



(12) 发明专利申请

(10) 申请公布号 CN 117795084 A

(43) 申请公布日 2024. 03. 29

(21) 申请号 202280054820.1

(22) 申请日 2022.07.14

(30) 优先权数据

63/221,611 2021.07.14 US

(85) PCT国际申请进入国家阶段日

2024.02.05

(86) PCT国际申请的申请数据

PCT/CA2022/051100 2022.07.14

(87) PCT国际申请的公布数据

W02023/000081 EN 2023.01.26

(71) 申请人 多伦多大学管理委员会

地址 加拿大

(72) 发明人 M·J·O·泰帕莱 N·阿莱拉苏尔

冷赫

(74) 专利代理机构 北京市铸成律师事务所

11313

专利代理师 陈艳娟 王艳波

(51) Int.Cl.

C12N 15/62 (2006.01)

C12N 9/14 (2006.01)

C12N 15/113 (2006.01)

C12N 15/861 (2006.01)

权利要求书2页 说明书60页

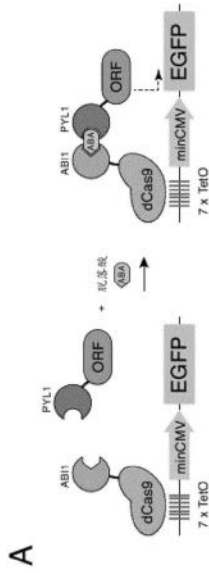
序列表(电子公布) 附图35页

(54) 发明名称

反式激活因子以及其方法和用途

(57) 摘要

一种异源转录激活因子,所述异源转录激活因子包含DNA靶向域,优选无催化活性的DNA靶向蛋白,例如CRISPR-Cas蛋白,以及包含至少一个本文所述的反式激活域或其功能变体的效应域。本文还提供了编码或表达所述转录激活因子的表达构建体、载体和细胞,以及用于靶基因的转录激活的系统和方法,以及用于其制备和使用的组合物、试剂盒和试剂。



1. 一种异源转录激活因子,所述转录激活因子包含:  
DNA靶向域,任选地无酶活性的CRISPR-CAS蛋白、锌指DNA结合域、tet阻遏物或转录激活因子样效应物(TALE)DNA结合域;和  
效应域,所述效应域包含:  
至少一种反式激活域(TAD)或其功能变体,所述TAD选自表2或表6、任选地表6中列出的TAD,或  
选自表1或表3中列出的TAD的至少两种TAD或其功能变体,优选地选自表4、或表5或表6中列出的TAD的至少一种TAD和/或其功能变体,  
其中所述DNA靶向域和所述效应域可操作地连接。
2. 如权利要求1所述的转录激活因子,其中所述效应域包含选自表1或表3中列出的TAD的至少三种或至少4种反式激活域或其功能变体。
3. 如权利要求1或权利要求2所述的转录激活因子,所述转录激活因子还包含至少一种相互作用组分。
4. 如权利要求1至3中任一项所述的转录激活因子,其中所述DNA靶向域和效应域是单一多肽的域。
5. 如权利要求3所述的转录激活因子,所述转录激活因子包含:  
包含所述DNA靶向域和第一相互作用组分的第一多肽,和  
包含效应域和第二相互作用组分的第二多肽,  
其中所述第一和第二相互作用组分在合适的条件下相互作用。
6. 如权利要求5所述的转录激活因子,其中所述第一和第二相互作用组分形成在诱导条件下相互作用的诱导型异源二聚体对,任选地ABI1和PYL1。
7. 如权利要求1至6中任一项所述的转录激活因子,其中所述DNA靶向域包含锌指域。
8. 如权利要求1至7中任一项所述的转录激活因子,其中所述效应域包含选自SEQ ID NO:103、104、105、106、167和185的TAD中的任一者的至少一种TAD,任选地选自SEQ ID NO:90、91、46、47、101-106、110、116-119、156、157、159、162、165、166、167和172的TAD中的任一者的至少一种TAD。
9. 如权利要求1至8中任一项所述的转录激活因子,所述转录激活因子还包含一种或多种核定位信号(NLS),任选地SV40 NLS。
10. 如权利要求1至9中任一项所述的转录激活因子,其中所述效应域包含SEQ ID NO:121、123、125、127、129、131、133、135、174、176、178、180或182,或与其中的TAD具有至少80%、85%、90%、95%或99%序列同一性的氨基酸序列。
11. 一种分离的核酸,所述核酸编码如权利要求1至10中任一项所述的转录激活因子或如权利要求1至10中任一项所述的效应域。
12. 一种表达构建体,所述表达构建体包含可操作地连接至一个或多个启动子和一个或多个转录终止位点的如权利要求11所述的核酸。
13. 一种载体,所述载体包含如权利要求11所述的核酸或如权利要求12所述的表达构建体,任选地其中所述载体是腺病毒或慢病毒载体。
14. 一种细胞,所述细胞包含如权利要求1至10中任一项所述的转录激活因子、如权利要求11所述的核酸、如权利要求12所述的表达构建体或如权利要求13所述的载体。

15. 一种转录激活系统,所述转录激活系统包含:

- a) 如权利要求1至10中任一项所述的异源转录激活因子,其中所述DNA靶向域包含CRISPR-Cas蛋白,和
- b) 至少一种gRNA。

16. 如权利要求15所述的转录激活系统,其中所述至少一种gRNA靶向基因的调节元件,任选地所述调节元件是启动子区域、增强子区域或远端调节位点。

17. 一种激活细胞中靶基因的转录的方法,所述方法包括:

- a) 将如权利要求1-10中任一项所述的转录激活因子、如权利要求11所述的核酸、如权利要求12所述的表达构建体或如权利要求13所述的载体引入所述细胞中;以及
- b) 在合适的条件下培养所述细胞以使得所述效应域激活所述靶基因的转录。

18. 如权利要求17所述的方法,其中所述DNA靶向域包含CRISPR-Cas蛋白,所述方法还包括将至少一种gRNA引入所述细胞中,以及在合适的条件下培养所述细胞以使得所述至少一种gRNA与所述CRISPR-Cas蛋白缔合,从而将所述转录激活因子引导至CRISPR靶位点。

19. 一种筛选测定,所述测定包括:

- a) 将如权利要求1至10中任一项所述的转录激活因子、如权利要求11所述的一种或多种核酸、如权利要求12所述的一种或多种表达构建体或如权利要求13所述的一种或多种载体引入多个细胞中,其中所述DNA靶向域包含CRISPR-Cas蛋白和多种gRNA;或将多种gRNA引入根据权利要求14所述的细胞的群中,其中所述DNA靶向域包含CRISPR-Cas蛋白;
- b) 培养所述多个细胞以使得所述一种或多种gRNA与所述CRISPR-Cas蛋白缔合并且将所述转录激活因子引导至CRISPR靶位点,以使得所述效应域激活靶基因的转录;
- c) 任选地用一定量的测试药物或毒素进行处理;
- d) 任选地培养所述多个细胞一段时间以允许gRNA丢失或富集;以及
- e) 收集所述多个细胞或其子集。

20. 如权利要求19所述的测定,其中所述方法还包括鉴定在所述多个细胞或其子集中过度表现或表现不足的一种或多种gRNA。

21. 一种组合物,所述组合物包含如权利要求1至10中任一项所述的转录激活因子、如权利要求11所述的核酸、如权利要求12所述的表达构建体、如权利要求13所述的载体或如权利要求14所述的细胞。

22. 一种试剂盒,所述试剂盒包括小瓶和如权利要求1至9中任一项所述的异源转录激活因子、如权利要求11所述的核酸、如权利要求12所述的表达构建体、如权利要求13所述的载体、如权利要求14所述的细胞或如权利要求21所述的组合物和任选的以下中的一者或多者:诱导剂、gRNA或gRNA表达构建体。

## 反式激活因子以及其方法和用途

[0001] 相关申请的交叉引用

[0002] 本公开要求2021年7月14日提交的美国专利申请第63/221,611号的优先权,其内容以引用的方式整体并入本文中。

### 技术领域

[0003] 本公开涉及用于转录激活的试剂和方法,并且特别涉及转录激活因子中的异源反式激活因子域用于靶向转录激活的用途。

### [0004] 引言

[0005] 蛋白质编码基因的转录通过以序列特异性方式结合DNA的转录因子(TF)、从启动子启动转录的RNA聚合酶II机制以及调节染色质结构并充当TF与RNA pol II之间的桥梁的多种染色质相关因子和复合物的协调相互作用来协调(Cramer,2019)。人类基因组编码参与转录调节各个阶段的数千种蛋白质,并且诸如ChIP-seq等方法的现成可用性揭示了不同条件下数百种因子的基因组结合位点(The ENCODE Project Consortium等人,2020)。与此同时,系统研究已经表征或推断了约四分之三的人类TF的DNA结合特异性(Badis等人,2009;Jolma等人,2013;Lambert等人,2018;Najafabadi等人,2015;Weirauch等人,2014)。类似地,相互作用蛋白质组学方法发现了许多染色质相关蛋白,并且表征了人类细胞中转录调节复合物的组成(Gao等人,2012;Huttlin等人,2020;Lambert等人,2019;Li等人,2015;Marcon等人,2014;Mashtalir等人,2018)。

[0006] 然而,由于这些方法提供的因果见解有限,TF和染色质相关因子是否以及如何促进转录激活或阻遏(或通过其它方式调节染色质状态)仍然很大程度上未知。例如,ChIP-seq实验中观察到的绝大多数基因组结合位点与转录事件没有因果关系。也就是说,给定转录调节因子的敲低或敲除不会影响所述调节因子所结合的大多数基因的转录。另一方面,基于序列的转录调节因子注释一直具有挑战性,因为大多数转录效应功能是由简并线性基序而不是折叠和保守的蛋白质域编码的(Arnold等人,2018;Erijman等人,2020;Sigler,1988;Staller等人,2021)。

[0007] 人工募集,也称为激活因子旁路,是在限定背景下表征不同蛋白质转录效应的有效方法(Ptashne和Gann,1997;Sadowski等人,1988)。在这种方法中,通过将蛋白质融合至特征明确的TF(例如Gal4或TetR)的DNA结合域,将蛋白质或其片段异位募集至报告基因。限定背景减轻了内源基因调节带来的挑战,在内源基因调节中,多种因子协同结合调节元件,阻碍了因果推理。传统上,人工募集已用于鉴定个体转录调节因子中的转录激活因子或反式激活域(TAD)(Ptashne和Gann,1997)。然而,最近的研究通过将果蝇或酵母中的大量调节因子单独系连至报告基因来表征其转录效应(Keung等人,2014;Stampfel等人,2015)。由于阵列格式的可扩展性有限,这些研究侧重于已知的调节因子,而不是潜在的新因子。此外,经典模式生物缺乏若干对哺乳动物基因表达至关重要的调节机制和层次,例如增强子(酵母)或DNA甲基化(酵母和果蝇)。最近,Tycko和其同事实施了一种无偏的合并筛选策略来表征注释的人类蛋白质域的转录激活潜力(Tycko等人,2020)。此研究强调了无偏方法在鉴定

新型转录调节因子方面的价值,例如变异KRAB域在转录激活而不是阻遏中的意想不到的作用。然而,由于大多数转录激活域是由无序区域编码(Arnold等人,2018;Dyson和Wright,2005),因此以域为中心的筛选可能会遗漏很大一部分反式激活因子。

[0008] 因此,尽管人们对转录调节复合物的组成和其基因组结合模式的了解有所增加,但仍缺乏对TF和多种染色质相关因子引起的下游效应的完整了解。

## 发明内容

[0009] 如本文所述,本发明人建立了一个平台来以无偏的方式系统地鉴定和表征人类蛋白质的转录调节潜力。通过以合并形式筛选13,000多种蛋白质,鉴定了数百种有效的激活因子,其中许多激活因子先前注释不充分。在命中当中系统地发现了反式激活域,包括一些不符合激活域的规范“酸性斑点”模型的域。此外,相互作用蛋白质组学与化学抑制剂相结合,描述了新型和已知转录激活因子的辅因子特异性,强调了即使具有几乎相同的DNA结合特异性的高度相关的TF也可通过不同的辅因子复合物激活转录。

[0010] 本发明人在本文描述了人类细胞中转录激活因子的首次系统筛选。鉴定了数百种转录激活因子,如所预期,其富含序列特异性转录因子和其它染色质相关蛋白。

[0011] 本文显示的结果表明,只有非常有限数量的TF是强转录激活因子。这在已知的人类TF的DNA结合特异性和染色质占据的背景下是有意义的(Jolma等人,2013;The ENCODE Project Consortium等人,2020;Yan等人,2013)。大多数TF识别短(约6-10bp)基序,并与基因组中的数千个位点相关。然而,大多数结合事件与转录输出无关。强转录激活域与DNA结合域的有限序列特异性相结合,会导致大量基因的虚假激活,可能会阻碍细胞适应性。本文鉴定的许多最强激活因子不是DNA结合转录因子,而是其它染色质相关蛋白。

[0012] 新颖且高效的人类反式激活域的发现也具有治疗意义。源自病毒蛋白的组分,例如VP16反式激活域或三联VPR激活因子可引发体内免疫反应并导致临床不良反应。从完全人类组分中设计合成转录调节因子预计将在治疗应用中具有优势(Israni等人,2021)。此外,本文还显示,组合来自多种不同人类蛋白质的激活域可产生“超级激活因子”,其即使在基因组的高度压缩区域中也能够稳健地上调基因。

[0013] 一个方面包括一种异源转录激活因子,其包含:

[0014] DNA靶向域,任选地无酶活性的CRISPR-CAS蛋白、锌指DNA结合域、tet阻遏物或转录激活因子样效应物(TALE)DNA结合域;和

[0015] 效应域,其包含:

[0016] 至少一种反式激活域(TAD)或其功能变体,所述TAD选自表1至6中的任一者、任选地表2或表6中列出的TAD,或

[0017] 选自表1-6中的任一者、任选地表1或表3中列出的TAD的至少两种TAD或其任一者的功能变体,优选地选自表4或表5或表6中列出的TAD的至少一种TAD或其功能变体,

[0018] 其中DNA靶向域和效应域可操作地连接。

[0019] 一个方面包括编码本文所述的效应域的分离的核酸

[0020] 一个方面包括编码本文所述的异源转录激活因子的分离的核酸。

[0021] 一个方面包括表达构建体,其包含可操作地连接至一个或多个启动子和一个或多个转录终止位点的本文所述的核酸。

[0022] 一个方面包括包含本文所述的核酸或表达构建体的载体,任选地其中该载体是腺病毒或慢病毒载体。

[0023] 一个方面包括包含本文所述的转录激活因子、核酸、表达构建体或载体的细胞。

[0024] 一个方面包括转录激活系统,其包含:本文所述的异源转录激活因子,其中DNA靶向域包含CRISPR-Cas蛋白和至少一种gRNA。

[0025] 一个方面包括激活细胞中靶基因的转录的方法,所述方法包括:a)将本文所述的转录激活因子、核酸、表达构建体或载体引入至细胞中;以及b)在合适的条件下培养细胞,以使得效应域激活靶基因的转录。

[0026] 一个方面包括筛选方法,所述方法包括:a)将转录激活因子、一种或多种核酸、一种或多种表达构建体或一种或多种本文所述的载体引入至多个细胞中,其中DNA靶向域包含CRISPR-Cas蛋白和多种gRNA;或将多种gRNA引入至本文所述的细胞群中,其中DNA靶向域包含CRISPR-Cas蛋白;b)培养所述多个细胞,以使得一种或多种gRNA与CRISPR-Cas蛋白缔合并将转录激活因子引导至CRISPR靶位点,以使得效应域激活靶基因的转录;c)任选地用一定量的测试药物或毒素进行处理;d)任选地培养所述多个细胞一段时间以允许gRNA丢失或富集;以及e)收集所述多个细胞或其子集。

[0027] 一个方面包括包含本文所述的转录激活因子、核酸、表达构建体、载体或细胞的组合物。

[0028] 一个方面包括试剂盒,其包括小瓶和本文所述的异源转录激活因子、核酸、表达构建体、载体、细胞或组合物,以及任选的以下中的一者或多者:诱导剂、gRNA或gRNA表达构建体。

[0029] 前面的章节仅作为实例提供,并且不意图限制本公开和所附权利要求的范围。根据本发明的权利要求、描述和实施例,本领域普通技术人员将理解与本公开的组合物和方法相关的附加目的和优点。例如,本公开的各个方面和实施方案可以以多种组合使用,本说明书明确地考虑了所有这些组合。这些额外的优势、目的和实施方案明确地包括在本公开的范围内。本文中用于说明本公开背景的出版物和其它材料,以及在特定情况下用于提供与实践相关的附加细节的出版物和其它材料通过引用并入,并且为了方便起见,在所附的参考文献章节中列出。

## 附图说明

[0030] 本公开的另外的目的、特征和优点从结合示出本公开的说明性实施方案的附图进行的以下详细描述中显而易见,在附图中:

[0031] 图1描绘了转录激活因子的合并ORFeome筛选。A,显示了用于表征人类细胞中的转录激活因子的化学诱导的二聚化系统的示意图。B,显示了当所示构建体转染至HEK293T报告细胞并用脱落酸或DMSO处理48小时时,高GFP细胞的百分比。C,转录激活因子的合并ORFeome筛选的概述。D,48小时ABA处理后,在合并ORFeome筛选中高GFP细胞的富集。E,与未分选的ORFeome相比,高GFP池中ORF的富集。F,阳性筛选命中当中基因本体类别的富集。G,阳性筛选命中当中InterPro域的富集。H,筛选命中当中CORUM复合物的富集。

[0032] 图2显示了转录因子家族的转录活性。以阵列方式单独测试转录调节因子的报告基因激活情况。DNA结合特异性(显示为序列标识)来自CisBP(Weirauch等人,2014)。星号表

示统计上显著的激活因子 ( $FDR < 0.05$ )。A, 同源盒家族蛋白B, 叉头盒蛋白, C, Kruppel样因子, D, SRY相关HMG盒 (SOX) 蛋白, E, 多梳家族环指 (PCGF) 蛋白。规范 (cPRC1) 和非规范 (ncPRC1) 复合物的组成如右侧所示。F, Spy1/RING0家族蛋白。统计显著性通过假定方差相等的不配对双尾t检验计算, 并使用Benjamini、Krieger和Yekutieli (Benjamini等人, 2006) 的错误发现率 (FDR) 方法校正多个假设。

[0033] 图3显示了利用TAD-seq对人类蛋白质中的反式激活域进行的系统发现。A, TAD-seq合并测定的概述。B, 已知反式激活域的实例。域组织显示在顶部。TAD-seq图显示了高GFP群体或中等GFP群体中RNAseq读段的富集倍数。每个圆圈显示60aa平铺 (tile) 的中点 (第30个氨基酸)。实心圆圈表示统计上显著的命中。灰色框表示先前描述的反式激活域。C, 新型反式激活域的实例。标签如图板B中所示。D, HOXA2的反式激活域的位置。E, HOXA2中鉴定的激活片段的序列。激活片段以粗体显示。在中等GFP群体中富集的所有三个片段共有的区域表示为激活片段的重叠。触角足样六肽序列的位置被表示为六肽。F, YAF2中反式激活域的位置。G, 与RYBP (PDB 3IXS) 的YAF2\_RYBP域和CBX7 (PDB 3GS2) 的CBX\_C域结合的RING1B RAWUL域的晶体结构。H, 分别测试所示蛋白质的YAF2\_RYBP域和CBX\_C域的转录活性。星号表示统计上显著的激活因子 ( $FDR < 0.05$ )。统计显著性通过假定方差相等的不配对双尾t检验计算, 并使用Benjamini、Krieger和Yekutieli (Benjamini等人, 2006) 的错误发现率 (FDR) 方法校正多个假设。右侧, 域对RING1B的体外亲和力 (Wang等人, 2008, 2010)。使用不配对双尾t检验计算统计显著性。

[0034] 图4显示了转录激活因子的辅因子特异性。A, 用BioID2鉴定指定转录调节因子的邻近配偶体。所选辅因子复合物的富集显示为热图。将平均光谱计数 (n+1) 标准化为EGFP和Nanoluc诱饵的背景光谱计数 (n+1)。B, 基于AP-MS研究的激活叉头转录因子的相互作用模式 (Li等人, 2015)。光谱计数按照图板A进行标准化。C, 左侧, A-485抑制p300/CBP对83个转录调节因子活性的影响。已知的p300相互作用因子显示为实心深色圆圈, 已知的NuA4相互作用因子显示为轮廓圆圈。右侧, p300相互作用因子比其它转录调节因子更容易受到A-485处理的影响。使用邓尼特氏多重检验校正通过单向方差分析计算统计显著性。D, 左侧, JQ1抑制BET溴域蛋白对83个转录调节因子活性的影响。已知的p300相互作用因子显示为实心深色圆圈, 已知的NuA4相互作用因子显示为轮廓圆圈。右侧, p300相互作用因子比其它转录调节因子更容易受到A-485处理的影响。使用邓尼特氏多重检验校正通过单向方差分析计算统计显著性。E, 转录激活因子基于其对不同抑制剂的敏感性的聚类。使用具有平均连锁的欧几里德距离 (Euclidian distance) 进行聚类。显示了富含已知CBP/p300相互作用因子 (实心圆圈) 和NuA4相互作用因子杂交的簇。

[0035] 图5显示了SRF-C3orf62融合物产生促进SRF/MRTF靶基因的表达的有效p300依赖性转录激活因子。A, 低级别子宫内膜间质肉瘤中发现的JAZF1-SUZ12融合物的示意图。指示了由TAD-seq鉴定的反式激活域。B, 肌纤维瘤/肌周细胞瘤中发现的SRF-C3orf62融合物的示意图。指示了由TAD-seq鉴定的反式激活域。C, 使用BioID2鉴定了JAZF1、SUZ12和JAZF1-SUZ12邻近相互作用因子。JAZF1-SUZ12与PRC2组分和NuA4组分相互作用, 生成超级复合物。D, 使用BioID鉴定了SRF、C3orf62、C3orf62-Cterm和SRF-C3orf62邻近相互作用因子。SRF-C3orf62融合物与CBP和p300相互作用。E, SRF-C3orf62是一种有效的转录激活因子。单独测试了指定PYL1融合物的基因组整合报告基因的激活情况。F, SRF-C3orf62在不存在辅因子

的情况下激活血清反应元件报告基因的表达。将C末端标记有3xFLAG-V5的指定构建体与SRE-Firefly荧光素酶报告基因和组成型Nanoluc报告基因共转染至NIH3T3细胞中。测量相对荧光素酶活性以评估每个构建体的活性。G, 稳定表达多西环素诱导型SRF-C3orf62-GFP或Nanoluc-GFP的NIH3T3细胞用多西环素处理46小时, 其中最后22小时在低血清条件(0.5%FCS)下。通过RNA-seq分析基因表达模式。显著上调(深色圆圈, 顶部)和下调(中等圆圈, 底部)基因(绝对log2倍数变化>1, FDR<0.05)。指出了SRF/MRTF(顶部, 标记圆圈)和SRF/TCF(底部, 标记圆圈)的明确特征目标。H, SRF-C3orf62-GFP表达细胞与Nanoluc-GFP表达细胞相比的基因集富集分析。

[0036] 图6显示了转录激活因子的ORFeome筛选。A, 合并质粒DNA和感染细胞中跨ORFeome的测序读段的分布。B, 质粒池和感染细胞中ORF大小的分布。C, 转录活性取决于脱落酸处理。用100 $\mu$ M ABA处理ORFeome-PYL1感染的细胞, 并随时间推移通过流式细胞仪测量高GFP细胞的分数。D, ABA浓度对转录活性的影响。用指定构建体转染的报告细胞用增加量的ABA处理48小时。E, 在不表达ORFeome-PYL1文库的ABA处理细胞中未观察到高GFP群体。F, 相互作用中心在激活筛选命中当中的富集。G, 酵母双杂交自动激活因子在激活筛选命中当中的富集。H, 激活筛选中鉴定的转录激活因子的单独验证。将指定构建体转染至报告细胞系中, 并在用ABA处理48小时后通过流式细胞术测量高GFP细胞分数。星号表示高GFP群体中统计上显著的ABA依赖性增加(FDR<5%)。统计显著性通过假定方差相等的不配对双尾t检验计算, 并使用Benjamini、Krieger和Yekutieli (Benjamini等人, 2006)的错误发现率(FDR)方法校正多个假设。

[0037] 图7显示转录因子和染色质相关蛋白家族的转录活性。以阵列方式单独测试转录调节因子的报告基因激活情况。DNA结合特异性(显示为序列标识)来自CisBP (Weirauch等人, 2014)。星号表示统计上显著的激活因子(FDR<5%)。统计显著性通过假定方差相等的不配对双尾t检验计算, 并使用Benjamini、Krieger和Yekutieli (Benjamini等人, 2006)的错误发现率(FDR)方法校正多个假设。A, ETS家族TF, B, 无调性相关bHLH因子, C, 肌源性因子, D, Twist/Hand bHLH因子, E, 酪蛋白激酶, F, 介体组分, G, 在初级激活筛选中鉴定的转录因子富含当异位表达时, 可重新编程人类iPS细胞(右)或小鼠ES细胞(左)的因子(Ng等人, 2021; Theodorou等人, 2009)。使用威尔科克森(Wilcoxon)秩和检验(左)或费歇尔(Fisher)精确检验(右)计算统计显著性。

[0038] 图8显示了转录因子的差异活性不能用表达水平来解释。A, 测量报告基因表达和效应蛋白表达的转录激活测定的示意图。B, Kruppel样因子的反式激活(左)与通过RFP荧光测量的每个因子的表达水平(右)的比较。背景RFP强度显示为虚线。C, 叉头TF活性。D, 同源域TF活性, E, SOX TF活性。

[0039] 图9显示了TAD-seq鉴定反式激活域。A, 用ABA向报告基因募集60aa片段48小时后, 通过流式细胞术评估高和中等GFP群体。高GFP和中等GFP细胞通过FACS进行分选, 并通过下一代测序鉴定池中富集的ORF。B, 与文库中的非活性片段相比, 鉴定的反式激活片段中氨基酸的富集和消耗。以粗体显示的氨基酸在统计上显著富集或耗尽。C, 活性片段中的预测反式激活域的富集。9aaTAD通过9aaTAD预测工具(<https://www.med.muni.cz/9aaTAD/>)使用“适度严格模式”进行预测。分析中仅考虑100%置信匹配。ADpred算法描述于(Erijman等人, 2020)中。用费歇尔精确检验计算统计显著性。D, 活性片段中的预测TAD比非活性片段中

的预测TAD长。统计显著性通过假定方差相等的双尾t检验来计算。E,片段在高GFP池与中等GFP池中的富集。显著的命中显示为轮廓圆圈。F,反式激活片段的单独验证。指定TAD与PYL1融合并以阵列形式转染至报告细胞中。星号表示统计上显著的激活因子(FDR<5%)。统计显著性通过假定方差相等的不配对双尾t检验计算,并使用Benjamini、Krieger和Yekutieli (Benjamini等人,2006)的错误发现率(FDR)方法校正多个假设。

[0040] 图10显示了利用TAD-seq对人类蛋白质中的反式激活域进行的系统发现。A,已知反式激活域的实例。域组织显示在顶部。TAD-seq图显示了高GFP群体(深色圆圈)或中等GFP群体(浅色圆圈)中RNAseq读段的富集倍数。每个圆圈显示60aa平铺的中点(第30个氨基酸)。实心圆圈表示统计上显著的命中。灰色框表示先前描述的反式激活域。B,新型反式激活域的实例。标签如图板A中所示。

[0041] 图11显示了SPDYE4和YAF2的反式激活域。A,五种Spy1/RING0家族蛋白的比对:SPDYE4(SEQ ID NO:142);SPDYE1(SEQ ID NO:143);SPDYE7P(SEQ ID NO:144);SPDYE2(SEQ ID NO:145);和SPDYC(SEQ ID NO:146)。通过TAD-seq筛选的四个片段(SPDYE4-2(SEQ ID NO:154);SPDYE4-3(SEQ ID NO:90);SPDYE4-4(SEQ ID NO:91);和SPDYE4-5(SEQ ID NO:155))显示于比对方,且活性片段以粗体指示。虚线框指示推断的最小激活域。应注意,最小激活域区域在SPDYC中并不保守,SPDYC是唯一不激活转录的Spy1/RING0家族成员。B,YAF2(SEQ ID NO:147)和RYBP(SEQ ID NO:148)的YAF2\_RYBP域以及CBX家族蛋白(SEQ ID NO:149-153)的CBX\_C域的比对。两个β片层的位置用箭头指示。

[0042] 图12显示转录激活因子的相互作用网络。A,转录激活因子的BioID2网络。诱饵蛋白(例如FAM90A1、SPDYE4、SS18L2、SOX7等)显示为浅灰色矩形。指示了BAF复合物成员、p300/CBP、NuA4复合物成员、介体组分和TFIID组分。边缘的宽度指示两次重复的平均光谱计数。为了清楚起见,从该可视化中删除了两种高度连接的猎物蛋白(ZNF518A和ZNF518B)。B,转录激活因子的AP-MS网络。标记如图板A中所示。为了清楚起见,从该可视化中删除了九种高度连接的猎物蛋白(APEH、ACTC1、ALDH1L1、PSDM4、LSS、FLII、PACSIN2、RPS3A、QPCTL)。

[0043] 图13显示了小分子抑制剂对83种转录调节因子的转录活性的影响。A,黄酮吡醇对CDK9抑制的影响。已知的p300相互作用因子显示为实心深色圆圈,已知的NuA4相互作用因子显示为轮廓圆圈。B,CX4945对酪蛋白激酶2抑制的影响。已知的p300相互作用因子显示为实心深色圆圈,已知的NuA4相互作用因子显示为轮廓圆圈。C,AZ191对DYRK1A/DYRK1B抑制的影响。已知的p300相互作用因子显示为实心深色圆圈,已知的NuA4相互作用因子显示为轮廓圆圈。突出显示了DYRK1A/DYRK1B相互作用因子DCAF7和DCAF7相互作用因子NCK1SPD。D,通过AP-MS和BioID表征的A-485抑制p300对转录调节因子的影响。星号表示统计上显著的激活因子(FDR<5%)。统计显著性通过假定方差相等的不配对双尾t检验计算,并使用Benjamini、Krieger和Yekutieli (Benjamini等人,2006)的错误发现率(FDR)方法校正多个假设。E,通过AP-MS和BioID表征的JQ1抑制BET溴域对转录调节因子的影响。星号表示统计上显著的激活因子(FDR<5%)。统计显著性通过假定方差相等的不配对双尾t检验计算,并使用Benjamini、Krieger和Yekutieli (Benjamini等人,2006)的错误发现率(FDR)方法校正多个假设。

[0044] 图14显示了SRF-C3orf62的转录激活。A,SRF、C3orf62、C3orf62-Cterm和SRF-C3orf62相互作用因子通过AP-MS进行表征。SRF-C3orf62融合物与CBP和p300相互作用。B,

与表达Nanoluc-GFP的细胞相比,表达SRF-GFP、C3orf62-GFP或SRF-C3orf62-GFP的NIH3T3细胞中差异表达基因的分析。C,表达SRF-C3orf62-GFP的NIH3T3细胞中显著上调和下调基因的基因本体富集分析。D,先前发表的SRF-C3orf62-GFP和SRF/MRTF或SRF/TCF靶基因显著上调的基因重叠(Esnault等人,2014;Gualdrini等人,2016)。通过超几何分布检验来计算统计显著性。

[0045] 图15显示了SPDYE4-CITED1-P65-HSF1融合物的每个单独组分的最小激活序列对EGFP报告基因的转录活性的影响。所有组分均与PYL1二聚化域融合,并通过添加1 $\mu$ M脱落酸(ABA)后维持48小时而募集至报告基因。每个融合物的序列显示于SEQ ID NO:121、123、127、129、131、133和135中。所测试的每个单独域的序列提供于SEQ ID NO:47、90、101-104、116和118中。

[0046] 图16显示了融合两种或多种反式激活域并将其靶向同一报告基因时的实例。添加SPDYE4激活域显示单独CITED1域的活性显著改善。这与SPDYE4单独与报告基因系连时的弱活性形成鲜明对比,表明当与CITED1激活域融合时具有协同活性。所有融合构建体都指示激活域而不是全长蛋白质,除了所指示的“全长CITED1/2”,对其进行测试以将活性与其相应的反式激活域进行比较。P300core是催化域EP300(氨基酸1048-1664)。报告细胞用1 $\mu$ M脱落酸(ABA)或相同体积的DMSO处理48小时,然后收集用于流式细胞术。误差条表示四次重复的S.D.。

[0047] 图17显示将来自人类SPDYE4、CITED1、p65和HSF1蛋白的激活域的不同组合和方向系连至HEK293T细胞中的EGFP报告基因(左)或CD133基因的启动子的活性。通过在收集细胞之前24或48小时添加1 $\mu$ M脱落酸(ABA)来诱导募集。

[0048] 图18显示了用不同的激活域替换多组分SPDYE4-CITED1-P65-HSF1(SCPH)激活因子的各部分的效果。SPDYE4激活域似乎是最不可或缺的,因为用其它单独更强的激活域替换它会破坏SPCH激活因子的活性。另一方面,用我们筛选的其它强人类反式激活域替换CITED1组分对活性没有有害影响,甚至似乎通过ZXDC和C3orf62激活域提高了其效力。通过添加1 $\mu$ M脱落酸后维持48小时,将所有构建体与EGFP报告基因系连。

[0049] 图19显示当与基于rTetR或dCas9的募集系统组合使用时,包含反式激活域或其片段的不同组合的117个不同效应域的影响。A,使用117个效应域与rTetR DNA靶向域的组合进行转录激活。B,使用117个效应域与dCas9 DNA靶向域的组合进行转录激活。C,基于rTetR和dCas9的募集系统之间转录激活的相关性。

## 具体实施方式

[0050] 以下是为帮助本领域技术人员实施本公开而提供的详细描述。除非另外定义,否则本文使用的所有技术术语和科学术语具有与本公开所属领域的普通技术人员通常所理解的相同的含义。本文中所使用的术语仅仅是为了描述具体实施方案,并且并不意图限制本公开。本文中提及的所有出版物、专利申请、专利、附图和其它参考文献的全文均以引用的方式明确地并入本文。

### [0051] I. 定义

[0052] 如本文所用,除非另外指明,否则以下术语可具有下面赋予它们的含义。然而,应当理解,本领域普通技术人员已知或理解的其它含义也是可能的,并且在本公开的范围内。

本文中提及的所有出版物、专利申请、专利和其它参考文献的全文均以引用的方式并入本文。在冲突的情况下,以本说明书(包括定义)为准。另外,所述材料、方法和实例仅是说明性的而不是意图限制。

[0053] 如本文所用的术语“核酸”、“寡核苷酸”、“引物”意指两个或更多个共价连接的核苷酸。除非上下文另有明确指示,否则该术语一般包括但不限于脱氧核糖核酸(DNA)和核糖核酸(RNA),它们可以是单链的(ss)或双链的(ds)。例如,本公开的核酸分子或多核苷酸可由单链和双链DNA、作为单链区和双链区的混合物的DNA、单链和双链RNA,以及作为单链区和双链区的混合物的RNA、包含DNA和RNA(其可以是单链的或更典型地是双链的,或单链区和双链区的混合物)的杂合分子组成。此外,核酸分子可由包含RNA或DNA或RNA和DNA两者的三链区组成。如本文所用的术语“寡核苷酸”一般是指长度为至多200个碱基对的核酸,并且可以是单链或双链的。本文提供的序列可以是DNA序列或RNA序列,但应理解,所提供的序列涵盖DNA和RNA,以及互补的RNA和DNA序列,除非上下文另有明确指示。例如,序列5'-GAATCC-3'被理解为包括5'-GAAUCC-3'、5'-GGATTC-3'和5'-GGAUUC-3'。

[0054] 如本文所用的术语“功能变体”包括本文公开的多肽序列的修饰,所述修饰以基本相同的方式执行与本文公开的多肽分子基本相同的功能。例如,功能变体可包括本文所述的多肽的活性片段,例如保留转录激活活性和/或共激活因子相互作用的N末端和/或C末端截短体。功能变体可包括具有一个或多个经取代的氨基酸和/或至少保留与未修饰或非变体序列的最小序列同一性的变体。例如,功能变体可包含每十个氨基酸至多1个、2个、3个或更多个氨基酸的取代。例如,功能变体可包含与本文公开的序列具有至少80%,或至少90%,或至少95%序列同一性的序列。功能变体还可包含本文公开的序列的经保守取代的氨基酸序列。取代性氨基酸变体是其中序列中的至少一个残基已被移除并在其位置插入了不同残基的变体。取代性氨基酸变体的一个实例是保守氨基酸取代。可例如使用本文所述的方法来鉴定保留转录激活活性和/或辅激活因子相互作用的功能变体,例如包括最小片段的活性片段,其保留转录激活活性和/或辅激活因子相互作用。

[0055] 如本文所用的“保守氨基酸取代”是其中一个氨基酸残基被另一个氨基酸残基替代而不消除蛋白质的所需特性的取代。合适的保守氨基酸取代可以通过将具有相似疏水性、极性和R-链长度的氨基酸彼此取代来进行。保守取代的实例包括以一种非极性(疏水)残基如丙氨酸、异亮氨酸、缬氨酸、亮氨酸或甲硫氨酸取代另一种,以一种极性(亲水)残基取代另一种(例如在精氨酸和赖氨酸之间的取代、在谷氨酰胺和天冬酰胺之间的取代、在甘氨酸和丝氨酸之间的取代),以一种碱性残基如赖氨酸、精氨酸或组氨酸取代另一种,或以一种酸性残基如天冬氨酸或谷氨酸取代另一种。短语“保守取代”还包括使用化学衍生的残基或非天然氨基酸代替非衍生的残基,条件是此类多肽展现出必需的活性。

[0056] 本文使用的术语“异源转录激活因子”或“本文所述的转录激活因子”意指工程化融合蛋白或工程化二聚体,其包含:效应域,其包含选自表2中列出的TAD和其功能变体的至少一种反式激活域(TAD),或选自表1、表2、表3、表4、表5和/或表6中列出的TAD和其任一者的功能变体的至少两种TAD,可操作地连接至DNA靶向域。

[0057] 本文使用的术语“可操作地连接”是指两种组分之间允许其以预期方式起作用的关系。例如,第一多肽可通过共价连接(例如作为融合蛋白)或通过一种或多种相互作用组分与第二多肽可操作地连接。类似地,当报告基因可操作地连接至启动子时,启动子启动报

告基因的表达。

[0058] 转录激活因子可进一步包含提供效应域和/或DNA靶向域和/或靶DNA之间的功能性相互作用的相互作用系统的一种或多种相互作用组分。术语“相互作用组分”在本文中用于涵盖相互作用系统的一种或多种组分,所述组分一起提供所述功能相互作用。如本文所用的术语“相互作用系统”意图涵盖允许共价或非共价相互作用和/或组成型或诱导型相互作用的相互作用组分。此类相互作用系统可包括例如肽接头,任选地为蛋白酶敏感肽接头;一种或多种二聚体、三聚体或更高阶多聚化组分,例如相互作用域,任选地为诱导型二聚体、三聚体或多聚化组分,任选地为诱导型相互作用域;和/或一种或多种可调节转录激活因子亚细胞定位的组分。所述相互作用系统可包括两种或更多种组分。

[0059] DNA靶向域和效应域可例如共价连接为单个多肽(例如融合蛋白)的域,或者可通过相互作用组分,例如在某些条件下相互作用的相互作用域连接(例如连接为二聚体)。因此,异源转录激活因子可包含单一多肽,或可包含含有DNA靶向域和第一相互作用组分(例如二聚体相互作用域)的第一多肽,和含有效应域和第二相互作用组分(例如二聚体相互作用域)的第二多肽,其中第一和第二二聚体相互作用域可例如在某些条件下相互作用。更高阶多聚化系统,例如SunTag系统(Tenenbaum等人,2014)也在本文中考考虑范围之内。

[0060] 效应域和/或DNA靶向域和/或靶DNA之间的相互作用可使用多种诱导型相互作用系统来控制。例如,效应域和DNA靶向域可通过蛋白酶敏感接头,例如在存在NS3抑制剂(例如格拉瑞韦(grazoprevir))时稳定的自裂解NS3蛋白酶域连接。在另一实例中,DNA靶向域和/或效应域向细胞核的定位可由相互作用组分,例如定位域,例如使用雌激素受体配体结合域变体的经他莫昔芬(tamoxifen)调节的细胞核定位来控制。在另一实例中,DNA靶向域可连接至第一相互作用组分(例如第一相互作用域)并且效应域可连接至第二相互作用组分(例如第二相互作用域),以使得第一和第二相互作用域相互作用。

[0061] 如本文所用,术语“相互作用域”意指第一多肽(例如第一二聚体相互作用域)中的序列基序,所述序列基序能够与包含第二多肽(例如第二二聚体相互作用域)中的序列基序的结合配偶体相互作用,以可操作地连接第一多肽和第二多肽。特别地,所述术语意图涵盖第一或第二相互作用二聚体域,其一起形成例如在合适的诱导条件下二聚化的异源二聚体对。其它相互作用域也被具体考虑并且可由技术人员根据期望的特征来鉴定。合适的诱导型相互作用域对包括但不限于:可用例如雷帕霉素或AP21967诱导的FKBP/FRB(FK506结合蛋白/FKBP雷帕霉素结合)、可用例如脱落酸诱导的PYL/ABI、可用例如赤霉素或赤霉酸诱导的GID1/GAI、和可由例如蓝光和/或温度诱导的pMag/nMag。

[0062] 如本文所用,“DNA靶向域”是指在DNA结合条件下结合DNA的多肽域,从而将多肽靶向所述DNA。DNA靶向域可以是任何合适的DNA结合域,例如无酶活性的序列特异性DNA靶向蛋白,例如CRISPR-Cas蛋白(例如dCas9、dCas12或其它Cas家族蛋白)、锌指DNA结合域、转录激活因子样效应物(TALE)DNA结合域、溴域、染色域、Tudor域、WD40域、PHD域、PWWP域或来自真核生物或原核生物的其他DNA结合域(DBD)(例如叉头、碱性螺旋-环-螺旋、亮氨酸拉链、同源域、核激素受体或tet阻遏物)或其变体。DNA靶向域可以序列特异性方式结合DNA(例如Cas家族蛋白、锌指DNA结合域、TALE DNA结合域)或可结合特定染色质修饰(例如溴域(对于乙酰化组蛋白)或染色质域、Tudor域、WD40域、PHD域、PWWP域等(对于甲基化组蛋白))。DNA靶向域可以是天然的(例如非工程化的)DNA结合域,例如在天然存在的(例如内源性的)转录

因子中发现的DNA结合域,或者DNA靶向域可被工程化以例如提供定制序列特异性(例如不同于非工程化DNA结合域的序列特异性)或改变的DNA结合亲和力。工程化例如锌指DNA结合域和TALE DNA结合域以提供定制DNA结合特异性的方法是本领域已知的,例如在Maeder等人2008和Sanjana等人2012中。有酶活性的Cas9在其将会导致阻遏时,例如在引导物是截短的引导物时(参见例如[24])也可被使用。DNA靶向域可具有固有的靶序列特异性,例如在锌指DNA结合域和TALE DNA结合域的情况下,或者靶序列特异性可由另外的序列特异性因子,例如引导RNA(在CRISPR-Cas蛋白的情况下)介导。合适的DNA结合条件取决于DNA靶向域并且可包括例如额外因素的存在,例如在tet阻遏物的情况下的四环素,或在Cas家族蛋白的情况下的指导RNA。

[0063] 本文使用的术语“效应域”是指包含以下的多肽域:至少一个本文所述的反式激活域(TAD),例如表1-5中列出的TAD和其功能变体,例如其活性片段。任选地,效应域可包含两个或更多个,例如两个、三个、四个或更多个本文所述的反式激活域。在本文所述的激活因子中,活性片段可以是约15个氨基酸、约20个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约60个氨基酸、约70个氨基酸或15与70个氨基酸之间的任何数量,或超过70个氨基酸。例如,对于HSF1,活性片段可包含对应于HSF1的氨基酸406至420的GFSVDTSALDLFSP(SEQ ID NO:104)。因此,举例来说,HSF1的活性片段可包含HSF1的氨基酸401至427。其它TAD的活性片段可通过任何合适的方法来鉴定,例如使用本文所述的方法。

[0064] 异源转录激活因子可以是效应N末端或C末端融合物,例如融合的顺序可以是效应域-DNA靶向域或DNA靶向域-效应域(参见例如[25]、[26]、[27]和[28])。效应域可通过接头融合至DNA靶向域。类似地,两种或更多种TAD可通过一个或多个接头融合在一起。例如,可使用甘氨酸和甘氨酸丝氨酸接头。实施例描述的转录激活因子使用多种甘氨酸丝氨酸接头,例如SGGSGGS(SEQ ID NO:6)、GGS、SGGS(SEQ ID NO:7)和/或GSGSGS(SEQ ID NO:8)。也可使用其它接头,例如INSRSSGS(SEQ ID NO:9)。

[0065] 如本文所用的术语“CRISPR-Cas”或“Cas”是指CRISPR成簇的规律间隔的短回文重复序列-CRISPR相关(CRISPR-Cas)蛋白,所述蛋白结合RNA并且通过其所结合的RNA靶向至特定DNA序列。CRISPR-Cas是II类单体Cas蛋白,例如II型Cas,例如Cas9。Cas9蛋白可以是来自化脓性链球菌(*Streptococcus pyogenes*)、新弗朗西斯菌(*Francisella novicida*)、耐苏氏放线菌(*A. Naeslundii*)、金黄色葡萄球菌(*Staphylococcus aureus*)或脑膜炎奈瑟菌(*Neisseria meningitidis*)的Cas9。任选地,Cas9来自化脓性链球菌。Cas蛋白也可以是例如来自氨基酸球菌属种(*Acidaminococcus* sp.)、毛螺菌科细菌(*Lachnospiraceae* bacterium)或土拉弗朗西斯菌(*Francisella tularensis*) (这些已被证明可充当dCas变体)的Cas12a(例如dCas12a),也可使用CasΦ(Cas12j)和CasX(Cas12e)。

[0066] 如本文所用,术语“dCas9”是指无酶活性(或死的)Cas9,该Cas9缺乏DNA核酸内切酶活性但保留了靶DNA结合活性。例如,dCas9包含RuvC1和HNH核酸酶域中的CAS9和D10A/H840A突变的序列。任选地,dCas9是这样的蛋白质,该蛋白质包含与由SEQ ID NO:1编码的蛋白质具有至少80%、至少90%、至少95%、至少99%或100%序列同一性的氨基酸序列并且包含D10A/H840A突变并且保留了Cas9靶DNA结合活性(例如结合gRNA和靶位点)。类似地,dCas12a是指无酶活性的Cas12a。

[0067] 如本文所用的术语“引导RNA”、“引导物”或“gRNA”是指与特定DNA序列杂交并且最

少包含间隔区序列的RNA分子。引导RNA可进一步包含结合CRISPR-Cas蛋白的蛋白结合区段。引导RNA的与特定DNA序列杂交的部分在本文中称为核酸靶向序列或间隔区序列。引导物的蛋白结合区段可包含例如tracrRNA和/或直接重复序列(direct repeat)。根据上下文,术语“引导物”或“引导RNA”可以指单独的间隔区序列,或包含间隔区序列和蛋白结合区段的RNA分子。引导RNA可以由相应的DNA序列表示。当酶是Cas9时,引导物可以是例如如[24]中描述的包含15个或更少的与靶位点互补的核苷酸的截短的引导物。例如,当Cas9与截短的引导物相互作用时,Cas9的DNA结合能力保持不变,同时其核酸分解活性被消除。可以使用保持Cas结合能力的任何长度的引导物。

[0068] 如本文所用的术语“间隔区”或“间隔区序列”是指引导物的与靶序列或靶序列的一部分形成或能够形成RNA-DNA双链体的部分。间隔区序列可以与特定的CRISPR靶序列互补或对应。间隔区序列的核苷酸序列可决定CRISPR靶序列并且可被设计或构造成靶向期望的CRISPR靶位点。

[0069] 如本文所用的术语“tracrRNA”是指“反式编码的crRNA”,该反式编码的crRNA可以例如与CRISPR-Cas蛋白(例如Cas9)相互作用并且可以连接至引导RNA或者形成引导RNA的一部分。tracrRNA可以是来自例如化脓性链球菌的tracrRNA。tracrRNA可具有例如5'-gtttcagagctatgctggaaacagcatagcaagttgaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcggtgc-3'(SEQ ID NO:2)的序列。也可使用其它tracrRNA。合适的tracrRNA可由本领域技术人员鉴定,包括例如5'-GTTTtagagctagaaatagcaagttaaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcggtgc-3'(SEQ ID NO:3)或5'-GTTTCAGAGCTACAGCAGAAATGCTGTAGCAAGTTGAAAT-3'(SEQ ID NO:4)。

[0070] 如本文所用的术语“CRISPR靶位点”或“CRISPR-Cas靶位点”意指激活的CRISPR-Cas蛋白(例如CRISPR-Cas蛋白,例如与引导RNA结合的dCas9)将在合适的条件下结合至的核酸。CRISPR靶位点包含原型间隔区相邻基序(protospacer-adjacent motif,PAM)和CRISPR靶序列(即与激活的CRISPR-Cas蛋白所结合至的引导物的间隔区序列对应)。PAM的序列和相对于CRISPR靶序列的相对位置将取决于CRISPR-Cas蛋白的类型。例如,Cas9或dCas9的CRISPR靶位点从5'至3'可包含15至25、16至24、17至23、18至22或19至21个核苷酸、任选地20个核苷酸的靶序列,随后是具有序列NGG的3个核苷酸的PAM。因此,Cas9靶位点可具有序列5'-N<sub>1</sub>NGG-3',其中N<sub>1</sub>的长度为15至25、16至24、17至23、18至22或19至21个核苷酸,任选地长度为20个核苷酸。

[0071] CRISPR靶位点可位于任何合适的基因组基因座中。例如,CRISPR靶位点可位于启动子、增强子、3' UTR或其它调节元件中,位于基因中,任选地内含子或外显子中,位于与非编码RNA对应的基因座中,或位于基因间区中。任选地,CRISPR靶位点位于启动子或增强子中。

[0072] 位于细胞核中的靶DNA需要可进入细胞核的转录激活因子。因此,转录激活因子可定位于细胞核和/或可包含例如一种或多种细胞核定位信号(NLS),任选一种或多种SV40 NLS。任选地,转录激活因子包含两种或更多种NLS。任选地,转录激活因子可包含一种或多种N末端NLS、一种或多种C末端NLS、一种或多种内部NLS,或一种或多种N末端NLS、一种或多种C末端NLS和/或一种或多种内部NLS。特别考虑了其它配置。在一个实施方案中,NLS是具有序列PKKKRKV(SEQ ID NO:22)的SV40 NLS。在一个实施方案中,NLS还包含N末端和/或C末

端接头,例如INSRSSGS (SEQ ID NO:9),并且任选地具有序列INSRSSGSPKKRKVGS (SEQ ID NO:141)。

[0073] 也可用标签来标记转录激活因子。例如,合适的标签包括但不限于Myc、FLAG、HA、V5、ALFA、T7、6xHis、VSV-G、S-tag、AviTag、StrepTagII、CBP、GFP、mCherry。标签可在N末端、C末端或异源转录激活因子的两个组分之间,例如DNA靶向域与效应域之间融合。

[0074] 在提供数值范围的情况下,应理解,在所述范围的上下限值之间的每个中间值(除非上下文另外清楚地指出,否则所述中间值达到下限值单位的十分之一)和在所述范围中的任何其它陈述值或中间值涵盖在说明书内。涵盖从任何下限值到任何上限值的范围。这些较小范围的可独立地包括于该较小范围中的上限值和下限值也涵盖在说明书内,受制于所述范围内任何特别排除的限值。在所陈述范围包括限值中的一个限值或两个限值的情况下,不包括那些所包括的限值中的任一个限值或两个限值的范围也包括在说明书中。

[0075] 必须指出,如本文和在所附权利要求书中所使用,单数形式“一个(a)/一种(an)”和“该/所述(the)”包括复数个指示物,除非上下文另外清楚地指出。

[0076] 本文的详细描述和权利要求书内的所有数值都通过“约”或“大约”指示值修改,并且考虑了本领域普通技术人员将预期的实验误差和变化。

[0077] 如本文在专利说明书和权利要求书中所用的短语“和/或(以及/或者,并且/或者)”应理解为意指如此结合的要素,即,在一些情况下结合存在并且在另一些情况下分离地存在的要素中的“任一个要素或两个要素”。用“和/或(以及/或者,并且/或者)”列出的多个要素应以相同方式解读,即,如此结合的要素中的“一个要素或多个要素”。除了通过“和/或(以及/或者,并且/或者)”句子具体标识的要素之外,其它要素也可以任选地存在,无论与具体标识的那些要素相关或无关。

[0078] 如本文在专利说明书和权利要求书中所使用,“或(或者)”应理解为具有与如上文所定义的“和/或(以及/或者,并且/或者)”相同的含义。例如,当将清单中的项分列时,“或(或者)”或“和/或(以及/或者,并且/或者)”应解释为包含性的,即,不仅包括许多要素或要素清单中的至少一个要素,而且包括许多要素或要素清单中的超过一个要素,以及任选的另外未列出的项。仅当术语有明确相反的指示(例如“.....中的仅一个”或“.....中的刚好一个”),或者当在权利要求书中使用时,“由.....组成”将指包括许多要素或要素清单中的刚好一个要素。一般来说,如本文中所用的术语“或(或者)”在后面存在排他性术语(例如“.....中的任一个”、“.....中的一个”、“.....中的仅一个”或“.....中的刚好一个”)时,仅应解释为指示排他性替代物(即,“一个或另一个,而非两个”)。

[0079] 在权利要求书以及以上专利说明书中,所有过渡性短语,例如“包含”、“包括”、“携有”、“具有”、“含有”、“涉及”、“容纳”、“由.....组成”等都应理解为开放式的(open-ended),即,意指包括(但不限于)。仅过渡性短语“由.....组成”和“基本上由.....组成”分别应是封闭或半封闭的过渡性短语。

[0080] 如本文在专利说明书和权利要求书中所使用,短语“.....中的至少一者”在提及一个或多个要素的清单时,应理解为意指选自要素清单中的任一个或多个要素的至少一个要素,但不一定包括要素清单内具体列举的每一个要素中的至少一者,并且不排除要素清单中的要素的任何组合。此定义还允许任选地存在短语“至少一个”所指的要素清单内具体标识的要素以外的要素,不论是否与那些具体标识的要素相关。

[0081] 类似地,本文特别设想的是,涉及一组要素的短语“一个或多个”包括所述组的至少一个成员,但不一定包括所述组的各一个成员。例如,当要素包括组成员A、B和/或C中的一者或多者时,所述要素可包括A;B;C;A和B;A和C;B和C。还可存在所述组中未具体列出的其它成员,例如参考以上实例,所述要素可另外包括未列出的成员D,并且相应地可包括A和D;B和D;A、C和D;等。

[0082] 如本文所用的术语“约”意指加或减所提及数字的10%-15%、5-10%,或任选地约5%。

[0083] 还应理解,在本文描述的包括多于一个步骤或动作的某些方法中,所述方法的步骤或动作的顺序不一定限于所述方法的步骤或动作被叙述的顺序,除非上下文另有指示。此外,在特定部分中描述的定义和实施方案意图适用于本文中描述的其适合于的其它实施方案,如本领域技术人员将理解。例如,在以下段落中,更详细地定义了本公开的不同方面。如此定义每个方面可与任何其它一个或多个方面组合,除非清楚地表明相反。具体地,本文描述的任何特征可与本文描述的任何其它特征组合。

[0084] 虽然还可在本公开的实践或测试中使用类似于或等同于本文所述的材料和方法的任何材料和方法,但现在描述以下材料和方法。

## [0085] II. 材料和方法

[0086] 本文描述异源转录激活因子的集合,其包含一种或多种反式激活域(TAD)和其组合(“效应域”),其可以可操作地连接至DNA靶向域以产生异源转录激活因子,并且其可用于激活所需基因(包括内源基因)的基因表达,例如用于治疗目的。在本文所述的异源转录激活因子中,效应域可操作地连接至DNA靶向域,所述DNA靶向域可指导融合构建体与基因组中的任何基因座的结合。如实施例中所证明,包含与效应域功能性关联的dCas9或rTetR的异源转录激活因子可用于激活靶基因的转录,所述效应域包含表1中列出的任一种TAD、其活性片段,例如SEQ ID NO:100-104、107-115、118-119、156-160、162、164和166-185的那些,以及两种或更多种TAD或其活性片段的组合,例如如表3中所列。TAD或其活性片段可在TAD或活性片段末端另外包含接头和/或另外的天然序列,例如1、2、3、4、5、6、7或更多个,例如至多5个、至多10个、至多20个、至多30个或至多40个(或之间的任何数量)N末端和/或C末端氨基酸。例如,表3中标记为“ZXDC短”的TAD(SEQ ID NO:107)包含在ZXDC-12和13中均发现的40个氨基酸片段,并且包含ZXDC-12天然序列的另外6个N末端氨基酸和ZXDC-13天然序列的另外5个C末端氨基酸。在一个实施方案中,活性片段可比所鉴定的片段短。例如,活性片段可为例如ZXDC-12和ZXDC-13两者中发现的40个氨基酸片段中的20个氨基酸或25个氨基酸,任选地例如SEQ ID NO:107的内部部分。较短活性片段例如以SEQ ID NO:119为例,其包含在HSF1-20和HSF1-21两者中发现的20个氨基酸片段。活性片段可如本文别处所述进行鉴定。例如,可通过比较活性片段来鉴定最小片段,例如对于ATF6,表5中所示的重叠片段是HRLDEDWDSALFAELGYFTDTDELQLEAANETYENNFDNL,并且对于KLF7,重叠片段是YFSALPSLEETWQQTCLERLYLQTEPRRISETFGEDLDC。

[0087] 因此,本公开的一个方面包括异源转录激活因子,其包含DNA靶向域和效应域,所述效应域包含至少一种选自表1-6中的任一者(任选地表1或任选地表2或表6)中列出的TAD中的任一者的TAD、其活性片段和其组合,例如选自表1或表3中列出的TAD的至少两种TAD或其功能变体,优选地选自表4、或表5或表6中列出的TAD的至少一种TAD,和/或其功能变体,

其中DNA靶向域和效应域可操作地连接。

[0088] 在一个实施方案中,至少一种TAD选自表1。

[0089] 在一个实施方案中,至少一种TAD选自表2。

[0090] 在一个实施方案中,至少一种TAD选自表3。

[0091] 在一个实施方案中,至少一种TAD选自表4。

[0092] 在一个实施方案中,至少一种TAD选自表5。

[0093] 在一个实施方案中,至少一种TAD选自表6。

[0094] 应理解,至少两种TAD或其功能变体可选自任何表,例如选自表3和表4中的每一者、两者或更多者,例如表3或表4、表3等。还预期分组可包括表1-6中的任一者中描述的TADS的任何子组合,例如可排除一种或多种TAD。

[0095] DNA靶向域和效应域可通过共价连接可操作地连接,例如作为单个多肽的域,和/或可通过一种或多种相互作用组分(例如相互作用域)可操作地连接和/或在某些条件下相互作用。因此,在一个实施方案中,异源转录激活因子是单一多肽。在另一实施方案中,异源转录激活因子还包含一对(即第一和第二)相互作用域,任选地二聚体相互作用域,任选地一对在合适条件下二聚化的诱导型二聚体相互作用域。例如,异源转录激活因子可包含含有DNA靶向域和第一二聚体相互作用域(任选地为诱导型二聚化域)的第一多肽,以及含有效应域和第二二聚体相互作用域(任选地为诱导型二聚化域)的第二多肽,其中第一二聚体相互作用域和第二二聚体相互作用域相互作用,任选地,第一诱导型二聚化域和第二诱导型二聚化域在一种或多种诱导剂存在下相互作用。

[0096] 如实施例中所示,包含ABI1和PYL1的异源转录激活因子的二聚化可通过添加脱落酸来诱导。因此,在一个实施方案中,转录激活因子包含第一和第二诱导型二聚化域,所述第一和第二诱导型二聚化域在诱导剂存在下提供可诱导的转录激活。技术人员可容易地鉴定和选择可一起使用的合适的诱导型二聚化域。可使用任何合适的诱导型二聚化域,例如可通过添加脱落酸来诱导ABI1和PYL1的二聚化。其它诱导型系统包括那些基于利用雷帕霉素、赤霉素/赤霉素的诱导的系统和基于分裂dCas9的系统。例如,GID1和GAI的二聚化可由赤霉素诱导,并且FKBP和FRB的二聚化可用雷帕霉素或其类似物(例如雷帕霉素类似物(rapalog))诱导。本文还考虑了更高阶多聚化系统,例如SunTag系统(Tenenbaum等人,2014)。

[0097] 也可使用其它诱导型系统来控制DNA靶向域与效应域之间的相互作用。其它系统(其不依赖于二聚化)包括格拉瑞韦诱导的稳定性(Tague等人,2018)或使用雌激素受体配体结合域变体的他莫昔芬调节的核定位。在格拉瑞韦诱导的稳定性(其抑制NS3活性)的情况下,DNA靶向域和效应域将通过自裂解NS3蛋白酶域连接。仅在存在格拉瑞韦(其抑制NS3活性)的情况下,DNA靶向域和效应域才会保持在一起并调节基因表达。

[0098] DNA靶向域可选自多种DNA结合域,例如锌指DNA结合域、转录激活因子样效应物(TALE)DNA结合域、dCas9、dCas12或其它Cas家族蛋白,或来自真核生物或原核生物的其他DNA结合域(DBD)(例如叉头、碱性螺旋-环-螺旋、亮氨酸拉链、同源域、细胞核激素受体或tet阻遏物),或其变体。DNA靶向域可以是天然的(例如非工程化的)DNA结合域,例如在天然存在的(例如内源性的)转录因子中发现的DNA结合域,或者DNA靶向域可被工程化以例如提供定制序列特异性(例如不同于非工程化DNA结合域的序列特异性)或改变的DNA结合亲和

力。工程化例如锌指DNA结合域和TALE DNA结合域以提供定制DNA结合特异性的方法是本领域已知的,例如在Maeder等人2008和Sanjana等人2012中。在本文所述的异源转录激活因子包含含有天然DNA结合域的DNA靶向域的情况下,效应域将靶向转录因子内源性结合至的所有基因座,从而增强/取代内源转录因子的功能。例如,已知用VP16替换Oct4反式激活域可提高将成纤维细胞重编程为iPS细胞的效率。类似地,包含可操作地连接至效应域的天然DNA结合域的异源转录激活因子可通过激活受内源转录因子调节的靶基因的转录来促进例如伤口愈合、转分化或组织再生。在工程化(例如定制序列特异性)锌指DNA结合域、TALE DNA结合域或Cas家族蛋白的情况下,可将效应域带到一个或多个特定基因座,或任选地以受控的方式带到基因组中的单个基因座。

[0099] 在一个实施方案中,DNA靶向域包含CRISPR-Cas蛋白,例如dCas9。可使用保留gRNA和靶DNA结合活性的无酶活性的CRISPR-Cas蛋白。例如,Cas9中的D10A/H840A的突变会在RuvC1和HNH核酸酶域中引入突变并导致失活。在一个实施方案中,CRISPR-Cas蛋白是dCas9,其具有SEQ ID NO:1的氨基酸序列或与SEQ ID NO:1具有至少80%、至少90%、至少95%或至少99%序列同一性的氨基酸序列,并且包含D10A/H840A,并且保留gRNA和靶DNA结合活性。技术人员可鉴定另外考虑的其它无酶活性的CRISPR-Cas蛋白。

[0100] 在一个实施方案中,DNA靶向域包含锌指DNA结合域。在一个实施方案中,锌指DNA结合域是工程化锌指DNA结合域,其已被工程化以结合特定DNA序列。

[0101] 效应域包含至少一种本文所述的反式激活域(TAD)或其活性片段。如实施例中所示,本文鉴定的各种全长ORF和TAD可单独或组合使用,以激活GFP报告构建体或内源基因,例如CD133的转录。实施例中还显示,效应域可包含表1中所示的至少一种TAD域,和/或其活性片段,例如表3中所示。因此,在一个实施方案中,效应域包含表2、表4和/或表5和/或表6中所示的至少一种TAD和/或其任一者的功能变体,或表1、表2、表3、表4、表5和/或表6中所示的两种或更多种TAD和其任一者的功能变体。在一个实施方案中,TAD包含具有与表1、表2、表3、表4、表5和/或表6中的任一TAD域和其任一者的功能变体具有至少80%、至少90%、至少95%或至少99%序列同一性的序列的多肽,并且其保留(例如至少80%有效)转录激活活性和/或与特定转录辅激活因子,例如CBP/p300、NuA4和/或BRD4的相互作用。

[0102] 也可使用其变体和组合。在一个实施方案中,效应域包含两个或更多个串联TAD,任选地两种TAD、三种TAD、四种TAD或超过四种TAD,例如5种TAD、10种TAD、15种TAD、20种TAD、25种TAD、30种TAD,或5种TAD与30种TAD之间的任何数量的TAD,或超过30种TAD。在一个实施方案中,效应域包含选自以下的两种或更多种TAD或其功能变体:表1、表2、表3、表4、表5和/或表6中所列的那些和其任一者的功能变体。在一个实施方案中,效应域包含选自表3中所列的那些的三种或四种TAD。在一个实施方案中,效应域包含SEQ ID NO:185、SEQ ID NO:103、SEQ ID NO:167、SEQ ID NO:105、SEQ ID NO:106和/或SEQ ID NO:104中的一者或多者。在一个实施方案中,效应域包含SEQ ID NO:185,任选地SEQ ID NO:90、91、102或157。在一个实施方案中,效应域包含SEQ ID NO:103,任选地SEQ ID NO:46、47或162。在一个实施方案中,效应域包含SEQ ID NO:167,任选地SEQ ID NO:101、110、166或172。在一个实施方案中,效应域包含SEQ ID NO:105,任选地SEQ ID NO:116、117或165。在一个实施方案中,效应域包含SEQ ID NO:106,任选地SEQ ID NO:116或117。在一个实施方案中,效应域包含SEQ ID NO:104,任选地SEQ ID NO:118、119或159。在一个实施方案中,效应域包含SPDYE4-

CITED1-RELA-HSF1 (SEQ ID NO:121);SPDYE4-CITED1-RELA (SEQ ID NO:123);HSF1-RELA-SPDYE4-CITED1 (SEQ ID NO:125);SPDYE4-CITED1-p65-miniHSF1 (SEQ ID NO:127);miniSPDYE4-CITED1-p65-HSF1 (SEQ ID NO:129);SPDYE4-miniCITED1-p65-HSF1 (SEQ ID NO:131);SPDYE4-CITED1-minip65(C)-HSF1 (SEQ ID NO:133);或SPDYE4-CITED1-minip65(N)-HSF1 (SEQ ID NO:135)。在一个实施方案中,效应域包含SPDYE4-CITED1-SERTAD2-HSF1;SPDYE4-CITED1-KLF6-HSF1;SPDYE4-CITED1-ZXDC-HSF1;SPDYE4-CITED1-ATF6-HSF1;SPDYE4-CITED1-FOXO1-HSF1;SPDYE4-CITED1-ATMIN-HSF1;SPDYE4-CITED1-p65-SERTAD2;SPDYE4-CITED1-p65-KLF6;SPDYE4-CITED1-p65-ZXDC;SPDYE4-CITED1-p65-ATF6;SPDYE4-CITED1-p65-FOXO1;SPDYE4-CITED1-p65-ATMIN;SPDYE4-C3orf62-p65-HSF1;SPDYE4-DDIT3-p65-HSF1;SPDYE4-FOXO1-p65-HSF1;SPDYE4-ATMIN-p65-HSF1;SPDYE4-ZXDC-p65-HSF1;C3orf62-CITED1-p65-HSF1;C11orf74-CITED1-p65-HSF1;KLF6-CITED1-p65-HSF1;ZXDC-CITED1-p65-HSF1;或SOX7-CITED1-p65-HSF1。在一个实施方案中,效应域包含SPDYE4-C3orf62.2-P\_AD-HSF1 (SEQ ID NO:174);SPDYE4-C3orf62.3-P\_AD-HSF1 (SEQ ID NO:176);SPDYE4-C3orf62\_MT-P\_AD-HSF1 (SEQ ID NO:178);SPDYE4-DDIT3\_MT-P\_AD-HSF1 (SEQ ID NO:180);SPDYE4-CITED1-P\_AD-HSF1\_MT (SEQ ID NO:182);或3x ZNF473\_KRAB (SEQ ID NO:184)。特别考虑了其它构造。

[0103] 效应域可包含具有不同转录辅激活因子偏好的两种或更多种TAD。例如,效应域可包含与CBP/p300组分相互作用的TAD,例如FOXO TAD,和与BET组分相互作用的TAD,例如SPDYE4 TAD。效应域可包含具有类似转录辅激活因子偏好的两种或更多种TAD。例如,效应域可包含与CBP/p300组分相互作用的两种TAD,例如效应域可包含FOXO1 TAD和CITED1 TAD。特别考虑了其它构造。

[0104] 对于功能变体,本文所用的“有效”意指与未修饰或非变体TAD (例如野生型或全长TAD) 相比,功能变体保留至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%、100%或大于100%转录激活活性和/或辅激活因子相互作用。变体 (例如截短体) 的转录激活活性和/或共激活因子相互作用可例如使用本文描述的方法来确定。例如,可使用实施例中描述的一个或多个GFP报告基因系统来确定转录激活活性。可在控制每个DNA靶向域 (例如dCas9) 的表达水平的同时将变体系连至相同的报告基因或内源环境。在报告基因或靶基因的所诱导表达中检测到的与亲本TAD相比的任何差异都可能促成变体的影响。辅激活因子相互作用可例如通过AP-MS和/或BioID来测定,例如如实施例中所示。

[0105] 表1-6以及SEQ ID NO:120-135和173-184中提供了示例性TAD和效应域核酸和多肽。在一个实施方案中,效应域可包含由所述核酸编码的氨基酸序列,或与由SEQ ID NO:120-135和173-184的TAD编码的氨基酸序列具有至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列。此类多肽的编码多肽 (融合或在表达和激活时) 的活性与例如SEQ ID NO:120-135和173-184一样有效 (例如提供至少80%有效的转录激活)。

[0106] 在一个实施方案中,效应域通过接头与DNA靶向域融合。在一个实施方案中,两种或更多种TAD通过一个或多个接头融合在一起。例如,可使用甘氨酸和甘氨酸丝氨酸接头。实施例中描述的转录激活因子使用多种甘氨酸丝氨酸接头,例如SGGSGGS (SEQ ID NO:6)、GGG、SGGS (SEQ ID NO:7) 和/或GSGSGS (SEQ ID NO:8)。也可使用其它接头,例如INSRSSGS

(SEQ ID NO:9)。

[0107] 在一个实施方案中,转录激活因子包含一种或多种细胞核定位信号(NLS)。可使用任何合适的NLS。任选地,NLS是SV40 NLS。所述一种或多种NLS可以是一种或多种N末端NLS、一种或多种C末端NLS、一种或多种内部NLS和/或其组合。任选地,NLS可包含SEQ ID NO:22的NLS。在一个实施方案中,NLS还包含N末端和/或C末端接头,例如INSRSSGS (SEQ ID NO:9),并且任选地具有序列INSRSSGSPKKRKVGS (SEQ ID NO:141)。

[0108] 如本文所述,转录激活因子或效应域可由核酸编码和/或由表达构建体表达。因此,本公开的一个方面是编码本文所述的转录激活因子的核酸。本公开的另一方面是编码本文所述的转录激活因子的效应域的核酸。例如,核酸可编码如表1至6中的任一者,任选地表2、4、5和/或6,任选地表2或表4或表6中所提供的TAD,或表1-6中所提供的两种或更多种TAD。在一个实施方案中,核酸可包含SEQ ID NO:120、122、124、126、128、130、132、134、173、175、177、179和180中的任一者的核酸,或与SEQ ID NO:120、122、124、126、128、130、132、134、173、175、177、179和180具有至少80%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性的序列,其中异源转录激活因子例如大约与由SEQ ID NO:120、122、124、126、128、130、132、134、173、175、177、179和180编码的效应域一样有效地激活转录,例如至少80%有效、至少85%有效、或至少90%有效、至少95%有效、至少96%有效、至少97%有效、至少98%有效、至少99%有效、至少100%有效、或多于100%有效,例如在本文描述的测定中所评估。序列同一性例如相对于完整效应域序列或相对于其中编码的一种或多种TAD或TAD片段。其它部分、接头、NLS等可完全不同。编码效应域的核酸可适合于产生编码本文所述的转录激活因子的核酸。例如,编码效应域的核酸可侧接有合适的克隆位点,或者包含所述核酸的表达构建体或载体可包含克隆位点以促进插入DNA靶向域以可操作地连接效应域和DNA靶向域。

[0109] 本文使用的术语“克隆位点”是指使用重组DNA技术(克隆)可将感兴趣的核酸分子插入其中或与感兴趣的核酸分子连接的核酸分子的一部分。在表达盒的情况下,克隆位点可位于启动子与聚腺苷酸化信号之间,使得感兴趣的核酸分子可被克隆至与启动子和聚腺苷酸化位点可操作连接的表达盒中。技术人员已知若干克隆技术,并且克隆位点将包括必要的特征(例如限制性核酸内切酶位点、重组酶识别位点、或平端或突出端)以允许在克隆位点插入感兴趣的核酸分子。克隆位点可以是例如多克隆位点(MCS)或多接头区域,其包含多个独特的限制酶识别位点以允许插入感兴趣的核酸分子。或者或另外,克隆位点可包括一个或多个重组酶识别位点以允许通过重组克隆插入DNA;使用位点特异性重组酶(例如整合酶或Cre重组酶)来催化DNA插入。重组克隆系统的实例包括Gateway®(整合酶)、Creator™(Cre重组酶)和Echo Cloning™(Cre重组酶)。对于一些克隆策略,表达盒或载体可作为线性分子提供,允许感兴趣的核酸分子的平端或突出端与表达盒或载体的平端或突出端连接,例如通过连接或聚合酶活性,从而形成环状分子。在此情况下,表达盒或载体的平端或突出端可一起被视为克隆位点。这种方法通常用于克隆PCR产物。

[0110] 相关方面是表达构建体,所述达构建体包含编码可操作地连接至启动子和转录终止位点的转录激活因子的核酸。可使用任何合适的启动子。合适的启动子可由本领域技术人员鉴定,并且可包括例如CMV、EF1A或PGK。例如,SEQ ID NO:25、26、27和/或28的启动子和/或增强子序列可用于表达构建体中。也可使用诱导型启动子。

[0111] 在一个实施方案中,构建体是载体。可使用任何合适的载体。合适的载体可由本领域技术人员鉴定,并且可包括病毒载体,任选地慢病毒载体或腺病毒载体。合适的载体可包括例如用于表达效应构建体的启动子、polyA尾、用于增加表达的稳定性3' UTR元件(如WPRE)、绝缘子序列、慢病毒包装信号、荧光蛋白和/或抗生素抗性标记。本领域技术人员可鉴定另外的合适的组分。

[0112] 在另一实施方案中,转录激活因子、核酸、构建体或载体在细胞中。可使用任何合适的细胞,并且可由技术人员基于期望的应用来确定。细胞可来自任何生物体。任选地,细胞是哺乳动物细胞,例如人类细胞或小鼠细胞。任选地,细胞是细胞系。细胞系可以是任何合适的细胞系。

[0113] 转录激活因子、核酸、构建体或载体可以任何合适的方式引入至细胞中,例如通过转染。合适的转染试剂和方法是本领域常规实践的并且可由技术人员鉴定。任选地,该构建体是病毒载体,任选地是慢病毒载体,并且通过转导被引入到细胞中。合适的转导方法是本领域常规实践的并且可由技术人员鉴定。

[0114] 在一些实施方案中,细胞稳定表达异源转录激活因子,任选地,细胞被稳定转导,例如使用包含编码异源转录激活因子的核酸的病毒制备。

[0115] 另一个方面是转录激活系统,所述转录激活系统包含本文所述的转录激活因子、编码所述转录激活因子的核酸,或包含所述核酸的构建体或载体,或表达所述转录激活因子的细胞。在基于CRISPR-Cas的系统的情况下,所述系统包含至少一种gRNA。在基于诱导型二聚化域的系统的情况下,所述系统任选地包含至少一种诱导剂。

[0116] 还提供了一种组合物,该组合物包含本文所述的异源转录激活因子、本文所述的核酸、本文所述的构建体、本文所述的载体、本文所述的细胞和/或本文所述的转录激活系统。所述组合物可包含载剂(例如BSA),或根据组合物组分合适的稀释剂(任选地为水或缓冲盐水)。所述组合物可包含多种组分,例如包含相同或不同元件的转录激活因子、核酸、构建体、载体或细胞。

[0117] 本文还提供了例如用于激活靶基因转录或执行本文所述方法的试剂盒,所述试剂盒包括本文所述的转录激活因子;核酸;表达构建体;或编码本文所述的转录激活因子的载体;或表达本文所述的转录激活因子的细胞;以及任选的容纳所述转录激活因子、核酸、表达构建体、载体、细胞或组合物的小瓶。该试剂盒可包括前面提到的组分中的一者或多者。任选地,试剂盒包括gRNA表达构建体、诱导剂和/或用于实施本文所述方法的说明书。

[0118] 本文还描述了激活细胞中靶基因的转录的方法。如实施例中所证明,本公开的转录激活因子可靶向于基因组基因座,例如启动子以激活细胞中靶基因的转录。

[0119] 本文鉴定的转录效应子可以是全长蛋白质、其片段(反式激活域)、其功能变体或反式激活域或其功能变体的组合。其涵盖多种不同的转录激活强度,从非常强到中等到弱。这可用于实现内源基因所需表达水平,特别是在表达过高可能导致有害表型的情况下。可通过选择不同的TAD或功能变体(例如活性片段)包含在效应域中来调整激活强度。效应域的激活强度可由技术人员确定,例如使用本文描述的实施例中所示的募集测定中的MFI或GFP阳性细胞百分比。例如,效应域的相对强度可通过比较与特定DNA靶向域和特定DNA靶标组合的特定效应域相对于对照,例如海肾(Renilla)与相同DNA靶向域和特定DNA靶标组合的MFI或GFP阳性细胞百分比来确定。高激活可被视为例如至少或高于50X对照、至少或高于

75X对照、至少或高于100X对照、或至少或高于150X对照。中等激活可被视为例如至少或高于10X、至少或高于20X、至少或高于30X、或至少或高于40X对照,以及至多50X、至多75X、至多100X、或至多150X对照。低激活可被视为例如至少2X、至少2.5X、至少3X、至少4X或至少5X对照以及至多10X、至多20X、至多30X或至多40X对照。例如,如图19A和19B中所示,相对于海肾对照,高激活可被视为>50倍,中等激活可以是20X至50X,并且低激活可以是至少3X至至多20X。可根据所需的应用来选择合适的激活水平。

[0120] 如本文所述,可利用化学诱导的与例如雷帕霉素类似物或脱落酸的二聚化来添加用于转录调节的另一水平的控制。在这种情况下,一半(例如DNA靶向域)将与FKBP或PYL1融合,而另一半(例如效应域)与FRB或ABI1融合。用雷帕霉素类似物或脱落酸处理将分别诱导FKBP与FRB或PYL1与ABI1之间的相互作用,从而导致暂时调节的基因表达。如实施例中所示,包含ABI1和PYL1的异源转录激活因子的二聚化可通过添加脱落酸来诱导。本领域技术人员可容易地鉴定和选择可一起使用的合适的诱导型二聚化域和诱导剂。可以使用蛋白质二聚化域和诱导剂的任何合适的诱导型组合,例如ABI1和PYL1的二聚化可以通过添加脱落酸来诱导。其它诱导型系统包括那些基于利用雷帕霉素、赤霉素/赤霉素的诱导的系统和基于分裂dCas9的系统。例如GID1和GAI的二聚化可由赤霉素诱导,并且FKBP和FRB的二聚化可用雷帕霉素或其类似物(例如雷帕霉素类似物(rapalog))诱导。本文还考虑了更高阶多聚化系统,例如SunTag系统(Tenenbaum等人,2014)。

[0121] 也可使用其它诱导型系统来控制DNA靶向域与效应域之间的相互作用。其它系统(其不依赖于二聚化)包括格拉瑞韦诱导的稳定性(Tague等人,2018)或使用雌激素受体配体结合域变体的他莫昔芬调节的核定位。在格拉瑞韦诱导的稳定性(其抑制NS3活性)的情况下,DNA靶向域和效应域将通过自裂解NS3蛋白酶域连接。只有在存在格拉瑞韦(其抑制NS3活性)的情况下,DNA结合域和效应域才会保持在一起并且调节基因表达。

[0122] 因此,本公开的一个方面是激活细胞中靶基因表达的方法,所述方法包括将本文所述的转录激活因子引入所述细胞中,以及在合适的条件下培养所述细胞以使得DNA靶向域将转录激活因子引导至靶位点并且效应域激活所述靶基因的转录。在一个实施方案中,靶基因是内源基因。在其中转录激活因子包含CRISPR-Cas的实施方案中,所述方法还包括将至少一种靶向细胞中期望基因组基因座的gRNA引入所述细胞中,以及在合适的条件下培养细胞以使得所述至少一种gRNA与CRISPR-Cas蛋白缔合并引导CRISPR-Cas蛋白将转录激活因子引导至CRISPR靶位点,以使得效应域激活靶基因的转录。在其中转录激活因子在DNA靶向域和效应域中的每一者中均包含诱导型二聚化域的实施方案中,所述方法还包括将至少一种诱导剂引入到该细胞中以及在使第一和第二诱导型二聚化域缔合从而使所述至少一个效应域激活靶基因的转录的合适条件下培养所述细胞。

[0123] 本文所述的方法可用于调节靶基因的基因表达,例如诱导内源基因的表达或调节基因组的限定区域中的染色质开放。举例来说,一些TAD可促进基因间区域(即不是启动子或增强子)的染色质开放,这可能导致染色质开放和染色体折叠重排。

[0124] 本文所述的方法可用于鉴定或筛选对于细胞活力或感兴趣的表型重要的一个或多个基因组基因座。举例来说,本文所述的方法可被用于筛选对于感兴趣的毒素(例如白喉毒素)的抗性或敏感性重要的基因或其调节元件。在另一实例中,本文所述的方法可用于鉴定对于感兴趣的蛋白质(例如CD81)的表达重要的调节元件。在另一实例中,本文所述的

方法可在高通量筛选方法中用于通过筛选在某些条件(例如随时间推移的药物治疗)下在细胞群中过度表现或表现不足的gRNA来鉴定细胞类型中的必需或非必需基因。其它应用可由本领域技术人员确定。

[0125] 上面的公开总体地描述了本申请。通过参考以下具体实施例可以得到更完整的理解。这些实施例仅出于说明的目的而描述,并且不意图限制本公开的范围。考虑了形式改变和等效物替代,因为环境可能会暗示或出现应急情况。尽管本文中已使用了特定术语,但此类术语意图处于描述性含义,而不是用于限制的目的。

[0126] 以下非限制性实施例是对本发明的举例说明:

### [0127] III. 实施例

[0128] 实施例1. 用于鉴定人类细胞中的转录激活因子的平台。

[0129] 为了以系统的方式鉴定转录调节因子,使用了化学诱导二聚化(CID)系统,其中无催化活性的Cas9(dCas9)在其N末端用蛋白磷酸酶ABI1进行标记,并且潜在转录激活因子与脱落酸受体PYL1融合(图1A)(Gao等人,2016;Liang等人,2011)。用脱落酸(ABA)处理细胞会诱导ABI1与PYL1之间的相互作用,从而将潜在的转录激活因子募集至由指导RNA(gRNA)限定的特定基因组基因座。作为报告基因,使用含有稳定整合构建体的HEK293T细胞系,所述构建体具有7xTetO阵列和驱动EGFP表达的基础CMV启动子(Gao等人,2016)。通过慢病毒感染产生表达ABI1-dCas9和靶向TetO的单一gRNA的单克隆细胞系。用与PYL1融合的已知转录激活因子VPR、VP64和p300转染此细胞系导致在ABA处理后强烈诱导GFP(图1B),这与之前的报告一致(Gao等人,2016)。将荧光素酶或RFP募集至启动子不会诱导GFP表达,表明单独的ABA不会激活报告基因(图1B)。

[0130] 为了扩大单个克隆的规模,在含有C末端PYL1的慢病毒载体中生成了人类ORFeome 8.1和ORFeome协作克隆的合并文库(ORFeome Collaboration,2016;Yang等人,2011)。这些开放阅读框(ORF)文库总共含有14,821个克隆,对应于13,571个独特基因。报告细胞系以低感染复数被合并文库感染,从而确保大多数细胞仅被一种慢病毒感染(图1C)。合并文库在感染前后分别覆盖96%和93%的ORFeome,表明尽管插入物大小存在显著差异,但文库的覆盖范围很广(图6A和B)。用ABA处理受感染的细胞后,在细胞子集中观察到GFP的明显诱导(图1D)。报告基因诱导取决于处理时间、ABA浓度和ORFeome的存在(图6C-6E)。此外,ABA的撤回导致高GFP群体的快速损失(图6C),表明激活需要持续的启动子占据。为了鉴定转录激活因子,将ABA处理细胞的前1%GFP阳性细胞一式两份分选,并通过测序鉴定高GFP群体中的ORF(图1D)。

[0131] 方法如实施例7中所述。

[0132] 实施例2. ORFeome范围的筛选鉴定已知和新颖的转录激活因子。

[0133] 使用5%错误发现率的截止值以及前1%GFP阳性细胞与未分选细胞之间的读段计数至少4倍的变化,鉴定了248个假定的转录激活因子(图1E)。基因本体(GO)分析揭示了与转录激活相关的多个功能类别的显著富集(图1F)。所述命中还高度富集于许多转录调节因子(图1G)、染色质相关蛋白复合物亚基(图1H)以及转录中枢相互作用因子,例如RNA聚合酶II和组蛋白乙酰转移酶CBP和p300(图6F)中发现的蛋白域中。此外,所述命中与在酵母双杂交测定中充当自动激活因子的人类蛋白质显著重叠(蛋白质组中的11%相对于命中当中的29%; $p < 0.0001$ ,费歇尔精确检验;图6G),所述自动激活因子是当异位募集至启动子时激活

酵母中的报告基因表达的蛋白质 (Luck等人, 2020)。筛选结果还通过将90个命中和非命中的集合单独转染至相同的报告细胞系中进行测定来验证。结果与筛选高度一致: 单独测试时几乎所有命中均得到再现, 相反, 大多数非命中并未激活报告基因 (图6H), 表明在此设置中假阳性和假阴性率较低。总之, 这些结果表明筛选以可再现的方式鉴定了功能相关的转录激活因子。

[0134] 单个筛选结果包括调节不同转录阶段的特征明确的因子 (图1E)。例如, 除了若干调节应激反应的主TF (例如HSF1、ATF6和DDIT3/CHOP (Vihervaara等人, 2018)) 之外, 序列特异性转录因子命中当中还包括已知的激活因子RELB和MYCL (Barrett等人, 1992; Ryseck等人, 1992)。本身不结合DNA但与TF相关的辅激活因子包括STAT2、CITED1和SERTAD1 (Boussoik和Montazeri Aliabadi, 2018; Hsu等人, 2001; Yahata等人, 2001)。TFIIE的两个亚基 (GTF2E1和GTF2E2) 都是调节前起始复合物组装的通用转录因子 (Cramer, 2019), 也是突出的命中, 促进RNA聚合酶II释放的蛋白质 (P-TEFb复合物亚基CDK9) 和促进转录延伸的蛋白质 (例如ELL3和MLLT1) 也如此 (Chen等人, 2018; Cramer, 2019)。其它突出的激活因子包括染色质修饰复合物的亚基, 例如SAGA (ATXN7L3、TADA3、SGF29) 和NuA4 (EPC1、MGRBP), 以及介导DNA去甲基化的GADD45家族蛋白 (GADD45A、GADD45B、GADD45G) (Barreto等人, 2007)。最后, 鉴定了文库中存在的15个介体亚基中的13者。这些突出的实例说明, 筛选发现了通过多种不同机制并在转录周期的不同阶段促进转录的因子。

[0135] 除了已知的转录调节因子外, 筛选中还鉴定出大量先前未与转录相关的蛋白质 (例如C11orf74/IFTAP、NCKIPSD、DCAF7、HFM1) 或完全未表征的蛋白质 (例如C3orf62、FAM90A1、SPDYE4、SS18L2、FAM9A、C21orf58) (图1H)。这些结果表明人类基因组编码许多先前未知的转录调节因子。下面描述了其中若干因子的详细特征。

[0136] 方法如实施例7中所述。

[0137] 实施例3. TF家族内独特的转录活性。

[0138] 转录因子包括多个特征在于其独特的DNA结合域和辅助域的家族 (Lambert等人, 2018)。属于同一家族的转录因子通常具有高度相似甚至相同的序列特异性 (Jolma等人, 2013; Lambert等人, 2018; Weirauch等人, 2014)。然而, 即使是高度相关的TF, 由于其独特的辅助域, 也可能对转录和染色质产生不同的影响。与此一致, 只有转录因子家族的一些成员被鉴定为合并筛选中的命中。为了排除合并方法的敏感性作为原因, 对叉头盒 (FOX)、SRY相关的HMG-盒 (SOX)、E-二十六 (ETS)、无调性相关碱性螺旋-环-螺旋 (bHLH)、Twist/Hand、Krüppel样因子 (KLF) 和同源盒蛋白 (HOX) 家族的成员均进行了单独测定 (图2A和7A)。还包括两个与转录和染色质相关的蛋白质家族 (多梳家族环指 (PCGF) 和酪蛋白激酶家族)、一个先前未与转录相关的家族 (Spy1/RINGO) (Gastwirt等人, 2007) 以及24个在进化上不相关, 但为同一保守复合物的组成部分的介体亚基。

[0139] 与合并筛选一致, 许多高度同源的转录因子的激活谱即使一次测试一个也有显著不同。例如, 37个叉头TF中仅有5个、14个SOX中有1个、14个KLF中有5个以及36个HOX蛋白中有2个激活报告基因表达 (图2A-2D和7)。为了排除活性差异是由表达引起的这种简单解释, 将若干TF作为RFP-PYL1融合物进行测定, 以解释融合蛋白表达水平。结果与仅PYL1的融合物高度一致, 并且与RFP水平不相关 (图8), 表明转录活性的差异反映了TF的内在潜力。

[0140] 值得注意的是, 筛选中鉴定的激活TF显著富集了异位表达时可诱导小鼠胚胎干细

胞和人类诱导多能细胞(iPSC)分化的因子(图7G)(Ng等人,2021;Theodorou等人,2009),表明测定中的激活差异与生物功能的差异有关。例如,在相关的bHLH家族转录因子中,NEUROG1、NEUROG2、NEUROD1和NEUROD2可诱导iPSC的神经元分化,而NEUROD6则不能(Goparaju等人,2017)。这种模式与其激活报告基因的能力一致(图7B)。类似地,只有四种HOX因子(HOXA1、HOXA2、HOXB1、HOXB2)可激活位于Hoxb1基因座中的b1-ARE自动调节元件(Di Rocco等人,1997),并且其中三个在测定中被表征为激活因子(合并筛选和阵列测定中的HOXA2和HOXB2,筛选中的HOXA1)(图2A)。此外,最近的工作揭示了HOX基因的阻遏潜力、表达模式和基因组位置之间存在惊人的共线性(Tycko等人,2020),其中同源盒簇的5'端的HOX转录因子具有阻遏作用。与此一致,测定中的HOX家族激活因子由其HOX基因簇中的最3'或倒数第二个3'基因编码。

[0141] 一个特别有趣的例子是PCGF家族蛋白,其为规范和非规范多梳阻遏复合物1(PRC1)的互斥组分(Gahan等人,2020;Gao等人,2012)(图2E)。尽管普遍认为PCGF3在PRC1的背景下在染色质压缩和基因沉默中起作用,但PCGF3在最初的筛选中被鉴定为激活因子。当单独测试时,PCGF3强烈激活报告基因,PCGF5(原始合并筛选中不存在)也是如此(图2E)。相比之下,其它三个PCGF家族成员(PCGF1、PCGF2和PCGF4/BMI1)在单独测试时既不是筛选命中也不激活报告基因(图2E)。PCGF5先前已被证明可调节转录激活(Gao等人,2014),但尚未描述PCGF3的这种作用。有趣的是,在PCGF家族系统发育中,PCGF3和PCGF5形成一个在动物进化早期出现的独特群体(Gahan等人,2020),表明其转录激活功能具有古老的起源。

[0142] 一些研究表明,介体复合物的亚基可具有不同的调节功能,一些亚基促进转录激活,而另一些亚基促进阻遏(Conaway和Conaway,2011;Stampfel等人,2015)。在测定中,24个测定的介体亚基中有20个强烈激活报告基因,在介体子模块之间没有差异(图7F)。例如,MED29被认为具有转录阻遏活性(Wang等人,2004),但其为合并筛选中的命中,并且在单独测试时被验证为激活因子(图7F)。因此,至少在报告系统的背景下,几乎所有介体亚基都促进转录激活。

[0143] 初步激活筛选鉴定了两种蛋白(SPDYE4和SPDYE7P),其属于细胞周期调节因子Spy1/RINGO(卵母细胞G2/M进展的快速诱导因子)家族。Spy1/RINGO蛋白以不依赖于细胞周期蛋白的方式结合并激活Cdk1和Cdk2,从而促进细胞周期进展(Gonzalez和Nebreda,2020)。然而,其先前并未涉及转录调节。由于最近的扩展(Chauhan等人,2012),人类基因组包含至少19个Spy1/RINGO家族基因和多个假基因。分别测试了五个Spy1/RINGO蛋白的转录激活情况,并且其中四个强烈激活了报告基因(图2F)。这些结果表明,Spy1/RINGO蛋白可能部分通过充当转录激活因子来促进细胞周期进展。

[0144] 方法如实施例7中所述。

[0145] 实施例4. TAD-seq揭示了新型人类反式激活域。

[0146] 转录因子通常通过与辅激活因子(例如介体、CBP/p300乙酰转移酶或TFIID)相互作用的反式激活域(TAD)来激活转录。大多数TAD是富含酸性和疏水性残基的短的、非结构化的序列(Sigler,1988)。尽管付出了巨大的努力,但尚未出现明确的共识主题,使得TAD的计算预测具有挑战性(Eri jman等人,2020;Ravarani等人,2018;Stal ler等人,2021)。几个小组最近实施了合并募集方法,以从已知的转录调节因子或随机序列中鉴定TAD(Arnold等人,2018;Eri jman等人,2020;Ravarani等人,2018;Sanborn等人,2020)。修改上述筛选平

台以鉴定命中当中负责转录激活的区域。此方法与之前发布的TAD-seq方法(Arnold等人, 2018)类似,不同之处在于使用合成片段而不是随机片段化DNA。

[0147] 生成了我们筛选中鉴定的75个激活因子的片段文库,每20个氨基酸使用60个氨基酸平铺,使得每个氨基酸都由三个不同的片段代表(图3A)。合并的片段文库与PYL1融合,并募集至原始ORFeome筛选中使用的相同GFP报告基因(图3A)。再次,在经片段文库感染的报告细胞子集中观察到ABA依赖性GFP表达(图9A)。将GFP阳性群体分为两个独立的重复,并通过测序对每个群体中的片段进行定量。然而,与ORFeome筛选相反,高GFP细胞(前1%)和中等GFP细胞(前2-5%)均经过分选以潜在地鉴定不同强度的反式激活因子。

[0148] 汇总方法揭示了39种不同蛋白质中的70个激活片段。如所预期,这些片段富含酸性和疏水性氨基酸,并且缺乏带正电(碱性)氨基酸(图9B),表明许多片段代表与“酸斑点”想法一致的“规范”TAD(Sigler,1988)。事实上,基于两种不同的算法,激活片段包含更多的预测TAD(Erijman等人,2020;Piskacek等人,2007)(图9C),并且激活片段中的预测TAD比非活性片段中的预测TAD更长(图9D)。然而,许多活性片段无法通过任何算法预测,凸显了实验方法的必要性。

[0149] 在70个活性片段中,大多数(44=63%)仅在高GFP或中等GFP群体中被鉴定,表明所述片段具有不同的激活潜力(图9E)。例如,与未分选的片段文库相比,VP64在高GFP群体中富集,但在中等GFP群体中耗尽(图9E)。这与VP16在转录激活中的异常效力一致(VP64由四个串联的VP16肽组成)(Sadowski等人,1988)。对筛选中鉴定的10个片段进行了单独测试,其中9个片段强烈激活了报告基因(图9F),表明假阳性率较低。

[0150] 片段筛选回收了若干已知TAD。例如,鉴定了KLF6、KLF7、KLF15、ATF6和CITED2中的已知TAD(图3B和10A)。另外,由于活性片段之间的重叠可精确定义活性所需的最小序列,因此先前鉴定的TAD(例如KLF6、KLF7和SERTAD2的那些)的活性所需的区域被缩小(图3B和10A)。在ORFeome范围的筛选中发现了规范转录调节因子和新型蛋白质中许多以前未知的TAD(图3C和10B)。在先前未表征的蛋白质SPDYE4、C3orf62、FAM22F和FAM90A1中发现了TAD(图3C)。有趣的是,SPDYE4中的新型TAD与其Spy1域相邻,所述域与Cdk2相互作用并激活Cdk2,表明潜在的转录活性和Cdk调节活性是由Speedy家族蛋白的不同域介导的(图3C和11A)(McGrath等人,2017)。此外,与Spy1/RINGO家族蛋白的转录激活谱一致,TAD区域在除SPDYC之外的所有家族成员中都是保守的,SPDYC是唯一在募集测定中无活性的Spy1/RINGO家族蛋白(图2F和11A)。

[0151] 有趣的是,一些未发现的TAD不具有典型反式激活域的特征。例如,H0XA2中激活转录的三个重叠片段跨越同源盒DNA结合域与触角足样六肽基序之间的聚丙氨酸伸长段(图3D和3E),所述基序与PBX1辅激活因子相互作用(Piper等人,1999)。然而,缺乏六肽基序的片段仍激活转录,表明所述活性不通过PBX1调节(图3E)。聚丙氨酸伸长段在果蝇的Ultrabithorax(Ubx)蛋白中具有重要的功能,并且最近其被认为在驱动例如H0XD13的许多TF的相分离中发挥作用(Basu等人,2020)。本文的结果表明,聚丙氨酸伸长段可能在转录激活中具有额外作用。

[0152] 另一个非规范激活区域来自YAF2,它是多梳阻遏复合物1(PRC1)的组分(Gao等人,2012),并且是原始ORFeome筛选中的一个突出命中(图3F)。此区域含有YAF2\_RYBP域,所述域折叠成结合RING1B(PRC1核心亚基)的反平行 $\beta$ 片层(Wang等人,2010)(图3G)。YAF2的

YAF2\_RYBP域和其密切同源物RYBP与CBX家族多梳蛋白中存在的CBX-C域具有序列和结构同源性(Wang等人,2010)(图11B)。含有CBX-C域的蛋白质和YAF2/RYPB以互斥的方式与RING1A和RING1B相互作用,以形成规范(cPRC1)和非规范(ncPRC1)多梳复合物(图3G)。

[0153] 为了测试是否所有CBX-C和YAF2\_RYBP域均可促进转录激活,YAF2(SEQ ID NO:96)和RYBP(SEQ ID NO:140)的YAF2\_RYBP域以及CBX2(SEQ ID NO:136)、CBX4、CBX6(SEQ ID NO:138)、CBX7和CBX8(SEQ ID NO:139)的CBX-C域被克隆并用报告系统测定活性。来自YAF2和RYBP的YAF2\_RYBP基序强烈激活报告基因,与YAF2的TADseq结果一致(图3H)。一些CBX-C域也是激活因子:来自CBX2的CBX-C是最有效的激活因子,而来自CBX6和CBX8的CBX-C对报告基因的激活较弱(图3H)。相比之下,CBX4和CBX7 CBX-C域对报告基因没有影响。这种模式反映了CBX-C域蛋白的进化祖先(图3H)。有趣的是,CBX4和CBX7的CBX-C域与RING1B的结合亲和力比激活转录的CBX蛋白的相同域或RYBP的YAF2\_RYBP域高约10倍( $p=0.008$ ,双尾t检验)(图3H)(Wang等人,2008,2010)。因此,CBX-C和YAF2\_RYBP域的转录活性的差异可能通过其对RING1B的结合亲和力或与其它因子的差异结合来解释。无论如何,这些结果表明CBX蛋白功能的差异(Morey等人,2012;Vincenz和Kerppola,2008)至少可部分地通过其CBX-C域的内在差异来解释。更广泛地说,这些结果揭示了非规范和规范PRC2复合物的组装和功能的另一层复杂性。

[0154] 方法如实施例7中所述。

[0155] 实施例5.新型转录激活因子与已知辅因子相互作用。

[0156] ORFeome范围的筛选揭示了若干有效的反式激活因子,所述反式激活因子的特征不明或完全无特征。为了解这些因子如何调节转录,建立了稳定的四环素诱导型HEK293细胞系,其表达九种特征不明的筛选命中(C3orf62、C11orf74/IFTAP、NCKIPSD、DCAF7、SS18L2、SPDY4、FAM90A1、FAM22F/NUTM2F、JAZF1)、五种已知的转录调节因子命中(SOX7、KLF6、KLF15、CTBP1、HOXA2和HOXB2)、两种合成反式激活因子(VP64和VPR)以及与来自风产液菌(*Aquifex aeolicus*)的生物素连接酶BirA和FLAG表位标签融合的阴性对照(EGFP和Nanoluc)。其蛋白质相互作用组通过亲和纯化联合质谱(AP-MS)和具有邻近依赖性生物素化(BioID2)的邻近配偶体进行表征(Kim等人,2016)。AP-MS是表征稳定蛋白质复合物的理想方法,而BioID擅长鉴定较弱或涉及难溶性蛋白质,例如与染色质紧密结合的蛋白质的相互作用(Lambert等人,2015)。

[0157] 转录激活因子的相互作用组揭示了两种模式。首先,他们指出,新命中的反式激活潜力可能反映其内源功能,而不是系连测定的假象。其次,相互作用组揭示了激活因子对特定辅助激活因子复合物的显著偏好,集中于五种不同的辅因子(CBP/p300、BAF、NuA4、介体和TFIID)上。

[0158] 支持新型激活因子在转录调节中的天然作用,九个特征不明的命中当中有八个与AP-MS、BioID或两者中已知的转录辅因子相关(图4A、图12)。C3orf62、DCAF7和FAM22F/NUTM2F与作为已知转录辅助激活因子的p300和/或CBP相互作用。相比之下,JAZF1与BioID中NuA4组蛋白乙酰转移酶复合物的多个亚基相关,表明它是这种高度保守复合物的新亚基。SS18L2反过来与BAF染色质重塑复合物相互作用,包括核心BAF成员SMARCA2和SMARCA4以及规范BAF(cBAF)和非规范BAF(ncBAF)特有的亚基(图4A和12)(Centore等人,2020)。SPDY4和FAM90A1与调节转录延伸的BET家族溴域蛋白BRD2、BRD3和BRD4相互作用

(Fujisawa和Filippakopoulos,2017)。NCKIPSD(也称为SPIN90)与运动神经元存活(SMN)复合物相互作用,所述复合物调节核糖核蛋白复合物的组装,但也与转录激活有关(Pellizzoni等人,2001;Singh等人,2017;Strasswimmer等人,1999)(图12)。有趣的是,NCKIPSD还与DCAF7相互作用(图12B)。DCAF7和NCKIPSD都与细胞质中肌动蛋白动力学的调节有关(Cao等人,2020;Morita等人,2006),表明这些蛋白质可能在细胞质和细胞核中具有不同的作用。

[0159] 已知的转录调节因子也与辅激活因子复合物相关(图4A和12)。KLF6诱饵将NuA4亚基EPC1鉴定为邻近相互作用因子,而HOXB2、KLF15、CTBP1和SOX7诱饵将CBP和p300作为邻近配偶体。另外,SOX7还与BAF亚基相关。与除SOX7之外的所有天然激活因子相比,强大的合成激活因子VP64和VPR将多种不同的辅激活因子鉴定为邻近配偶体。VP64由病毒VP16反式激活基序的四个串联拷贝组成,与CBP/p300和介体亚基MED14和MED15相关,与先前的报道一致(Kundu等人,2000;Yang等人,2004)。VPR是VP64、人类p65/RELA反式激活域和爱泼斯坦-巴尔病毒(Epstein-Barr virus)R反式激活蛋白的融合物(Chavez等人,2015),与更多辅因子相关,包括CBP/p300以及介体和TFIID的多个亚基(图4A和12A)。与多种辅因子的此类关联可能解释了当VPR与dCas9融合时出色的反式激活效力(Chavez等人,2015,2016)。

[0160] 这些结果表明激活转录因子和其它转录调节因子对特定辅因子具有强烈的内在偏好。为了进一步研究这一点,对先前发布的叉头家族TF的AP-MS相互作用数据集(Li等人,2015)进行了分析,并与此测定中的转录激活结果进行了比较。值得注意的是,只有那些在募集至报告基因时激活转录的叉头转录因子才能与AP-MS中的辅激活因子相互作用(图4B)。另外,与质谱结果类似,激活叉头TF具有不同的辅因子偏好:FOXO1和FOXO3与CBP和p300特异性相互作用,FOXN1与p300和BAF复合物的亚基相互作用,而FOXR1和FOXR2更喜欢NuA4复合物(图4B)。这些数据强烈表明,鉴定高度相似的序列并以几乎相同的程度激活转录的相关转录因子可通过不同的辅激活因子复合物促进转录。

[0161] 为了在功能上研究转录激活因子与辅因子之间的关联,阵列了一组83个强大的激活因子。使用报告基因测定测试了其激活潜力,但现在存在针对多个转录辅调节因子的小分子抑制剂。采用三种靶向转录激酶的激酶抑制剂(针对CDK9的黄酮吡醇、针对酪蛋白激酶2的CX-4945以及针对DYRK1A和DYRK1B的AZ191),以及两种抑制转录辅因子的化合物(针对CBP/p300的A-485和针对BET家族溴域蛋白的JQ1)。

[0162] 所述三种激酶抑制剂几乎影响所有激活因子,尽管程度不同。用黄酮吡醇抑制CDK9导致所有激活因子的活性几乎完全丧失,与p-TEFb在启动子清除中的关键作用一致(图13A)。对调节转录延伸的酪蛋白激酶2的抑制(Basnet等人,2014)也导致转录激活的普遍但适度的减弱(图13B)。DYRK1A/DYRK1B抑制比其它两种化合物具有更微妙的影响,略微减弱了大多数激活因子的活性(图13C)。有趣的是,不受DYRK1A/DYRK1B抑制剂AZ191影响的两种激活因子是DCAF7,其与DYRK1A形成保守复合物(Breitkreutz等人,2010;Yu等人,2019),以及NCKIPSD,其与DCAF7相互作用(图13C)。

[0163] 与对转录具有广泛影响的激酶抑制剂相比,抑制CBP/p300的乙酰转移酶活性会强烈影响一些但不是所有反式激活因子的活性(图4C)。重要的是,这些效应与我们的AP-MS和BioID结果一致:六个CBP/p300相互作用因子中有四个的活性显著降低,而七个非相互作用因子中只有一个受到A-485的影响(图13D)。例如,A-485抑制KLF15的活性,而其对相关的

Kruppel样因子KLF6没有影响(图13D)。更广泛地说,与非相互作用的反式激活因子相比,A-485显著更强地抑制已知与CBP/p300相互作用的蛋白质的活性(图4C)。相比之下,NuA4复合物的相互作用因子并未受到CBP/p300抑制的显著影响(图4D和13D)。

[0164] 与CBP/p300抑制类似,JQ1对BET家族的抑制对某些反式激活因子有明显的影响。有趣的是,在大多数情况下,JQ1处理导致报告基因活性增加(图4D)。尽管JQ1处理在许多情况下会导致BRD4靶基因快速下调(Lovén等人,2013;Muhar等人,2018),但其也与报告基因的转录激活有关(Sdelci等人,2016),这可能解释了观察到的影响。然而,并非所有反式激活因子对JQ1处理的反应都相似。特别地,与NuA4复合物的亚基相互作用的因子不受JQ1处理的影响(例如JAZF1和KLF6;图4D和13D)。事实上,与CBP/p300相互作用因子或其它反式激活因子相比,NuA4相互作用因子受JQ1处理的影响明显较小(图4D)。尽管NuA4相互作用因子以独特的方式对JQ1处理作出反应的机制需要进一步研究,但这些结果证明了转录激活因子如何在单个启动子的背景下通过不同的辅助激活因子复合物促进转录。事实上,基于激活因子对五种化合物的敏感性的分层聚类揭示了多个不同的组(图4E)。许多旁系同源因子(例如CITED1和CITED2,或PCGF3和PCGF5)彼此相邻聚类,表明聚类产生了功能相关的基因。此外,已知的CBP/p300相互作用因子主要位于两个不同的簇中,NuA4相互作用因子也是如此(图4E)。这些簇的其它成员很可能类似地使用CBP/p300或NuA4作为辅激活因子,例如用于p300的YAF2或MYOG,或用于NuA4的NOM1。

[0165] 方法如实施例7中所述。

[0166] 实施例6.SRF-C3orf62融合物与CBP/p300相互作用并促进SRF/MRTF转录程序。

[0167] 涉及转录调节因子的融合蛋白是某些癌症,例如白血病和肉瘤的常见标志。我们的ORFeome范围筛选的命中显著丰富了COSMIC数据库(cancer.sanger.ac.uk)中记录的作为不同癌症融合配偶体的基因( $p=0.019$ ;超几何分布测试)。其中包括特征明确的融合配偶体,例如ERG,其在尤文氏肉瘤(Ewing sarcoma)中与EWSR1融合并且在前列腺癌中与TMPRSS2融合;DDIT3/CHOP,其在粘液样脂肪肉瘤中与EWSR1或FUS融合;CRIC1,其在粘液表皮样癌中与MAML2融合;和ENL/MLLT1,其在混合谱系白血病中与MLL融合。另外,文献中已将若干命中描述为融合配偶体,但未进行功能表征。例如,BTBD18和NCKIPSD被鉴定为KMT2A/MLL融合配偶体白血病(Alonso等人,2010;Sano等人,2000)。此外,与MLL融合的BTBD18片段含有TAD-seq鉴定的反式激活域(图10B),暗示TAD与融合产物的致癌潜力有关。大多数MLL融合物涉及调节转录延伸的基因,例如超级延伸复合物(Winters和Bernt,2017)。有趣的是,BTBD18也被证明可促进转录延伸(Zhou等人,2017)。

[0168] 为了更深入地了解激活筛选命中可能作为融合配偶体促进肿瘤发生的机制,选择了两个特征不明的融合物JAZF1-SUZ12和SRF-C3orf62进行进一步表征。JAZF1-SUZ12融合物是低级别子宫内膜间质肉瘤(LG-ESS)的标志(Hrzenjak,2016),将多梳蛋白SUZ12和JAZF1结合在一起(图5A)(Piunti等人,2019)。最近在一例肌纤维瘤/肌周细胞瘤儿科病例中描述了SRF-C3orf62(Antonescu等人,2017)。在此融合物中,血清反应因子(SRF)的DNA结合域与C3orf62的C末端融合(图5B)。值得注意的是,在这两种情况下,TAD-seq鉴定的反式激活域(图3C和10B)均保留在融合构建体中(图5A和5B)。

[0169] SUZ12、SRF、JAZF1-SUZ12、SRF-C3orf62以及SRF融合的C3orf62的C末端片段(C3orf62-Cterm)用BirA-FLAG进行标记,并使用BioID和AP-MS分析其相互作用组,以补充

我们为JAZF1和C3orf62获得的数据。如所预期,SUZ12邻近配偶体包括PRC2复合物的其它组分,例如EZH2、MTF2/PCL2和C10orf12 (Alekseyenko等人,2014) (图5C)。引人注目的是,JAZF1-SUZ12融合蛋白具有PRC2和NuA4亚基两者作为邻近配偶体(图5C),表明此融合物组装成两种染色质相关复合物的超级复合物,所述两种染色质相关复合物通常与相反的转录活性相关。与此一致,也与LG-ESS相关的EPC1-PHF1融合物类似地组装PRC2-NuA4超级复合物,从而导致多梳靶标的异常表达(Sudarshan等人,2021)。此外,JAZF1-SUZ12和EPC1-PHF1最近都被证明是强转录激活因子(Sudarshan等人,2021)。因此,NuA4整合至致癌超级复合物中可超越PRC2复合物的正常阻遏功能。

[0170] SRF邻近配偶体包括多种转录因子,例如ELK1,其与血清反应元件(SRE)上的SRF形成三元复合物(Buchwalter等人,2004) (图5D)。然而,它不与任何转录辅激活因子相关。相比之下,C3orf62-Cterm构建体和SRF-C3orf62融合物都强有力地将CBP和p300鉴定为邻近配偶体(图5D)。AP-MS将CBP和p300类似地鉴定为突出的SRF-C3orf62相互作用因子(图14A)。与这些结果一致,当与报告基因系连时,C3orf62-Cterm和SRF-C3orf62都强烈激活GFP报告基因(图5E)。此活性对A-485的CBP/p300抑制高度敏感(图5F),与SRF-C3orf62的相互作用模式一致。

[0171] SRF与三元复合因子(TCF;例如ELK1)或心肌素相关转录因子(MRTF;例如MAL)一起发挥作用,以调节靶基因表达(Buchwalter等人,2004;Olson和Nordheim,2010)。结果表明,SRF-C3orf62可在没有此类辅因子的情况下激活SRF靶基因。为了进一步测试这一点,使用基于荧光素酶的血清反应元件报告基因在NIH3T3成纤维细胞中测定了不同构建体的活性(Vartiainen等人,2007)。单独的SRF或C3orf62不会激活报告基因,而SRF-C3orf62强有力地激活报告基因(图5G),进一步证实与癌症相关的融合物绕过了转录激活中对SRF辅因子的需求。

[0172] 受MAP激酶信号传导调节的SRF/TCF途径调节立即早期基因的表达(Gualdrini等人,2016),而肌动蛋白-Rho信号传导依赖的SRF/MRTF途径靶向参与细胞运动和粘附的基因(Miralles等人,2003)。为了测试SRF-C3orf62是否可调节这两种途径的靶基因,生成了稳定表达SRF、C3orf62、SRF-C3orf62和与GFP融合的Nanoluc的稳定多西环素诱导型NIH3T3细胞系。用多西环素诱导转基因表达24小时,并通过RNA-seq分析基因表达的变化。虽然与Nanoluc相比,GFP标记的C3orf62或SRF的表达对转录组的影响非常有限或无影响(图14B),但SRF-C3orf62-GFP的表达导致564个基因上调和471个基因下调( $\log_2$ 倍数变化 $>1$ ,FDR $<0.05$ ;图5H和14B)。基因集富集分析(GSEA)和基因本体分析显示,上调基因高度富集于涉及DNA复制、细胞周期进展和有丝分裂的基因(图5I和14C),而下调基因与细胞外基质相关(图14C)。上调基因在E2F转录因子(细胞增殖的关键调节因子)的靶标中也显著富集(图5I)。这些特征与SRF-C3orf62融合物的致癌性质一致。另外,许多上调最多的基因与肌动蛋白动力学和肌生成有关。例如,上调最多的基因之一是平滑肌肌动蛋白(ACTA2),它是SRF/MRTF的已知靶标,并且是SRF融合阳性肌纤维瘤/肌周细胞瘤的标志(Antonescu等人,2017) (图5H)。与此一致,上调基因的另一显著特征是肌生成(图5I)。相反,许多特征明确的SRF/TCF立即早期靶基因,例如FOS、EGR1或EGR2不受异位SRF-C3orf62表达的影响(图5H)。有趣的是,SRF与VP16的融合可有效上调FOS,表明不同TAD激活靶基因表达的能力存在内在差异(Schratt等人,2002)。事实上,SRF-C3orf62上调的基因与先前报道的SRF/MRTF的靶基因之

间存在显著重叠,但与SRF/TCF的靶基因不存在显著重叠(图14D)(Esnault等人,2014;Gualdrini等人,2016)。综上所述,这些结果表明,SRF-C3orf62表达导致增殖和肌源性基因表达特征,并且SRF/MRTF靶基因比SRF/TCF靶标优先表达。

[0173] 方法如实施例7中所述。

[0174] 实施例7.方法:

[0175] 细胞培养。将包括用于筛选的pTRE3G-EGFP报告基因细胞系(来自斯坦福大学Lei Stanley Qi实验室的礼物)在内的所有HEK293T细胞维持在含有10%胎牛血清(FBS)的DMEM中。NIH-3T3细胞获自Sachdev Sidhu博士的实验室(多伦多大学),并保存在含有10%小牛血清(BCS)的达尔伯克改良伊格尔培养基(DMEM)中。所有培养基均补充有1%青霉素-链霉素。将细胞维持在37℃、5%CO<sub>2</sub>的加湿培养箱中,并常规测试支原体污染。

[0176] 慢病毒生产。通过用pLX301-ORFs/TADs-PYL1、psPAX2(Addgene#12260)和pVSV-G(Addgene#8454)以8:6:1的比率转染293T细胞,产生了含有合并ORFeome和反式激活文库的慢病毒颗粒。根据制造商的方案,使用XtremeGENE 9(Roche)在15cm培养皿上进行转染。转染后6-8小时将培养基更换为收获培养基(DMEM+1.1g/100mL BSA)。转染后72小时,上清液经过滤(0.45μm),合并并收集。当建立单个稳定细胞系时,小规模病毒生产遵循类似的方案,其中使用Lipofectamine 2000(Thermo Fisher Scientific,11668019)试剂在6孔板上进行转染。

[0177] 细胞系生成。生成表达ABI-dCas9(blastacidin,6μg/mL)和靶向pTRE3G启动子的gRNA(SEQ ID NO:10)(共表达EBFP2)的EGFP报告基因系的克隆系。对单细胞进行分选(FACS Aria IIIu,BD)并扩增,并选择显示出被强转录激活因子诱导的克隆用于后续实验。为了生成表达多西环素诱导型EGFP标记蛋白的NIH3T3细胞,从hORFeome集合中挑选入门克隆,并亚克隆至Gateway相容的pSTV6-Tet0-ccdB-EGFP慢病毒质粒(Payman Samavarchi-Tehrani的善意礼物)中。NIH-3T3细胞在8μg/mL聚凝胺存在下被感染,并在感染后24小时用2μg/mL嘌呤霉素进行选择。

[0178] 合并ORFeome文库生成。来自人类ORFeome集合(v8.1)的入门克隆被收集至40个标准化子池中,每个子池含有约384个ORF,并克隆至慢病毒Gateway相容的目的载体pLX301-DEST-PYL1中。LR反应一式两份进行,每个入门ORF子池各150ng,与1μl Gateway LR克隆酶II混合,反应体积总计5μl,并在室温下在TE缓冲液中培育过夜。接下来的两天,向每个反应中添加含1μl额外LR酶的4μl TE和150ng目的载体。将菌落转化为化学感受态DH5α大肠杆菌,并在30℃下铺展在含有羧苄青霉素(100μg/μl)的LB琼脂平板上过夜。对菌落进行计数以确保>200倍的覆盖度,将其收集在冰上的SOC中,沉淀并根据干沉淀的重量在多个柱上进行大规模制备。

[0179] 激活域平铺文库生成。平铺文库由ORFeome筛选中鉴定为激活因子的75种蛋白质生成。含有5' 衔接子GGAAGTCAGGGTAGCGGAAGTATG(SEQ ID NO:23)和3' 衔接子GGAGGTAGTGTGAACGCGAAGGC(SEQ ID NO:24)以生成5' 衔接子(GSQGSGSM)(SEQ ID NO:11)和3' 衔接子(GGSVEREG)(SEQ ID NO:12)的寡核苷酸被合成为合并文库(Twist Biosciences)。使用NEBNext Ultra II Q5主混合物(New England Biolabs)以5nM寡核苷酸作为模板建立6×50μl PCR反应。PCR条件经过优化,以找到具有预期300bp长度的干净可见产物的最低循环。热循环条件为初始98℃持续30秒,接着是98℃持续10秒、63℃持续20秒和

72℃持续15秒的2个循环,随后再进行10个98℃持续10秒和72℃持续30秒的循环,最后在72℃下延伸5分钟。引物被设计为具有与Gateway相容的侧接序列。所得文库在以60V负载于2%TAE凝胶上2小时后,通过QIAgen凝胶提取试剂盒进行凝胶提取,且随后使用20个单独的BP反应(总共5μl反应)克隆至pDONR221中。过夜反应后,将入门质粒池转化至DH5α感受态大肠杆菌中,并在含有卡那霉素(100μg/μl)的LB琼脂板上培育过夜。收集菌落并纯化质粒DNA。接着按照上一节所述建立20个LR反应。将每个反应转化至NEB 10-β化学感受态大肠杆菌中,并在30℃下在含有羧苄青霉素(100μg/μl)的LB琼脂板上生长过夜。对菌落进行计数,以确保克隆、合并和大规模制备的每个步骤的覆盖率>200倍。

[0180] 合并激活筛选。ORFeome和C末端标记有PYL1的反式激活平铺文库被包装成慢病毒颗粒。稳定共表达ABI-dCas9的克隆EGFP报告细胞系和靶向启动子的gRN以低感染复数(MOI)转导,在嘌呤霉素(1μg/mL)选择后,细胞存活率大约为30%。相同条件下未转导的细胞被完全消除。转导足够的细胞以维持文库的>500倍覆盖率。通过用100μM脱落酸(ABA, Sigma)处理细胞48小时来诱导募集。同时,用等总体积的DMSO处理对照批次的细胞。接着用PBS洗涤细胞,用解离缓冲液(1mM EDTA、10mM KCl、150mM NaCl、5mM碳酸氢钠、0.1%葡萄糖)处理,并再悬浮于流动缓冲液(5mM EDTA、25mM HEPES pH 7、1%BSA、PBS)中。对每个文库的高GFP群体(ORFeome的前1%以及TAD筛选的前1%和接下来的4%的两个箱)进行分选,并使用QIAmp DNA Blood Mini Kit(QIAGEN)直接提取其基因组DNA。

[0181] ORFeome测序。使用来自分选群体的所有纯化基因组DNA或来自预分选群体的至少5μg基因组DNA进行巢式PCR。使用靶向T7启动子和PYL1的引物从基因组DNA中扩增目标ORFeome区域。汇集每个样品的此反应产物,并通过靶向Gateway attB位点之外的引物进一步扩增另外10个循环。随后在1%琼脂糖凝胶上分离扩增子,并对除引物二聚体之外的任何可见PCR产物进行凝胶纯化。使用Quant-iT 1X dsDNA HS试剂盒(Thermo Fisher Scientific, Q33232)对DNA进行定量后,使用Illumina DNA Prep, (M) Tagmentation试剂盒(Illumina, 20018705)对每个样品50ng进行处理,并进行6个扩增循环。每个纯化的最终文库取2μl在Agilent TapeStation HS D1000 ScreenTape (Agilent Technologies, 5067-5584)上运行。使用Quant-iT 1X dsDNA HS试剂盒(Thermo Fisher Scientific, Q33232)对文库进行定量,并在大小调整后以等摩尔比合并。使用Illumina的NEBNext Library Quant Kit(New England Biolabs, E7630L)对最终池进行定量,并在Illumina MiSeq上进行配对末端测序。

[0182] TAD测序。在第一步中使用靶向骨架载体的T7启动子和PYL1的引物对纯化的基因组DNA进行巢式PCR,产生约470bp的产物。接着汇集第一反应的产物,并使用靶向Gateway位点之外的引物进行另外10个步骤的扩增,产生约300bp产物。在Qubit dsDNA Broad Range试剂盒上对文库进行定量,并在Illumina MiSeq上使用定制PAGE纯化的R1测序引物进行配对末端测序。

[0183] 对合并激活筛选的测序数据进行的分析。ORFeome参考序列的索引是使用STAR比对器v2.7.8a创建的,预索引字符串的长度设置为11,以考虑较小的“基因组”大小。来自ORFeome文库的读段用STAR比对器比对,允许最多3个不匹配。对于TAD测序读段,首先使用cutadapt从两端去除克隆衔接子序列,对于5'衔接子使用CCAGTGTGGTGAATTCTGCAGATATCAACAAGTTTGTACAAAAAAGTTGGCGGAAGTCAGGGTAGCGGAAGT (SEQ ID NO:20),并且对于3'衔接子

使用CCGCCACTGTGCTGGATATCAACCACTTTGTACAAGAAAGTTGGGTAGCCTTCGCGTTC AACACTACCTCC (SEQ ID NO:21)。生成Bowtie参考,并使用Bowtie v1.2.3映射读段,允许0个不匹配。为了鉴定激活因子,使用edgeR包(Robinson等人,2010)通过比较分选样品与未分选细胞的计数变化来计算每个ORF的log2倍数变化、p值和错误发现率(FDR)。

[0184] 阵列募集测定。将稳定共表达ABI-dCas9和TetO gRNA的报告细胞接种在48孔或96孔板(Sarstedt)上,以在转染当天达到50-70%铺满。将150ng待测试的每种构建体用聚乙烯亚胺(PEI)以0.6 $\mu$ l试剂的比率转染。转染后第二天,通过ABA(100 $\mu$ M)处理来诱导募集。对于抑制剂存在下的系连报告基因测定,转染后第二天用100 $\mu$ M ABA类似地诱导募集,但在抑制剂或相同体积的任何额外DMSO存在下进行。将抑制剂溶解于DMSO中至储备浓度为10mM。使用的最终浓度为:黄酮吡醇为100nM,JQ1为300nM,A-485为1 $\mu$ M,CX-4945为2.5 $\mu$ M,且AZ191为3 $\mu$ M。所有抑制剂都是结构基因组学联盟(SGC)的一份善意礼物。诱导后48小时,使用液体处理机器人(TECAN)将细胞解离并再悬浮于流动缓冲液中,并通过LSR Fortessa (BD) 进行分析。在高EBFP2和RFP上对细胞进行门控,分别作为gRNA和转染对照的量度。使用FlowJo (v10)分析流式细胞术数据。

[0185] SRF报告基因测定。用8ng SRF报告基因(p3DA.luc)、20ng参考报告基因(pcDNA3.1-Nanoluc-3xFLAG-V5)和50ng 3xFLAG标记构建体(Addgene#87063)转染24孔板上的30,000个NIH-3T3细胞。使用Lipofectamine 3000试剂(Thermo Fisher Scientific, L3000001)根据制造商的方案进行转染。荧光素酶构建体是Maria Vartiainen博士(University of Helsinki)的一份善意礼物。将细胞在低血清培养基(0.5%BCS)中维持18小时,并刺激7小时(15%BCS),然后测量荧光素酶活性。使用来自四次独立转染的数据将萤火虫荧光素酶标准化为海肾荧光素酶活性。

[0186] RNA测序和分析。用1 $\mu$ g/mL多西环素诱导NIH-3T3细胞24小时,所述细胞具有在C末端经EGFP标记的SRF、C3orf62、SRF-C3orf62或Nanoluc的稳定整合。使用RNeasy纯化试剂盒(Qiagen)从维持在低血清条件(0.5%小牛血清)下22小时的细胞中提取RNA,并在柱上用DNA酶处理。从6孔板中诱导并收集技术重复形式的样品。使用NEBNext Ultra II定向RNA-seq和Poly-A选择试剂盒制备文库,合并并且在100循环NovaSeq 6000SP上测序。使用STARaligner v2.7.8a将读段与Gencode小鼠初级装配(GRCm39)进行比对。使用Gencode vM26转录本注释生成每个基因的计数。使用edgeR包对与表达Nanoluc-EGFP的细胞相比基因表达的变化进行量化(Robinson等人,2010)。

[0187] 质谱分析样品。入门克隆来自人类ORFeome集合(Yang等人,2011)。使用Gateway重组酶将克隆转移至携带C末端BioID2-FLAG标签的pDEST-pcDNA5载体(Kim等人,2016)。如先前所报道地生成稳定的HEK293 Flp-In T-REx细胞系(Piette等人,2021)。

[0188] 对于AP-MS,细胞在150mm培养皿上生长至70%铺满,然后用1 $\mu$ g/mL四环素诱导诱导表达24小时。接着用1xPBS洗涤细胞一次,刮擦、沉淀、快速冷冻,并储存在-80℃下直至处理。AP-MS如先前所述进行(Lambert等人,2015)。简而言之,使用1:4沉淀重量:体积比将细胞再悬浮于冷裂解缓冲液(50mM HEPES-NaOH pH 8.0、100mM KCl、2mM EDTA、0.1%NP40、10%甘油、1mM PMSF、1mM DTT、15nM Calyculin A和蛋白酶抑制剂混合液(Sigma-Aldrich P8340))中。通过一轮冻融裂解细胞,并在4℃下使用35%振幅的三次10秒突发并暂停2秒而对裂解物进行超声处理。将超声处理的裂解物用100U全能核酸酶在4℃下处理30分钟,然后

通过在4℃下以20,000g离心20分钟进行澄清。将一批内处理的所有样品的等量上清液转移至含有25μL预洗抗FLAG磁性珠粒50%浆液(Sigma,M8823)的管中,并在4℃下培育两小时。通过磁化回收珠粒并丢弃上清液。将珠粒在裂解缓冲液中洗涤一次,并在含有2mM CaCl<sub>2</sub>的20mM Tris-HCl pH8.0中洗涤一次,并在两个阶段中用胰蛋白酶消化珠粒(1μg胰蛋白酶持续4小时,然后向上清液中添加0.5μg胰蛋白酶并在37℃下培育过夜),如前所述(Taipale等人,2014)。最后,将样品用5%甲酸(最终浓度)酸化并储存在-80℃下。

[0189] 对于BioID,细胞在150mm培养皿中生长至70%铺满,然后用1μg/mL四环素诱导基因表达18小时。接着将50μM生物素添加至每个板后维持6小时。按照AP-MS收集细胞沉淀,并使用1:10沉淀重量:体积比再悬浮于裂解缓冲液(50mM Tris-HCl pH 7.5、150mM NaCl、0.1%SDS、1%Igepal CA-630、1mM EDTA、1mM MgCl<sub>2</sub>、蛋白酶抑制剂混合液(Sigma-Aldrich P8340,1:500)和0.5%脱氧胆酸钠)中。超声处理后,每个样品用250U Turbonuclease (BioVision 9207-50KU)和1μL RNA酶A溶液(Sigma-Aldrich R6148)处理,并在4℃下培育30分钟。接着添加SDS至最终浓度为0.25%,并且在混合后,将样品在4℃下再培育10分钟,然后以20,000g离心20分钟。将上清液转移至含有30μl预洗包装链霉亲和素珠粒(GE Healthcare,17-5113-01)的管中。链霉亲和素下拉在4℃下进行3小时。将珠粒在1ml SDS缓冲液(2%SDS/50mM Tris-HCl pH7.5)中洗涤一次,在1ml裂解缓冲液中洗涤一次,并在TAP缓冲液(50mM HEPES-KOH pH 8.0、100mM KCl、10%甘油、2mM EDTA、0.1%Igepal CA-630)中洗涤一次,然后用50mM碳酸氢铵pH 8.0进行三次1ml洗涤。洗涤后,将珠粒再悬浮于含有1μg胰蛋白酶的ABC缓冲液中,并在37℃下培育过夜。第二天,收集上清液,并用50μl水洗涤链霉亲和素珠粒,将其与第一个上清液级分合并。将0.5μg胰蛋白酶添加至合并的上清液样品中,接着将其在37℃下培育4小时。接着将珠粒减速旋转并回收上清液。使用ABC缓冲液冲洗珠粒两次,并将这些冲洗液与原始上清液合并。通过离心蒸发干燥合并的上清液。

[0190] 质谱数据采集和分析。如前所述(Piette等人,2021),使用数据依赖型采集(DDA)在TripleTOF 5600仪器(AB SCIEX,Concord,Ontario,Canada)上对样品进行分析。如前所述(Piette等人,2021),使用ProHits v4.0(Liu等人,2016)、Mascot和Comet(Eng等人,2013)中实现的Proteowizard(Adusumilli和Mallick,2017)对数据进行处理和分析。随后使用iProphet(Shteynberg等人,2011)通过Trans-Proteomic Pipeline对结果进行分析,并对iProphet概率≥0.95的蛋白质进行进一步分析。

[0191] 使用SAINTexpress鉴定了重要的相互作用因子(Teo等人,2014)。EGFP-BioID2-FLAG和EGFP-BioID2-FLAG用作阴性对照,使用2倍压缩以达到严格性,如先前所述(Mellacheruvu等人,2013)。SAINTexpress分析使用默认参数,并且如果猎物蛋白符合≤5%的计算得出的贝叶斯FDR截止值,则被视为显著的。点图是使用ProHits-viz网络服务器生成的(Knight等人,2017)。

[0192] 实施例8.可组合多种TAD来激活转录

[0193] 将多种TAD与PYL1融合在一起,并使用上述PYL1-ABI EGFP报告系统测试转录活性。使用以下TAD片段:CITED1-8(SEQ ID NO:47)、CITED2-12(SEQ ID NO:49)、C3orf62-12(SEQ ID NO:45)、BRD8-25(SEQ ID NO:40)、ZXDC-12(SEQ ID NO:97)、KLF7-1(SEQ ID NO:72)、ATXN7L3-1(SEQ ID NO:35)、FAM90A1-20(SEQ ID NO:60)、SPDYE4-3(SEQ ID NO:90)、YAF2-7(SEQ ID NO:96)。

[0194] 如图16中所示,与单独的TAD相比,SPDYE4-CITED1 TAD融合蛋白使转录激活显著增加。

[0195] 从p65和HSF1中选出的至多2个另外的TAD与SPDYE4-CITED1 TAD以不同的顺序组合,并测试激活PYL1-ABI EGFP报告系统(图17,左图板)或内源CD133基因(图17,右图板)的转录的能力。SPDYE4-CITED1-P65-HSF1(SCPH)效应子在测试的两个系统中均提供最强大的激活。

[0196] 为了测试另外的TAD组合,将多组分SPDYE4-CITED1-P65-HSF1(SCPH)效应子的每个部分单独替换为实施例4中鉴定的另一TAD(参见例如表1)。结果在图18中示出。

[0197] 为了测试较短的TAD序列是否可在SCPH效应子中使用,每种TAD都被独立地替换为短或迷你TAD组分。结果在图15中示出。Mini-TAD序列选择如下:

[0198] miniSPDYE4(SEQ ID NO:102):基于来自我们的SPDYE4全长蛋白质平铺筛选的两个富集片段之间的重叠来选择序列。最小序列被设计为包括富含酸性区域或随后的 $\beta$ 发夹。

[0199] miniHSF1:HSF1的150个氨基酸的C末端区域,含有其激活域,所述激活域由两个已知的TAD构成。较长的TAD位于氨基酸431-529之间,并且较短的TAD(“mini-HSF1(401-420)”或“H(mini)”);SEQ ID NO:119)位于氨基酸401-420之间。我们选择富含疏水残基的较短域进行测试,因为其尺寸更紧凑,并且先前报道其相比于较长的TAD具有更强的效力。(Newton等人,1996;10.1128/MCB.16.3.839)

[0200] miniP65:已知RelA或p65在其C末端内有两个不同的反式激活域。第一种TAD(“P(mini N-term)”);SEQ ID NO:105)包含氨基酸428-520,并且第二个(“P(mini C-term)”);SEQ ID NO:106)是前面的521-551个残基。(Schmitz和Baeuerle,1991;10.1002/j.1460-2075.1991.tb04950.x)

[0201] 选择miniCITED1(“C(mini)”);SEQ ID NO:103)以包括CITED1-7与CITED1-8之间的重叠区域内的C末端富含酸性区域。

[0202] 方法:将细胞接种于48孔板上。第二天,在细胞达到70-90%铺满后,使用(Thermo Fisher Scientific,11668019)试剂将250ng的每种构建体转染至每个孔中。每个构建体直接融合TagRFP,或直接融合至每个组分,或从相同的质粒共表达,用作转染的量度。在每个实验中,RFP+细胞的相同门用于控制每个构建体的表达水平对活性的影响。EGFP报告细胞从克隆系扩增,以控制ABI-dCas9和靶向启动子上游TetO7位点的gRNA的水平。对于CD133诱导实验,所有激活因子均使用稳定表达ABI-dCas9的293T细胞系和靶向启动子的5个gRNA池。使用与APC结合的CD133抗体(Miltenyi Biotec,130-113-668)。所有激活因子构建体均在其C末端与PYL1融合,并通过用1 $\mu$ M脱落酸处理细胞24或48小时将其募集至其靶标。

[0203] 实施例9.

[0204] 另外117种不同的激活域组合或单个激活域的片段通过将其与PYL1(用dCas9-ABI1系统进行测试)或rTetR(一种在多西环素存在下结合DNA的转录因子)融合进行了测定。所有构建体均使用相同的TetO报告构建体进行测试。在这些构建体的活性中观察到很大的差异,范围从无活性至非常高的效力(图19A-B)。一般而言,基于rTetR和dCas9的募集系统的结果之间存在良好相关性(图19C),表明大多数激活域在两种情况下的功能相似。然而,例如,VPR作为rTetR融合物的活性远低于作为dCas9融合物的活性。与先前的实验一致,SCPH和其变体仍然是系统中最具活性的构建体。例如,含有C3orf62的小型化激活域而不是

CITED1的激活域(SCPH构建体中的“C”)和较短型式的p65激活域(“P”)的构建体比原始SCPH构建体更有效,尽管显著较小(1059bp相对于1611bp)(图19A-B)。

[0205] 方法:将200ng在C末端与rTetR融合的表达质粒的超级激活因子和50ng DsRed转染至稳定表达7xTetO-EGFP报告基因的HEK293T细胞中。转染一天后,用1 $\mu$ g/ml多西环素处理细胞48小时。处理后,通过流式细胞术分析细胞的EGFP表达

[0206] 尽管已参考目前被认为是优选的实施例对本申请进行了描述,但应理解,本申请不限于所公开的实施例。与此相反,本申请意图涵盖各种修改和包括在所附权利要求书的精神和范围内的等效布置。权利要求书的范围不应受优选实施例和实例限制,而应给予与整个说明书一致的最广泛的解释。

[0207] 所有出版物、专利和专利申请均以引用的方式整体并入本文中,其程度如同每个单独出版物、专利或专利申请被具体地和单独地指示为以引用的方式整体并入。

[0208] 序列表

[0209] 甘氨酸丝氨酸接头(SEQ ID NO:6)SGGSGGS

[0210] 甘氨酸丝氨酸接头(SEQ ID NO:7)SGGS

[0211] 甘氨酸丝氨酸接头(SEQ ID NO:8)GSGSGS

[0212] 接头(SEQ ID NO:9)INSRSSGS

[0213] NLS(SEQ ID NO:22)PKKKRKV

[0214] 连接子+NLS(SEQ ID NO:141)INSRSSGSPKKRKVGS

[0215] >YAF2的YAF2\_RYBP域(SEQ ID NO:147)

[0216] RPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKT

[0217] >RYBP的YAF2\_RYBP域(SEQ ID NO:148)

[0218] RPRLKNVDRSTAQQLAFTVGNTVTIITDFKEKT

[0219] >CBX2的CBX-C域(SEQ ID NO:149)

[0220] QDWKPTRSLIEHVFVTDVTANLITVTVKESPTSV

[0221] >CBX6的CBX-C域(SEQ ID NO:150)

[0222] GDWRPEMSPCSNVVVDVTSNLLTVTIKEFCNPE

[0223] >CBX8的CBX-C域(SEQ ID NO:151)

[0224] ESWSPSLTNLEKVVVDVTSNFLTVTIKESNTDQ

[0225] >CBX4的CBX-C域(SEQ ID NO:152)

[0226] SEFKPFFGNIIITDVTANCLTVTFKEYVTV

[0227] >CBX7的CBX-C域(SEQ ID NO:153)

[0228] PPWTPALPSSEVTVDITANSITVTFREAQAAE

[0229] >SPDYE4-2(SEQ ID NO:154)

[0230] VRSPEVVVDDEVPGPSAPWIDPSPQPSLGLKRKSEWSDESEEELEEELELERAPEPEDT

[0231] >SPDYE4-5(SEQ ID NO:155)

[0232] WVVETLCGLKMKLKRKRASSVLPEHHEAFNRLLGDPVVQKFLAWDKDLRVSDKYLLAMVI

[0233]

表 1: 通过 TADseq 鉴定为 TAD 的片段。标有\*的片段经过独立验证。

| 基因      | 片段             | 序列                                                                 | SEQ ID NO: | logFC 高 GFP     | FDR 高 GFP      | 命中高 GFP | logFC 中等 GFP    | FDR 中等 GFP      | 命中中等 GFP |
|---------|----------------|--------------------------------------------------------------------|------------|-----------------|----------------|---------|-----------------|-----------------|----------|
| ATF6    | ATF6-1         | MGEPAGVAGTMESPSPGLFHLRDEWDSDALFAE<br>LGYFTDTDELQLEAANETYENNFDNL    | 29         | 3.14527<br>6698 | 0.02121<br>147 | 命中      | 1.5632<br>96241 | 6.04E-0<br>6    | 命中       |
| ATF6    | ATF6-2*        | HRLEDDWSALFAELGYFTDTDELQLEAANETVE<br>NNFDNLDFDLVPWESDIWDINNQI      | 30         | 4.68756<br>189  | 2.94E-0<br>5   | 命中      | 1.4772<br>63549 | 4.41E-0<br>5    | 命中       |
| ATF6B   | ATF6B-1        | MAELMLLSEIADPTRFFTDNLLSPEDWGLQNSTL<br>YSGLDEVAEEQTQLFRCPEQDVFPDG   | 31         | 1.91689<br>6584 | 0.00665<br>441 | 命中      | 0.9930<br>94738 | 0.00102<br>4406 | 命中       |
| ATMIN   | ATMIN-2<br>4   | NPGPDTQLPSGPAQNPGIDFIEFFSASNIQTQ<br>TESELSTMTTEPVLESLDIETQDF       | 32         | 1.23264<br>9648 | 0.05344<br>855 |         | 0.9880<br>0755  | 0.00096<br>5273 | 命中       |
| ATMIN   | ATMIN-3<br>2   | ETQTMSSGFETLGSFFTSNETQTAMDDFLLADL<br>AWNTMESQFSSVETQTSAPHTVSNF     | 33         | 0.63551<br>575  | 0.30551<br>37  |         | 0.8197<br>59214 | 0.00905<br>0433 | 命中       |
| ATOH1   | ATOH1-1        | MSRLHAEWEAEVKELGDHHRQPQPHLPQPPPP<br>PQPPATLQAREHPVYPPELSLLDSTD     | 34         | 0.72182<br>2453 | 0.30506<br>874 |         | 0.9193<br>51522 | 0.01713<br>0359 | 命中       |
| ATXN7L3 | ATXN7L3<br>-1* | MKMEMSLSGLDNSKLEATAQEIYADLVEDSCLG<br>FCFEVHRAVKCGYFFLDLDDTPDSMKD   | 35         | 1.63790<br>1718 | 0.02573<br>555 | 命中      | 1.6838<br>93477 | 1.80E-0<br>5    | 命中       |
| ATXN7L3 | ATXN7L3<br>-2  | QEIYADLVEDSCLGFCFEVHRAVKCGYFFLDLDDTD<br>PDSMKDFEIVDQPGLDIFGQVFNQWK | 36         | 0.76543<br>1613 | 0.28306<br>469 |         | 0.8910<br>59001 | 0.00244<br>0708 | 命中       |
| ATXN7L3 | ATXN7L3<br>-4  | FEIVDQPGLDIFGQVFNQWKSKEVCPCNCSRSIA<br>ASRFAPHLEKCLGMGRNSSRIANRRI   | 37         | 0.76410<br>1911 | 0.20450<br>075 |         | 0.7514<br>39597 | 0.02241<br>0022 | 命中       |
| BRD8    | BRD8-23        | KEECFRSGVAEAPVGSKAPSIDGKEELDLAEKMD<br>IAVSYTGEELDFETVGDIIIIIEDKV   | 38         | 0.58842<br>0958 | 0.35787<br>782 |         | 0.5914<br>94577 | 0.03298<br>8887 | 命中       |

[0234]

|              |                 |                                                                   |    |                 |                |    |                  |                 |    |
|--------------|-----------------|-------------------------------------------------------------------|----|-----------------|----------------|----|------------------|-----------------|----|
| BRD8         | BRD8-24         | IDGKEELDIAEKMDIAVSYTGEELDFETVGDIIA<br>IIEDKVDDHPEVLDAAVEAALSFCE   | 39 | 1.59319<br>3215 | 0.04662<br>626 | 命中 | 1.5656<br>45971  | 2.69E-0<br>5    | 命中 |
| BRD8         | BRD8-25<br>*    | GEELDFETVGDIIAIIEDKVDDHPEVLDAAVEA<br>ALSFCEENDDPQSLPGWEHP IQER    | 40 | 1.56747<br>8812 | 0.02072<br>139 | 命中 | 1.8337<br>47802  | 6.04E-0<br>6    | 命中 |
| BTBD18       | BTBD18-<br>24   | IDCREPYAFDTALLEQCEAEERYITSAATSEL<br>EEILDFMLCGSDIEPPIGSLESPGAE    | 41 | 1.24617<br>6755 | 0.15033<br>388 |    | 0.8561<br>07889  | 0.00321<br>7538 | 命中 |
| BTBD18       | BTBD18-<br>25   | EEYRITSAATSELEBILDFMLCGSDIEPPIGSL<br>ESPGAEGCRTPTYHLTETGKNWIEGE   | 42 | 1.47063<br>5679 | 0.07002<br>876 |    | 1.4133<br>28465  | 7.30E-0<br>5    | 命中 |
| C11orf7<br>4 | C11orf7<br>4-1  | MSAHMSGLEIMDEDQLIKDVLDKFLNCHEQTYDE<br>EFLNTFTHLSQLLLLPGEVEQDVST   | 43 | 2.35635<br>6482 | 0.01211<br>4   | 命中 | 1.5512<br>34476  | 7.99E-0<br>6    | 命中 |
| C3orf62      | C3orf62<br>-11  | NHMGHCQDSPFKEEAWALLMDKSPQKATDADPG<br>SLKQAFDDHNIIVETVLDLEEDYNVMT  | 44 | 6.02271<br>451  | 7.93E-0<br>6   | 命中 | -0.526<br>078624 | 0.22147<br>5633 |    |
| C3orf62      | C3orf62<br>-12* | QDSPFKEEAWALLMDKSPQKATDADPGSLKQAFD<br>DHNIVETVLDLEEDYNVMTSFKYQIE  | 45 | 6.49295<br>0845 | 6.25E-0<br>6   | 命中 | -0.606<br>664208 | 0.12470<br>1907 |    |
| CITED1       | CITED1-<br>7    | LQNWDFGAQAGGAESLSPSAGAQSPAIIDS DPVD<br>EEVMSLVVELGLDRANELPELWLGQ  | 46 | 4.31381<br>114  | 8.10E-0<br>5   | 命中 | 1.5966<br>38925  | 0.00050<br>5651 | 命中 |
| CITED1       | CITED1-<br>8*   | ESLSPSAGAQSPAIIDS DPVDEEVMSLVVELGL<br>DRANELPELWLGQNEFDFTADFPSSC  | 47 | 6.82626<br>0035 | 6.25E-0<br>6   | 命中 | -1.165<br>531748 | 0.00024<br>2099 |    |
| CITED2       | CITED2-<br>11   | MPASVAHVPAAMLPPNVIDTDFIDEEVMSLVIE<br>MGLDR IKELPELWLGQNEFDFTDFV   | 48 | 6.16497<br>7472 | 6.48E-0<br>6   | 命中 | -0.949<br>095547 | 0.00056<br>3063 |    |
| CITED2       | CITED2-<br>12*  | AMLPPNV IDTDFIDEEVMSLV IEMGLDR IKELP<br>ELWLGQNEFDFTDFVCKQPPSRVSC | 49 | 6.14757<br>9295 | 7.93E-0<br>6   | 命中 | -0.900<br>478069 | 0.00244<br>0708 |    |
| CSRN1        | CSRN1-<br>28    | DNIEAPHFPLPGLSPPDASSCFLESLMGFSEPA<br>AEALDPFIDSQFEDTVPASLMPEVPV   | 50 | 0.50286<br>9407 | 0.46018<br>996 |    | 0.6955<br>35008  | 0.01577<br>9736 | 命中 |
| DDIT3        | DDIT3-1         | MAAESLPFSFGTLSSWELEAWYEDLQEVLSSEN                                 | 51 | 2.37268         | 0.00646        | 命中 | 1.5017           | 6.04E-0         | 命中 |

[0235]

|         |                 |                                                                   |    |                 |                |    |                 |                 |    |
|---------|-----------------|-------------------------------------------------------------------|----|-----------------|----------------|----|-----------------|-----------------|----|
|         |                 | GGTYVSPPGNEEESKIFTTLDPASL                                         |    | 4704            | 538            |    | 78634           | 6               |    |
| DDIT3   | DDIT3-2         | WYEDLQEVLSSENGGTYVSPPGNEEESKIFTT<br>LDPASLAWLTEEEPEAEVTSQS        | 52 | 1.13825<br>3747 | 0.11502<br>125 |    | 1.2799<br>79309 | 0.00033<br>0392 | 命中 |
| ELF4    | ELF4-1          | MAITLQPSDLIFEFASNGMDDDIHQLEDPSVPPA<br>VIVEQVYPDLLHLVSGLELDDVHNG   | 53 | 0.66255<br>5474 | 0.25787<br>399 |    | 0.7287<br>08968 | 0.01552<br>4531 | 命中 |
| ELF4    | ELF4-2          | DDIHQLEDPSVFPVAVIEQVYPDLLHLVSGLEL<br>DDVHNGIITDGTLCMTQDQILEGSFL   | 54 | 1.46675<br>7956 | 0.02072<br>139 | 命中 | 1.3636<br>97693 | 6.52E-0<br>5    | 命中 |
| FAM22F  | FAM22F-<br>19   | PRPQRPAAETNAHLPPRPQRPAAETKVPEEIPPEV<br>VQEVVDIMEELLGSHPGDTGEPEGQR | 55 | 0.99983<br>1925 | 0.17807<br>384 |    | 0.9883<br>53867 | 0.01577<br>9736 | 命中 |
| FAM22F  | FAM22F-<br>20   | PAETKVPEEIPPEVVQEVVDIMEELLGSHPGDTG<br>EPGQREKQKVEQPQEEEDGITSDPGL  | 56 | 3.21433<br>5785 | 0.00094<br>772 | 命中 | 2.1010<br>12353 | 6.76E-0<br>7    | 命中 |
| FAM22F  | FAM22F-<br>22   | EKGKVEQPQEEEDGITSDPGLLSYIDKLCSQEDFV<br>TKVEAVIHPRFLEELLSPDPQMDFLA | 57 | 4.76013<br>3648 | 8.76E-0<br>6   | 命中 | 2.3248<br>79314 | 1.46E-0<br>5    | 命中 |
| FAM22F  | FAM22F-<br>23   | LSYIDKLCSQEDFVTKVEAVIHPRFLEELLSPDP<br>QMDFLALSQELEQEGLTLAQLVEKR   | 58 | 5.82223<br>0419 | 6.25E-0<br>6   | 命中 | 0.2736<br>85646 | 0.46620<br>8757 |    |
| FAM90A1 | FAM90A1<br>-19  | RQPPHSRPLCLPTAQACTMSHHPAASHDGAQPLRV<br>LFRRLNGRWSSSLLTAPSFHSPEKP  | 59 | 2.05146<br>2115 | 0.03110<br>723 | 命中 | 1.5567<br>96738 | 7.06E-0<br>5    | 命中 |
| FAM90A1 | FAM90A1<br>-20* | HPAASHDGAQPLRVLFRRLNGRWSSSLLTAPSF<br>HSPEKPGAFLAQSPHVSEKSEGPCVR   | 60 | 3.01655<br>1155 | 0.00281<br>867 | 命中 | 1.6161<br>64006 | 0.00029<br>23   | 命中 |
| H0XA2   | H0XA2-4         | PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPE<br>YPMWKEKKAACKTALLPAAAAATAA   | 61 | 1.66442<br>7181 | 0.23442<br>147 |    | 1.5948<br>16728 | 0.02180<br>5285 | 命中 |
| H0XA2   | H0XA2-5         | RGSPVPAGALQPPEYPMWKEKKAACKTALLPAAA<br>AAATAAATGPACLSHESLEIADGSG   | 62 | 1.68406<br>4934 | 0.24906<br>108 |    | 2.3114<br>8338  | 0.00038<br>4794 | 命中 |

[0236]

|       |         |                                                                    |    |                  |                 |    |                  |                  |    |
|-------|---------|--------------------------------------------------------------------|----|------------------|-----------------|----|------------------|------------------|----|
| HOXA2 | HOXA2-6 | KKAAKKTALLPAAAAATAAATGPACLSHKESLE<br>IADGSGGSRRLRTAYTNTQLLELEK     | 63 | 1. 95100<br>1869 | 0. 24906<br>108 |    | 1. 7377<br>95426 | 0. 01577<br>9736 | 命中 |
| HSF1  | HSF1-20 | KNELSDHLDAMDSNLDNLQTMLSHGFSVDTSAL<br>LDFSPSVTVPDMSLPDLDSLASIQ      | 64 | 0. 62335<br>7487 | 0. 30974<br>031 |    | 0. 7077<br>35867 | 0. 02614<br>1316 | 命中 |
| HSF1  | HSF1-21 | MLSSHGFSVDTSALLDLFSPSVTVPDMSLPDLDS<br>SLASIQELLSPQEP RPPEAENSSPD   | 65 | 0. 43198<br>2684 | 0. 45359<br>875 |    | 0. 8991<br>90622 | 0. 00249<br>6242 | 命中 |
| HSF1  | HSF1-24 | SGKQLVHYTAQPLFLDPGSDVTGSNDLPVLFEL<br>GEGSYFSEG DGFAEDPTISLLTGSEP   | 66 | 0. 67628<br>0443 | 0. 54433<br>041 |    | 0. 8092<br>18277 | 0. 00814<br>9744 | 命中 |
| HSF1  | HSF1-25 | AQPLFLDPGSDVTGSNDLPVLFELGEGSYFSEG<br>DGFAEDPTISLLTGSEPPKAKDPTVS    | 67 | 0. 77254<br>8781 | 0. 22014<br>105 |    | 1. 1525<br>82348 | 0. 00029<br>23   | 命中 |
| JAZF1 | JAZF1-1 | MTGIAAASFFSNTCRGCGGLHFTLADLIEHIE<br>DNHIDTDP RVLEKQELQPTYVALSY     | 68 | 0. 28475<br>8647 | 0. 67306<br>89  |    | 0. 6638<br>77674 | 0. 03199<br>1955 | 命中 |
| KLF15 | KLF15-6 | GPVAGPWRRAAAPVKGHFCLPEFPLGDPDDVP<br>RPFQPTLEEIEEFLEENMEPGVKEVP     | 69 | 2. 46655<br>7755 | 0. 00312<br>371 | 命中 | 1. 2018<br>5423  | 0. 00020<br>3701 | 命中 |
| KLF6  | KLF6-1  | MDVLPMC SIFQELQIVHETGYFSALPSLEETWQQ<br>TCLELERYLQSEPCVVSASEIKFDSQ  | 70 | 5. 45310<br>7649 | 6. 25E-0<br>6   | 命中 | 1. 8677<br>53689 | 6. 47E-0<br>6    | 命中 |
| KLF7  | KLF7-1* | MDVLASYSIFQELQLVHDTGYFSALPSLEETWQQ<br>TCLELERYLQTEPRRISETFGEDLDC   | 72 | 5. 99766<br>1437 | 6. 25E-0<br>6   | 命中 | 1. 2896<br>03101 | 0. 00220<br>6825 | 命中 |
| KLF7  | KLF7-2  | YFSALPSLEETWQQTCLELERYLQTEPRRISETF<br>GEDLDCFLHASPPPCIEESFRRLDPL   | 73 | 1. 43095<br>7654 | 0. 17621<br>033 |    | 1. 1651<br>59905 | 0. 00066<br>6098 | 命中 |
| LIN9  | LIN9-25 | SRLTAILLQIKCLAEGLNSFEFKSLTDSLNDI<br>KSTIDASNISCFQNNVEIHVAHIQSG     | 74 | 2. 09396<br>4257 | 0. 03526<br>736 | 命中 | 1. 0799<br>6726  | 0. 07536<br>2785 |    |
| MYCBP | MYCBP-1 | MAHYKAADSKREQFRRYLEKSGVLDTLTKVLVAL<br>YEEPEKPN S ALDFLKHHLGAATPENP | 75 | 1. 01658<br>798  | 0. 11503<br>602 |    | 0. 7299<br>80644 | 0. 00711<br>2247 | 命中 |

[0237]

|         |               |                                                                   |    |                 |                |    |                  |                 |    |
|---------|---------------|-------------------------------------------------------------------|----|-----------------|----------------|----|------------------|-----------------|----|
| MYCL1   | MYCL1-1       | MDYDSYQHYFYDYDCGEDFYRSTAPSEDIWKKFE<br>LVPSPTSPPWGLPGAGDPAGIGP     | 76 | 1.13677<br>0705 | 0.05636<br>799 |    | 0.7967<br>82271  | 0.01156<br>406  | 命中 |
| MYOD1   | MYOD1-2       | CSFATDDFYDDPCFDSPLRFFEDLDPRLMHVG<br>ALLKPEESHFPAAVHPAGAREDEH      | 77 | 0.70253<br>9998 | 0.30142<br>347 |    | 0.8486<br>02054  | 0.00249<br>6242 | 命中 |
| NEUROD1 | NEUROD1<br>-6 | LRRMKANARERNMHGNAALDNLKVVPCVSKT<br>QKLSKIETLRKAKNYIWLSEILRSG      | 78 | 0.67079<br>4426 | 0.29541<br>863 |    | 0.9533<br>38619  | 0.00135<br>4227 | 命中 |
| NEUROG3 | NEUROG3<br>-5 | RRSRKKANDRERNMHNLSALDALRGVLPTEP<br>DDAKLTKIETLRFHNYIWTQTTLR       | 79 | 0.06038<br>7059 | 0.91793<br>719 |    | 0.6713<br>31782  | 0.03839<br>6908 | 命中 |
| NPAS4   | NPAS4-3<br>1  | DVPLVPEGLLTPEASPVKQSFHSEKEQNEIDR<br>LIQQISQLAQGMDRPFSAEAGTGGL     | 80 | 1.22008<br>4575 | 0.14208<br>522 |    | 0.8563<br>3903   | 0.03471<br>3147 | 命中 |
| NPAS4   | NPAS4-3<br>4  | PLGGLEPLDSNLSGAGPPVLSLDLKPWKQCEL<br>DFLADPDNMFLEETPVEDIFMDLSTP    | 81 | 0.94597<br>6876 | 0.14289<br>484 |    | 0.9186<br>42828  | 0.00319<br>0486 | 命中 |
| NPAS4   | NPAS4-3<br>6  | DNMFLEETPVEDIFMDLSTDPSEEWGSGDPEAE<br>GPGGAPSPCNNLSPEDHSFLEDLATY   | 82 | 0.67098<br>6209 | 0.25560<br>82  |    | 1.2093<br>79094  | 6.61E-0<br>5    | 命中 |
| NPAS4   | NPAS4-3<br>7  | DPSEEWGSGDPEAEGPGGAPSPCNNLSPEDHSFL<br>EDLATYETAFETGVSAPFYDGTDEL   | 83 | 1.64517<br>4283 | 0.01967<br>705 | 命中 | 0.7018<br>78979  | 0.05583<br>9582 |    |
| NPAS4   | NPAS4-3<br>9  | CNNLSPEDHSFLEDLATYETAFETGVSAPFYDGF<br>TDELHQLQSQVQDSFIHEDGSGGEPTF | 84 | 0.79804<br>9889 | 0.38225<br>13  |    | 0.7683<br>90862  | 0.01927<br>7289 | 命中 |
| RBPJ    | RBPJ-13       | GMALPRLIRKVDKQTALLDADDPVSQLHKCAFY<br>LKDTERMYLCLSQRILIQFQATPCPK   | 85 | 1.43103<br>5271 | 0.02573<br>555 | 命中 | -0.031<br>481589 | 0.97421<br>3707 |    |
| RELB    | RELB-2        | RRVARPPAAPELGALCSPDLSSLAVSRSTDEL<br>EIIIDEYIKENGFGLDGQPGGEGLP     | 86 | 2.14654<br>0297 | 0.00281<br>867 | 命中 | 1.7108<br>67835  | 6.99E-0<br>6    | 命中 |
| RELB    | RELB-3        | SSLSLAVSRSTDELEIIDEYIKENGFGLDGQPG<br>PGEGLPRLVSRGAASLSTVTLGPVAP   | 87 | 1.76839<br>2783 | 0.01126<br>942 | 命中 | 1.8029<br>01174  | 6.76E-0<br>7    | 命中 |
| SERTAD1 | SERTAD1       | IDTSMYDNELWAPASEGLKPGPEDGPGKEEAPEL                                | 88 | 0.23494         | 0.84161        |    | 0.8385           | 0.03298         | 命中 |

[0238]

|                |                |                                                                   |    |                 |                |    |                  |                 |    |
|----------------|----------------|-------------------------------------------------------------------|----|-----------------|----------------|----|------------------|-----------------|----|
|                | -10            | DEAELDYLMDVLVGTAQALRPPGGR                                         |    | 6196            | 029            |    | 07882            | 8887            |    |
| SERTAD2        | SERTAD2-11     | SEAGTQKLDGPQESRADDSKLDLPGNFEITTS<br>TGFLTDLTLDLILFADIDTSMYDFDP    | 89 | 2.48048<br>9855 | 0.00665<br>441 | 命中 | 1.8217<br>66531  | 4.96E-0<br>6    | 命中 |
| SPDYE4         | SPDYE4-3*      | DPSPPQSLGLKRSKSEWSESEEELELELERA<br>PEPEDTWVETLCGLKMKLKRKRASS      | 90 | 4.82049<br>5878 | 2.26E-0<br>5   | 命中 | 2.4093<br>6057   | 1.30E-0<br>7    | 命中 |
| SPDYE4         | SPDYE4-4       | SEEELEEELELERAPEPTWVETLCGLKMKLK<br>RKRASSVLPEHHEAFNRLGDPVVQK      | 91 | 4.31606<br>0433 | 0.00020<br>711 | 命中 | 1.9142<br>91641  | 6.04E-0<br>6    | 命中 |
| SS18L2         | SS18L2-2       | QETIQRLLLEENDQLIRCIVEYQNKGRGNECVQVQ<br>HVLHRNLIYLATIAADAPTSTSKAME | 92 | 1.01153<br>592  | 0.09853<br>126 |    | 1.1661<br>48268  | 0.00038<br>6437 | 命中 |
| TP53-AD<br>1-2 | TP53-AD<br>1-2 | PQSDPSVEPPLSQETFSDLWKLLPENNVLSPPLPS<br>QAMDDLMLSPDDIEQWFTEDPGPDEA | 93 | 1.10856<br>2999 | 0.28306<br>469 |    | 0.9855<br>78432  | 0.01758<br>1223 | 命中 |
| VP64           | VP64           | DALDDFDLMDLGSADLDDFDLMDLGSADLDDFDL<br>DMLGSADLDDFDLMDLGSVEREG     | 94 | 5.47585<br>036  | 6.48E-0<br>6   | 命中 | -2.374<br>266586 | 1.76E-0<br>7    |    |
| YAF2           | YAF2-6         | EKKDKVEKEKSEKETTSKKNSHKTRPRLKNVDR<br>SSAQHLEVTVGDLTVIITDFKEKTKS   | 95 | 4.31599<br>9866 | 0.00046<br>506 | 命中 | 1.7714<br>87421  | 0.00061<br>2967 | 命中 |
| YAF2           | YAF2-7*        | SHKTRPRLKNVDRSSAQHLEVTVGDLTVIITDF<br>KEKTKSPPASSAASADQHSQSGSSD    | 96 | 4.98686<br>9359 | 1.91E-0<br>5   | 命中 | 1.8283<br>75134  | 7.30E-0<br>5    | 命中 |
| ZXDC           | ZXDC-12*       | SPAEQHGAQDTELSAGTGNFYLESGGSARTDYRA<br>IQLAKEKKQRGAGSNAGASQSTQRKI  | 97 | 3.96574<br>8255 | 0.01967<br>705 | 命中 | 2.4960<br>61078  | 8.08E-0<br>8    | 命中 |
| ZXDC           | ZXDC-13        | YLESGGARTDYRAIQLAKEKKQRGAGSNAGASQ<br>STQRKIKEGKMSPPHFHASQNSWLCG   | 98 | 2.37728<br>1872 | 0.00307<br>646 | 命中 | 1.7342<br>31119  | 8.57E-0<br>6    | 命中 |
| ZXDC           | ZXDC-16        | SLVVPSSGGRPGPAPAGVQCGAQGVQVLVQDDP<br>SGEGVLPSARGPATFLPFLTVDLVPVY  | 99 | 0.47849<br>6057 | 0.41728<br>761 |    | 0.7292<br>56459  | 0.01792<br>5843 | 命中 |

[0239] 表2:表1中通过TADseq鉴定为TAD的片段子集。

[0240]

| 基因      | 片段        | 序列                                                               | SEQ ID NO: |
|---------|-----------|------------------------------------------------------------------|------------|
| ATMIN   | ATMIN-24  | NPGPDTQLPSGPAQNPGIDFDIEEFFSASNIQTQTEES<br>ELSTMTTEPVLDSLDIETQTDF | 32         |
| ATMIN   | ATMIN-32  | ETQTMSSGFETLGSFFTSNETQTAMDDFLLADLAWNT<br>MESQFSSVETQTSAPHTVSNF   | 33         |
| ATOH1   | ATOH1-1   | MSRLLHAEWEAEVKELGDHHRQPQPHHLPQPPPPQPP<br>ATLQAREHPVYPPELSLLDSTD  | 34         |
| ATXN7L3 | ATXN7L3-1 | MKMEEMSLSGLDNSKLEAIAQEIYADLVEDSCLGFCFE<br>VHRAVKCGYFFLDDTDPDSMKD | 35         |
| ATXN7L3 | ATXN7L3-2 | QEIYADLVEDSCLGFCFEVHRAVKCGYFFLDDTDPDSM<br>KDFEIVDQPGLDIFGQVFNQWK | 36         |
| ATXN7L3 | ATXN7L3-4 | FEIVDQPGLDIFGQVFNQWKSKECVCPNCSRSIAASRF<br>APHLEKCLGMGRNSSRIANRRI | 37         |
| BRD8    | BRD8-23   | KEECFRSGVAEAPVGSKAPSIDGKEELDLAEKMDIAVS<br>YTGEELDFETVGDIIIIIEDKV | 38         |
| BRD8    | BRD8-24   | IDGKEELDLAEKMDIAVSYTGEELDFETVGDIIAIIED<br>KVDDHPEVLDVAAVEAALSFCF | 39         |
| BRD8    | BRD8-25   | GEELDFETVGDIIAIIEDKVDDHPEVLDVAAVEAALS<br>CEENDDPQSLPGPWEHPIQQR   | 40         |
| BTBD18  | BTBD18-24 | IDCREPYAFDTALLEQPCEAEEYRITSAAATSELEEIL<br>DFMLCGSDIEPPIGSLESPGAE | 41         |
| BTBD18  | BTBD18-25 | EEYRITSAAATSELEEILDFMLCGSDIEPPIGSLESPG<br>AEGCRTPTYHLTETGKNWIEGE | 42         |
| CSRNP1  | CSRNP1-28 | DNIEAPHFPLPGLSPPGDASSCFLESLMGFSEPAEAL<br>DPFIDSQFEDTVPASLMEVPV   | 50         |
| HOXA2   | HOXA2-4   | PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPM<br>KEKKAACKTALLPAAAAAATAA  | 61         |
| HOXA2   | HOXA2-5   | RGSPVPAGALQPPEYPMKEKKAACKTALLPAAAAAAT<br>AAATGPACLSHKESLEIADGSG  | 62         |
| HOXA2   | HOXA2-6   | KKAACKTALLPAAAAAATAAATGPACLSHKESLEIADG<br>SGGGSRRRLRTAYTNTQLELEK | 63         |
| JAZF1   | JAZF1-1   | MTGIAAASFFSNTCRFGGCGLHFPTLADLIEHIEDNHI<br>DTPRVLEKQELQQPTYVALSY  | 68         |
| LIN9    | LIN9-25   | SRLTAILLQIKCLAEGGDLNSFEFKSLTDSLNDIKSTI<br>DASNISCFQNNVEIHVAHIQSG | 74         |
| MYCBP   | MYCBP-1   | MAHYKAADSKREQFRRYLEKSGVLDLTKVLVALYEPP<br>EKPNSALDFLKHHLGAATPENP  | 75         |
| MYCL1   | MYCL1-1   | MDYDSYQHYFYDYDCGEDFYRSTAPSEDIWKKFELVPS<br>PPTSPWGLPGGAGDPAPGIGP  | 76         |
| NEUROD1 | NEUROD1-6 | LRRMKANARERNRMHGLNAAALDNLRKVVPCYSKTQKLS                          | 78         |

[0241]

|         |           |                                                                  |    |
|---------|-----------|------------------------------------------------------------------|----|
|         |           | KIETLRLAKNYIWALSEILRSG                                           |    |
| NEUROG3 | NEUROG3-5 | RRSRKKANDRERNRMHNLNSALDALRGVLPPTFPDDAK<br>LTKIETLRFHNYIWALTQTLR  | 79 |
| RBPJ    | RBPJ-13   | GMALPRLIIRKVDKQTALLDADDPVSQLHKCAFYLKDT<br>ERMYLCLSQERIIQFQATPCPK | 85 |

[0242] 表3:TAD和其功能变体和活性片段的示例性序列。

[0243]

| TAD                                        | SEQ ID NO: | 序列                                                                  |
|--------------------------------------------|------------|---------------------------------------------------------------------|
| YAF2 短                                     | 100        | SHKKTRPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKTKSPPASS<br>A                  |
| C3orf62 短                                  | 101        | MDKSPQKATDADPGSLKQAFDDHNIVETVLDLEEDYNVMTSFKYQ<br>IE                 |
| “miniSPDYE4”<br>或“S(mini)”                 | 102        | SDESEEEEELELERAPEPEDTWVETL                                          |
| “miniCITED1”<br>或“C(mini)”                 | 103        | DSDPVDEEVLMSLVVELGLDRANEL                                           |
| HSF1 短                                     | 104        | GFSVDTSALLDLFSP                                                     |
| “p65 短(N末端)”;<br>“P(mini N末端)”;<br>或“P_Nt” | 105        | PTQAGEGTLSEALLQLQFDDDLGALLGNSTDPAVFTDLASVDNS<br>EFQQ                |
| “p65 短(C末端)”;<br>“P(mini C末端)”;<br>或“P_Ct” | 106        | APGLPNGLLSGDEDFSSIADMDFSALL                                         |
| ZXDC 短                                     | 107        | AGTGNFYLES GGSA RTDYRAIQLAKEKKQRGAGSNAGASQSTQRK<br>IKEGKM           |
| RELB 短                                     | 108        | SSL SLAVSRSTDELEI IDEYIKENG FGLDGGQPGPGGLP                          |
| FAM22F 短                                   | 109        | LSYIDKLCSQEDFVTKVEAVIHPRFLEELLSPDPQMDFLA                            |
| KLF6-TAD                                   | 71         | DVLPMCSIFQELQIVHETGYFSALPSLEEWQQTCELELERYLQSE<br>PCYVSASEIKFDSQ     |
| “C3orf62-TAD”<br>; 或 “miniC3orf62”         | 110        | QDSPFKEEAWALLMDKSPQKATDADPGSLKQAFDDHNIVETVLDL<br>EEDYNVMTSFKYQIEGSS |
| SOX7 C末端                                   | 111        | LGQLSPPPEHPGFDALDQLSQVELLGMDRNEFDQYLNTPGHPDS<br>ATGAMALSGHVPVSQV    |

[0244]

|                                               |     |                                                                                                                                                                                                                                                                                      |
|-----------------------------------------------|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| FOX1<br>(697-762)                             | 112 | DVPKPGSPEPQVSGLAANRSLTEGLVLDTMNDSLSKILLDISFPG<br>LDEDPLGPDNINWSQFIPELQ                                                                                                                                                                                                               |
| ATF6                                          | 113 | HRLDEDWDSALFAELGYFTDTDELQLEAANETYENNFDNLDLDFLD<br>LVPWESDIWDINNQIGSS                                                                                                                                                                                                                 |
| “<br>SERTAD2-NAHL<br>”                        | 89  | SEAGTQKLDGPQESRADD SKLMDSLPGNFEITTTSTGFLDTLTD<br>ILFADIDTSMYDFDP                                                                                                                                                                                                                     |
| “DDIT3-1” ;<br>或 “miniDDIT3<br>”              | 51  | MAAESLPFSFGTLSSWELEAWYEDLQEVLSSENGGTYVSPPGNE<br>EESKIFTTLDPASL                                                                                                                                                                                                                       |
| ELF4                                          | 54  | DDIHQLEDPSVFPVIVEQVPYDLLHLYSGLELDDVHNGIITDG<br>TLCMTQDQILEGSFL                                                                                                                                                                                                                       |
| C11orf74                                      | 114 | SAHMSGLEIMDEDQLIKDVLDFLNCHQTYDEEFLNTFTHLSQD<br>LLLLPGEVEQDVST                                                                                                                                                                                                                        |
| “miniATMIN<br>”                               | 32  | NPGPDTQLPSGPAQNP GIDFDIEFFSASNIQTQTEESELSTMTT<br>EPVLESLDIETQTDF                                                                                                                                                                                                                     |
| “miniZDC”                                     | 115 | YLESGSARTDYRAIQLAKEKKQRGAGSNAGASQSTQRKI                                                                                                                                                                                                                                              |
| “SPDYE4” ;<br>或 “S”                           | 90  | DPSPQPQSLGLKRKSEWSDESEEEEELELERAPEPEDTWVET<br>LCGLKMKLKRKRASS                                                                                                                                                                                                                        |
| “CITED1” 或<br>“C”                             | 47  | ESLSPSAGAQSPA I IDSDPVDEEVMSLVVELGLDRANELPELWL<br>GQNEFDFTADFPSSC                                                                                                                                                                                                                    |
| “RELA (p65)<br>” ; “RELA”<br>或 “p65”          | 116 | QYLPDTPDRHRIEKKRRTYETFKSIMKKSPFSGPTDPRPPRR I<br>AVPSRSSASVPKPAPQYPFTSSLSTINYDEFPTMVFPSPGQISQA<br>SALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAV<br>APPAPKPTQAGEGTLSEALLQLQFDDDLGALLGNSTDPVFTDL<br>ASVDNSEFQQLNQIPVAPHTTEPMLMEYPEA ITRLVTGAQRPP<br>DPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALL |
| “p65-AD” ;<br>“P-短” ; 或<br>“P”                | 117 | PTQAGEGTLSEALLQLQFDDDLGALLGNSTDPVFTDLASVDNS<br>EFQQLNQIPVAPHTTEPMLMEYPEA ITRLVTGAQRPPDPAPAP<br>LGAPGLPNGLLSGDEDFSSIADMDFSALL                                                                                                                                                         |
| “HSF1-TAD”<br>; “H-TAD” ;<br>“HSF1” 或 “H<br>” | 118 | GFSVDTSALLDLFSPSVTVPMSLPDLSSLASIQELLSPQEPPR<br>PPEAENSSPDGKQLVHYTAQPLFLDPGSVDTGSDNLPVLFELG<br>EGSYFSEGDFGAEPTISLLTGSEPPKAKDPTVS                                                                                                                                                      |
| “mini-HSF1<br>(401-420)” 或<br>“H(mini)”       | 119 | MLSSHGFSVDTSALLDLFSP                                                                                                                                                                                                                                                                 |
| CBX2 (CBX-C)                                  | 136 | SASPPSTGQNPSVSVQTSQDWKPTRSLIEHVFVTDVTANLITVT<br>KESPTSVGFFNLRHY                                                                                                                                                                                                                      |
| CBX6 (CBX-C)                                  | 138 | GASSEPEAGDWRPEMSPCSNVVVDVTSNLLTVTIKEFCNPEDFE<br>KVAAGVAGAAGGGGSGASK                                                                                                                                                                                                                  |
| CBX8 (CBX-C)                                  | 139 | RPSLIARIPVARILGDPEEESWSPSLTNLEKVVVDVTSNFLT<br>VTKESNTDQGGFFKEKR                                                                                                                                                                                                                      |
| RYBP-C                                        | 140 | IQSANATTKTSETNHTSRPRLKNVDRSTAQQLA VTVGNVTVIITD                                                                                                                                                                                                                                       |

[0245]

|                |     |                                                                                          |
|----------------|-----|------------------------------------------------------------------------------------------|
| (YAF2_RYBP)    |     | FKEKTRSSSTSSSTVTSSAGSEQQNQS                                                              |
| DDIT3_MT       | 156 | SFGTLSSWELEAWYEDLQEVLSSENGGTYVSPPGNEEEESKIFT<br>TLDPASLAWLT                              |
| SPDYE4_MT      | 157 | ERAPEPEDTWVETLCGLKMKLKRKRASS                                                             |
| ZXDC_MT        | 158 | TGNFYLES GG SARTDYRAIQLAKEKKQRGAGSNAGASQSTQRKI                                           |
| HSF1_MT        | 159 | GFSVDTSALLDLFSPSVTVPDMSLPDL DSSLASI QELLSPQEPPR                                          |
| SERTAD2_MT     | 160 | LMDSLPGNF EIT TSTGFLTDLTLDDILFADIDTSMYDFDP                                               |
| ZNF473         | 161 | MAEEFVTLKDVGMDFTLGDWEQLGLEQGDTFWDTALDNCQDLFLL<br>DPPRPNLTSHPDGSEDL EPLAGGSPEATS          |
| CITED1_MT      | 162 | AIIDSDPVDEEVLMSLVVELGLDRANELPELWLGQNEFDFTADF                                             |
| ZNF496         | 163 | MQQVTTLQLPPSRVSPFKDMILCFSEEDWSLLDPAQTGFYGEFII<br>GEDYGVSMPPNDLAAQPDLSQGEENEPRVPELQDLQKGE |
| CITED2_MT      | 164 | VIDTDFIDEEVLMSLVIEMGLDRIKELPELWLGQNEFDFMTDFV                                             |
| mini-p65_MT    | 165 | PTQAGEGTLSEALLQLQFDD EDLGALLGNSTDP AVFTDLASVDNS<br>EFQQLLNQG                             |
| C3orf62-2      | 166 | ATDADPGSLKQAFDDHNIVETVLDLEEDYNVMT                                                        |
| C3orf62-3      | 167 | DDHNIVETVLDLEEDYNVMT                                                                     |
| miniC3orf62-1  | 168 | DADPGSLKQAFDDHNIVETVLDLEED                                                               |
| miniC3orf62-2  | 169 | DSPFKEEAWALLMDKSPQKATDADPGSL                                                             |
| miniDDIT3_1    | 170 | DLQEVLSSENGGTYVSPPGNEEEESKIFTTLD                                                         |
| miniDDIT3_2    | 171 | LSSWELEAWYEDLQEVLSSE                                                                     |
| miniC3orf62-MT | 172 | ATDADPGSLKQAFDDHNIVETVLDLEEDYNVMTSFKYQIEGSS                                              |
| miniSPDYE4_1   | 185 | ERAPEPEDTWVETL                                                                           |

[0246] 表4:表1中通过TADseq鉴定为TAD的片段子集。

[0247]

| 基因      | 片段        | 序列                                                                  | SEQ ID NO: |
|---------|-----------|---------------------------------------------------------------------|------------|
| ATMIN   | ATMIN-24  | NPGPDTQLPSGPAQNPGIDFDIEEFFSASNIQTQTEESE<br>LSTMTTEPVLES LD IETQ TDF | 32         |
| ATMIN   | ATMIN-32  | ETQTMSSGFETLGSLFFTSNETQTAMDDFLADLAWNTM<br>ESQFSSVETQ TSAEPHTVSNF    | 33         |
| ATOH1   | ATOH1-1   | MSRLLHAEWEAEVKELGDHHRQPQPHHLPQPPPPQPPA<br>TLQAREHPVYPPELSLLDSTD     | 34         |
| ATXN7L3 | ATXN7L3-1 | MKMEEMSLSGLDNSKLEAIAQE IYADLVEDSCLGFCFEV<br>HRAVKCGYFFLDDTD PDSMKD  | 35         |
| ATXN7L3 | ATXN7L3-2 | QE IYADLVEDSCLGFCFEVHRAVKCGYFFLDDTD PDSMK<br>DFEIVDQPGLDIFGQVFNQWK  | 36         |
| ATXN7L3 | ATXN7L3-4 | FEIVDQPGLDIFGQVFNQWKSKECVCPCNCSRSIAASRFA<br>PHLEKCLGMGRNSSRIANRRI   | 37         |

[0248]

|              |                |                                                                    |    |
|--------------|----------------|--------------------------------------------------------------------|----|
| BRD8         | BRD8-23        | KEECFRSGVAEAPVGSKAPSIDGKEELDLAEKMDIAVS<br>YTGEELDFETVGDIIIIIEDKV   | 38 |
| BRD8         | BRD8-24        | IDGKEELDLAEKMDIAVSYTGEELDFETVGDIIAIIEDK<br>VDDHPEVLDVAAVEAALSFC    | 39 |
| BRD8         | BRD8-25        | GEELDFETVGDIIAIIEDKVDDHPEVLDVAAVEAALSFC<br>EENDDPQSLPGPWEHPIQQR    | 40 |
| BTBD18       | BTBD18-<br>24  | IDCREPYAFDTALLEQPCEAEEYRITSAAATSELEEILD<br>FMLCGSDIEPPIGSLESPGAE   | 41 |
| BTBD18       | BTBD18-<br>25  | EEYRITSAAATSELEEILDFMLCGSDIEPPIGSLESPGA<br>EGCRTPTYHLTETGKNWIEGE   | 42 |
| C11orf<br>74 | C11orf7<br>4-1 | MSAHMSGLEIMDEDQLIKDVLDFLNCHEQTYDEEFLNT<br>FTHLSQDLLLLPGEVEQDVST    | 43 |
| C3orf6<br>2  | C3orf62<br>-11 | NHMCQHCQDSPFKEEAWALLMDKSPQKATDADPGSLKQA<br>FDDHNIVETVLDLEEDYNVMT   | 44 |
| C3orf6<br>2  | C3orf62<br>-12 | QDSPFKEEAWALLMDKSPQKATDADPGSLKQAFDDHNIV<br>ETVLDLEEDYNVMTSFKYQIE   | 45 |
| FAM22F       | FAM22F-<br>19  | PRPQRPAETNAHLPPRPQRPAETKVPEEIPPEVVQEYV<br>DIMEELLGSHPGDTGEPEGQR    | 55 |
| FAM22F       | FAM22F-<br>20  | PAETKVPEEIPPEVVQEYVDIMEELLGSHPGDTGEPEGQ<br>REKGVQEQEEDGITSDPGL     | 56 |
| FAM22F       | FAM22F-<br>22  | EKGKVEQPQEEEDGITSDPGLLSYIDKLCSEQEDFVTKVEA<br>VIHPRFLEELLSPDPQMDFLA | 57 |
| FAM22F       | FAM22F-<br>23  | LSYIDKLCSEQEDFVTKVEAVIHPRFLEELLSPDPQMDFL<br>ALSQEELEEEGLTLAQLVEKR  | 58 |
| FAM90A<br>1  | FAM90A1<br>-19 | RQPPHSRCLPTAQAQCTMSHHPAASHDGAQPLRVLFRRLL<br>ENGRWSSSLLTAPSFHSPEKP  | 59 |
| FAM90A<br>1  | FAM90A1<br>-20 | HPAASHDGAQPLRVLFRRLENGRWSSSLLTAPSFHSPEK<br>PGAFLAQSPHVSEKSEGPCVR   | 60 |
| HOXA2        | HOXA2-4        | PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPMWK<br>EKKAAKKTALLPAAAAAATAA   | 61 |
| HOXA2        | HOXA2-5        | RGSPVPAGALQPPEYPMWKEKKAAKKTALLPAAAAAATA<br>AATGPACLSHKESLEIADGSG   | 62 |
| HOXA2        | HOXA2-6        | KKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGS<br>GGGSRRLRTAYTNTQLLELEK   | 63 |
| JAZF1        | JAZF1-1        | MTGIAAASFFSNTCRFGGCGLHFPTLADLIEHIEDNHID<br>TDPRVLEKQELQQPTYVALSY   | 68 |
| LIN9         | LIN9-25        | SRLTALLQIKCLAEAGDLNSFEFKSLTDSLNDIKSTID<br>ASNISCFQNNVEIHVAHIQSG    | 74 |
| MYCBP        | MYCBP-1        | MAHYKAADSKREQFRRYLEKSGVLDLTKVLVALYEEPE<br>KPNSALDFLKHHLGAATPENP    | 75 |
| MYCL1        | MYCL1-1        | MDYDSYQHYFYDYDCGEDFYRSTAPSEDIWKKFELVPSP<br>PTSPPWGLGPGAGDPAPGIGP   | 76 |
| NEUROD<br>1  | NEUROD1<br>-6  | LRRMKANARERNRMHGLNAALDNLRKVPCYSKTQKLSK<br>IETRLAKNYIWALSEILRSG     | 78 |

[0249]

|         |           |                                                                |    |
|---------|-----------|----------------------------------------------------------------|----|
| NEUROG3 | NEUROG3-5 | RRSRRKKANDRERNRMHNLNSALDALRGVLPFPDDAKLTKIETLRFHNYIWALTQTLR     | 79 |
| RBPJ    | RBPJ-13   | GMALPRLIIRKVDKQTALLDADDPVSQLHKCAFYLKDTERMYLCLSQERIIQFQATPCPK   | 85 |
| SPDYE4  | SPDYE4-3  | DPSPQPQSLGLKRRKSEWSDESEEELEEELELERAPEPEDTWVVETLCGLKMKLKRKRASS  | 90 |
| SPDYE4  | SPDYE4-4  | SEEELEEELELERAPEPEDTWVVETLCGLKMKLKRKRASSVLPEHHEAFNRLLGDPVVQK   | 91 |
| SS18L2  | SS18L2-2  | QETIQRLLLEENDQLIRCIVEYQNKGRGNECVQYQHVLRNLIYLATIADASPTSTSKAME   | 92 |
| YAF2    | YAF2-6    | EKKDKVEKEKSEKETTSKKNSHKKTRPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKTKS   | 95 |
| YAF2    | YAF2-7    | SHKKTRPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKTKSPPASSAASADQHSQSGSSSD   | 96 |
| ZXDC    | ZXDC-12   | SPAEQHGAQDTELSAGTGNFYLESGGSAARTDYRAIQLAKEKKQRGAGSNAGASQSTQRK I | 97 |
| ZXDC    | ZXDC-13   | YLESGGSAARTDYRAIQLAKEKKQRGAGSNAGASQSTQRK IKEGKMSPPHFHASQNSWLCG | 98 |
| ZXDC    | ZXDC-16   | SLVVPSSGGRPGPAPAAGVQCGAQGVQVQLVQDDPSGEGVLPSARGPATFLPFLTVDLVPVY | 99 |

[0250] 表5:通过TADseq鉴定为TAD的活性片段子集。

[0251]

| 基因     | 片段        | 序列                                                                                          | SEQ ID NO: | 先前鉴定的TAD 位置 |
|--------|-----------|---------------------------------------------------------------------------------------------|------------|-------------|
| ATF6   | ATF6-1    | MGEPAGVAGTMESPFSPGLF <i>HR</i> LDEDWDSA<br>LFAELGYFTD <i>DELQ</i> LEAANETYENNFDML           | 29         | aa 1-150    |
| ATF6   | ATF6-2    | <i>HR</i> LDEDWDSALFAELGYFTD <i>DELQ</i> LEAAN<br>ETYENNFDNLDFDLVPWESDIWDINNQI              | 30         | aa 1-150    |
| ATF6B  | ATF6B-1   | MAELMLLSEIADPTRFFTDNLLSPEDWGLQ<br>NSTLYSGLDEVAEEQTQLFRCPEQDVPFDG                            | 31         | aa 2-86     |
| CSRNP1 | CSRNP1-28 | DNIEAPHFPLPGLSPPGDASSCFLESLMGF<br>SEPAAEALDPFIDSQFEDTVPASLMPEVPV                            | 50         | aa 493-583  |
| KLF6   | KLF6-1    | MDVLP MCSIFQELQIVHETGYFSALPSLEE<br>YWQQT <i>CLELERYLQ</i> SEPCYVSASEIKFDSQ                  | 70         | aa 1-200    |
| KLF7   | KLF7-1    | MDVLASY SIFQELQLVHDTG <i>YF</i> SALPSLEE<br><i>TW</i> QQT <i>CLELERYLQ</i> TEPRRISETFGEDLDC | 72         | aa 1-75     |
| KLF7   | KLF7-2    | <i>YF</i> SALPSLEET <i>TW</i> QQT <i>CLELERYLQ</i> TEPRRI<br>SETFGEDLDCFLHASPPPCIEESFRRLDPL | 73         | aa 1-75     |

[0252] 表6:鉴定为TAD的片段子集。

[0253]

| 基因 | 片段 | 序列 | SEQ ID |
|----|----|----|--------|
|----|----|----|--------|

[0254]

|         |           |                                                                   | NO: |
|---------|-----------|-------------------------------------------------------------------|-----|
| ATMIN   | ATMIN-24  | NPGPDTQLPSGPAQNPGIDFDIEEFFSASNIQTQTEESE<br>LSTMTTEPVLESLEDIETQTD  | 32  |
| ATMIN   | ATMIN-32  | ETQTMSSGFETLGSFFTSNETQTAMDDFLADLAWNTM<br>ESQFSSVETQTSAPHTVSNF     | 33  |
| ATOHI   | ATOHI-1   | MSRLLHAEWEAEVKELGDHHRQPQPHLPQPPPPQPPA<br>TLQAREHPVYPPELSLLDSTD    | 34  |
| ATXN7L3 | ATXN7L3-1 | MKMEEMSLSGLDNSKLEAIAQEIIYADLVEDSCLGFCFEV<br>HRAVKCGYFFLDDTDPDSMKD | 35  |
| ATXN7L3 | ATXN7L3-2 | QEIIYADLVEDSCLGFCFEVHRAVKCGYFFLDDTDPDSMK<br>DFEIVDQPGLDIFGQVFNQWK | 36  |
| ATXN7L3 | ATXN7L3-4 | FEIVDQPGLDIFGQVFNQWKSKECVCPCNSRSIAASRFA<br>PHLEKCLGMGRNSSRIANRRI  | 37  |
| BRD8    | BRD8-23   | KEECFRSGVAEAPVSGKAPSIDGKEELDLAEKMDIAVSY<br>TGEELDFETVGDIIIIIEDKV  | 38  |
| BRD8    | BRD8-24   | IDGKEELDLAEKMDIAVSYTGEELDFETVGDIIAIIEDK<br>VDDHPEVLDVAAVEAALSFC   | 39  |
| BRD8    | BRD8-25   | GEELDFETVGDIIAIIEDKVDDHPEVLDVAAVEAALSFC<br>EENDDPQSLPGPWEHPIQER   | 40  |
| BTBD18  | BTBD18-24 | IDCREPYAFDTALLEQPCEABEYRITSAAATSELEEILD<br>FMLCGSDIEPPIGSLESPGAE  | 41  |
| BTBD18  | BTBD18-25 | EEYRITSAAATSELEEILDFMLCGSDIEPPIGSLESPGA<br>EGCRTPTYHLTETGKNWIEGE  | 42  |
| FAM22F  | FAM22F-19 | PRPQRPAETNAHLPPRPQRPAETKVPEEIPPEVVQEYV<br>DIMEELGSHPGDTGEPEGQR    | 55  |
| FAM22F  | FAM22F-20 | PAETKVPEEIPPEVVQEYVDIMEELGSHPGDTGEPEGQ<br>REKKGVEQPQEEDGITSDPGL   | 56  |
| HOXA2   | HOXA2-4   | PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPMK<br>EKKAACKTALLPAAAAAATAA   | 61  |
| HOXA2   | HOXA2-5   | RGSPVPAGALQPPEYPMKEKKAACKTALLPAAAAAATA<br>AATGPACLSHKESLEIADGSG   | 62  |
| HOXA2   | HOXA2-6   | KKAACKTALLPAAAAAATAAATGPACLSHKESLEIADGS<br>GGGSRRRLRTAYTNTQLELEK  | 63  |
| JAZF1   | JAZF1-1   | MTGIAAASFFSNTCRFGGCGLHFPTLADLIEHIEDNHID<br>TDPRVLEKQELQQPTYVALSY  | 68  |
| LIN9    | LIN9-25   | SRLTAILLQIKCLAEAGDLNSFEFKSLTDSLNDIKSTID<br>ASNISCFQNNVEIHVAHIQSG  | 74  |
| MYCBP   | MYCBP-1   | MAHYKAADSKREQFRRYLEKSGVLDLTKVLVALYEEPE<br>KPNSALDFLKHHLGAATPENP   | 75  |
| MYCL1   | MYCL1-1   | MDYDSYQHYFYDYDCGEDFYRSTAPSEDIWKKFELVPSP<br>PTSPPWGLGPGAGDPAPGIGP  | 76  |
| NEUROD1 | NEUROD1-6 | LRRMKANARERNRMHGLNAALDNLRKVVPYCYSKTQKLSK<br>IETLRLAKNYIWALSEILRSG | 78  |
| NEUROG3 | NEUROG3-  | RRSRRKKANDRERNRMHNLNSALDALRGVLPFTFPDDAKL                          | 79  |

[0255]

|        |          |                                                                    |    |
|--------|----------|--------------------------------------------------------------------|----|
|        | 5        | TKIETLRF AHNYIWALTQTLR                                             |    |
| RBPJ   | RBPJ-13  | GMALPRLIIRKV DKTALLDADDPVSQLHKCAFYLKDTE<br>RMYLCLSQERIIQFQATPCPK   | 85 |
| SS18L2 | SS18L2-2 | QETIQRLL EENDQLIRCI VEYQNKGRGNECVQYQHVLHR<br>NLIYLATIADASPTSTSKAME | 92 |
| YAF2   | YAF2-6   | EKKDKVEKEKSEKETTSKKNSHKKTRPRLKNVDRSSAQH<br>LEVTVGDLTVIITDFKEKTKS   | 95 |
| YAF2   | YAF2-7   | SHKKTRPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKTK<br>SPPASSAASADQHSQSGSSSD   | 96 |
| ZXDC   | ZXDC-12  | SPAEQHGAQDTELSAGTGNFYLES GGSARTDYRAIQ LAK<br>EKKQRGAGSNAGASQSTQRKI | 97 |
| ZXDC   | ZXDC-13  | YLES GGSARTDYRAIQ LAKEKKQRGAGSNAGASQSTQRK<br>IKEGKMSPPHFHASQNSWLCG | 98 |
| ZXDC   | ZXDC-16  | SLVVP SGGRPGPAPAAGVQCGAQGVQVQLVQDDPSGEGV<br>LPSARGPATFLPFLTVDLPVY  | 99 |

[0256] SPDYE4-CITED1-RELA-HSF1核苷酸序列 (SEQ ID NO:120)

[0257] SPDYE4-CITED1-RELA-HSF1蛋白质序列 (SEQ ID NO:121)

[0258] MDPSPQPQSLGLKRKSEWSDESEEELELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSGGSE  
SLSPSAGAAQSPAI IDSDPVDEEVLM SLVVELGLDRANELPELWL GQNEFDFTADFPSSCGGSINSRSSGSPKKKR  
KVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKS PFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPFTSSLSTIN  
YDEFPTMVFP SGQISQASALAPAPPQVLPQAPAPAPAMVSALAQA PAPVPVLAPGPPQAVAPPAPKPTQAGEGTL  
SEALLQLQFDD EDLGALLGNSTDP AVFTDLASVDNSEFQQLLNQ GIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDP  
APAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSGFSVDT SALLDLFSPSVTPDMSLPDL DSSLASIQELL  
SPQEPPRPPEAENSSPD SGKQLVHYTAQPLFLLDPGSVDTGSNDLPVLFELGEGSYFSEG DGFAEDPTISLLTGSEP  
PKAKDPTVS

[0259] 接头和NLS序列用下划线表示

[0260] SPDYE4-CITED1-RELA核苷酸序列 (SEQ ID NO:122)

[0261] SPDYE4-CITED1-RELA蛋白质序列 (SEQ ID NO:123)

[0262] MDPSPQPQSLGLKRKSEWSDESEEELELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSGGSE  
SLSPSAGAAQSPAI IDSDPVDEEVLM SLVVELGLDRANELPELWL GQNEFDFTADFPSSCGGSINSRSSGSPKKKR  
KVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKS PFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPFTSSLSTIN  
YDEFPTMVFP SGQISQASALAPAPPQVLPQAPAPAPAMVSALAQA PAPVPVLAPGPPQAVAPPAPKPTQAGEGTL  
SEALLQLQFDD EDLGALLGNSTDP AVFTDLASVDNSEFQQLLNQ GIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDP  
APAPLGAPGLPNGLLSGDEDFSSIADMDFSALL

[0263] 接头和NLS序列用下划线表示

[0264] HSF1-RELA-SPDYE4-CITED1核苷酸序列 (SEQ ID NO:124)

[0265] HSF1-RELA-SPDYE4-CITED1蛋白质序列 (SEQ ID NO:125)

[0266] MGFSVDT SALLDLFSPSVTPDMSLPDL DSSLASIQELLSPQEPPRPPEAENSSPD SGKQLVHYTAQPL  
FLLDPGSVDTGSNDLPVLFELGEGSYFSEG DGFAEDPTISLLTGSEP PKAKDPTVSINSRSSGSPKKKRKVGSQYLP  
DTDDRHRIEEKRKRTYETFKSIMKKS PFSGPTDPRPPPRRIAVPSRSSASVPKPAPQYPFTSSLSTIN YDEFPTMV  
FPSGQISQASALAPAPPQVLPQAPAPAPAMVSALAQA PAPVPVLAPGPPQAVAPPAPKPTQAGEGTL SEALLQLQ

FDDEDLGALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAP  
GLPNGLLSGDEDFSSIADMDFSALLGSGSGSDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETL  
CGLKMKLKRKRASSGSGSGSESLSPSAGAQSPAIIIDSDPVDEEVLM SLVVELGLDRANELPELWLQGNEFDFTADFP  
SSCGGS

[0267] 接头和NLS序列用下划线表示

[0268] SPDYE4-CITED1-p65-miniHSF1核苷酸序列 (SEQ ID NO:126)

[0269] SPDYE4-CITED1-p65-miniHSF1蛋白质序列 (SEQ ID NO:127)

[0270] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGSGSGSES  
LSPSAGAQSPAIIIDSDPVDEEVLM SLVVELGLDRANELPELWLQGNEFDFTADFPSSCGGSINSRSSGSPKKRKRKVG  
SPTQAGEGTLSEALLQLQFDDEDLGALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRL  
VTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSMLSSHGFSVDTSALLDLFSP

[0271] 接头和NLS序列用下划线表示

[0272] miniSPDYE4-CITED1-p65-HSF1核苷酸序列 (SEQ ID NO:128)

[0273] miniSPDYE4-CITED1-p65-HSF1蛋白质序列 (SEQ ID NO:129)

[0274] MSDESEEELEEELELERAPEPEDTWVETLSGSGSGSESLSPSAGAQSPAIIIDSDPVDEEVLM SLVVEL  
GLDRANELPELWLQGNEFDFTADFPSSCGGSINSRSSGSPKKRKRKVGSPTQAGEGTLSEALLQLQFDDEDLGALLGN  
STDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDF  
SSIADMDFSALLGSGSGSGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDGK  
QLVHYTAQPLFLDPGSVDTGSDNLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

[0275] 接头和NLS序列用下划线表示

[0276] SPDYE4-miniCITED1-p65-HSF1核苷酸序列 (SEQ ID NO:130)

[0277] SPDYE4-miniCITED1-p65-HSF1蛋白质序列 (SEQ ID NO:131)

[0278] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGSGSGGS

[0279] DSDPVDEEVLM SLVVELGLDRANELGGINSRSSGSPKKRKRKVGSPTQAGEGTLSEALLQLQFDDEDLG  
ALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLL  
SGDEDFSSIADMDFSALLGSGSGSGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSS  
PDGKQLVHYTAQPLFLDPGSVDTGSDNLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

[0280] 接头和NLS序列用下划线表示

[0281] SPDYE4-CITED1-minip65(C)-HSF1核苷酸序列 (SEQ ID NO:132)

[0282] SPDYE4-CITED1-minip65(C)-HSF1蛋白质序列 (SEQ ID NO:133)

[0283] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGSGSGGSE  
SLSPSAGAQSPAIIIDSDPVDEEVLM SLVVELGLDRANELPELWLQGNEFDFTADFPSSCGGSINSRSSGSPKKR  
KVGSAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQELLSP  
QEPPRPPEAENSSPDGKQLVHYTAQPLFLDPGSVDTGSDNLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPK  
AKDPTVS

[0284] 接头和NLS序列用下划线表示

[0285] SPDYE4-CITED1-minip65(N)-HSF1核苷酸序列 (SEQ ID NO:134)

[0286] SPDYE4-CITED1-minip65(N)-HSF1蛋白质序列 (SEQ ID NO:135)

[0287] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGSGSGGSE

SLSPSAGAQSPAIDSDPVDEEVLMSLVVELGLDRANELPELWLGQNEFDFTADFPSSCGGSINSRSSGSPKKRK  
KVGSPQTQAGEGTLSEALLQLQFDDDEDLGALLGNSTDPVFTDLASVDNSEFQQGSGSGSGFSVDT  
SALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDGKQLVHYTAQPLFLDPGSVDTGSNDLPVLFELGEGSYF  
SEGDGFAEDPTISLLTGSEPPKAKDPTVS

[0288] 接头和NLS序列用下划线表示

[0289] >S-C3orf62.2-P\_AD-H核苷酸序列 (SEQ ID NO:173)

[0290] >S-C3orf62.2-P\_AD-H蛋白质序列 (SEQ ID NO:174)

[0291] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSATDAD  
PGSLKQAFDDHNIVETVLDLEEDYNVMTGGSINSRSSGSPKKRKVGSPQTQAGEGTLSEALLQLQFDDDEDLGALLGN  
STDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDED  
FSSIADMDFSALLGSGSGSGFSVDT  
SALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDGKQLVHYTAQPLFLDPGSVDTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

[0292] 接头和NLS序列用下划线表示

[0293] >S-C3orf62.3-P\_AD-H核苷酸序列 (SEQ ID NO:175)

[0294] >S-C3orf62.3-P\_AD-H蛋白质序列 (SEQ ID NO:176)

[0295] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSDDHNI  
VETVLDLEEDYNVMTGGSINSRSSGSPKKRKVGSPQTQAGEGTLSEALLQLQFDDDEDLGALLGNSTDPVFTDLASV  
DNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALL  
GSGSGSGFSVDT  
SALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDGKQLVHYTAQPLFLDPGSVDTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

[0296] 接头和NLS序列用下划线表示

[0297] >S-C3orf62\_MT-P\_AD-H核苷酸序列 (SEQ ID NO:177)

[0298] >S-C3orf62\_MT-P\_AD-H蛋白质序列 (SEQ ID NO:178)

[0299] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSATDAD  
PGSLKQAFDDHNIVETVLDLEEDYNVMTSFYQIEGSSGGSINSRSSGSPKKRKVGSPQTQAGEGTLSEALLQLQFDD  
DEDLGALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGL  
PNGLLSGDEDFSSIADMDFSALLGSGSGSGFSVDT  
SALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDGKQLVHYTAQPLFLDPGSVDTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

[0300] 接头和NLS序列用下划线表示

[0301] >S-DDIT3\_MT-P\_AD-H核苷酸序列 (SEQ ID NO:179)

[0302] >S-DDIT3\_MT-P\_AD-H蛋白质序列 (SEQ ID NO:180)

[0303] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGSGSGSSFG  
TGLSSWELEAWYEDLQEVLSSENGGTYVSPPGNEEEESKIFTTLDPASLAWLTGGSINSRSSGSPKKRKVGSPQTQ  
AGEGTLSEALLQLQFDDDEDLGALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGA  
QRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSGFSVDT  
SALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDGKQLVHYTAQPLFLDPGSVDTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

[0304] 接头和NLS序列用下划线表示

[0305] >S-C-P\_AD-HSF1\_MT核苷酸序列 (SEQ ID NO:181)

[0306] >S-C-P\_AD-HSF1\_MT蛋白质序列 (SEQ ID NO:182)

[0307] MDPSPQPQSLGLKRKSEWSDESEEELEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSGGSE  
SLSPSAGAQSPAI IDSDPVDEEVLM SLVVELGLDRANELPELWLGQNEFDFTADFPSSCGGSINSRSSGSPKKKR  
KVGSPQTQAGEGTLSEALLQLQFDDDELGALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAI  
TRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSGFSVDTSALLDLFSPSVTPDMSLP  
DLDSSLASIQELLSPQEPPR

[0308] 接头和NLS序列用下划线表示

[0309] >3x ZNF473\_KRAB核苷酸序列 (SEQ ID NO:183)

[0310] >3x ZNF473\_KRAB蛋白质序列 (SEQ ID NO:184)

[0311] MAEEFVTLKDVGMDFTLGDWEQLGLEQGDTFWD TALDNCQDLFLLDPPRPNLTSHPDGSEDLEPLAGGS  
PEATSM AEEFVTLKDVGMDFTLGDWEQLGLEQGDTFWD TALDNCQDLFLLDPPRPNLTSHPDGSEDLEPLAGGSPEA  
TSMAEEFVTLKDVGMDFTLGDWEQLGLEQGDTFWD TALDNCQDLFLLDPPRPNLTSHPDGSEDLEPLAGGSPEATS

[0312] 参考文献:

[0313] Adusumilli, R., and Mallick, P. (2017). Data Conversion with ProteoWizard msConvert. *Methods Mol. Biol.* Clifton NJ 1550, 339-368.

[0314] Alekseyenko, A. A., Gorchakov, A. A., Kharchenko, P. V., and Kuroda, M. I. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc. Natl. Acad. Sci.* 111, 2488-2493.

[0315] Ali, R. H., and Rouzbahman, M. (2015). Endometrial stromal tumours revisited: an update based on the 2014 WHO classification. *J. Clin. Pathol.* 68, 325-332.

[0316] Alonso, C. N., Meyer, C., Gallego, M. S., Rossi, J. G., Mansini, A. P., Rubio, P. L., Medina, A., Marschalek, R., and Felice, M. S. (2010). BTBD18: A novel MLL partner gene in an infant with acute lymphoblastic leukemia and inv(11)(q13;q23). *Leuk. Res.* 34, e294-e296.

[0317] Antonescu, C., Sung, Y.-S., Zhang, L., Agaram, N., and Fletcher, C. (2017). Recurrent SRF-RELA Fusions Define a Novel Subset of Cellular Myofibroblast/Myopericytoma: A Potential Diagnostic Pitfall With Sarcomas With Myogenic Differentiation. *Am. J. Surg. Pathol.* 41, 677-684.

[0318] Arnold, C. D., Nemčko, F., Woodfin, A. R., Wienerroither, S., Vlasova, A., Schleiffer, A., Pagani, M., Rath, M., and Stark, A. (2018). A high-throughput method to identify trans-activation domains within transcription factor sequences. *EMBO J.* 37.

[0319] Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* 324, 1720-1723.

[0320] Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S. K., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F., et al. (2007). Gadd45a promotes epigenetic

gene activation by repair-mediated DNA demethylation. *Nature* 445,671-675.

[0321] Barrett, J., Birrer, M. J., Kato, G. J., Dosaka-Akita, H., and Dang, C. V. (1992). Activation domains of L-Myc and c-Myc determine their transforming potencies in rat embryo cells. *Mol. Cell. Biol.* 12, 3130-3137.

[0322] Basnet, H., Su, X. B., Tan, Y., Meisenhelder, J., Merkurjev, D., Ohgi, K. A., Hunter, T., Pillus, L., and Rosenfeld, M. G. (2014). Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation. *Nature* 516, 267-271.

[0323] Basu, S., Mackowiak, S. D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., Geertsema, H., Ali, S., Jerković, I., Ewers, H., et al. (2020). Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* 181, 1062-1079.e30.

[0324] Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 491-507.

[0325] Bousoik, E., and Montazeri Aliabadi, H. (2018). "Do We Know Jack" About JAK? A Closer Look at JAK/STAT Signaling Pathway. *Front. Oncol.* 8.

[0326] Breitskreutz, A., Choi, H., Sharom, J. R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitskreutz, B.-J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. *Sci. N.Y. NY* 328, 1043-1046.

[0327] Brien, G. L., Remillard, D., Shi, J., Hemming, M. L., Chabon, J., Wynne, K., Dillon, E. T., Cagney, G., Van Mierlo, G., Baltissen, M. P., et al. (2018). Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma. *ELife* 7, e41305.

[0328] Buchwalter, G., Gross, C., and Wasylyk, B. (2004). Ets ternary complex transcription factors. *Gene* 324, 1-14.

[0329] Cao, L., Yonis, A., Vaghela, M., Barriga, E. H., Chugh, P., Smith, M. B., Maufront, J., Lavoie, G., Méant, A., Ferber, E., et al. (2020). SPIN90 associates with mDia1 and the Arp2/3 complex to regulate cortical actin organization. *Nat. Cell Biol.* 22, 803-814.

[0330] Centore, R. C., Sandoval, G. J., Soares, L. M. M., Kadoch, C., and Chan, H. M. (2020). Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends Genet.* 36, 936-950.

[0331] Chauhan, S., Zheng, X., Tan, Y. Y., Tay, B.-H., Lim, S., Venkatesh, B., and Kaldis, P. (2012). Evolution of the Cdk-activator Speedy/RINGO in vertebrates. *Cell. Mol. Life Sci.* 69, 3835-3850.

[0332] Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., Iyer, E. P. R., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326-328.

[0333] Chavez, A., Tuttle, M., Pruitt, B. W., Ewen-Campen, B., Chari, R., Ter-

- Ovanesyan,D.,Haque,S.J.,Cecchi,R.J.,Kowal,E.J.K.,Buchthal,J.,et al. (2016) .Comparison of Cas9 activators in multiple species.Nat.Methods 13,563-567.
- [0334] Chen,F.X.,Smith,E.R.,and Shilatifard,A. (2018) .Born to run:control of transcription elongation by RNA polymerase II.Nat.Rev.Mol.Cell Biol.19,464-478.
- [0335] Conaway,R.C.,and Conaway,J.W. (2011) .Function and regulation of the Mediator complex.Curr.Opin.Genet.Dev.21,225-230.
- [0336] Core,L.,and Adelman,K. (2019) .Promoter-proximal pausing of RNA polymerase II:a nexus of gene regulation.Genes Dev.33,960-982.
- [0337] Cramer,P. (2019) .Organization and regulation of gene transcription.Nature 573,45-54.
- [0338] Davis,R.L.,Weintraub,H.,and Lassar,A.B. (1987) .Expression of a single transfected cDNA converts fibroblasts to myoblasts.Cell 51,987-1000.
- [0339] Di Rocco,G.,Mavilio,F.,and Zappavigna,V. (1997) .Functional dissection of a transcriptionally active,target-specific Hox-Pbx complex.EMBO J.16,3644-3654.
- [0340] Dyson,H.J.,and Wright,P.E. (2005) .Intrinsically unstructured proteins and their functions.Nat.Rev.Mol.Cell Biol.6,197-208.
- [0341] Eng,J.K.,Jahan,T.A.,and Hoopmann,M.R. (2013) .Comet:an open-source MS/MS sequence database search tool.PROTEOMICS 13,22-24.
- [0342] Erijman,A.,Kozlowski,L.,Sohrabi-Jahromi,S.,Fishburn,J.,Warfield,L.,Schreiber,J.,Noble,W.S.,Söding,J.,and Hahn,S. (2020) .A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning.Mol.Cell 78,890-902.e6.
- [0343] Esnault,C.,Stewart,A.,Gualdrini,F.,East,P.,Horswell,S.,Matthews,N.,and Treisman,R. (2014) .Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts.Genes Dev.28,943-958.
- [0344] Fujisawa,T.,and Filippakopoulos,P. (2017) .Functions of bromodomain-containing proteins and their roles in homeostasis and cancer.Nat.Rev.Mol.Cell Biol.18,246-262.
- [0345] Gahan,J.M.,Rentzsch,F.,and Schnitzler,C.E. (2020) .The genetic basis for PRC1 complex diversity emerged early in animal evolution.Proc.Natl.Acad.Sci.117,22880-22889.
- [0346] Gao,Y.,Xiong,X.,Wong,S.,Charles,E.J.,Lim,W.A.,and Qi,L.S. (2016) .Complex transcriptional modulation with orthogonal and inducible dCas9 regulators.Nat.Methods 13,1043-1049.
- [0347] Gao,Z.,Zhang,J.,Bonasio,R.,Strino,F.,Sawai,A.,Parisi,F.,Kluger,Y.,and Reinberg,D. (2012) .PCGF Homologs,CBX Proteins,and RYBP Define Functionally

Distinct PRC1 Family Complexes. *Mol. Cell* 45,344-356.

[0348] Gao, Z., Lee, P., Stafford, J.M., von Schimmelmann, M., Schaefer, A., and Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature* 516,349-354.

[0349] Gastwirt, R.F., McAndrew, C.W., and Donoghue, D.J. (2007). Speedy/RINGO Regulation of CDKs in Cell Cycle, Checkpoint Activation and Apoptosis. *Cell Cycle* 6,1188-1193.

[0350] Gerritsen, M.E., Williams, A.J., Neish, A.S., Moore, S., Shi, Y., and Collins, T. (1997). CREB-binding protein/p300 are transcriptional coactivators of p65. *Proc. Natl. Acad. Sci. U.S.A.* 94,2927-2932.

[0351] Gonzalez, L., and Nebreda, A.R. (2020). RINGO/Speedy proteins, a family of non-canonical activators of CDK1 and CDK2. *Semin. Cell Dev. Biol.* 107,21-27.

[0352] Goparaju, S.K., Kohda, K., Ibata, K., Soma, A., Nakatake, Y., Akiyama, T., Wakabayashi, S., Matsushita, M., Sakota, M., Kimura, H., et al. (2017). Rapid differentiation of human pluripotent stem cells into functional neurons by mRNAs encoding transcription factors. *Sci. Rep.* 7,42367.

[0353] Gualdrini, F., Esnault, C., Horswell, S., Stewart, A., Matthews, N., and Treisman, R. (2016). SRF Co-factors Control the Balance between Cell Proliferation and Contractility. *Mol. Cell* 64,1048-1061.

[0354] Guo, Z., Zhang, L., Wu, Z., Chen, Y., Wang, F., and Chen, G. (2014). In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. *Cell Stem Cell* 14,188-202.

[0355] Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., and Stark, A. (2019). Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* 570,122-126.

[0356] Hirai, H., Tani, T., Katoku-Kikyo, N., Kellner, S., Karian, P., Firpo, M., and Kikyo, N. (2011). Radical acceleration of nuclear reprogramming by chromatin remodeling with the transactivation domain of MyoD. *Stem Cells Dayt. Ohio* 29, 1349-1361.

[0357] Horb, M.E., Shen, C.-N., Tosh, D., and Slack, J.M.W. (2003). Experimental Conversion of Liver to Pancreas. *Curr. Biol.* 13,105-115.

[0358] Hrzenjak, A. (2016). JAZF1/SUZ12 gene fusion in endometrial stromal sarcomas. *Orphanet J. Rare Dis.* 11,15.

[0359] Hsu, S.I., Yang, C.M., Sim, K.G., Hentschel, D.M., O'Leary, E., and Bonventre, J.V. (2001). TRIP-Br: a novel family of PHD zinc finger- and bromodomain-interacting proteins that regulate the transcriptional activity of E2F-1/DP-1. *EMBO J.* 20,2273-2285.

[0360] Huttlin, E.L., Bruckner, R.J., Navarrete-Perea, J., Cannon, J.R., Baltier, K., Gebreab, F., Gygi, M.P., Thornock, A., Zarraga, G., Tam, S., et al. (2020). Dual

Proteome-scale Networks Reveal Cell-specific Remodeling of the Human Interactome (Systems Biology).

[0361] Israni, D.V., Li, H.-S., Gagnon, K.A., Sander, J.D., Roybal, K.T., Joung, J.K., Wong, W.W., and Khalil, A.S. (2021). Clinically-driven design of synthetic gene regulatory programs in human cells (Synthetic Biology).

[0362] Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327-339.

[0363] Karanian, M., Pissaloux, D., Gomez-Brouchet, A., Chevenet, C., Le Loarer, F., Fernandez, C., Minard, V., Corradini, N., Castex, M.-P., Duc-Gallet, A., et al. (2020). SRF-FOXO1 and SRF-NCOA1 Fusion Genes Delineate a Distinctive Subset of Well-differentiated Rhabdomyosarcoma. *Am. J. Surg. Pathol.* 44, 607-616.

[0364] Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J., and Khalil, A.S. (2014). Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* 158, 110-120.

[0365] Kim, D.I., Jensen, S.C., Noble, K.A., Kc, B., Roux, K.H., Motamedchaboki, K., and Roux, K.J. (2016). An improved smaller biotin ligase for Biold proximity labeling. *Mol. Biol. Cell* 27, 1188-1196.

[0366] Knight, J.D.R., Choi, H., Gupta, G.D., Pelletier, L., Raught, B., Nesvizhskii, A.I., and Gingras, A.-C. (2017). ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nat. Methods* 14, 645-646.

[0367] Kundu, T.K., Palhan, V.B., Wang, Z., An, W., Cole, P.A., and Roeder, R.G. (2000). Activator-dependent transcription from chromatin in vitro involving targeted histone acetylation by p300. *Mol. Cell* 6, 551-561.

[0368] Lambert, J.-P., Tucholska, M., Go, C., Knight, J.D.R., and Gingras, A.-C. (2015). Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J. Proteomics* 118, 81-94.

[0369] Lambert, J.-P., Picaud, S., Fujisawa, T., Hou, H., Savitsky, P., Uusküla-Reimand, L., Gupta, G.D., Abdouni, H., Lin, Z.-Y., Tucholska, M., et al. (2019). Interactome Rewiring Following Pharmacological Targeting of BET Bromodomains. *Mol. Cell* 73, 621-638.e17.

[0370] Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. *Cell* 172, 650-665.

[0371] Lecoq, L., Raiola, L., Chabot, P.R., Cyr, N., Arseneault, G., Legault, P., and Omichinski, J.G. (2017). Structural characterization of interactions between transactivation domain 1 of the p65 subunit of NF- $\kappa$ B and transcription regulatory factors. *Nucleic Acids Res.* 45, 5564-5576.

- [0372] Li, X., Wang, W., Wang, J., Malovannaya, A., Xi, Y., Li, W., Guerra, R., Hawke, D.H., Qin, J., and Chen, J. (2015). Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol. Syst. Biol.* 11.
- [0373] Liang, F.-S., Ho, W.Q., and Crabtree, G.R. (2011). Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci. Signal.* 4, rs2-rs2.
- [0374] Liu, G., Knight, J.D.R., Zhang, J.P., Tsou, C.-C., Wang, J., Lambert, J.-P., Larsen, B., Tyers, M., Raught, B., Bandeira, N., et al. (2016). Data Independent Acquisition analysis in ProHits 4.0. *J. Proteomics* 149, 64-68.
- [0375] Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers *Cell* 153, 320-334.
- [0376] Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charlotiaux, B., et al. (2020). A reference map of the human binary protein interactome. *Nature* 580, 402-408.
- [0377] Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichtinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Müller-Lerch F, Fu F, Pearlberg J, Göbel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, lafrate AJ, Dobbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell.* 2008 Jul 25; 31(2):294-301.
- [0378] Marcon, E., Ni, Z., Pu, S., Turinsky, A.L., Trimble, S.S., Olsen, J.B., Silverman-Gavrila, R., Silverman-Gavrila, L., Phanse, S., Guo, H., et al. (2014). Human-Chromatin-Related Protein Interactions Identify a Demethylase Complex Required for Chromosome Segregation. *Cell Rep.* 8, 297-310.
- [0379] Mashtalir, N., D'Avino, A.R., Michel, B.C., Luo, J., Pan, J., Otto, J.E., Zullo, H.J., McKenzie, Z.M., Kubiak, R.L., St Pierre, R., et al. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* 175, 1272-1288.e20.
- [0380] McGrath, D.A., Fifield, B.-A., Marceau, A.H., Tripathi, S., Porter, L.A., and Rubin, S.M. (2017). Structural basis of divergent cyclin-dependent kinase activation by Spyl/RINGO proteins. *EMBO J.* 36, 2251-2262.
- [0381] Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.-P., St-Denis, N.A., Li, T., Miteva, Y.V., Hauri, S., Sardi, M.E., Low, T.Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* 10, 730-736.
- [0382] Miralles, F., Posern, G., Zaromytidou, A.-I., and Treisman, R. (2003). Actin Dynamics Control SRF Activity by Regulation of Its Coactivator MAL. *Cell* 113, 329-342.
- [0383] Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di

Croce,L.(2012).Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells.Cell Stem Cell 10,47-62.

[0384] Morita,K.,Celso,C.L.,Spencer-Dene,B.,Zouboulis,C.C.,and Watt,F.M.(2006).HAN11 binds mDial and controls GLI1 transcriptional activity.J.Dermatol.Sci.44,11-20.

[0385] Muhar,M.,Ebert,A.,Neumann,T.,Umkehrer,C.,Jude,J.,Wieshofer,C.,Rescheneder,P.,Lipp,J.J.,Herzog,V.A.,Reichholf,B.,et al.(2018).SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis.Science 360, 800-805.

[0386] Najafabadi,H.S.,Mnaimneh,S.,Schmitges,F.W.,Garton,M.,Lam,K.N.,Yang,A.,Albu,M.,Weirauch,M.T.,Radovani,E.,Kim,P.M.,et al.(2015).C2H2 zinc finger proteins greatly expand the human regulatory lexicon.Nat.Biotechnol.33,555-562.

[0387] Narayan,S.,Bryant,G.,Shah,S.,Berrozpe,G.,and Ptashne,M.(2017).OCT4 and SOX2 Work as Transcriptional Activators in Reprogramming Human Fibroblasts.Cell Rep.20,1585-1596.

[0388] Nasrin,N.,Ogg,S.,Cahill,C.M.,Biggs,W.,Nui,S.,Dore,J.,Calvo,D.,Shi,Y.,Ruvkun,G.,and Alexander-Bridges,M.C.(2000).DAF-16 recruits the CREB-binding protein coactivator complex to the insulin-like growth factor binding protein 1 promoter in HepG2 cells.Proc.Natl.Acad.Sci.97,10412-10417.

[0389] Ng,A.H.M.,Khoshakhlagh,P.,Rojo Arias,J.E.,Pasquini,G.,Wang,K.,Swiersy,A.,Shipman,S.L.,Appleton,E.,Kiaee,K.,Kohman,R.E.,et al.(2021).A comprehensive library of human transcription factors for cell fate engineering.Nat.Biotechnol.39,510-519.

[0390] Olson,E.N.,and Nordheim,A.(2010).Linking actin dynamics and gene transcription to drive cellular motile functions.Nat.Rev.Mol.Cell Biol.11, 353-365.

[0391] ORFeome Collaboration(2016).The ORFeome Collaboration:a genome-scale human ORF-clone resource.Nat.Methods 13,191-192.

[0392] Pellizzoni,L.,Charroux,B.,Rappsilber,J.,Mann,M.,and Dreyfuss,G.(2001).A Functional Interaction between the Survival Motor Neuron Complex and RNA Polymerase II.J.Cell Biol.152,75-86.

[0393] Piette,B.L.,Alerasool,N.,Lin,Z.-Y.,Lacoste,J.,Lam,M.H.Y.,Qian,W.W.,Tran,S.,Larsen,B.,Campos,E.,Peng,J.,et al.(2021).Comprehensive interactome profiling of the human Hsp70 network highlights functional differentiation of J domains.Mol.Cell 0.

[0394] Piper,D.E.,Batchelor,A.H.,Chang,C.-P.,Cleary,M.L.,and Wolberger,C.(1999).Structure of a HoxB1-Pbx1 Heterodimer Bound to DNA:Role of the Hexapeptide and a Fourth Homeodomain Helix in Complex Formation.Cell 96,587-

597.

[0395] Piskacek, S., Gregor, M., Nemethova, M., Grabner, M., Kovarik, P., and Piskacek, M. (2007). Nine-amino-acid transactivation domain: Establishment and prediction utilities. *Genomics* 89, 756-768.

[0396] Piunti, A., Smith, E.R., Morgan, M.A.J., Ugarenko, M., Khaltyan, N., Helmin, K.A., Ryan, C.A., Murray, D.C., Rickels, R.A., Yilmaz, B.D., et al. (2019). CATACOMB: An endogenous inducible gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. *Sci. Adv.* 5, eaax2887.

[0397] Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. *Nature* 386, 569-577.

[0398] Ravarani, C.N., Erkina, T.Y., De Baets, G., Dudman, D.C., Erkin, A.M., and Babu, M.M. (2018). High-throughput discovery of functional disordered regions: investigation of transactivation domains. *Mol. Syst. Biol.* 14, e8190.

[0399] Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma Oxf. Engl.* 26, 139-140.

[0400] Ryseck, R.P., Bull, P., Takamiya, M., Bours, V., Siebenlist, U., Dobrzanski, P., and Bravo, R. (1992). RelB, a new Rel family transcription activator that can interact with p50-NF-kappa B. *Mol. Cell. Biol.* 12, 674-684.

[0401] Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335, 563-564.

[0402] Sanborn, A.L., Yeh, B.T., Feigerle, J.T., Hao, C.V., Townshend, R.J.L., Aiden, E.L., Dror, R.O., and Kornberg, R.D. (2020). Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. *BioRxiv* 2020.12.18.423551.

[0403] Sanjana, N., Cong, L., Zhou, Y., et al. A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 7, 171-192 (2012).

[0404] Sano, K., Hayakawa, A., Piao, J.-H., Kosaka, Y., and Nakamura, H. (2000). Novel SH3 protein encoded by the AF3p21 gene is fused to the mixed lineage leukemia protein in a therapy-related leukemia with t(3;11) (p21;q23). *Blood* 95, 1066-1068.

[0405] Schratt, G., Philippar, U., Berger, J., Schwarz, H., Heidenreich, O., and Nordheim, A. (2002). Serum response factor is crucial for actin cytoskeletal organization and focal adhesion assembly in embryonic stem cells. *J. Cell Biol.* 156, 737-750.

[0406] Sdelci, S., Lardeau, C.-H., Tallant, C., Klepsch, F., Klaiber, B., Bennett, J., Rathert, P., Schuster, M., Penz, T., Fedorov, O., et al. (2016). Mapping the chemical chromatin reactivation landscape identifies BRD4-TAF1 cross-

talk.Nat.Chem.Biol.12,504-510.

[0407] Shteynberg,D.,Deutsch,E.W.,Lam,H.,Eng,J.K.,Sun,Z.,Tasman,N.,Mendoza,L.,Moritz,R.L.,Aebersold,R.,and Nesvizhskii,A.I. (2011) .iProphet:multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates.Mol.Span Classnocasespan Cell.Proteomics MCP 10,M111.007690.

[0408] Sigler,P.B. (1988) .Transcriptional activation.Acid blobs and negative noodles.Nature 333,210-212.

[0409] Singh,R.N.,Howell,M.D.,Ottesen,E.W.,and Singh,N.N. (2017) .Diverse role of survival motor neuron protein.Biochim.Biophys.Acta BBA-Gene Regul.Mech.1860,299-315.

[0410] Skapek,S.X.,Ferrari,A.,Gupta,A.A.,Lupo,P.J.,Butler,E.,Shipley,J.,Barr,F.G.,and Hawkins,D.S. (2019) .Rhabdomyosarcoma.Nat.Rev.Dis.Primer 5,1-19.

[0411] Staller,M.V.,Ramirez,E.,Holehouse,A.S.,Pappu,R.V.,and Cohen,B.A. (2021) .Design principles of acidic transcriptional activation domains.BioRxiv 2020.10.28.359026.

[0412] Stampfel,G.,Kazmar,T.,Frank,O.,Wienerroither,S.,Reiter,F.,and Stark,A. (2015) .Transcriptional regulators form diverse groups with context-dependent regulatory functions.Nature 528,147-151.

[0413] Strasswimmer,J.,Lorson,C.L.,Breiding,D.E.,Chen,J.J.,Le,T.,Burghes,A.H.M.,and Androphy,E.J. (1999) .Identification of Survival Motor Neuron as a Transcriptional Activator-Binding Protein.Hum.Mol.Genet.8,1219-1226.

[0414] Sudarshan,D.,Avvakumov,N.,Lalonde,M.-E.,Alerasool,N.,Jacquet,K.,Mameri,A.,Rousseau,J.,Lambert,J.-P.,Paquet,E.,Setty,S.T.,et al. (2021) .Recurrent chromosomal translocations in sarcomas create a mega-complex that mislocalizes NuA4/TIP60 to Polycomb target loci.BioRxiv 2021.03.26.436670.

[0415] Tague,E.P.,Dotson,H.L.,Tunney,S.N.et al.Chemogenetic control of gene expression and cell signaling with antiviral drugs.Nat Methods 15,519-522 (2018) .

[0416] Taipale,M.,Tucker,G.,Peng,J.,Krykbaeva,I.,Lin,Z.-Y.,Larsen,B.,Choi,H.,Berger,B.,Gingras,A.-C.,and Lindquist,S. (2014) .A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways.Cell 158,434-448.

[0417] Tanenbaum ME,Gilbert LA,Qi LS,Weissman JS,Vale RD.A protein-tagging system for signal amplification in gene expression and fluorescence imaging.Cell.2014;159(3) :635-646.doi:10.1016/j.cell.2014.09.039

[0418] Teo,G.,Liu,G.,Zhang,J.,Nesvizhskii,A.I.,Gingras,A.-C.,and Choi,H. (2014) .SAINTexpress:Improvements and additional features in Significance Analysis of INteractome software.J.Proteomics 100,37-43.

- [0419] The ENCODE Project Consortium, Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shores, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583, 699-710.
- [0420] Theodorou, E., Dalembert, G., Heffelfinger, C., White, E., Weissman, S., Corcoran, L., and Snyder, M. (2009). A high throughput embryonic stem cell screen identifies Oct-2 as a bifunctional regulator of neuronal differentiation. *Genes Dev.* 23, 575-588.
- [0421] Tycko, J., DelRosso, N., Hess, G.T., Aradhana, Banerjee, A., Mukund, A., Van, M.V., Ego, B.K., Yao, D., Spees, K., et al. (2020). High-Throughput Discovery and Characterization of Human Transcriptional Effectors. *Cell* 0.
- [0422] Vannam, R., Sayilgan, J., Ojeda, S., Karakyriakou, B., Hu, E., Kreuzer, J., Morris, R., Herrera Lopez, X.l., Rai, S., Haas, W., et al. (2021). Targeted degradation of the enhancer lysine acetyltransferases CBP and p300. *Cell Chem. Biol.* 28, 503-514.e12.
- [0423] Vartiainen, M.K., Guettler, S., Larijani, B., and Treisman, R. (2007). Nuclear Actin Regulates Dynamic Subcellular Localization and Activity of the SRF Cofactor MAL. *Science* 316, 1749-1752.
- [0424] Vihervaara, A., Duarte, F.M., and Lis, J.T. (2018). Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.* 19, 385-397.
- [0425] Vincenz, C., and Kerppola, T.K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc. Natl. Acad. Sci.* 105, 16572-16577.
- [0426] Wang, R., Ilango, U., Robinson, A.K., Schirf, V., Schwarz, P.M., Lafer, E.M., Demeler, B., Hinck, A.P., and Kim, C.A. (2008). Structural transitions of the RING1B C-terminal region upon binding the polycomb cbox domain. *Biochemistry* 47, 8007-8015.
- [0427] Wang, R., Taylor, A.B., Leal, B.Z., Chadwell, L.V., Ilango, U., Robinson, A.K., Schirf, V., Hart, P.J., Lafer, E.M., Demeler, B., et al. (2010). Polycomb Group Targeting through Different Binding Partners of RING1B C-Terminal Domain. *Structure* 18, 966-975.
- [0428] Wang, Y., Li, Y., Zeng, W., Zhu, C., Xiao, J., Yuan, W., Wang, Y., Cai, Z., Zhou, J., Liu, M., et al. (2004). IXL, a new subunit of the mammalian Mediator complex, functions as a transcriptional suppressor. *Biochem. Biophys. Res. Commun.* 325, 1330-1338.
- [0429] Wang, Y., Chen, J., Hu, J.-L., Wei, X.-X., Qin, D., Gao, J., Zhang, L., Jiang, J., Li, J.-S., Liu, J., et al. (2011). Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep.* 12, 373-378.
- [0430] Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe,

P.,Najafabadi,H.S.,Lambert,SA.,Mann,I.,Cook,K.,et al.(2014).Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity.Cell 158, 1431-1443.

[0431] Winters,A.C.,and Bernt,K.M.(2017).MLL-Rearranged Leukemias--An Update on Science and Clinical Approaches.Front.Pediatr.5.

[0432] Yahata,T.,Shao,W.,Endoh,H.,Hur,J.,Coser,K.R.,Sun,H.,Ueda,Y.,Kato,S.,Isselbacher,K.J.,Brown,M.,et al.(2001).Selective coactivation of estrogen-dependent transcription by CITED1 CBP/p300-binding protein.Genes Dev.15,2598-2612.

[0433] Yan,J.,Enge,M.,Whittington,T.,Dave,K.,Liu,J.,Sur,I.,Schmierer,B.,Jolma,A,Kivioja,T.,Taipale,M.,et al.(2013).Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites.Cell 154,801-813.

[0434] Yang,F.,DeBeaumont,R.,Zhou,S.,and Naar,A.M.(2004).The activator-recruited cofactor/Mediator coactivator subunit ARC92 is a functionally important target of the VP16 transcriptional activator.Proc.Natl.Acad.Sci.U.S.A.101,2339-2344.

[0435] Yang,X.,Boehm,J.S.,Yang,X.,Salehi-Ashtiani,K.,Hao,T.,Shen,Y.,Lubonja,R.,Thomas,S.R.,Alkan,O.,Bhimdi,T.,et al.(2011).A public genome-scale lentiviral expression library of human ORFs.Nat.Methods 8,659-661.

[0436] Yu,D.,Cattoglio,C.,Xue,Y.,and Zhou,Q.(2019).A complex between DYRK1A and DCAF7 phosphorylates the C-terminal domain of RNA polymerase II to promote myogenesis Nucleic Acids Res.47,4462-4475.

[0437] Zhou,L.,Canagarajah,B.,Zhao,Y.,Baibakov,B.,Tokuhiro,K.,Maric,D.,and Dean,J.(2017).BTBD18 Regulates a Subset of piRNA-Generating Loci through Transcription Elongation in Mice.Dev.Cell 40,453-466e5.

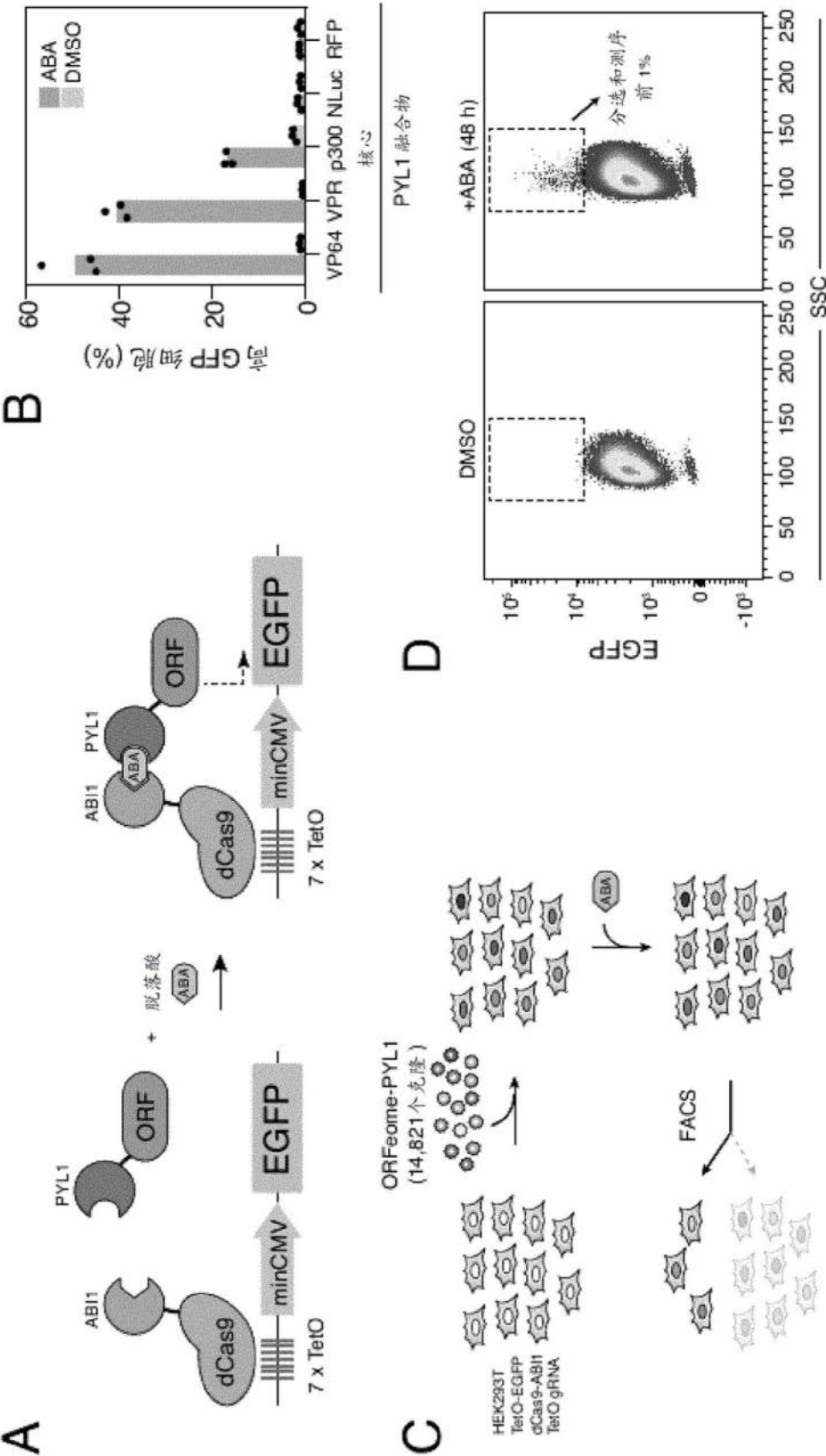
[0438] 24.Kiani S et al.Cas9 gRNA engineering for genome editing,activation and repression.Nature Methods 12,1051(2015).

[0439] 25.Beerli R.R.et al Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks PNAS 95,14628(1995).

[0440] 26.Cong,L.,Zhou,R.,Kuo,Yc.et al.Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains.Nat Commun 3,968(2012).

[0441] 27.Gilbert LA,Larson MH,Morsut L,Liu Z,Brar GA,Torres SE,Stern-Ginossar N,Brandman O,Whitehead EH,Doudna JA,Lim WA,Weissman JS,Qi LS Cell.2013 Jul 9.pii:S0092-8674(13)00826-X.

[0442] 28.Sanson,K.R.,Hanna,R.E.,Hegde,M.et al.Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities.Nat Commun 9,5416(2018).



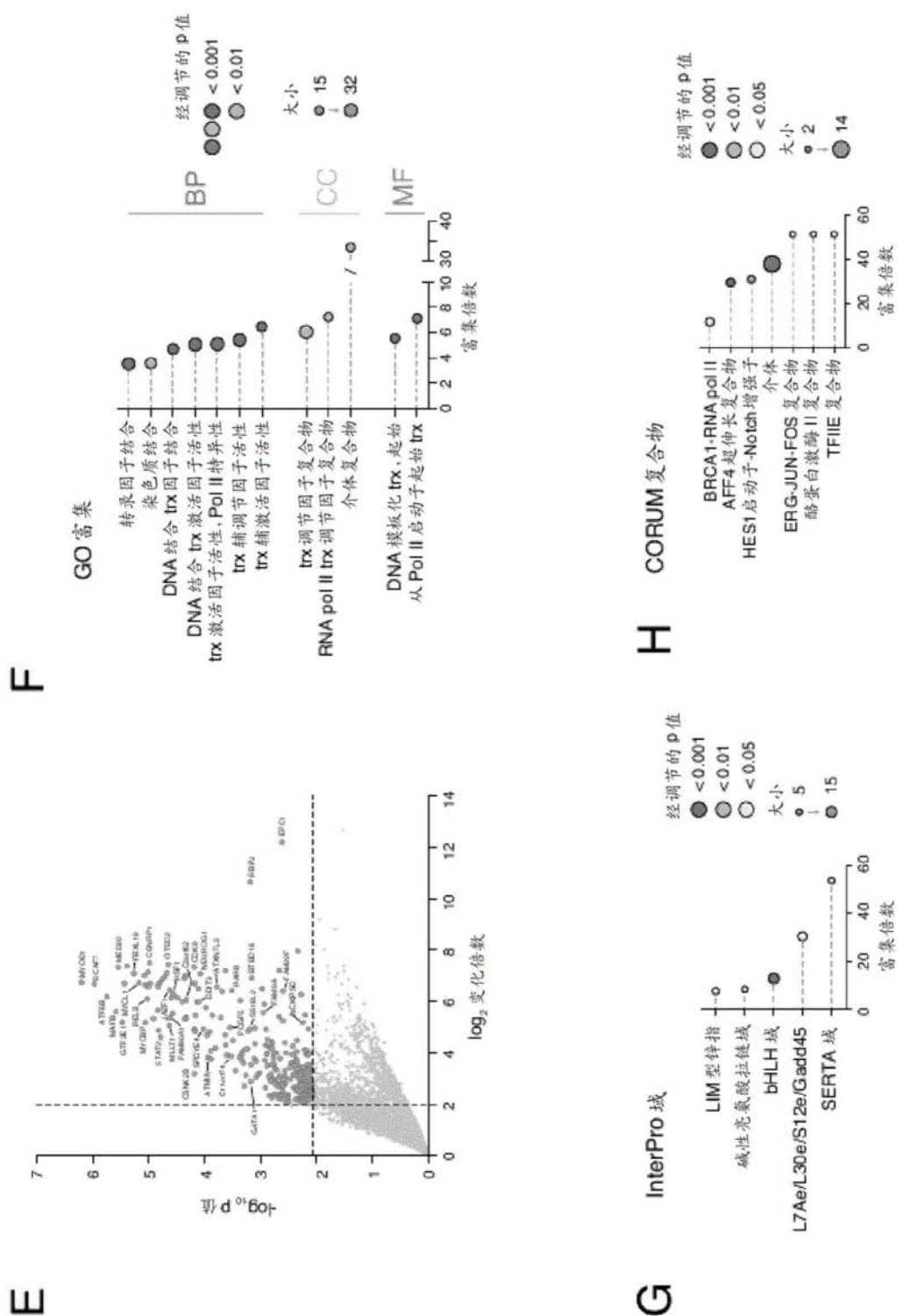
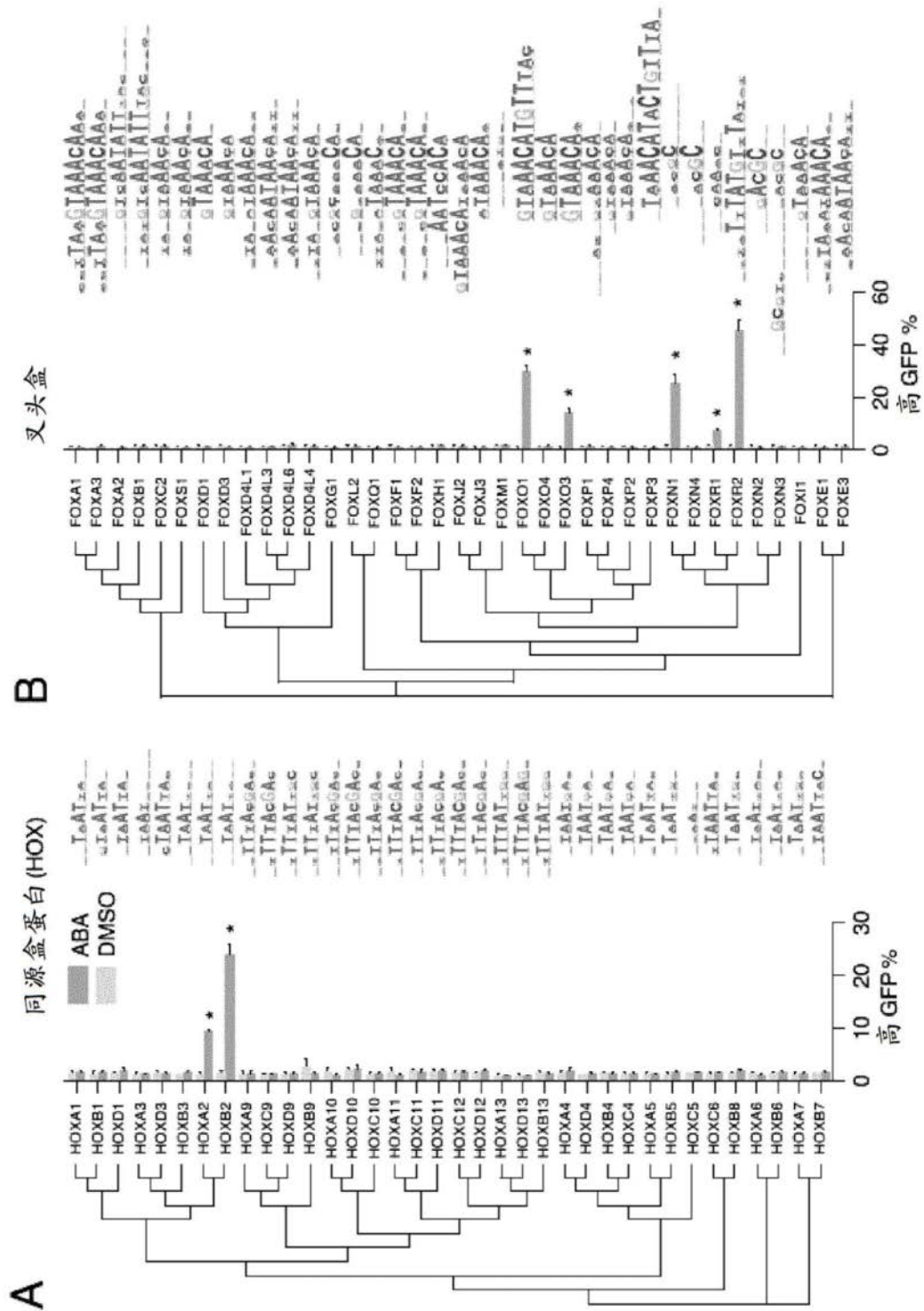


图1 (续)



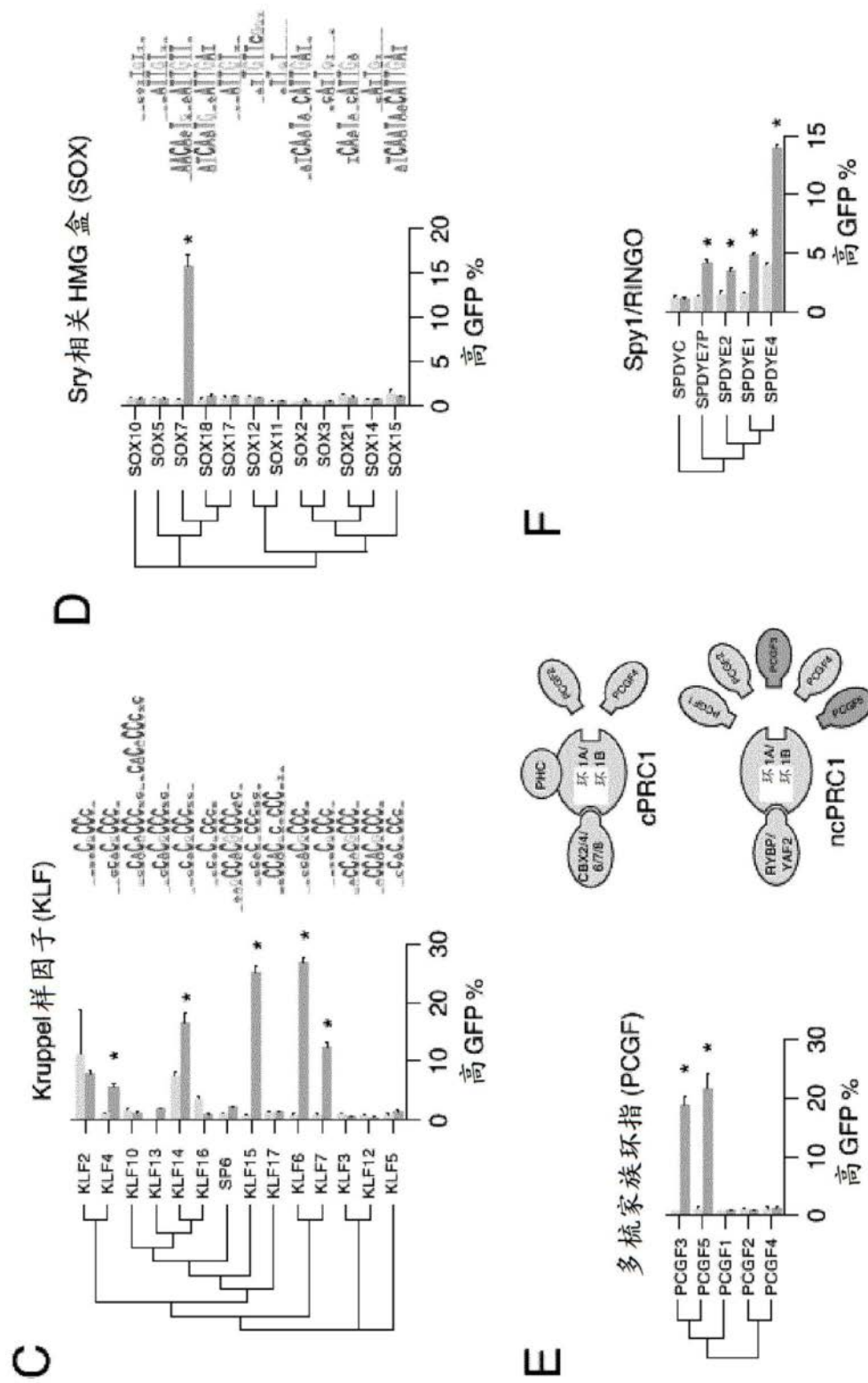


图2(续)

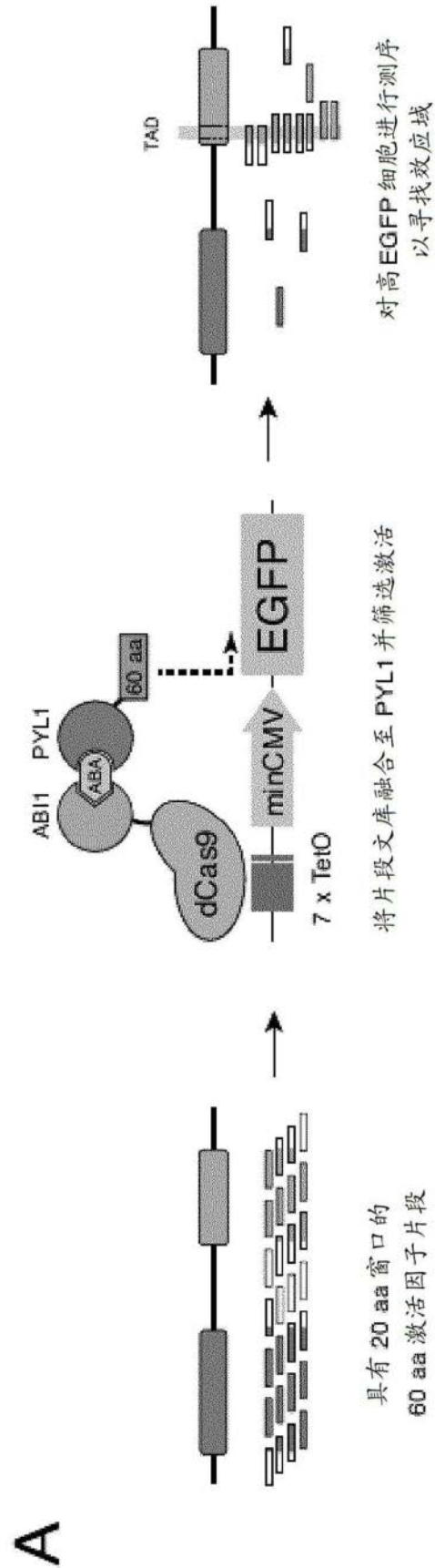


图3

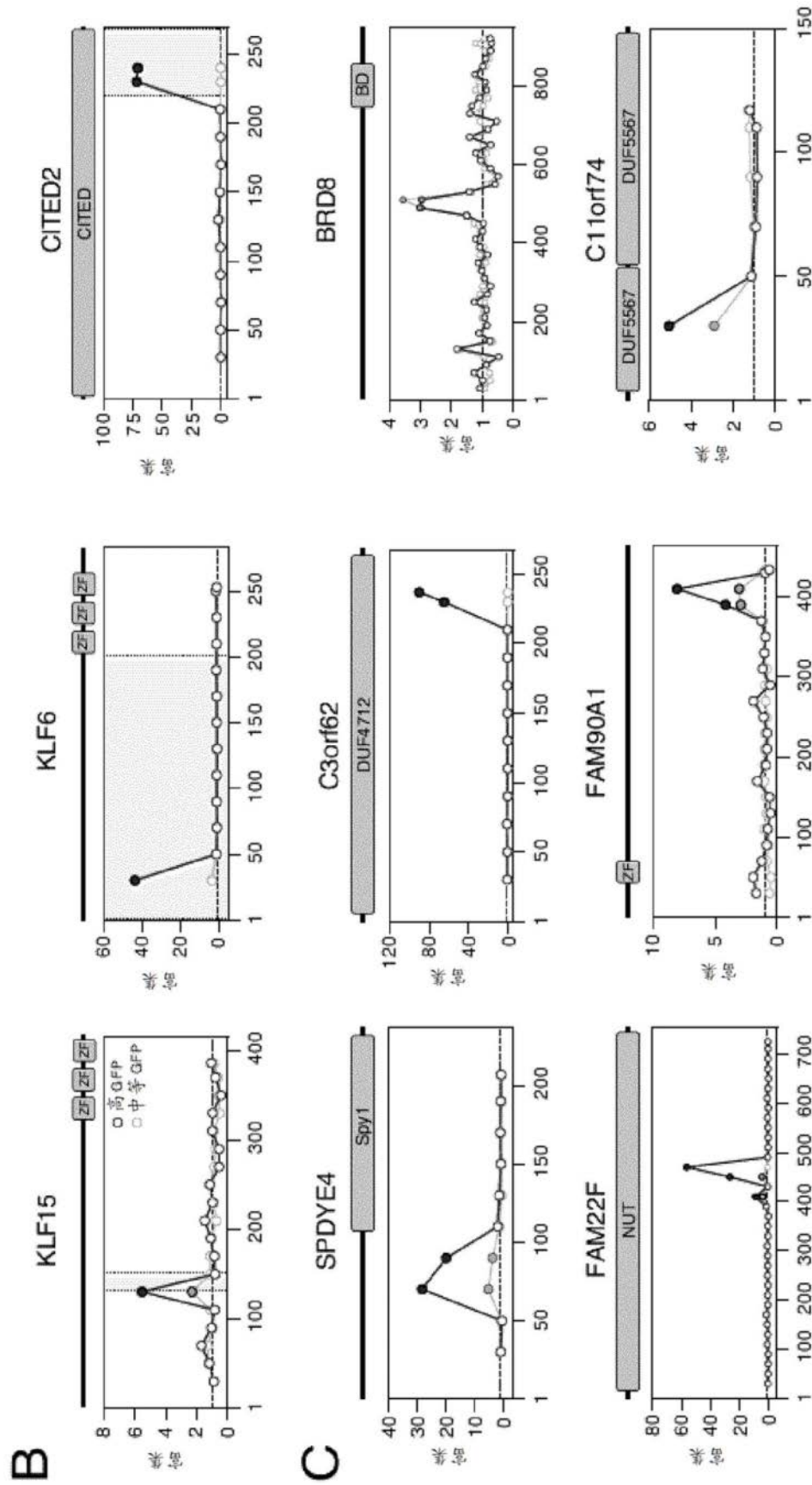


图3(续)





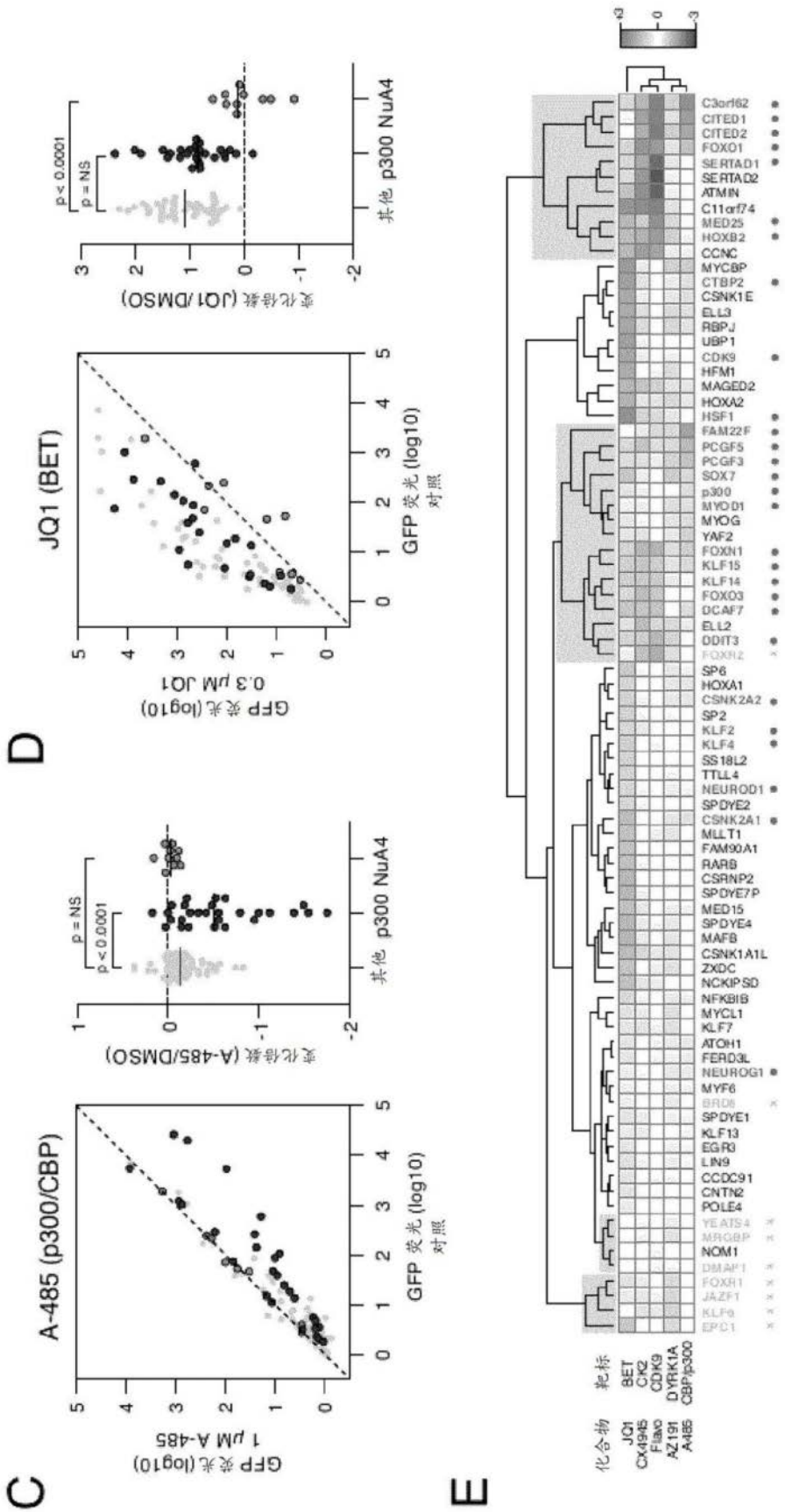


图4(续)

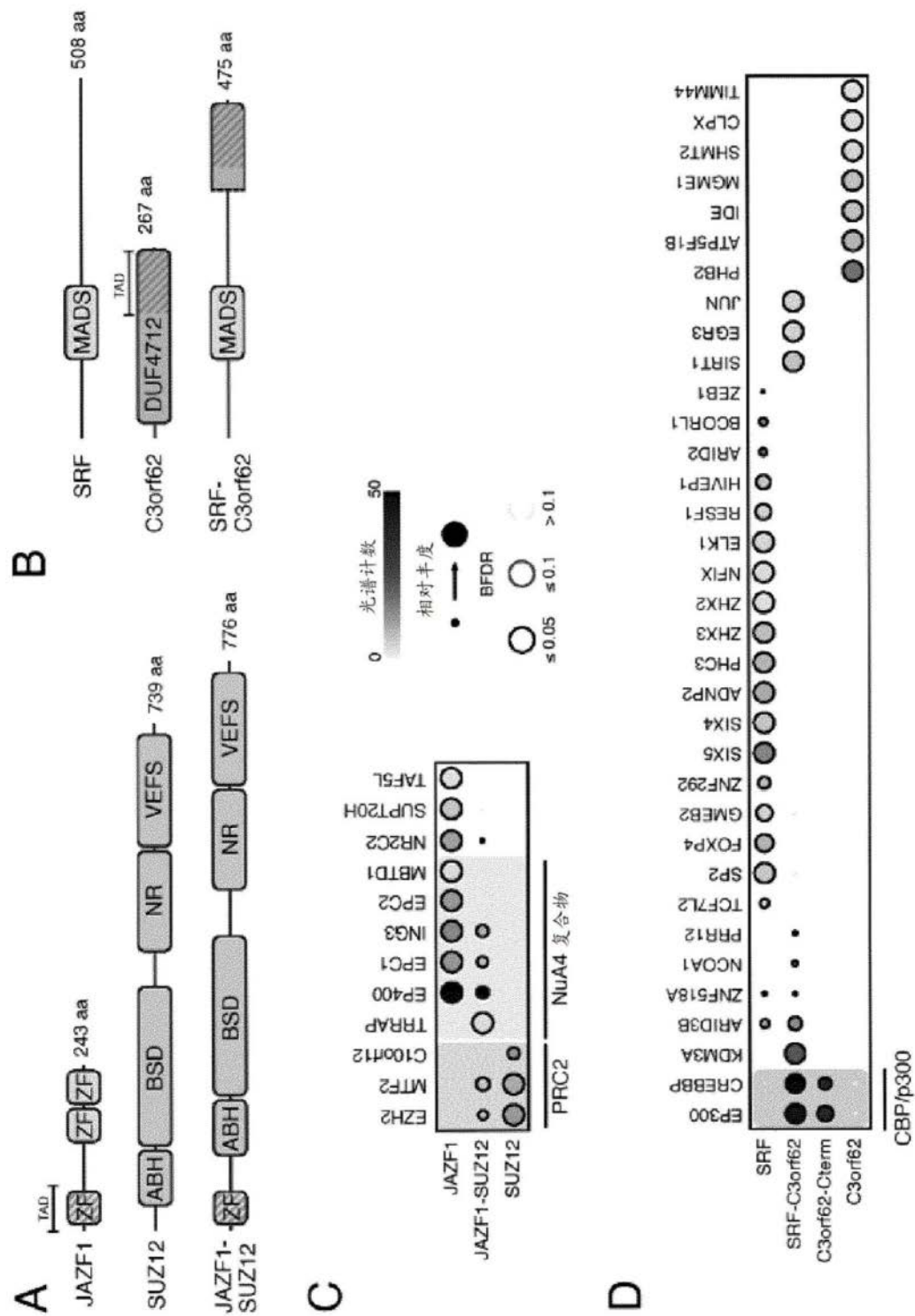


图5

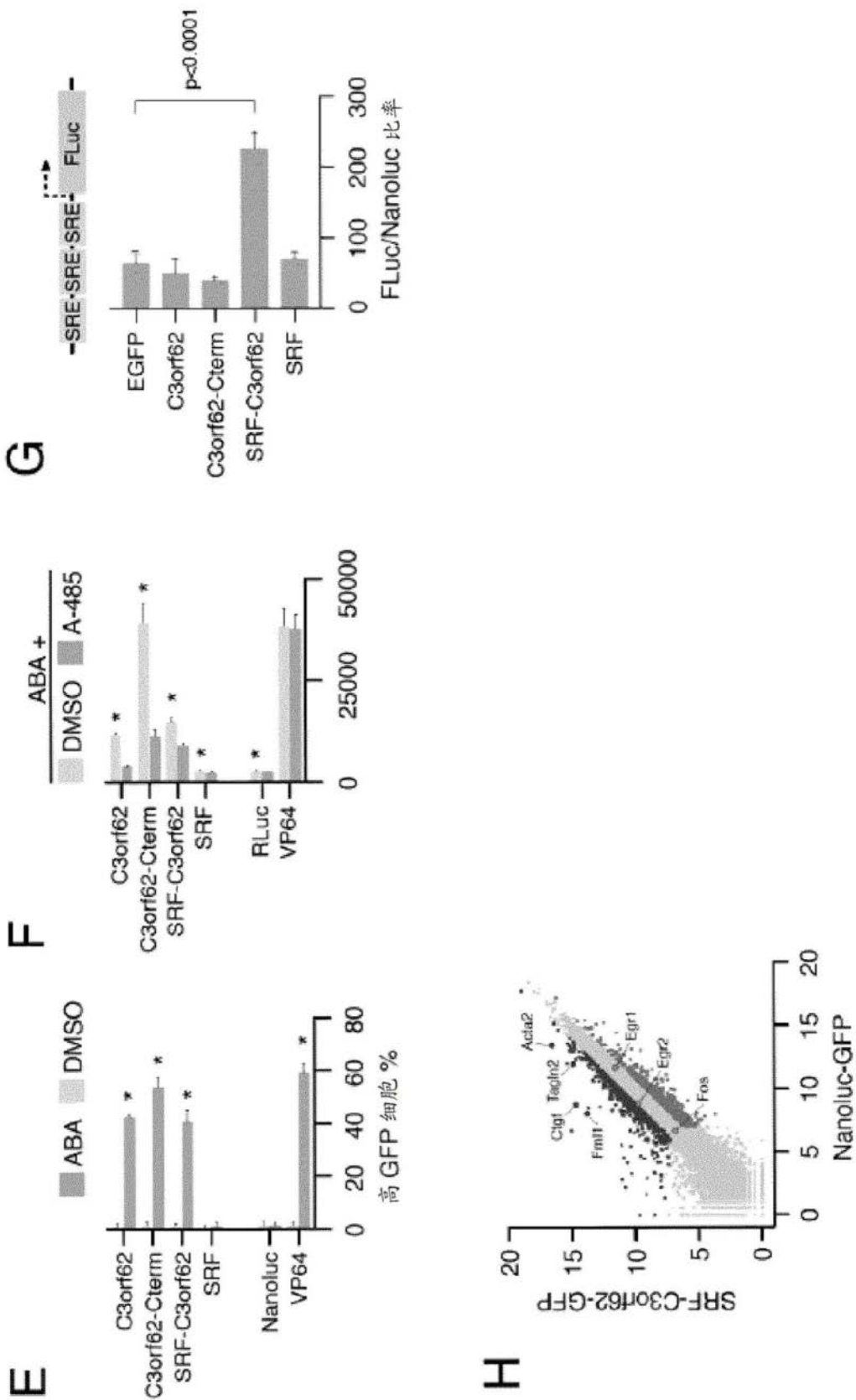


图5 (续)

I

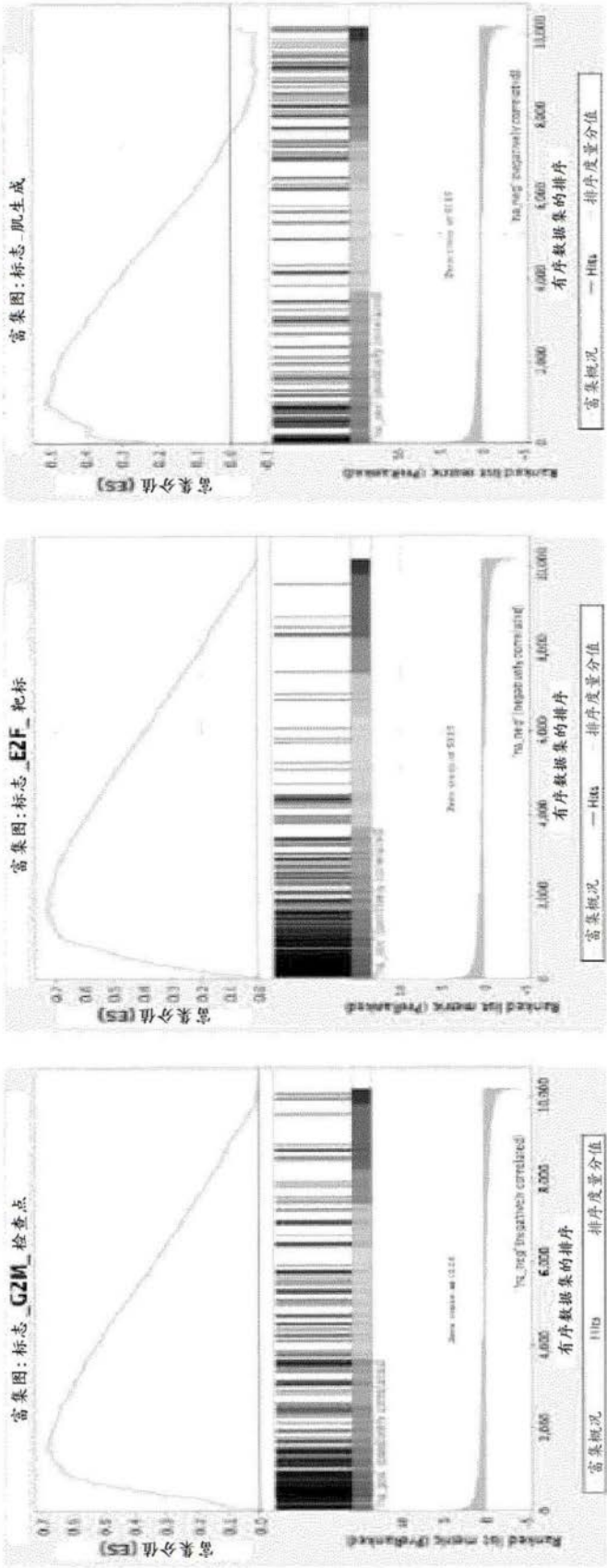


图5 (续)

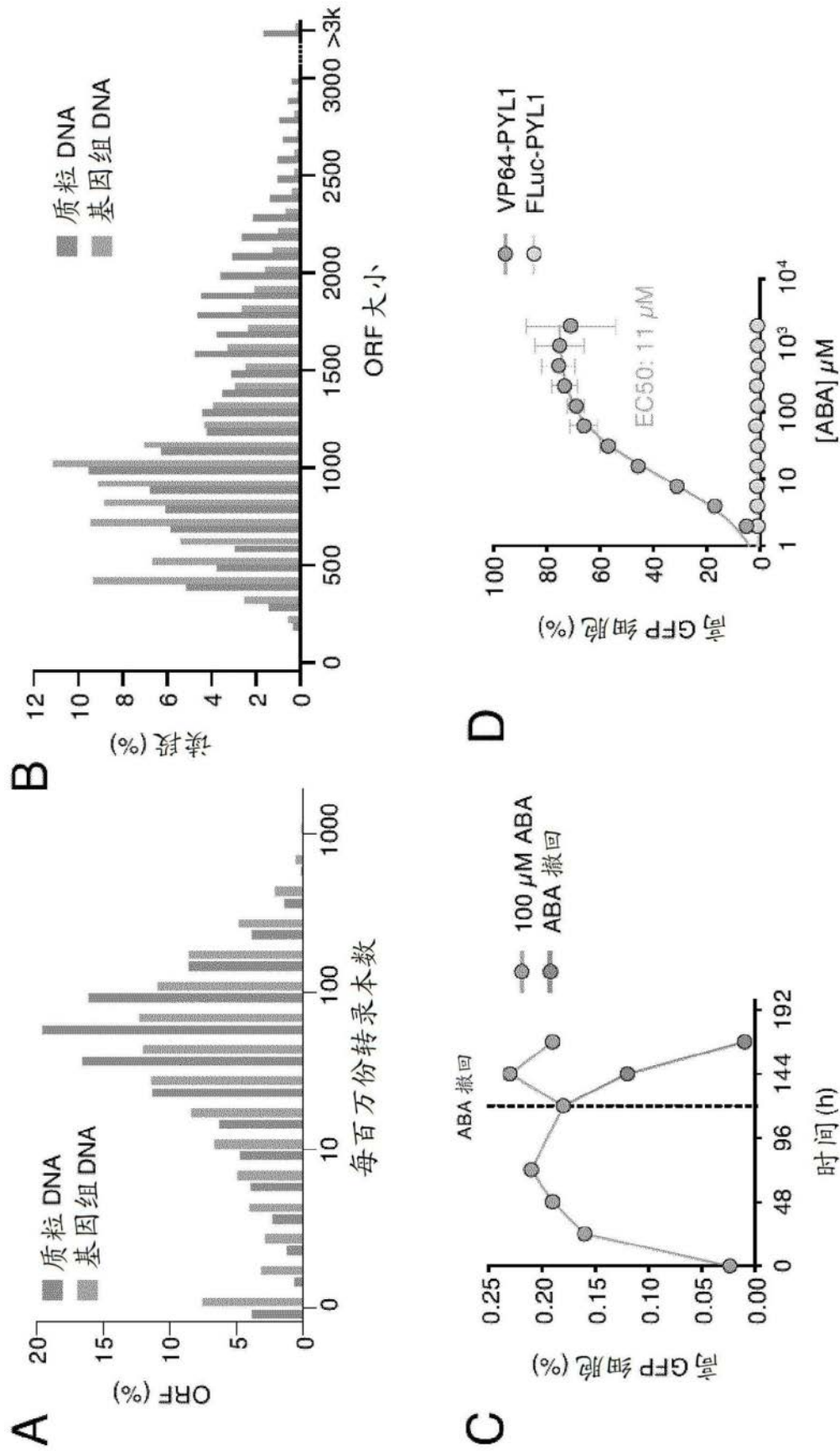


图6

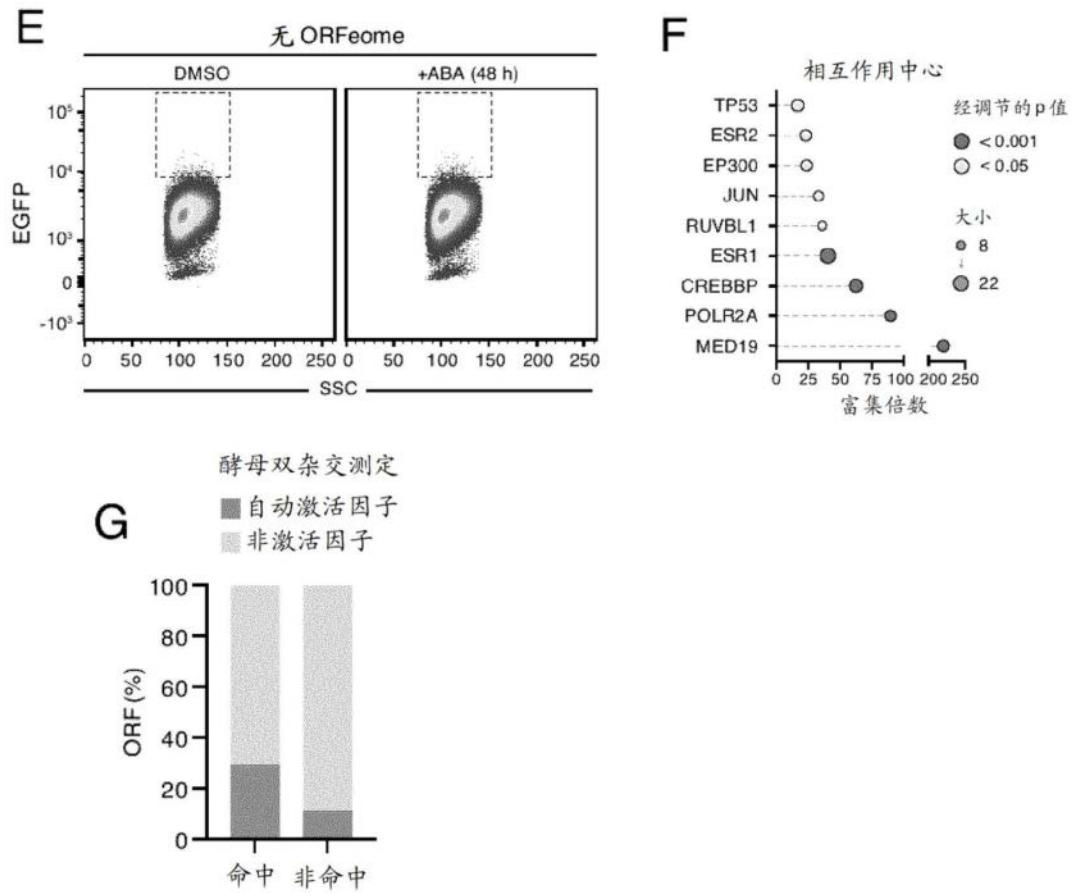


图6 (续)

H

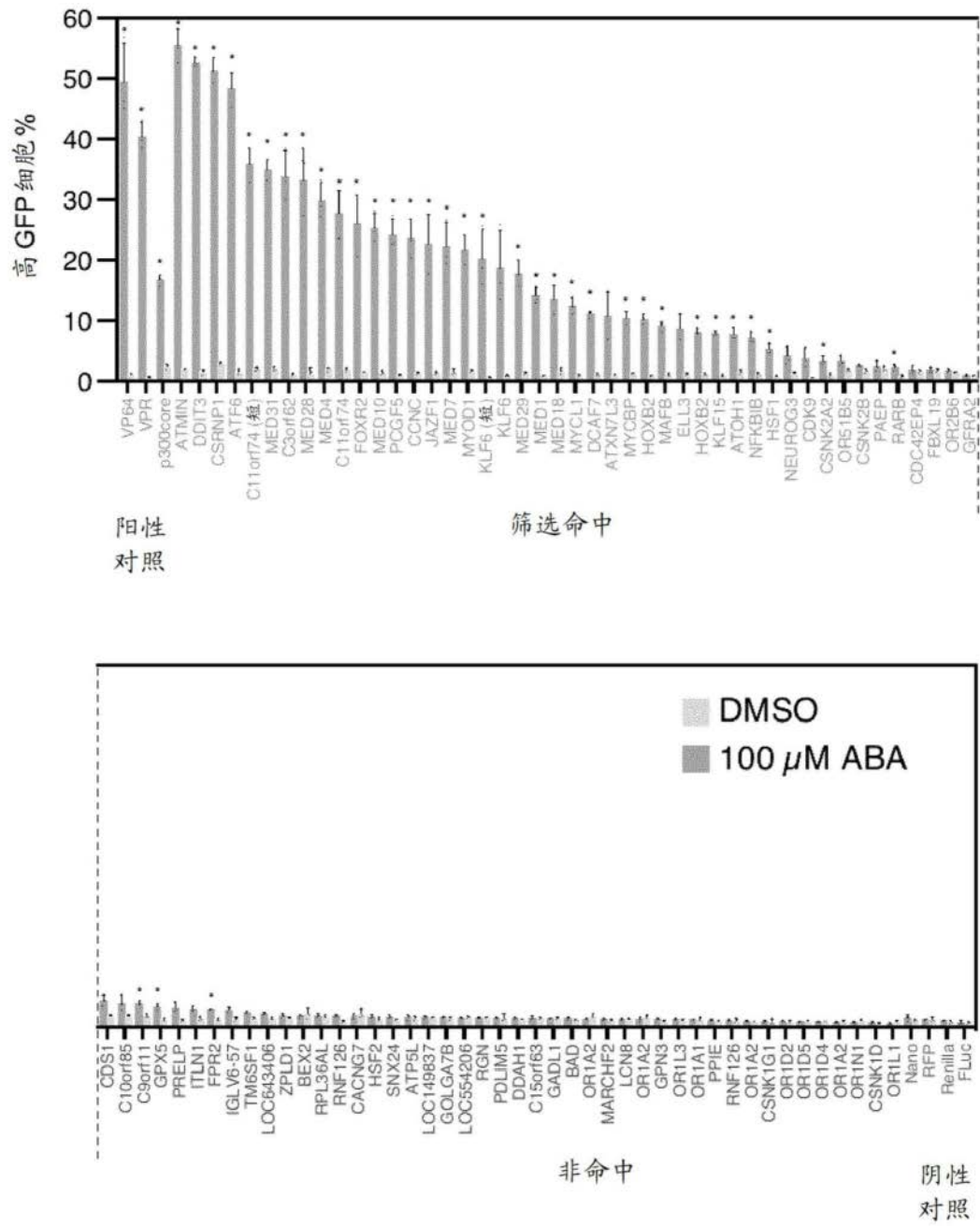


图6(续)

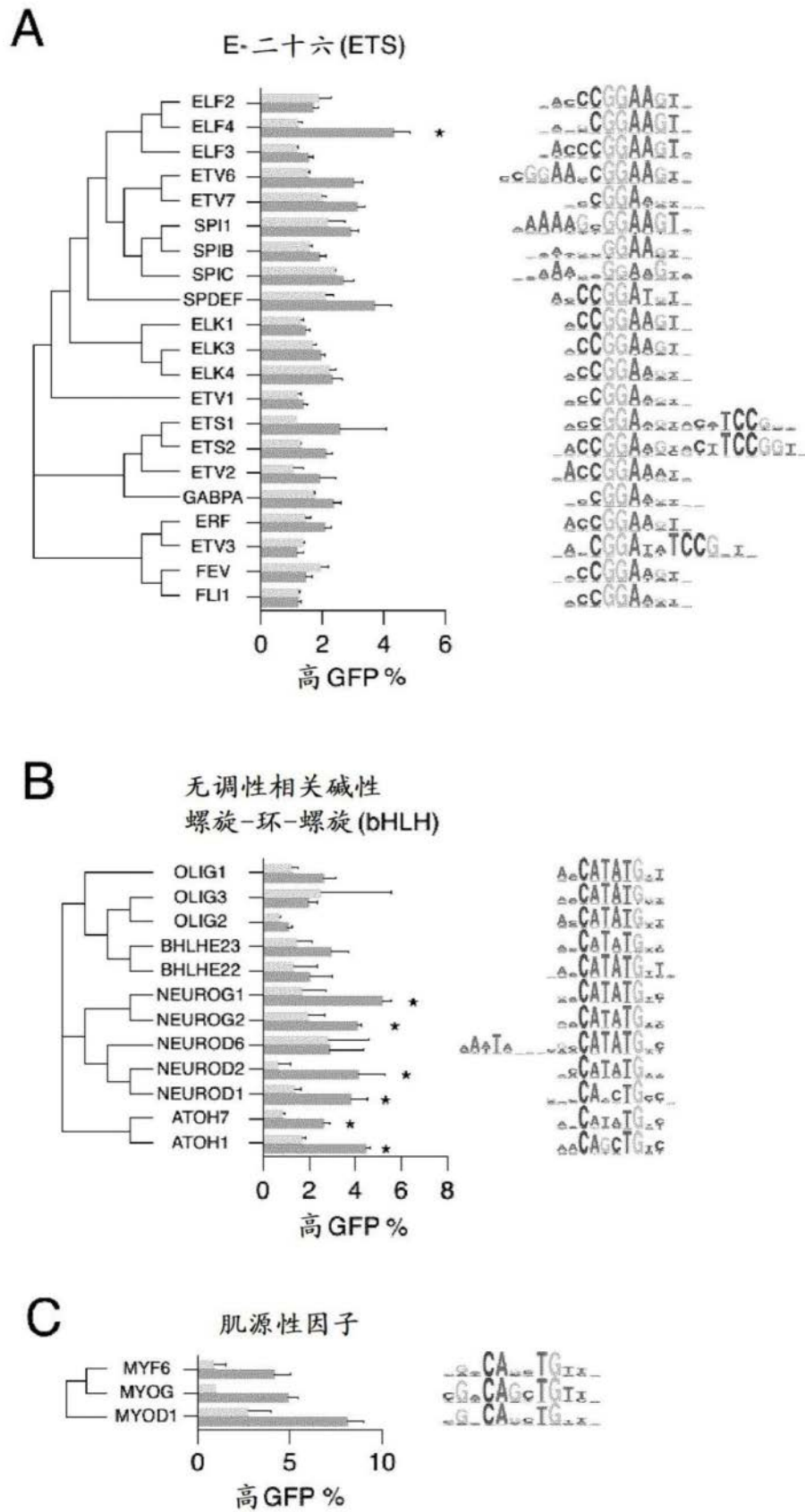


图7

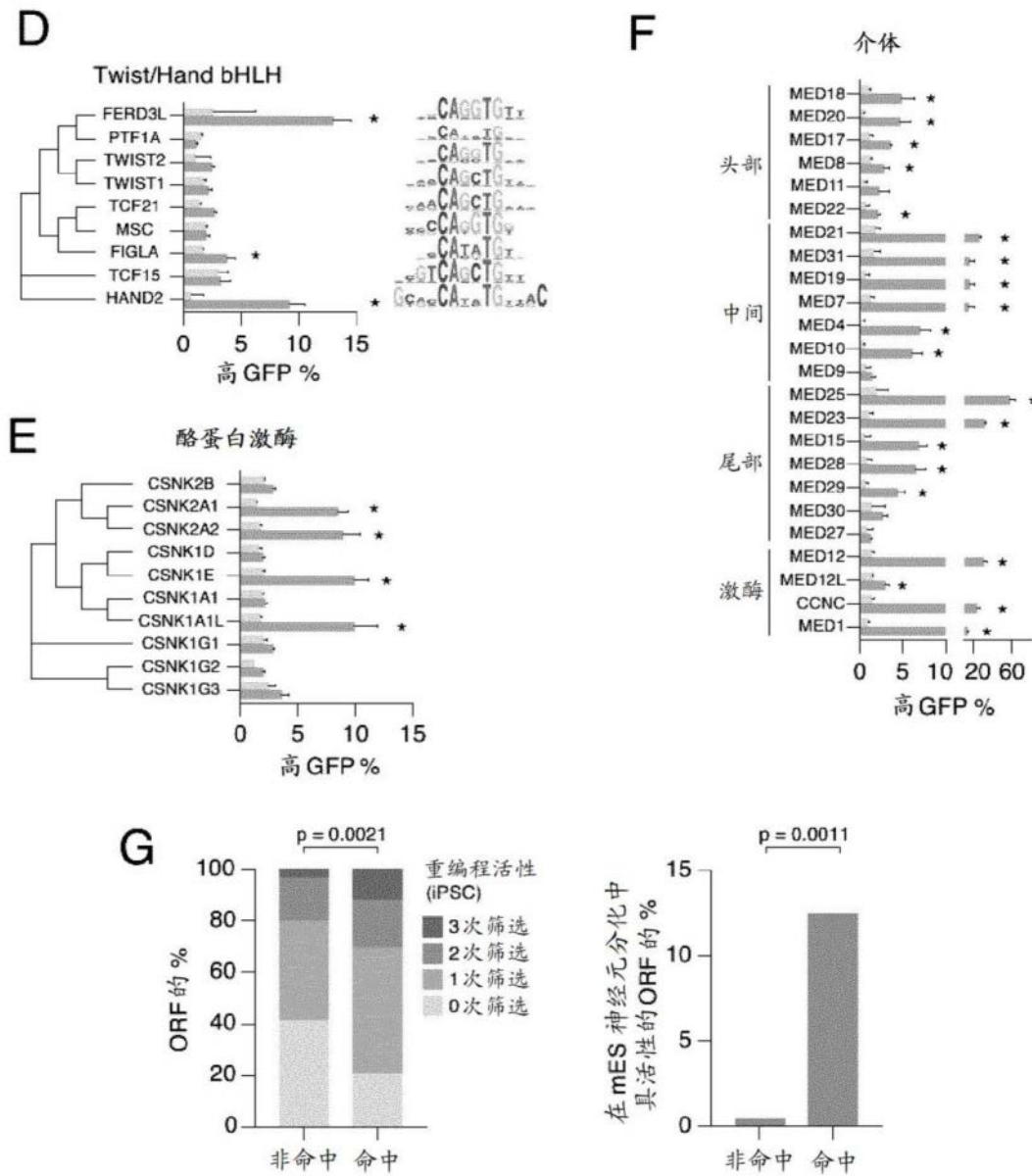


图7(续)



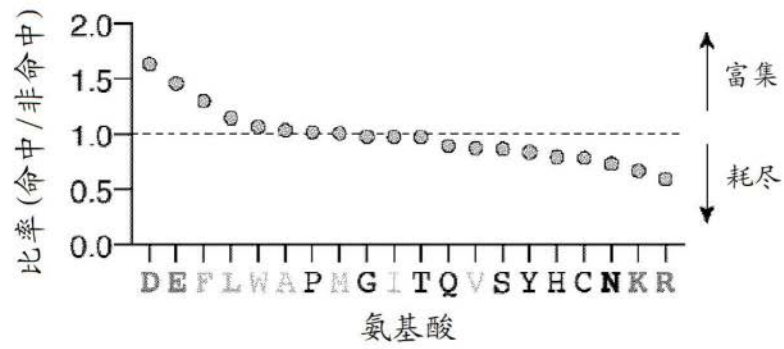
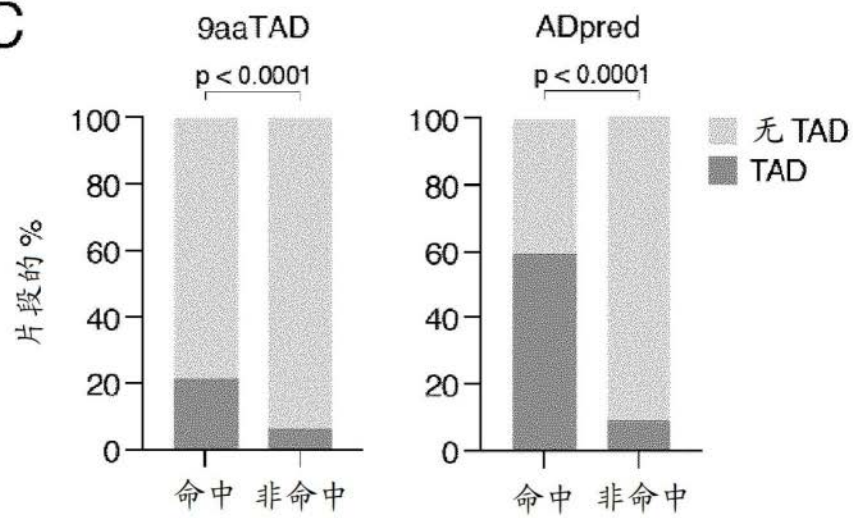
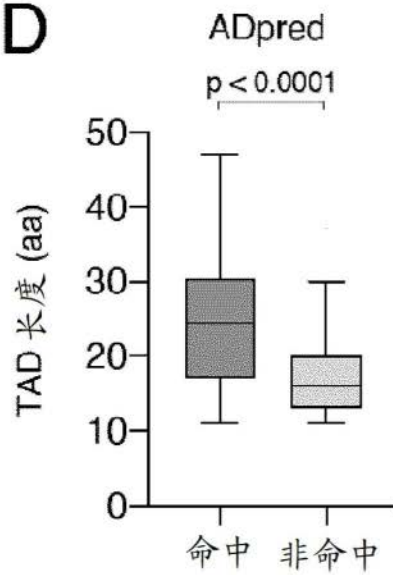
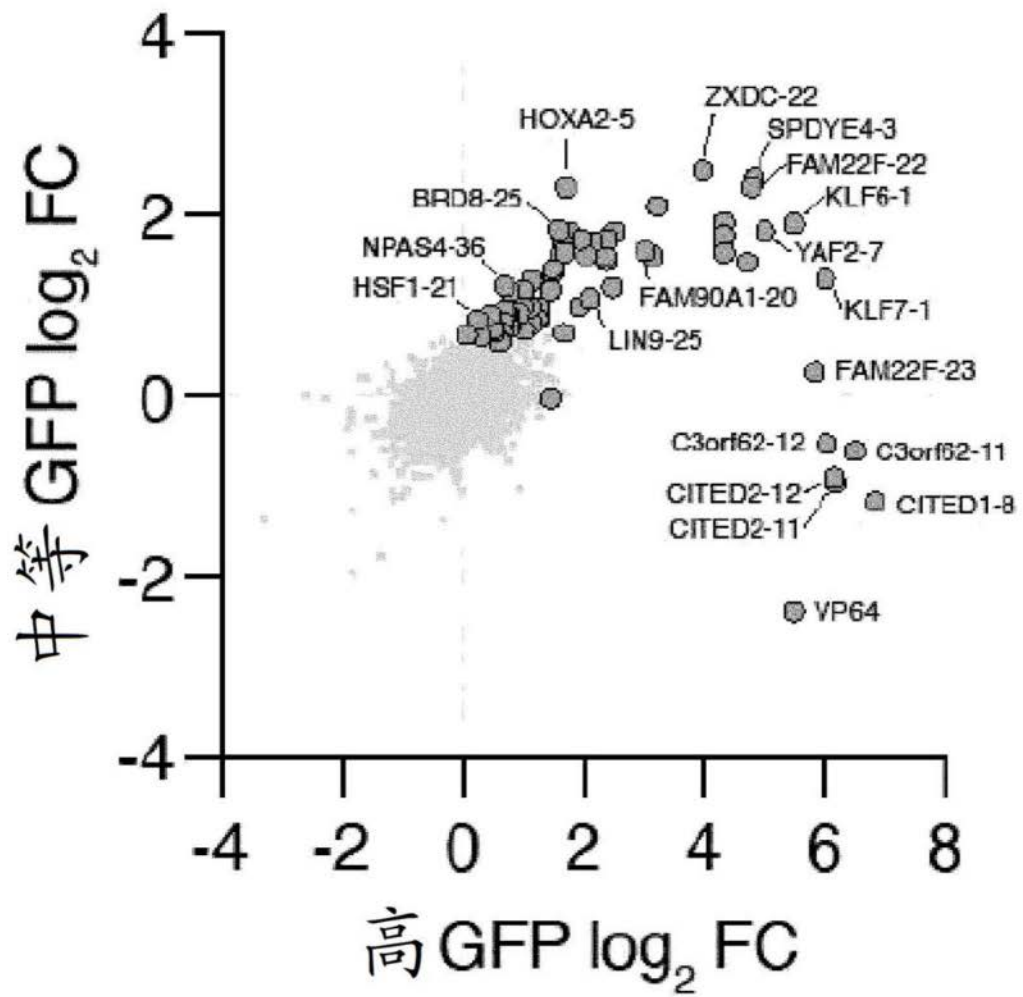
**B****C****D**

图9(续)

E



F

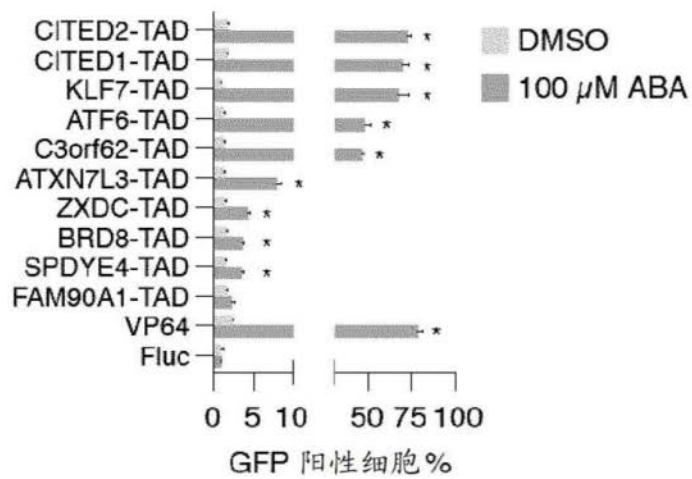


图9(续)

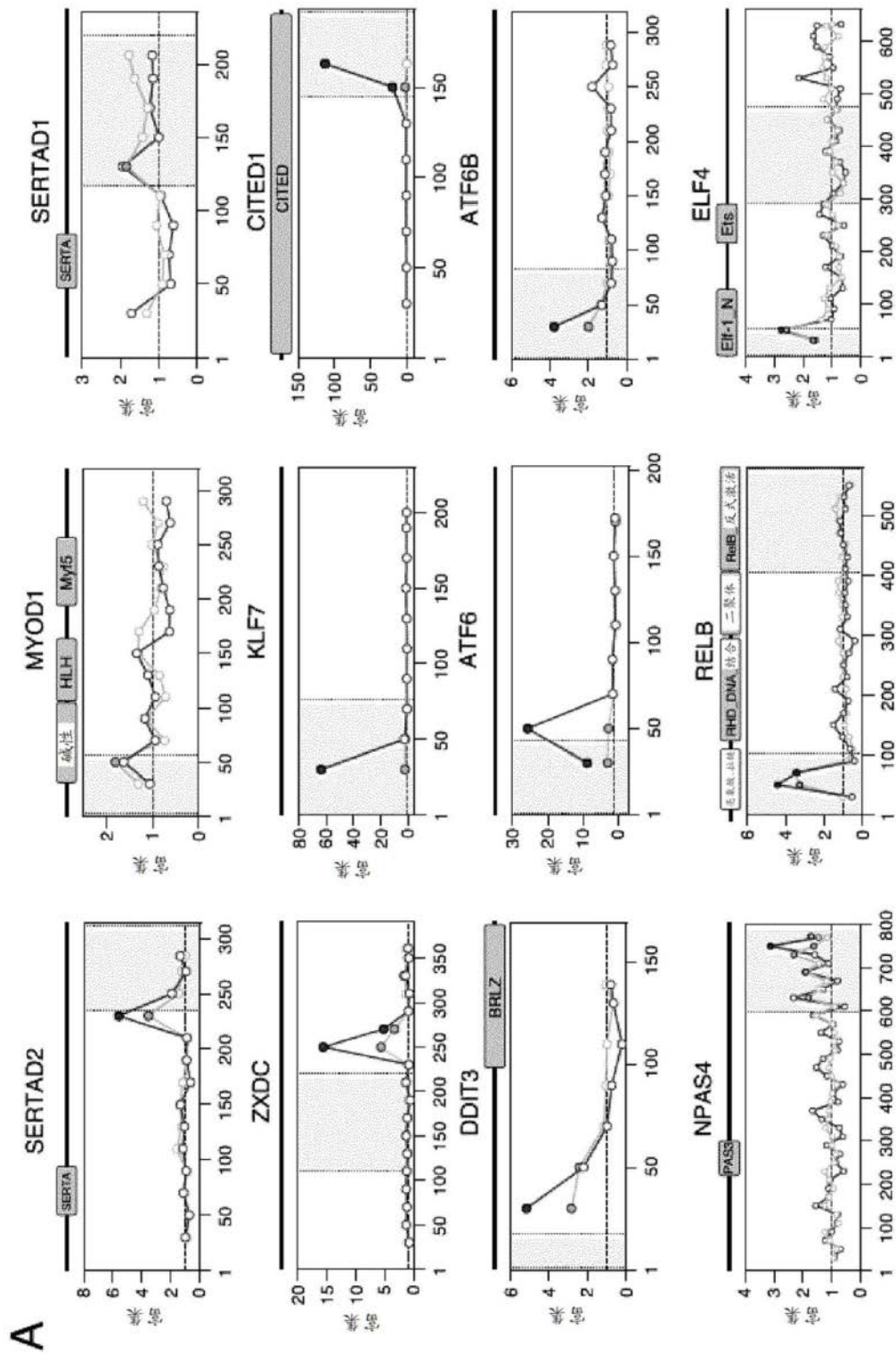


图10

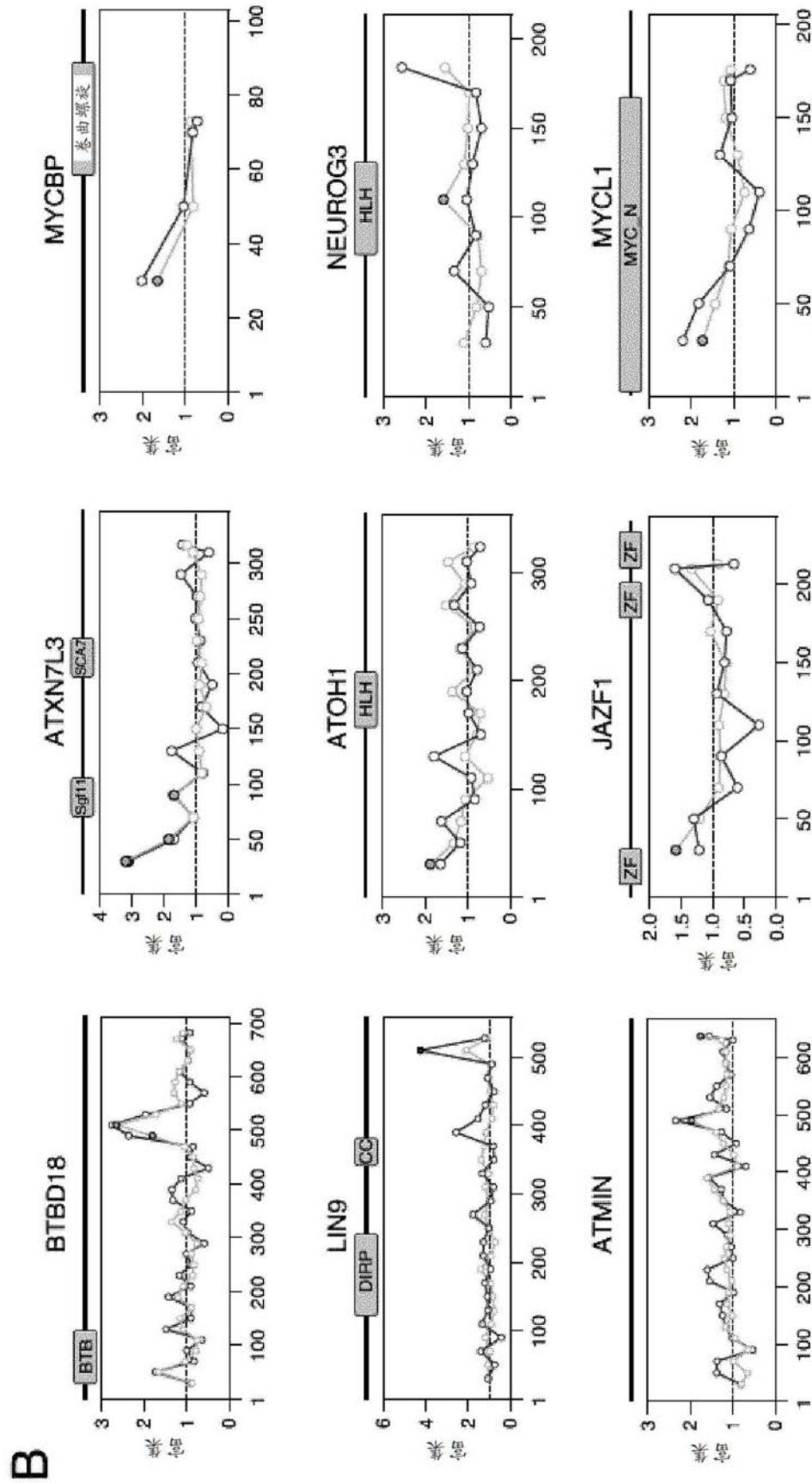


图10(续)

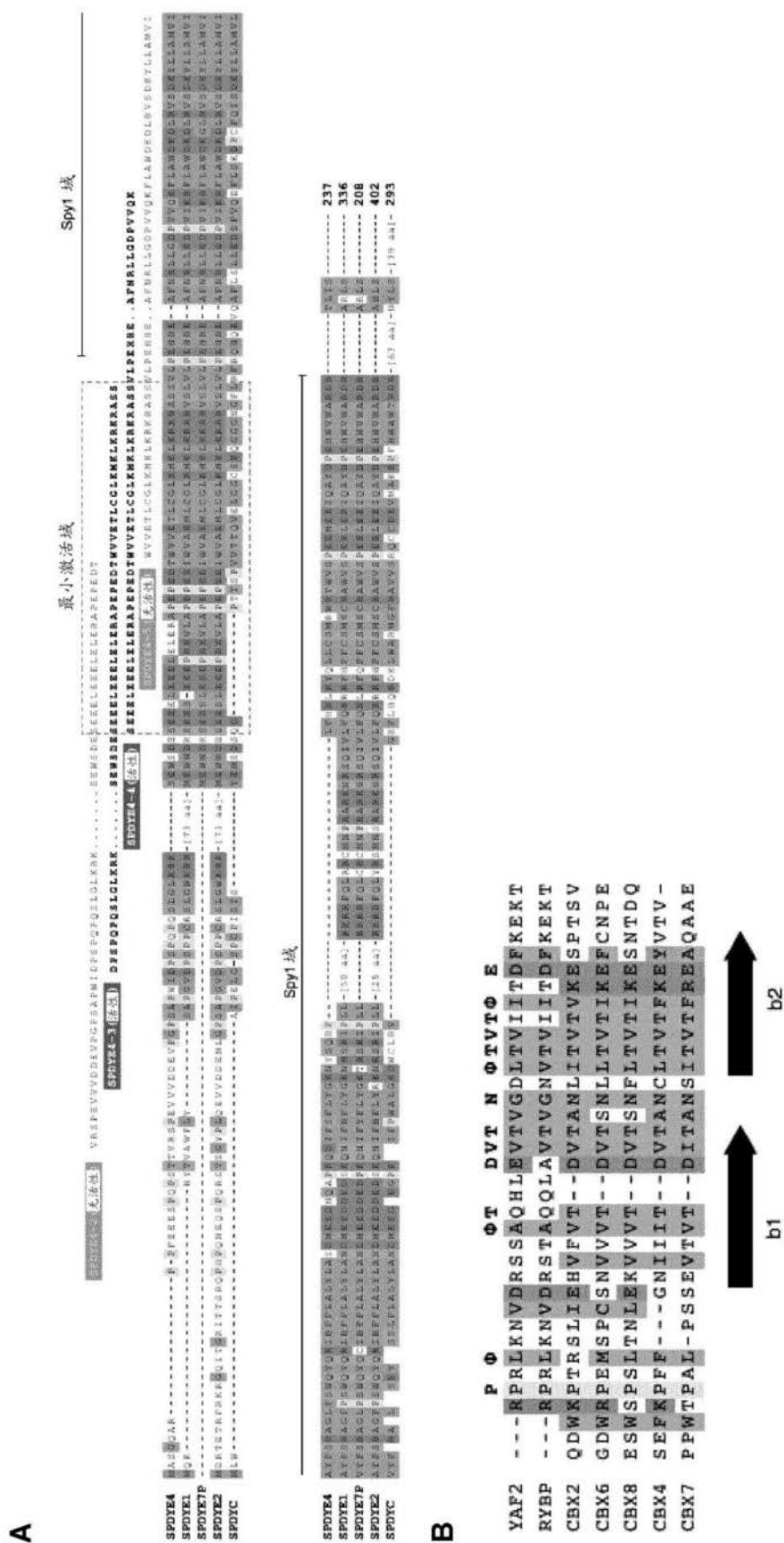


图11

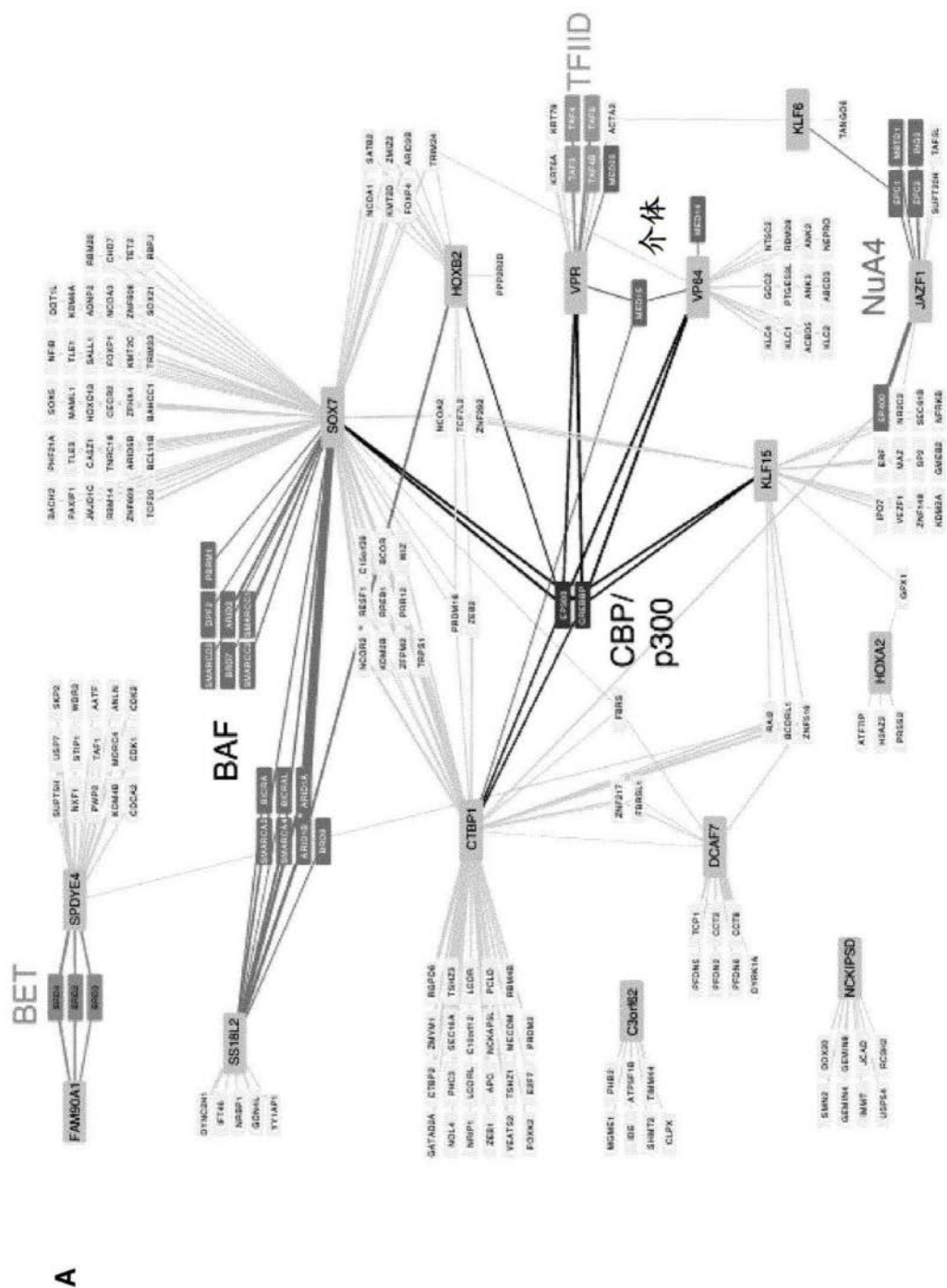


图12

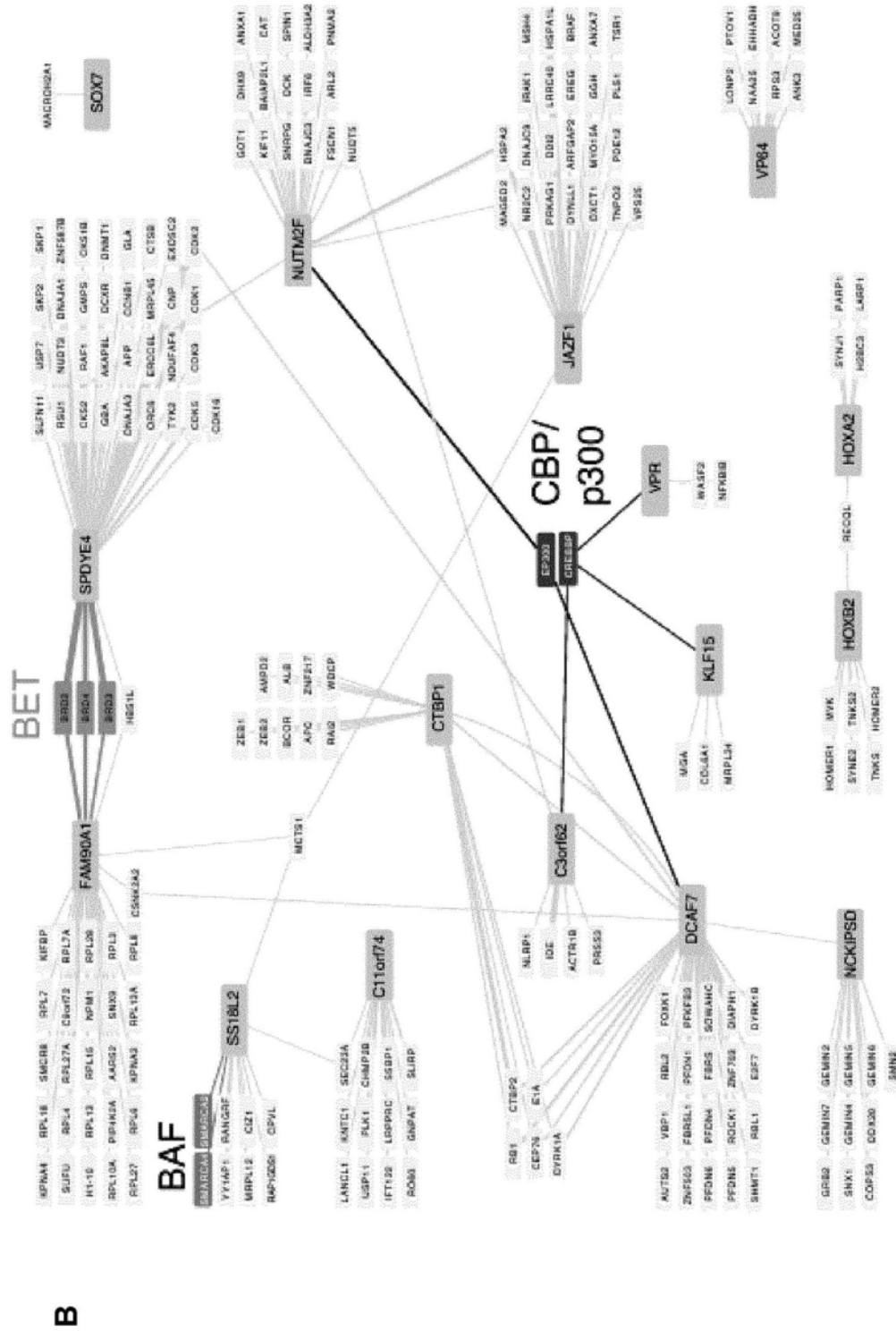


图12(续)

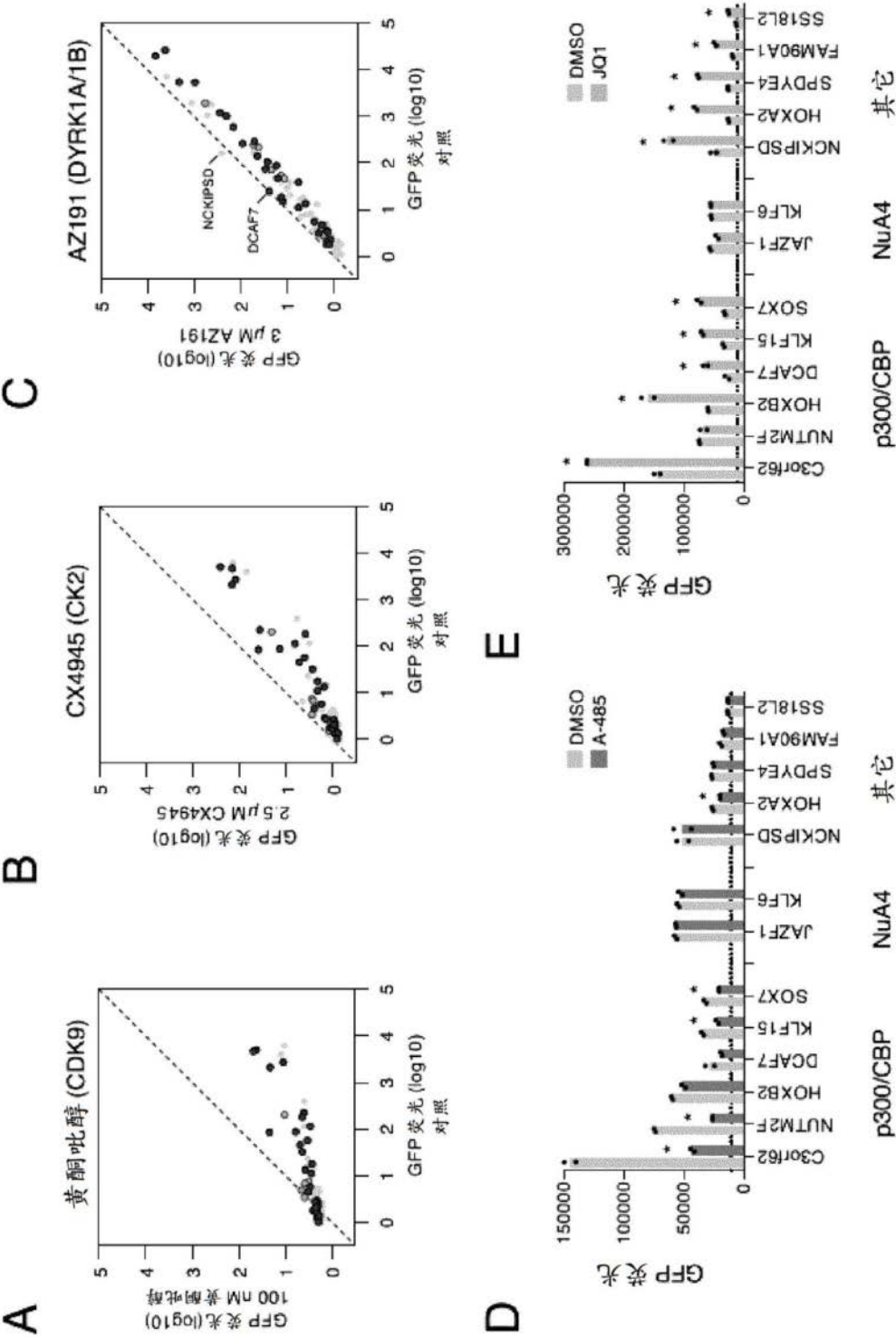


图13

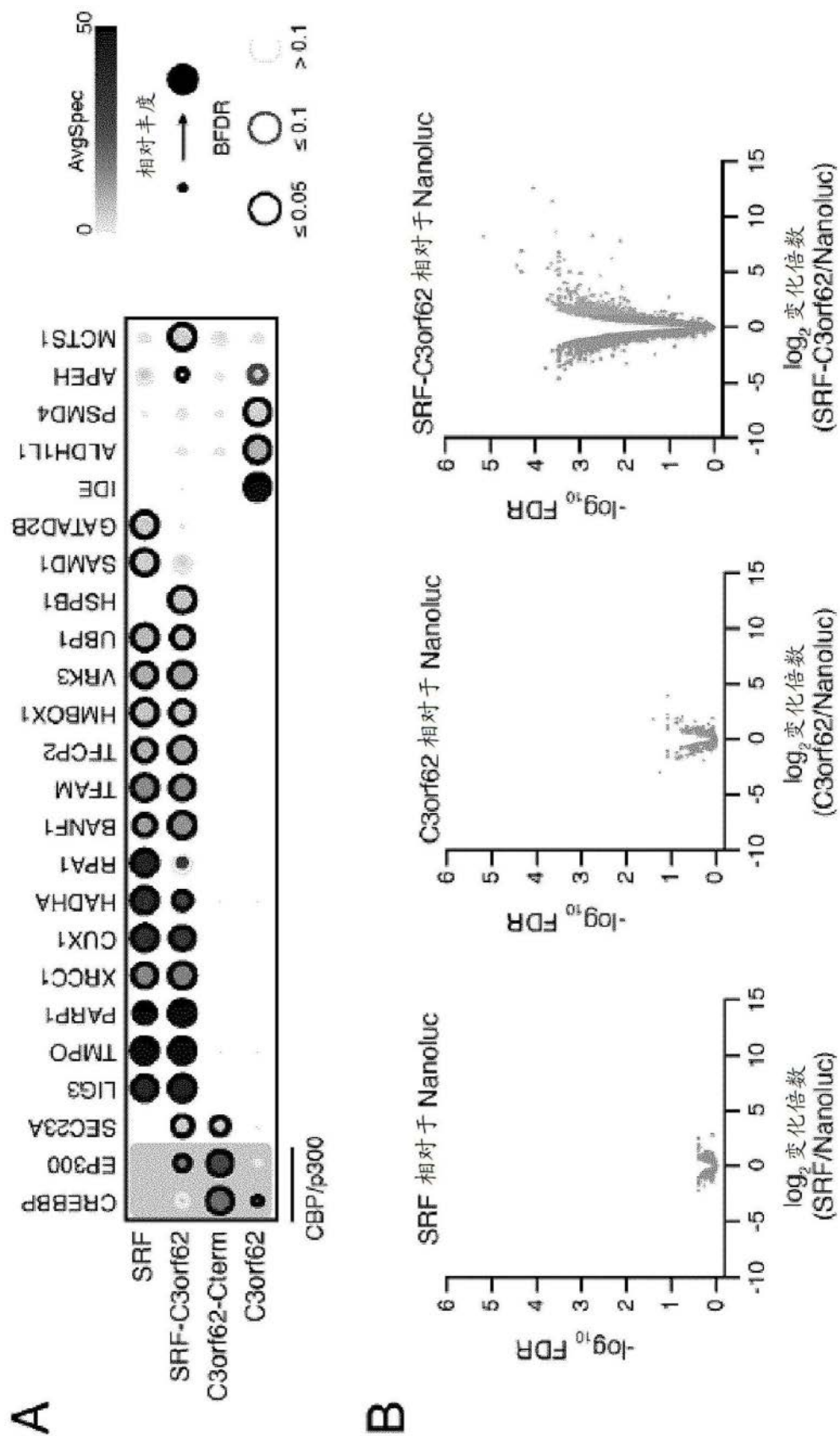


图14

C

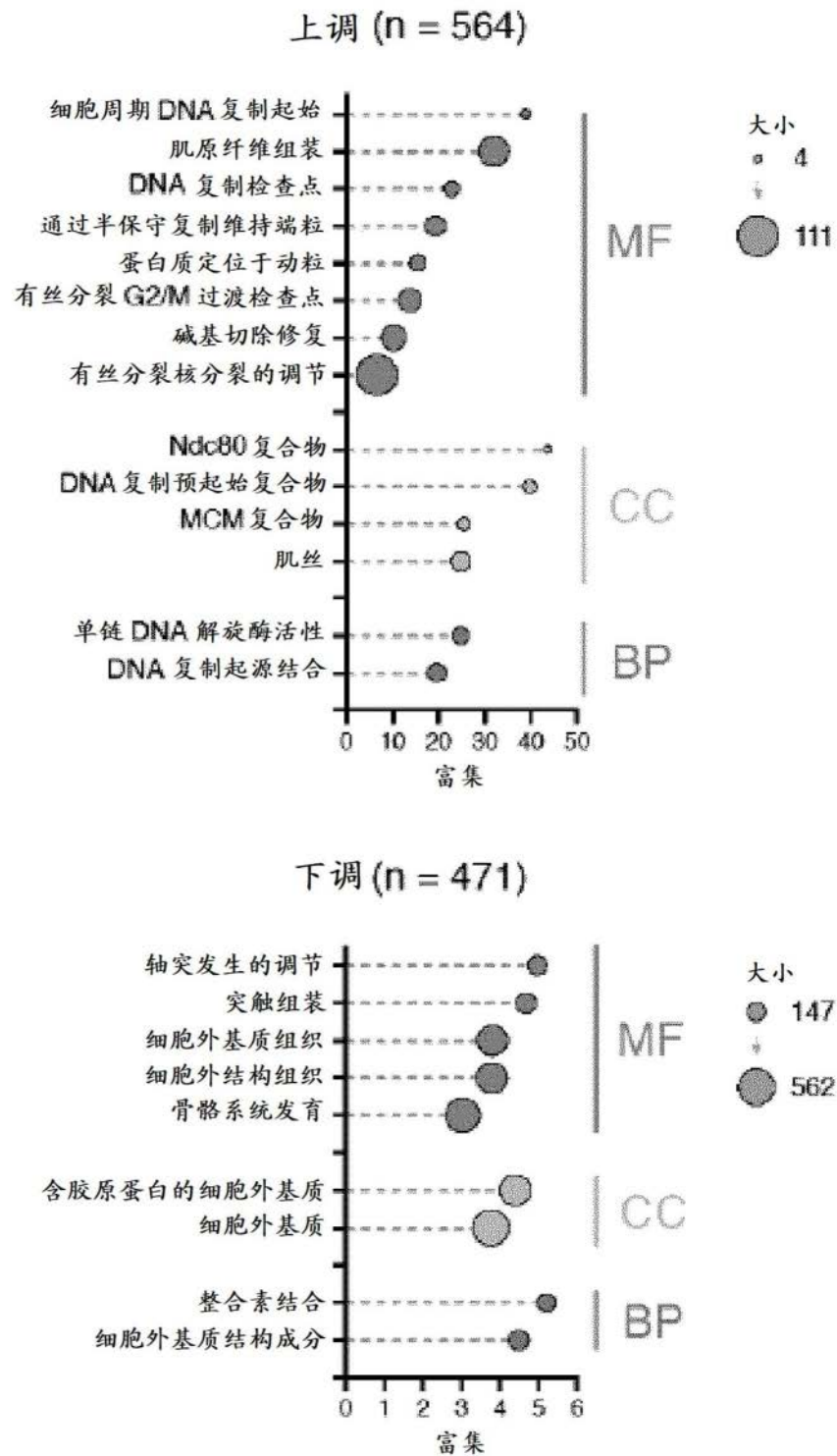


图14(续)

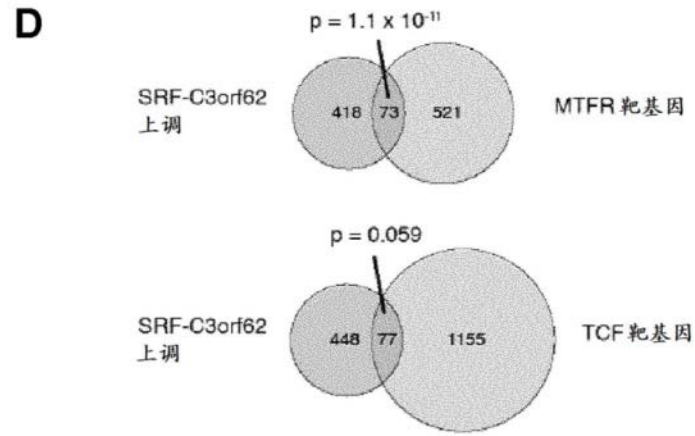


图14(续)

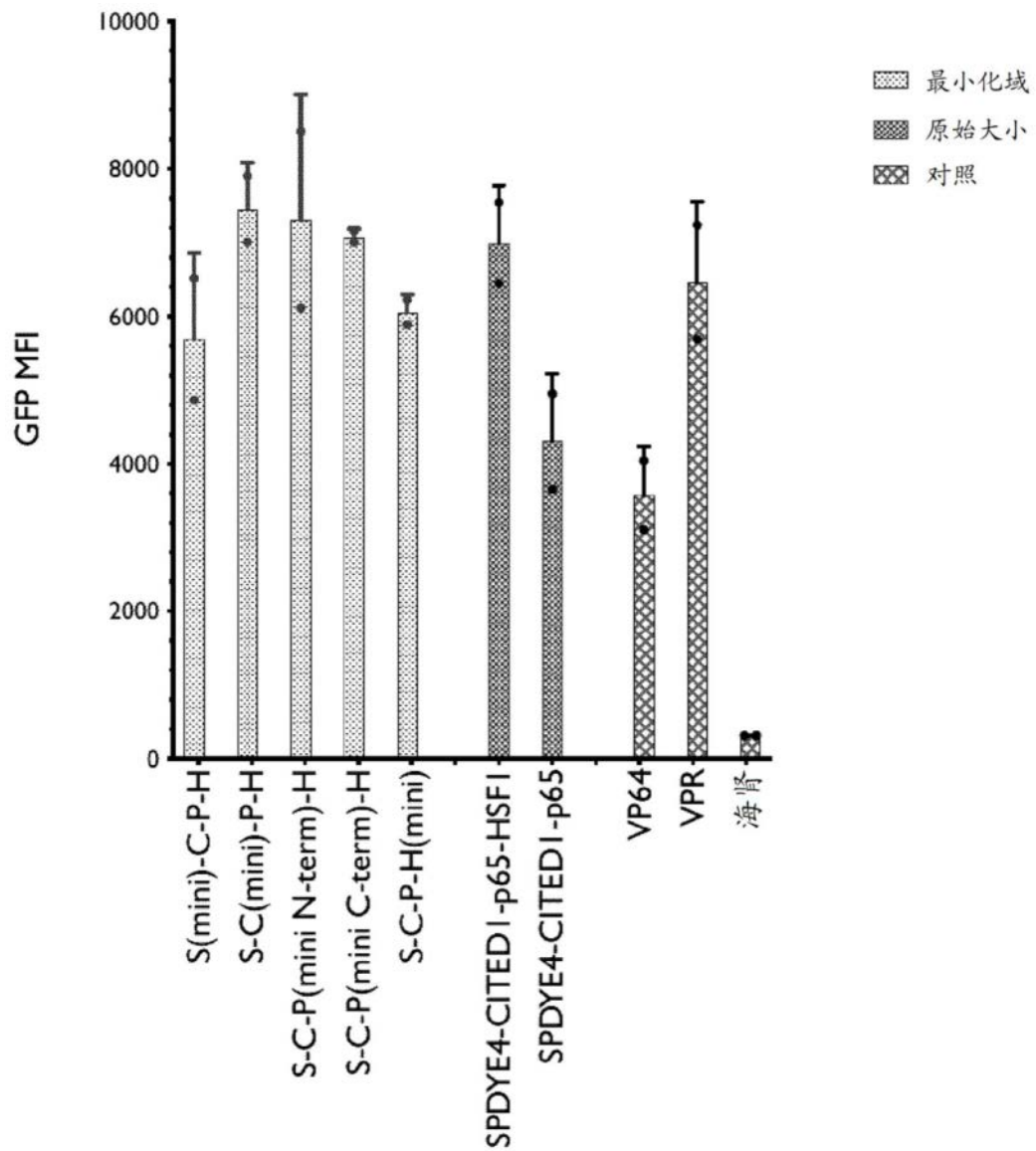


图15

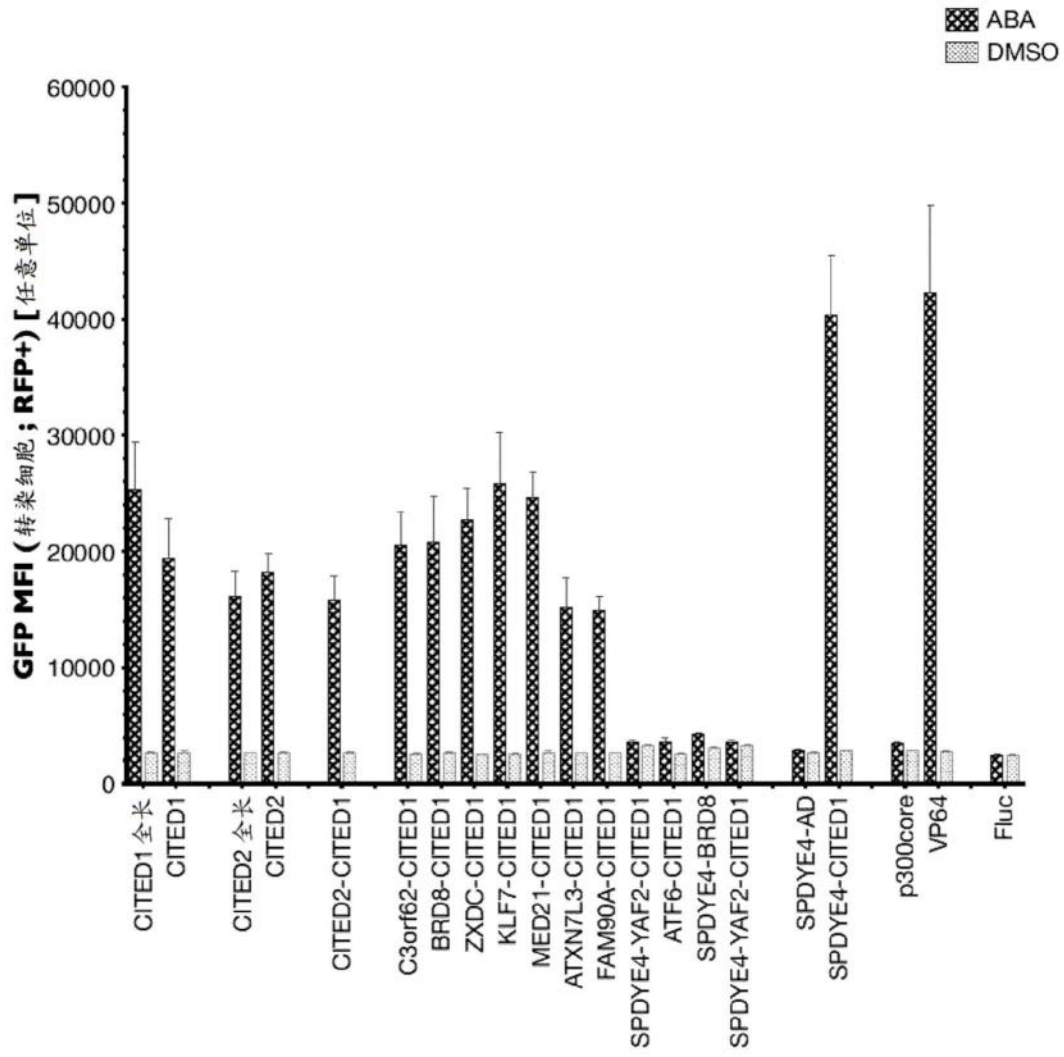


图16



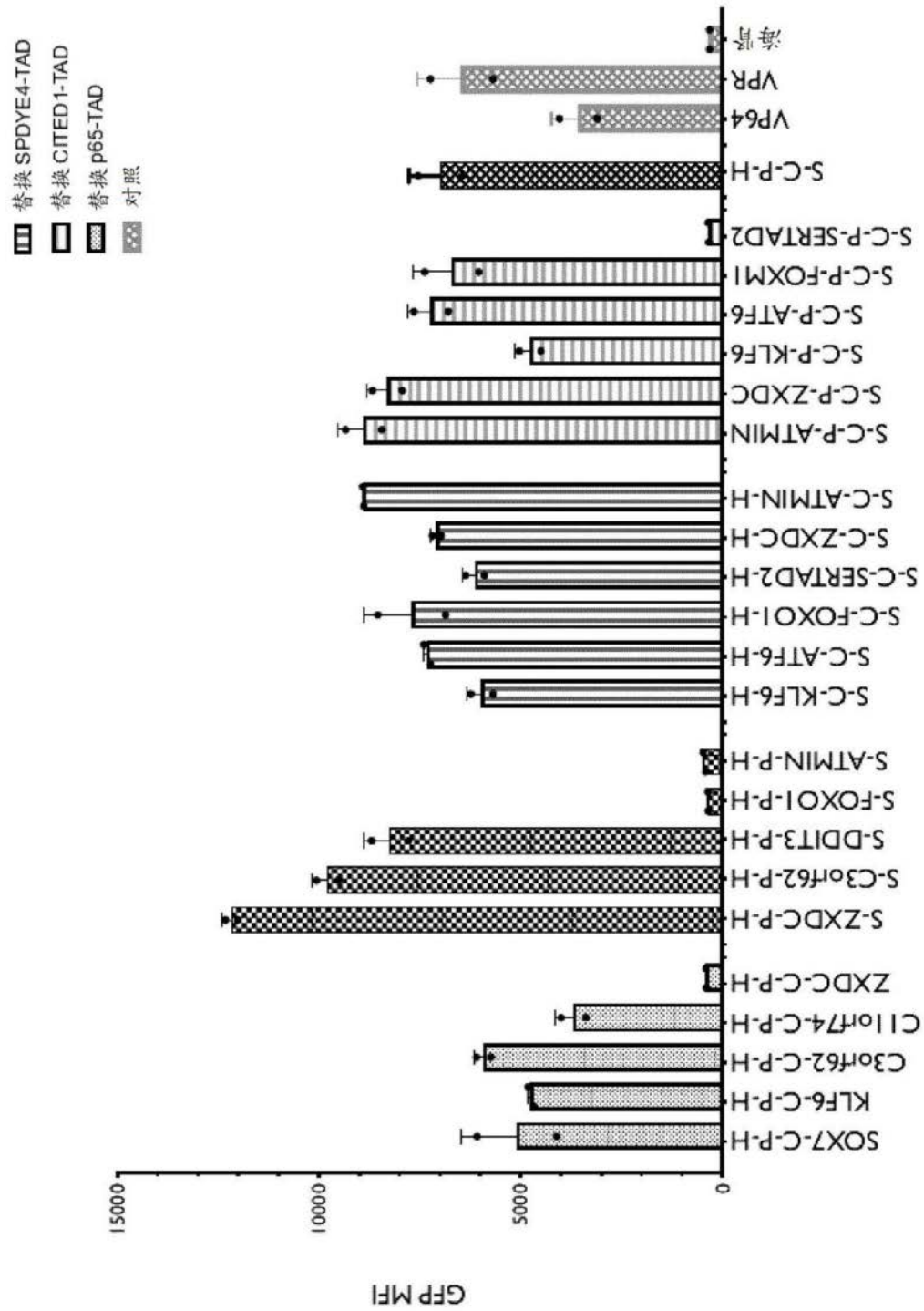
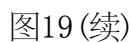


图18





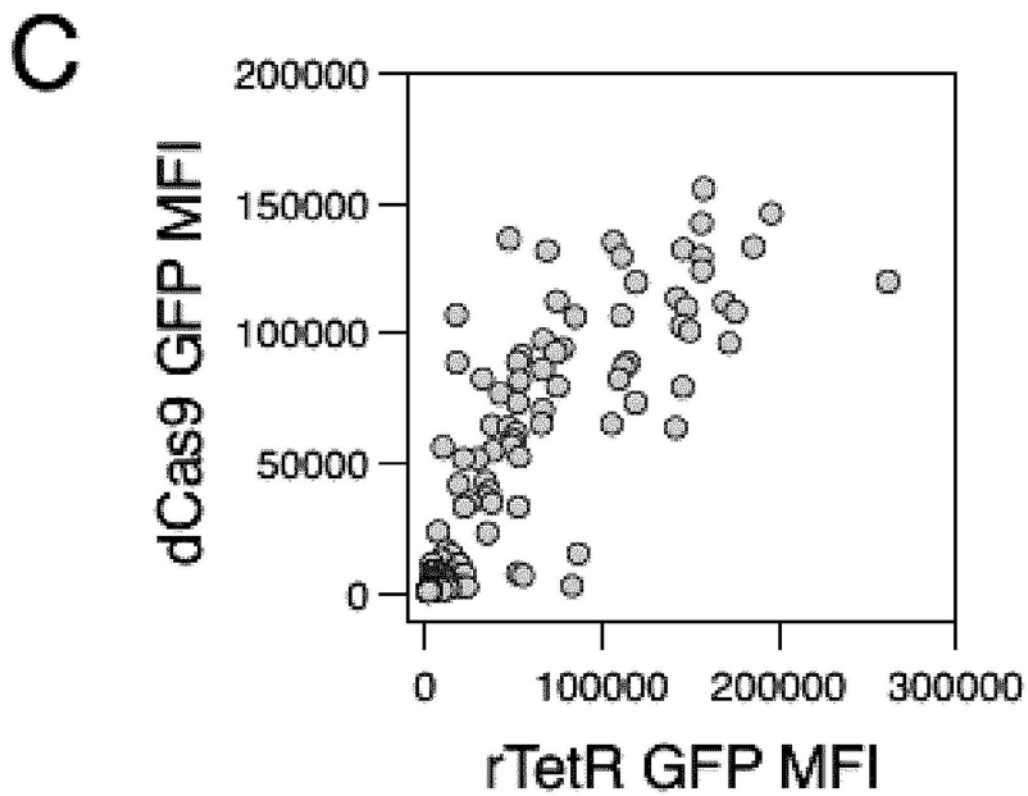


图19(续)