	
tggtcagggt aaacagggtg atg	23
<210> 45	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> TBP正向引物	
<400> 45	
tgtatccaca gtgaatcttg gttg	24
<210> 46	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	

<223> TBP反向引物	
<400> 46	
ggttcgtggc tctcttatcc tc	22
<210> 47	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> CNTNAP2正向引物	
<400> 47	
tcacacagac caagatgagc caa	23
<210> 48	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> CNTNAP2反向引物	
<400> 48	
taggaagcga acctcgtgcc a	21
<210> 49	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> KCNQ1正向引物	
<400> 49	
ggacaaagac aatgggggtga ct	22
<210> 50	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> KCNQ1反向引物	
<400> 50	
gtgttgggct cttccttaca gaa	23
<210> 51	
<211> 22	
<212> DNA	
<213> Artificial Sequence	

<220>	
<223> TCF4B cDNA正向引物	
<400> 51	
atgcatcacc aacagcgaat gg	22
<210> 52	
<211> 32	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> TCF4B cDNA反向引物	
<400> 52	
atcctcgagt tacatctgtc ccatgtgatt cg	32
<210> 53	
<211> 59	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> minP TCF4	
<400> 53	
gaattcagag ggtatataat ggaagctcga cttccagatg catcaccaac agcgaatgg	59
<210> 54	
<211> 131	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> E-box-x6-minP TCF4	
<400> 54	
gaattcaaca cctgcaagaa cacctgcaag aacacctgca agaacctg caagaacacc	60
tgcaagaaca cctgcaagag agggtatata atggaagctc gacttccaga tgcattacca	120
acagcgaatg g	131
<210> 55	
<211> 200	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> E-box-x12-MinP TCF4	
<400> 55	
gaattcacct gcaagaacac ctgcaagaac acctgcaaga acacctgcaa gaacctgctc	60
aagaacacct gcaagaacac ctgcaagaac acctgcaaga acacctgcaa gaacctgctc	120

aagaacacct gcaagaacac ctgcaagaga gggtatataa tggaagctcg acttccagat	180
gcatcaccaa cagcgaatgg	200

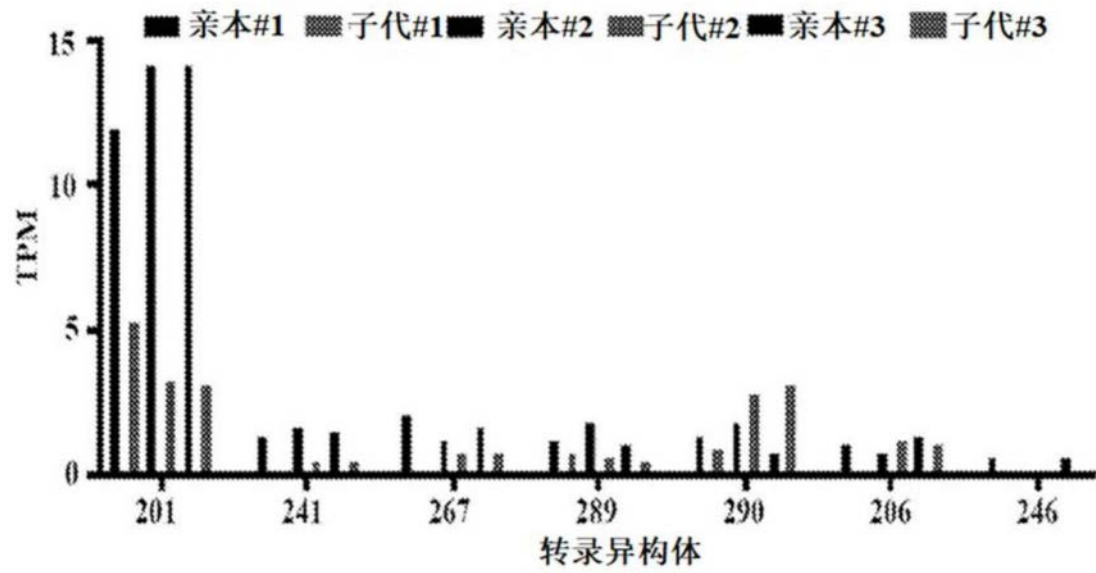


图1A

转录本名称	蛋白异构体
201	TCF4-B+
241	TCF4-F-
267	TCF4-B+
289	TCF4-B+
290	TCF4-D+
206	TCF4-D-
246	TCF4-A-

图1B

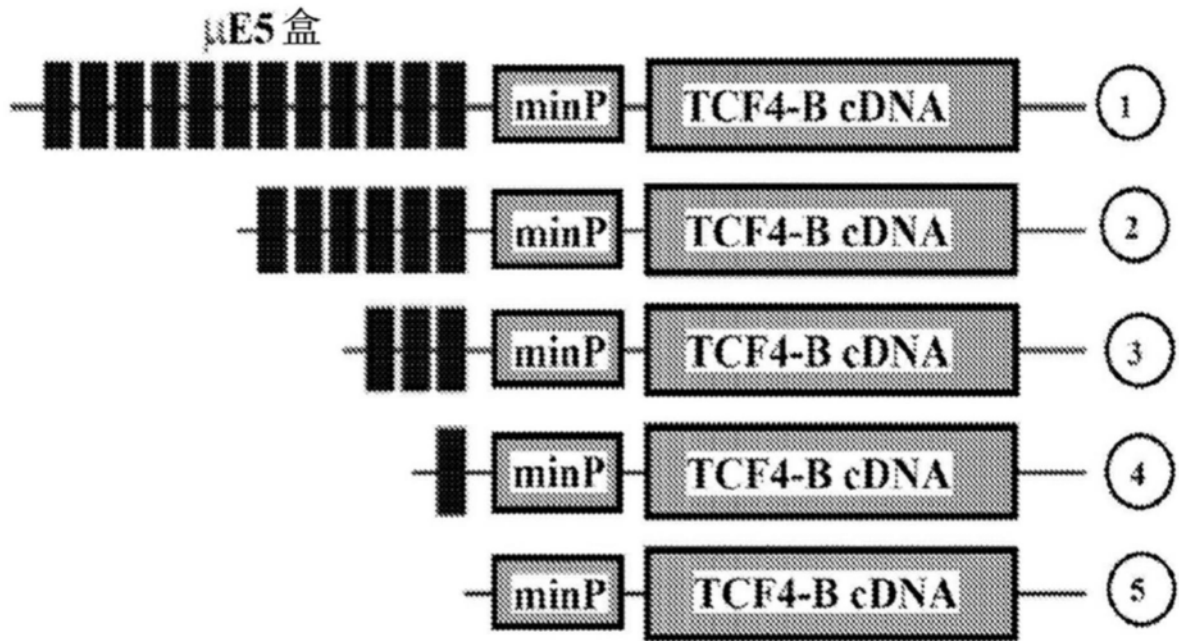


图2A

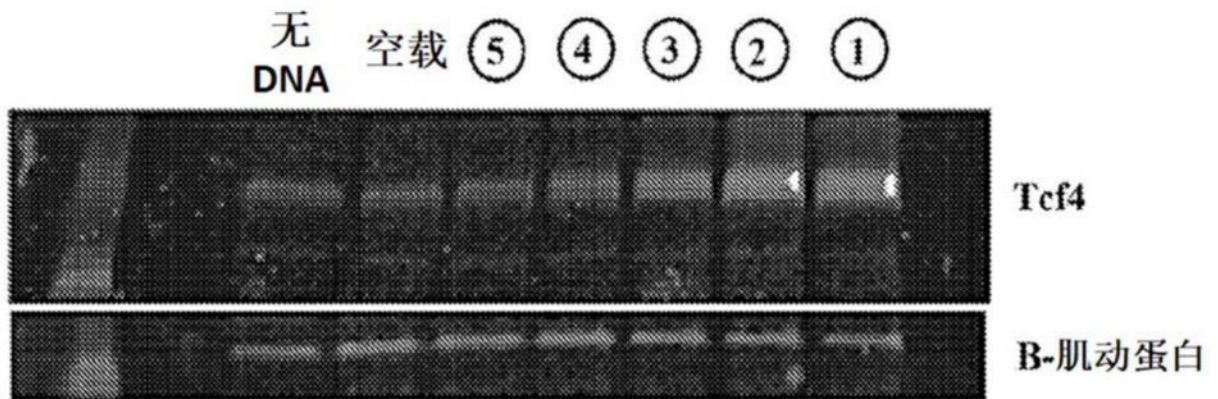


图2B



构建体#1；质粒名：ARN65**: 12 microE5盒 + minP + TCF4 编码序列 (SEQ ID NO:4)**

```

gaattcacctgcaagaacacctgcaagaacacctgcaagaacacctgcaagaacacctgcaagaaca
cctgcaagaacacctgcaagaacacctgcaagaacacctgcaagaacacctgcaagaacacctgcaa
gaacacctgcaagagaggggtatataatggaagctcgacttccagATGCATCACCAACAGCGAATGGC
TGCCTTAGGGACGGACAAAGAGCTGAGTGATTTACTGGATTTCAGTGCGATGTTTTACCTCCTGTG
AGCAGTGGGAAAAATGGACCAACTTCTTTGGCAAGTGGACATTTTACTGGCTCAAATGTAGAAGACA
GAAGTAGCTCAGGGTCTGGGGGAATGGAGGACATCCAAGCCCGTCCAGGAATATGGAGATGGGAC
TCCCTATGACCACATGACCAGCAGGGACCTTGGGTACATGACAATCTCTCTCCACCTTTTGTCAAT
TCCAGAATACAAAGTAAAACAGAAAGGGGCTCATACTCATCTTATGGGAGAGAATCAAACCTTACAGG
GTTGCCACCAGCAGAGTCTCCTTGGAGGTGACATGGATATGGGCAACCCAGGAACCTTTTCGCCAC
CAAACCTGGTTCCCAGTACTATCAGTATTCTAGCAATAATCCCCGAAGGAGGCCTCTTCACAGTAGT
GCCATGGAGGTGCAGACAAAGAAAGTTGAAAAGTTCCCTCCAGGTTTGCCATCTTCAGTCTATGCTC
CATCAGCAAGCACTGCCGACTACAATAGGGACTCGCCAGGCTATCCTTCCCTCAAACAGCAACCAG
CACTTTCCCTAGCTCCTTCTTCATGCAAGATGGCCATCACAGCAGTGACCCTTGGAGCTCCTCCAGT
GGGATGAATCAGCCTGGCTATGCAGGAATGTTGGGCAACTCTTCTCATATTCACAGTCCAGCAGCT
ACTGTAGCCTGCATCCACATGAACGTTTGAGCTATCCATCACACTCCTCAGCAGACATCAATTCCAG
TCTTCTCCGATGTCCACTTTCATCGTAGTGGTACAAACCATTACAGCACCTCTTCTGTACGCCT
CCTGCCAACGGGACAGACAGTATAATGGCAAATAGAGGAAGCGGGGAGCCGGCAGCTCCAGACTG
GAGATGCTCTGGGAAAGCACTTGCTTCGATCTATTCTCCAGATCACACTAACAACAGCTTTTCATC
AAACCTTCAACTCCTGTTGGCTCTCCTCCATCTCTCTCAGCAGGCACAGCTGTTTGGTCTAGAAAT
GGAGGACAGGCCTCATCGTCTCCTAATTATGAAGGACCCTTACACTCTTTGCAAAGCCGAATTGAAG
ATCGTTTAGAAAGACTGGATGATGCTATTTCATGTTCTCCGGAACCATGCAGTGGGCCCATCCACAGC
TATGCCTGGTGGTCATGGGGACATGCATGGAATCATTTGGACCTTCTCATAATGGAGCCATGGGTGGT
CTGGGCTCAGGGTATGGAACCGGCCTTCTTTCAGCCAACAGACATTCACTCATGGTGGGGACCCATC
GTGAAGATGGCGTGGCCCTGAGAGGCAGCCATTCTCTTCTGCCAAACCAGGTTCCGGTTCCACAGCT
TCCTGTCCAGTCTGCGACTTCCCCTGACCTGAACCCACCCAGGACCCTTACAGAGGCATGCCACCA
GGACTACAGGGGAGAGTGTCTCTCTGGCAGCTCTGAGATCAAATCCGATGACGAGGGTGATGAGA
ACCTGCAAGACACGAAATCTTCGAGGACAAGAAATTAGATGACGACAAGAAGGATATCAAATCAAT
TACTAGGTCAAGATCTAGCAATAATGACGATGAGGACCTGACACCAGAGCAGAAGGCAGAGCGTGAG
AAGGAGCGGAGGATGGCCAACAATGCCGAGAGCGTCTGCGGGTCCGTGACATCAACGAGGCTTTCA
AAGAGCTCGGCCGCATGGTGCAGCTCCACCTCAAGAGTGACAAGCCCCAGACCAAGCTCCTGATCCT
CCACCAGGCGGTGGCCGTCATCCTCAGTCTGGAGCAGCAAGTCCGAGAAAGGAATCTGAATCCGAAA
GCTGCGTGTCTGAAAAGAAGGGAGGAAGAGAAGGTGTCCTCAGAGCCTCCCCCTCTCTCCTTGGCCG
GCCCACACCTTGAATGGGAGACGCATCGAATCACATGGGACAGATGTAAc-cgaggata

```

特征：

名称	位置	长度	注释
人工启动子	1-178	178	小写
microE5 box	6-11	6	
microE5 box	18-23	6	
microE5 box	30-35	6	
microE5 box	42-47	6	
microE5 box	54-59	6	
microE5 box	66-71	6	
microE5 box	78-83	6	
microE5 box	90-95	6	
microE5 box	102-107	6	
microE5 box	114-119	6	
microE5 box	126-131	6	
microE5 box	138-143	6	
minP 微型启动子	148-178	31	下划线
TCF4 编码序列	179-2194	2016	大写

图4A

构建体#2**质粒名: ARN60****minP + TCF4 编码序列, 无microE5 盒****(SEQ ID NO: 5)**

```

gaattcagagggtatataatggaagctcgacttccagATGCATCACCAACAGCGAATGGCTGCCTTA
GGGACGGACAAAGAGCTGAGTGATTACTGGATTTCAGTGCGATGTTTTACCTCCTGTGAGCAGTG
GGAAAAATGGACCAACTTCTTTGGCAAGTGGACATTTTACTGGCTCAAATGTAGAAGACAGAAGTAG
CTCAGGGTCTTGGGGGAATGGAGGACATCCAAGCCCGTCCAGGAACTATGGAGATGGGACTCCCTAT
GACCACATGACCAGCAGGGACCTTGGGTACATGACAATCTCTCTCCACCTTTTGTCAATTCCAGAA
TACAAAGTAAAACAGAAAGGGGCTCATACTCATCTTATGGGAGAGAATCAAACCTTACAGGGTTGCCA
CCAGCAGAGTCTCCTTGGAGGTGACATGGATATGGGCAACCAGGAACCTTTTCGCCACCAAACCT
GGTTCCTCAGTACTATCAGTATTCTAGCAATAATCCCCGAAGGAGGCTCTTCACAGTAGTGCCATGG
AGGTGCAGACAAAGAAAGTTGCAAAAGTTCTCCAGGTTTGCCATCTTCAGTCTATGCTCCATCAGC
AAGCACTGCCGACTACAATAGGGACTCGCCAGGCTATCCTTCCTCCAAACCAGCAACCAGCACTTTC
CCTAGCTCCTTCTTCATGCAAGATGGCCATCACAGCAGTGACCCTTGGAGCTCCTCCAGTGGGATGA
ATCAGCCTGGCTATGCAGGAATGTTGGGCAACTCTTCTCATATTCCACAGTCCAGCAGCTACTGTAG
CCTGCATCCACATGAACGTTTGAGCTATCCATCACACTCCTCAGCAGACATCAATTCAGTCTTCCT
CCGATGTCCACTTTCCATCGTAGTGGTACAAACCATTACAGCACCTCTTCCTGTACGCCTCCTGCCA
ACGGGACAGACAGTATAATGGCAAATAGAGGAAGCGGGGCAGCCGGCAGCTCCAGACTGGAGATGC
TCTGGGGAAAGCACTTGCTTCGATCTATTCTCCAGATCACACTAACAACAGCTTTTCATCAAACCTT
TCAACTCCTGTTGGCTCTCCTCCATCTCTCTCAGCAGGCACAGCTGTTTGGTCTAGAAATGGAGGAC
AGGCCTCATCGTCTCCTAATTATGAAGGACCTTACACTCTTTGCAAAGCCGAATTGAAGATCGTTT
AGAAAGACTGGATGATGCTATTTCATGTTCTCCGGAACCATGCAGTGGGCCCATCCACAGCTATGCCT
GGTGGTTCATGGGGACATGCATGGAATCATTGGACCTTCTCATTAATGGAGCCATGGGTGGTCTGGGCT
CAGGGTATGGAACCGGCCCTTCTTTCAGCCAAACAGACATTCATCTCATGGTGGGGACCCATCGTGAAGA
TGGCGTGGCCCTGAGAGGCAGCCATTCTCTTCTGCCAAACCAGGTTCCGGTTCCACAGCTTCCTGTC
CAGTCTGCGACTTCCCCTGACCTGAACCCACCCAGGACCTTTACAGAGGCATGCCACCAGGACTAC
AGGGGCAGAGTGTCTCCTCTGGCAGCTCTGAGATCAAATCCGATGACGAGGGTGATGAGAACCTGCA
AGACACGAAATCTTCGGAGGACAAGAAATTAGATGACGACAAGAAGGATATCAAATCAATTACTAGG
TCAAGATCTAGCAATAATGACGATGAGGACCTGACACCAGAGCAGAAGGCAGAGCGTGAGAAGGAGC
GGAGGATGGCCAACAATGCCCCGAGAGCGTCTGCGGGTCCGTGACATCAACGAGGCTTTCAAAGAGCT
CGGCCGCATGGTGCAGCTCCACCTCAAGAGTGACAAGCCCCAGACCAAGCTCCTGATCCTCCACCAG
GCGGTGGCCGTCATCCTCAGTCTGGAGCAGCAAGTCCGAGAAAGGAATCTGAATCCGAAAGCTGCGT
GTCTGAAAAGAAGGGAGGAAGAGAAGGTGTCTCAGAGCCTCCCCCTCTCTCCTTGGCCGGCCACA
CCCTGGAATGGGAGACGCATCGAATCACATGGGACAGATGTAAActcgaggata

```

特征:

名称	位置	长度	注释
人工启动子	1-37	37	小写
minP 微型启动子	7-37	31	下划线
TCF4 编码序列	38-2053	2016	大写

图4B

质粒名: ARN61**1 microE5 盒 + minP + TCF4 编码序列****(SEQ ID NO: 6)**

gaattcaacacctgcaagagaggggtatataatggaagctcgacttccagATGCATCACCAACAGCGA
 ATGGCTGCCTTAGGGACGGACAAAGAGCTGAGTGATTTACTGGATTTCAGTGCGATGTTTTACCTC
 CTGTGAGCAGTGGGAAAAATGGACCAACTTCTTTGGCAAGTGGACATTTTACTGGCTCAAATGTAGA
 AGACAGAAGTAGCTCAGGGTCCTGGGGGAATGGAGGACATCCAAGCCCGTCCAGGAACTATGGAGAT
 GGGACTCCCTATGACCACATGACCAGCAGGGACCTTGGGTACATGACAATCTCTCTCCACCTTTTG
 TCAATTCCAGAATACAAAGTAAAACAGAAAGGGGGCTCATACCTCATCTTATGGGAGAGAATCAAACCTT
 ACAGGGTTGCCACCAGCAGAGTCTCCTTGGAGGTGACATGSATATGGGCAACCCAGGAACCTTTTCG
 CCCACCAAACCTGGTTCCCAGTACTATCAGTATTCTAGCAATAATCCCCGAAGGAGGCCTCTTCACA
 GTAGTGCCATGGAGGTGCAGACAAAGAAAGTTCGAAAAGTTCCTCCAGGTTTGCCATCTTCAGTCTA
 TGCTCCATCAGCAAGCACTGCCGACTACAATAGGGACTCGCCAGGCTATCCTTCCTCCAAACCAGCA
 ACCAGCACTTTCCCTAGCTCCTTCTTCATGCAAGATGGCCAACACAGCAGTGACCCTTGGAGCTCCT
 CCAGTGGGATGAATCAGCCTGGCTATGCAGGAATGTTGGGCAACTCTTCTCATATTCCACAGTCCAG
 CAGCTACTGTAGCCTGCATCCACATGAACGTTTGAGCTATCCATCACACTCCTCAGCAGACATCAAT
 TCCAGTCTTCTCCGATGTCCACTTTCCATCGTAGTGGTACAAACCATTACAGCACCTCTTCCTGTA
 CGCTCCTGCCAACGGGACAGACAGTATAATGGCAAATAGAGGAAGCGGGGCAGCCGGCAGCTCCCA
 GACTGGAGATGCTCTGGGGAAAGCACTTGCTTCGATCTATTCTCCAGATCACACTAACAACAGCTTT
 TCATCAAACCTTCAACTCCTGTTGGCTCTCCTCCATCTCTCTCAGCAGGCACAGCTGTTTGGTCTA
 GAAATGGAGGACAGGCCTCATCGTCTCCTAATTATGAAGGACCCTTACACTCTTTGCAAAGCCGAAT
 TGAAGATCGTTTAGAAAGACTGGATGATGCTATTTCATGTTCTCCGGAACCATGCAGTGGGCCCATCC
 ACAGCTATGCCTGGTGGTTCATGGGGACATGCATGGAATCATGGACCTTCTCATAATGGAGCCATGG
 GTGGTCTGGGCTCAGGGTATGGAACCGGCCTTCTTTAGCCAAACAGACATTCACTCATGGTGGGGAC
 CCATCGTGAAGATGGCGTGGCCCTGAGAGGCAGCCATTCTCTCTGCCAAACCAGGTTCCGGTTCCA
 CAGCTTCCTGTCCAGTCTGCGACTTCCCCTGACCTGAACCCACCCAGGACCCTTACAGAGGCATGC
 CACCAGGACTACAGGGGCAGAGTGTCTCCTCTGGCAGCTCTGAGATCAAATCCGATGACGAGGGTGA
 TGAGAACCTGCAAGACACGAAATCTTCGGAGGACAAGAAATAGATGACGACAAGAAGGATATCAAA
 TCAATTACTAGGTCAAGATCTAGCAATAATGACGATGAGGACCTGACACCAGAGCAGAAGGCAGAGC
 GTGAGAAGGAGCGGAGGATGGCCAACAATGCCCGAGAGCGTCTGCGGGTCCGTGACATCAACGAGGC
 TTTCAAAGAGCTCGGCCGCATGGTGCAGCTCCACCTCAAGAGTGACAAGCCCCAGACCAAGCTCCTG
 ATCCTCCACCAGGCGGTGGCCGTATCCTCAGTCTGGAGCAGCAAGTCCGAGAAAGGAATCTGAATC
 CGAAAGCTGCGTGTCTGAAAAGAAGGGAGGAAGAGAAGGTGCTCTCAGAGCCTCCCCCTCTCTCCTT
 GGCCGGCCACACCCTGGAATGGGAGACGCATCGAATCACATGGGACAGATGTAAActcgaggata

特征:

名称	位置	长度	注释
人工启动子	1-49	49	小写
microE5 box	9-14	6	
minP 微型启动子	19-49	31	下划线
TCF4 编码序列	50-2065	2016	大写

图4C

构建体#4**质粒名: ARN62****3 microE5盒 + minP + TCF4 编码序列 (SEQ ID NO:7)**

gaattcaacacctgcaagaacacctgcaagaacacctgcaagagaggggtatataaatggaagctcgac
 ttccagATGCATCACCAACAGCGAATGGCTGCCTTAGGGACGGACAAAGAGCTGAGTGATTTACTGG
 ATTTTCAGTGCGATGTTTTACCTCCTGTGAGCAGTGGGAAAAATGGACCAACTTCTTTGGCAAGTGG
 ACATTTTACTGGCTCAAATGTAGAAGACAGAAGTAGCTCAGGGTCCTGGGGGAATGGAGGACATCCA
 AGCCCGTCCAGGAACCTATGGAGATGGGACTCCCTATGACCACATGACCAGCAGGGACCTTGGGTCCAC
 ATGACAATCTCTCTCCACCTTTTGTCAATTCCAGAATACAAAGTAAACAGAAAGGGGCTCATACTC
 ATCTTATGGGAGAGAATCAAACCTTACAGGGTTGCCACCAGCAGAGTCTCCTTGGAGGTGACATGGAT
 ATGGGCAACCCAGGAACCTTTTCCCCACCAAACCTGGTTCCAGTACTATCAGTATTCTAGCAATA
 ATCCCCGAAGGAGGCCTCTTCACAGTAGTGCCATGGAGGTGCAGACAAAGAAAGTTTCGAAAAGTTCC
 TCCAGGTTTGGCATCTTCAGTCTATGCTCCATCAGCAAGCACTGCCGACTACAATAGGGACTCGCCA
 GGCTATCCTTCCCTCCAAACCAGCAACCAGCACTTTCCTAGCTCCTTCTTCATGCAAGATGGCCATC
 ACAGCAGTGACCCTTGGAGCTCCTCCAGTGGGATGAATCAGCCTGGCTATGCAGGAATGTTGGGCAA
 CTCTTCTCATATTCCACAGTCCAGCAGCTACTGTAGCCTGCATCCACATGAACGTTTGAGCTATCCA
 TCACACTCCTCAGCAGACATCAATTCCAGTCTTCCTCCGATGTCCACTTTCCATCGTAGTGGTACAA
 ACCATTACAGCACCTCTTCCTGTACGCCTCCTGCCAACGGGACAGACAGTATAATGGCAAATAGAGG
 AAGCGGGGACGCCGAGCTCCCAGACTGGAGATGCTCTGGGGAAAGCACTTGCTTCGATCTATTCT
 CCAGATCACACTAACAACAGCTTTTCATCAAACCTTCAACTCCTGTTGGCTCTCCTCCATCTCTCT
 CAGCAGGCACAGCTGTTTGGTCTAGAAATGGAGGACAGGCCCTCATCGTCTCCTAATTATGAAGGACC
 CTTACACTCTTTGCAAAGCCGAATTGAAGATCGTTTAGAAAGACTGGATGATGCTATTCTATGTTCTC
 CGGAACCATGCAGTGGGCCCATCCACAGCTATGCCTGGTGGTTCATGGGGACATGCATGGAATCATTG
 GACCTTCTCATAATGGAGCCATGGGTGGTCTGGGCTCAGGGTATGGAACCGGCCTTCTTTAGCCAA
 CAGACATTCACTCATGGTGGGGACCCATCGTGAAGATGGCGTGGCCCTGAGAGGCAGCCATTCTCTT
 CTGCCAAACCAGGTTCCGGTTCACAGCTTCCTGTCCAGTCTGCGACTTCCCCTGACCTGAACCCAC
 CCCAGGACCCTTACAGAGGCATGCCACCAGGACTACAGGGGACAGAGTGTCTCCTCTGGCAGCTCTGA
 GATCAAATCCGATGACGAGGGTGATGAGAACCTGCAAGACACGAAATCTTCGGAGGACAAGAAATTA
 GATGACGACAAGAAGGATATCAAATCAATTACTAGGTCAAGATCTAGCAATAATGACGATGAGGACC
 TGACACCAGAGCAGAAGGCAGAGCGTGAGAAGGAGCGGAGSATGGCCAACAATGCCCAGAGCGTCT
 GCGGGTCCGTGACATCAACGAGGCTTTCAAAGAGCTCGGCCGCATGGTGCAGCTCCACCTCAAGAGT
 GACAAGCCCCAGACCAAGCTCCTGATCCTCCACCAGGCGGTGGCCGTCTCCTCAGTCTGGAGCAGC
 AAGTCCGAGAAAGGAATCTGAATCCGAAAGCTGCGTGTCTGAAAAGAAGGGAGGAAGAGAAGGTGTC
 CTCAGAGCCTCCCCCTCTCTCCTTGGCCGGCCACACCCTGGAATGGGAGACGCATCGAATCACATG
 GGACAGATGTAAActcgaggata

特征:

名称	位置	长度	注释
人工启动子	1-73	73	小写
microE5 box	9-14	6	
microE5 box	21-26	6	
microE5 box	33-38	6	
minP 微型启动子	43-73	31	下划线
TCF4 编码序列	74-2089	2016	大写

图4D

构建体#5**质粒名: ARN63****6 microE5 盒 + minP + TCF4 编码序列 (SEQ ID NO:8)**

```

gaattcaacacacctgcaagaacacacctgcaagaacacacctgcaagaacacacctgcaaga
acacctgcaagagagaggtatataatggaagctcgacttccagATGCATCACCAACAGCGAATGGCTG
CCTTAGGGACGGACAAAGAGCTGAGTGATTTACTGGATTTTCACTGCGATGTTTTACCTCCTGTGAG
CAGTGGGAAAAATGGACCAACTTCTTTGGCAAGTGGACATTTTACTGGCTCAAATGTAGAAGACAGA
AGTAGCTCAGGGTCTTGGGGGAATGGAGGACATCCAAGCCCGTCCAGGAAGTATGGAGATGGGACTC
CCTATGACCACATGACCAGCAGGGACCTTGGGTACATGACAATCTCTCTCCACCTTTTGTCAATTC
CAGAATACAAAGTAAACAGAAAGGGGCTCATACTCATCTTATGGGAGAGAATCAAACCTTACAGGGT
TGCCACCAGCAGAGTCTCCTTGGAGGTGACATGGATATGGGCAACCCAGGAACCTTTTCGCCACCA
AACCTGGTTCCCGTACTATCAGTATTCTAGCAATAATCCCGAAGGAGGCTCTTCACAGTAGTGC
CATGGAGGTGCAGACAAAGAAAGTTTCGAAAAGTTCTCCAGGTTTGCCATCTTCAGTCTATGCTCCA
TCAGCAAGCACTGCCGACTACAATAGGGACTCGCCAGGCTATCCTTCCCTCCAAACCAGCAACCAGCA
CTTTCCTAGCTCCTTCTTCATGCAAGATGGCCATCACAGCAGTGACCCTTGGAGCTCCTCCAGTGG
GATGAATCAGCCTGGCTATGCAGGAATGTTGGGCAACTCTTCTCATATTCCACAGTCCAGCAGCTAC
TGATGCCTGCATCCACATGAACGTTTGAGCTATCCATCACACTCCTCAGCAGACATCAATTCCAGTC
TTCTTCCGATGTCCACTTTCCATCGTAGTGGTACAAACCAATACAGCACCTCTTCTGTACGCCTCC
TGCCAACGGGACAGACAGTATAATGGCAAATAGAGGAAGCGGGGAGCCGGCAGCTCCAGACTGGA
GATGCTCTGGGGAAAGCACTTGCTTCGATCTATTCTCCAGATCACACTAACAACAGCTTTTCATCAA
ACCCTTCAACTCCTGTTGGCTCTCCTCCATCTCTCTCAGCAGGCACAGCTGTTTGGTCTAGAAATGG
AGGACAGGCCTCATCGTCTCCTAATTATGAAGGACCCTTACACTCTTTGCAAAGCCGAATTGAAGAT
CGTTTAGAAAGACTGGATGATGCTATTTCATGTTCTCCGGAACCATGCAGTGGGCCCATCCACAGCTA
TGCCTGGTGGTTCATGGGGACATGCATGGAATCATTTGGACCTTCTCATAATGGAGCCATGGGTGGTCT
GGGCTCAGGGTATGGAACCGGCCTTCTTTTACGCCAACAGACATTCATCATGGTGGGGACCCATCGT
GAAGATGGCGTGGCCCTGAGAGGCAGCCATTCTCTTCTGCCAAACAGGTTCCGGTTCCACAGCTTC
CTGTCCAGTCTGCGACTTCCCCTGACCTGAACCCACCCACAGGACCCTTACAGAGGCATGCCACCAGG
ACTACAGGGGCAGAGTGTCTCCTCTGGCAGCTCTGAGATCAAATCCGATGACGAGGGTGATGAGAAC
CTGCAAGACACGAAATCTTCGGAGGACAAGAAATTAGATGACGACAAGAAGGATATCAAATCAATTA
CTAGGTCAAGATCTAGCAATAATGACGATGAGGACCTGACACCAGAGCAGAAGGCAGAGCGTGAGAA
GGAGCGGAGGATGGCCAAACATGCCCGAGAGCGTCTGCGGGTCCGTGACATCAACGAGGCTTTCAA
GAGCTCGGCGCATGGTGCAGCTCCACCTCAAGAGTGACAAGCCCCAGACCAAGCTCCTGATCCTCC
ACCAGGCGGTGGCCGTATCCTCAGTCTGGAGCAGCAAGTCCGAGAAAGGAATCTGAATCCGAAAGC
TGCGTGTCTGAAAAGAAGGGAGGAAGAGAAGGTGTCCTCAGAGCCTCCCCCTCTCTCCTTGCCGGC
CCACACCCTGGAATGGGAGACGCATCGAATCACATGGGACAGATGTAAActcgaggata

```

特征:

名称	位置	长度	注释
人工启动子	1-109	109	小写
microE5 box	9-14	6	
microE5 box	21-26	6	
microE5 box	33-38	6	
microE5 box	45-50	6	
microE5 box	57-62	6	
microE5 box	69-74	6	
minP微型启动子	79-109	31	下划线
TCF4编码序列	110-2125	2016	大写

图4E

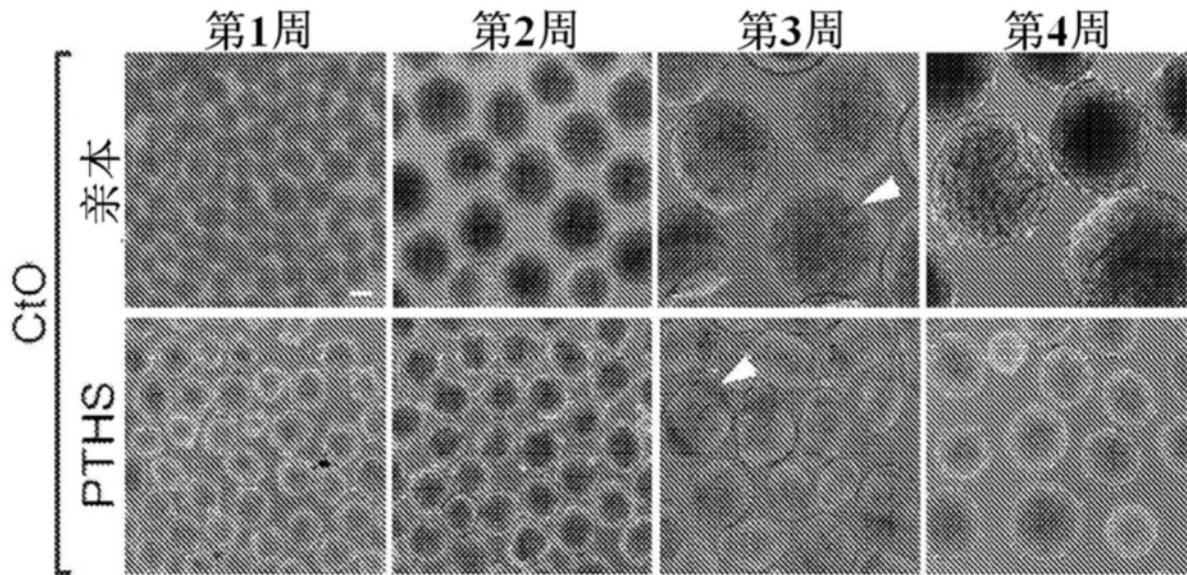


图5A

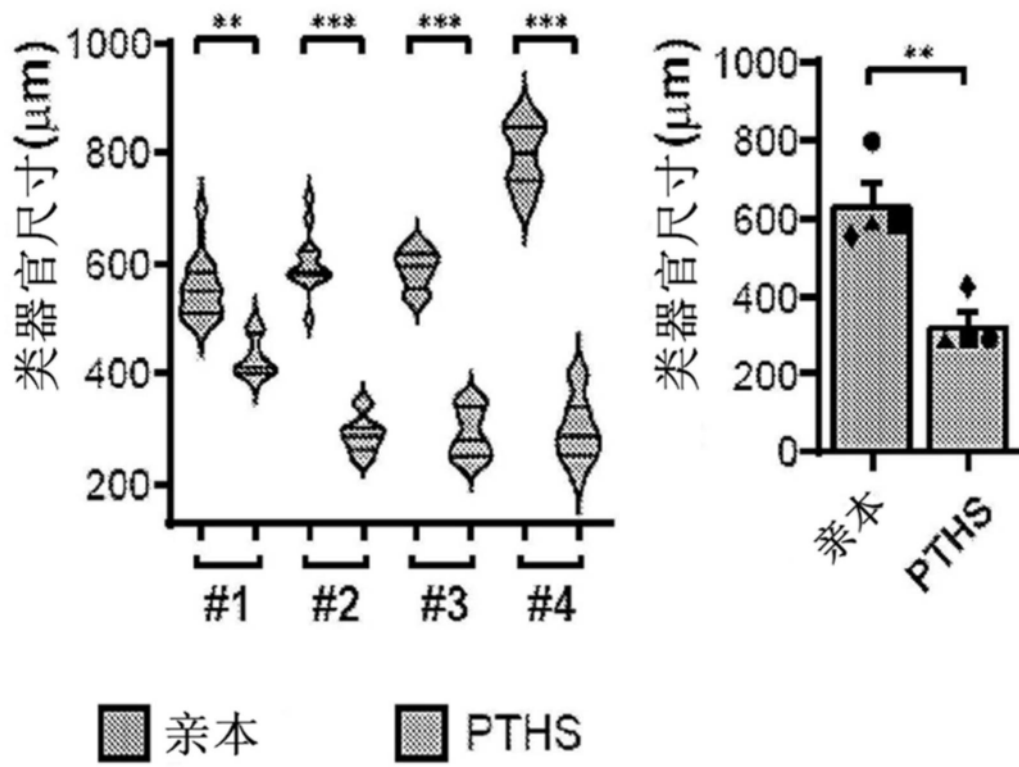


图5B

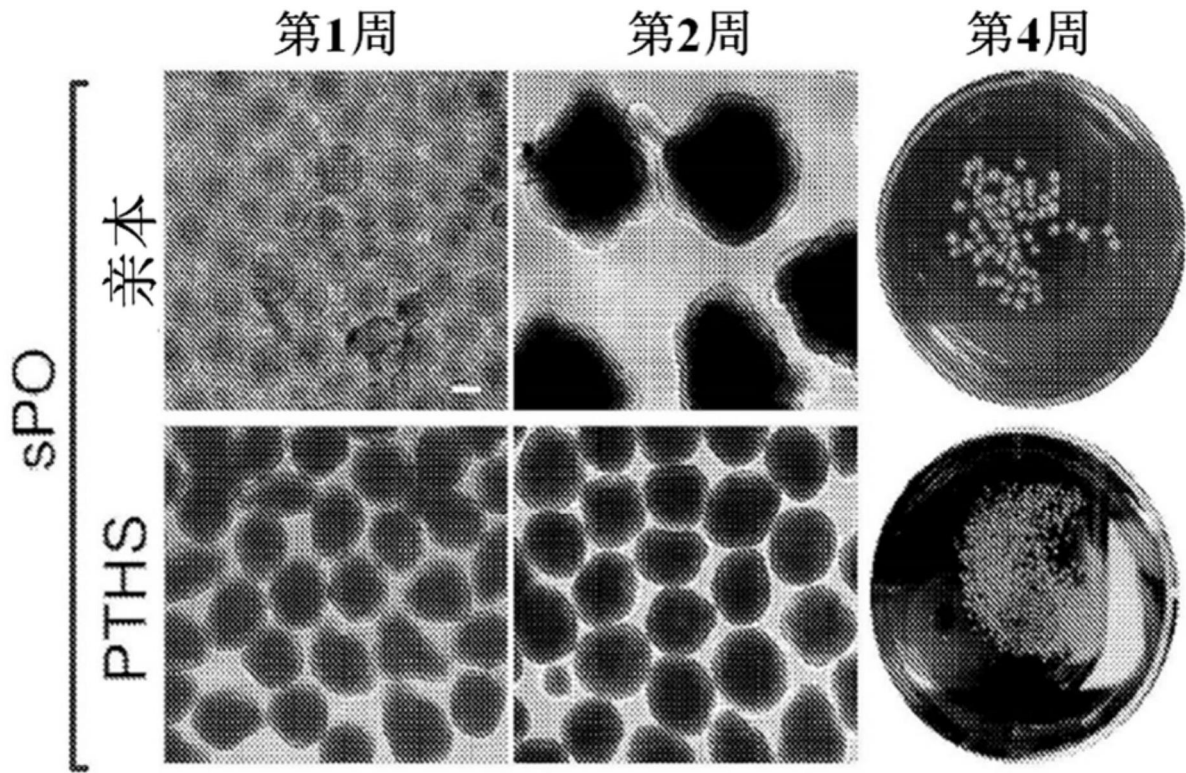


图5C

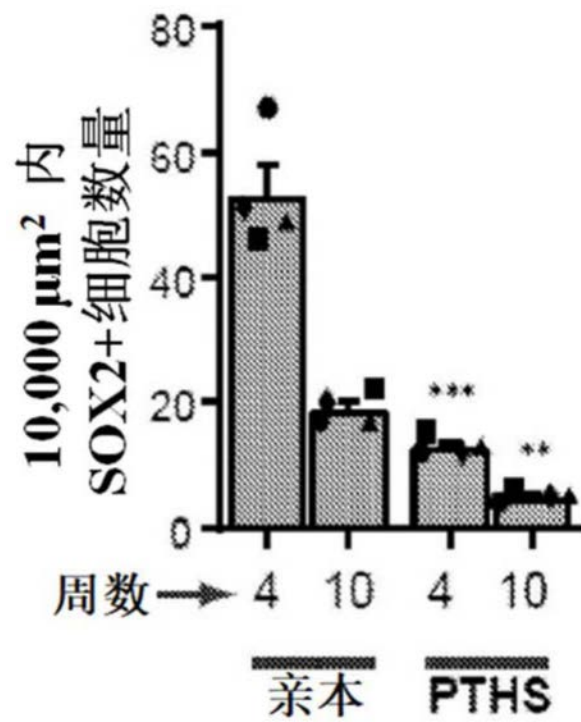


图5D

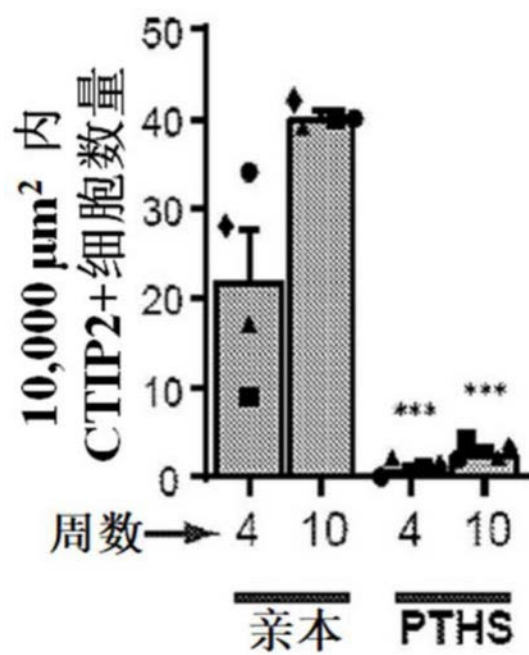


图5E

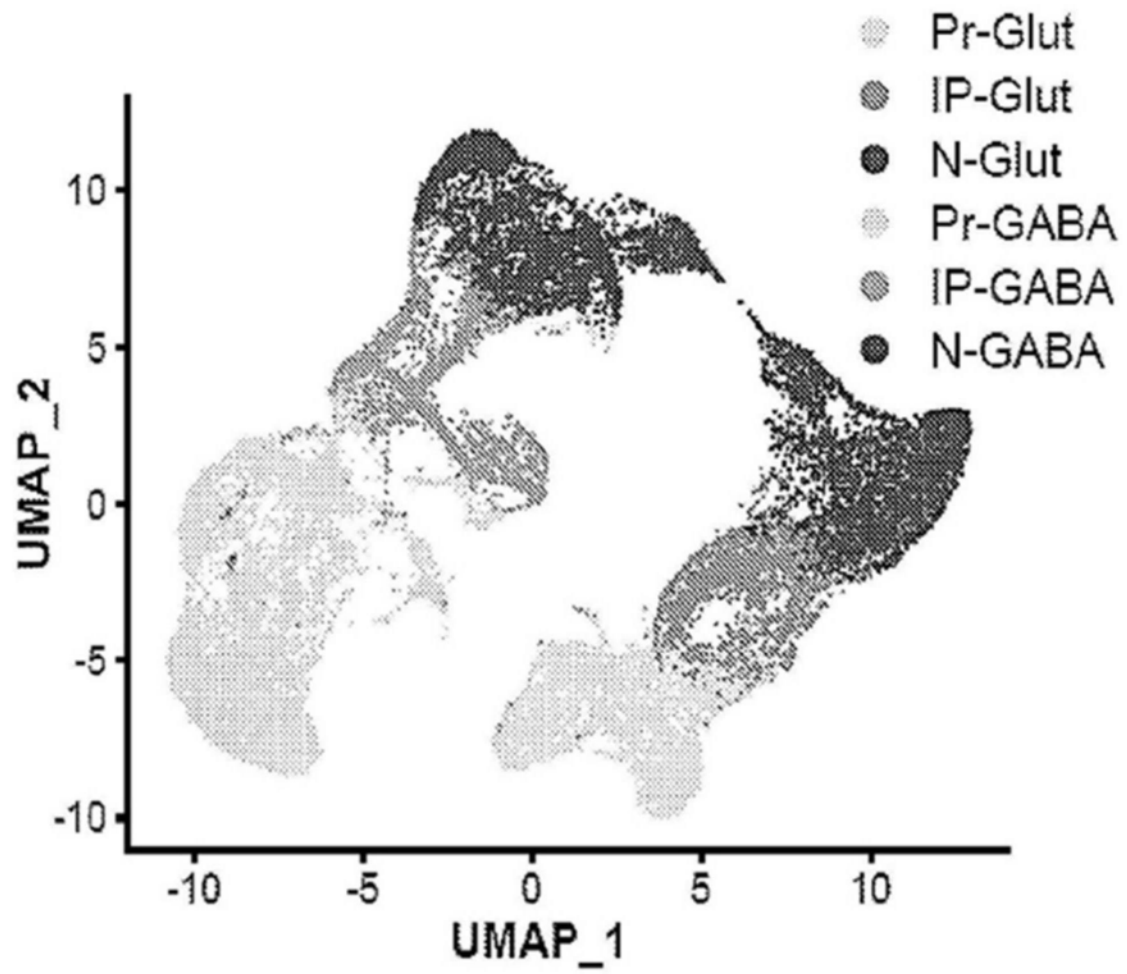


图6A

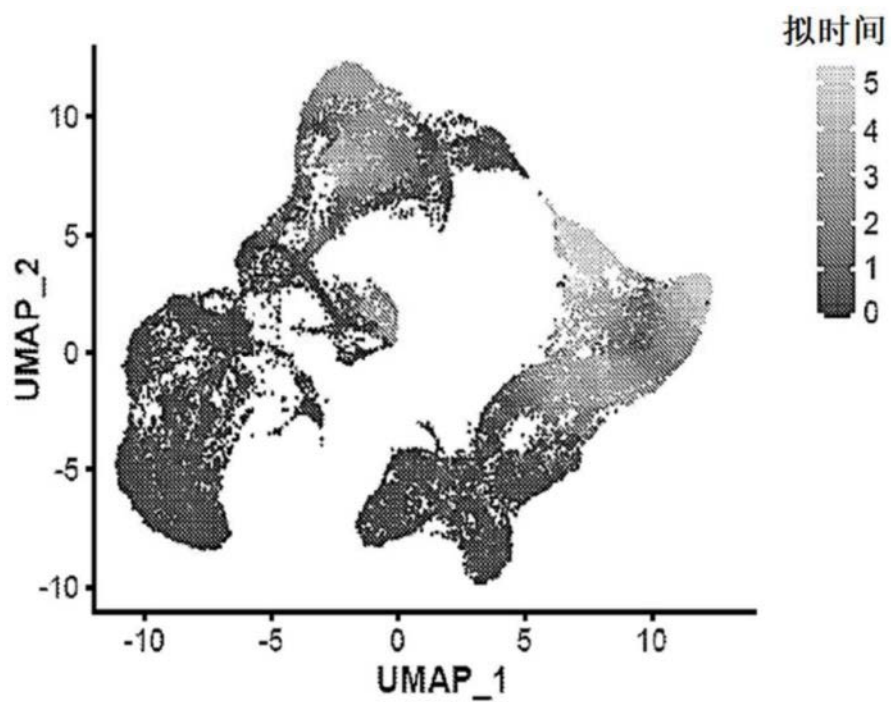


图6B

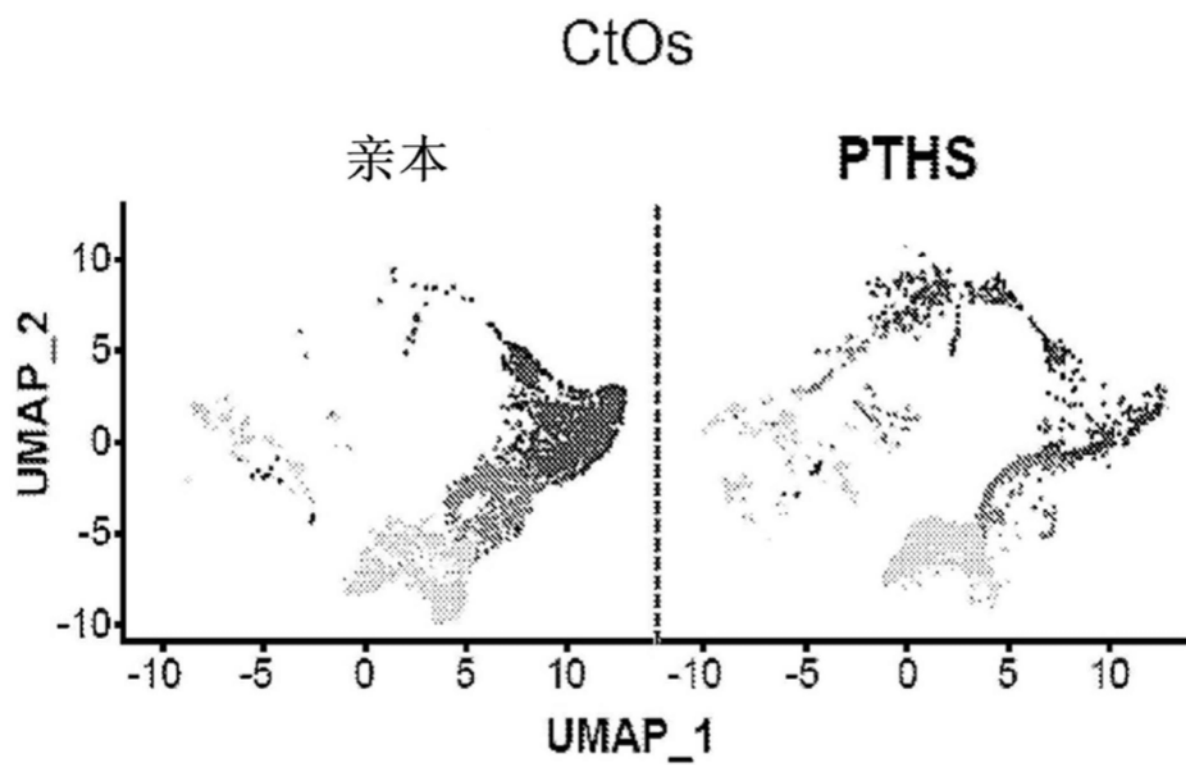


图6C

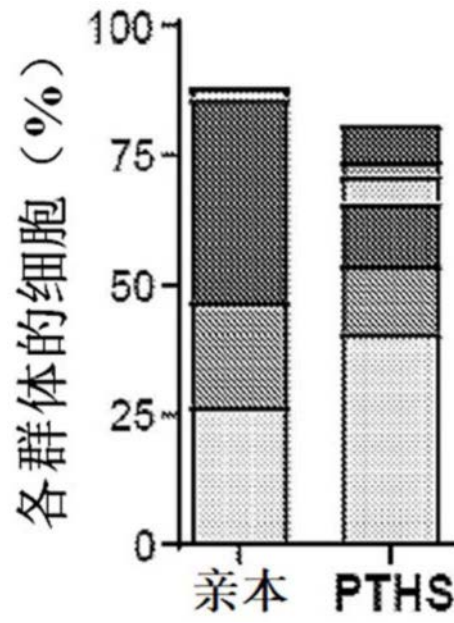


图6D

Pr-Glut中SOX2

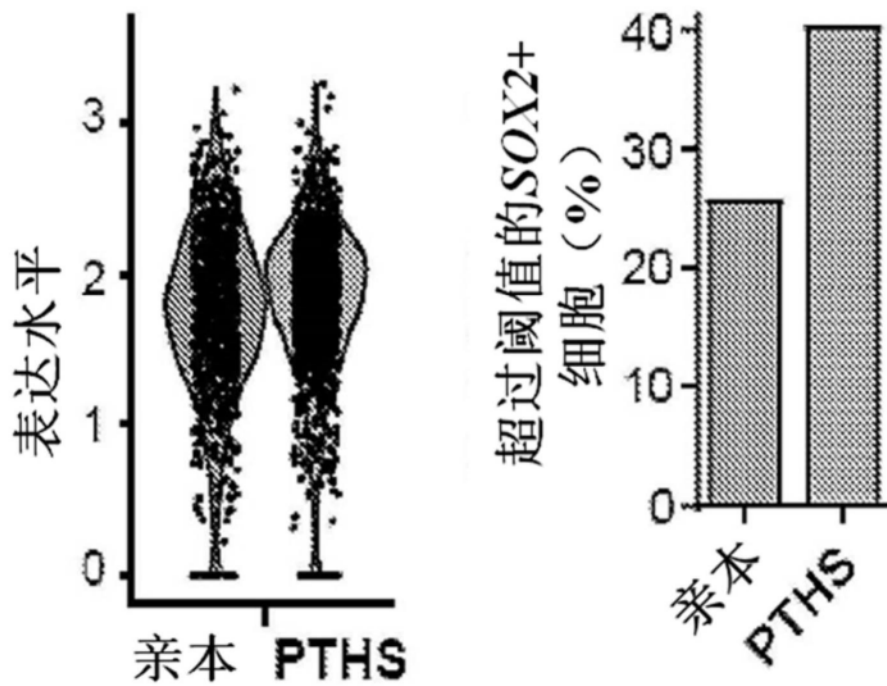


图6E

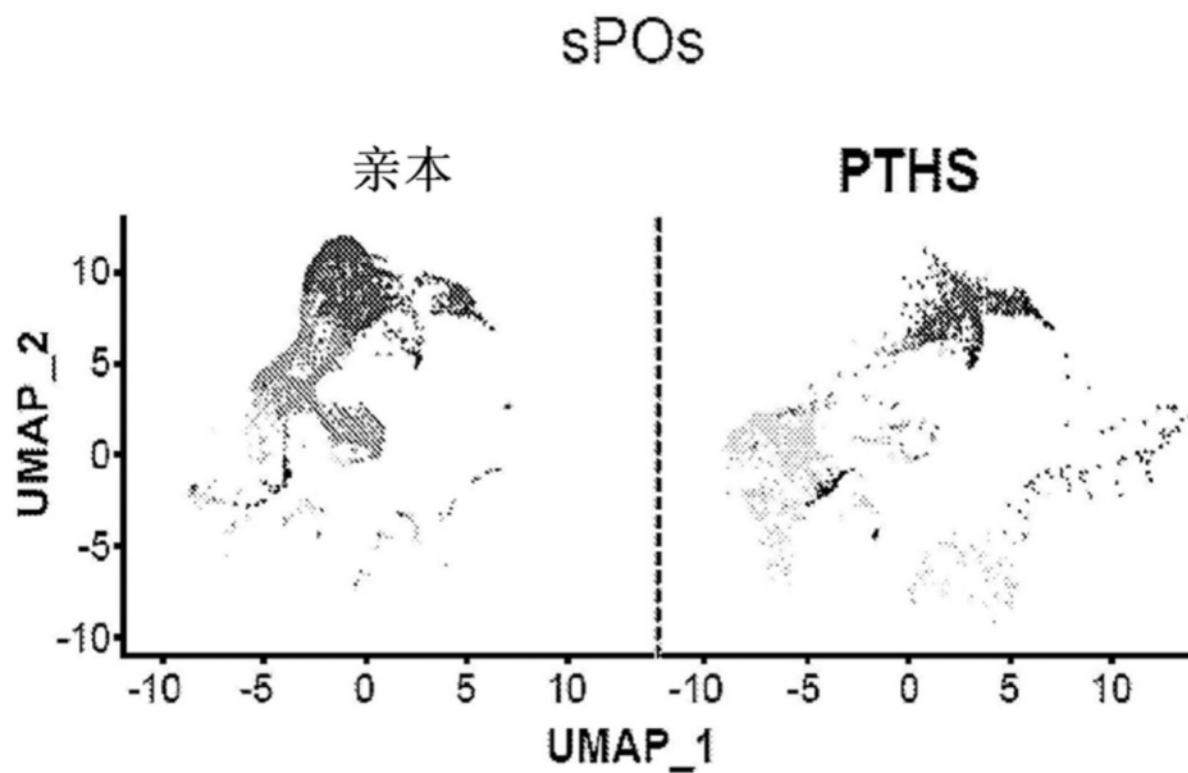


图6F

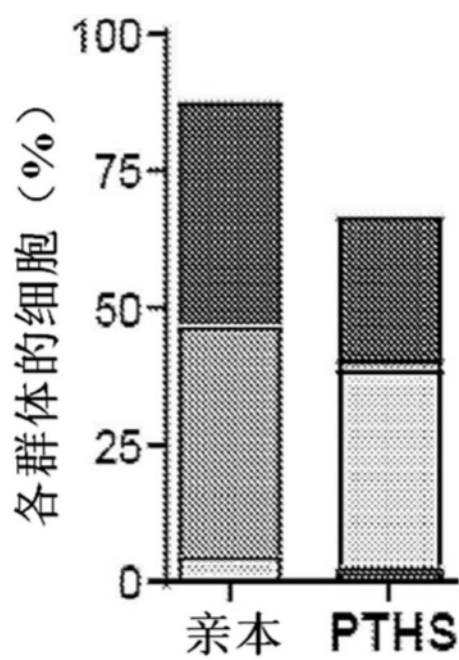


图6G

Pr-GABA中SOX2

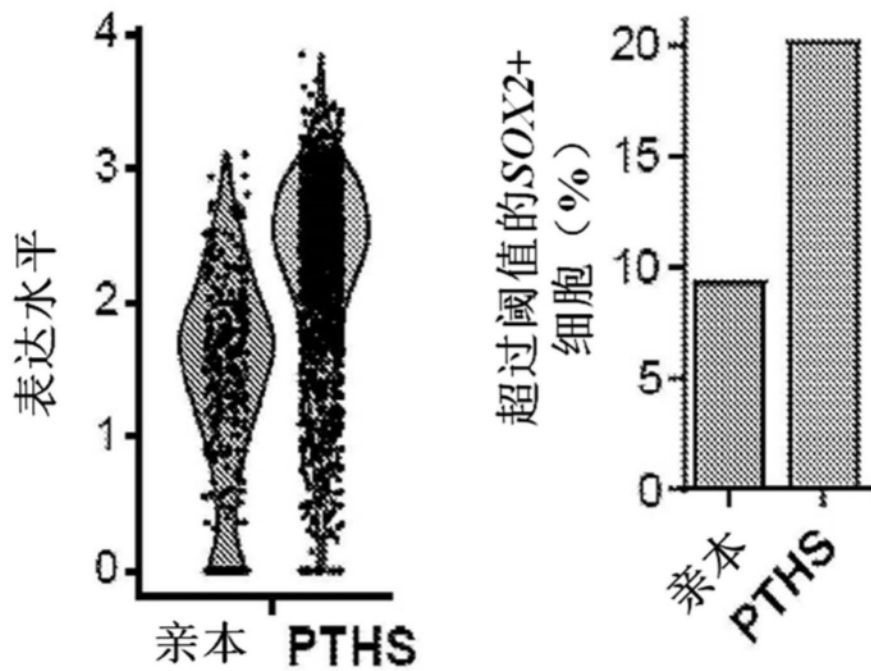
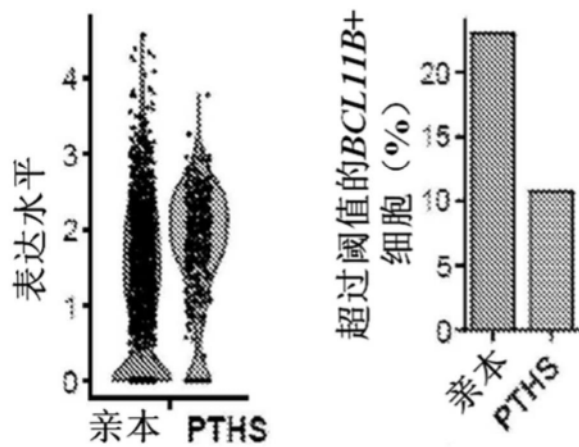


图6H

N-Glut (CtO中) 中 *BCL11B* (CTIP2)



N-Glut (CtO中) 中 *SATB2*

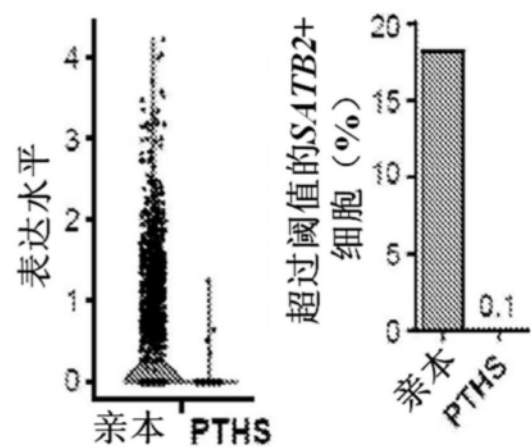


图6I

N-GABA (sPO中) 中GAD2

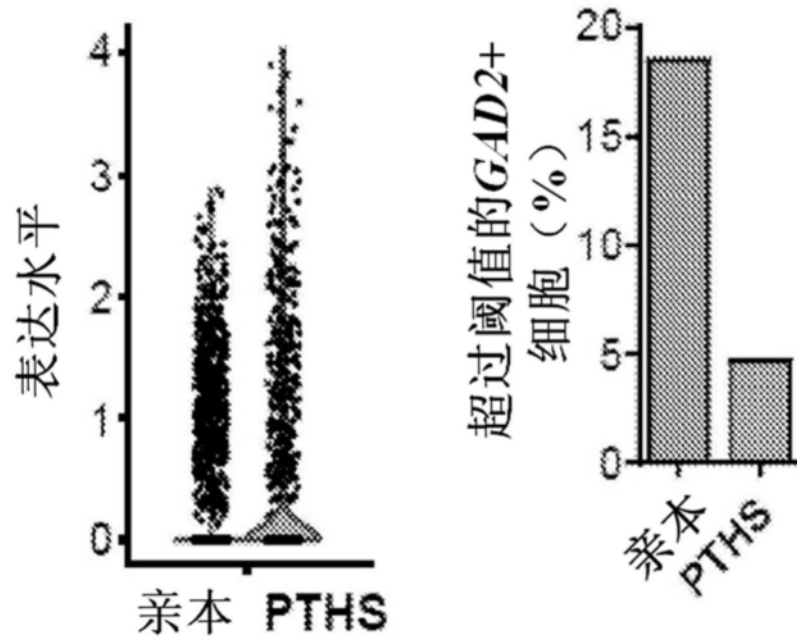


图6J

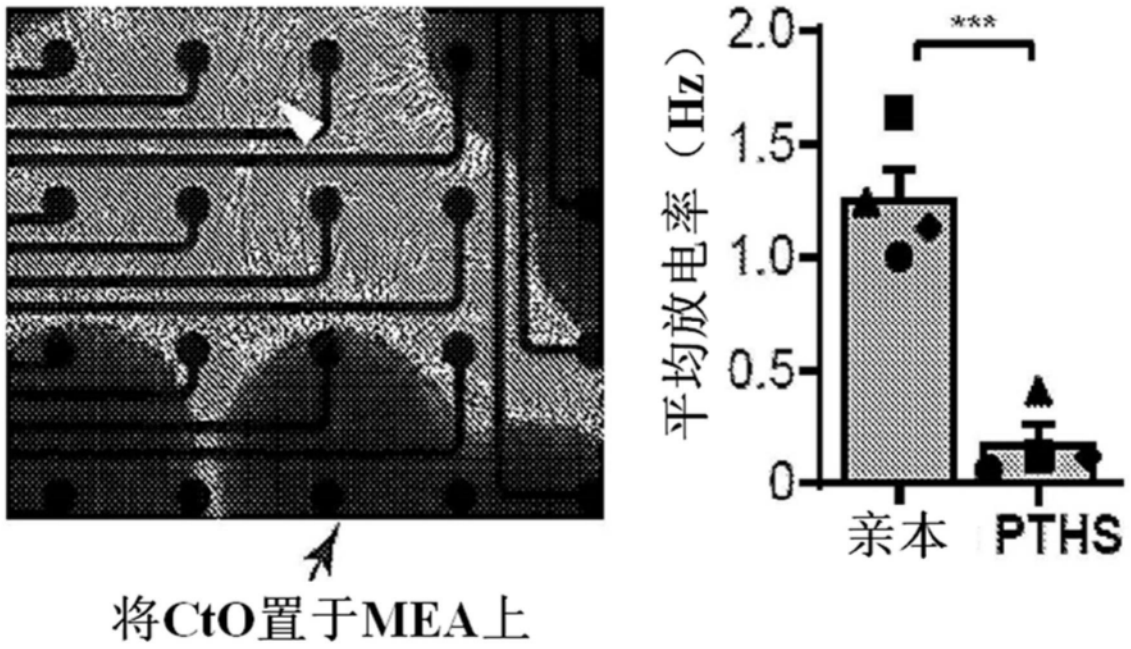


图7A

神经元（2D培养）

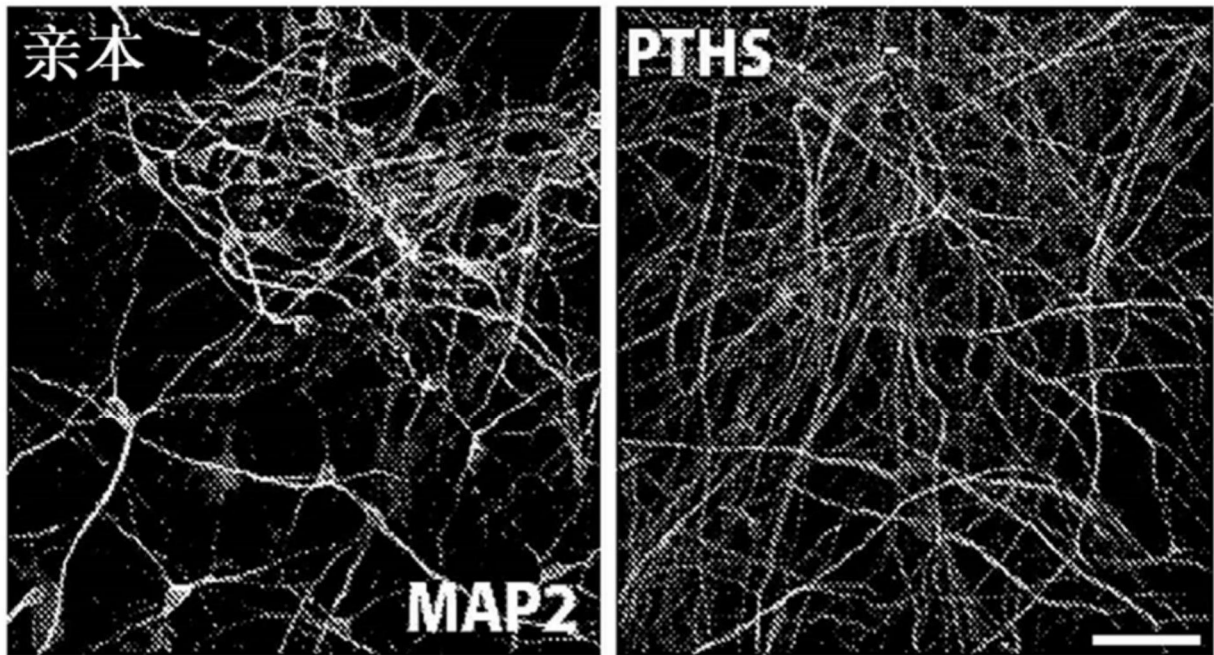


图7B

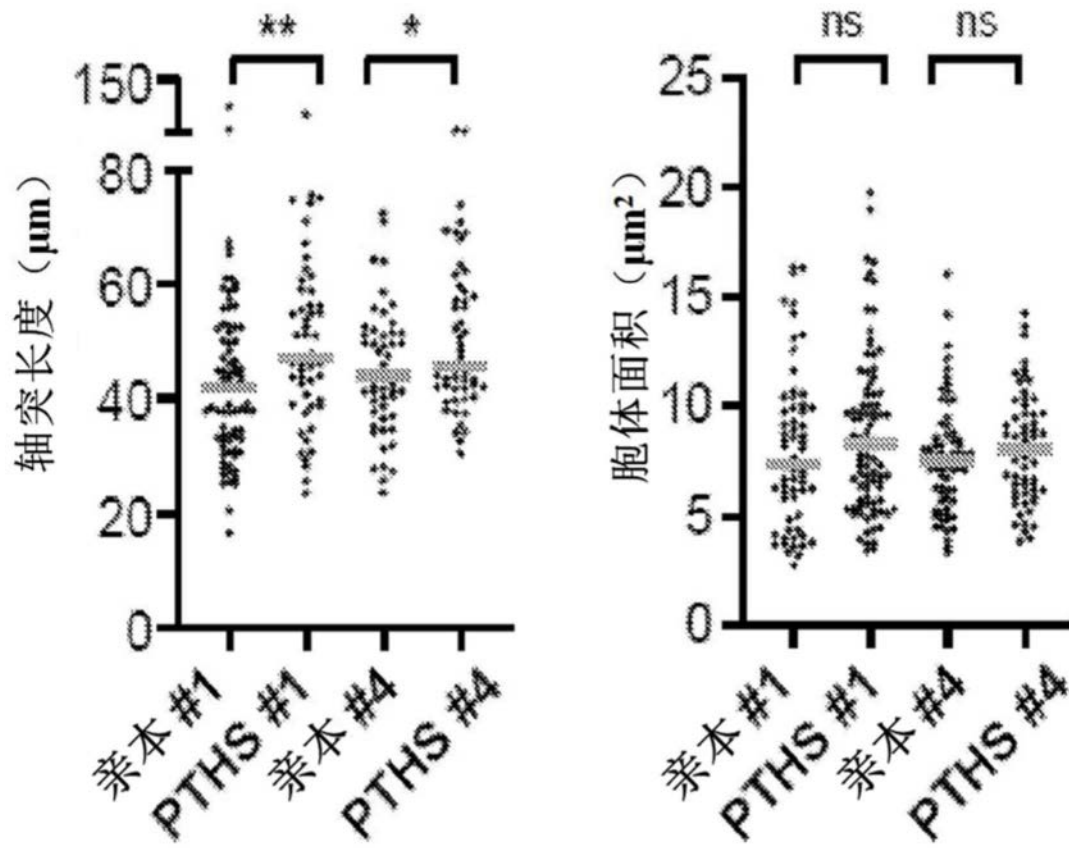


图7C

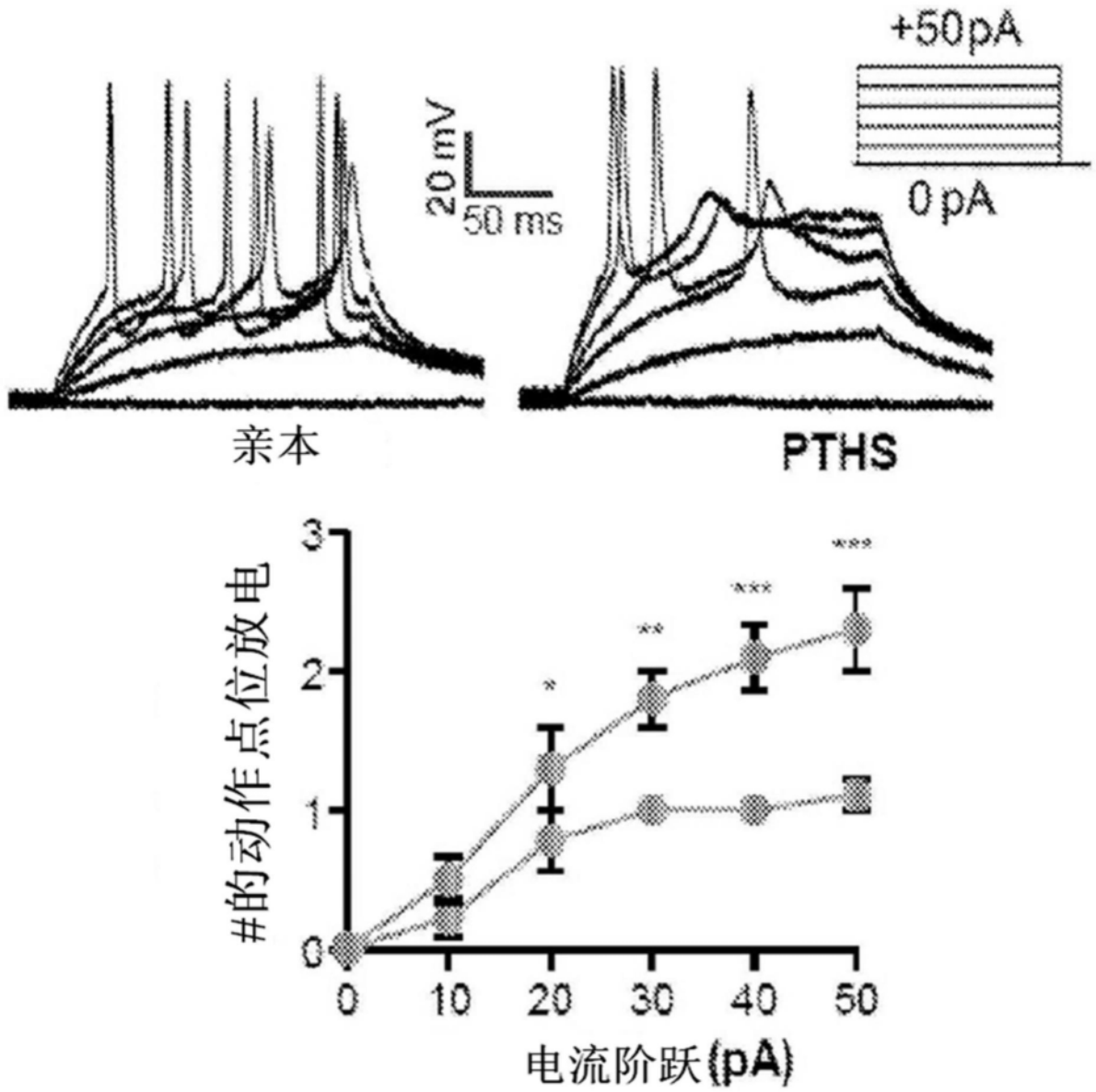


图7D

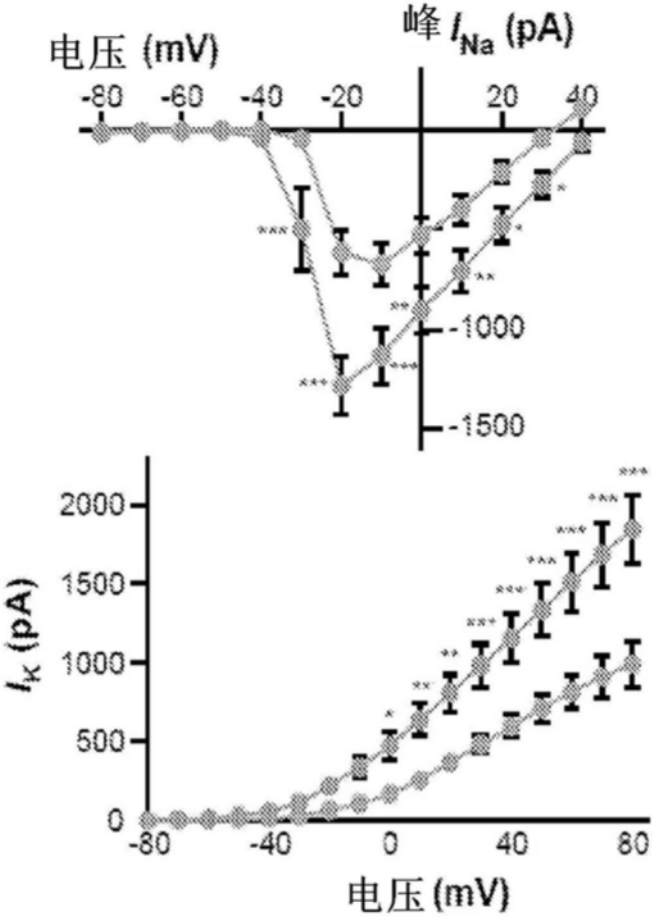


图7E

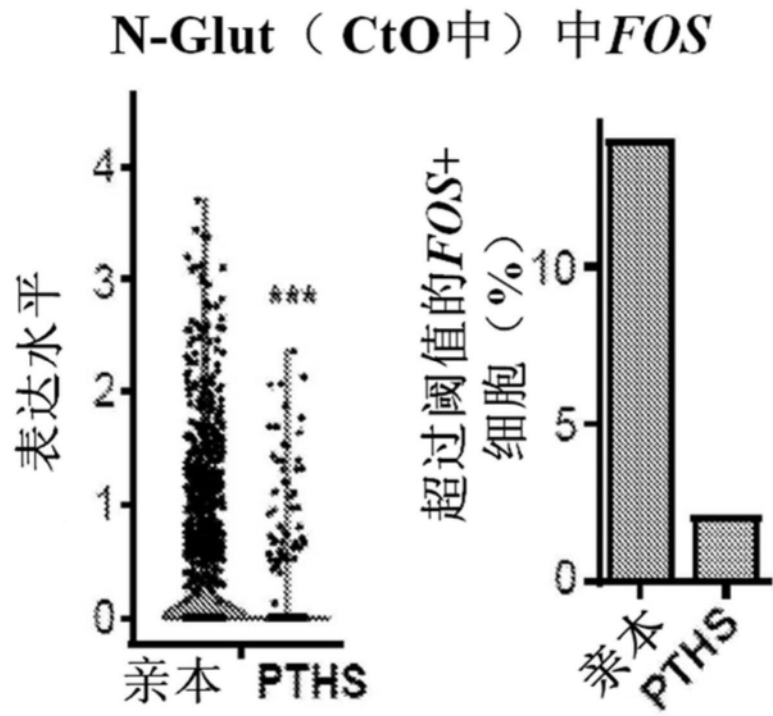


图7F

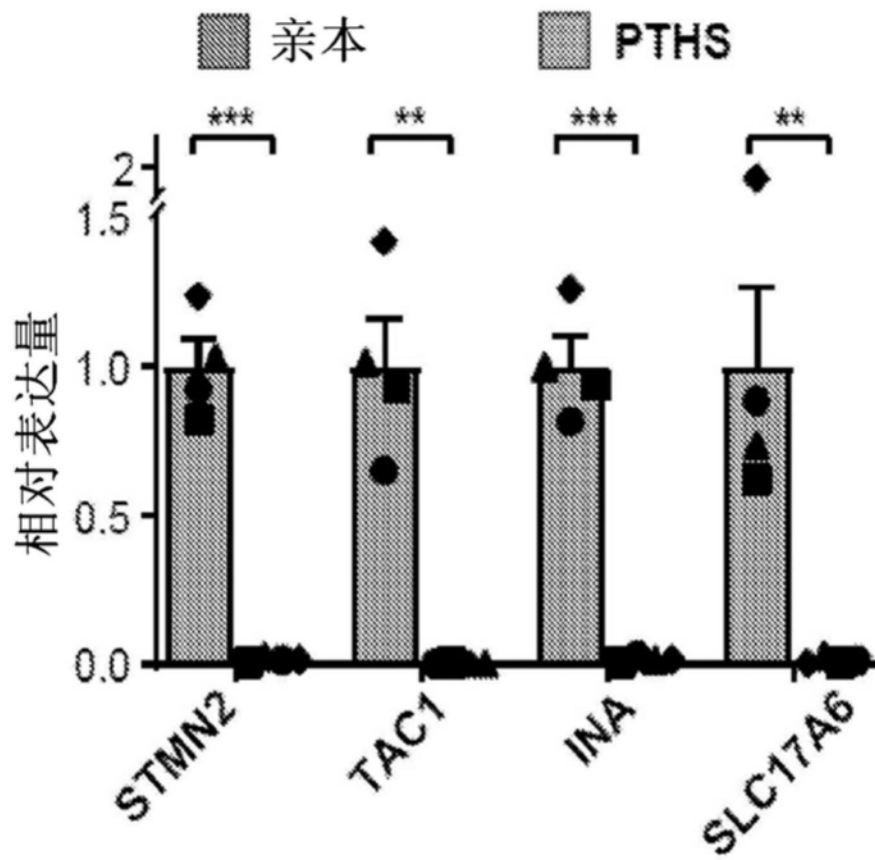


图7G

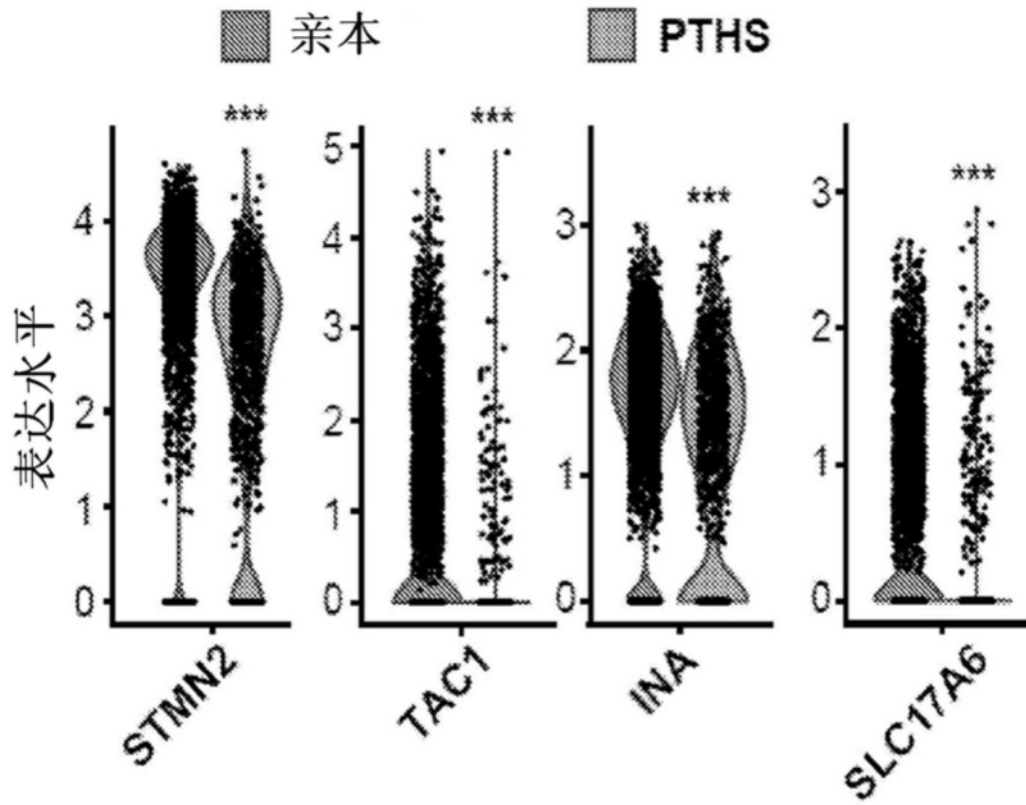


图7H

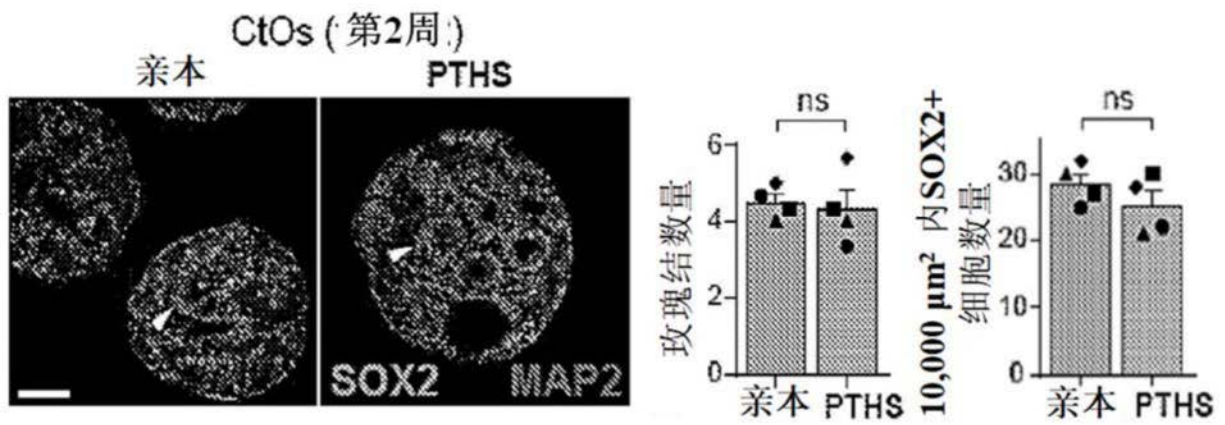


图8A

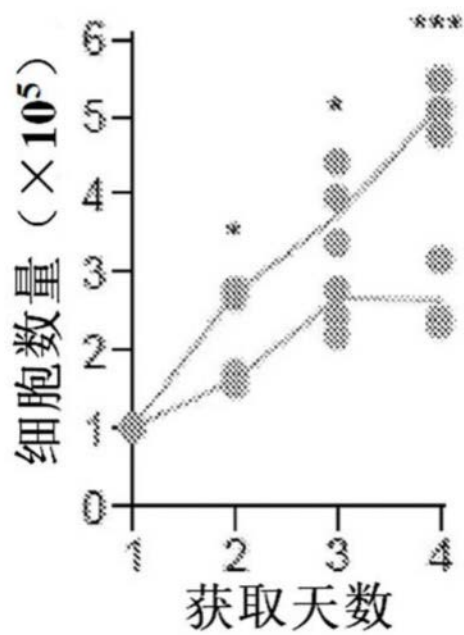


图8B

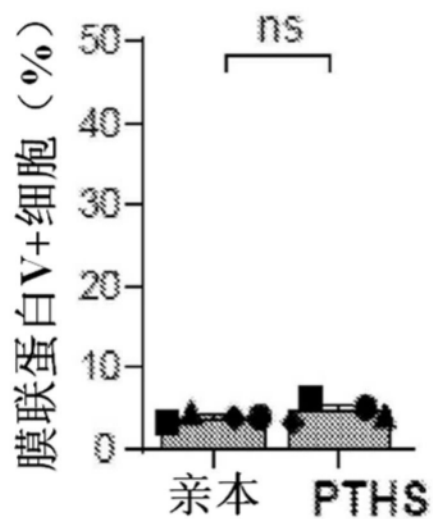


图8C

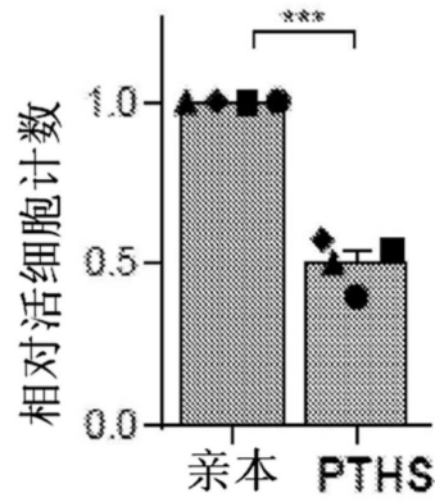


图8D

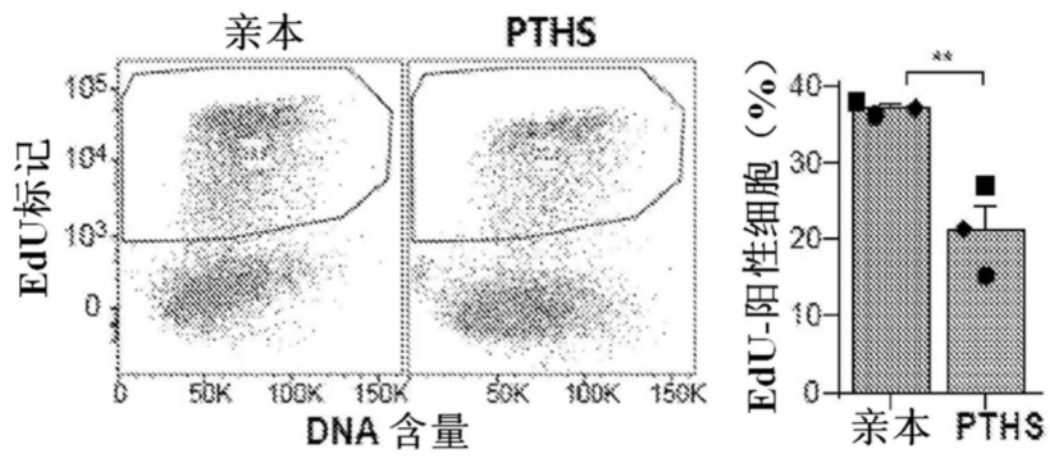


图8E

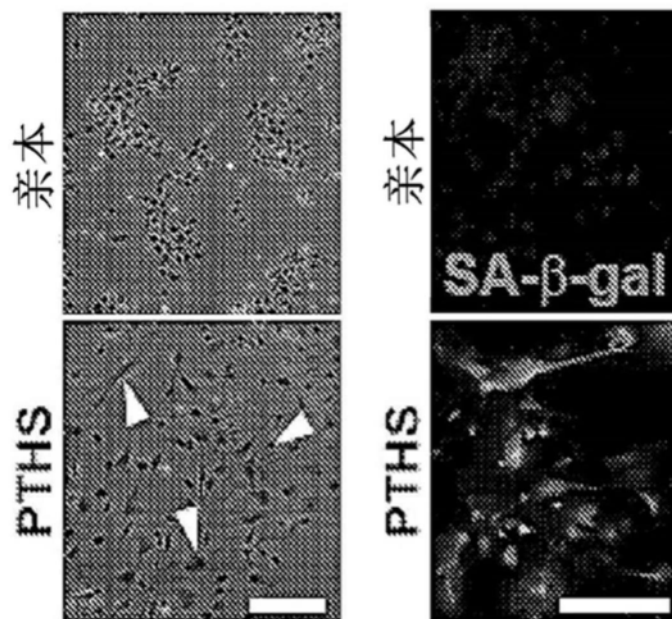


图8F

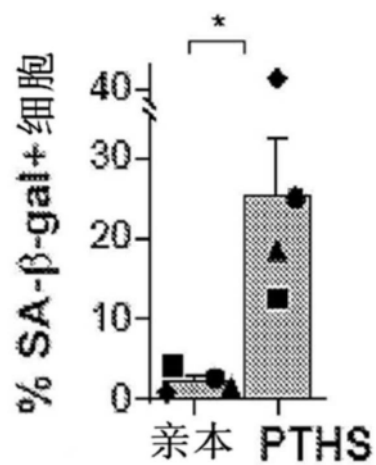


图8G

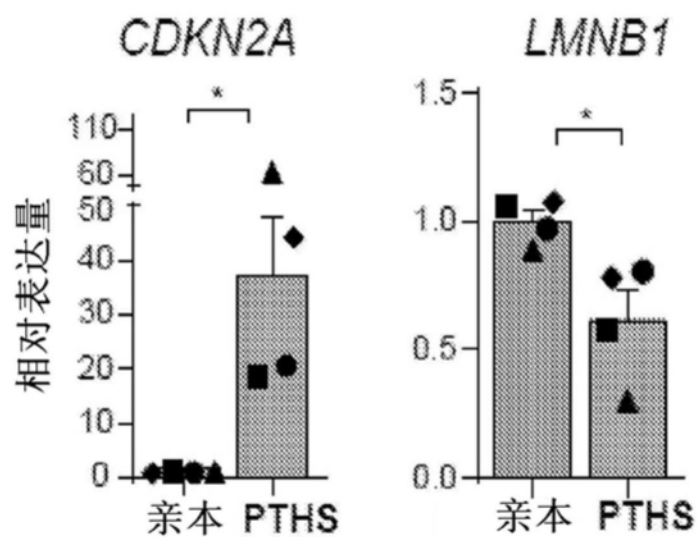


图8H

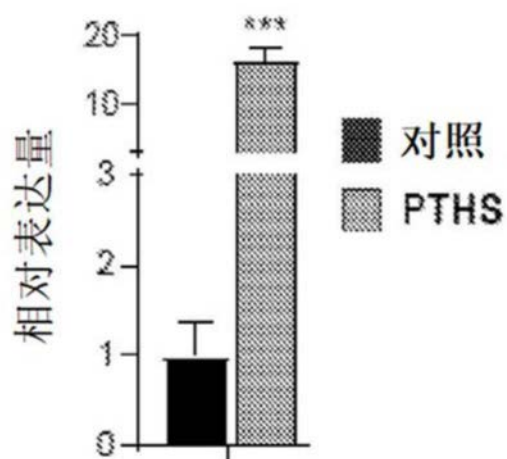
死后样本中*CDKN2A*

图8I

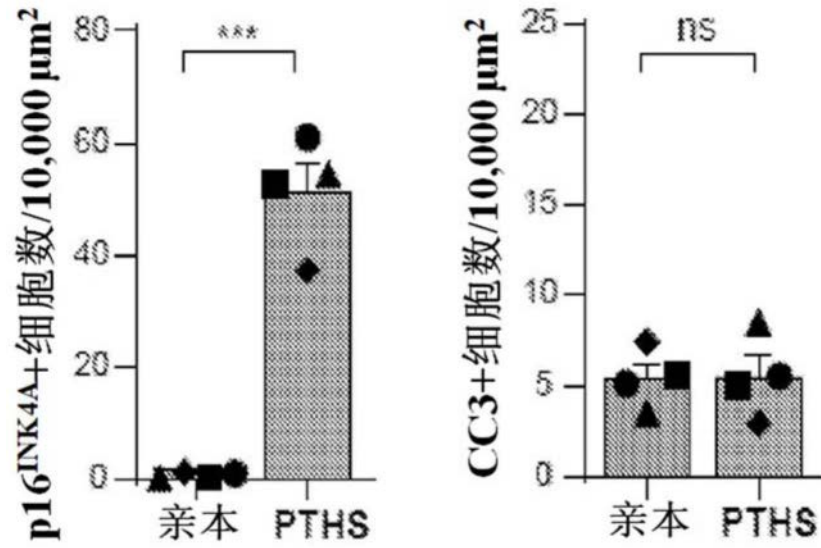


图8J

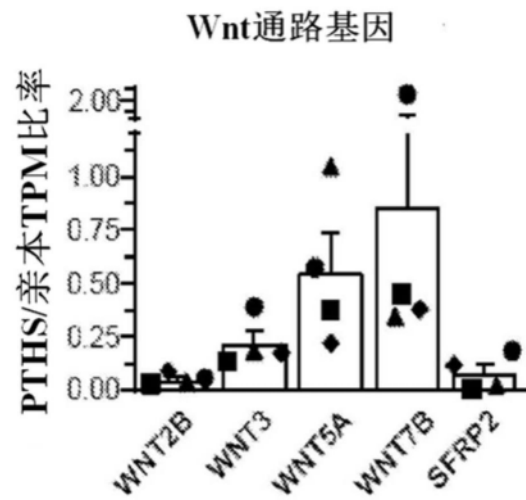


图9A

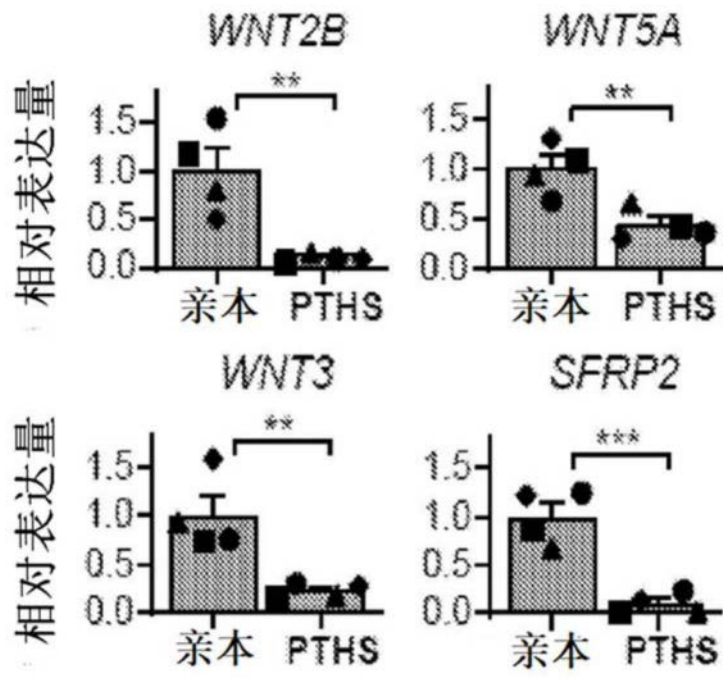


图9B

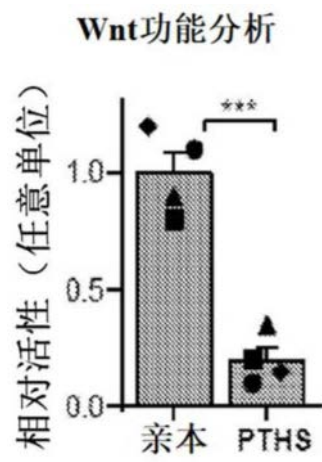


图9C

死后样本中WNT信号基因

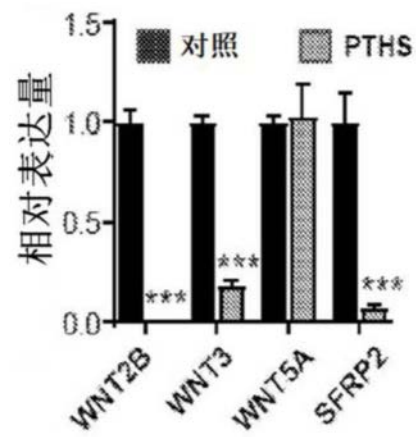


图9D

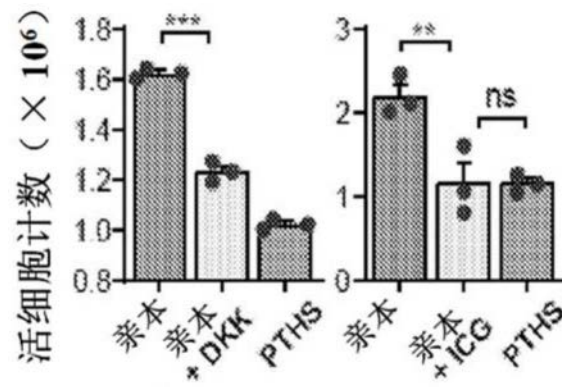


图9E

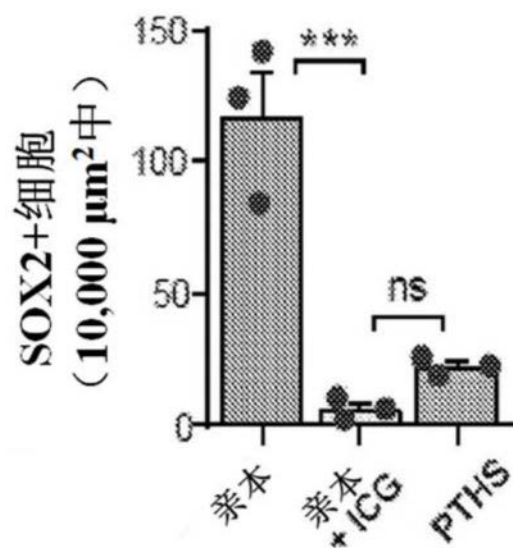


图9F

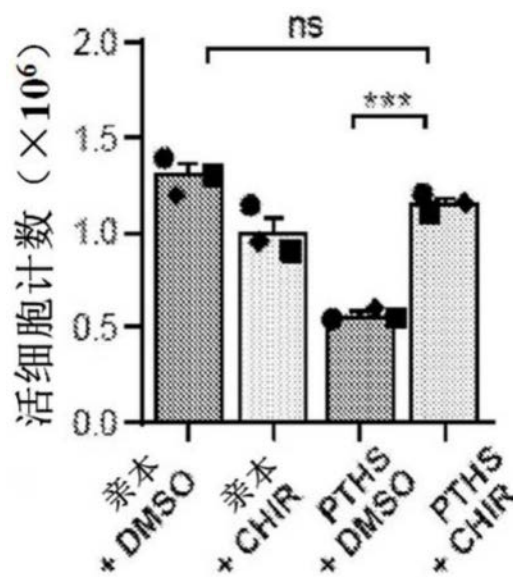


图9G

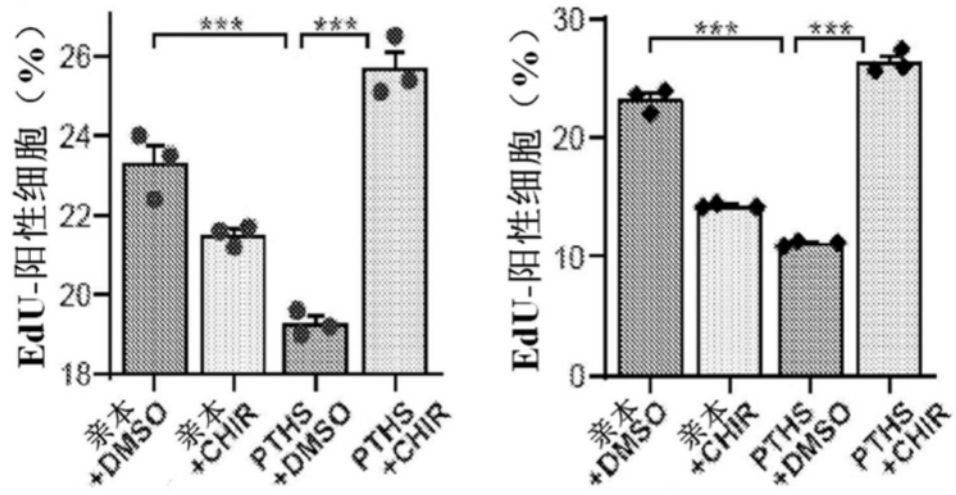


图9H

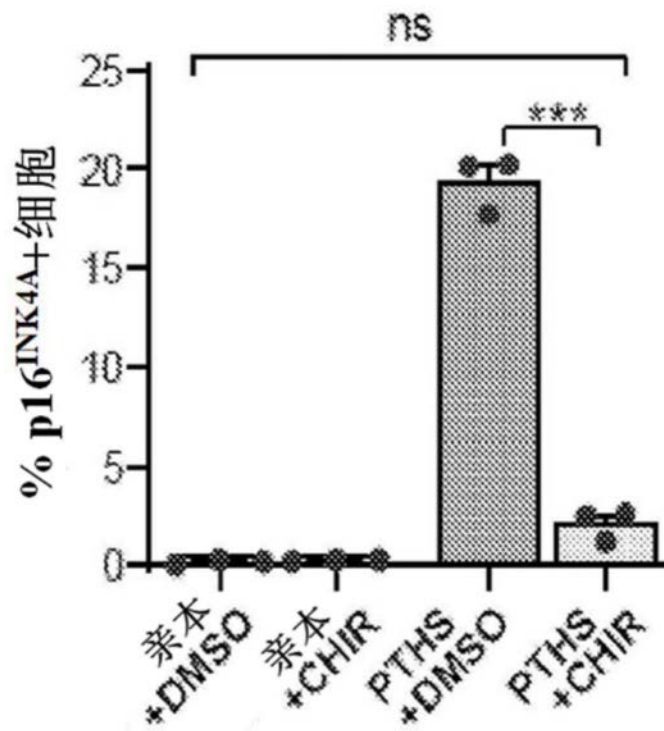


图9I

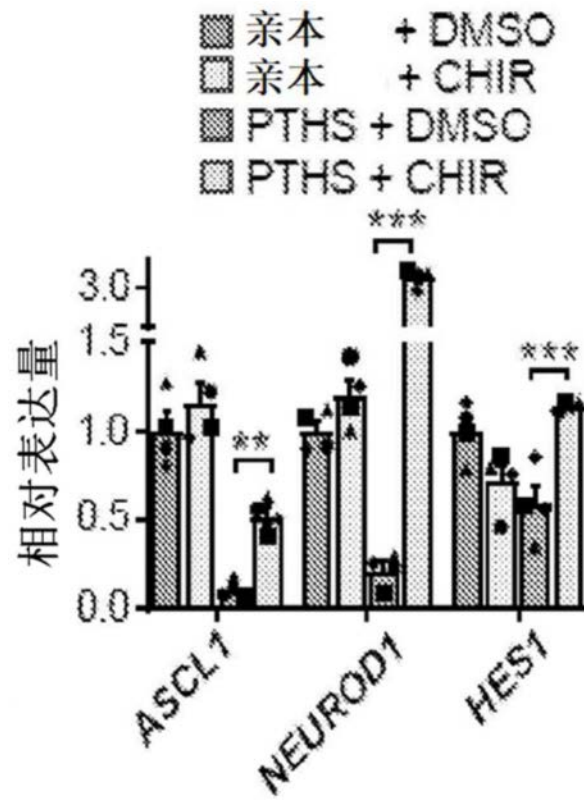


图9J

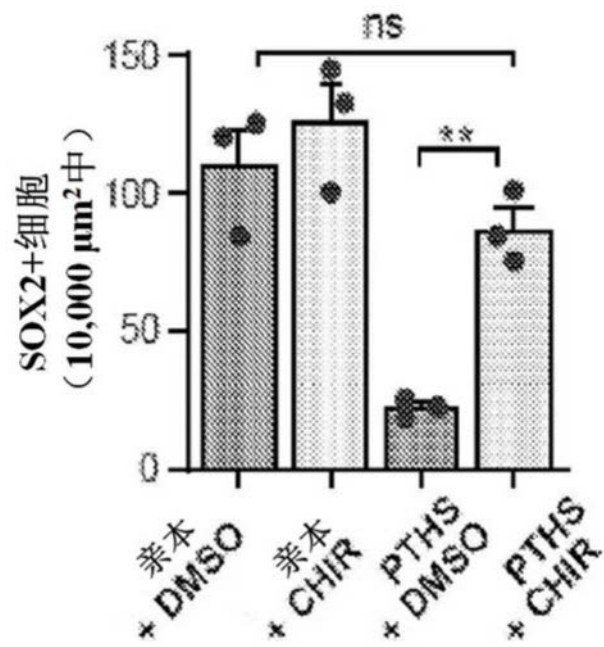


图9K

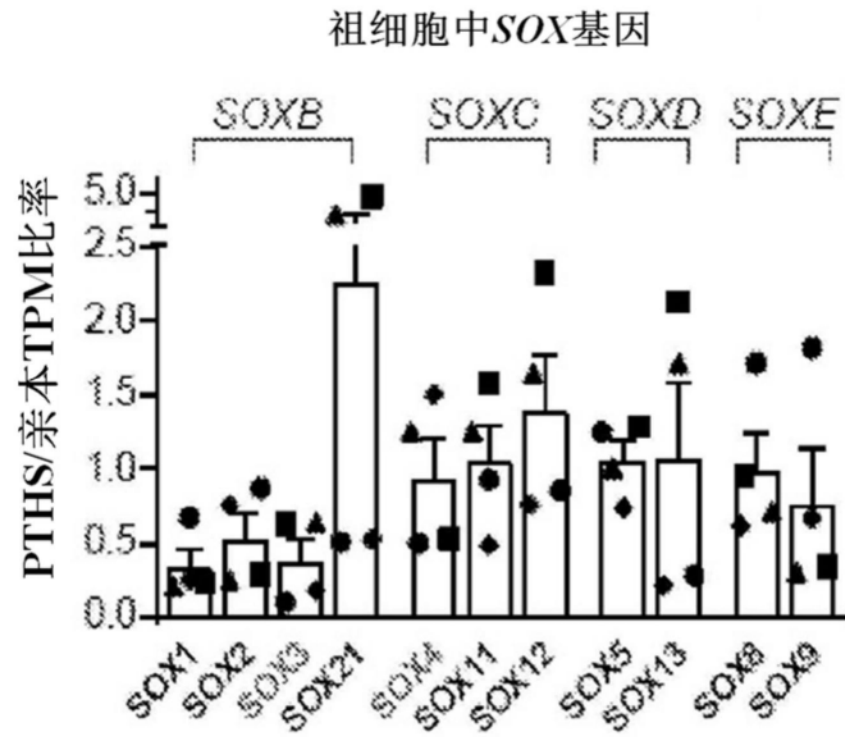


图10A

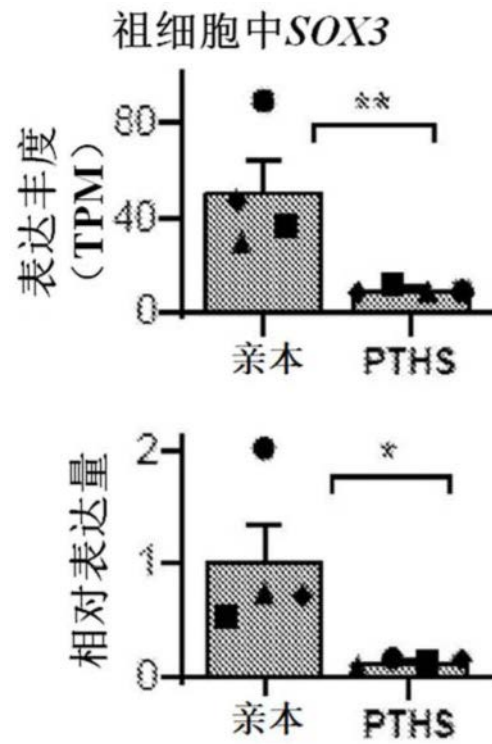


图10B

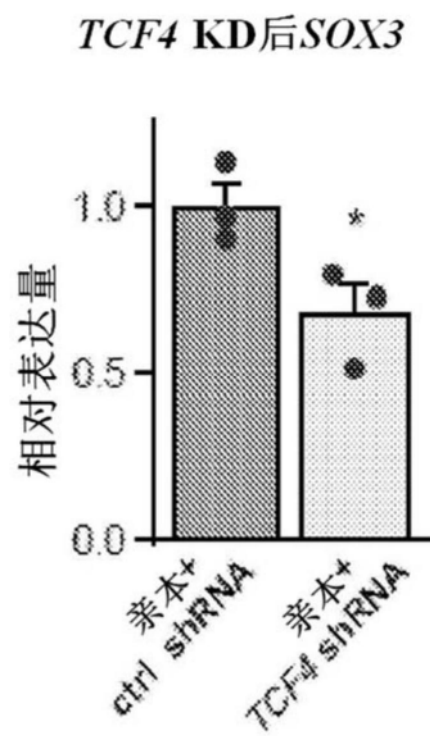


图10C

死后样本中SOX3

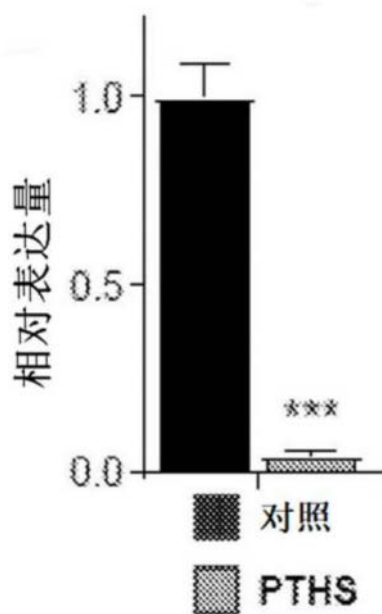


图10D

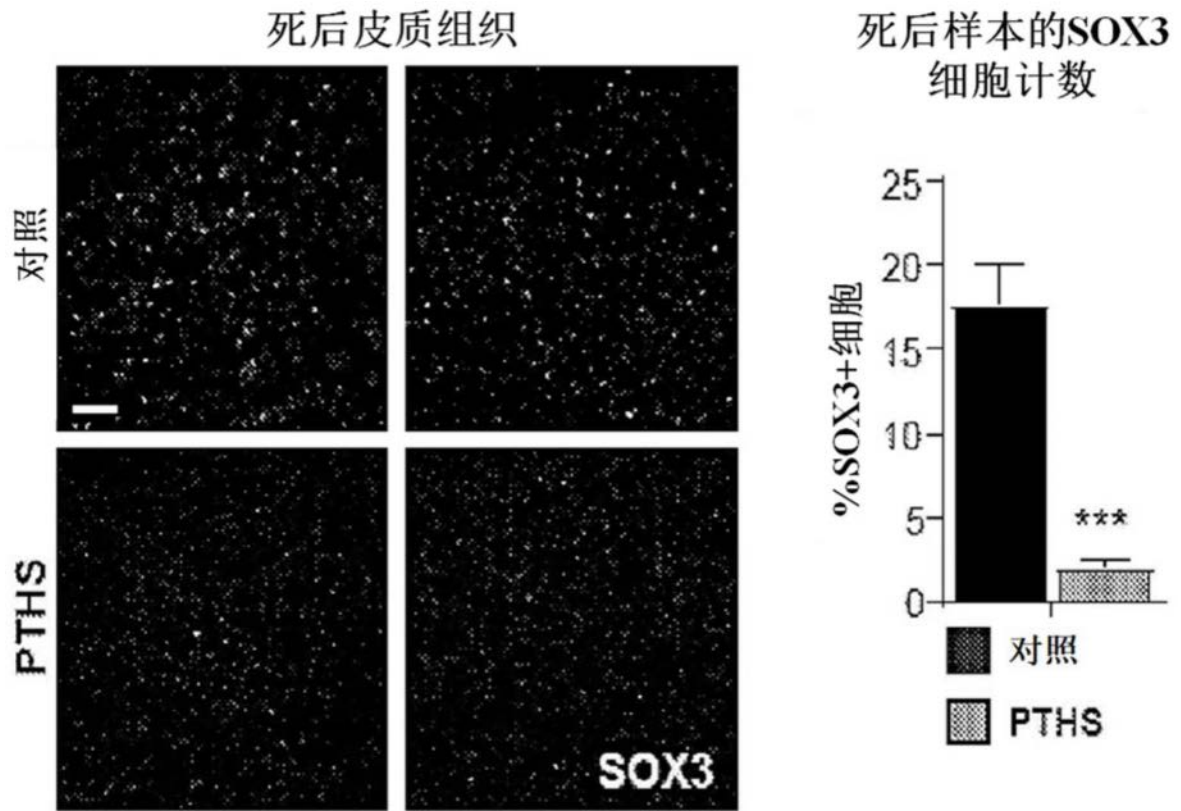


图10E

Wnt激动剂后SOX3

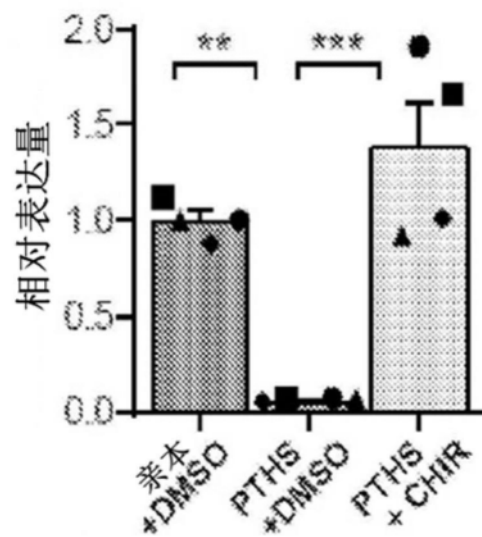


图10F

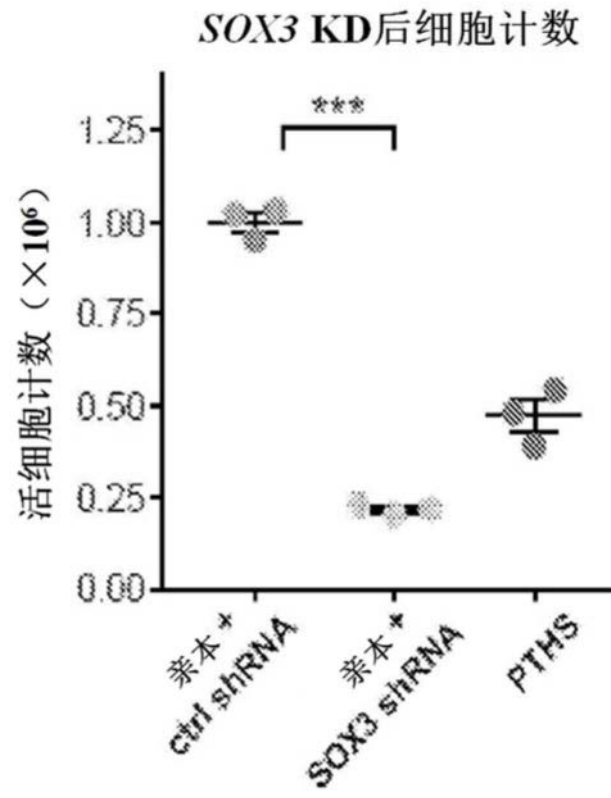


图10G

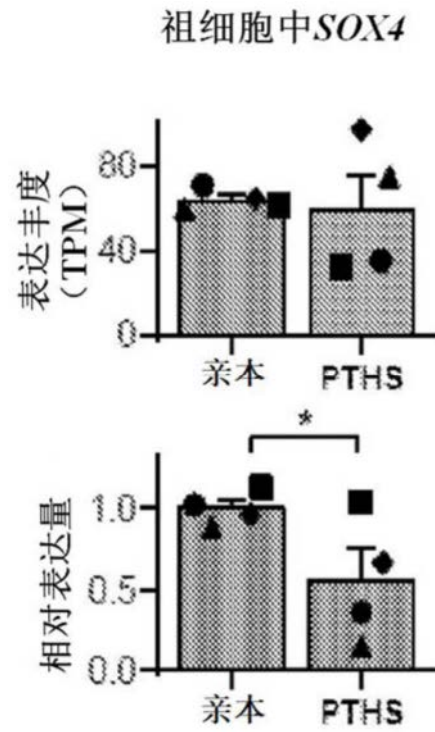


图10H

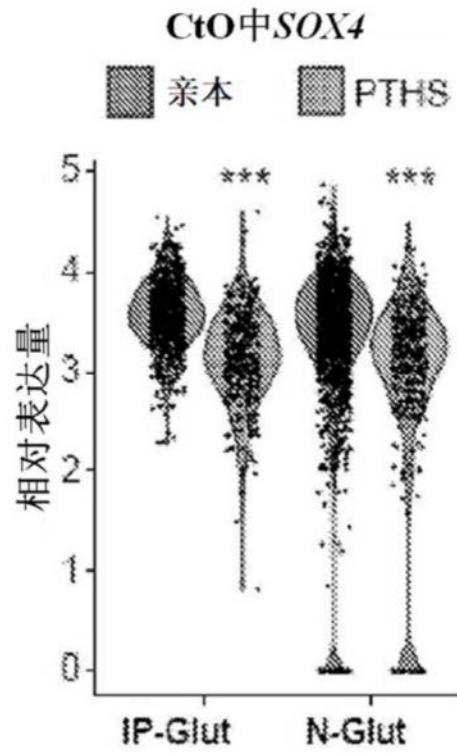


图10I

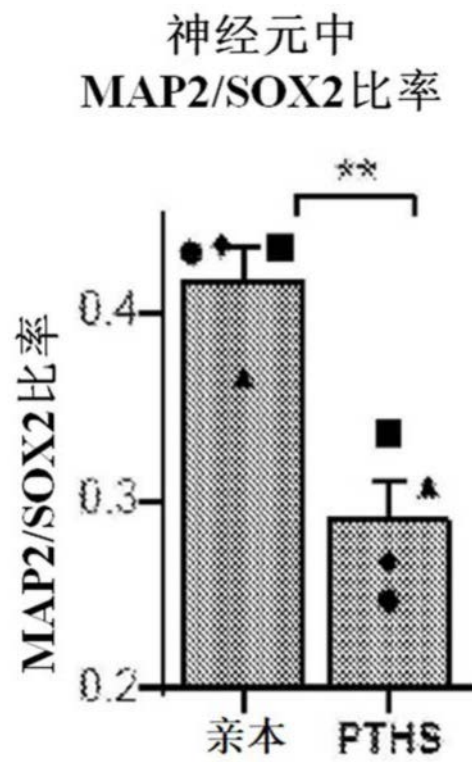


图10J

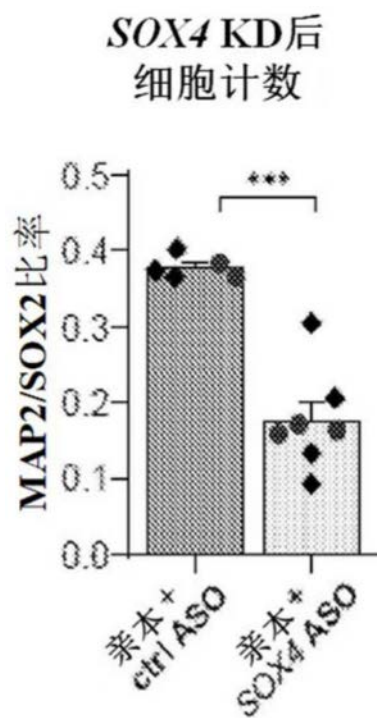


图10K

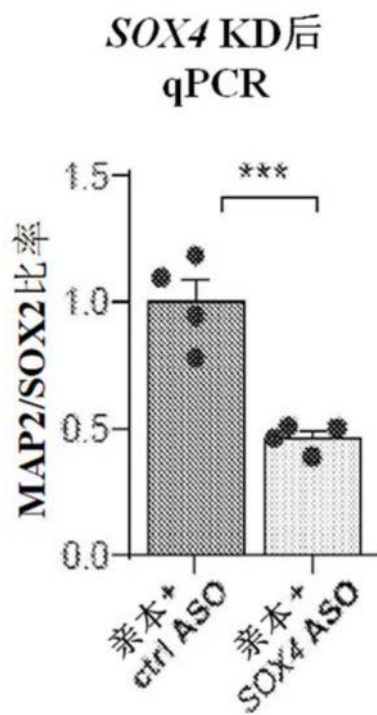


图10L

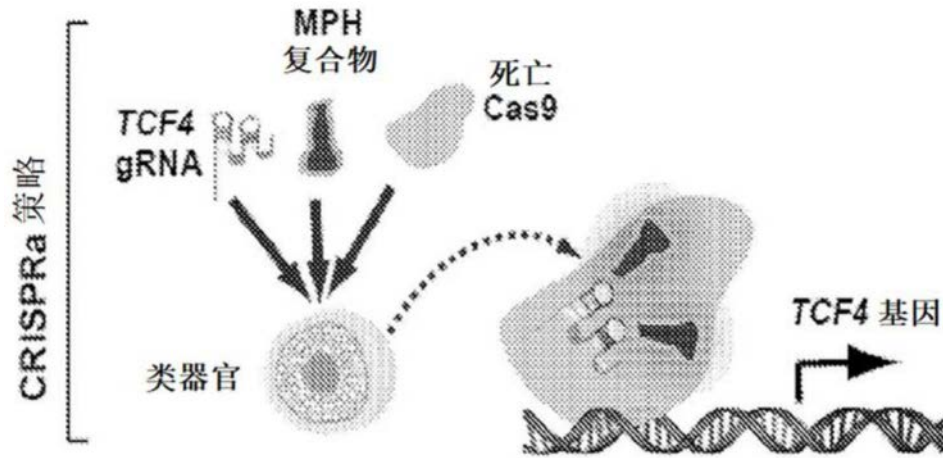


图11A

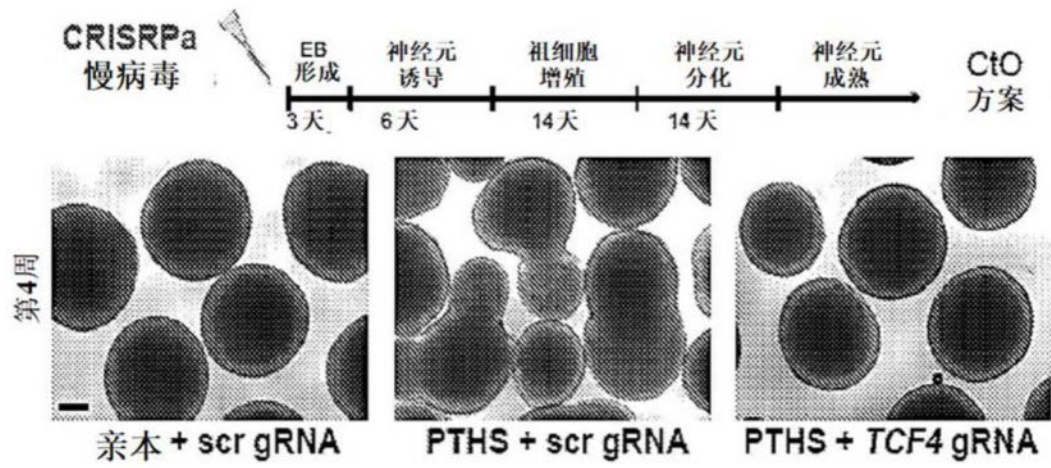


图11B

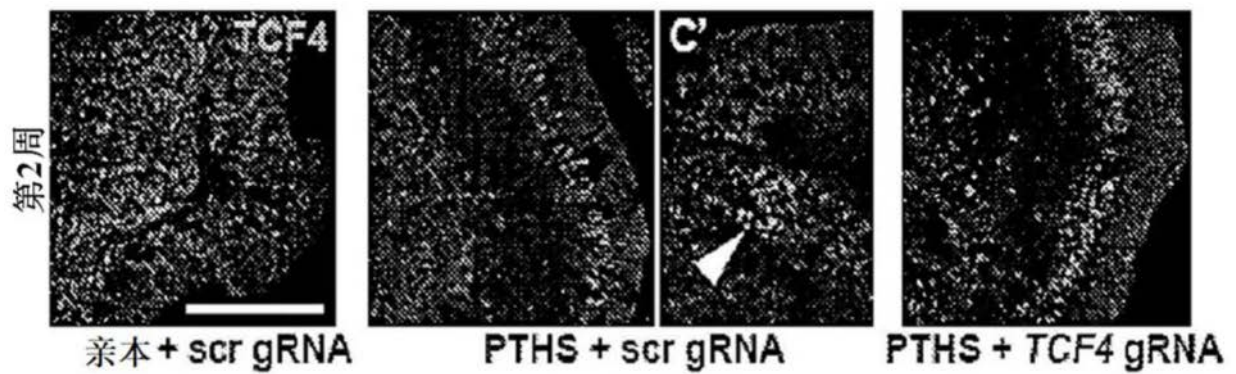


图11C

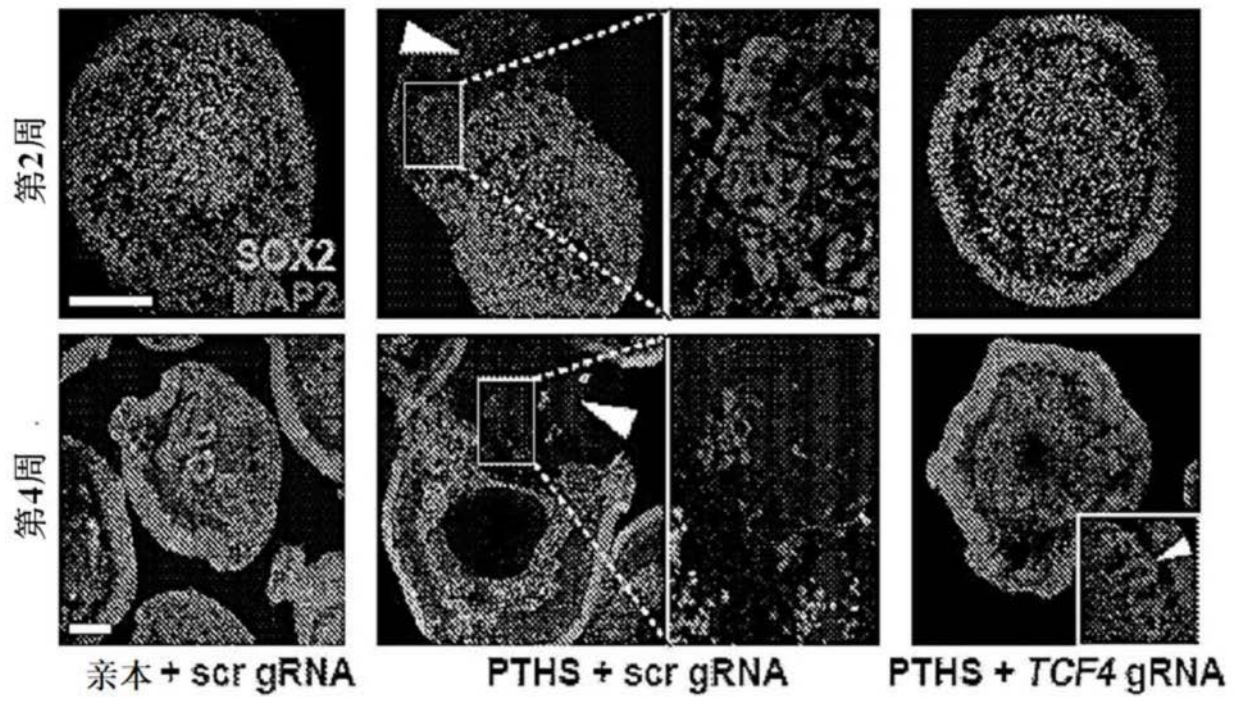


图11D

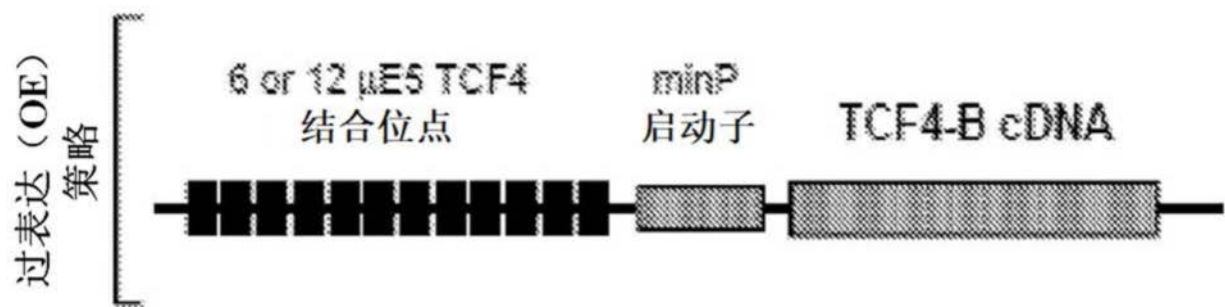


图11E

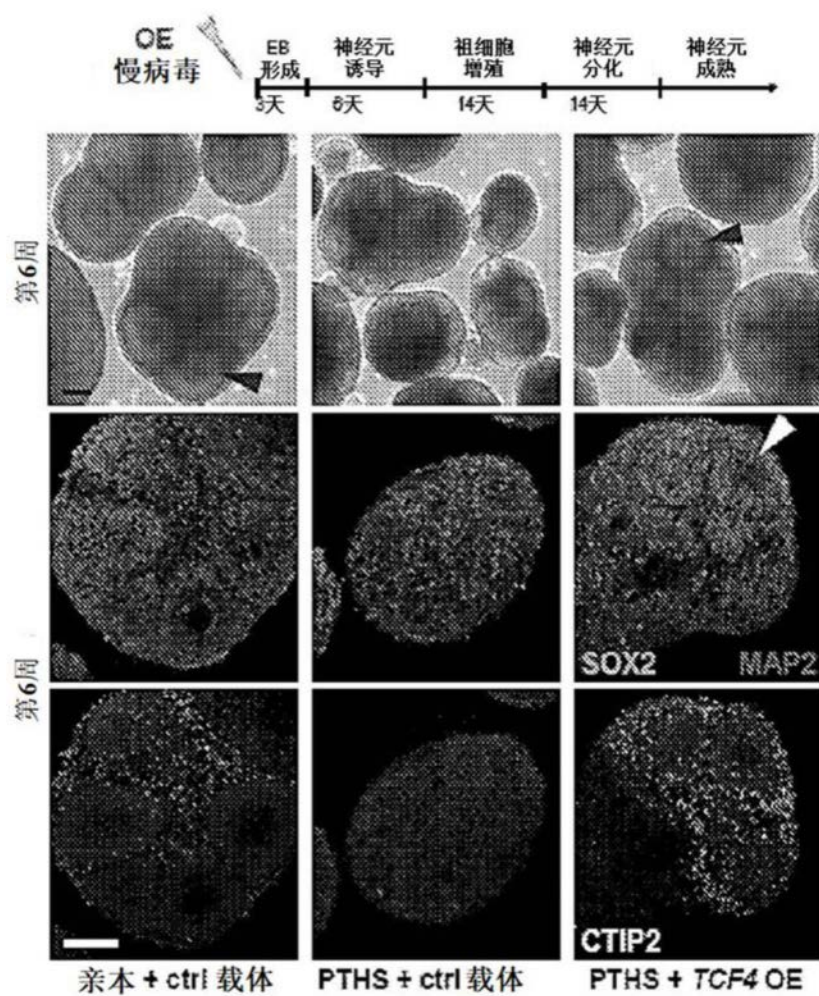


图11F

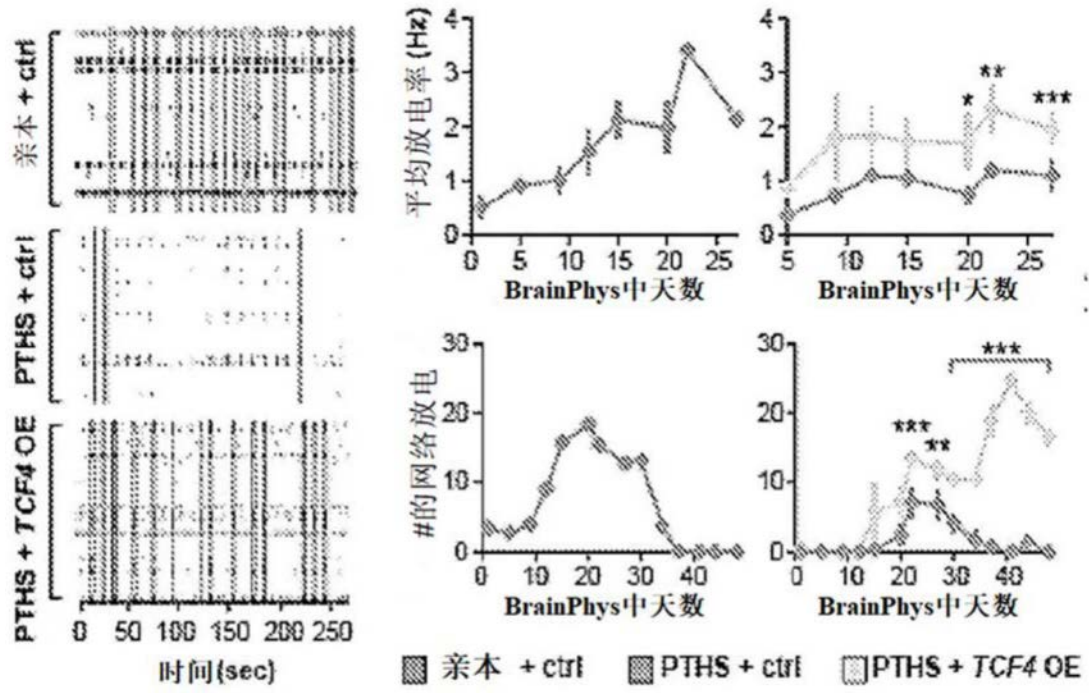


图11G

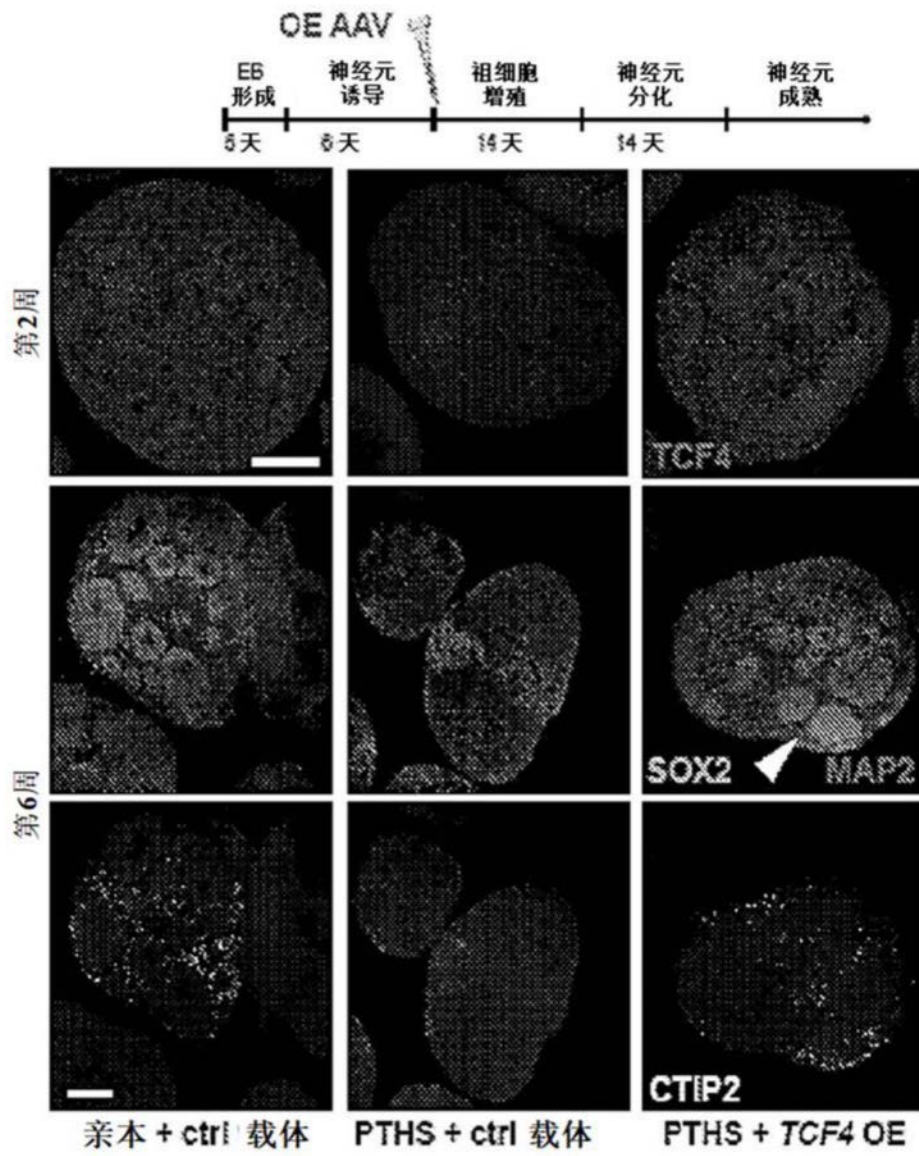


图11H

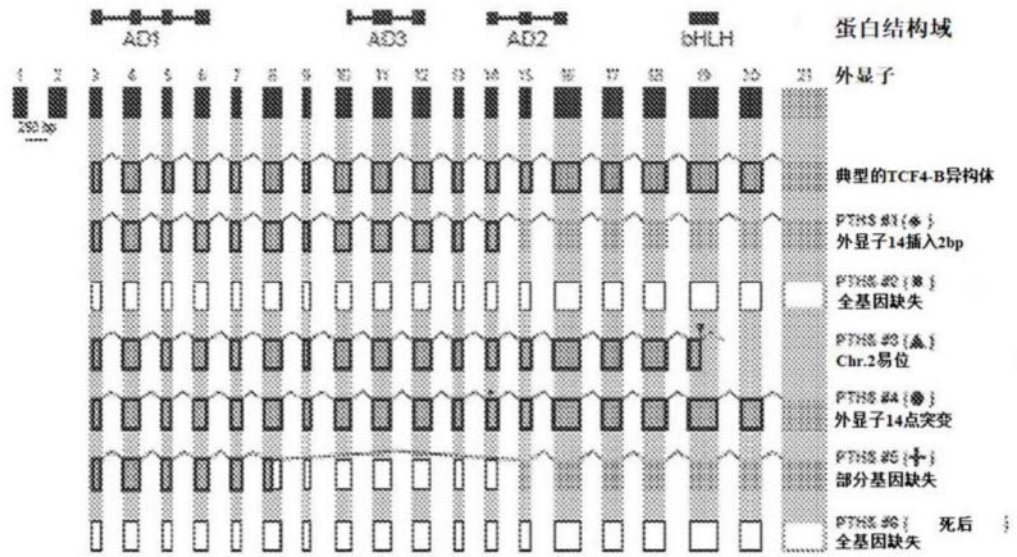


图12A

PTHS #2

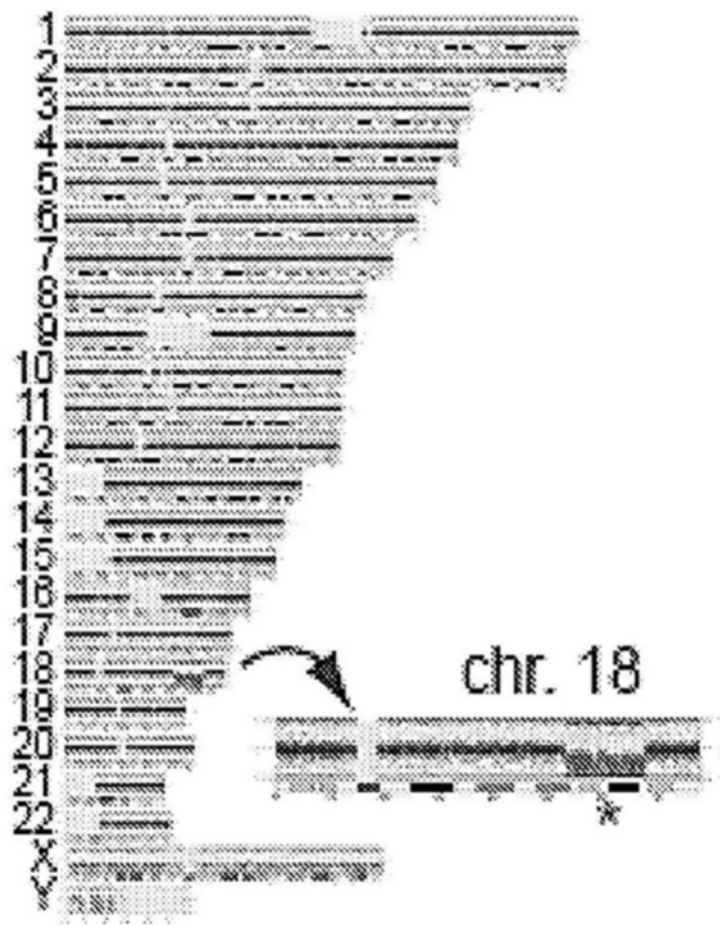


图12B

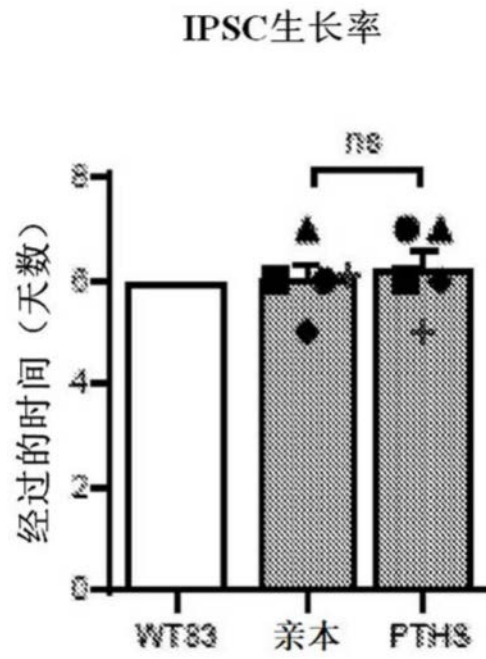


图12C

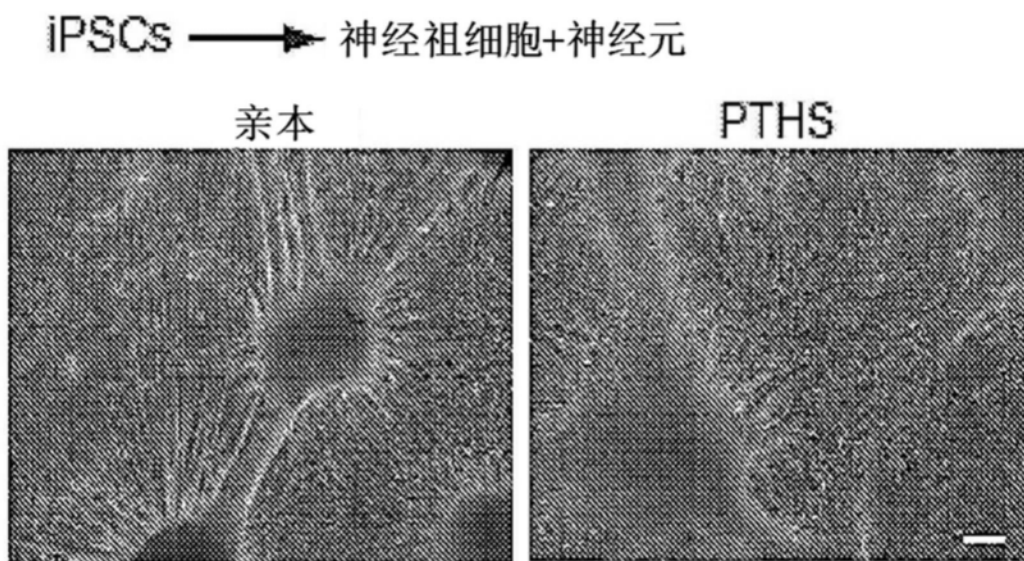


图12D

神经元中SOX2和MAP2

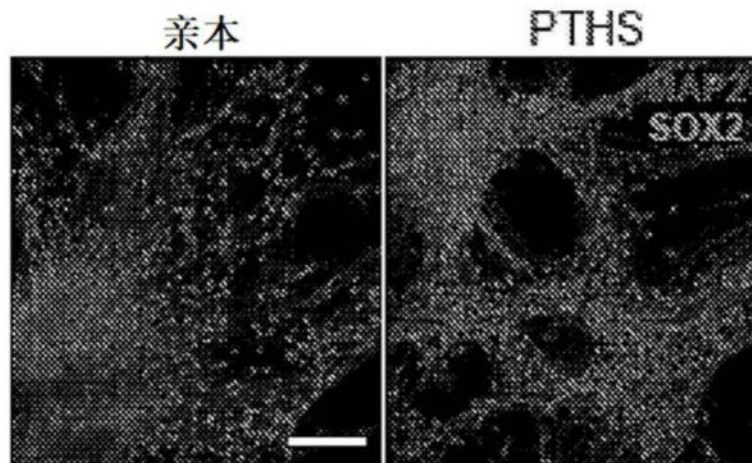


图12E

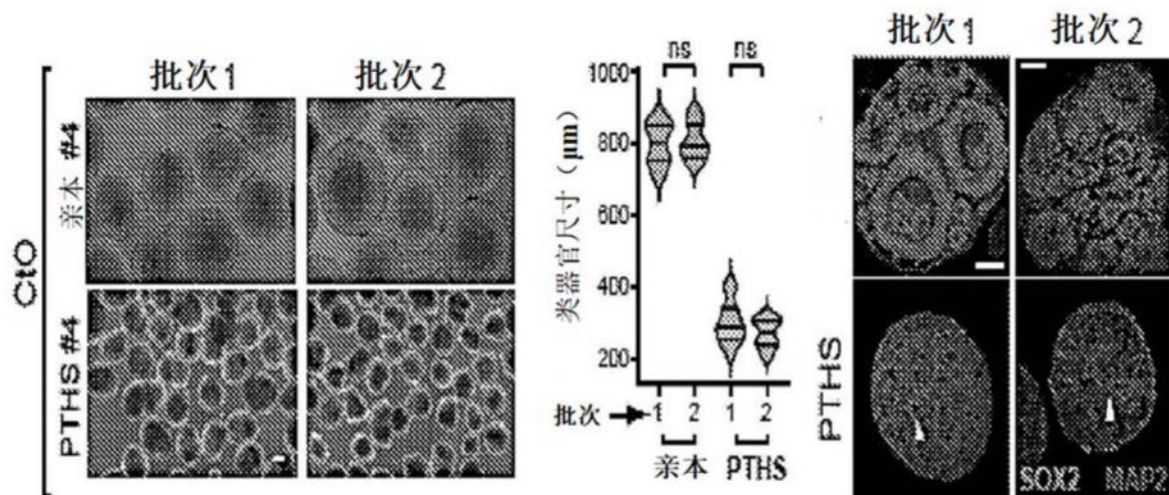


图12F

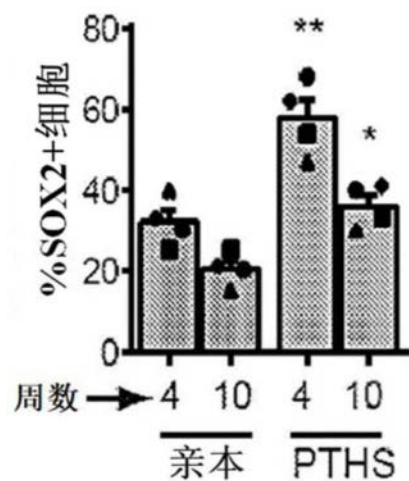


图12G

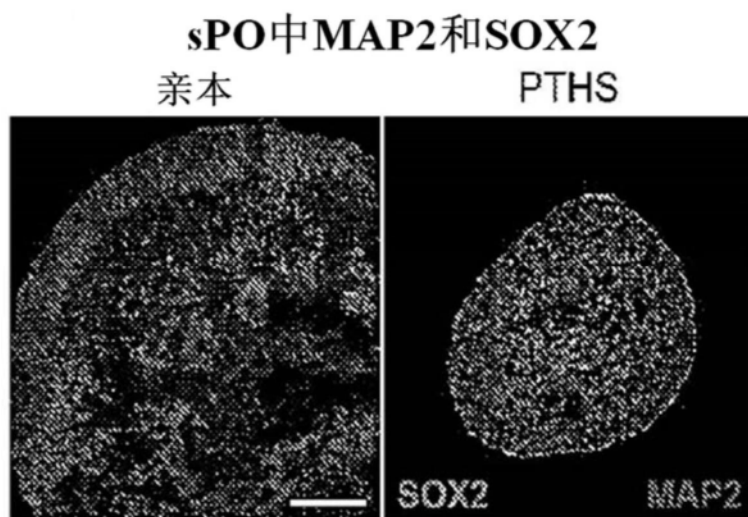


图12H

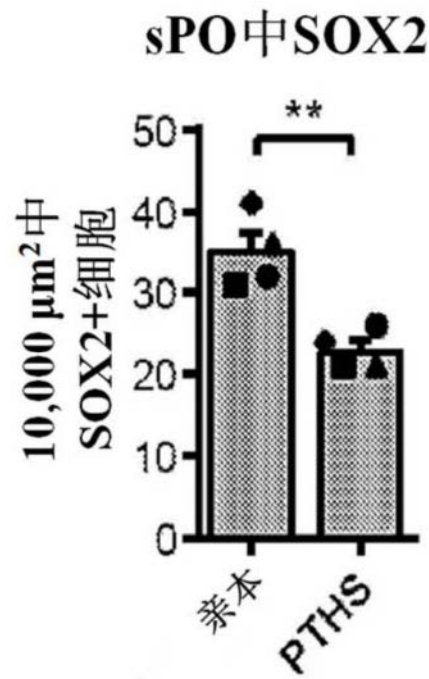


图12I

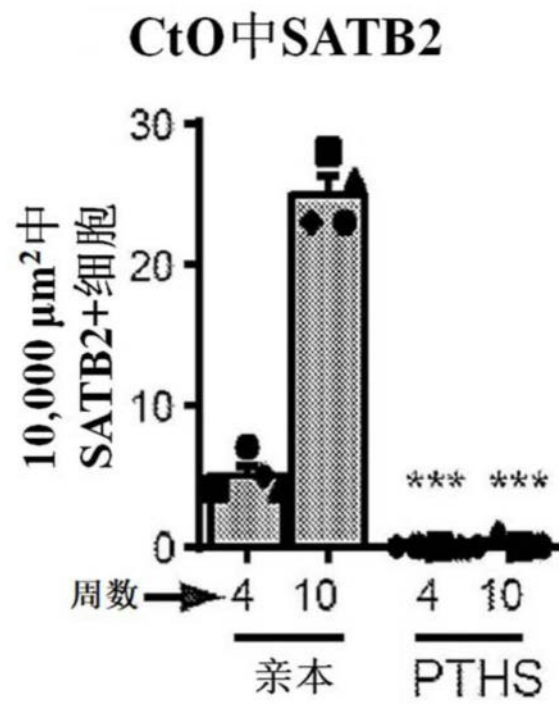


图12J

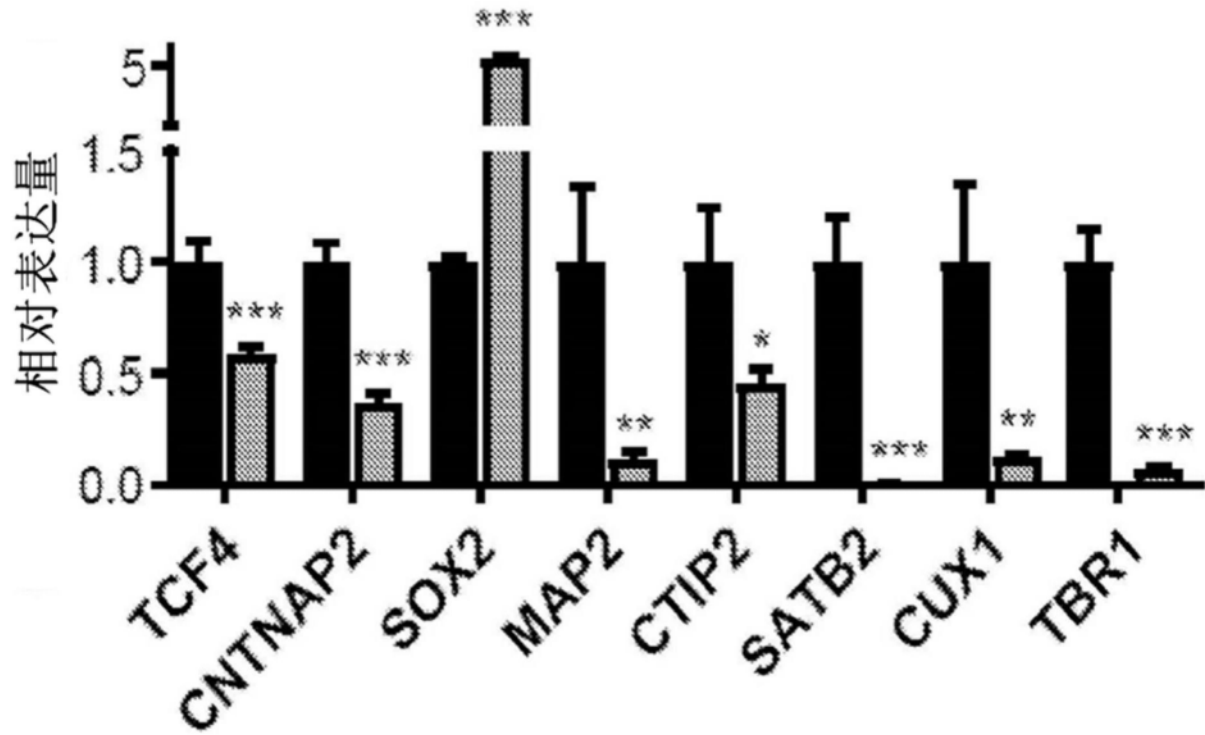


图12K

死后样本中
CTIP2细胞计数

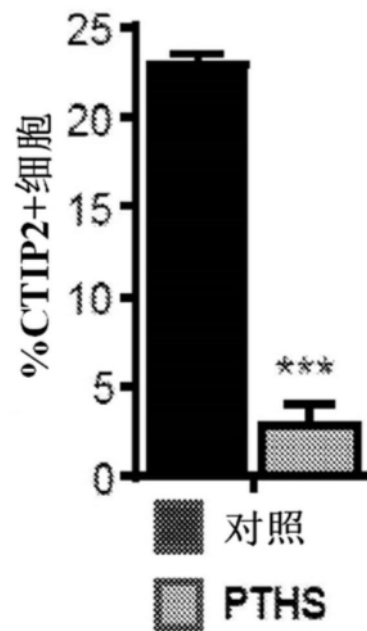


图12L

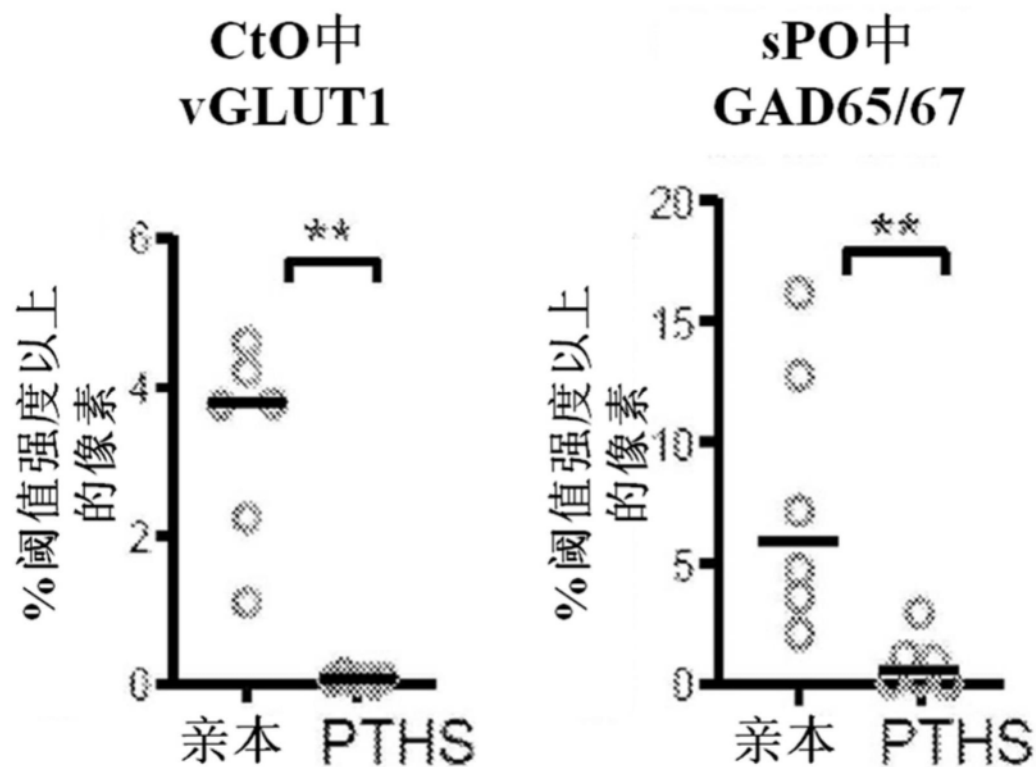


图12M

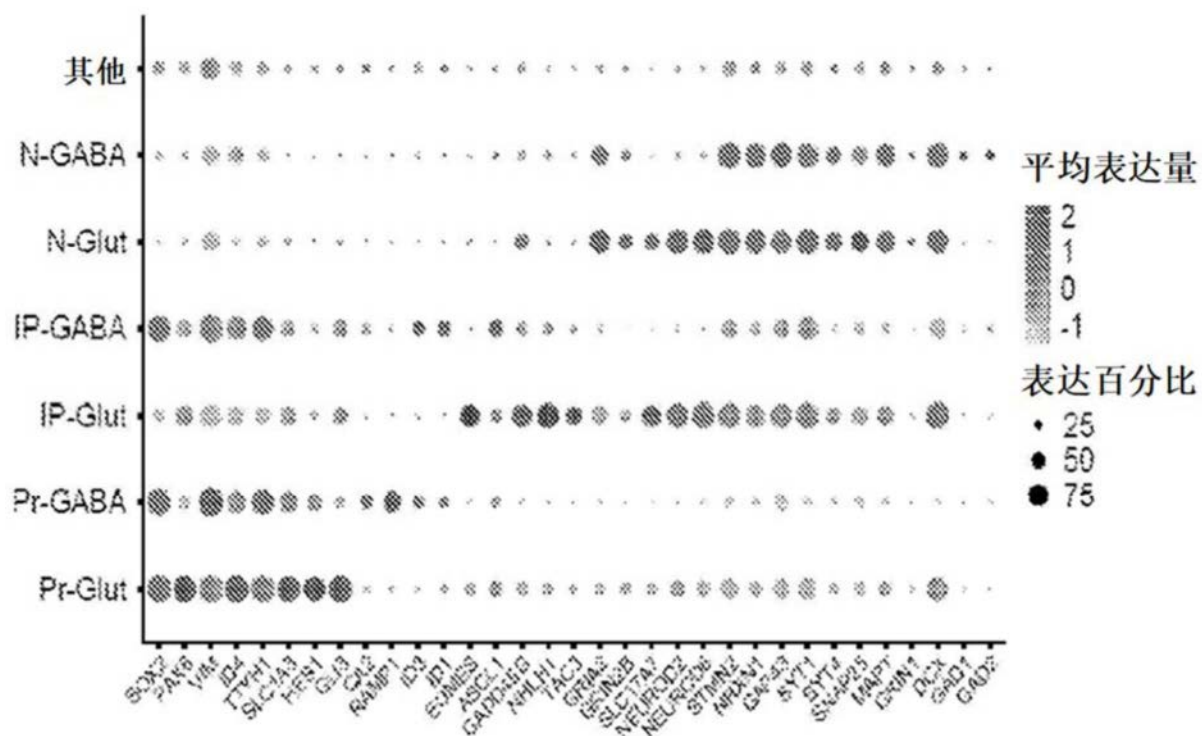


图13A

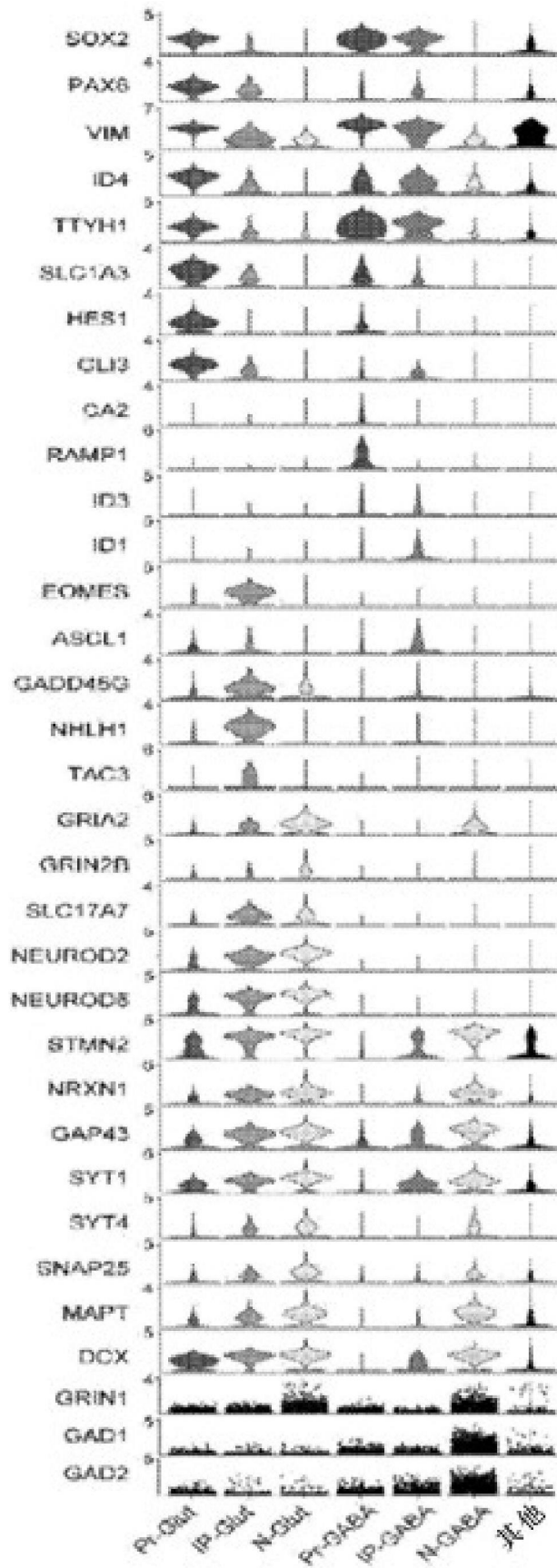


图13B

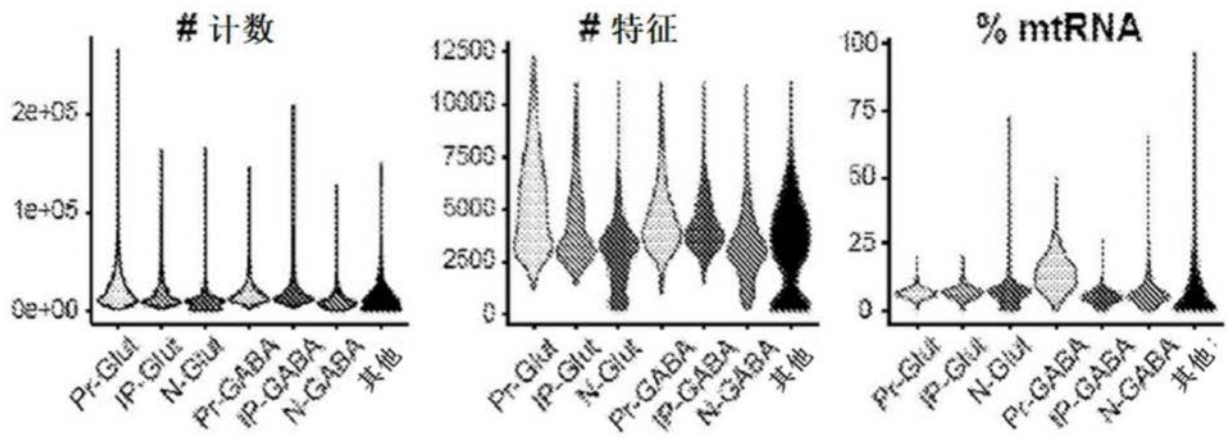


图13C

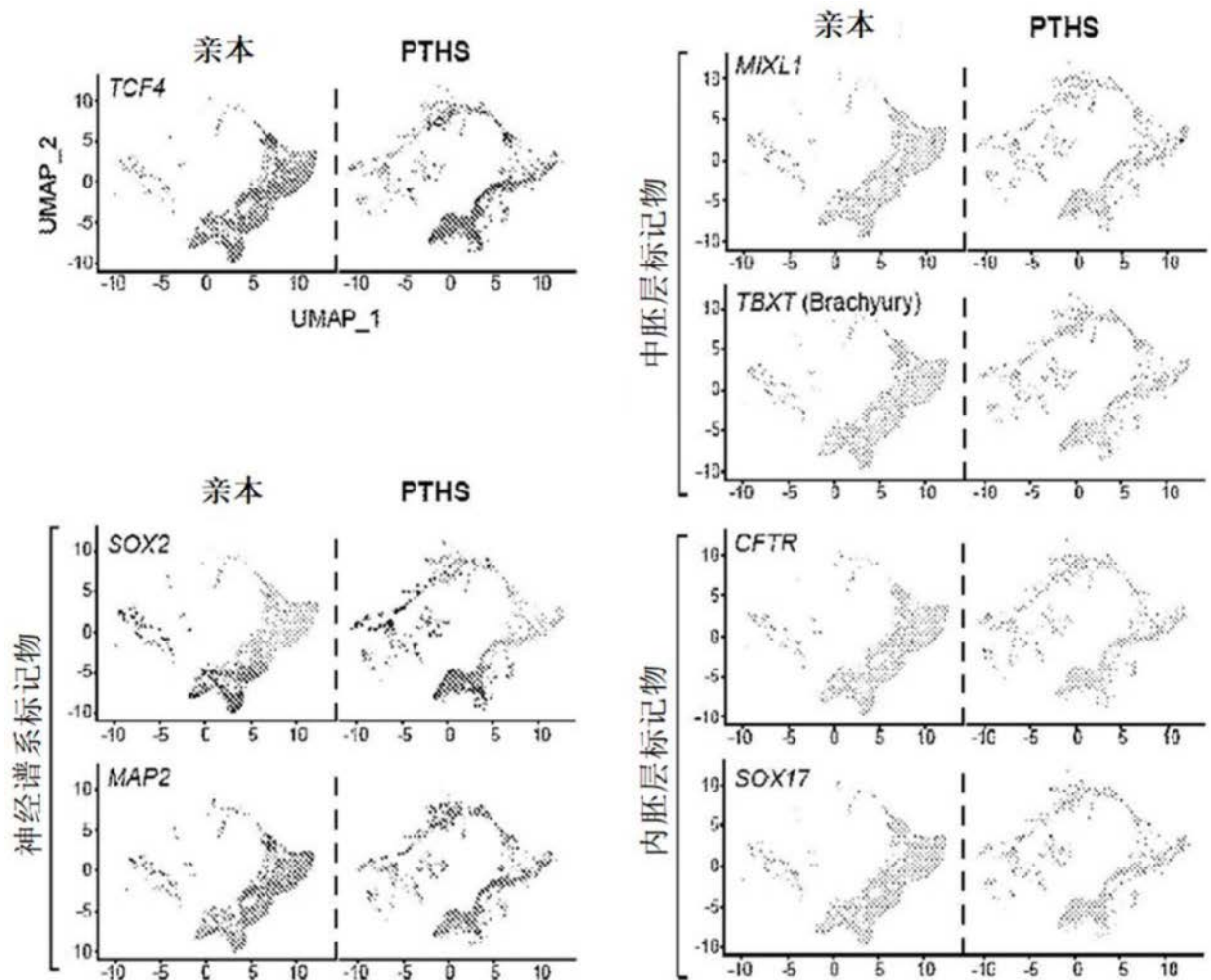


图13D

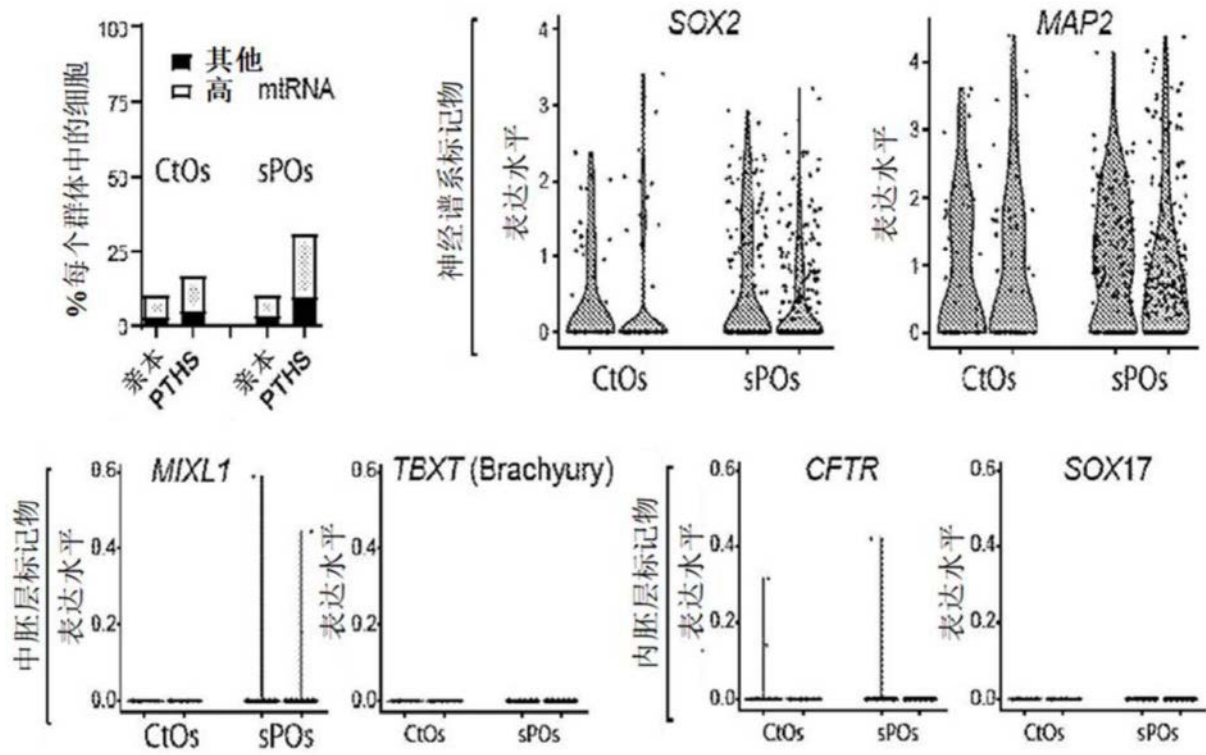


图13E

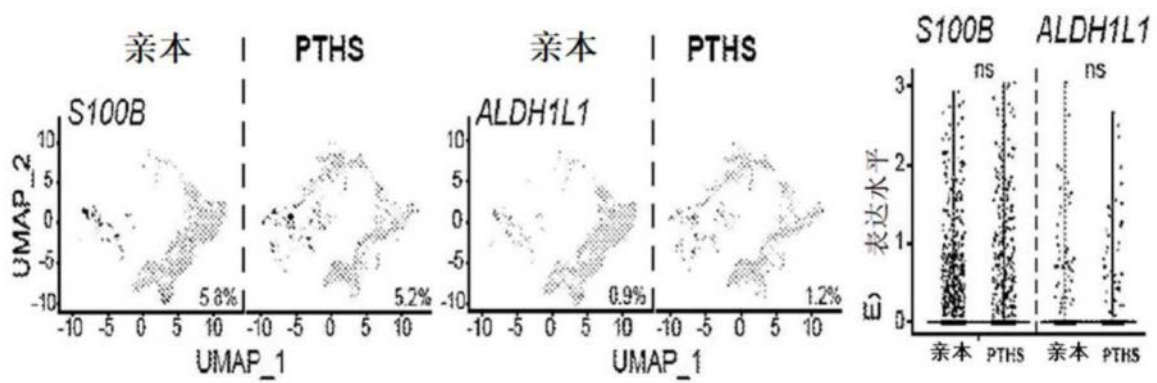


图13F

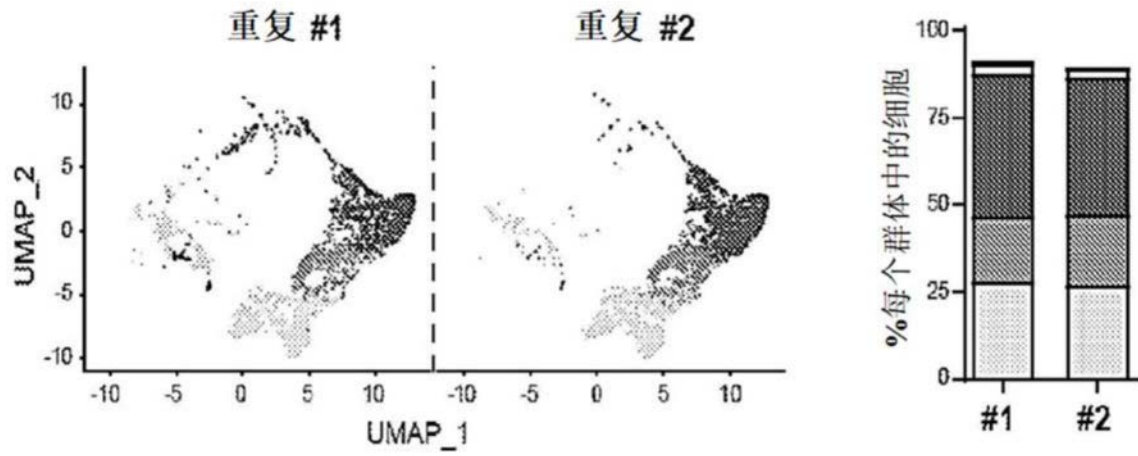


图13G

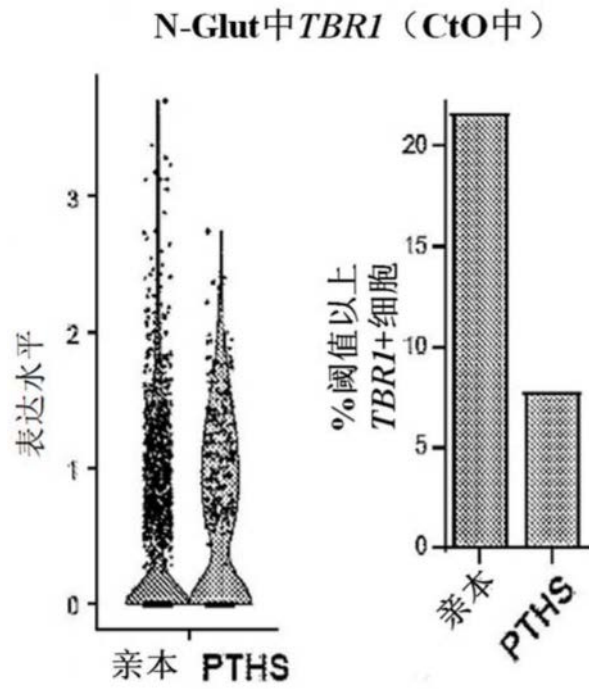


图13H

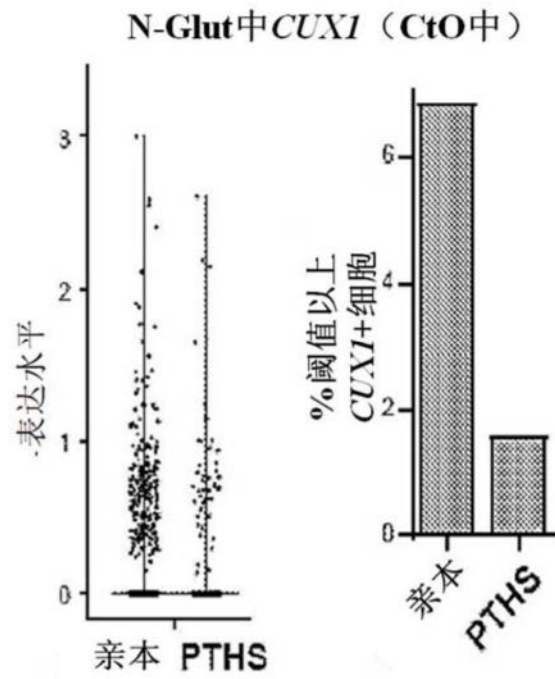


图13I

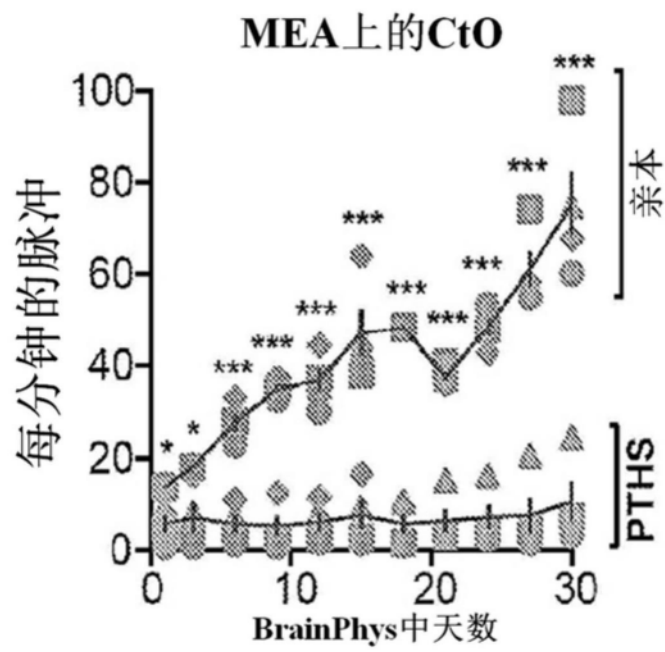


图14A

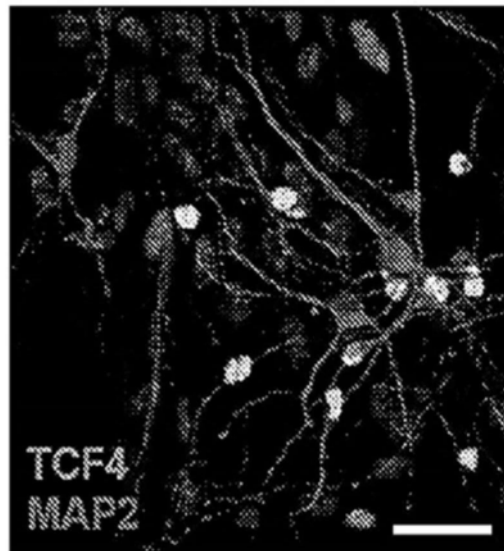


图14B

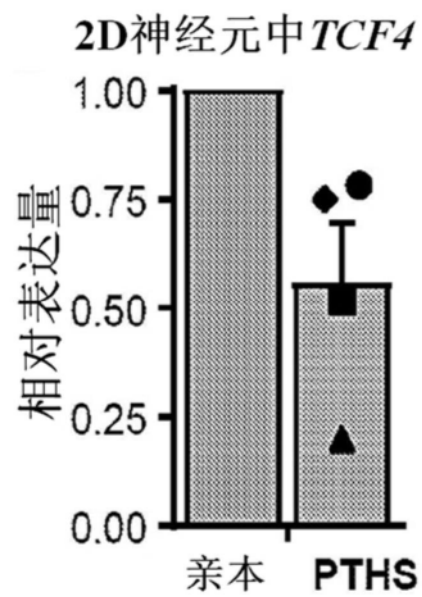


图14C

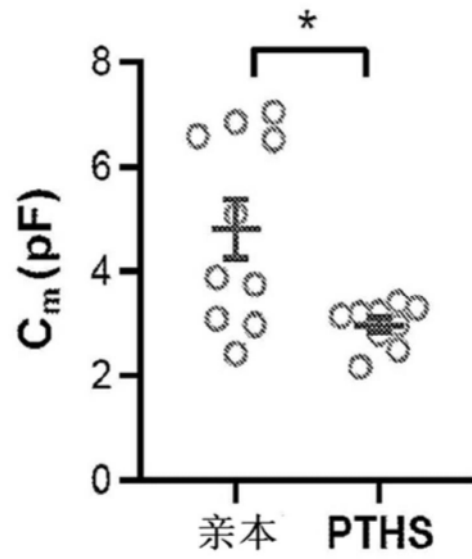


图14D

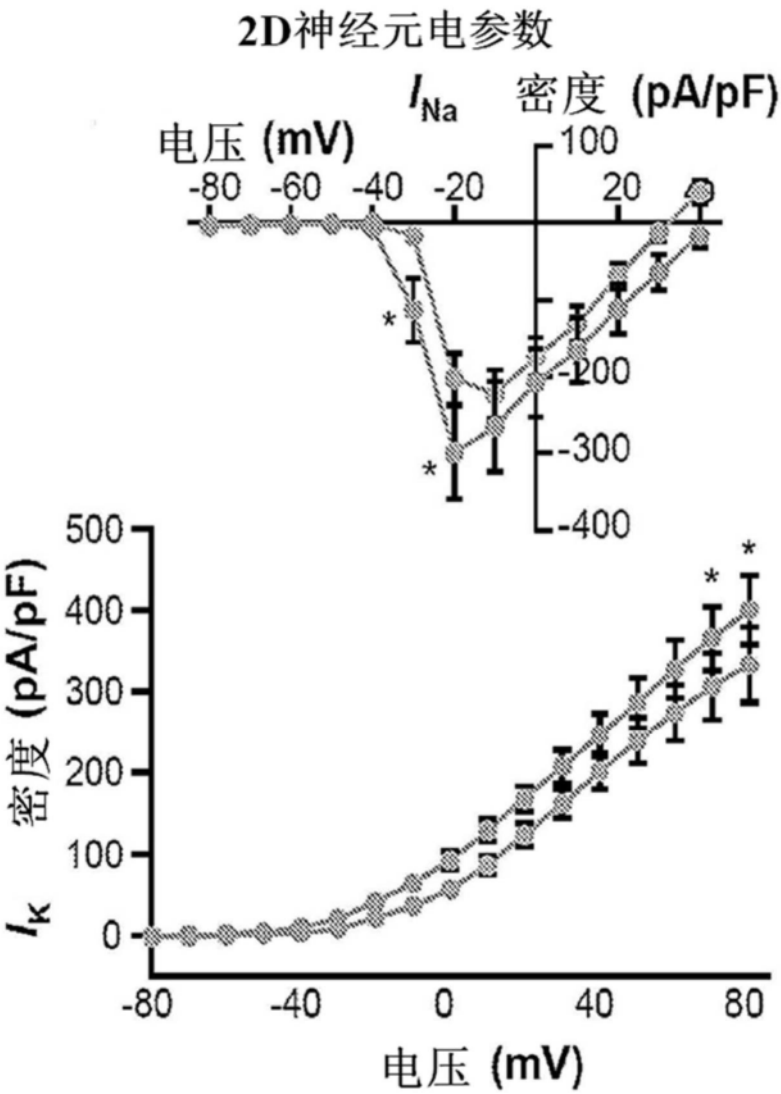


图14E

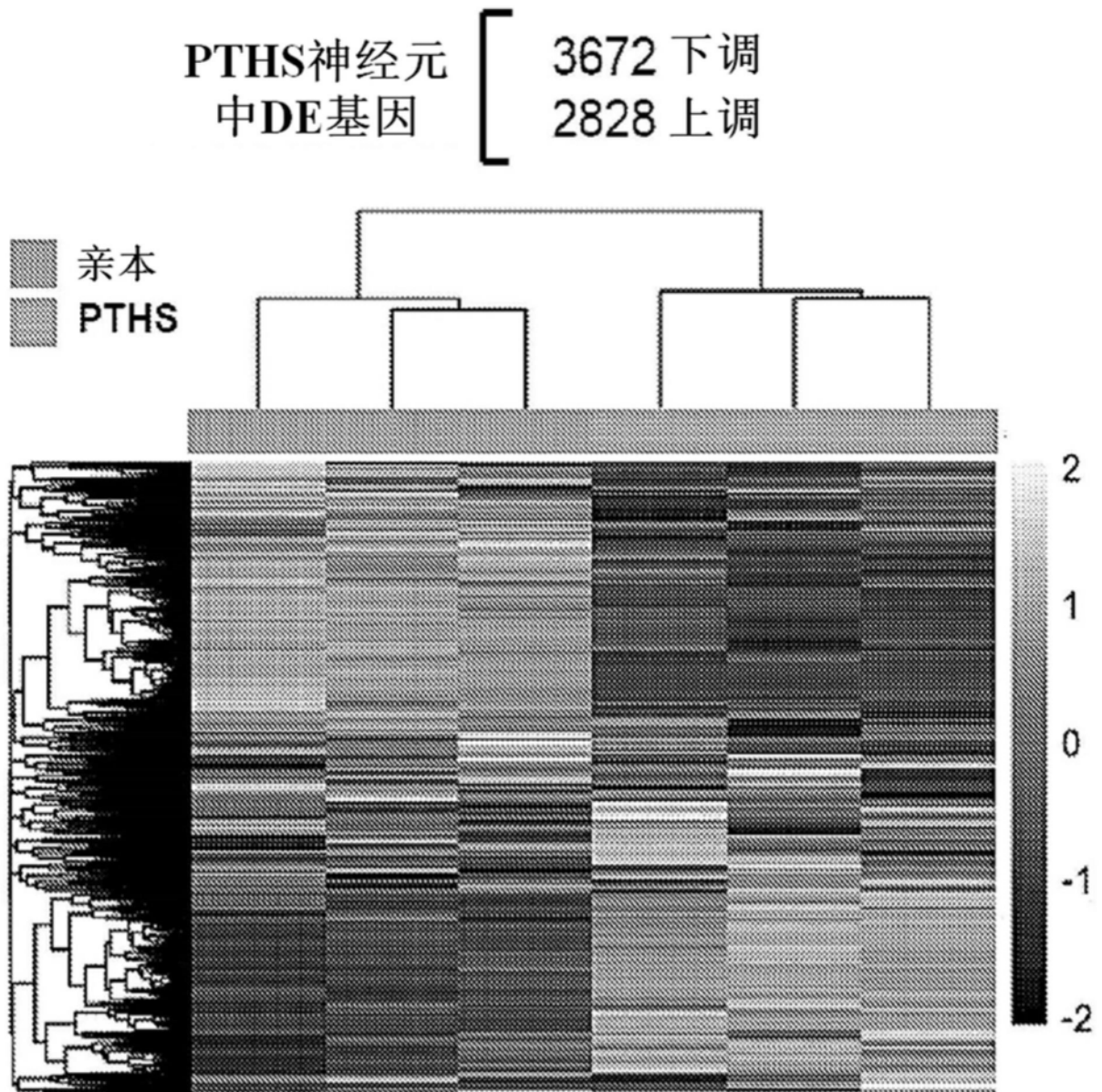


图14F

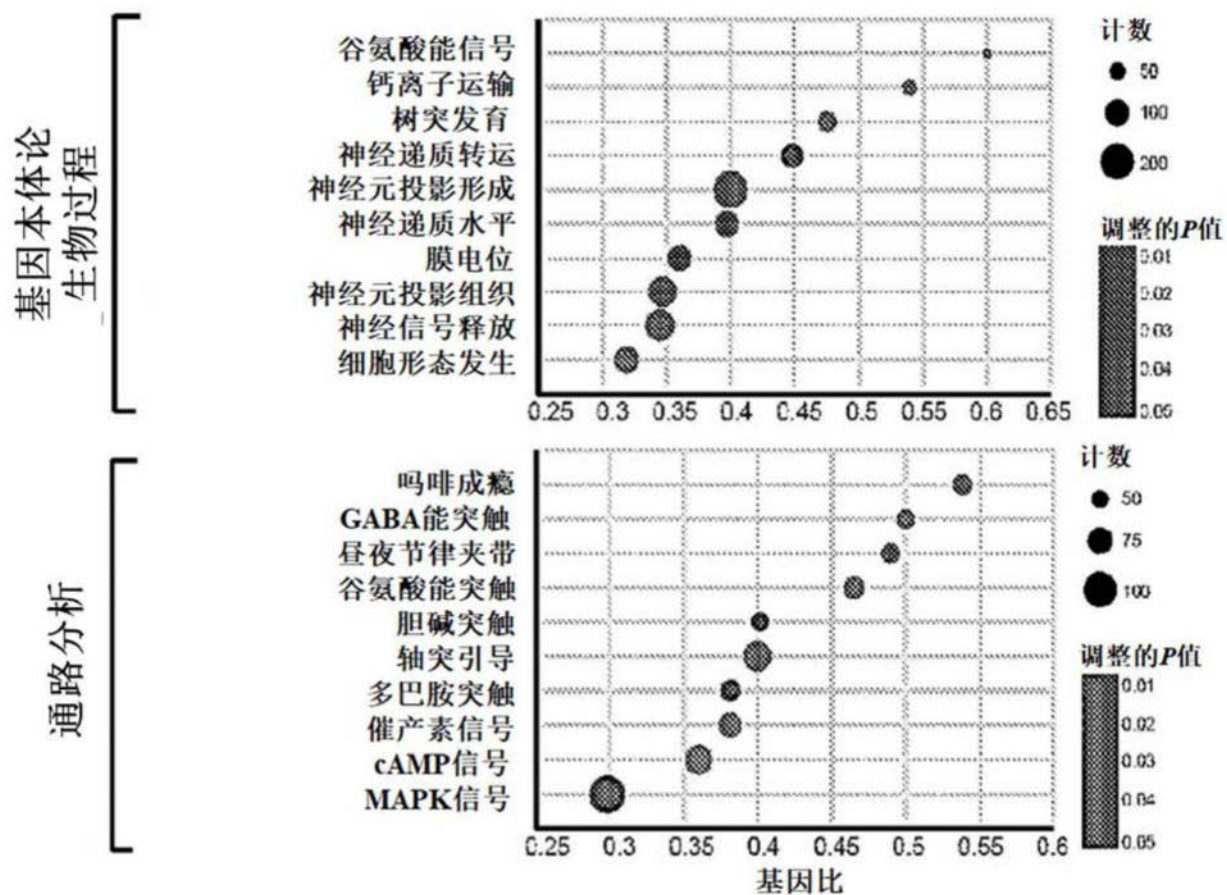


图14G

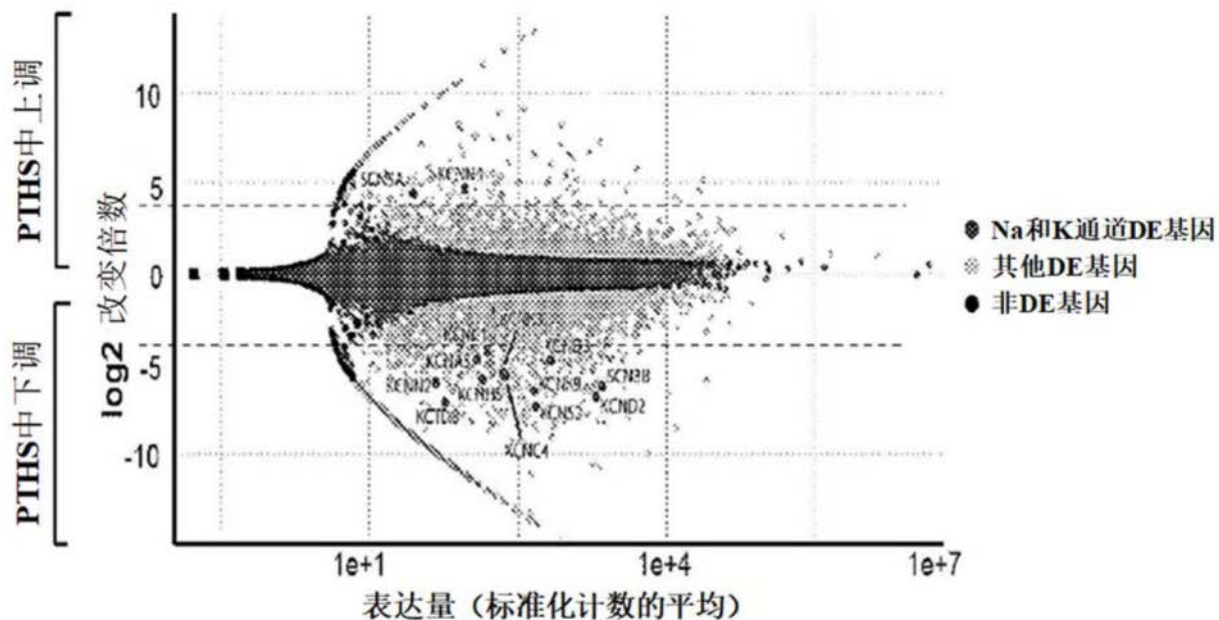


图14H

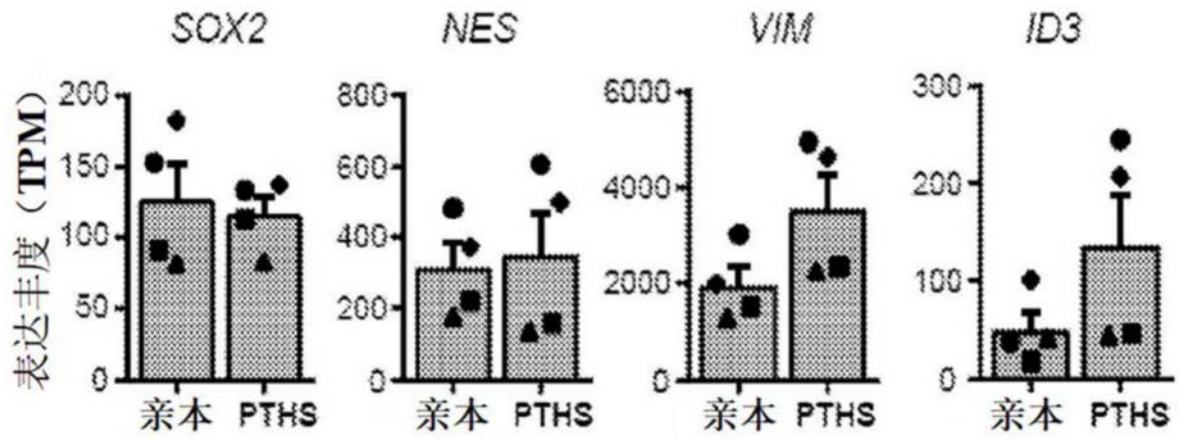


图15A

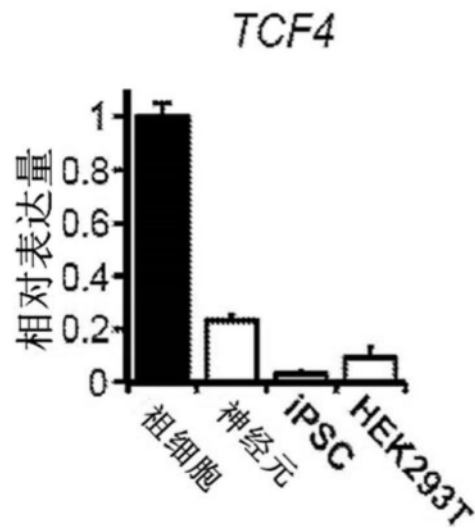


图15B

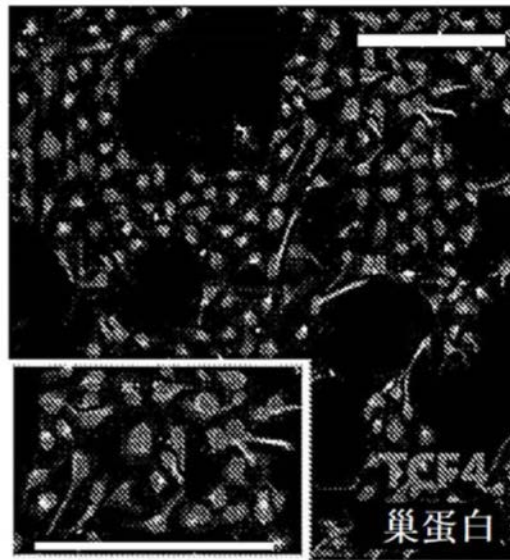


图15C

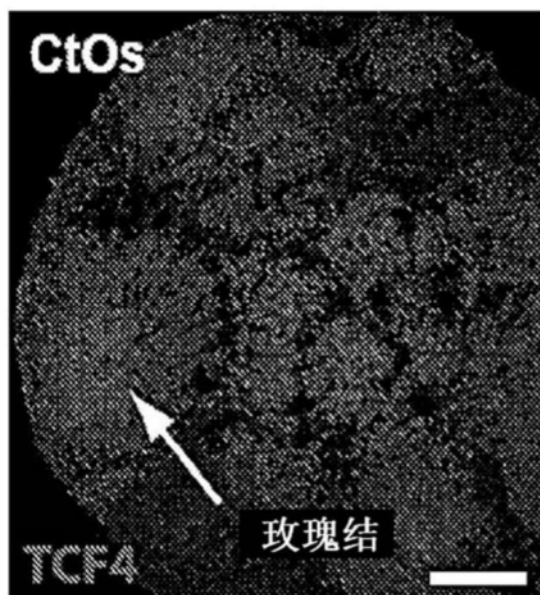


图15D

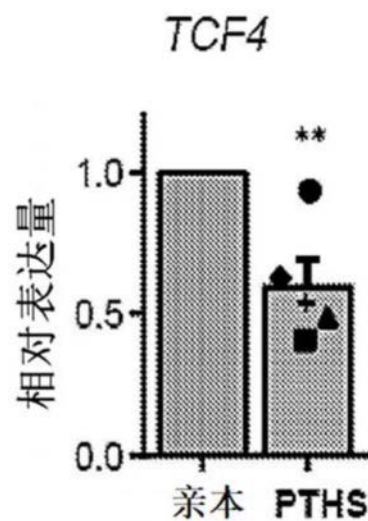


图15E

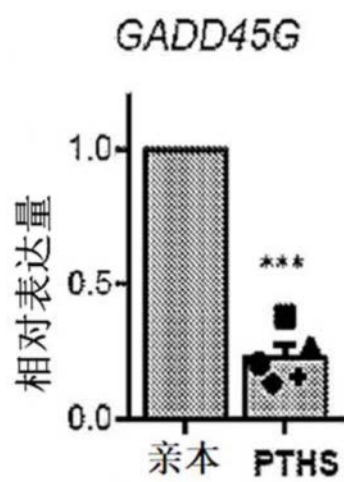


图15F

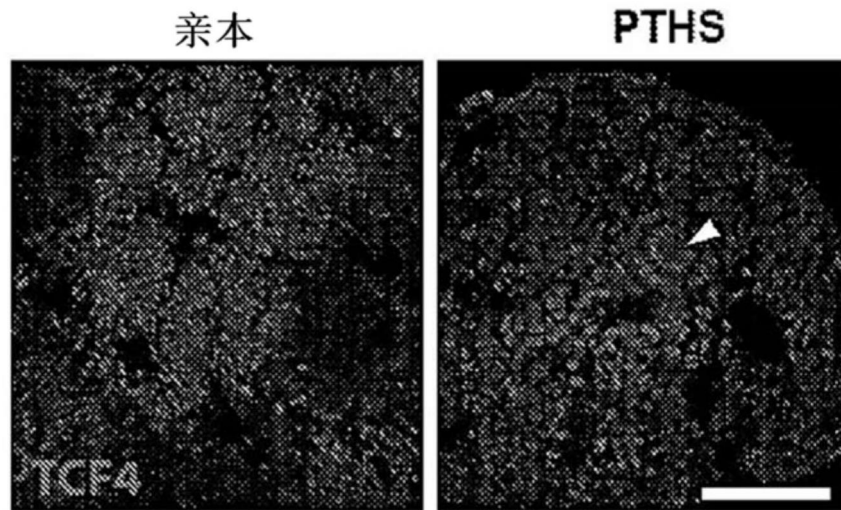


图15G

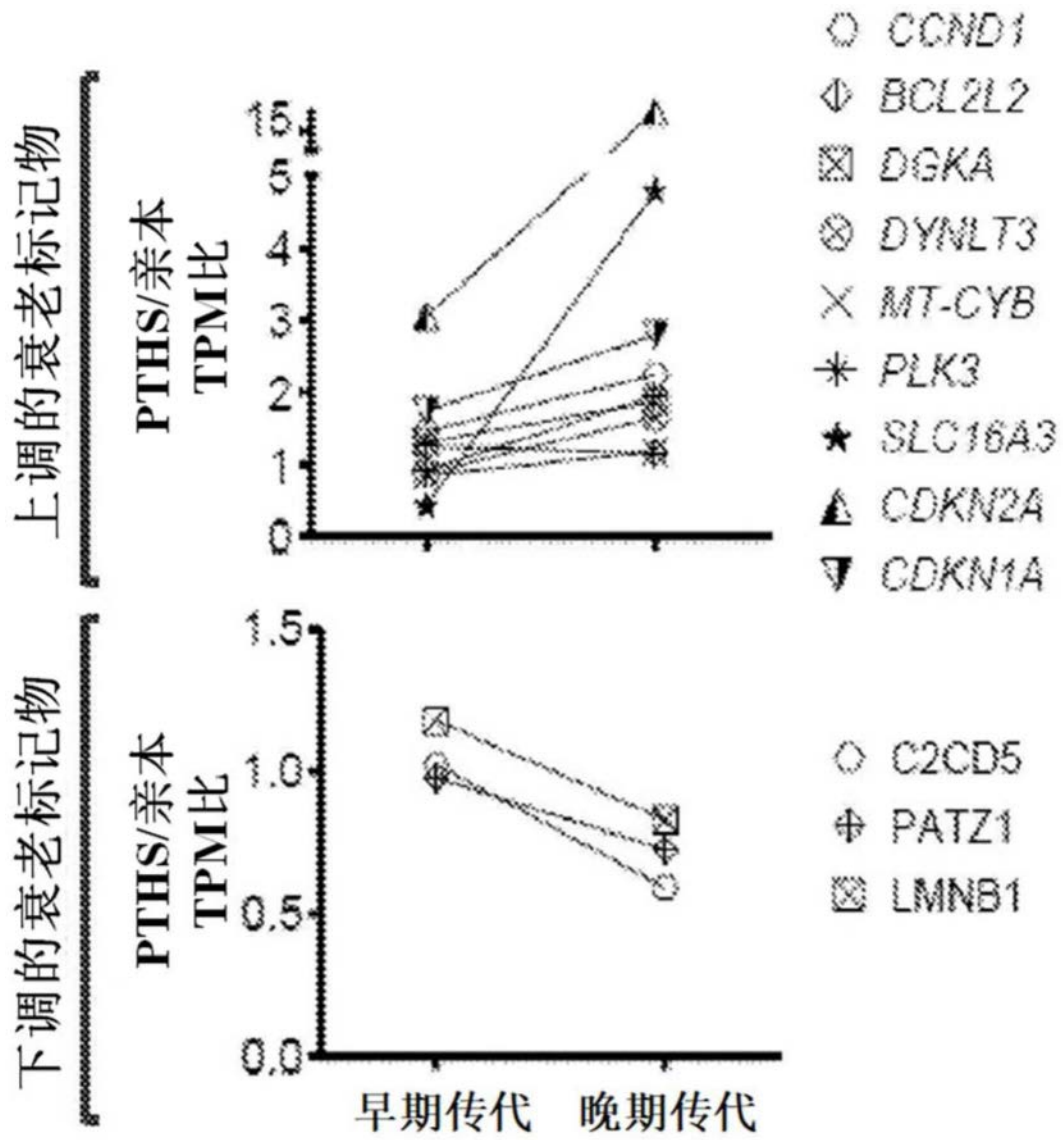


图15H

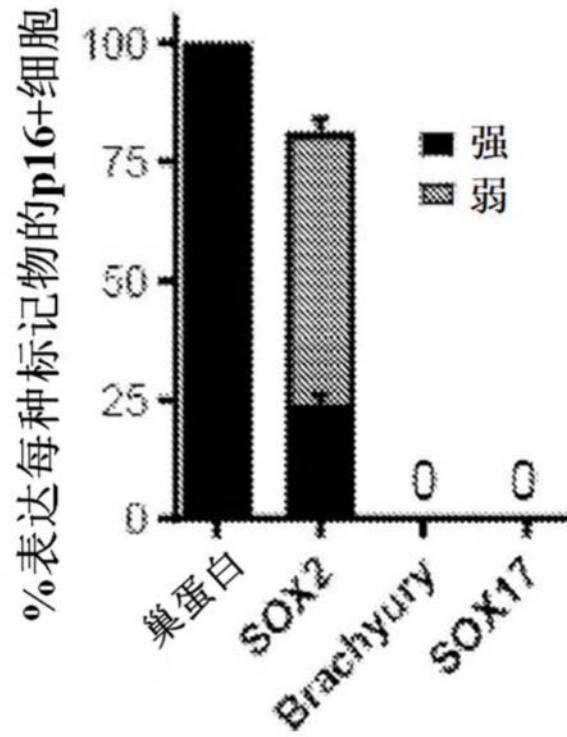


图15I

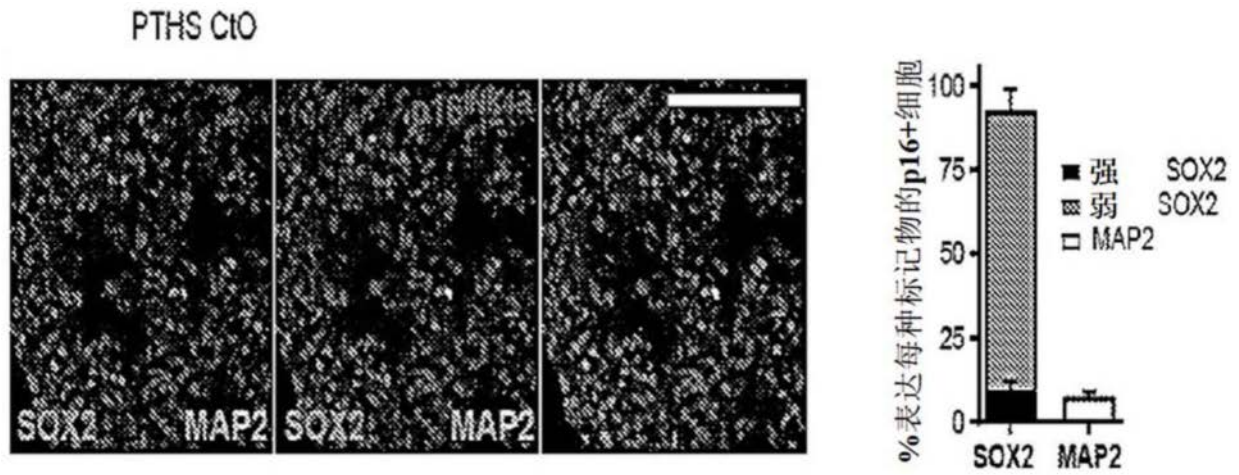


图15J

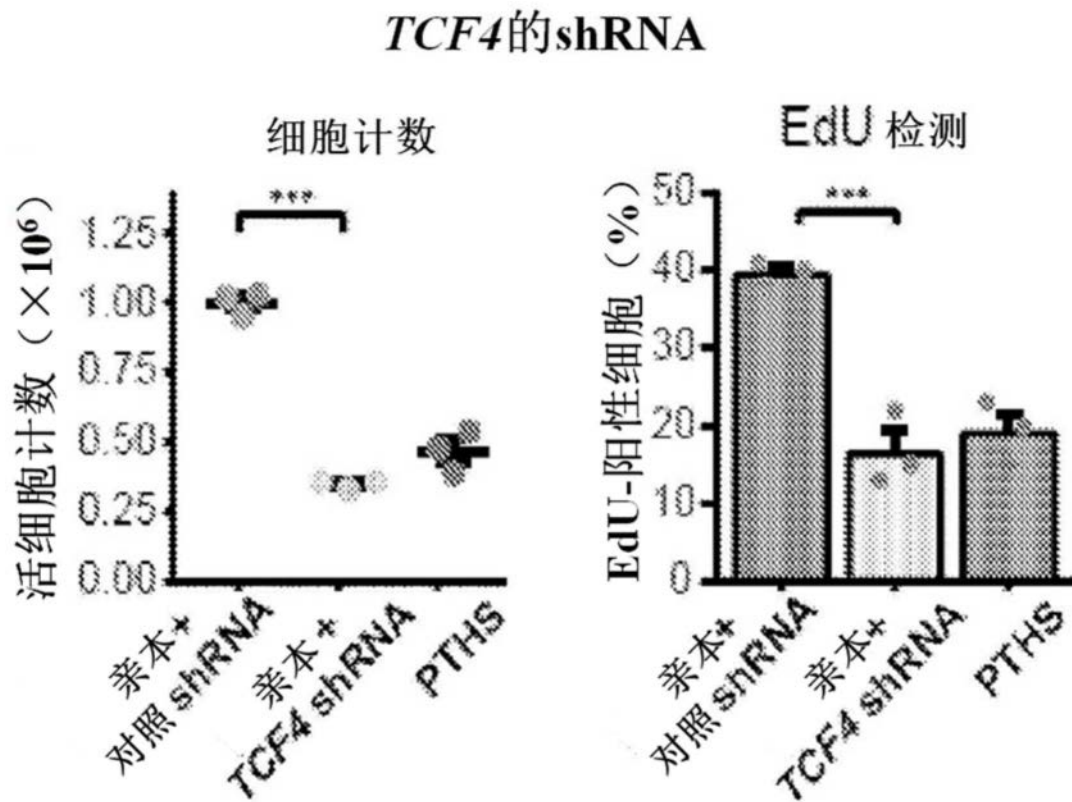


图15K

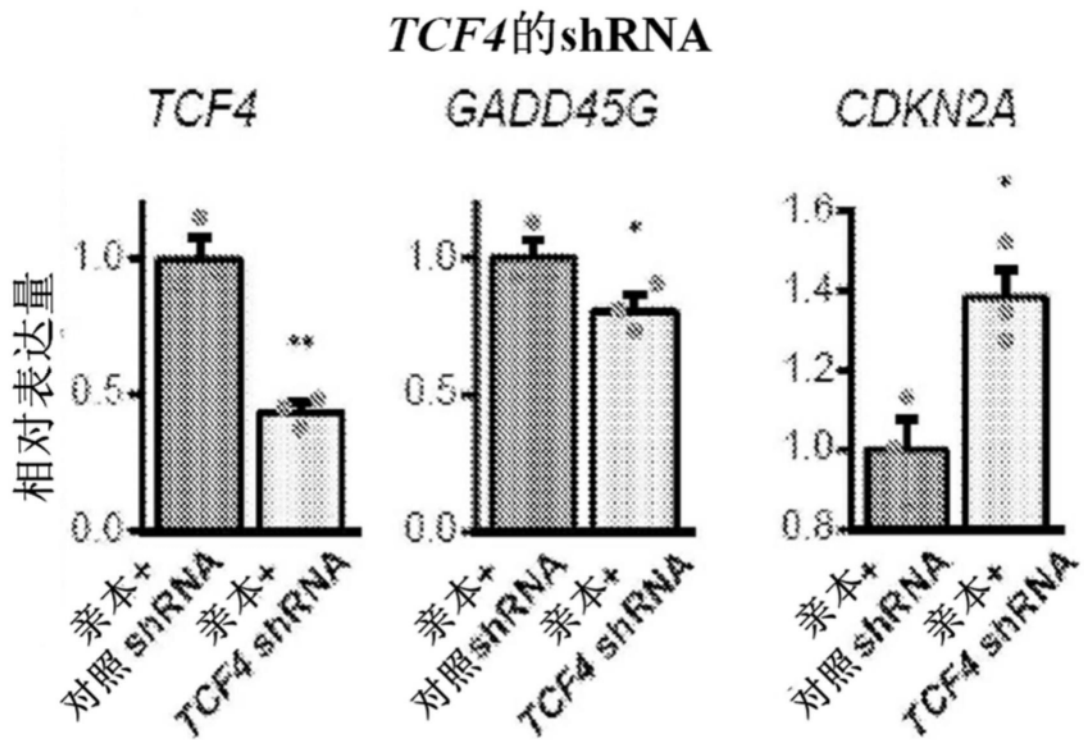


图15L

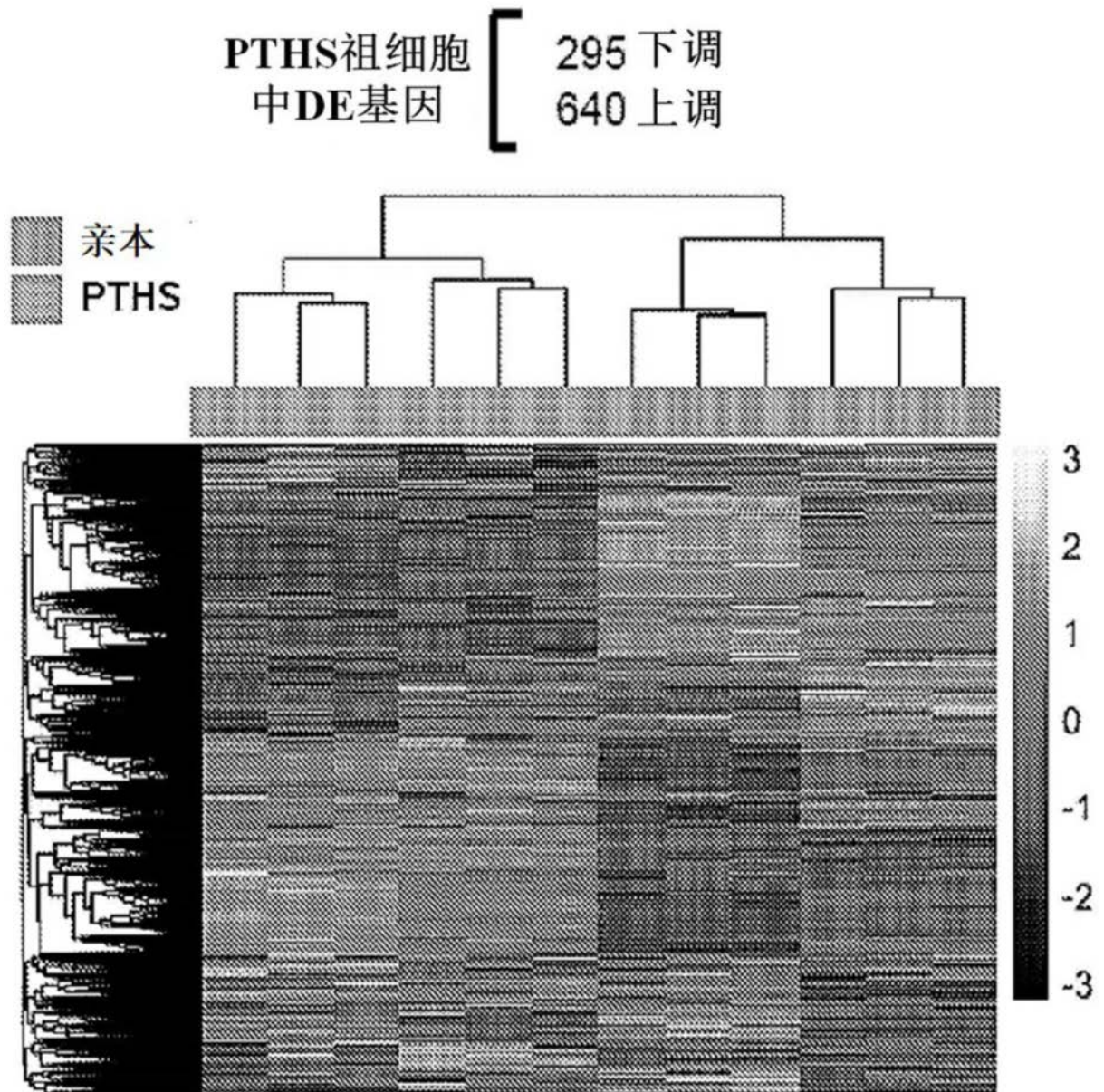


图15M

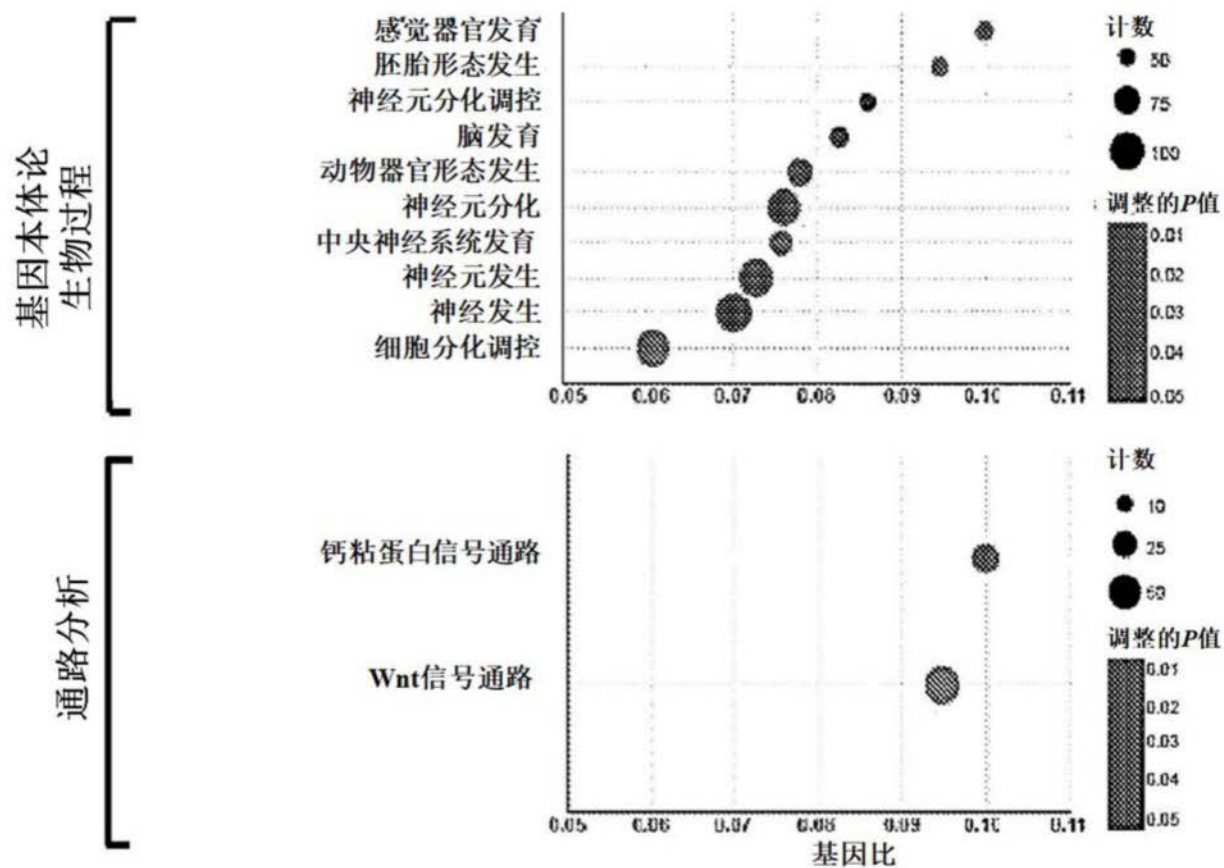


图15N

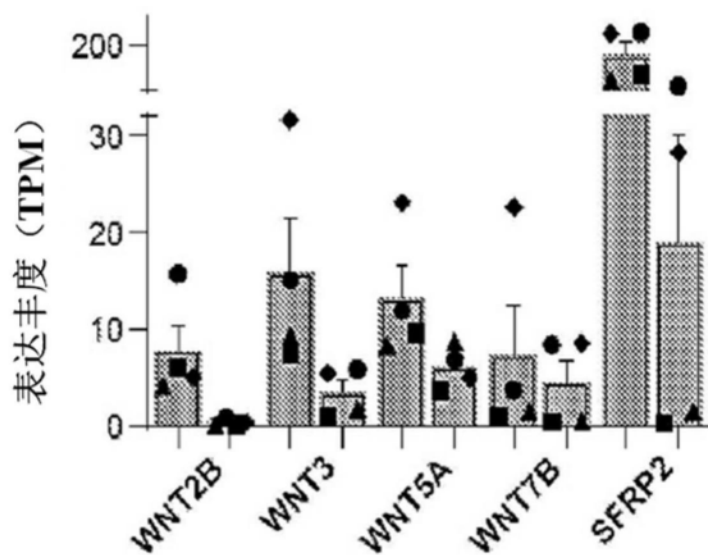


图16A

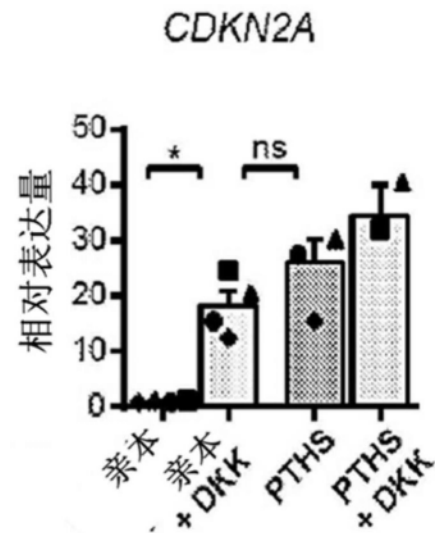


图16B

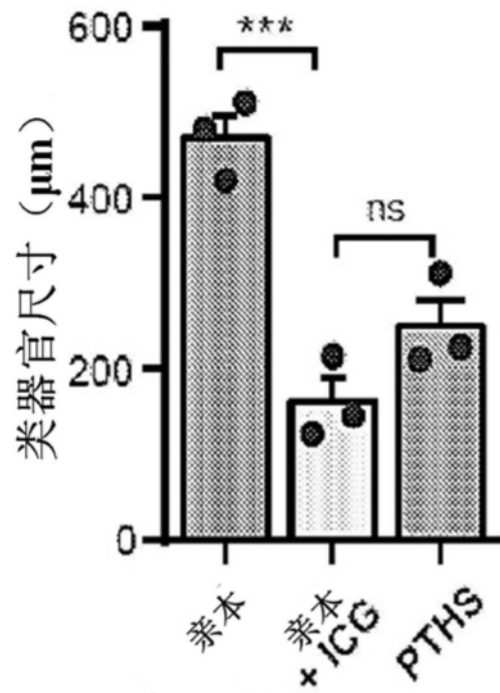


图16C

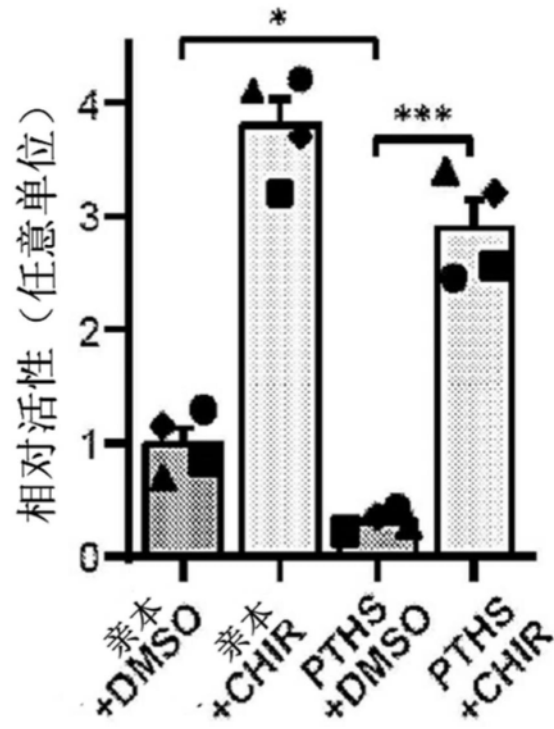


图16D

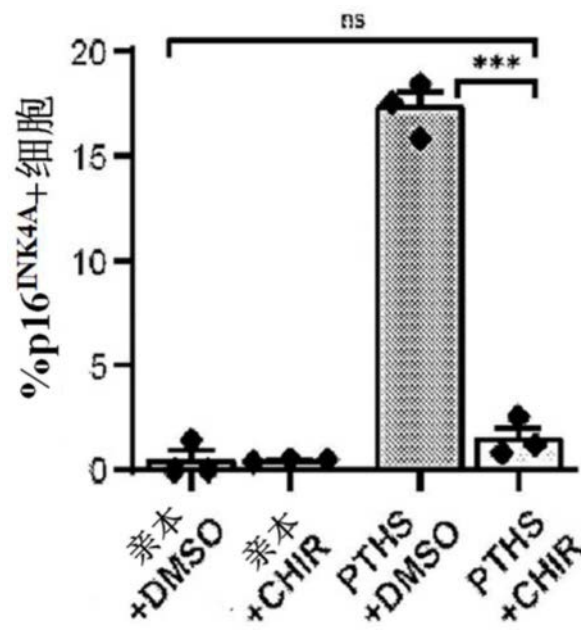


图16E

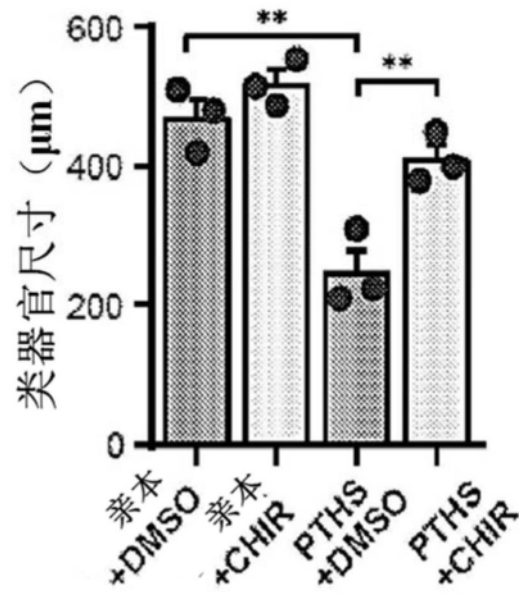


图16F

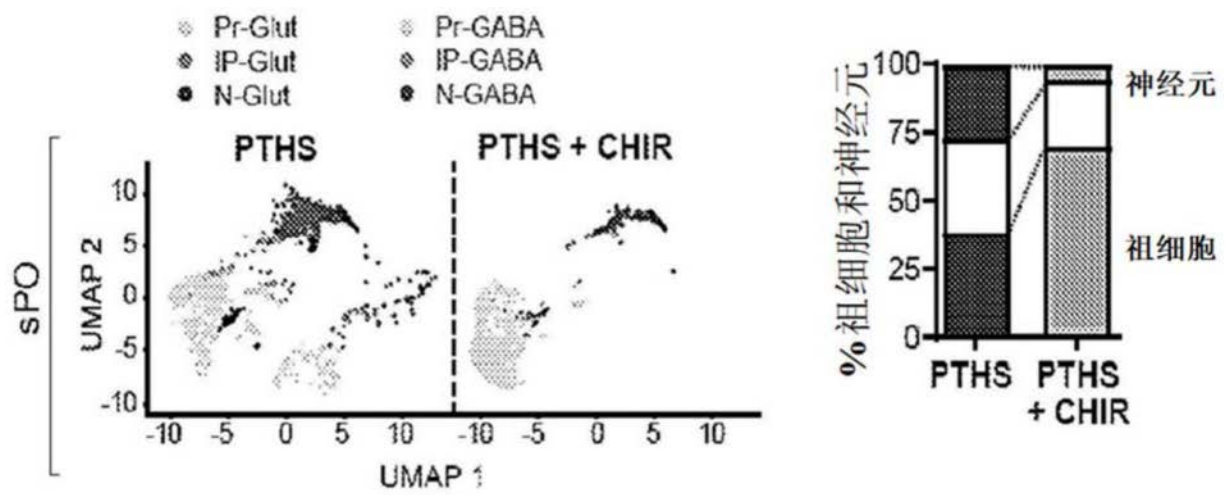


图16G

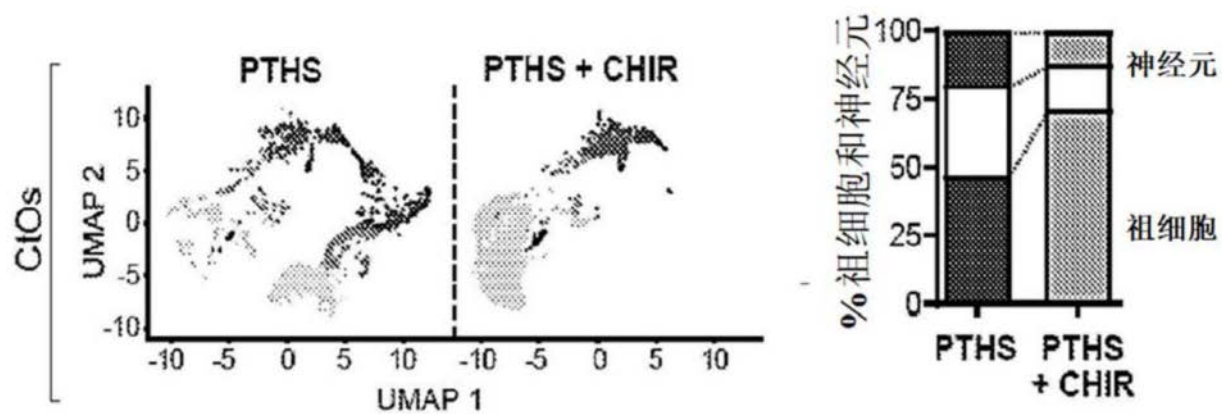


图16H

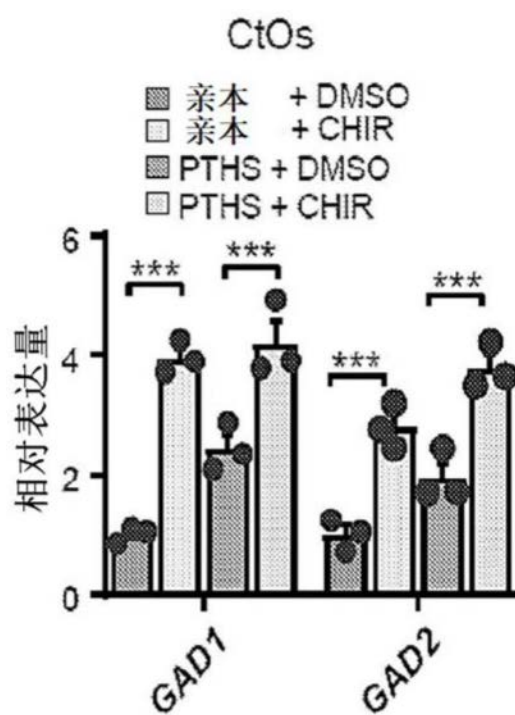


图16I

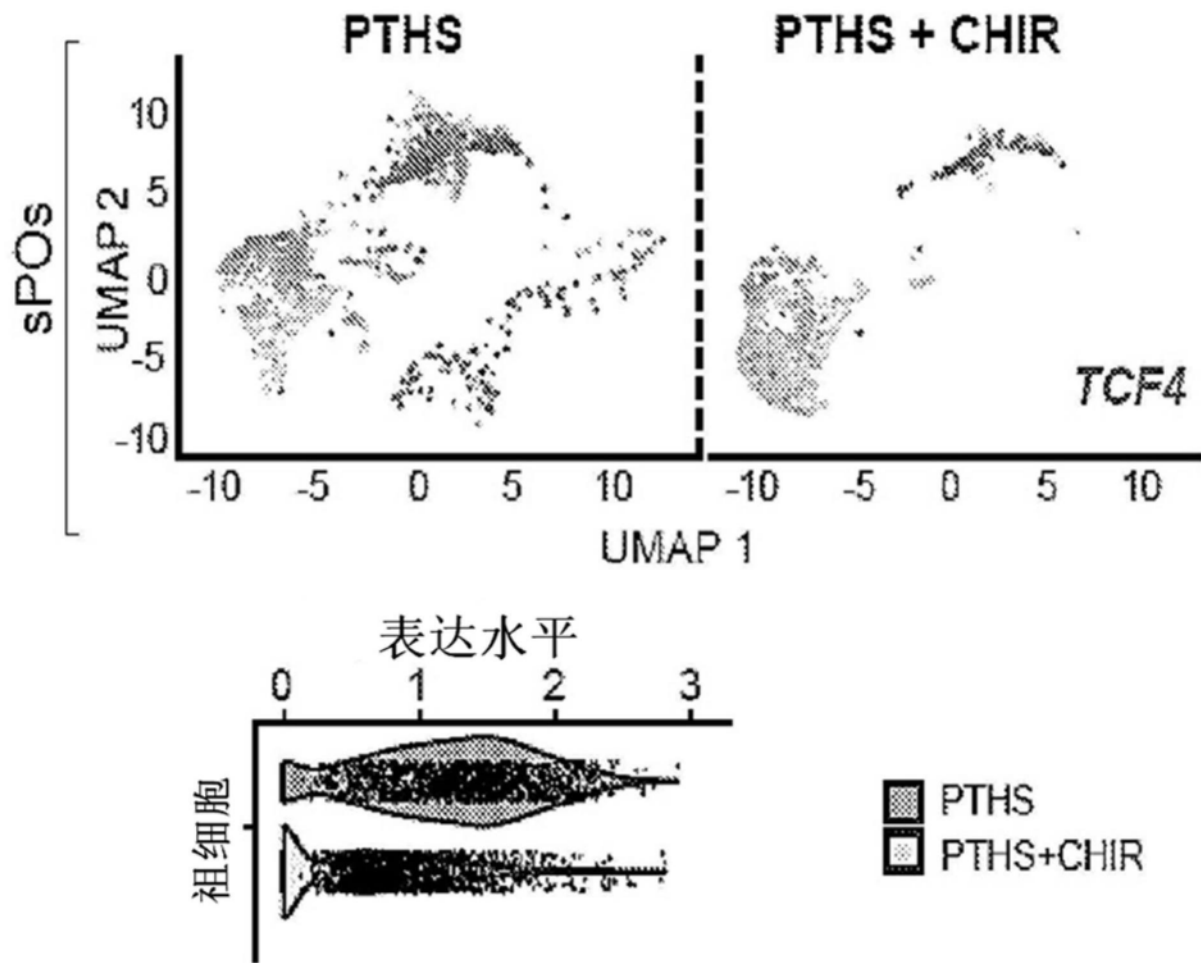


图16J

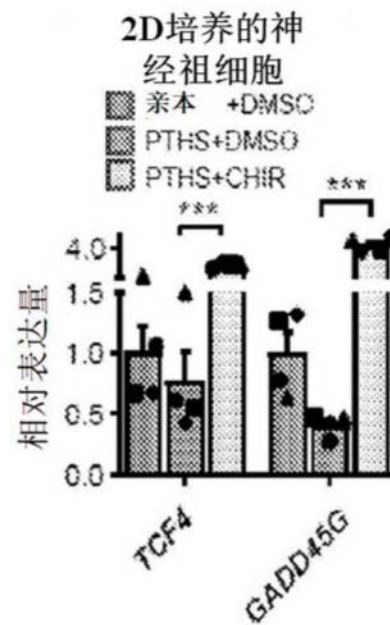


图16K

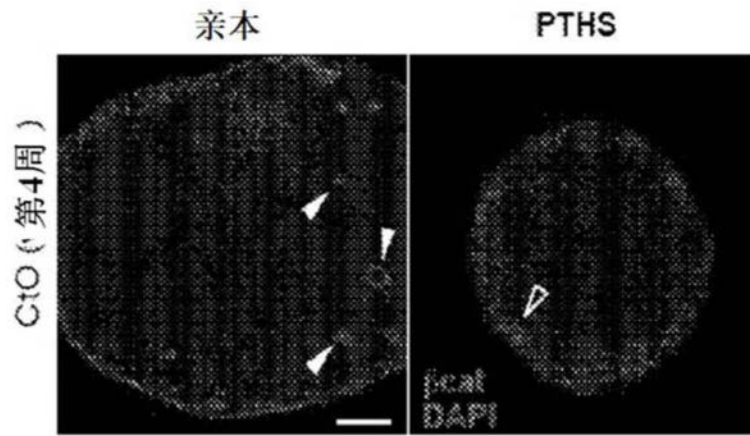


图16L

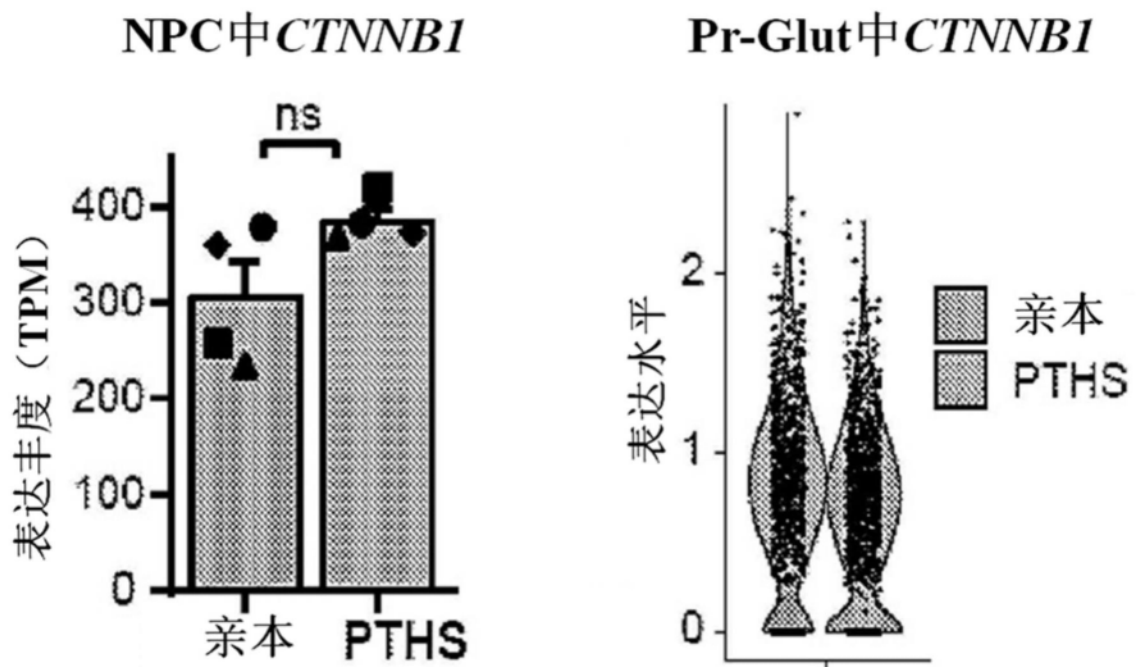


图16M

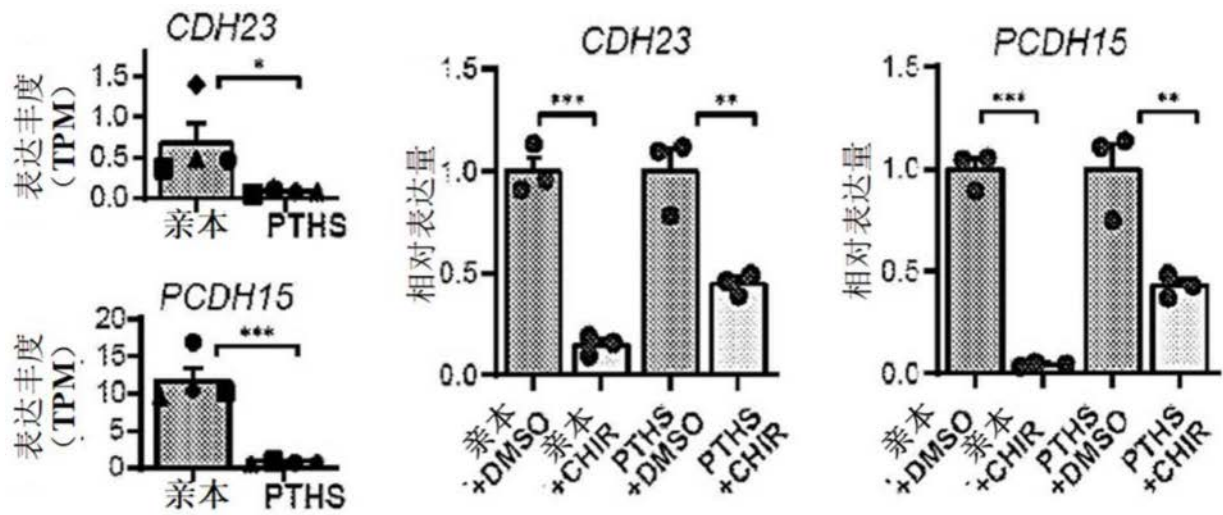


图16N

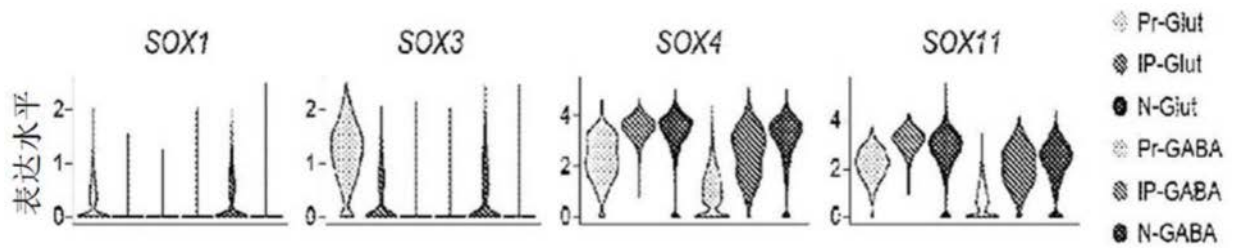


图17A

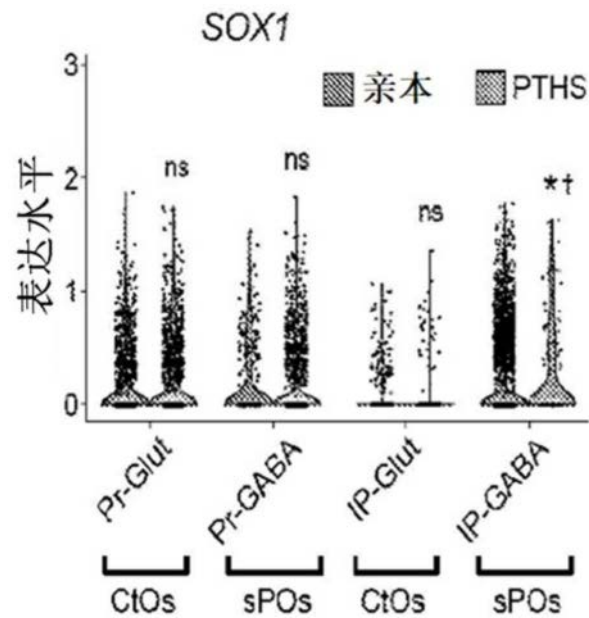


图17B

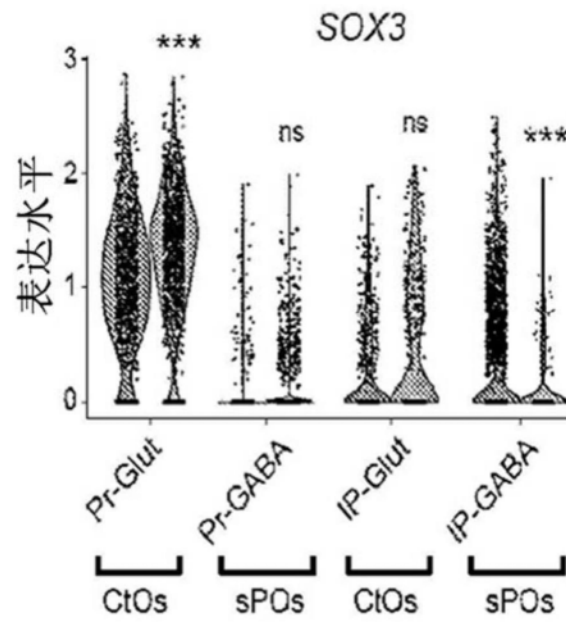


图17C

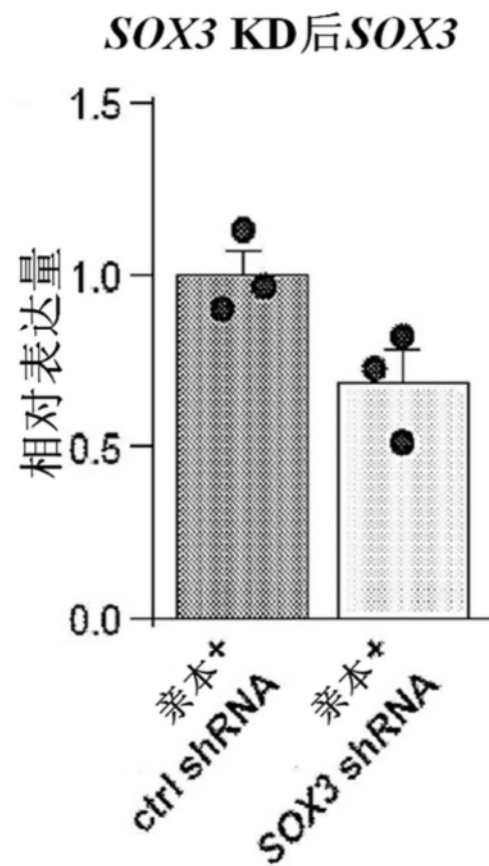


图17D

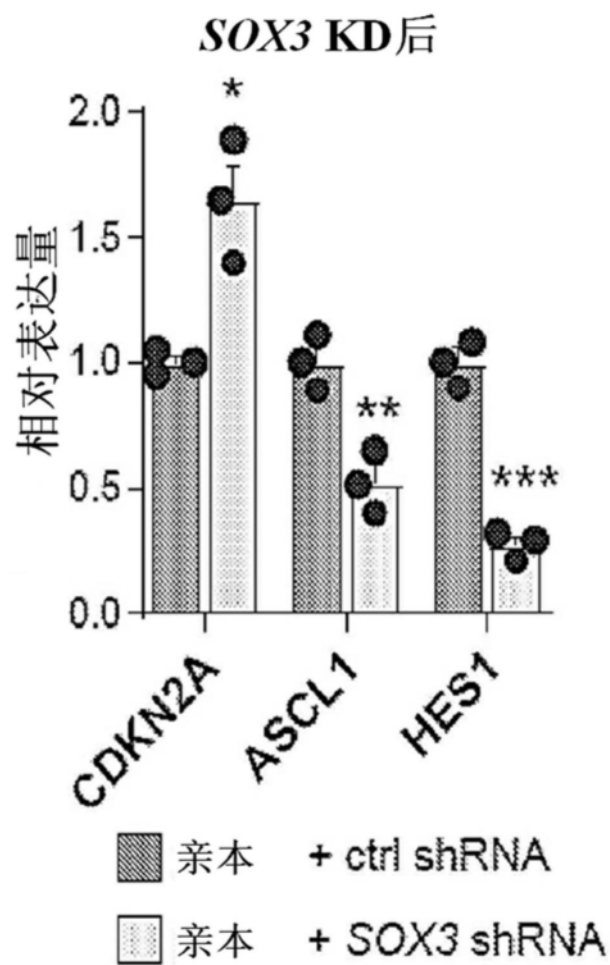


图17E

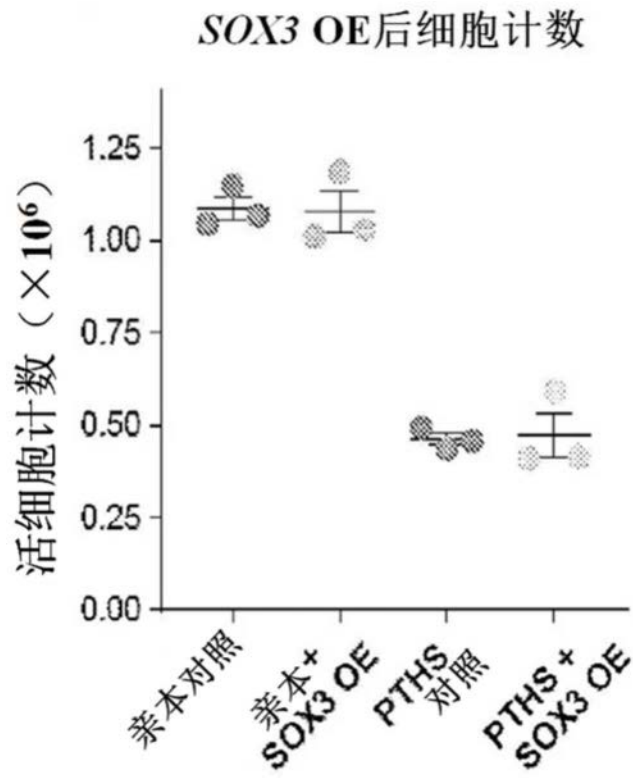


图17F

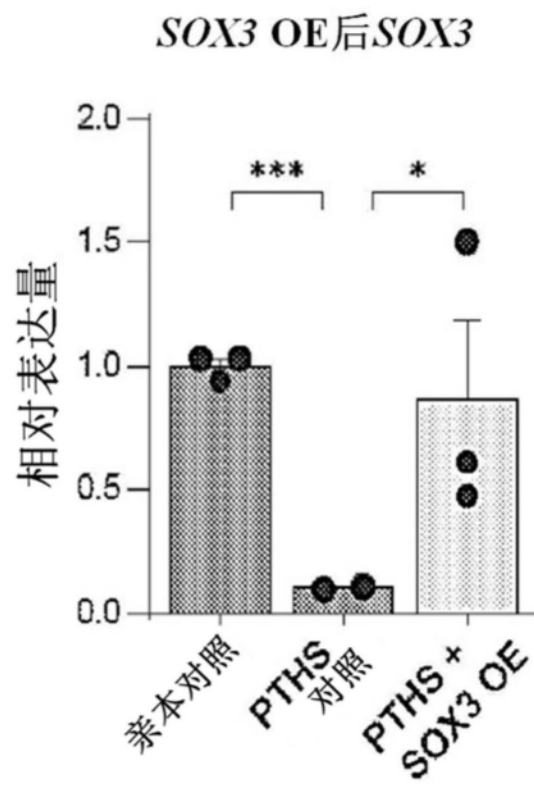


图17G

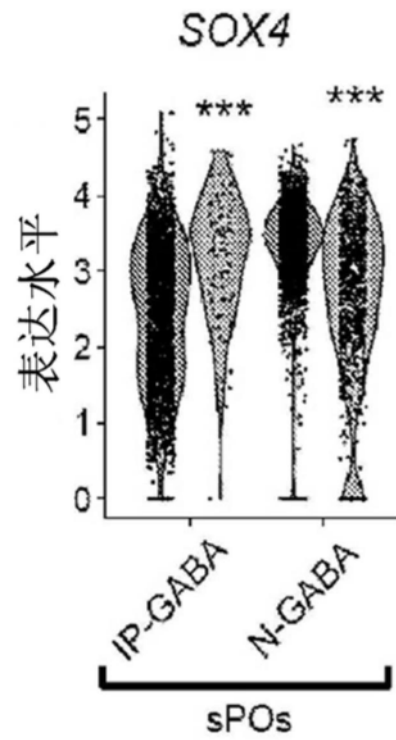


图17H

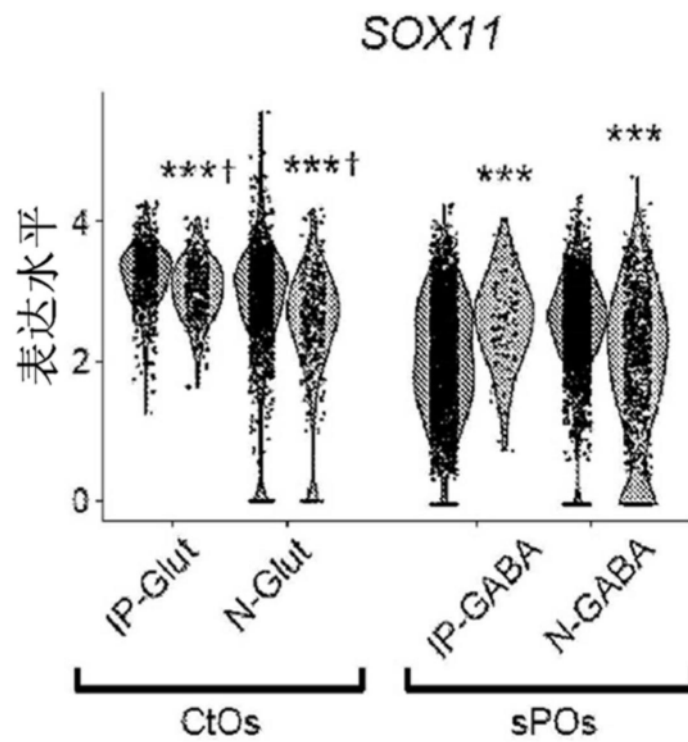


图17I

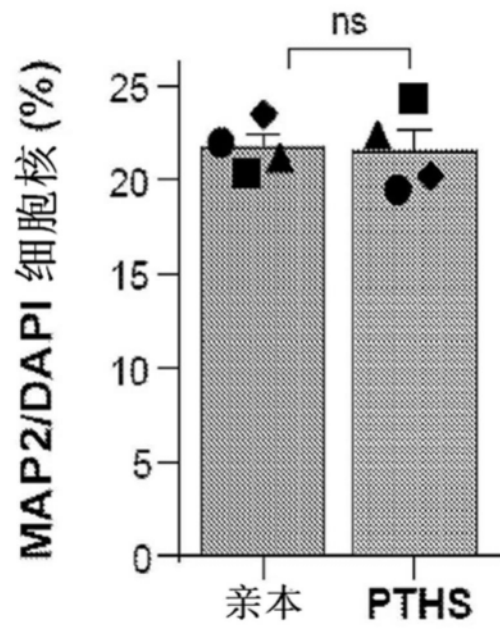


图17J

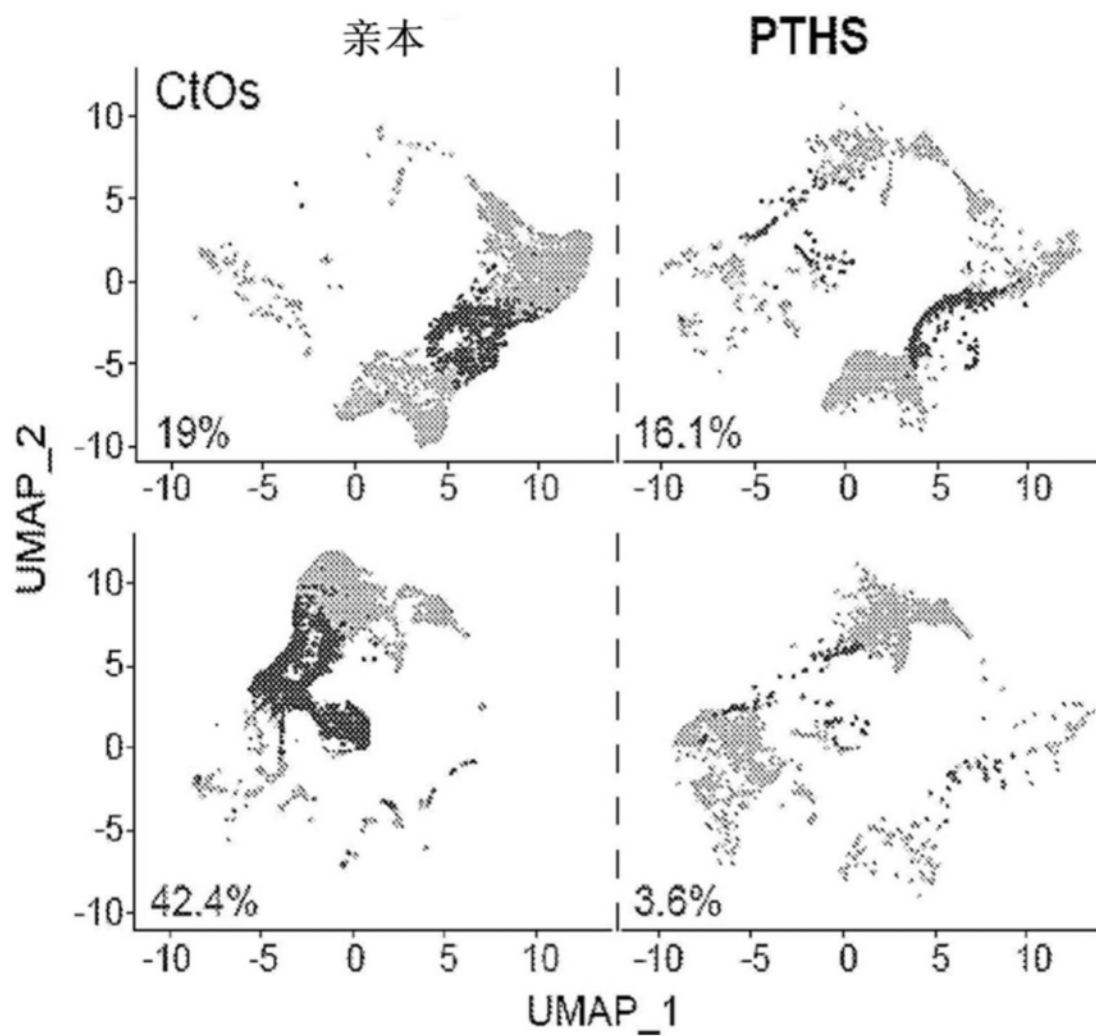


图17K

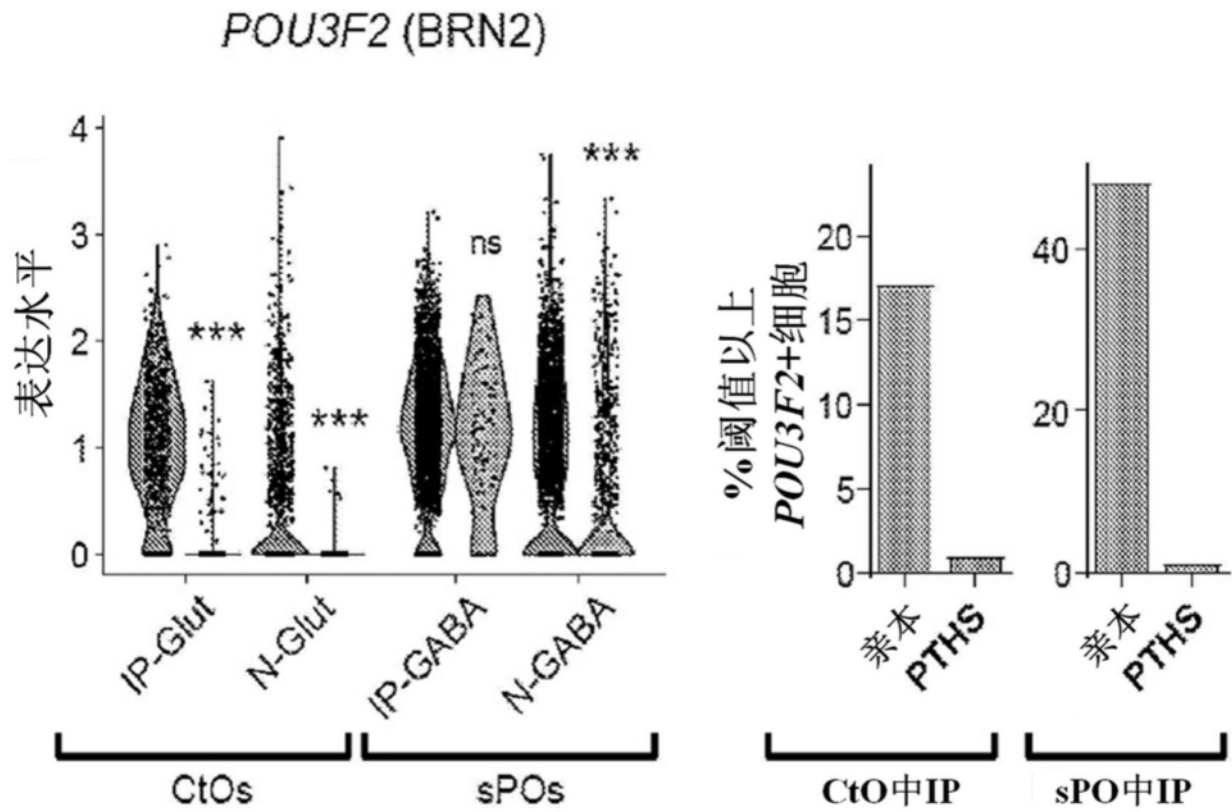


图17L

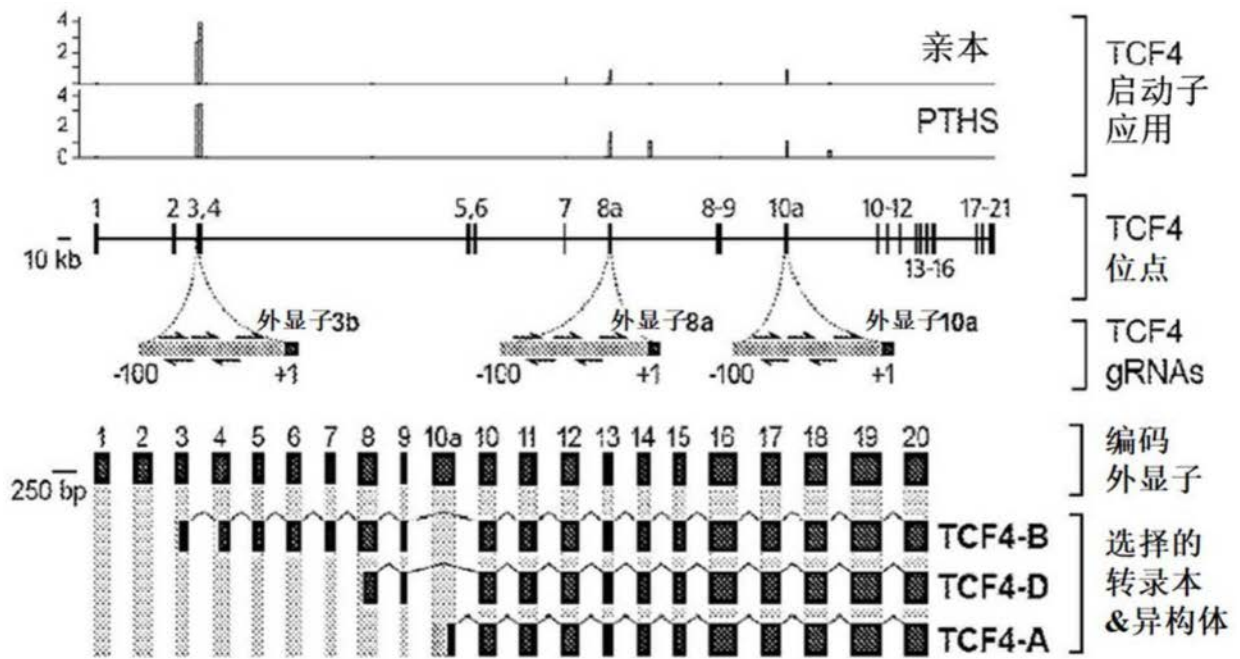


图18A

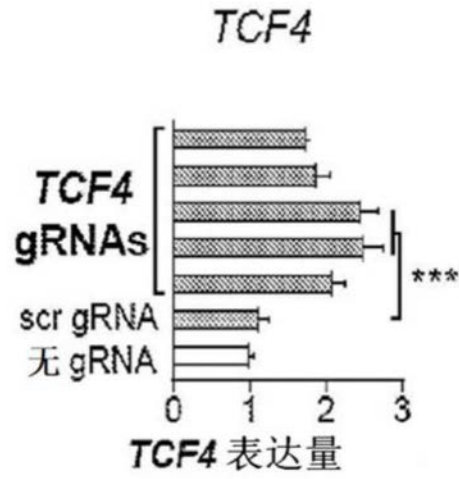


图18B

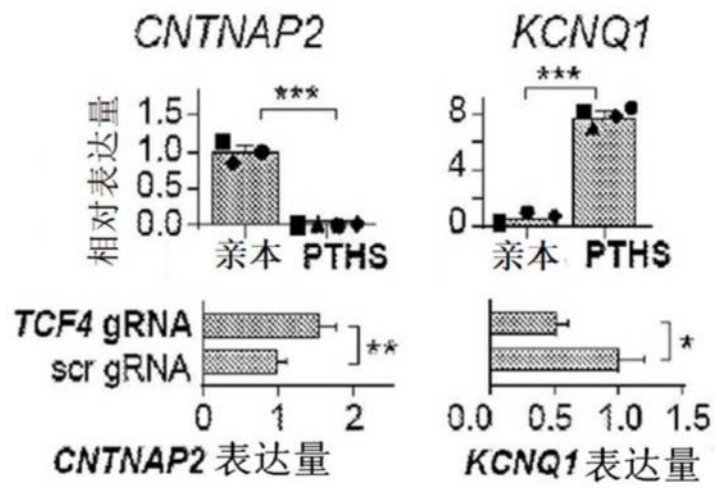


图18C

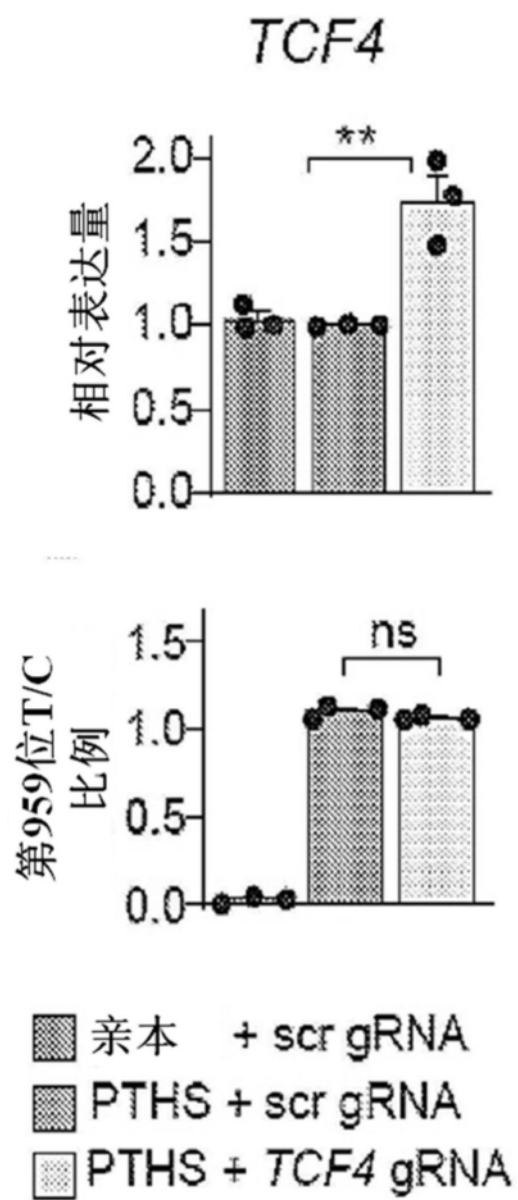


图18D

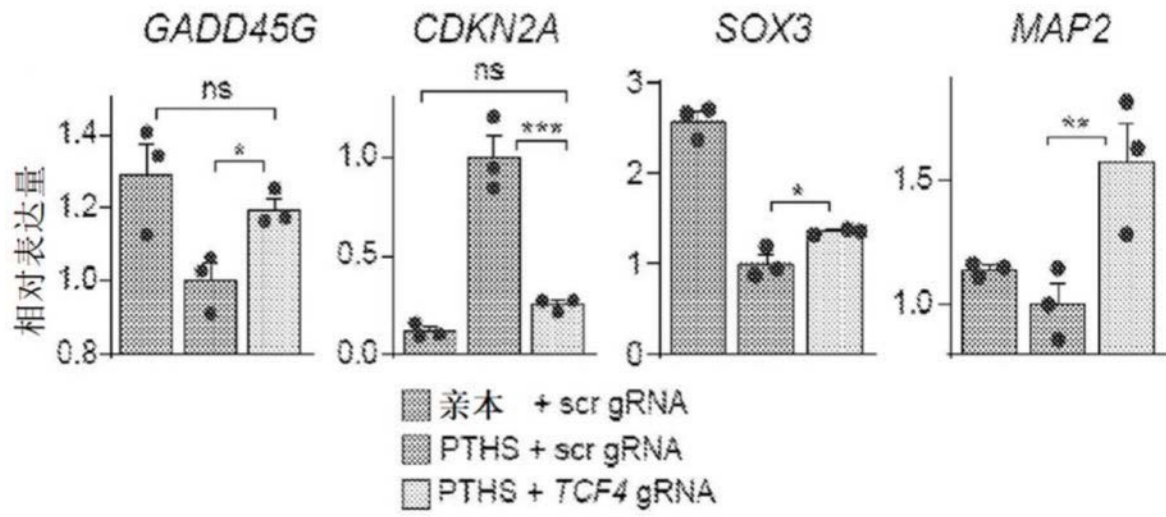


图18E

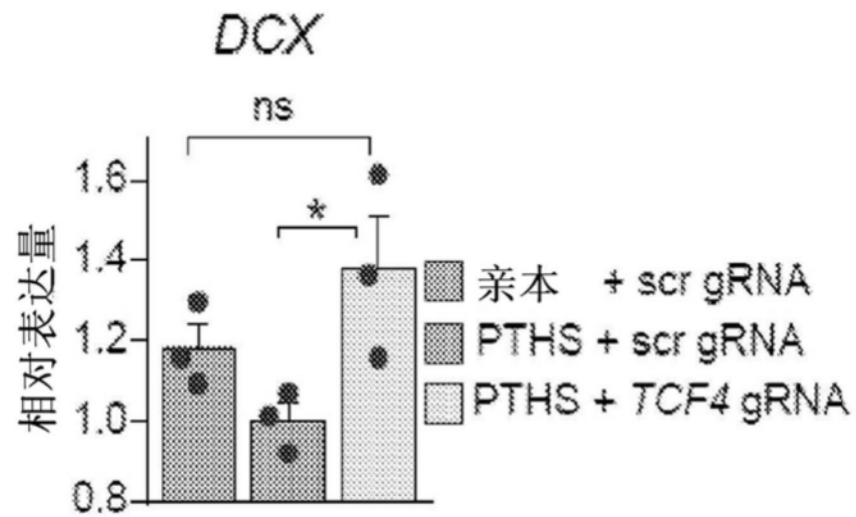


图18F

死后组织中 中*DCX*

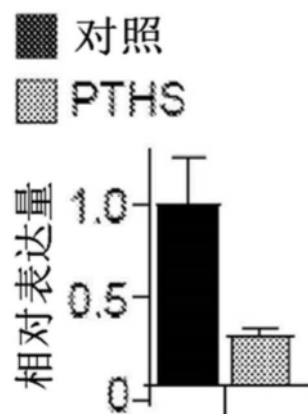


图18G

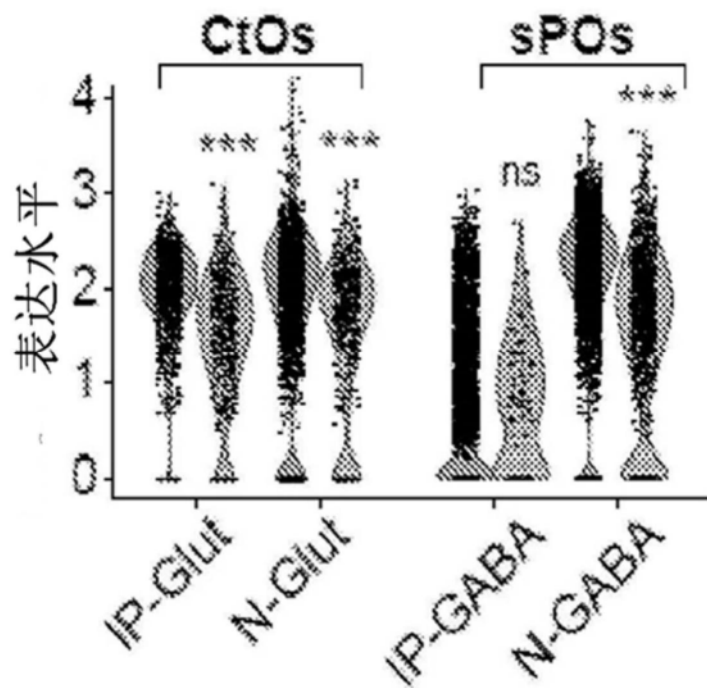


图18H

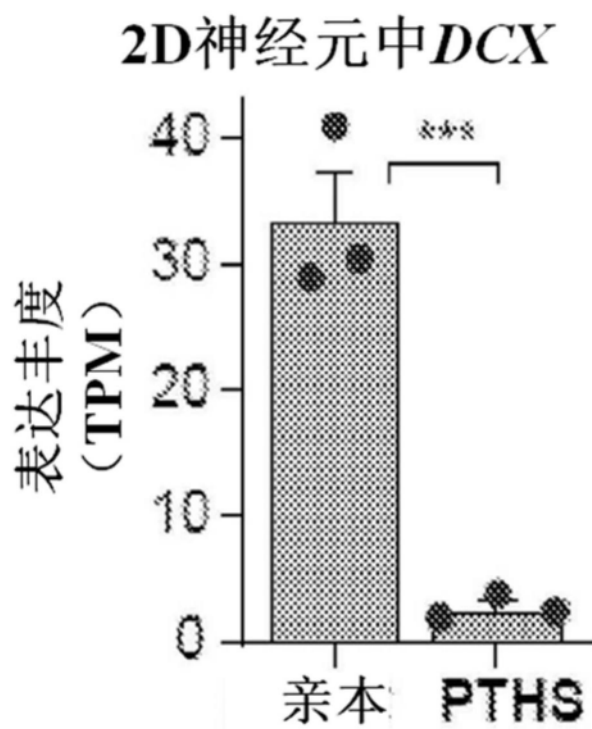


图18I

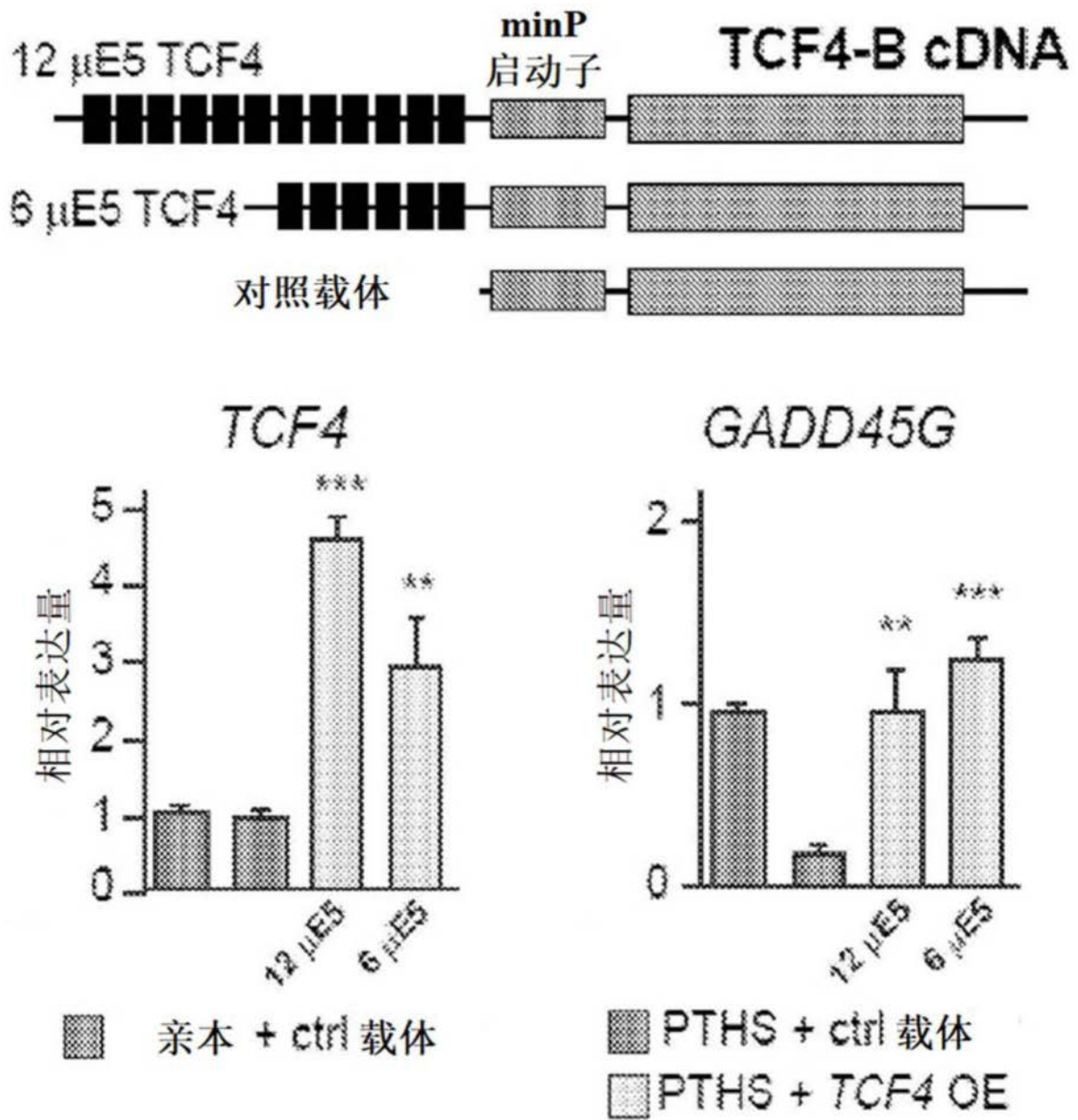


图18J

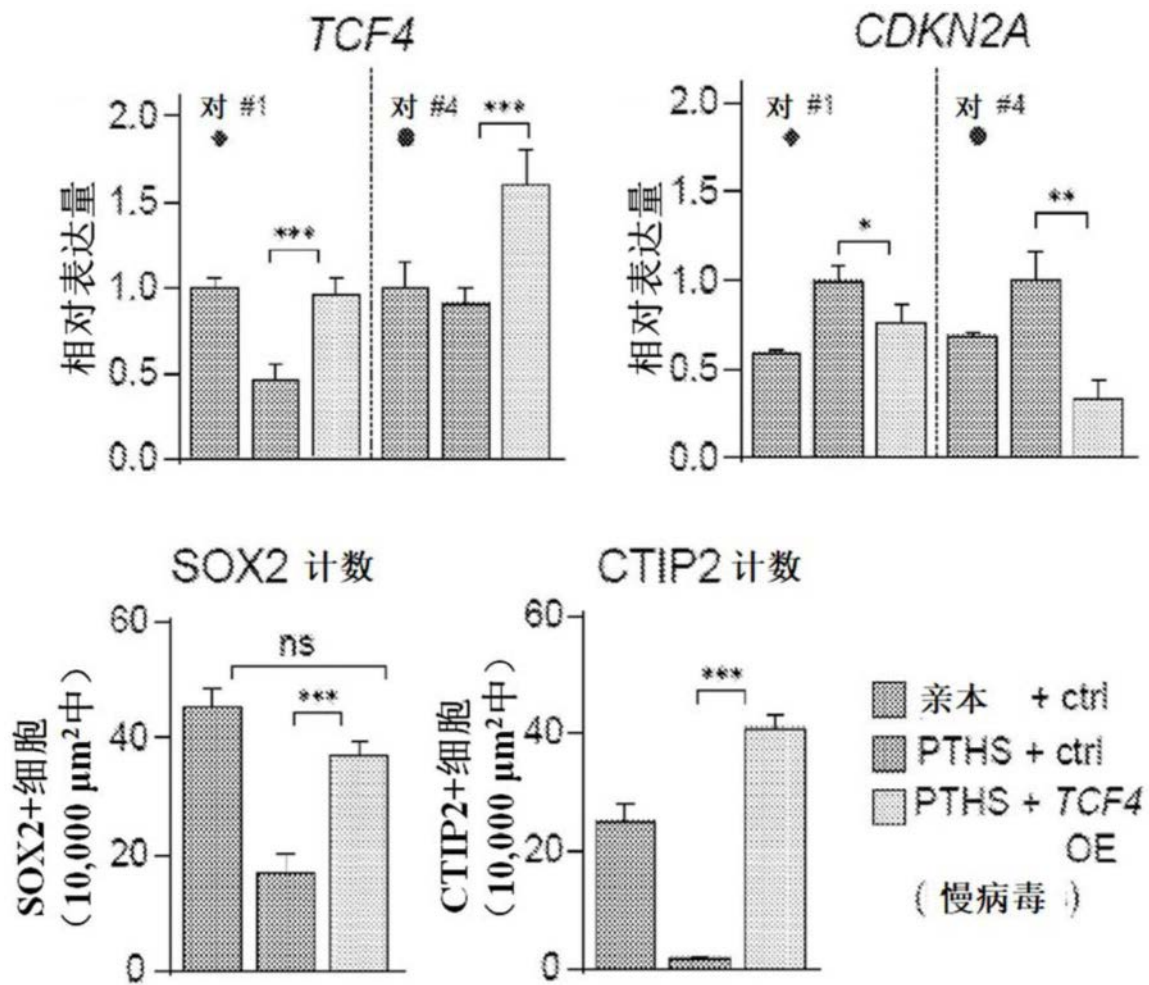


图18K

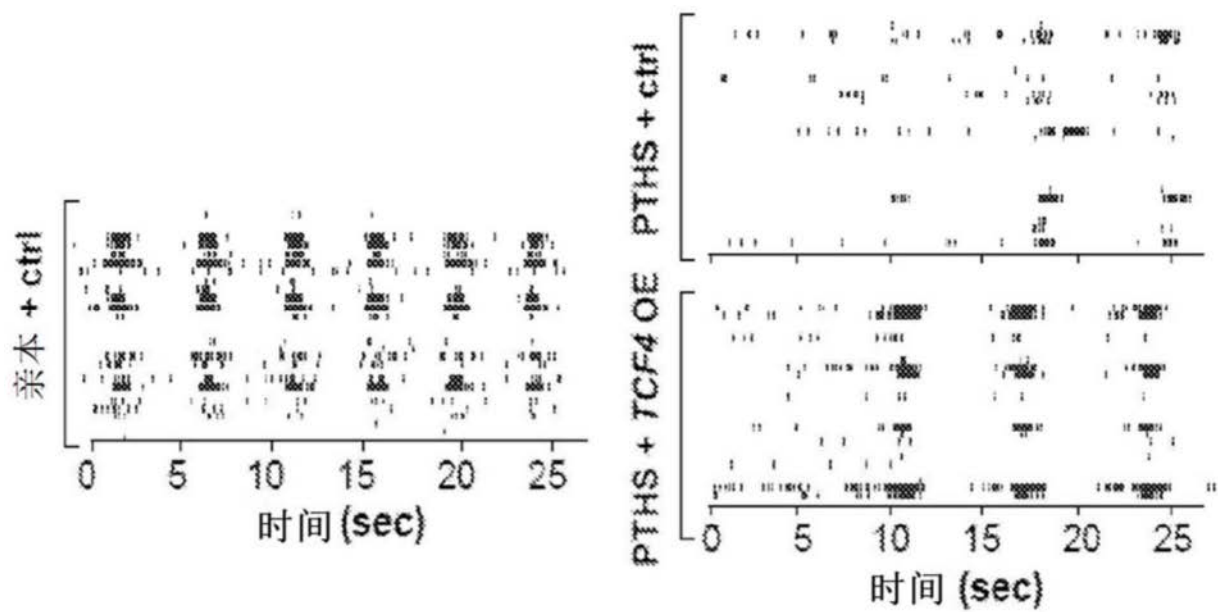


图18L

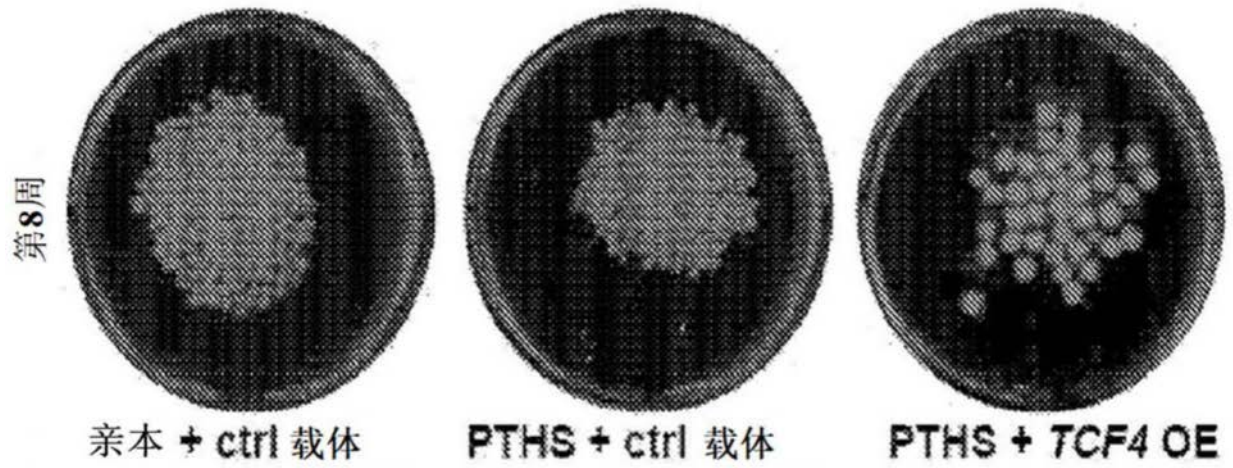


图18M

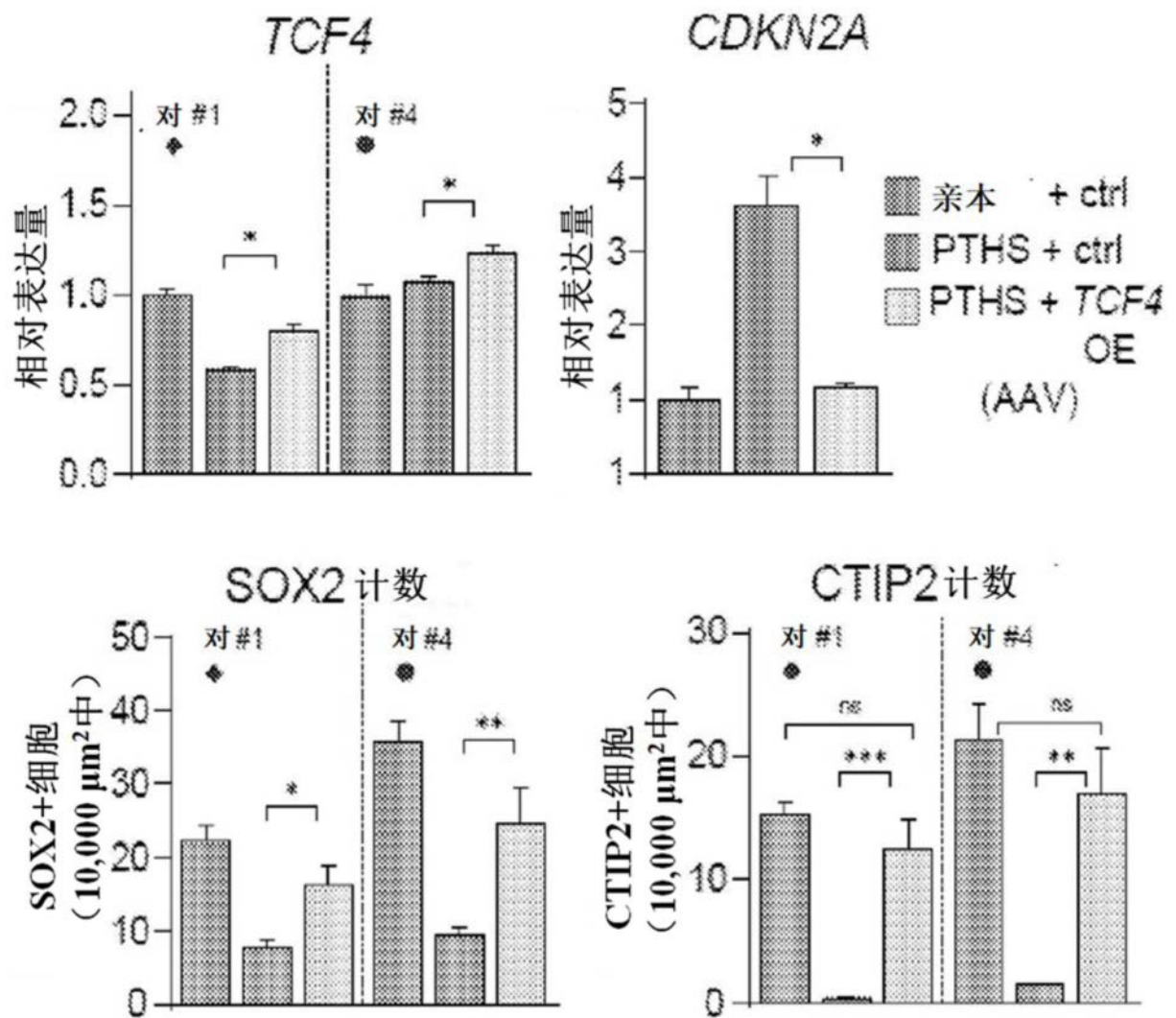


图18N

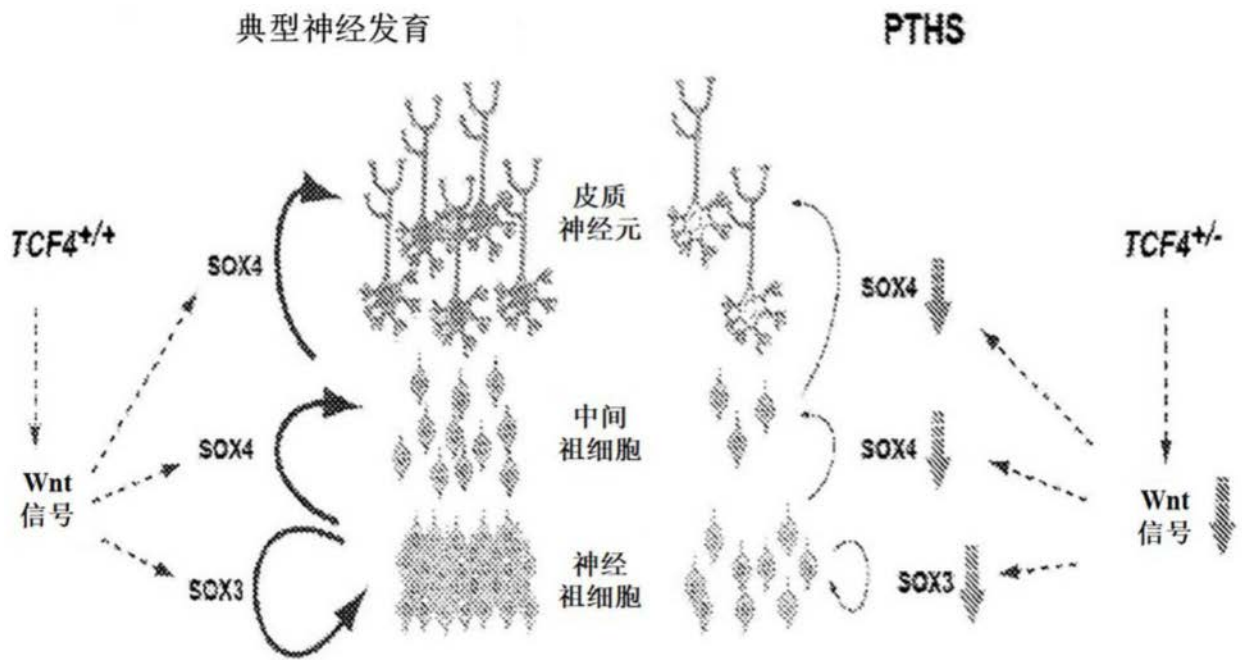


图180