(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116234902 A
(43)申请公布日 2023.06.06

(21)申请号 202180047231.6

(22)申请日 2021.05.04

(30)优先权数据
63/019706 2020.05.04 US
63/074793 2020.09.04 US

(85)PCT国际申请进入国家阶段日
2022.12.30

(86)PCT国际申请的申请数据
PCT/US2021/030643 2021.05.04

(87)PCT国际申请的公布数据
WO2021/226077 EN 2021.11.11

(71)申请人 利兰·斯坦福青年大学托管委员会
地址 美国加利福尼亚州

(72)发明人 M·C·巴斯克 J·泰科
G·T·赫斯 L·宾图

(74)专利代理机构 中国专利代理(香港)有限公司 72001
专利代理师 权陆军 初明明

(51)Int.Cl.
C12N 9/64(2006.01)

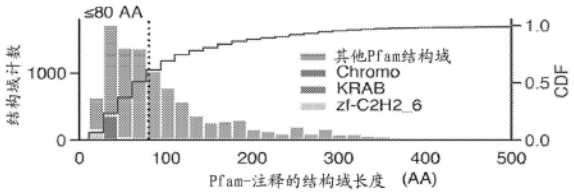
权利要求书4页 说明书130页 附图91页

(54)发明名称

用于激活和沉默基因表达的效应物结构域的生成、鉴定和表征的组合物、系统和方法

(57)摘要

本文提供了用于激活和沉默基因表达的效应物结构域的生成、鉴定和表征的组合物、系统和方法。特别地,提供了高通量系统以发现和表征效应物结构域。



CN 116234902 A

1.用于鉴定转录阻抑蛋白结构域的方法,包括:

a)制备包含多种核酸序列的结构域文库,每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的蛋白结构域的融合蛋白;

b)用结构域文库转化报道细胞,其中报道细胞包含在强启动子控制下的两部分报道基因,所述两部分报道基因包含表面标记和荧光蛋白,其中在用配置为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录阻抑蛋白结构域沉默;

c)用试剂处理所述报道细胞达细胞中的蛋白和mRNA降解所必需的时间长度;

d)基于表面标记、荧光蛋白或其组合的存在或不存在分离报道细胞;

e)对来自分离的报道细胞的蛋白结构域进行测序;

f)对于每种蛋白结构域序列计算来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率;和

g)将蛋白结构域鉴定为转录阻抑蛋白。

2.根据权利要求1所述的方法,进一步包括停止用所述试剂处理所述报道细胞并重复步骤d-g一次或多次。

3.根据权利要求2所述的方法,其中在停止用所述试剂处理所述报道的细胞后至少48小时重复步骤d-g。

4.根据权利要求1-3任一项所述的方法,其中每种蛋白结构域少于或等于80个氨基酸。

5.根据权利要求1-4任一项所述的方法,其中所述蛋白结构域来自核定位蛋白。

6.根据权利要求1-5任一项所述的方法,其中所述蛋白结构域包含来自核定位蛋白的野生型蛋白结构域的氨基酸序列。

7.根据权利要求1-5任一项所述的方法,其中所述蛋白结构域包含来自核定位蛋白的蛋白结构域的突变氨基酸序列。

8.根据权利要求1-7任一项所述的方法,其中所述诱导型DNA结合结构域包含标签。

9.根据权利要求1-8任一项所述的方法,进一步包括测量蛋白结构域的表达水平。

10.根据权利要求9所述的方法,其中所述表达水平通过测量标签在DNA结合结构域上的相对存在或不存在来确定。

11.根据权利要求1-10任一项所述的方法,其中所述报道细胞用试剂处理至少3天。

12.根据权利要求1-11任一项所述的方法,其中所述报道细胞用试剂处理5天。

13.根据权利要求1-12任一项所述的方法,其中所述当所述比率的log2离开表达不佳的阴性对照的平均值至少两个标准差时,所述蛋白结构域被鉴定为转录阻抑蛋白。

14.用于鉴定转录激活蛋白结构域的方法,包括:

a)制备包含多种核酸序列的结构域文库,每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的蛋白结构域的融合蛋白;

b)用结构域文库转化报道细胞,其中所述报道细胞包含在弱启动子控制下的两部分报道基因,所述两部分报道基因包含表面标记和荧光蛋白,其中在用配置为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录激活蛋白结构域激活;

c)用试剂处理所述报道细胞达细胞中的蛋白和mRNA产生所必需的时间长度;

d)基于表面标记、荧光蛋白或其组合的存在或不存在分离报道细胞;

e)对来自分离的报道细胞的蛋白结构域进行测序;

f)对于每种蛋白结构域序列计算来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率;和

g)将蛋白结构域鉴定为转录阻抑蛋白。

15.根据权利要求14所述的方法,进一步包括停止用所述试剂处理所述报道细胞并重复步骤d-g一次或多次。

16.根据权利要求15所述的方法,其中在停止用所述试剂处理所述报道的细胞后至少48小时重复步骤d-g。

17.根据权利要求14-16任一项所述的方法,其中每种蛋白结构域少于或等于80个氨基酸。

18.根据权利要求14-17任一项所述的方法,其中所述蛋白结构域来自核定位蛋白。

19.根据权利要求14-18任一项所述的方法,其中所述蛋白结构域包含来自核定位蛋白的野生型蛋白结构域的氨基酸序列。

20.根据权利要求14-19任一项所述的方法,其中所述蛋白结构域包含来自核定位蛋白的蛋白结构域的突变氨基酸序列。

21.根据权利要求14-20任一项所述的方法,其中所述诱导型DNA结合结构域包含标签。

22.根据权利要求14-21任一项所述的方法,进一步包括测量蛋白结构域的表达水平。

23.根据权利要求22所述的方法,其中所述表达水平通过测量标签在DNA结合结构域上的相对存在或不存在来确定。

24.根据权利要求14-23任一项所述的方法,其中所述报道细胞用试剂处理至少24小时。

25.根据权利要求14-24任一项所述的方法,其中所述报道细胞用试剂处理48小时。

26.根据权利要求14-25任一项所述的方法,其中所述当所述比率的log2离开弱表达阴性对照的平均值至少两个标准差时,所述蛋白结构域被鉴定为转录激活蛋白。

27.合成转录因子,其包含与异源DNA结合结构域融合的一个或多个转录激活蛋白结构域、一个或多个转录阻抑蛋白结构域或其组合,

其中所述一个或多个转录激活蛋白结构域中的至少一个或所述一个或多个转录阻抑蛋白结构域中的至少一个包含与SEQIDNOs:1-896中的任一个具有至少70%同一性的氨基酸序列。

28.根据权利要求27所述的合成转录因子,其包含与异源DNA结合结构域融合的两个或更多个转录激活蛋白结构域或两个或更多个转录阻抑蛋白结构域。

29.根据权利要求27-28任一项所述的合成转录因子,其中所述一个或多个转录激活蛋白结构域中的至少一个包含与SEQIDNOs:563-664中的任一个具有至少70%同一性的氨基酸序列。

30.根据权利要求27-29任一项所述的合成转录因子,其中所述一个或多个转录激活蛋白结构域中的至少一个选自表2中发现的那些。

31.根据权利要求27-30任一项所述的合成转录因子,其中所述一个或多个转录阻抑蛋白结构域中的至少一个包含与SEQIDNOs:1-562和665-896中的任一个具有至少70%同一性的氨基酸序列。

32.根据权利要求27-31任一项所述的合成转录因子,其中所述一个或多个转录阻抑蛋白结构域中的至少一个选自表1、3或4的任一个中发现的那些。

33.根据权利要求27-32任一项所述的合成转录因子,其中所述一个或多个转录激活蛋白结构域或所述一个或多个转录阻抑蛋白结构域通过根据权利要求1-26任一项所述的方法鉴定。

34.根据权利要求27-33任一项所述的合成转录因子,其中所述异源DNA结合结构域包含可编程的DNA结合结构域。

35.根据权利要求27-34任一项所述的合成转录因子,其中所述DNA结合结构域衍生自规律间隔成簇短回文重复序列相关(Cas)蛋白。

36.编码根据权利要求27-35任一项所述的合成转录因子的核酸。

37.根据权利要求36所述的核酸,其中所述核酸处于诱导型启动子的控制之下。

38.根据权利要求36所述的核酸,其中所述核酸处于组织特异性启动子的控制之下。

39.根据权利要求36-39任一项所述的核酸,其编码至少一种额外的转录因子。

40.根据权利要求39所述的核酸,其中所述至少一种额外的转录因子包含根据权利要求27-35任一项所述的合成转录因子。

41.包含根据权利要求36-40任一项所述的核酸的载体。

42.细胞,其包含根据权利要求27-35任一项所述的合成转录因子、根据权利要求36-40任一项所述的核酸或根据权利要求41所述的载体。

43.根据权利要求42所述的细胞,其中所述细胞包含两种或更多种合成转录因子、核酸或载体。

44.组合物或系统,其包含根据权利要求27-35任一项所述的合成转录因子、根据权利要求36-40任一项所述的核酸、根据权利要求41所述的载体或根据权利要求42或43所述的细胞。

45.根据权利要求44所述的组合物或系统,其中所述组合物包含两种或更多种合成转录因子、核酸、载体或细胞。

46.根据权利要求44或45所述的组合物或系统,其进一步包含指导RNA或编码指导RNA的核酸。

47.试剂盒,其包括至少一种根据权利要求27-36任一项所述的合成转录因子、根据权利要求36-40任一项所述的核酸、根据权利要求41所述的载体、根据权利要求42或43所述的细胞或根据权利要求44-46所述的组合物或系统。

48.调节细胞中至少一种靶基因的表达的方法,所述方法包括将至少一种根据权利要求27-35任一项所述的合成转录因子、根据权利要求36-40任一项所述的核酸、根据权利要求41所述的载体或根据权利要求44-46所述的组合物或系统引入所述细胞中。

49.根据权利要求48所述的方法,其中所述合成转录因子包含Cas蛋白DNA结合结构域且所述方法进一步包括使细胞与至少一种指导RNA接触。

50.根据权利要求48或49所述的方法,其中所述细胞在主体中。

51.根据权利要求50所述的方法,其中所述方法包括向所述主体施用所述至少一种合成转录因子、核酸、载体或组合物或系统。

52.根据权利要求48-51任一项所述的方法,其中调节至少两种基因的基因表达。

53.根据权利要求48-52任一项所述的方法,其中当所述至少一种靶基因的基因表达水平与所述至少一种靶基因的正常基因表达水平相比增加或降低时,所述至少一种靶基因的基因表达被调节。

54.根据权利要求27-35任一项所述的合成转录因子、根据权利要求36-40任一项所述的核酸、根据权利要求41所述的载体或根据权利要求44-46所述的组合物或系统用于调节细胞中至少一种靶基因的表达的用途。

55.根据权利要求54所述的用途,其中所述合成转录因子包含Cas蛋白DNA结合结构域。

56.根据权利要求54或55所述的用途,其中调节至少两种基因的基因表达。

57.根据权利要求54-56任一项所述的用途,其中当所述至少一种靶基因的基因表达水平与所述至少一种靶基因的正常基因表达水平相比增加或降低时,所述至少一种靶基因的基因表达被调节。

用于激活和沉默基因表达的效应物结构域的生成、鉴定和表征的组合物、系统和方法

[0001] 与相关申请的相互参照

本申请要求2020年5月4日提交的美国临时申请No.63/019,706和2020年9月4日提交的美国临时申请No.63/074,793的利益,其各自的内容均整体引入本文作为参考。

技术领域

本文提供了用于激活和沉默基因表达的效应物结构域的生成、鉴定和表征的组合物、系统和方法。特别地,提供了高通量系统以发现和表征效应物结构域。

[0003] 关于联邦资助的研究的声明

本发明是利用在美国国立卫生研究院(NationalInstitutesofHealth)授予的合同GM128947项下的政府支持作出的。政府对本发明享有某些权利。

背景技术

以前人工改造合成转录因子的努力已经从以前发现的效应物结构域的小工具箱获得了激活和阻抑蛋白结构域。需要新的方法来扩展所述工具箱。

发明内容

本文提供了用于激活和沉默基因表达的效应物结构域的生成、鉴定和表征的组合物、系统和方法。特别地,提供了高通量系统以发现和表征效应物结构域。在一些实施方案中,本文提供了高通量方法以发现和表征效应物结构域,其极大地扩展了工具箱。这些结构域满足了人工改造增强的合成转录因子用于基因和细胞治疗、合成生物学和功能基因组学中的应用的关键需要。

[0006] 在一些实施方案中,用于鉴定效应物结构域的方法包括:a)制备包含多种核酸序列的结构域文库,每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的蛋白结构域的融合蛋白;b)用结构域文库转化报道细胞,其中报道细胞包含在强启动子控制下的两部分报道基因,所述两部分报道基因包含表面标记和荧光蛋白,其中在用配置为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录阻抑蛋白结构域沉默;c)用试剂处理所述报道细胞达细胞中的蛋白和mRNA降解所必需的时间长度;d)基于表面标记、荧光蛋白或其组合的存在或不存在分离报道细胞;e)对来自分离的报道细胞的蛋白结构域进行测序;f)对于每种蛋白结构域序列计算来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率;和g)将蛋白结构域鉴定为转录阻抑蛋白。

[0007] 在一些实施方案中,用于鉴定效应物结构域的方法包括:a)制备包含多种核酸序列的结构域文库,每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的蛋白结构域的融合蛋白;b)用结构域文库转化报道细胞,其中所述报道细胞包含在弱启动子控制下的两部分报道基因,所述两部分报道基因包含表面标记和荧光蛋白,其中在用配置

为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录激活蛋白结构域激活;c)用试剂处理所述报道细胞达细胞中的蛋白和mRNA产生所必需的时间长度;d)基于表面标记、荧光蛋白或其组合的存在或不存在分离报道细胞;e)对来自分离的报道细胞的蛋白结构域进行测序;f)对于每种蛋白结构域序列计算来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率;和g)将蛋白结构域鉴定为转录激活蛋白。

[0008] 在一些实施方案中,方法进一步包括停止用试剂处理报道细胞并重复步骤d-g一次或多次。在一些实施方案中,在停止用试剂处理报道的细胞后至少48小时重复步骤d-g。

[0009] 在一些实施方案中,每种蛋白结构域少于或等于80个氨基酸。在一些实施方案中,蛋白结构域来自核定位蛋白。在一些实施方案中,蛋白结构域包含来自核定位蛋白的野生型蛋白结构域的氨基酸序列。在一些实施方案中,蛋白结构域包含来自核定位蛋白的蛋白结构域的突变氨基酸序列。

[0010] 在一些实施方案中,诱导型DNA结合结构域包含标签。

[0011] 在一些实施方案中,方法进一步包括测量蛋白结构域的表达水平。在一些实施方案中,表达水平通过测量标签在DNA结合结构域上的相对存在或不存在来确定。

[0012] 在一些实施方案中,报道细胞用试剂处理至少3天。在一些实施方案中,报道细胞用试剂处理至少5天。在一些实施方案中,报道细胞用试剂处理至少24小时。在一些实施方案中,报道细胞用试剂处理至少48小时。

[0013] 在一些实施方案中,当比率的log2离开(例如,高于)表达不佳的阴性对照的平均值至少两个标准差时,蛋白结构域被鉴定为转录阻抑蛋白。

[0014] 在一些实施方案中,当比率的log2离开(例如,低于)弱表达阴性对照的平均值至少两个标准差时,蛋白结构域被鉴定为转录激活蛋白。

[0015] 本文还提供了合成转录因子,其包含与异源DNA结合结构域融合的一个或多个转录激活蛋白结构域、一个或多个转录阻抑蛋白结构域或其组合。在一些实施方案中,一个或多个转录激活蛋白结构域中的至少一个或一个或多个转录阻抑蛋白结构域中的至少一个包含与SEQ ID NOs:1-896中的任一个具有至少70%同一性的氨基酸序列。

[0016] 在一些实施方案中,合成转录因子包含与异源DNA结合结构域融合的两个或更多个转录激活蛋白结构域或两个或更多个转录阻抑蛋白结构域。

[0017] 在一些实施方案中,一个或多个转录激活蛋白结构域中的至少一个包含与SEQ ID NOs:563-664中的任一个具有至少70%同一性的氨基酸序列。在一些实施方案中,一个或多个转录激活蛋白结构域中的至少一个选自表2中发现的那些。

[0018] 在一些实施方案中,一个或多个转录阻抑蛋白结构域中的至少一个包含与SEQ ID NOs:1-562和665-896中的任一个具有至少70%同一性的氨基酸序列。在一些实施方案中,一个或多个转录阻抑蛋白结构域中的至少一个选自表1、3或4的任一个中发现的那些。

[0019] 在一些实施方案中,一个或多个转录激活蛋白结构域或一个或多个转录阻抑蛋白结构域通过本文公开的方法鉴定。

[0020] 在一些实施方案中,异源DNA结合结构域包含可编程的DNA结合结构域。在一些实施方案中,DNA结合结构域衍生自规律间隔成簇短回文重复序列(Clustered Regularly Interspaced Short Palindromic Repeats)相关(Cas)蛋白。在一些实施方案中,DNA结合

结构域衍生自转录激活蛋白样效应物(TALEs)结构域。

[0021] 本文还提供编码合成转录因子或效应物结构域的核酸,如本文所公开的。在一些实施方案中,核酸处于诱导型启动子的控制之下。在一些实施方案中,核酸处于组织特异性启动子的控制之下。在一些实施方案中,核酸编码至少一种额外的转录因子或效应物结构域。

[0022] 本文进一步提供包含如本文所公开的合成转录因子、核酸、载体或细胞的组合物或系统。在一些实施方案中,组合物包含两种或更多种合成转录因子、核酸、载体或细胞。在一些实施方案中,组合物进一步包含指导RNA或编码指导RNA的核酸。

[0023] 此外,提供了调节细胞中至少一种靶基因的表达的方法。方法包括将如本文所述的至少一种合成转录因子、核酸、载体或组合物或系统引入细胞中。当至少一种靶基因的基因表达水平与所述至少一种靶基因的正常基因表达水平相比增加或降低时,所述至少一种靶基因的基因表达被调节。在一些实施方案中,合成转录因子包含Cas蛋白DNA结合结构域且方法进一步包括使细胞与至少一种指导RNA接触。

[0024] 在一些实施方案中,细胞在体外(例如,先体外后体内)或在主体中。

[0025] 在一些实施方案中,调节至少两种基因的基因表达。

附图说明

图1A-1G显示高通量募集测量了来自核定位蛋白的数千个Pfam-注释的结构域的转录阻抑蛋白活性。图1A-定位至细胞核的人蛋白中Pfam-注释的结构域的长度。选择≤80个氨基酸的结构域用于包括在文库中。图1B-进行筛选以鉴定转录阻抑蛋白的示意图。阻抑报道基因使用强的pEF启动子,其可以通过dox-介导的阻抑蛋白结构域的募集来沉默。细胞用多西环素处理5天,ON和OFF细胞被磁分离,并对结构域进行测序。去除Dox并在第9天和第13天取额外的时间点。图1C-显示了来自独立转导的生物重复实验的log2(OFF:ON)比率的可再现性,并且选择的结构域家族是着色的。图1D-通过家族内结构域第5天的最大阻抑蛋白强度排序的最前面阻抑蛋白结构域家族的箱线图(Boxplot)。图1E-通过流式细胞术测量的命中RYBP结构域的个别确认时间过程。图1F-阻抑蛋白结构域实验对象组的额外确认时间过程。结构域长度列在括号中,这是因为一些结构域被作为来自文库的精确80AA序列测试,而一些被作为调整到通过Pfam注释为结构域的区域的较短的序列测试。在第0天添加1000ng/mldox,并在第5天去除。图1G-筛选测量结果与KRAB效应物结构域集合的个别确认流式细胞术测量结果的相关性。

[0027] 图2A-2D显示阻抑性KRAB结构域位于较年轻的KRAB-锌指蛋白中,它们共定位并结合到KAP1辅阻抑蛋白。图2A-KRAB沉默功能与结构域天然存在于其中的KRAB锌指蛋白构造进行比较。图2B-KRAB沉默功能与KRAB锌指基因进化年龄进行比较,如通过使用其全部DNA-结合锌指阵列序列找到所述基因的最近代直系同源物所确定的(年龄公开于Trono 2017)。图2C-KRAB结构域被归类为沉默基因或非沉默基因且其在ChIP-seq数据集中的基因组定位与辅阻抑蛋白KRAB-相关蛋白1(KAP1)的定位进行比较。图2D-通过KRAB结构域的KRAB锌指基因是否在质谱分析法(mass spec)数据集中与辅阻抑蛋白KAP1显著相互作用分类的KRAB结构域的阻抑强度分布(Helleboid 2019)。点颜色是KRAB结构域表达水平的五分值(quintile)。

[0028] 图3A-3G显示ZNF10 KRAB结构域的深度突变扫描鉴定降低或增强阻抑蛋白活性的取代。图3A-深度突变扫描文库包括来自ZNF10的KRAB结构域中的所有单一和连续的双重和三重取代。通过改变密码子选择，将DNA寡核苷酸(oligos)设计成比蛋白序列更不同。红色残基不同于WT序列。图3B-所有单取代和三重取代变体相对于WT的阻抑蛋白测量结果显示在KRAB结构域的示意图下面。图3C-与来自所有人KRAB结构域的多序列比对的序列保守性(用ConSurf计算的)相比，第9天的对阻抑的平均突变影响。图3D-在不同细胞类型中高通量测量结果与以前公开的使用CAT测定的低通量数据的相关性。图3E-KRAB突变体的个别时间过程确认了A/B-框和N-末端中取代的作用。图3F-对于图3B中每个时间点的每个位置，将所有单取代的分布与野生型作用的分布进行比较(Wilcoxon秩和检验)。在第5天加log10(p)<-5记号的位置用红色着色(沉默中高度显著地降低)，在第9天而不是第5天加log10(p)<-5记号的位置用绿色着色，且在第13天具有log10(p)>5的位置W8用蓝色着色(高度显著地增加)。水平虚线显示命中阈值。序列保守性ConSurf分数以橙色显示。图3G-当突变时在第5天消除沉默的残基被作图到小鼠KRABA-框的NMR结构的有序区域上(PDB:1v65)。

[0029] 图4A和4B显示同源异型域阻抑强度与Hox基因组织线性对应(colinear)。图4A-同源异型框基因家族或类别通过在第5天的中值阻抑强度的排序。含有最强同源异型域阻抑蛋白的ANTP类别同源异型域的HOXL和NKL亚类以及PRD和LIM类别被分成个别的基因家族(Holland BMC2007)，而剩余的类别被群聚。点颜色是如在HT-表达测定中所测量的同源异型域表达水平的五分值。图4B-来自Hox基因家族的同源异型域在第5天的阻抑蛋白强度。箭标代表在四个人Hox基因座中发现的基因，并指向Hox基因转录的方向。灰色条分隔基因家族。计算了遍及所有Hox基因的基因数目和阻抑蛋白强度之间的关系的Spearman氏ρ和p-值。过滤数据以去除在任一种第5天测序样品中具有少于10计数的任何结构域。

[0030] 图5A-5F显示高通量募集发现了激活蛋白结构域，包括ZNF473中的有效、酸性和趋异的KRAB结构域变体。图5A-使用可通过dox-介导的激活结构域募集激活的弱minCMV启动子的激活报道基因的示意图，以及激活筛选的示意图。细胞集合体用多西环素处理48小时，ON和OFF细胞用ProG Dynabeads磁分离，并对结构域进行测序。图5B-显示了来自独立转导的生物重复实验的log2(OFF:ON)比率的可再现性，其中已知的激活蛋白结构域家族(FOXO-TAD,MybLMSTEN,TORC_C)是着色的。图5C-含有具有低于阈值的激活强度的结构域的基因的GO术语富集。图5D-激活蛋白结构域(红色)比非命中(灰色)更加酸性。图5E-通过平均激活强度排序的结构域家族列表。图5F-KRAB结构域通过序列比对和聚类，从而提供与Helleboid 2019中的分类相似的结果。最趋异的KRAB序列的聚簇被以绿色标记为变体KRABs。来自的筛选的结果在下面显示在热图(heatmaps)中。如果表达良好，则标准KRABs起阻抑蛋白的作用。变体KRABs显示出在筛选中作为阻抑蛋白、激活蛋白和无转录效果的混合效果。

[0031] 图6A-6F显示叠瓦式(tiling)文库发现了大染色质调节子蛋白中的新的自主阻抑蛋白结构域。图6A-文库的图形描绘，其中80AA叠瓦(tiles)覆盖蛋白序列，具有10AA滑动窗口(sliding window)。图6B-显示了来自独立转导的生物重复实验的log2(OFF:ON)比率的可再现性。图6C-将第5天的阻抑与MGA蛋白的已知结构域构造进行比较。在以前注释的区域之外发现了两个阻抑蛋白结构域。图6D-流式细胞术时间过程确认作为80AA叠瓦的各个MGA效应物。图6E-通过选择在筛选中显示阻抑蛋白活性的叠瓦之间共同共享的序列将效应物

最小化到10和30AA亚叠瓦(subtiles)。这些最小化的序列利用流式细胞术时间过程个别确认。图6F-来自叠瓦筛选的额外的80AA阻抑蛋白命中的个别确认。通过慢病毒(lentivirus)将rTetR-叠瓦融合物递送到K562报道细胞且细胞用100ng/mldox处理5天，然后去除dox。通过流式细胞术分析细胞，并通过根据其柠檬色(citrine)表达水平对细胞进行门控来测量细胞OFF的分数。

[0032] 图7A-7D显示募集测定测量通过具有荧光报道基因的慢病毒rTetR-结构域融合物的基因沉默。图7A-慢病毒载体的示意图。图7B-在K562报道细胞中进行的小规模试验(pilot test)，显示了对于克隆到pJT050上的ZNF10KRAB随着时间过去的柠檬色OFF:ON FACS直方图。第0天添加1000ng/ml dox，并在第5天去除。图7C-随着时间的过去的细胞ON的分数。图7D-报道系统也在HEK293T细胞中确立。用编码rTetR-KRAB或pOri对照的质粒转染细胞，并在用流式细胞术进行分析之前用或不用1000ng/mldox处理2(顶部)和4(底部)天。

[0033] 图8A-8E显示通过FLAG染色、分选和测序的结构域表达的高通量测量。图8A-测量文库中每个结构域融合物的表达水平的高通量方法的示意图。图8B-结构域表达测量结果的可再现性。图8C-用蛋白印迹确认。图8D-子文库的稳定性-随机是去稳定化的，叠瓦类似于Pfam结构域。图8E-与归类为无序促进(disorder promoting)的残基和多种残基的净电荷相关的稳定性。

[0034] 图9A-9E显示了用于阻抑蛋白功能的Pfam结构域的筛选。图9A-磁分离之前和之后细胞文库的流式细胞术。图9B-最高的10个logP的稳定对瞬时阻抑蛋白的PANTHER蛋白类别富集。图9C-通过第5天的阻抑蛋白强度排序的结构域家族的全部列表。图9D-rTetR-SUMO融合物使报道基因沉默。SUMO缀合位点(GG91AA)中的突变降低沉默速度，而SUMO-相互作用非共价结合位点中的突变降低沉默记忆。图9E-具有阻抑蛋白活性的未知功能结构域(DUFs)的确认。

[0035] 图10A-10C显示了KRAB深度突变扫描。图10A-来自在第5、9和13天来自ZNF10的KRAB结构域的深度突变文库的两个生物重复实验的OFF:ON分数。图10B-KRAB变体表达水平的FLAG-标签染色：非沉默的被降解。B-框突变体是稳定的。图10C-FLAG-标签染色与FLAG-标签蛋白印迹相关。

[0036] 图11A-11C显示了激活蛋白筛选数据。图11A-小规模试验，将rTetR-VP64电穿孔至K562 minCMV报道细胞。添加多西环素后，报道细胞变成ON，如通过对于柠檬色表达的流式细胞术所测量的。图11B-通过流式细胞术分析的在激活蛋白筛选期间合并的文库的磁分离。图11C-使用具有两种不同报道基因启动子的Pfam结构域文库的HT-募集转录调节测量的比较。每个结构域都是一个点，且点的大小是如在FLAG筛选中所测量的表达四分位数。

[0037] 图12A-12D显示了在数千个Pfam结构域的筛选中发现的数百个阻抑蛋白。图12A-通过家族内任何结构域在第5天的最大阻抑蛋白强度排序的最高阻抑蛋白结构域家族的箱线图。线显示中值，须延伸到高四分位数和低四分位数之外的四分位数的间距的1.5倍，且异常值用菱形显示。虚线显示命中阈值。正文中突出的结构域家族的箱是着色的。图12B-通过流式细胞术测量的RYBP结构域和两个具有阻抑蛋白活性的未知功能结构域(DUF)的个别确认。未处理的细胞分布以浅灰色显示，且多西环素处理的细胞以彩色显示，每种条件中都有两个独立转导的生物重复实验。垂直线显示了用于确定细胞OFF的分数的柠檬色门控。图12C-确认时间进程符合基因沉默模型：具有速率ks的指数沉默，继之以指数再激活。在第0

天添加多西环素(1000ng/ml)并在第5天去除(N＝2个生物重复实验)。具有柠檬色报道基因OFF的mCherry阳性细胞的分数通过流式细胞术确定,如在图12B中,并对于使用未处理的时间匹配的对照的背景沉默进行归一化。图12D-在第5天的高通量测量结果与沉默速率ks的相关性(R^2＝0.86,n＝15个结构域,N＝2-3个生物重复实验)。水平误差棒是拟合的速率的标准差,垂直误差棒是筛选生物重复实验的范围,且虚线是线性回归的95％置信区间。

[0038] 图13A-13E显示Hox同源异型域阻抑强度与Hox基因组织线性对应并与正电荷有关。图13A-在第5天通过其同源异型域的中值阻抑强度对同源异型框基因类别的排序。水平线显示命中阈值。来自CERS类别的5个同源异型域均未良好表达。图13B-来自Hox基因家族的同源异型域。(顶部)沿前后轴的Hox基因表达模式在修改的胚胎图像上根据Hox旁系同源物数目着色(Hueber等人,2010)。Hox11和12在后端和沿肢的近侧-远侧轴两者均表达(Wellik和Capecchi,2003)。(中部)5天dox后的阻抑强度。点根据Hox聚簇着色,且旁系同源物数目如在胚胎图像中那样着色。计算了遍及所有Hox基因的旁系同源物数目和阻抑蛋白强度之间的关系的Spearman氏ρ和p-值。(底部)着色的箭标代表在四个人Hox聚簇中发现的基因,并指向从5'到3'的Hox基因转录的方向。灰色条将基因序列相似性组分开,如以前所分类的(Hueber等人,2010)。图13C-Hox同源异型域的多序列比对,其中更强的阻抑蛋白在顶部(通过第5天的OFF:ON比率排序的),从而显示以红色突出的RKKR基序。N-末端臂内的其他碱性残基以淡紫色着色。图13D-每个Hox同源异型域的螺旋1上游的N-末端臂中带阳电的残基数目与第5天的平均阻抑之间的相关性。点颜色显示旁系同源物数目。图13E-从PDB ID:2L7Z检索的HOXA13同源异型域的NMR结构,其中RKKR基序以红色突出。显示了使用来自多序列比对的坐标的从G15到S81的序列。

[0039] 图14A-14G显示激活蛋白结构域的发现。图14A-激活报道基因的示意图,其使用可被与rTetR融合的激活效应物结构域的多西环素介导的募集激活的弱minCMV启动子。图14B-来自两个独立转导的生物重复实验的高通量激活蛋白测量结果的可再现性。含有(图14A)中的激活报道基因的细胞集合体用核内结构域文库转导并用多西环素处理48小时;ON和OFF细胞被磁分离,且结构域被测序。显示了良好表达的结构域的OFF对ON细胞的测序读长的比率。Pfam-注释的激活蛋白结构域家族(FOXO-TAD、MybLMSTEN、TORC_C)以红色阴影着色。向最强的命中,来自ZNF473的KRAB结构域划一条线。命中阈值是低于表达不佳的结构域分布的平均值两个标准差所划的虚线。图14C-具有至少一个激活蛋白命中的结构域家族的排序列表。以前在Pfam中被注释为激活蛋白的家族是红色的。虚线表示命中阈值,如在图14B中。仅显示良好表达的结构域。图14D-计算为每个氨基酸的净电荷的来自Pfam文库的效应物结构域的酸性。(左)非命中、良好表达的Pfam结构域(除KRAB和注释的激活蛋白之外)与激活蛋白命中的比较。Pfam-注释的激活蛋白结构域家族作为组作为阳性对照显示(橙色)。(右)来自KRAB结构域家族的激活蛋白命中和非命中的比较。来自Mann-Whitney检验的P-值显示为比较的组之间的条。n.s.＝不显著(p>0.05)。图14E-所有良好表达的KRAB结构域的系统树,其中序列趋异变体KRAB聚簇以绿色显示(顶部)。第5天阻抑的高通量募集测量结果以蓝色(中部)显示,且激活的测量结果以红色(底部)显示。水平虚线表示命中阈值。来自ZNF10的实例阻抑蛋白KRAB、来自ZFP28的阻抑蛋白KRAB_1以及所有激活蛋白KRAB结构域都由较大的标签引出。KRAB结构域起始位置写在括号中。图14F-变体KRAB激活蛋白结构域的个别确认。rTetR(SE-G72P)-结构域融合物通过慢病毒递送至K562报道细胞,并用杀稻瘟

素选择，细胞用1000ng/ml多西环素处理2天，且然后通过流式细胞术测量柠檬色报道基因水平。未处理的细胞分布以浅灰色显示，且多西环素处理的细胞以彩色显示，每种条件中都有两个独立转导的生物重复实验。垂直线显示用于确定细胞ON分数的柠檬色门控，并显示了多西环素处理的细胞的平均分数ON。图14G-KRAB锌指蛋白的ChIP峰位置离开活性染色质标记H3K27ac的最近峰的距离。KRAB蛋白通过其作为在第5天的阻抑蛋白筛选中的命中(蓝色)或非命中(绿色)的状态分类(左)。此外，个别显示了含有阻抑蛋白命中KRAB的ZNF10(黑色)、含有激活蛋白命中KRAB的ZNF473(红色)和含有激活蛋白命中和阻抑蛋白命中KRAB两者的ZFP28(黄色)的数据(右)。每个点显示40个碱基对箱中的峰的分数。ChIP-seq和ChiP-exo数据检索自(ENCODE Project Consortium等人,2020;Imbeault等人,2017;Najafabadi等人,2015;Schmitges等人,2016)。对于集合的数据仅包括其中单个KRAB锌指结合的单独峰(蓝点和绿点,左)，但对于个别蛋白包括所有峰，这是因为对于每个个别蛋白的单独峰数目是低的(红色、黑点和黄点,右)。

[0040] 图15A-15I显示了在核内蛋白中发现的紧凑阻抑蛋白结构域。图15A-覆盖组织的238种核定位蛋白的组的80AA叠瓦式文库的示意图。使用与图1中相同的工作流程，将这些叠瓦融合到rTetR并募集到报道基因，以测量阻抑强度。图15B-对于每个叠瓦用点显示的第5天通过最大阻抑蛋白功能排序的叠瓦的基因。命中是具有高于阴性对照的平均值≥2个标准差的log2(OFF:ON)的叠瓦。具有命中叠瓦的基因以渐变着色，且没有任何命中叠瓦的基因以灰色着色。图15C-叠瓦式CTCF。图表显示了从UniProt检索的蛋白注释。水平条显示每个叠瓦跨越的区域，且垂直误差条显示来自筛选的两个生物重复实验的标准误。最强的命中叠瓦以垂直渐变突出并注释为阻抑蛋白结构域(橙色)。图15D-叠瓦式BAZ2A(也称为TIP5)。图15E-个别确认。慢病毒rTetR(SE-G72P)-叠瓦融合物被递送至K562报道细胞，细胞用100ng/ml多西环素处理5天(垂直虚线之间)，且然后去除多西环素。通过流式细胞术分析细胞，确定柠檬色报道基因OFF的细胞的分数，并且数据符合基因沉默模型(N＝2个生物重复实验)。两个KRAB阻抑蛋白结构域显示为阳性对照。与底部显示的确认(蓝色曲线)相对应的叠瓦筛选数据显示在图22中。图15F-叠瓦式MGA。在以前注释的区域之外发现了两个阻抑蛋白结构域，并标记为阻抑蛋白1和2(深红色、紫色)。在命中叠瓦重叠处最小化的阻抑蛋白区域以窄的红色垂直渐变突出。图15G-来自MGA中两个峰的最大强度阻抑蛋白叠瓦用图15E中描述的方法个别确认(N＝2个生物重复实验)。图15H-MGA阻抑蛋白1序列通过选择峰中所有命中叠瓦之间共同共享的区域最小化，在垂直虚线之间显示并以红色加阴影。蛋白序列保守性ConSurf分数在下方以橙色线显示，且置信区间(推断的进化速率分布的第25个和第75个百分位)以灰色显示。星号标记通过ConSurf预测具有功能的(高度保守和暴露的)残基。使用相同的方法将阻抑蛋白2序列最小化，并且也与具有预测的功能残基的区域重叠(数据未显示)。图15I-MGA效应物被最小化为10和30AA亚叠瓦，如图15H中所示的，作为慢病毒rTetR(SE-G72P)-叠瓦融合物被克隆，并被递送至K562报道细胞。选择后，用100或1000ng/ml多西环素处理细胞5天，并通过流式细胞术测量具有沉默的柠檬色报道基因的细胞的百分比(N＝2个生物重复实验)。

[0041] 图16A-16C显示了慢病毒募集测定和基因沉默的双报道基因的确认。图16A-具有用于产生效应物结构域与dox-诱导型DNA结合结构域rTetR的融合物的Golden Gate克隆位点的慢病毒募集载体的示意图。组成型pEF启动子驱动被T2A自切割肽隔开的rTetR-效应物

融合物和mCherry-BSD(杀稻瘟素S脱氨酶抗性基因)的表达。图16B-(顶部)rTetR-KRAB融合物募集到双报道基因的示意图。报道基因通过TALEN-介导的同源性指导的修复整合到AAVS1基因座中,且PuroR抗性基因由内源性AAVS1启动子驱动。双报道基因由合成的表面标记(Igκ-hIgG1-Fc-PDGFRβ)和柠檬色荧光蛋白组成。(底部)在K562报道细胞中的小规模试验。报道细胞通过TALEN-介导的同源性指导的修复产生,以将报道基因整合到AAVS1基因座中,且然后用嘌呤霉素进行选择。然后用慢病毒旋转感染(spinfect)细胞以递送rTetR-KRAB,且然后使其不处理或用1000ng/ml多西环素处理以诱导rTetR在TetO位点与DNA结合。未处理的细胞分布以浅灰色显示,且多西环素处理的细胞以黑色或橙色显示,每种条件中都有两个独立转导的生物重复实验。慢病毒处理的细胞根据作为递送标记的mCherry进行门控。使用来自人ZNF10的KRAB结构域。图16C-使用结合合成的表面标记的ProG Dynabeads对OFF细胞与ON细胞的磁分离的证实。一千万个细胞经受使用30μl珠的磁分离,并通过流式细胞术之前和之后测量柠檬色报道基因表达。右侧显示了经受磁分离的混合ON和OFF细胞的图解。

[0042] 图17A-17F显示了通过FLAG染色、分选和测序的结构域表达的高通量测量。图17A-(顶部)用于测量文库中每个结构域的表达水平的高通量策略的示意图。低于80AA长的结构域使用其天然蛋白序列在两侧延伸以达到80AA,从而使得所有合成的文库元件长度相同。(中部)文库被克隆到加入FLAG标签的构建体中,并通过慢病毒以低感染复数递送至K562细胞,从而使得大多数细胞都表达单个文库成员。mCherry-BSD融合蛋白使得在没有使用第二2A组分的情况下的用于递送和选择效率的杀稻瘟素选择和荧光标记成为可能。(底部)表达是通过用抗-FLAG染色细胞、分选高表达和低表达群体、对结构域进行测序并计算log2(FLAG高:FLAG低)比率来测量的。图17B-分选到两个箱之前和之后通过流式细胞术测量的FLAG染色水平的分布(N=用重叠阴影区域显示的细胞文库的2个生物重复实验)。图17C-来自结构域表达筛选的生物重复实验的可再现性($r^2=0.82$)。选择高于阈值(高于随机对照的中值一个标准差的虚线)的表达良好的结构域用于在转录调节筛选中进行进一步分析。图17D-KRAB结构域实验对象组的表达水平的确认。通过慢病毒将个别rTetR-3XFLAG-KRAB构建体递送至K562细胞。用杀稻瘟素选择细胞,并通过流式细胞术确认>80%mCherry阳性。表达水平通过使用抗-FLAG抗体的蛋白印迹来测量。抗组蛋白H3用作用于归一化的加载对照。使用ImageJ定量水平。图17E-表达的高通量测量结果与蛋白印迹蛋白水平的比较。这6个KRAB结构域是使用来自Pfam结构域文库的精确80AA序列个别克隆的。图17F-不同类别的文库成员的表达水平分布。与遍及DMD蛋白或Pfam结构域的叠瓦相比,随机对照的表达不佳($p<1e-5$,Mann Whitney检验)。虚线显示了表达水平的阈值,如在图17C中。

[0043] 图18A-18K显示具有阻抑蛋白功能的结构域的鉴定。图18A-流式细胞术显示在使用与合成的表面标记结合的ProG DynaBeads进行磁分离之前和之后,表达Pfam结构域文库的细胞集合体中的柠檬色报道基因水平分布。显示了两个生物重复实验的重叠直方图。细胞OFF的平均百分比显示在显示柠檬色水平门控的垂直线的左侧。在第0天添加1000ng/ml多西环素并在第5天去除。图18B-当与具有文库中包括的结构域的所有核内蛋白的背景组相比时,含有具有更强或更弱记忆的阻抑蛋白结构域的核内蛋白的PANTHER蛋白类别富集。图18C-rTetR-SUMO确认时间过程符合基因沉默模型。以SUMO3的Rad60-SLD结构域和调整的结构域为中心的80AA序列单独克隆到慢病毒中并递送至报道细胞。在第0天添加1000ng/ml

多西环素，并在第5天去除(N＝2个生物重复实验)。具有柠檬色报道基因OFF的mCherry-阳性细胞的分数通过流式细胞术确定，并对于使用未处理的、时间匹配的对照的背景沉默进行归一化。图18D-利用在筛选中使用的完全80AA序列和进行调整以匹配Pfam和UniProt注释的序列的HUSH复合物成员MPP8 Chromo结构域确认。图18E-利用进行调整以匹配Pfam注释的52AA序列的CBX1 Chromoshadow结构域确认。图18F-利用进行调整以匹配Pfam注释的65AA序列的Polycomb 1组分SCMH1 SAM1结构域(也称为SPM)确认。图18G-利用在筛选中使用的完全80AA序列和进行调整以匹配Pfam注释的72AA序列的HERC2 Cyt-b5结构域确认。图18H-BIN1 SH3_9结构域确认。图18I-利用进行调整以匹配Pfam注释的39AA序列的Polycomb 1组分PCGF2 zf-C3HC4_2结构域确认。图18J-利用在筛选中使用的完全80AA序列和进行调整以匹配Pfam注释的68AA序列的TOX HMG框结构域确认。图18K-起阻抑蛋白作用的随机的80AA序列的确认。

[0044] 图19A-19D显示rTetR(SE-G72P)减轻人细胞中的渗漏KRAB沉默。图19A-通过rTetR-KRAB融合物的沉默，显示了对于KRAB结构域的亚型在没有多西环素处理的情况下的渗漏沉默(高灰条)。在第0天通过慢病毒将构建体递送至报道细胞，在第3天和第11天之间使用杀稻瘟素选择细胞，在第11天将细胞分成多西环素处理或未处理的条件，并在第16天通过流式细胞术测量报道基因水平。在对于mCherry阳性细胞进行门控后显示结果。KRAB结构域是基于它们在筛选中的测量结果从三个类别中选择的，在右侧进行标记。条显示平均值，且误差棒显示标准差(N＝3个独立转导的生物重复实验)。图19B-渗漏可以通过使用rTetR(SE-G72P)或在rTetR和来自ZNF823的KRAB结构域之间引入3XFLAG来减轻。在第0天通过慢病毒将构建体递送至报道细胞，在第4天将细胞分成多西环素处理或未处理的条件，并在第7天通过流式细胞术测量报道基因水平。在对于mCherry阳性细胞进行门控后显示结果。来自ZNF140的非泄漏KRAB结构域用作对照。条显示平均值，且误差棒显示标准差(N＝2个独立转导的生物重复实验)。图19C-具有作为与rTetR或rTetR(SE-G72P)的融合物克隆的来自ZNF823的渗漏KRAB结构域或来自ZNF140的非渗漏阻抑蛋白KRAB结构域的稳定慢病毒表达的K562报道细胞系用不同剂量的多西环素处理。四天后通过流式细胞术测量报道基因水平，且显示具有柠檬色报道基因OFF的mCherry阳性细胞的百分比(N＝2个独立转导的生物重复实验)。使用PRISM统计分析软件，通过最小二乘法将剂量反应拟合非线性变斜率S形曲线。图19D-KRAB结构域的所有个别确认的沉默和记忆动态符合基因沉默模型。rTetR(SE-G72P)-KRAB融合物被通过慢病毒递送至K562报道细胞，用杀稻瘟素选择，且然后在第0天添加10ng/ml多西环素，并在第5天去除(N＝2个生物重复实验)。具有柠檬色报道基因OFF的mCherry阳性细胞的分数通过流式细胞术确定，并对于使用未经处理的、时间匹配的对照的背景沉默进行归一化。10ng/ml dox用于在动态范围内起作用，在所述范围内更容易测量快速KRAB沉默结构域之间的沉默和记忆能力中的差异。使用1000ng/ml多西环素，所有阻抑蛋白命中KRAB结构域(绿色和橙色)都在5天内以无法区分的动态完全沉默报道基因(数据未显示)。值得注意的是，当与rTetR(SE-G72P)融合时，在rTetR上泄漏的KRABs(橙色)不显示与未泄漏的KRABs(绿色)显著不同的记忆动态。重要的是，在未处理的条件中，rTetR(SE-G72P)-KRAB融合物均未显示出显著的渗漏沉默。

[0045] 图20A-20H显示了CRISPRi中使用的ZNF10 KRAB的深度突变扫描。图20A-流式细胞术显示在使用与合成的表面标记结合的ProG DynaBeads进行磁分离之前和之后，具有合并

的KRAB文库的细胞中的柠檬色报道基因水平。显示了两个生物重复实验的重叠直方图。细胞OFF的平均百分比显示在显示柠檬色水平门控的垂直线的左侧。图20B-来自第5、9和13天的ZNF10 KRAB结构域的深度突变文库的两个生物重复实验的OFF:ON富集。前5天用1000ng/ml多西环素处理细胞。灰色对角线显示平均log2(OFF:ON)是WT结构域(黑点)的中值之处。黑色对角线显示拟合线性模型。图20C-人ZNF10 KRAB与NMR结构中使用的小鼠KRAB(PDB:1v65)和重组蛋白结合测定中使用的KRAB-O(Peng等人,2009)的比对。图3中使用有序区域,且含有所有12个必需残基的比对的区域用于(图20D)。第5天沉默所必需的残基在ZNF10和PDB:1v65序列中以红色着色。在KRAB-O序列中,结合重组KAP1所必需的残基以红色着色,而结合KAP1不必需的残基以灰色着色,从而总结了以前公开的结果(Peng等人,2009)。图20D-KRAB NMR结构的20个状态的集合(PDB:1v65)。第5天沉默所必需的残基以红色着色。图20E-KRAB ZNF10突变体的所有个别确认的沉默和记忆动态,符合基因沉默模型。(顶部)rTetR-KRAB融合物通过慢病毒递送至K562报道细胞,用杀稻瘟素选择,且然后在第0天添加1000ng/ml多西环素,并在第5天去除(N＝2个生物重复实验。(底部)rTetR(SE-G72P)-KRAB融合物通过慢病毒递送至K562报道细胞,用杀稻瘟素选择,且然后在第0天添加10ng/ml多西环素并在第5天去除(N＝2个生物重复实验)。列标记描述了KRAB结构域内的变异位置以及对效应物功能的影响。具有柠檬色报道基因OFF的mCherry阳性细胞的分数通过流式细胞术确定,并对于使用未经处理的、时间匹配的对照的背景沉默进行归一化。所有rTetR(SE-G72P)-KRAB融合物也在用1000ng/ml多西环素处理的5天内进行测量,且结果与具有rTetR的那些不能区别,其中除不沉默的EEW25AAA变体以外所有KRAB变体完全沉默报道基因(数据未显示)。图20F-rTetR-KRAB融合物表达水平与来自Pfam结构域文库的第13天沉默分数的相关性。仅包括通过IP/MS显示与辅阻抑蛋白KAP1相互作用的KRAB结构域(Helleboid等人,2019)。图20G-遍及Pfam结构域和对照的文库的氨基酸频率与结构域表达水平的相关性(显示Pearson氏的r值)。图20H-慢病毒递送至K562后加入FLAG标签的rTetR-KRAB融合物的蛋白印迹。用杀稻瘟素选择用于递送的细胞,并通过流式细胞术确认>80％mCherry阳性。使用ImageJ定量相对于H3加载对照的表达水平。

[0046] 图21A-21C显示HT-募集至最小启动子发现了激活蛋白结构域。图21A-磁分离之前和之后激活报道细胞中Pfam结构域的合并的文库的流式细胞术。细胞ON的百分比显示在用垂直线划的柠檬色水平门控的右侧。显示了具有重叠的阴影区域的1-2个生物重复实验。图21B-与计数过滤后含有文库中良好表达的结构域的所有蛋白的背景组相比,含有命中激活结构域的基因的GO术语富集。显示了原始p-值,并且所有显示的术语都低于10％错误发现率。图21C-激活蛋白结构域的个别确认。rTetR(SE-G72P)-结构域融合物通过慢病毒递送至K562报道细胞,并用杀稻瘟素进行选择。细胞用1000ng/ml多西环素处理2天,且然后通过流式细胞术测量柠檬色报道基因水平。未处理的细胞分布以浅灰色显示,且多西环素处理的细胞以彩色显示,每种条件中都有两个独立转导的生物重复实验。垂直线显示了用于确定细胞ON的分数的柠檬色门控,且显示了多西环素处理的细胞的平均分数ON。VP64是阳性对照。除具有最小延伸的Med9和DUF3446之外,这是因为Pfam注释的区域分别为75和69AA,每个结构域都作为来自文库的延伸的80AA序列或调整的Pfam-注释的结构域序列进行测试。KRAB结构域的80AA文库序列的相应结果显示于图14中。

[0047] 图22A-22H显示用叠瓦筛选鉴定核内蛋白中的紧凑阻抑蛋白结构域。图22A-流式

细胞术显示在使用与合成的表面标记结合的ProG DynaBeads进行磁分离之前和之后，表达叠瓦式文库的细胞集合体中的柠檬色报道基因水平分布。显示了两个生物重复实验的重叠直方图。细胞OFF的平均百分比显示在显示柠檬色水平门控的垂直线的左侧。第0天添加1000ng/ml多西环素并在第5天去除。图22B-在多西环素处理第5天和第13天(多西环素去除后8天)来自核内蛋白叠瓦式文库的两个生物重复实验的高通量募集测量结果。命中调入阈值高于随机和DMD叠瓦对照的平均值2个标准差。图22C-KRAB锌指蛋白ZNF57和ZNF461的叠瓦结果。每个条都是80AA叠瓦，且垂直误差棒是来自2个生物重复实验的范围。蛋白注释来源于UniProt。图22D-叠瓦式RYBP。图表显示了使用写在顶部的UniProt ID检索的蛋白注释。垂直误差棒显示来自两个生物重复实验的标准误。图22E-叠瓦式REST。图22F-叠瓦式CBX7。图22G-叠瓦式DNMT3B。图22H(顶部)叠瓦式DMD。(底部)DMD命中叠瓦募集后的沉默和记忆动态。前5天用1000ng/ml多西环素处理细胞，并通过流式细胞术测量柠檬色报道基因水平。细胞OFF的百分比被归一化以说明背景沉默，并且数据(点)符合基因沉默模型(曲线)(N＝2个生物重复实验)。

具体实施方式

提供了生成紧凑转录效应物结构域目录的系统和方法。进一步地，在一些实施方案中，所述结构域目录被融合到DNA结合结构域上以人工改造合成转录因子。这些发现用于执行真核(或其他)细胞中基因表达的靶向的和可调的调节。所述技术利用高通量平台以筛选和表征细胞中数以万计的合成转录因子。这些合成转录因子是DNA结合结构域和转录效应物结构域之间的融合物。所述系统已被用于生成数百个短效应物结构域(例如，80个氨基酸)和用于将它们进一步缩短到最小限度足够的序列(例如，10个氨基酸)的高通量方法，这对递送有利(例如，包装在病毒载体中)。这些融合物的靶向产生mRNA转录的局部调节，取决于效应物结构域可以是正调节或负调节。这些合成转录因子中的一些介导了在因子本身已从靶中释放后继续存在的长期的外遗传调节。

[0049] 以前，有限数目的转录效应物结构域可用于合成转录因子的人工改造。为了处理所述限制，本文提供了高通量方法以筛选和定量转录效应物结构域的功能。所述方法使得能够发现数百个效应物结构域，所述效应物结构域当融合到DNA结合结构域上时可以以靶向的方式上调或下调转录。所述方法还用于鉴定具有增强的活性的效应物结构域突变体。这些效应物结构域用于人工改造用于应用于基因和细胞治疗、合成生物学和功能基因组学中的合成转录因子。

[0050] 示例性应用包括，但不限于：

利用可编程DNA结合结构域(例如，dCas9、dCas12a、锌指、TALE)与转录效应物结构域的融合物的内源基因的靶向阻抑/激活。

[0051] 基因和细胞治疗(例如，以沉默患者中的致病转录物)或在研究中。

[0052] 合成转录因子用于同时干扰多个基因的表达(例如，以使用多种指导RNAs利用CRISPRi/a筛选进行高通量遗传相互作用作图)。

[0053] 用于遗传回路(genetic circuits)例如诱导型基因表达或更复杂的回路中的合成转录因子中。这些回路用于基因治疗(例如，抗体的AAV递送)和细胞治疗(例如，CAR-T胞的先体外后体内人工改造)，以实现响应环境和小分子输入的治疗性基因表达输出。

[0054] 本文提供的新转录效应物结构域对于依赖于合成转录因子的应用具有几个优点。鉴定了短结构域(例如,80个氨基酸或更少),并产生了用于将它们进一步缩短到最小限度足够的序列的高通量方法,这对递送有利(例如,包装在病毒载体中)。在一些情况下,鉴定了短至10个氨基酸的有效的效应物结构域。在一些实施方案中,结构域是从人蛋白中提取的,与病毒效应物结构域相比,其提供了降低免疫原性的优点。大多数生成的结构域以前尚未被报道为转录效应物。此外,提供了用于测试这些结构域中的突变以鉴定增强的变体的高通量方法。人工细胞表面标记的开发更容易帮助高通量方法,所述人工细胞表面标记提供了使用磁分离对这些文库的更有效、廉价和快速的筛选。这是比起基于荧光报道基因表达的更传统的分选文库的方法来的优点。

[0055] 鉴定的结构域的集合庞大且多样,且平台容易使结构域的新组合能够高通量地作为融合物进行测试,以产生具有新性质的合成转录因子(例如,两个阻抑蛋白结构域的组合以实现快速沉默和永久沉默的组合)。

[0056] 数百个以前未表征或未知的效应物结构域,其可以沉默或激活转录,并可以融合到DNA结合结构域上。例如,提供了在人细胞中使用慢病毒筛选来筛选单个结构域和结构域对的高通量方法。人工细胞表面标记的开发更容易使得能够进行高通量方法,所述人工细胞表面标记提供了使用磁分离对这些文库的更有效、廉价和快速的筛选。

[0057] 1.定义

如本文所用的,术语“包含”、“包括”、“具有(having)”、“具有(has)”、“可以”、“含有”及其变化形式意图是不排除额外的行为或结构的可能性的开放式的过渡短语、术语或词语。单数形式“一个”、“一种”和“所述(the)”包括复数参考,除非上下文另有明确指示。本公开内容还预期“包含”本文提供的实施方案或要素、“由”其“组成”和“基本上由”其“组成”的其他实施方案,无论是否明确陈述。

[0058] 对于本文数值范围的叙述,明确预期其间具有相同精确度的每个中间数字。例如,对于6-9的范围,除了6和9以外,预期数字7和8,并且对于6.0-7.0的范围,明确预期数字6.0、6.1、6.2、6.3、6.4、6.5、6.6、6.7、6.8、6.9和7.0。

[0059] 除非本文另有定义,否则与本公开内容有关的科学和技术术语应具有本领域普通技术人员通常理解的含义。例如,与本文所述的细胞和组织培养、分子生物学、免疫学、遗传学以及蛋白和核酸化学和杂交有关使用的任何术语和技术是本领域众所周知和常用的那些。术语的含义和范围应是清楚的;然而,万一存在任何潜在的歧义,则本文提供的定义优于任何字典或外部定义。进一步地,除非上下文另有要求,否则单数术语应包括复数,且复数术语应包括单数。

[0060] 如本文所用的,术语“抗体”指由免疫系统内源性地使用以鉴定和中和异物例如细菌和病毒的蛋白。一般,抗体是包含至少一个互补性决定区(CDR)的蛋白。CDRs形成抗体的“高变区”,其负责抗原结合(下文进一步讨论)。完整抗体一般由四条多肽组成:两个相同的重(H)链多肽拷贝和两个相同的轻(L)链多肽拷贝。每条重链含有一个N-末端可变(V_H)区和三个C-末端恒定(C_{H1}、C_{H2}和C_{H3})区,且每条轻链含有一个N-末端可变(V_L)区和一个C-末端恒定(C_L)区。抗体的轻链可基于其恒定结构域的氨基酸序列分配给两个不同类型之一,kappa(κ)或lambda(λ)。在典型的抗体中,每条轻链通过二硫键与重链连接,且两条重链通过二硫键相互连接。轻链可变区与重链可变区对齐,且轻链恒定区与重链第一恒定区对齐。重链的

剩余恒定区相互对齐。每对轻链和重链的可变区形成抗体的抗原结合部位。V_H和V_L区具有相同的一般结构，每个区域包含四个构架(FW或FR)区。如本文所用的，术语“构架区”指可变区内位于CDRs之间的相对保守的氨基酸序列。每个可变结构域中有四个构架区，其分别称为FR1、FR2、FR3和FR4。构架区形成提供可变区结构构架的β折叠(参见，例如，C.A.Janeway等人(编)，Immunobiology，第5版，Garland Publishing，New York，N.Y.(2001))。构架区由三个CDRs连接。如上文所讨论的，称为CDR1、CDR2和CDR3的三个CDRs形成负责抗原结合的抗体的“高变区”。CDRs形成连接由构架区形成的β折叠结构且在一些情况下包含由构架区形成的β折叠结构的一部分的环。虽然轻链和重链的恒定区不直接与抗体与抗原的结合有关，但恒定区可以影响可变区的定向。恒定区还表现出各种效应子功能，例如经由与效应物分子和细胞的相互作用参与抗体依赖性补体介导的裂解或抗体依赖性细胞毒性。

[0061] 术语抗体的“抗体的片段”、“抗体片段”和“抗原结合片段”在本文中可互换使用以指抗体的一个或多个片段，其保留特异性结合抗原的能力(参见，一般来说，Holliger等人，Nat.Biotech.，23(9):1126-1129(2005))。本文所述抗体的任何抗原结合片段都在本发明的范围内。抗体片段期望地包含例如一个或多个CDRs、可变区(或其部分)、恒定区(或其部分)或其组合。抗体片段的例子包括，但不限于，(i)Fab片段，它是由V_L、V_H、C_L和C_{H1}结构域组成的单价片段，(ii)$F(ab')_2$片段，它是包含通过在铰链区的二硫键连接的两个Fab片段的二价片段，(iii)由抗体单臂的V_L和V_H结构域组成的Fv片段，(iv)Fab’片段，其由使用温和还原条件断裂$F(ab')_2$片段的二硫键产生，(v)二硫键稳定的Fv片段(dsFv)，和(vi)结构域抗体(dAb)，它是特异性结合抗原的抗体单可变区结构域(V_H或V_L)多肽。

[0062] 如本文所用的，“核酸”或“核酸序列”指嘧啶和/或嘌呤碱基优选分别为胞嘧啶、胸腺嘧啶和尿嘧啶以及腺嘌呤和鸟嘌呤的多聚体或寡聚体(参见Albert L.Lehninger，Principles of Biochemistry，第793-800页(Worth Pub.1982))。本技术预期任何脱氧核糖核苷酸、核糖核苷酸或肽核酸组分，及其任何化学变体，例如这些碱基的甲基化、羟甲基化或糖基化形式等。多聚体或寡聚体在组成上可以是异质的或同质的并且可以从天然存在的来源分离或者可以是人工或合成产生的。此外，核酸可以是DNA或RNA，或它们的混合物，并且可以以单链或双链形式永久或短暂存在，包括同源双链体、异源双链体和杂交状态。在一些实施方案中，核酸或核酸序列包含其他种类的核酸结构，如例如，DNA/RNA螺旋、肽核酸(PNA)、吗啉代核酸(参见，例如，Braasch和Corey，Biochemistry，41(14):4503-4510(2002))和美国专利No.5，034，506)、锁定核酸(LNA；参见Wahlestedt等人，Proc.Natl.Acad.Sci.U.S.A.，97:5633-5638(2000))、环己烯基核酸(参见，Wang，J.Am.Chem.Soc.，122:8595-8602(2000))和/或核酶。因此，术语“核酸”或“核酸序列”还可以包括链，所述链包含非天然核苷酸、修饰的核苷酸和/或可表现出与天然核苷酸相同功能的非核苷酸结构单元(例如，“核苷酸类似物”)；进一步地，如本文所使用的术语“核酸序列”指寡核苷酸、核苷酸或多核苷酸及其片段或部分，以及基因组或合成起源的DNA或RNA，其可以是单链或双链的，并且代表有义链或反义链。术语“核酸”、“多核苷酸”、“核苷酸序列”和“寡核苷酸”可互换使用。它们指任何长度的核苷酸的聚合形式，或是脱氧核糖核苷酸或是核糖核苷酸，或其类似物。

[0063] “肽”或“多肽”是由肽键连接的两个或更多个氨基酸的连接的序列。肽或多肽可以是天然的、合成的或天然和合成的修饰或组合。多肽包括蛋白，例如结合蛋白、受体和抗体。

可以通过添加糖、脂质或不包括在氨基酸链中的其他部分来修饰蛋白。术语“多肽”和“蛋白”在本文中可互换使用。

[0064] 如本文所用的，术语“百分比序列同一性”指在比对两个序列并在必要时引入缺口以实现最大百分比同一性后，与参考序列中相应核苷酸或氨基酸相同的核酸序列中核苷酸或核苷酸类似物或氨基酸序列中氨基酸的百分比。因此，假如根据所述技术的核酸比参考序列长，则对于确定序列同一性不考虑核酸中不与参考序列比对的额外的核苷酸。许多用于获得最佳比对和计算两个或更多个序列之间的同一性的数学算法是已知的，并且被并入许多可用的软件程序中。这种程序的示例包括CLUSTAL-W，T-Coffee和ALIGN(用于核酸和氨基酸序列的比对)，BLAST程序(例如，BLAST 2.1、BL2SEQ其新近版本)和FASTA程序(例如，FASTA3x、FAS™和SSEARCH)(用于序列比对和序列相似性搜索)。序列比对算法也公开于，例如，Altschul等人，J.Molecular Biol.，215(3):403-410(1990)，Beigert等人，Proc.Natl.Acad.Sci.USA，106(10):3770-3775(2009)，Durbin等人，编，Biological Sequence Analysis:Probabilistic Models of Proteins and Nucleic Acids，Cambridge University Press，Cambridge，UK(2009)，Soding，Bioinformatics，21(7):951-960(2005)，Altschul等人，Nucleic Acids Res.，25(17):3389-3402(1997)，和Gusfield，Algorithms on Strings，Trees and Sequences，Cambridge University Press，Cambridge UK(1997))。

[0065] “载体”或“表达载体”是复制子，例如质粒、噬菌体、病毒或黏粒，另一个DNA片段，例如“插入片段”，可以与其附着或合并，以便引起附着的片段在细胞中的复制。

[0066] 术语“野生型”指具有当从天然存在的来源分离时所述基因或基因产物的特征的基因或基因产物。野生型基因是在群体中最常常观察到的基因，且因此被人为地称为基因的“正常”或“野生型”形式。比较起来，术语“修饰的”、“突变的”或“多态的”指当与野生型基因或基因产物相比时展示序列和或功能性质中的修饰(例如，改变的特征)的基因或基因产物。要指出的是，可以分离天然存在的突变体；这些是通过当与野生型基因或基因产物相比时它们具有改变的特征的事实来鉴定的。

[0067] 2.鉴定转录修饰结构域的方法

本文公开了用于鉴定转录效应物(例如，激活蛋白和阻抑蛋白)结构域的方法。在一些实施方案中，所述方法包括：制备包含多种核酸序列的结构域文库，每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的来自核定位蛋白的蛋白结构域的融合蛋白；用结构域文库转化报道细胞，其中所述报道细胞包含在启动子控制下的两部分报道基因，所述两部分报道基因包含表面标记和荧光蛋白，其中在用配置为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录效应物结构域调节；用试剂处理所述报道细胞达细胞中的蛋白和mRNA水平被改变(例如，由于产生而增加的或由于降解而降低的)所必需的时间长度；对来自分离的报道细胞的蛋白结构域进行测序；对于每种蛋白结构域序列计算来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率；和将蛋白结构域鉴定为转录阻抑蛋白或激活蛋白。

[0068] 所述方法包括制备包含多种核酸序列的结构域文库，每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的来自核定位蛋白的蛋白结构域的融合蛋白。蛋白

结构域可以少于或等于80个氨基酸。在一些实施方案中，蛋白结构域可以是约75个氨基酸、约70个氨基酸、约65个氨基酸、约60个氨基酸、约55个氨基酸、约50个氨基酸、约45个氨基酸、约40个氨基酸、约35个氨基酸、约30个氨基酸、约25个氨基酸、约20个氨基酸、约15个氨基酸、约10个氨基酸或约5个氨基酸。

[0069] 蛋白结构域可以衍生自任何已知的蛋白。在一些实施方案中，蛋白结构域来自核定位蛋白。核定位蛋白包括在蛋白的生活周期期间完全或部分定位于或可以定位于细胞核的那些蛋白。在一些实施方案中，蛋白结构域包含来自核定位蛋白的野生型蛋白结构域的氨基酸序列。在一些实施方案中，蛋白结构域包含来自核定位蛋白的蛋白结构域的突变氨基酸序列。

[0070] 诱导型DNA结合结构域可以使用用于诱导DNA结合的任何系统，包括，但不限于四环素Tet，/DOX诱导型系统、光诱导型系统、脱落酸(ABA)诱导型系统、cumate系统、4OHT/雌激素诱导型系统、基于蜕皮素的诱导型系统和FKBP12/FRAP(FKBP12-雷帕霉素复合物)诱导型系统。

[0071] 在一些实施方案中，诱导型DNA结合结构域包含标签。标签可以包括本领域已知的任何标签，包括通过化学或酶促方法可去除的标签。供本方法之用的合适标签包括几丁质结合蛋白(CBP)、麦芽糖结合蛋白(MBP)、Strep-标签、谷胱甘肽-S-转移酶(GST)、聚组氨酸(PolyHis)标签、ALFA-标签、V5-标签、Myc-标签、血凝素(HA)-标签、Spot-标签、T7-标签、NE-标签、钙调蛋白标签、聚谷氨酸标签、聚精氨酸标签、FLAG标签等。

[0072] 方法包括用结构域文库转化报道细胞，其中所述报道细胞包含在启动子控制下的两部分报道基因，所述两部分报道基因包含表面标记和荧光蛋白，其中在用配置为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录效应物结构域调节。

[0073] 启动子可以赋予高转录率(强启动子)或赋予低转录率(弱启动子)。许多启动子文库已用实验方法确立，且启动子和启动子强度的选择取决于细胞类型。在一些实施方案中，当鉴定转录激活蛋白结构域时，可以使用弱启动子。在一些实施方案中，当鉴定转录阻抑蛋白结构域时，可以使用强启动子。

[0074] 细胞表面标记包括附着至细胞膜的蛋白和碳水化合物。对于多种细胞类型，细胞表面标记是本领域中通常已知的，并且可以基于已知的分子生物学方法在选择的报道细胞中表达。表面标记可以是合成的表面标记，其包含附着至跨膜结构域的标记多肽。例如，标记多肽可以包括附着至跨膜结构域的抗体或其片段(例如，Fc区)。在一些实施方案中，标记多肽是人IgG1Fc区并且合成表面标记包含附着至跨膜结构域的人IgG1Fc区。

[0075] 荧光蛋白在本领域中是众所周知的并且包括适合于在各种细胞区室中并且作为入射光的不同波长的结果而发荧光的蛋白。荧光蛋白的例子包括藻胆蛋白、青色荧光蛋白(cyan fluorescent protein)(CFP)、绿色荧光蛋白(GFP)、黄色荧光蛋白(YFP)、增强型橙色荧光蛋白(OFP)、增强型绿色荧光蛋白(eGFP)、修饰的绿色荧光蛋白(emGFP)、增强型黄色荧光蛋白(eYFP)和/或单体红色荧光蛋白(mRFP)及其衍生物和变体。

[0076] 方法包括基于表面标记、荧光蛋白或其组合的存在或不存在分离报道细胞。许多细胞分离技术是本领域已知的适合于供本文公开的方法使用，包括，例如，免疫磁细胞分离、荧光激活细胞分选(FACS)和微观流体细胞分选。在一些实施方案中，细胞分离包括免疫

磁细胞分离。

[0077] 在一些实施方案中，方法进一步包括停止用试剂处理报道细胞并重复分离、测序、计算和鉴定步骤一次或多次。在一些实施方案中，在停止用试剂处理报道细胞后至少48小时重复所述步骤。

[0078] 在一些实施方案中，方法进一步包括测量蛋白结构域的表达水平。可以使用本领域中已知的任何方法确定蛋白结构域的表达水平，包括对于蛋白本身或其任何标签或标记的免疫印迹和免疫测定。在一些实施方案中，表达水平通过测量标签在DNA结合结构域上的相对存在或不存在来确定。

[0079] 在一些实施方案中，方法鉴定转录阻抑蛋白结构域。在一些实施方案中，方法包括，a)制备包含多种核酸序列的结构域文库，每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的蛋白结构域的融合蛋白；b)用结构域文库转化报道细胞，其中报道细胞包含在强启动子控制下的两部分报道基因，所述两部分报道基因包含表面标记和荧光蛋白，其中在用配置为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录阻抑蛋白结构域沉默；c)用试剂处理所述报道细胞达细胞中的蛋白和mRNA降解所必需的时间长度；d)基于表面标记、荧光蛋白或其组合的存在或不存在分离报道细胞；e)对来自分离的报道细胞的蛋白结构域进行测序；f)对于每种蛋白结构域序列计算来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率；和g)将蛋白结构域鉴定为转录阻抑蛋白。

[0080] 在一些实施方案中，报道细胞用试剂处理至少3天。例如，报道细胞可以用试剂处理至少3天、至少4天、至少5天、至少6天、至少7天、至少8天、至少9天、至少10天、至少14天或更多。在一些实施方案中，报道细胞用试剂处理3-12天、3-10天、3-7天或3-5天。

[0081] 当来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率的log2离开(例如，大于)阴性对照的平均值至少两个标准差时，蛋白结构域被鉴定为转录阻抑蛋白(例如，参见图1C)。

[0082] 在一些实施方案中，方法鉴定转录激活蛋白结构域。在一些实施方案中，方法包括，a)制备包含多种核酸序列的结构域文库，每种所述核酸序列被配置为表达包含连接至诱导型DNA结合结构域的蛋白结构域的融合蛋白；b)用结构域文库转化报道细胞，其中所述报道细胞包含在弱启动子控制下的两部分报道基因，所述两部分报道基因包含表面标记和荧光蛋白，其中在用配置为诱导所述诱导型DNA结合结构域的试剂处理后所述两部分报道基因能够被推定的转录激活蛋白结构域激活；c)用试剂处理所述报道细胞达细胞中的蛋白和mRNA产生所必需的时间长度；d)基于表面标记、荧光蛋白或其组合的存在或不存在分离报道细胞；e)对来自分离的报道细胞的蛋白结构域进行测序；f)对于每种蛋白结构域序列计算来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率；和g)将蛋白结构域鉴定为转录阻抑蛋白。

[0083] 在一些实施方案中，报道细胞用试剂处理至少24小时。例如，报道细胞可用试剂处理至少24小时(1天)、至少36小时、至少48小时(2天)、至少60小时、至少72小时(3天)、至少94小时、至少106小时(4天)或更多。在一些实施方案中，报道细胞被处理24至72小时或36至

60小时。

[0084] 当来自不具有表面标记、荧光蛋白或其组合的报道细胞的测序计数与来自具有表面标记、荧光蛋白或其组合的报道细胞的测序计数的比率的log2离开(例如,小于)阴性对照的平均值至少两个标准差时,蛋白结构域被鉴定为转录激活蛋白。(例如参见图5B)。

[0085] 3.转录因子

本公开内容还提供合成转录因子,其包含与异源DNA结合结构域融合的一个或多个转录效应物结构域。如本文所用的,术语"转录因子"指蛋白或多肽,其直接或间接地与和感兴趣的基因组基因座或基因有关的特定DNA序列相互作用以阻断或募集RNA聚合酶活性至基因或一组基因的启动子位点。

[0086] 在一些实施方案中,合成转录因子包含与异源DNA结合结构域融合的一个或多个转录激活蛋白结构域、一个或多个转录阻抑蛋白结构域或其组合。在一些实施方案中,一个或多个转录激活蛋白结构域中的至少一个或一个或多个转录阻抑蛋白结构域中的至少一个包含与SEQ ID NOs:1-896中的任一个具有至少70%(例如,至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、99%)同一性的氨基酸序列。在一些实施方案中,一个或多个转录激活蛋白结构域、一个或多个转录阻抑蛋白结构域或其组合通过本文公开的方法鉴定。

[0087] 在一些实施方案中,合成转录因子包含与异源DNA结合结构域融合的两个或更多个转录效应物结构域(例如,转录激活蛋白结构域、转录阻抑蛋白结构域或其组合)。在一些实施方案中,合成转录因子包含与异源DNA结合结构域融合的两个或更多个转录激活蛋白结构域或两个或更多个转录阻抑蛋白结构域。两个或更多个效应物结构域可以任何定向融合到DNA结合结构域,并且可以用氨基酸接头彼此分开。

[0088] 在一些实施方案中,当合成转录因子包含不止一个转录效应物结构域时,合成转录因子可包含如本文所公开的至少一个转录激活蛋白结构域或至少一个转录阻抑蛋白结构域以及至少一个本领域已知的额外的效应物结构域。参见例如,TyckoJ.等人,Cell.2020年12月23日;183(7):2020-2035,其整体引入本文作为参考。在一些实施方案中,一个或多个转录激活蛋白结构域、一个或多个转录阻抑蛋白结构域通过本文所述的方法鉴定。

[0089] 在一些实施方案中,当合成转录因子包含不止一个转录效应物结构域时,一个或多个转录激活蛋白结构域中的至少一个包含与SEQ ID NOs:563-664中的任一个具有至少70%同一性的氨基酸序列。在一些实施方案中,一个或多个转录激活蛋白结构域中的至少一个包含与SEQ ID NOs:563-596中的任一个具有至少70%同一性的氨基酸序列。在一些实施方案中,一个或多个转录激活蛋白结构域中的至少一个选自表2中发现的那些。

[0090] 在一些实施方案中,当合成转录因子包含不止一个转录效应物结构域时,一个或多个转录阻抑蛋白结构域中的至少一个包含与SEQ ID NOs:1-562和665-896中的任一个具有至少70%同一性的氨基酸序列。在一些实施方案中,一个或多个转录阻抑蛋白结构域中的至少一个包含与SEQ ID NO:666中的任一个具有至少70%同一性的氨基酸序列。在一些实施方案中,一个或多个转录阻抑蛋白结构域中的至少一个选自表1、3或4中的任一个中发现的那些。

[0091] DNA结合结构域是能够通常或具有序列特异性地结合双链或单链DNA的任何多肽。DNA结合结构域包括具有螺旋-转角-螺旋基序、锌指、亮氨酸拉链、HMG-框(高速泳动族框)

结构域、翼状螺旋区、翼状螺旋-转角-螺旋区、螺旋-环-螺旋区、免疫球蛋白折叠、B3结构域、Wor3结构域、TAL效应物DNA-结合结构域等的那些多肽。异源DNA结合结构域可以是天然结合结构域。在一些实施方案中，异源DNA结合结构域包含可编程的DNA结合结构域，例如，人工改造的DNA结合结构域，例如通过改变天然DNA结合结构域的一个或多个氨基酸以结合预定的核苷酸序列。

[0092] 在一些实施方案中，DNA结合结构域能够直接结合靶DNA序列。

[0093] DNA结合结构域可衍生自天然存在的转录激活蛋白样效应物(TALEs)中发现的结构域，例如AvrBs3、Hax2、Hax3或Hax4(Bonas等人.1989.Mol Gen Genet 218(1):127-36；Kay等人2005Mol Plant Microbe Interact 18(8):838-48)。TALEs具有由重复的残基序列组成的模块式DNA-结合结构域；每个重复区域由34个氨基酸组成。每个重复区域第12位和第13位的一对残基决定了核苷酸特异性，并且组合所述区域允许合成序列特异性TALE DNA-结合结构域。在一些实施方案中，TALE DNA结合结构域可以使用已知方法人工改造以提供对任何靶序列具有选择的特异性的DNA结合结构域。DNA结合结构域可包含多个(例如，2、3、4、5、6、10、20个或更多个)Tal效应物DNA结合基序。特别地，可以组合任何数目的核苷酸特异性Tal效应物基序以形成在本发明的转录因子中使用的序列特异性DNA-结合结构域。

[0094] 在一些实施方案中，DNA结合结构域与外源因子一致地结合靶DNA。

[0095] 在一些实施方案中，DNA结合结构域衍生自规律间隔成簇短回文重复序列相关(Cas)蛋白(例如，催化死亡Cas9)并通过指导RNA与靶DNA结合。gRNA本身包含与DNA靶序列的一条链互补的序列和将Cas9结合并募集到靶DNA序列的支架序列。本文所述的转录因子可用于CRISPR干扰(CRISPRi)或CRISPR激活(CRISPRa)。

[0096] 指导RNA(gRNA)可以是crRNA、crRNA/tracrRNA(或单指导RNA，sgRNA)。gRNA可以是非天然存在的gRNA。术语“gRNA”、“指导RNA”和“指导序列”自始至终可互换使用，且指包含决定Cas蛋白结合特异性的序列的核酸。gRNA与DNA靶序列杂交(部分或完全与其互补)。

[0097] 与靶核酸(靶位点)杂交的gRNA或其部分可以是选择性杂交所必需的任何长度。gRNAs或sgRNA(s)可以为约5至约100个核苷酸长，或更长(例如，长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、5960、61、62、63、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、9192、93、94、95、96、97、98、99或100个核苷酸，或更长)。

[0098] 为了促进gRNA设计，已经开发了许多计算工具(参见Prykhozhij等人(PLoS ONE,10(3):(2015))；Zhu等人(PLoS ONE,9(9)(2014))；Xiao等人(Bioinformatics.1月21日(2014))；Heigwer等人(Nat Methods,11(2):122-123(2014))。Zhu(Frontiers in Biology,10(4)第289-296页(2015))讨论了指导RNA设计的方法和工具，其引入本文作为参考。另外，有许多可用于促进sgRNA(s)的设计的公开可用的软件工具；包括但不限于，Genscript Interactive CRISPR gRNA设计工具(Design Tool)、WU-CRISPR和Broad Institute GPP sgRNA Designer。还有公开可用的预先设计的(pre-designed)gRNA序列以靶向许多物种(人、小鼠、大鼠、斑马鱼、秀丽隐杆线虫(C.elegans))基因组内的许多基因和位置，包括但不限于，IDT DNA预先设计的(Predesigned)Alt-R CRISPR-Cas9指导RNAs、

Addgene确认的(Validated)gRNA靶序列和GenScript全基因组gRNA数据库。

[0099] 本公开内容还提供了编码合成转录因子或转录效应物(例如,激活蛋白或阻抑蛋白)结构域的核酸,如本文所公开的。例如,效应物结构域可以由表1-3中公开的核酸编码。在一些实施方案中,效应物结构域可以由与SEQ ID NOs:897-1329中的任一个具有至少70%同一性的核酸编码。在一些实施方案中,核酸编码一种或多种合成转录因子或一种或多种效应物结构域。

[0100] 本公开内容的核酸可包含本领域已知的多种启动子中的任一种,其中启动子是组成型的、可调节的或诱导型的、细胞类型特异性的、组织特异性的或物种特异性的。除了足以指导转录的序列之外,本发明的启动子序列还可以包括与调节转录有关的其他调节元件的序列(例如,增强子、Kozak序列和内含子)。许多用于驱动基因的组成型表达的启动子/调节序列在本领域中是可用的并且包括,但不限于,例如,CMV(巨细胞病毒启动子)、EF1a(人延伸因子1α启动子)、SV40(猿猴空泡病毒40启动子)、PGK(哺乳动物磷酸甘油酸激酶启动子)、Ubc(人泛素C启动子)、人β-肌动蛋白启动子、啮齿类动物β-肌动蛋白启动子、CBh(鸡β-肌动蛋白启动子)、CAG(杂合启动子含有CMV增强子、鸡β肌动蛋白启动子和兔β-珠蛋白剪接受体)、TRE(四环素应答元件启动子)、H1(人聚合酶III RNA启动子)、U6(人U6小核启动子(human U6 small nuclear promoter))等。可用于表达本系统组分的额外的启动子包括,但不限于,巨细胞病毒(CMV)立即早期启动子,病毒LTR,例如劳斯肉瘤病毒LTR,HIV-LTR,HTLV-1 LTR,莫洛尼鼠白血病病毒(MMLV)LTR,骨髓增生性肉瘤病毒(myeoloproliferative sarcoma virus)(MPSV)LTR,脾病灶形成性病毒(spleen focus-forming virus)(SFFV)LTR,猿猴病毒40(SV40)早期启动子,单纯疱疹tk病毒启动子,具有或不具有EF1-α内含子的延伸因子1-α(EF1-α)启动子。额外的启动子包括任何组成型活性启动子。另一方面,可以使用任何可调节的启动子,从而使得可以在细胞内调节其表达。

[0101] 此外,可通过将编码这种分子的核酸置于诱导型启动子/调节序列的控制下来达到诱导型表达。本领域中众所周知的启动子可以响应于诱导剂如金属、糖皮质激素、四环素、激素等而被诱导,也预期供本发明使用。因此,将理解的是,本公开内容包括使用本领域已知的能够驱动与之可操作地连接的所期望的蛋白表达的任何启动子/调节序列。

[0102] 本公开内容还提供了含有核酸的载体和含有核酸或其载体的细胞。载体可用于在合适的细胞中繁殖核酸和/或允许从核酸表达(例如,表达载体)。本领域的普通技术人员将知道可用于核酸序列的繁殖和表达的各种载体。

[0103] 为了构建表达本转录因子的细胞,可经由常规方法构建用于稳定或瞬时表达本系统的表达载体并将其引入细胞。例如,可以将编码公开转录因子的组分或其他核酸或蛋白的核酸克隆到合适的表达载体中,例如与合适的启动子可操作地连接的质粒或病毒载体。表达载体/质粒/病毒载体的选择应适合于在真核细胞中整合和复制。

[0104] 在某些实施方案中,本公开内容的载体可以使用哺乳动物表达载体驱动一种或多种序列在哺乳动物细胞中的表达。哺乳动物表达载体的例子包括pCDM8(Seed,Nature(1987)329:840,引入本文作为参考)和pMT2PC(Kaufman等人,EMBO J.(1987)6:187,引入本文作为参考)。当用于哺乳动物细胞中时,表达载体的控制功能一般由一种或多种调节元件提供。例如,常用的启动子衍生自多瘤、腺病毒2、巨细胞病毒、猿猴病毒40以及本文公开和本领域已知的其他启动子。对于原核细胞和真核细胞两者的其他合适的表达系统,参见,例

如,引入本文作为参考的Sambrook等人,MOLECULAR CLONING:A LABORATORY MANUAL.第2版,Cold Spring Harbor Laboratory,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.,1989的第16和17章。

[0105] 本公开内容的载体可以指导核酸在特定细胞类型中的表达(例如,组织特异性调节元件用于表达核酸)。这种调节元件包括可能是组织特异性或细胞特异性的启动子。当应用于启动子时术语“组织特异性”指能够在相对没有相同的感兴趣的核苷酸序列在不同类型的组织中表达的情况下指导所述感兴趣的核苷酸序列至特定类型的组织(例如,种子)的选择性表达的启动子。当应用于启动子时术语“细胞类型特异性”指能够在相对没有相同的感兴趣的核苷酸序列在相同组织中不同类型的细胞中表达的情况下指导所述感兴趣的核苷酸序列在特定类型的细胞中的选择性表达的启动子。当应用于启动子时术语“细胞类型特异性”还意指能够促进感兴趣的核苷酸序列在单个组织内的区域中的选择性表达的启动子。启动子的细胞类型特异性可以使用本领域中众所周知的方法评估,例如,免疫组织化学染色。

[0106] 另外,载体可以含有例如下述的一些或全部:用于在宿主细胞中选择稳定或瞬时转染子的选择标记基因;转录终止和RNA加工信号;5’-和3’-非翻译区;内部核糖体结合位点(IRESes),通用多克隆位点;和用于评估嵌合受体表达的报道基因。合适的载体和用于产生含有转基因的载体的方法是本领域中众所周知和可用的。选择标记包括氯霉素抗性、四环素抗性、壮观霉素抗性、新霉素、链霉素抗性、红霉素抗性、利福平抗性、博来霉素抗性、热适应卡那霉素抗性、庆大霉素抗性、潮霉素抗性、三甲氧苄二氨嘧啶抗性、二氢叶酸还原酶(DHFR)、GPT;啤酒糖酵母(S.cerevisiae)的URA3、HIS4、LEU2和TRP1基因。

[0107] 当引入细胞中时,载体可以作为自主复制序列或染色体外元件维持或可以整合到宿主DNA中。

[0108] 因此,本公开内容进一步提供包含如本文所公开的合成转录因子、核酸或载体的细胞。

[0109] 可以使用常规的基于病毒和非基于病毒的基因转移方法将核酸引入细胞、组织或主体。这种方法可用于将核酸施用于培养物中或宿主生物中的细胞。非病毒载体递送系统包括DNA质粒、黏粒、RNA(例如,本文所述载体的转录物)、核酸和与递送媒介物复合的核酸。

[0110] 病毒载体递送系统包括DNA和RNA病毒,其在递送至细胞后具有附加型或整合基因组。多种病毒构建体可用于将本核酸递送至细胞、组织和/或主体。病毒载体包括,例如,反转录病毒、慢病毒、腺病毒、腺伴随和单纯疱疹病毒载体。这种重组病毒的非限制性实例包括重组腺伴随病毒(AAV)、重组腺病毒、重组慢病毒、重组反转录病毒、重组单纯疱疹病毒、重组痘病毒、噬菌体等。本公开内容提供了能够整合到宿主基因组中的载体,例如反转录病毒或慢病毒。参见,例如,引入本文作为参考的Ausubel等人,Current Protocols in Molecular Biology,John Wiley&Sons,New York,1989;Kay,M.A.等人,2001Nat.Medic.7(1):33-40;和Walther W.和Stein U.,2000Drugs,60(2):249-71。

[0111] 可以通过任何合适的方式递送核酸或转录因子。在某些实施方案中,核酸或其蛋白在体内递送。在其他实施方案中,将核酸或其蛋白在体外或先体外后体内递送至分离的/培养的细胞以提供可用于体内递送至患有疾病或状况的患者的修饰的细胞。

[0112] 根据本公开内容的载体可以被转化、转染或以其他方式引入多种宿主细胞中。转

染指细胞对载体的吸收，不管实际上是否表达了任何编码序列。许多转染方法是普通技术人员已知的，例如，lipofectamine、磷酸钙共沉淀、电穿孔、DEAE-葡聚糖处理、显微注射、病毒感染和本领域已知的其他方法。转导指病毒进入细胞以及由病毒载体基因组递送的序列的表达(例如，转录和/或翻译)。在重组载体的情况下，“转导”通常指重组病毒载体进入细胞以及由载体基因组递送的感兴趣的核酸的表达。

[0113] 将载体递送至细胞的方法是本领域众所周知的并且可以包括DNA或RNA电穿孔、转染试剂例如脂质体或纳米颗粒以递送DNA或RNA；通过机械变形递送DNA、RNA或蛋白(参见，例如，引入本文作为参考的Sharei等人Proc.Natl.Acad.Sci.USA(2013)110(6):2082-2087)；或病毒转导。在一些实施方案中，载体通过病毒转导递送至宿主细胞。核酸可以作为较大结构的一部分进行递送，例如质粒或病毒载体，或直接递送，例如，通过电穿孔、脂囊泡、病毒转运蛋白、显微注射和生物射弹(高速粒子轰击)。类似地，含有一种或多种转基因的构建体可以通过适合于将核酸引入细胞中的任何方法来递送。在一些实施方案中，编码本系统组分的构建体或核酸是DNA分子。在一些实施方案中，编码本系统组分的核酸是DNA载体并且可以被电穿孔到细胞中。在一些实施方案中，编码本系统组分的核酸是RNA分子，其可以被电穿孔到细胞中。

[0114] 另外，可以使用递送媒介物，例如基于纳米颗粒和脂质的递送系统。递送媒介物的进一步实例包括慢病毒载体、核糖核蛋白(RNP)复合物、基于脂质的递送系统、基因枪、流体动力学、电穿孔或核转染(nucleofection)显微注射和生物射弹。引入本文作为参考的Nayerossadat等人(Adv Biomed Res.2012;1:27)和Ibraheem等人(Int J Pharm.2014年1月1日;459(1-2):70-83)详细讨论了各种基因递送方法。

[0115] 因此，本公开内容提供了包含本文公开的一种或多种载体或一种或多种核酸的分离的细胞。优选的细胞是可以容易且可靠地生长、具有相当快的生长速率、具有良好表征的表达系统并且可以容易且有效地转化或转染的那些细胞。合适的原核细胞的实例包括，但不限于，来自属芽孢杆菌属(Bacillus)(例如枯草芽孢杆菌(Bacillus subtilis)和短芽孢杆菌(Bacillus brevis))、埃希氏菌属(Escherichia)(如大肠杆菌(E.coli))、假单胞菌属(Pseudomonas)、链霉菌属(Streptomyces)、沙门氏菌属(Salmonella)和欧文氏菌属(Envinia)的细胞。合适的真核细胞是本领域已知的并且包括，例如，酵母细胞、昆虫细胞和哺乳动物细胞。合适的酵母细胞的实例包括来自属克鲁维酵母属(Kluyveromyces)、毕赤酵母属(Pichia)、鼻孢子菌属(Rhino-sporidium)、糖酵母属(Saccharomyces)和裂殖酵母属(Schizosaccharomyces)的那些。示例性昆虫细胞包括Sf-9和HIS(Invitrogen,Carlsbad,Calif.)，并且描述于，例如，引入本文作为参考的Kitts等人,Biotechniques,14:810-817(1993)；Lucklow,Curr.Opin.Biotechnol.,4:564-572(1993)；和Lucklow等人,J.Virol.,67:4566-4579(1993)中。期望地，细胞是哺乳动物细胞，并且在一些实施方案中，细胞是人细胞。许多合适的哺乳动物和人宿主细胞是本领域已知的，并且许多可从美国典型培养物保藏中心(ATCC,Manassas,Va.)获得。合适的哺乳动物细胞的例子包括，但不限于，中国仓鼠卵巢细胞(CHO)(ATCC No.CCL61)、CHO DHFR-细胞(Urlaub等人，Proc.Natl.Acad.Sci.USA,97:4216-4220(1980))、人胚肾(HEK)293或293T细胞(ATCC No.CRL1573)和3T3细胞(ATCC No.CCL92)。其他合适的哺乳动物细胞系是猴COS-1(ATCC No.CRL1650)和COS-7细胞系(ATCC No.CRL1651)，以及CV-1细胞系(ATCC No.CCL70)。进一

步的示例性哺乳动物宿主细胞包括灵长类动物、啮齿类动物和人细胞系，包括转化的细胞系。正常二倍体细胞、衍生自初生组织以及初级外植体的体外培养物的细胞株也是合适的。其他合适的哺乳动物细胞系包括，但不限于，小鼠成神经细胞瘤N2A细胞、HeLa、HEK、A549、HepG2、小鼠L-929细胞和BHK或HaK仓鼠细胞系。

[0116] 选择合适的哺乳动物细胞的方法和用于细胞的转化、培养、扩增、筛选和纯化的方法是本领域中已知的。

[0117] 本发明还涉及包含如本文所述的合成转录因子、核酸、载体或细胞的组合物或系统。在一些实施方案中，组合物或系统包含两种或更多种合成转录因子、核酸、载体或细胞。

[0118] 在一些实施方案中，组合物或系统进一步包含gRNA。gRNA可以在与合成转录因子相同的核酸或不同的核酸上编码。在一些实施方案中，编码合成转录因子的载体可以在相同或不同的启动子下进一步编码gRNA。在一些实施方案中，gRNA在其自身的载体上编码，与转录因子的载体分开。

[0119] 4.调节基因表达的方法

本公开内容还提供调节细胞中至少一种靶基因的表达的方法，所述方法包括将至少一种如本文所述的合成转录因子、核酸、载体或组合物或系统引入所述细胞中。在一些实施方案中，调节至少两种基因的基因表达。

[0120] 表达的调节包括与靶基因的正常基因表达相比增加或减少基因表达。当至少两种基因的基因表达被调节时，两种基因都可能具有增加的基因表达，两种基因都可能具有减少的基因表达，或者一种基因可能具有增加的基因表达而另一种可能具有减少的基因表达。

[0121] 细胞可以是原核或真核细胞。在优选的实施方案中，细胞是真核细胞。在一些实施方案中，细胞是体外的。在一些实施方案中，细胞是先体外后体内的。

[0122] 在一些实施方案中，细胞在生物或宿主中，从而使得将公开的系统、组合物、载体引入细胞包括对主体的施用。方法可包括体内或通过移植先体外后体内处理的细胞向主体提供或施用至少一种如本文所述的合成转录因子、核酸、载体或组合物或系统。

[0123] “主体”可以是人或非人，并且可以包括，例如，用作用于研究目的的“模型系统”的动物品系或物种，例如如本文所述的小鼠模型。同样，主体可以包括成年的或幼年的(例如，儿童)。此外，主体可以意指任何活的生物，优选哺乳动物(例如，人或非人)，其可以受益于施用本文预期的组合物。哺乳动物的例子包括，但不限于，哺乳动物纲的任何成员：人，非人灵长类动物如黑猩猩和其他猿和猴物种；农畜，如牛、马、绵羊、山羊、猪；家畜，如兔、狗和猫；实验室动物包括啮齿类动物，例如大鼠、小鼠和豚鼠等。非哺乳动物的例子包括，但不限于，鸟、鱼等。在本文提供的方法和组合物的一个实施方案中，哺乳动物是人。

[0124] 如本文所用的，术语“提供”、“施用”、“引入”在本文中可互换使用并且指将本公开内容的系统通过导致系统至少部分定位到所期望的位点的方法或途径放置于主体中。系统可以通过任何适当的途径施用，其导致递送至主体中期望的位置。

[0125] 5.试剂盒

也在本公开内容的范围内的是试剂盒，其包括下述的至少一种或全部：至少一种编码效应物结构域或DNA结合结构域或其组合的核酸，至少一种合成转录因子或其编码核酸，编码至少一种效应物结构域或至少一种合成转录因子的载体，如本文所述的组合物或

系统,包含效应物结构域、DNA结合结构域、合成转录因子或编码其任何一种的核酸的细胞,如本文所述的报道细胞和如本文所述的两部分报道基因或编码其的核酸。

[0126] 试剂盒还可以包括使用试剂盒组分的说明书。说明书是与试剂盒有关系的相关材料或方法。材料可能包括下述的任意组合:背景信息、组分清单、使用组合物的简要或详细规程、故障检修、参考文献、技术支持和任何其他相关文件。说明书可以与试剂盒一起提供或作为单独的成员组分提供,或者是纸质形式或者是电子形式,其可以在计算机可读存储设备上提供或可从互联网网站下载,或作为记录的影像(recorded presentation)。

[0127] 应当理解,所公开的试剂盒可以与所公开的方法共同使用。试剂盒可包括用于本文所述的任何方法的说明书。说明书可以包括组分用于鉴定阻抑蛋白结构域的方法或调节基因表达的方法的用途的描述。

[0128] 本文提供的试剂盒处于合适的包装中。合适的包装包括,但不限于,小瓶、瓶、广口瓶、软包装(flexible packaging)等。

[0129] 试剂盒任选地可以提供额外的组分,例如缓冲液和解释信息。通常,试剂盒包括容器和在容器上或与伴随容器的标签或一种或多种包装插页。在一些实施方案中,本公开内容提供包括上述试剂盒内容物的制品。

[0130] 试剂盒可进一步包括用于容纳或施用本系统或组合物的设备。设备可包括输注设备、静脉内溶液袋、皮下注射针、小瓶和/或注射器。

[0131] 本公开内容还提供了用于在体外执行方法或生产组分的试剂盒。试剂盒可包括本系统的组分。试剂盒的任选组分包括以下一项或多项:(1)缓冲液成分,(2)对照质粒,(3)测序引物。

[0132] 6.实施例

人基因表达受数千种激活或阻抑转录的蛋白的调节。我们缺乏对这些蛋白的效应物结构域,足以介导基因表达的变化的结构域的完整和定量描述。为了系统地测量人细胞中的转录效应物结构域,本文提供了一种高通量测定,其中蛋白结构域文库与DNA结合结构域融合并募集到报道基因。然后通过报道基因表达水平分离细胞,并对蛋白结构域文库进行测序。报道基因是一种合成表面标记,其便于使用磁珠将数千万个细胞简单分离为高表达和低表达群体。

[0133] 在募集所有≤80个氨基酸的核内蛋白结构域后,对基因沉默和外遗传记忆进行定量。使用对>300个KRAB结构域和>200个同源异型域的完整家族的测量,发现了转录因子的阻抑蛋白结构域强度与其进化史和发育作用之间的关系。进一步地,与CRISPRi中使用的KRAB结构域相比,ZNF10 KRAB效应物功能的深度突变扫描鉴定了具有增强的稳定性和阻抑的取代。为了寻找以前注释的区域之外的效应物结构域,对238种阻抑蛋白复合物蛋白的序列进行了叠瓦,并在大染色质调节子的未注释区域中发现了短至10个氨基酸的新阻抑蛋白结构域,包括非规范的polycomb 1.6募集蛋白MGA。对大于20种阻抑蛋白进行了个别表征,且发现它们全部都在单细胞水平上以非完全有效即完全无效的方式沉默报道基因,但具有不同的沉默和外遗传记忆动态。

[0134] 此外,发现了核内蛋白中的新激活蛋白结构域,包括高度趋异的酸性KRAB结构域变体。

[0135] 这些结果共同显示了一种用于系统测量人细胞中转录效应物结构域活性的策略,

并扩大了可应用于合成转录和外遗传干扰技术的紧凑转录效应物结构域的数目。

[0136] 本技术解决的问题：

i.不知道哪些基因具有效应物功能

ii.在已知的TF/CR基因中,经常不知道哪些结构域具有所述功能

iii.在包括已知效应物结构域的结构域家族内,不知道哪些家族成员具有所述功能

iv.在已知的效应物结构域内,不知道哪些残基是必需的,以及突变如何降低或增强功能本文提供的系统和方法可以测量激活蛋白和阻抑蛋白的调节结构域改变从报道基因启动子的输出的能力。在历史上,这需要低通量工作,从而已测量了相对少的效应物结构域。本文提供的系统和方法提供了备择的高通量测定。

[0137] 所述系统和方法用于例如:a.理解基因调节,预测这些蛋白结合的非编码调节元件的功能;和b.鉴定用于外因基因组(epigenome)干扰工具的效应物结构域。

[0138] 以前,有限数目的转录效应物结构域可用于合成转录因子的人工改造。为了处理所述限制,本文提供了高通量方法以筛选和定量转录效应物结构域的功能。所述方法使得能够发现数百个效应物结构域,所述效应物结构域当融合到DNA结合结构域上时可以以靶向的方式上调或下调转录。所述方法还鉴定具有增强的活性的效应物结构域突变体。这些效应物结构域可用于人工改造用于应用于基因和细胞治疗、合成生物学和功能基因组学中的合成转录因子。

[0139] 本文提供的新转录效应物结构域对于依赖于合成转录因子的应用具有几个优点。我们鉴定了短结构域(≤80个氨基酸),和用于将它们进一步缩短到最小限度足够的序列的高通量方法,这对递送有利(例如,包装在病毒载体中)。在一些情况下,我们鉴定了短至10个氨基酸的有效的效应物结构域。结构域是从人蛋白中提取的,与病毒效应物结构域相比,其提供了降低免疫原性的优点。大多数这些结构域以前尚未被报道为转录效应物。

[0140] 通过使用Pfam结构域文库针对强pEF启动子和弱minCMV启动子两者进行高通量募集,能够测量阻抑蛋白和激活蛋白结构域两者。发现更多阻抑蛋白的一个可能原因是,它们更经常是满足结构域的Pfam定义的自主稳定折叠序列,而TADs更经常是未注释为结构域的无序或低复杂性区域。另一个可能的原因可能是辅激活蛋白比辅阻抑蛋白更局限于细胞核中(Gillespie Mol Cell2020),这意味着激活蛋白结构域的较低表达可导致更大的激活强度,但这种效应将不预期完全掩盖筛选中的信号。叠瓦转录因子或集中于具有TAD-样特征(例如,酸性)的区域的新文库设计将发现额外的激活蛋白结构域。

[0141] 此外,本文公开了用于测试这些结构域中的突变以鉴定增强的变体的高通量方法。人工细胞表面标记的开发更容易使得高通量方法成为可能,所述人工细胞表面标记提供了使用磁分离对这些文库的更有效、廉价和快速的筛选。这是比起基于荧光报道基因表达的更传统的分选文库的方法来的优点。

[0142] 实施例1

HT-募集(HT-recruit)鉴定了人蛋白中的数百个阻抑蛋白结构域

为了将经典的募集报道基因测定转化为转录结构域的高通量测定,解决了两个问题:(1)报道基因的修饰以使其与数万个结构域的文库的快速筛选相容,以及(2)开发策略以生成候选效应物结构域文库。为了改进以前公开的荧光报道基因(Bintu等人,2016),人

工改造了合成表面标记以使得能够进行大量细胞的容易的磁分离，并且报道基因被整合到适合于在大体积旋转瓶(spinner flask)中的细胞培养的悬浮细胞系中。具体而言，生成了K562报道细胞，其具有在强组成型pEF1a启动子的上游的9xTetO结合位点，所述启动子驱动由合成表面标记(与Igκ前导序列和PDGFRβ跨膜结构域连接的人IgG1Fc区)和柠檬色荧光蛋白(fluorescent citrine protein)组成的两部分报道基因的表达(图1)。流式细胞术证实已知的阻抑蛋白结构域即来自锌指转录因子ZNF10的KRAB结构域在TetO位点的募集在5天内以多西环素依赖性方式使所述报道基因沉默(图7和16A和16B)。利用结合合成表面标记的ProG Dynabeads的磁分离将具有报道基因ON的细胞与OFF的细胞分离(图7和16C)。

[0143] 从可定位于细胞核的人蛋白(包括非排他性核定位蛋白)中Pfam-注释的结构域的UniProt数据库获得序列。总起来检索到14,657个结构域。在这些中，72％小于或等于80个氨基酸(AA)长(图1)，这使得它们与作为300碱基寡核苷酸的合并合成相容。对于短于80AA的结构域，结构域序列在两端用来自天然蛋白序列的相邻残基延伸，以达到80AA的长度并避免PCR扩增偏差。添加了861个阴性对照，其或者是随机的80AA序列或者是以10AA叠瓦窗口沿着DMD蛋白叠瓦的80AA序列。DMD蛋白不定位于细胞核中(Chevron等人,1994)，且因此不太可能以具有转录活性的结构域为特色。文库被克隆用于作为与单独的rTetR多西环素诱导型DNA结合结构域或与加入3X-FLAG标签的rTetR的融合蛋白的慢病毒表达(图17A和8)并递送至K562报道细胞(图1)。

[0144] 在测定转录活性之前，使用高通量方法确定哪些蛋白结构域在K562细胞中良好表达(图17A和8)。细胞文库用抗-FLAG荧光标记的抗体染色，将细胞分选到两个箱中(图17B和8)，提取基因组DNA，并通过扩增子测序计数每个结构域的频率。测序计数用于计算对于每个结构域FLAG高对FLAG低群体的富集比率，作为表达水平的量度。这些测量结果在单独转导的生物重复实验之间是可再现的(r^2＝0.82，图17C和8)，并且与通过蛋白印迹测量的个别结构域融合物表达水平高度相关(r^2＝0.92，图17D和17E和8)。天然Pfam结构域显著比随机序列对照更好地表达($p<1e-5$，Mann Whitney检验)，而Pfam结构域和DMD叠瓦对照类似地良好表达(图17F和8)。设置阈值以鉴定具有高于随机对照中值一个标准差的FLAG高:FLAG低比率的良好表达的结构域。通过所述定义，66％的Pfam结构域良好表达；这些结构域是进一步分析的焦点。

[0145] 对Pfam结构域文库进行转录阻抑蛋白的筛选。合并的细胞文库用多西环素处理5天，这在转录沉默后给出足够的时间用于报道基因mRNA和蛋白由于细胞分裂而降解和稀释，从而结果产生清晰的‘ON’和‘OFF’细胞的双峰混合物(图18A和9)。然后，进行磁细胞分离(图18A和9)和结构域测序，然后使用未结合和珠结合的群体中的读长计数计算每个文库成员的log2(OFF:ON)比率(图1)。为清楚起见，珠结合的群体被称为‘ON’，且未结合的群体被称为‘OFF’。测量结果在单独转导的生物重复实验之间高度可再现(r^2＝0.96，图1)。当结构域引起超过表达不佳的阴性对照的平均值不止2个标准差的阻抑时，它们被称为命中。这在第5天结果产生446个阻抑蛋白命中，具有来自63个结构域家族的结构域(图12A)。这些阻抑蛋白结构域见于451种人蛋白中，这是因为在一些情况下，完全相同的结构域序列出现在多个基因中。来自被Pfam描述为阻抑蛋白或辅阻抑蛋白结合结构域的10个结构域家族的已知阻抑蛋白结构域(例如，来自人ZNF10的KRAB，来自CBX5的Chromoshadow)在所述命中中。为了测量外遗传记忆，在第9天和第13天采取额外的时间点。当与文库中使用的所有核内蛋

白相比时，含有命中的蛋白的组关于转录因子和染色质调节子显著富集，但不同类别的蛋白在通过其记忆水平分类时有差别地富集(图18B和9)。具体而言，在第13天具有高记忆(保持OFF的细胞)的阻抑蛋白最关于包括KRAB ZNF蛋白的C2H2锌指转录因子富集，而具有低记忆的阻抑蛋白最关于包括Hox蛋白的同源异型域转录因子富集。总的说来，非常高的可再现性和命中中预期的阳性对照阻抑蛋白结构域的鉴定提示称为HT-募集的筛选方法产生了可靠的结果。在核Pfam结构域文库中鉴定的阻抑蛋白的氨基酸和核酸序列显示在表1中，其中较高的分数表示增加的阻抑。

[0146] 最强的命中之一是YAF2_RYBP，存在于RING1-和YY1-结合蛋白(RYBP)及其旁系同源物(paralog)YY1-相关因子2(YAF2)中的结构域，它们都是polycomb阻抑复合物1(PRC1)的组分(Chittock等人，2017；García等人，1999)。如由Pfam所注释的来自RYBP蛋白的结构域(其仅32个氨基酸，因此比在80AA结构域文库中合成的版本更短)被单独测试并证实了报道基因的快速沉默(图12B)。RYBP-介导的沉默也在小鼠胚胎干细胞中全长RYBP蛋白募集的最近报道中得到证实(Moussa等人，2019；Zhao等人，2020)。结果确定，已通过表面等离振子共振显示为结合polycomb组蛋白修饰酶(histone modifier enzyme)RING1B的最小所需结构域的32AARYBP结构域(Wang等人，2010)足以介导细胞中的沉默。

[0147] 为了定量阻抑动力学，门控柠檬色水平分布以在未处理的细胞中均匀低水平的背景沉默归一化的情况下计算沉默细胞的百分比，然后将数据拟合到具有在多西环素处理期间的指数式沉默率和多西环素去除后的指数式衰减(或再激活)的模型，所述模型在细胞的恒定不可逆沉默百分比达到平稳状态(图12C)。使用所述方法，也确认了SUMO3、来自MPP8的Chromo结构域、来自CBX1的Chromoshadow结构域和来自SCMH1的SAM_1/SPM结构域的阻抑蛋白功能(图18C-18F和9)，它们都具有以前来自募集或辅阻抑蛋白结合测定的关于阻抑蛋白功能的支持(Chang等人，2011；Chupreta等人，2005；Frey等人，2016；Lechner等人，2000)。来自所有单独测量的沉默率(对于上面的阻抑蛋白命中和下面讨论的其他命中，图18C-18K和9)与第5天的沉默的高通量测量结果良好地相关(R^2＝0.86，图12D)。这些个别确认是使用DNA结合结构域rTetR的新变体(SE-G72P)进行的，所述变体被人工改造以在酵母中没有多西环素的情况下减轻渗漏(Roney等人，2016)，并且发现它在人细胞中不渗漏(图19A和19B)，从而使其成为哺乳动物合成生物学的有用工具。所述新的rTetR变体在最大多西环素募集时具有与原始rTetR相同的沉默强度(图19C)，这也由个别确认和筛选分数之间的高相关性证明(图12D)。这些确认实验共同证实，HT-募集既成功地鉴定了真正的阻抑蛋白，又以与单独的流式细胞术实验不相上下的准确性定量了每个结构域的阻抑强度。

[0148] 实施例2

阻抑转录的未知功能结构域的鉴定

超过22％的Pfam结构域家族被标记为未知功能结构域(DUFs)，而其他未使用所述标记命名但仍然是DUFs(El-Gebali等人，2019)。这些结构域具有可识别的序列保守性，但缺乏实验表征。因此，本文描述的高通量结构域筛选提供了将初始功能与DUFs联系起来的机会。首先，DUF3669结构域被鉴定为阻抑蛋白命中并通过流式细胞术个别确认(图12A-12C)。这些DUFs天然见于KRAB锌指蛋白中，其是含有许多阻抑转录因子的基因家族。最近公开了在两个DUF3669家族结构域募集后显示转录阻抑的一致的结果(AlChiblak等人，2019)，且高通量结果将所述发现扩展到包括四个剩余的未测试的DUF3669序列。HNF3C-末

端结构域HNF_C是另一种DUF,尽管它有更具体的名称,这是因为它仅见于肝细胞核因子3α和β(也称为FOXA1和2)中。来自FOXA1和2的HNF_C结构域也作为阻抑蛋白命中被发现。它们都包含EH1(engrailed同源性1)基序,其特征在于FxIxxIL序列,其已被指定为候选阻抑蛋白基序(Copley,2005)。

[0149] 所有三个IRF-2BP1_2N-末端锌指结构域(Childs和Goodbourn,2003),在干扰素调节因子2(IRF2)辅阻抑蛋白IRF2BP1、IRF2BP2和IRF2BPL中发现的未表征的结构域,都是阻抑蛋白命中。DNA修复因子HERC2E3连接酶中的Cyt-b5结构域(Mifsud和Bateman,2002)是另一个被确认为强阻抑蛋白命中的功能未表征的结构域(图18G和9)。BIN1中的SH3_9结构域是SH3蛋白结合结构域的一个基本上未表征的变体,其也被确认为阻抑蛋白(图18H和9)。BIN1是Myc-相互作用蛋白和肿瘤阻抑因子(Elliott等人,1999),其也与阿尔茨海默病风险有关(Nott等人,2019)。与结果一致的是,全长BIN1和Myc-相互作用结构域缺失突变体两者以前均显示在HeLa细胞中的Gal4募集测定中阻抑转录(Elliott等人,1999),并且BIN1酵母同源物(homolog)hob1已与转录阻抑和组蛋白甲基化有联系(Ramalingam和Prendergast,2007)。此外,确认了来自转录因子TOX的HMG_框结构域和来自polycomb组分PCGF2的zf-C3HC4_2RING指结构域的阻抑蛋白活性(图18I和18J)。最后,在CHD染色质重塑蛋白(chromatin remodeler)中发现了DUF1087,且尽管其高通量测量结果刚好低于筛选显著性阈值(图12A),但CHD3DUF1087被单独的流式细胞术确认为弱阻抑蛋白(图12B和12C)。这些结果共同证实,高通量蛋白结构域筛选可以将初始功能分配给DUFs,并扩展对未完全表征的结构域功能的理解。

[0150] 实施例3

具有强阻抑蛋白活性的随机序列

以前尚未测试随机序列的阻抑蛋白活性。令人惊讶地,作为阴性对照设计的随机80AA序列中的一个是具有平均log2(OFF:ON)=4.0的强阻抑蛋白命中,尽管具有低于阈值的弱表达水平。通过流式细胞术的个别确认证实所述序列在5天具有中等外遗传记忆的募集后完全沉默报道细胞群体最多到多西环素去除后两周(图18K和9)。一个额外的随机序列显示在一定程度上高于命中阈值的阻抑分数。

[0151] 实施例4

阻抑蛋白KRAB结构域见于较年轻的蛋白中

数据提供了分析最大的转录因子家族中所有效应物结构域:KRAB结构域的功能的机会。KRAB基因家族包括一些已知最强的阻抑蛋白结构域(例如ZNF10中的KRAB)。以前对阻抑KRAB结构域亚型的研究揭示它们可以通过与辅阻抑蛋白KAP1相互作用来阻抑转录,所述辅阻抑蛋白KAP1本身又与诸如SETDB1和HP1的染色质调节子相互作用(Cheng等人,2014)。然而,仍然不清楚有多少KRAB结构域是阻抑蛋白,以及KAP1的募集是否对于遍及所有KRABs的阻抑是必要或足够的。

[0152] 文库包括335个人KRAB结构域,且在关于良好表达的结构域过滤后,92.1%作为阻抑蛋白命中被发现。9个阻抑蛋白命中和2个非命中KRAB结构域通过流式细胞术个别确认,并且这些分类在每种情况下都得到确认(图19D)。然后,将结构域募集结果与以前公开的由全长KRAB蛋白拉下(pulldowns)产生的免疫沉淀法质谱分析法数据(Helleboid等人,2019)进行了比较,且除了一个以外所有的非阻抑性KRABs都在不与KAP1相互作用的蛋白中(所述

一个例外的KRAB表达低)，且所有阻抑蛋白命中KRAB结构域都是KAP1相互作用蛋白(interactors)(p<1e-9，Fisher氏精确检验，图2)。此外，分析了可用的ChiP-seq和ChIP-exo数据集(ENCODEProjectConsortium等人，2020；Imbeault等人，2017；Najafabadi等人，2015；Schmitges等人，2016)，且阻抑性KRAB结构域来自与KAP1共定位的KRAB锌指蛋白，与非阻抑性KRAB结构域形成对照(图2)。

[0153] 有趣的是，阻抑性KRAB结构域主要见于具有仅由KRAB结构域和锌指阵列组成的最简单的结构域构造的蛋白中，而非阻抑性KRAB结构域主要见于还包含DUF3669或SCAN结构域的基因中(图2)。事实上，含有DUF3669的基因ZNF783中只有一个KRAB是阻抑蛋白。ZNF783是未表征的含有DUF3669-KRAB的基因，它独特地缺乏锌指阵列(尽管其名称如此)，从而提示其在这类转录因子中在效应物功能和其定位到靶的模式方面都是特殊的。

[0154] 包含SCAN或DUF3669的复合结构域构造在进化的老KRAB基因中更常见(Imbeault等人，2017)。在这里，观察到KRAB基因的进化年龄与KRAB阻抑蛋白强度之间的清楚关系，其中来自早于有袋类-人共同祖先的基因的KRAB结构域没有阻抑蛋白活性，而来自较后进化的基因的KRAB结构域一致地起强阻抑蛋白的作用(图2)。这些结果共同支持了这样的模型，即非阻抑蛋白KRAB基因的古代生成继之以募集KAP1以沉默基因组靶的阻抑蛋白KRAB基因的更近代大规模扩张。

[0155] 实施例5

CRISPRiZNF10 KRAB效应物的深度突变扫描鉴定调节基因沉默的突变

来自ZNF10的KRAB结构域已广泛用于基因阻抑的合成生物学应用中，并在称为CRISPR干扰的可编程外遗传和转录控制工具中与dCas9融合(Gilbert等人，2014)。为了更好地理解其序列-功能关系，使用HT-募集对所述KRAB结构域进行了深度突变扫描(DMS)。设计了具有所有可能的单取代和所有连续的双重和三重取代的文库(图3)。为了提高明确比对测序读长的能力，使用可变密码子选择以在结构域编码序列中实现沉默条形码，从而使得DNA序列比氨基酸序列更独特(图3)。HT-募集是使用图1中的报道基因和工作流程进行的：5天多西环素诱导和第5、9和13天ON和OFF细胞的磁分离(图20A和10)。如所预期的那样，这些测量结果高度可再现，并且显示了随着突变长度从单一到三重的逐渐增加有害性逐渐增加的一般趋势(图20B和10)。进一步地，将这些结果与KRAB氨基酸保守性进行比较，且在保守性和突变的有害性之间发现了显著的相关性(图3)。鉴定的KRAB阻抑蛋白突变体的氨基酸和核酸序列显示在表3中。每个阻抑蛋白突变体分数相对于野生型序列的0显示，其中较高的分数表示更加增强的KRAB转录阻抑。

[0156] ZNF10 KRAB效应物有3个组分：结合KAP1所必需的A-框(Peng等人，2009)、被认为增强KAP1结合的B-框(Peng等人，2007)以及天然见于KRAB结构域上游的单独外显子上的N-末端延伸(图3)。相对于野生型序列，A-框中多个位置的突变显著降低了阻抑蛋白活性(图3)。这些突变中的几个以前已在COS和3T3细胞中用募集CAT测定进行测试；那些数据与来自K562细胞中深度突变扫描的测量结果良好相关(图3)。A-框KRAB突变体中沉默功能的完全缺乏也被个别确认(图3)。遍及A框的突变影响看来似乎是周期性的，从而提示这些残基沿α螺旋的角可能在功能上相关(图3)。这些残基被指定为沉默所必需的(p<1e-5，比较第5天所有取代相对于野生型的分布的Wilcoxon秩和检验)且发现了12个在A-框中具有强突变影响的必需残基和一个在B-框中具有显著但弱的影响的残基(图3)。

[0157] 这些取代被作图到比对的小鼠KRABA-框结构上(PDB:1v65,A-框中55%同一性,69%相似性[V13-Y54],图20C和10),且发现必需的残基在3D空间中类似地定向,提示结合界面(图3和20D,红色,和10)。这些残基对于KAP1结合可能是重要的,这是因为这些A-框残基的12个中有10个实际上显示在以前使用KRAB-O的重组蛋白结合测定(Peng等人,2009)中促进KAP1结合,所述KRAB-O与ZNF10 KRAB12-71(50%同一性,75%相似性)在包含所有12个必需残基的区域中比对(红色KRAB-O残基,图20C和10)。以前发现对于结合非必需的8个残基中剩余的8个对于DMS中的阻抑也不是必需的($p<1e-4$,Fisher氏精确检验,灰色KRAB-O残基,图20C和10)。对于结合测定中使用的单一、双重和三重丙氨酸取代检查了DMS第5天沉默分数,且发现了完美的一致性:除去结合的突变也消除了沉默(与野生型分布相比,Z-分数<-4),并且不影响结合的突变也不影响沉默(|Z-分数|<0.6)($p<0.01$,$n=12$个突变,Fisher氏精确检验)。所述高确认率及其在3D结构中的定位提示来自DMS的12个必需A-框残基中剩余的2个(V41和N45)也可能与KAP1结合有关。

[0158] 与A-框形成对照的是,B-框突变在募集结束时(第5天)显示相对少的影响,其中只有一个统计学显著的位置(P59)显示出一致但弱的作用。同时,P59和4个其他位置(K58、I62、L65、E66)显示出如在第9天所测量的多西环素去除后对记忆的显著作用(图3)。对4个重要位置进行了个别确认,并且如在高通量实验中那样,B-框突变体在募集第5天后是强的基因沉默基因,但在多西环素释放后显示减少的记忆(图3和20E和10)。为了解释所述结果,考虑了以前提出的基因沉默模型,其中沉默的细胞在进入"不可逆沉默"状态之前通过"可逆沉默"状态(Bintu等人,2016)。B-框突变体记忆减少可能是中等沉默速度降低的结果,从而导致更少的细胞在第5天前定型至不可逆沉默状态,并且对沉默速度的突变影响被掩盖,这是因为可逆沉默和不可逆沉默细胞在第5天时不能区别。为了测试所述可能性,用1/100的较低剂量的多西环素重复沉默时间过程,以调低募集强度。在所述情况下,B-框突变在第5天之前降低了沉默速度(图20E和10)。所述结果表明B-框对KRAB沉默速度有部分贡献。

[0159] 最后,KRABN-末端含有许多取代相对于野生型一致地增强沉默的残基(图3,蓝色,第13天图片)。特别地,在第13天(这是具有最大动态范围以检测高于野生型的沉默水平的时间点)几乎所有在位置8对色氨酸的取代都导致相对于野生型更大数目的沉默的细胞。这是增强的沉默的唯一重要位置(图3)。用高多西环素募集个别确认了这些突变体中排序最高的两个(WSR8EEE和AW7EE)的记忆增强(图3和20E和10)。

[0160] 所述沉默增强可能是增强的KRAB蛋白表达水平的结果。为了研究蛋白表达水平与KRAB沉默强度之间的关系,检查了KAP1-结合KRAB结构域的组的高通量FLAG-标签表达水平测量结果,并发现了KRAB表达水平与第13天的沉默之间的显著相关性($r^2=0.49$,图20F和10)。与深度突变扫描结果最相关的是,ZNF10 KRAB与显示出更高的第13天沉默水平的其他KRAB结构域相比具有较低的表达水平,从而意味着其可以经由突变得到改善。值得注意的是,N-末端的保守性非常差(图3),并且实际上通过BLAST独特地在来自ZNF10的KRAB中发现,从而提示N-末端中的稳定性改善突变将不太可能与KRAB功能相干扰。此外,在整个结构域表达中,在与表达水平呈负相关的结构域中观察到较高的色氨酸(W)频率,而较高的谷氨酸(E)频率与表达水平正相关(图20G和10)。所述氨基酸组成趋势进一步提示N-末端KRAB突变体增强可能是由于提高的表达水平,这是因为从KRAB位置8取代掉色氨酸增强了其效应物功能,并且所述增强在用谷氨酸取代时最为明显。ZNF10 KRAB变体的蛋白印迹证实,N-末

端谷氨酸取代突变体比野生型更高地表达(图20H和10)。这些结果共同证明了深度突变扫描用于对人转录阻抑蛋白的序列对功能进行作图以及通过将表达增强性取代并入保守性差的位置来改善效应物两者的用途。

[0161] 实施例6

同源异型域阻抑蛋白强度与Hox基因组织线性对应(colinear)

筛选中包括阻抑蛋白命中的第二最大结构域家族是同源异型域家族。同源异型域由3个螺旋组成,且是通过螺旋3进行碱基接触的序列特异性DNA结合结构域(Lynch等人,2006)。在一些情况下,它们也已知充当阻抑蛋白(Holland等人,2007;Schnabel和Abate-Shen,1996)。文库包括来自216个人基因的同源异型域,且26%是阻抑蛋白命中。在同源异型域的11个亚类中的4个中发现了阻抑蛋白:PRD、NKL、HOXL和LIM(图13A)。这些募集测定结果提示转录阻抑可能是同源异型域转录因子的广泛的(尽管不是普遍存在的)功能。

[0162] 然后,更仔细地检查了HOXL亚类结果。所述亚类含有Hox基因,作为细胞命运的主要调节物并在胚胎发生过程中沿前后轴确定机体发育布局区域的39个同源异型域转录因子的亚型。这些基因见于四个Hox旁系同源聚簇(A到D)中,其对应于它们沿前后轴表达的时间顺序和空间图式形成从3'到5'共线排列(Gilbert,1971)。有趣的是,它们的同源异型域的阻抑蛋白强度也与其在Hox聚簇中的排列线性对应,从而使得更5'的基因同源异型域是更强的阻抑蛋白(Spearman氏ρ=0.82,图13B)。所述相关性提示同源异型域阻抑蛋白功能与Hox基因表达的时间选择和前后轴空间图式形成之间的可能联系。

[0163] Hox同源异型域的多序列比对揭示了存在于11个最强阻抑蛋白结构域的N-末端臂中的RKKR(SEQ ID NO:1330)基序(图13C)。基序存在于最强阻抑蛋白中的碱性背景中,而排序较低的结构域缺乏所述基序,但在无序的N-末端臂中仍含有一些碱性残基,从而导致阻抑强度与带阳电的氨基酸精氨酸和赖氨酸的数目之间的显著相关性(R^2=0.85,图13C-13E)。

[0164] 在Hox同源异型域之外,Pfam核内蛋白结构域文库中99.5%的阻抑蛋白命中不含有RKKR(SEQ ID NO:1330)基序,而许多非命中含有。同样,当考虑全部结构域文库时,第5天的净结构域电荷与阻抑强度之间没有相关性(R^2=0.04)。这些结果共同提示RKKR(SEQ ID NO:1330)基序和电荷有助于募集测定中的Hox同源异型域阻抑,但当在其他结构域的背景中发现时它们对于阻抑是不足够的。

[0165] 实施例7

通过HT-募集到最小启动子发现转录激活蛋白

已确定具有弱最小CMV(minCMV)启动子的报道K562系可以在rTetR和激活结构域之间的融合物募集时被激活(图14A)。为了进行激活蛋白筛选,使用慢病毒以将核Pfam结构域文库递送至这些报道细胞,用多西环素诱导rTetR-介导的募集达48小时,将细胞(图21A)磁分离,并对两个所得到的细胞群体中的结构域进行测序。计算每个结构域的珠结合的(ON)和未结合的(OFF)群体中的测序计数的富集比率作为转录激活强度的量度,并且命中超过表达不佳的阴性对照的平均值两个标准差(图14B)。命中包括文库中存在的三个以前已知的转录激活结构域家族:来自FOXO1/3/6的FOXO-TAD、来自Myb/Myb-A的LMSTEN和来自CRTC1/2/3的TORC_C。命中的激活强度测量结果在单独转导的生物重复实验之间高度可再现(r^2=0.89,图14B)。利用短核内结构域文库的所述第二筛选确定了HT-募集可用于通过

改变报道基因的启动子来测量激活或阻抑。在核Pfam结构域文库中鉴定的激活蛋白的氨基酸和核酸序列示于表2中，其中较低的分数表示较强的激活蛋白。

[0166] 总计发现了来自26个结构域家族的48个命中。除了上述三个已知的激活蛋白结构域家族之外，具有激活蛋白命中的剩余家族以前未在Pfam上注释为激活蛋白结构域(图14C)。总的说来，发现了比阻抑蛋白更少的激活蛋白，这可能仅仅是因为激活蛋白经常是无序或低复杂性区域(Liu等人,2006)，其常常不被注释为Pfam结构域。然而，含有激活蛋白结构域的蛋白对于基因本体论术语例如‘转录的正调节’显著富集，且最强的富集是对于‘信号传导’，这反映其来源蛋白的许多都是激活因子(图21B)。进一步地，命中比非命中显著更酸性(p≤1e-5,Mann Whitney检验，图14D)，这是激活结构域中的共同性质(Mitchell和Tjian,1989;Staller等人,2018)。

[0167] 几种命中不是来源于预期经典激活蛋白结构域的序列特异性转录因子，而是来自辅激活蛋白和转录机器蛋白包括Med9、TFIIEβ和NCOA3的非经典激活蛋白。特别地，其直向同源物直接结合酵母中的其他介质复合物组分的Med9结构域(Takahashi等人,2009)是具有平均log2(OFF:ON)＝-5.5的强激活蛋白，尽管其表达水平弱。非经典激活蛋白以前已被报道在酵母中单独起作用(Gaudreau等人,1999)，但在哺乳动物细胞中个别募集时仅仅弱地起作用(Nevado等人,1999)。一个例外是TATA-结合蛋白(Dorris和Struhl,2000)。通过筛选更多非经典序列，发现了更多所述概念的例外。

[0168] 对于所有测试的结构域，报道基因的多西环素依赖性激活使用来自文库的延伸的80AA序列和调整的Pfam-注释的结构域两者进行确认(图21C)。以前注释的FOXO-TAD和LMSTEN在其延伸版本和调整版本中都是强激活蛋白。也证实了来自转录因子EGR3的DUF3446和来自SWI/SNF家族SMARCA2蛋白的基本上未表征的QLQ结构域的激活蛋白功能。进一步地，已证实Dpy-30基序结构域，Dpy-30蛋白中发现的DUF，是弱激活蛋白。Dpy-30是写成H3K4me3的组蛋白甲基转移酶复合物的核心亚基(Hyun等人,2017)，与转录活性染色质区域有关的染色质标记(Sims等人,2003)。测试了总计11个命中结构域(包括来自NCOA3的非经典命中Med9和Nuc_rec_co-act)，且当使用来自文库的延伸的80AA序列时，发现全部都显著激活报道基因。筛选和确认共同证实，可以有结果地重新筛选无偏倚的核内蛋白结构域文库以发现具有不同功能的结构域，并且除经典激活结构域(且包括DUFs)以外的一组不同的结构域可以在募集时激活转录。

[0169] 实施例8

KRAB激活蛋白结构域的发现

令人惊讶地，文库中最强的激活蛋白是来自ZNF473的KRAB结构域(图5B)。三个其他KRAB结构域(来自ZFP28、ZNF496和ZNF597)也是激活蛋白命中，所有这些都稳定表达且不是阻抑蛋白。这些结构域中的一个(来自ZNF496)以前当在HT1080细胞中单独募集时已被报道为激活蛋白(Losson和Nielsen,2010)。有趣的是，ZFP28含有两个KRAB结构域;KRAB_1是阻抑蛋白，而KRAB_2是激活蛋白。以前对全长ZFP28进行的亲和纯化/质谱分析法鉴定了与阻抑蛋白和激活蛋白两者的显著相互作用(Schmitges等人,2016)。激活蛋白KRAB结构域比非激活蛋白KRABs显著更酸性(p＝0.01,Mann Whitney检验，图14D)。序列分析显示它们与共有的KRAB序列趋异，同时彼此共享同源性，并形成变体KRAB亚聚簇(subcluster)(图14E)。以前的系统发育分析已将变体KRAB聚簇与缺乏KAP1结合和较老的进化年龄联系起来

(Helleboid等人,2019)。更具体地,激活蛋白KRAB来源蛋白中的两种(ZNF496和ZNF597)以前已经用共免疫沉淀质谱分析法进行了测试,但未发现与KAP1相互作用(Helleboid等人,2019)。

[0170] 使用以文库中使用的KRAB结构域为中心的相同80AA序列,个别确认了作为强激活蛋白的来自ZNF473的KRAB和作为中等强度激活蛋白的来自ZFP28的KRAB_2(图14F)。进一步地,调整的来自ZNF473的41AA KRAB对于强激活是足够的,而来自ZFP28的调整的37AA KRAB_2没有激活,从而意味着对于激活需要一些周围序列(图21C)。然后,检查了可用的ChiP-seq和ChIP-exo数据集(ENCODE Project Consortium等人,2020;Imbeault等人,2017;Najafabadi等人,2015;Schmitges等人,2016)并且发现ZNF473与活性染色质标记H3K27ac共定位,与阻抑性ZNF10形成对照(图14G)。人工检查时,在基因(CASC3、STAT6、WASF2、ZKSCAN2)和lncRNA(LINC00431)的转录起始位点附近发现了最显著的ZNF473峰。同时,ZFP28不与H3K27ac共定位,可能表明其KAP1-结合阻抑蛋白KRAB_1结构域是通常比起其中等强度激活蛋白KRAB_2结构域来占优势的效应物。从这些个别KRAB蛋白向远处看,含有阻抑蛋白KRAB的锌指蛋白不与H3K27ac共定位,而作为一个组的非阻抑性KRAB蛋白确实包括共定位的峰(图14G)。结果共同支持变体KRAB蛋白在功能上是各种各样的,有时起转录激活蛋白的作用。

[0171] 实施例9

叠瓦式文库发现了核内蛋白未注释的区域中的效应物结构域

Pfam注释提供了一种过滤核内蛋白组以生成相对紧凑的文库的有用方法,但Pfam目前很可能遗漏许多人效应物结构域。为了发现蛋白的未注释区域中的效应物结构域,通过组织来自沉默基因复合物的238种蛋白的列表并将它们的序列以由10个氨基酸的叠瓦窗口隔开的80个氨基酸叠瓦来设计叠瓦式文库(图15A)。进行对强pEF报道基因的高通量募集,并在多西环素5天后采取时间点以测量沉默,并在第13天(多西环素释放后8天)再次采取时间点以测量外遗传记忆(图22A)。4.3%的叠瓦在第5天得分为命中(图15B),并且它们的阻抑蛋白强度测量结果是可再现的(r^2=0.72,图22B)。总而言之,叠瓦筛选在141/238种蛋白中发现了短阻抑蛋白结构域。这些命中中的一些包括与注释的结构域重叠的阳性对照:例如,通过叠瓦ZNF57和ZNF461,这些转录因子的KRAB结构域被鉴定为阻抑性效应物,而不是序列的剩余部分(图22C)。类似地,叠瓦策略鉴定了由Pfam注释的RYBP阻抑性结构域,并且80AA叠瓦和32AAPfam结构域两者在个别确认中以相似的强度和外遗传记忆沉默(图22D)。也鉴定和确认了REST(与CoREST结合结构域重叠(Ballas等人,2001))、DNMT3b(与DNMT1和DNMT3a结合结构域重叠(Kim等人,2002))和CBX7(与募集PRC1的PcBox重叠(Li等人,2010))中的阻抑蛋白(图22E-22G)。另一类叠瓦命中在Pfam中未注释为结构域,但在文献中发现了其阻抑蛋白功能的以前的报道。例如,CTCF的氨基酸121-220在筛选中和个别确认时具有强的阻抑功能(图15C和15E),与以前在HeLa、HEK293和COS-7细胞中的募集研究一致(Drueppel等人,2004)。这些结果共同确定,蛋白叠瓦的高通量募集是鉴定真正阻抑蛋白结构域的有效策略。叠瓦式文库中鉴定的阻抑蛋白的氨基酸序列示于表4中,其中较高的分数表示增加的阻抑。

[0172] 还发现了新的未注释的阻抑蛋白结构域。例如,BAZ2A(也称为TIP5)是介导一些rDNA的转录沉默的核重塑复合物(NoRC)组分(Guetg等人,2010),但不具有任何注释的效应

物结构域。BAZ2A叠瓦数据显示了富谷氨酰胺区域中阻抑蛋白功能的峰，并且其被个别确认为中等强度阻抑蛋白(图15D和15E)。在三种TET DNA脱甲基酶(TET1/2/3)的未注释的区域中发现了阻抑蛋白叠瓦。出乎意料地，在对照蛋白DMD中也鉴定了阻抑蛋白叠瓦，其通过流式细胞术确认(图22H)。

[0173] 被认为通过在E-框基序处结合基因组并募集非规范的polycomb1.6复合物来阻抑转录的MGA(Blackledge等人，2014；Jolma等人，2013；Stielow等人，2018)叠瓦实验揭示了两个具有阻抑蛋白功能的结构域，位于两个已知的DNA结合结构域附近，此处称为阻抑蛋白1和阻抑蛋白2(图15F)。这些阻抑蛋白结构域被个别确认，并且观察到不同的沉默动态和记忆程度；第一结构域(氨基酸341-420)以缓慢沉默但强的记忆为特色，而第二结构域(氨基酸2381-2460)以具有快速再激活的情况下的快速沉默但弱的记忆为特色(图15G)。这些看来似乎是从ncPRC1.6沉默复合物中的蛋白分离的第一个效应物结构域。

[0174] 然后，试图通过检查包括显示阻抑蛋白功能的蛋白区域的所有叠瓦中的重叠并确定所有阻抑性叠瓦中存在哪些邻接的氨基酸序列来鉴定每个独立结构域中阻抑蛋白功能的最小必要序列(图15H)。使用所述方法，生成了MGA的两个候选最小化的效应物结构域：10氨基酸序列MGA[381-390]和30氨基酸序列MGA[2431-2460]，它们都与具有ConSurf-预测的功能暴露的残基的保守区域重叠。个别确认实验证实两种最小化的候选物都可以有效地使报道基因沉默(图15I)。

[0175] 材料和方法

细胞系和细胞培养

所有实验均在K562细胞(ATCCCCL-243)中进行。细胞在37℃和5％CO2的受控的湿润培养箱中，在补加有10％FBS(Hyclone)、青霉素(10,000I.U./mL)、链霉素(10,000ug/mL)和L-谷氨酰胺(2mM)的RPMI 1640(Gibco)培养基中培养。HEK293FT和HEK293T-LentiX细胞在补加有10％FBS(Hyclone)、青霉素(10,000I.U./mL)和链霉素(10,000ug/mL)的DMEM(Gibco)培养基中生长，并用于生产慢病毒。报道细胞系通过TALEN-介导的同源性指导的修复产生，以将供体构建体整合到AAVS1基因座中，如下所示：1.2×10^{6}个K562细胞在Amaxa溶液(Lonza Nucleofector 2b，设置T0-16)中用1000ng报道基因供体质粒和500ng的每种TALEN-L(Addgene#35431)和TALEN-R(Addgene#35432)质粒(分别上游和下游靶向预期的DNA切割位点)电穿孔。7天后，用1000ng/mL嘌呤霉素抗生素处理细胞5天，以选择其中供体稳定整合到预期的基因座中的群体，其提供了表达PuroR抗性基因的启动子。通过显微术和流式细胞术(BDAccuri)测量荧光报道基因表达。

[0176] 核内蛋白Pfam结构域文库设计

在UniProt数据库(UniProt Consortium,2015)查询了可以定位到细胞核的人基因。UniProt上的亚细胞定位信息根据出版物或者在只有关于相似基因(例如，直向同源物)的出版物的情况下‘通过相似性’确定，并进行人工审查。然后使用ProDy searchPfam功能检索Pfam-注释的结构域(Bakan等人，2011)。过滤得到80个氨基酸或更短的结构域，并且高度丰富、重复的C2H2锌指DNA-结合结构域被排除且预期不起转录效应物的作用。检索到注释的结构域的序列，并将其在每一侧等同地延伸以达到总计80个氨基酸。去除重复序列，然后对于人密码子选择进行密码子优化，从而去除BsmBI位点并将GC含量约束在每50个核苷酸的窗口中20％-75％(使用DNA chisel执行(Zulkower和Rosser,2020))。计算地生成499

个缺少终止密码子的80个氨基酸的随机对照作为对照。也包括以具有10氨基酸滑动窗口(sliding window)的80氨基酸叠瓦对DMD蛋白进行叠瓦的362个元件作为对照,这是因为DMD不被认为是转录调节物。总起来,文库由5,955个元件组成。

[0177] 沉默基因叠瓦式文库设计

从转录调节物的数据库组织了216种与转录沉默有关的蛋白(Lambert等人,2018)。人工添加了32种很可能与转录沉默有关的蛋白,且然后生成了无偏倚的蛋白叠瓦式文库。为了做到这一点,使用Python API从Ensembl BioMart(Kinsella等人,2011)检索每个基因的规范转录物。如果没有发现规范的转录物,则检索具有CDS的最长转录物。编码序列被分成在叠瓦之间具有10氨基酸滑动窗口的80氨基酸叠瓦。对于每个基因,包括最终叠瓦,从最后一个残基上游的80个氨基酸横跨到所述最后一个残基,从而使得C-末端区域将包括在文库中。去除重复的蛋白序列,且对于人密码子选择进行密码子优化,从而去除BsmBI位点并将GC含量约束在每50个核苷酸的窗口中20%-75%(使用DNA chisel执行(Zulkower和Rosser,2020))。如在以前的文库设计中那样,包括361个DMD叠瓦阴性对照,从而结果产生总计15,737个文库元件。

[0178] KRAB深度突变扫描文库设计

设计具有相同氨基酸的所有可能的单取代和所有连续的双重和三重取代(例如,用AAA取代)的如CRISPRi(Gilbert等人,2014)中所使用的ZNF10 KRAB结构域序列的深度突变扫描。使用概率密码子优化算法将这些氨基酸序列反向翻译成DNA序列,从而使得每个DNA序列都含有除取代的残基以外的一些变异,这提高了将测序读长与独特文库成员明确比对的能力。此外,包括在InterPro上发现的来自人KRAB基因的所有Pfam-注释的KRAB结构域,类似地如在以前的核Pfam结构域文库中那样。对于5个KRAB锌指基因也包括如在以前的叠瓦式文库中所设计的叠瓦序列。包括来自DMD基因的300个随机对照序列和200个叠瓦作为阴性对照。在密码子优化过程中,去除BsmBI位点并将GC含量约束在每80个核苷酸的窗口中30%-70%(使用DNA chisel执行(Zulkower和Rosser,2020))。总文库大小为5,731个元件。

[0179] 结构域文库克隆

具有最多到300个核苷酸的长度的寡核苷酸被作为合并的文库合成(Twist Biosciences),且然后进行PCR扩增。在干净的PCR通风橱中设立6x50ul反应以避免扩增污染DNA。对于每个反应,使用5ng模板、0.1μl的各100μM引物、1μl HerculaseII聚合酶(Agilent)、1μl DMSO、1μl10nM dNTPs和10μl 5x Herculase缓冲液。热循环规程是于98℃3分钟,然后为98℃20秒、61℃20秒、72℃30秒的循环,以及然后为72℃3分钟的最后步骤。默认循环数为29x,并且这对于每个文库进行最优化以找到结果产生用于凝胶提取的干净可见产物的最低循环(实际上,25个循是最小值)。PCR后,通过加载≥4个泳道的2%TBE凝胶、在预期长度(约300bp)切除条带并使用QIAgen凝胶提取试剂盒对所得到的dsDNA文库进行凝胶提取。在各自于37℃消化和于16℃连接5分钟的30个循环继之以于37℃的最后5分钟消化和然后于70℃热灭活20分钟的情况下,以4x10μl GoldenGate反应(75ng预消化和凝胶提取的主链质粒,5ng文库,0.13μl T4 DNA连接酶(NEB,20000U/μl),0.75μl Esp3I-HF(NEB)和1μl 10x T4 DNA连接酶缓冲液)将文库克隆到慢病毒募集载体pJT050中。然后将反应合并并用MinElute柱(QIAgen)纯化,在6ul ddH2O中洗脱。按照制造商的说明书,将每管2μl转

化到两管50μl电感受态(electrocompetent)细胞(Lucigen DUO)中。恢复后,将细胞平板接种在3-7个具有羧苄青霉素的大10"x10"LB平板上。于37℃过夜生长后,将细菌菌落刮入收集瓶中,并用HiSpeed Plasmid Maxiprep试剂盒(QIAgen)提取质粒集合体。用稀释的转化的细胞平行制备2-3个小平板,以计数菌落并确认转化效率足以维持至少30x文库覆盖度。为确定文库的质量,利用具有包括Illumina衔接子的延伸的引物通过PCR从质粒集合体和原始寡核苷酸(oligo)集合体扩增结构域,并测序。PCR和测序规程与下文对于从基因组DNA测序描述的相同,只是这些PCRs使用10ng输入DNA和17个循环。如下所述分析这些测序数据集以确定文库的覆盖率和合成质量的均匀性。此外,对来自转化的20-30个菌落进行桑格测序(Quintara),以估计克隆效率和集合体中空主链质粒的比例。

[0180] 高通量募集以测量阻抑蛋白活性

进行了K562细胞的大规模慢病毒生产和旋转感染(spinfection)。为了产生足够的慢病毒以将文库感染到K562细胞中,将HEK293T细胞平板接种在四个15-cm组织培养平板上。在每个平板上,将9×105个HEK293T细胞平板接种在30mL的DMEM中,生长过夜,且然后使用50μl聚乙烯亚胺(PEI,Polysciences#23966)由8μg三种第三代包装质粒的等摩尔混合物和8μg rTetR-结构域文库载体进行转染。48小时和72小时温育后,收获慢病毒。合并的慢病毒通过0.45-μm PVDF过滤器(Millipore)过滤以去除任何细胞碎片。对于核Pfam结构域阻抑蛋白筛选,4.5×10^7个K562报道细胞通过旋转感染用慢病毒文库感染2小时,具有两次独立的感染的生物重复实验。感染的细胞生长3天,且然后用杀稻瘟素(10μg/mL,Sigma)选择细胞。每天使用流式细胞术监控感染和选择效率以测量mCherry(BD AccuriC6)。每天通过将细胞浓度稀释回到5×10^5个细胞/mL,将细胞以对数生长条件维持在旋转瓶中,每次重复实验保持总计至少1.5×10^8个细胞,从而使得最低的维持覆盖度是每个文库元件>25,000×个细胞(补偿来自不完全的杀稻瘟素选择、文库制备和文库合成错误的损失的非常高覆盖度水平)。在感染后第6天,通过用1000ng/ml多西环素(Fisher Scientific)处理细胞5天来诱导募集,然后将细胞离心(spin down)失去多西环素和杀稻瘟素,并在未处理的RPMI培养基中再维持8天,最多到从添加多西环素计数的第13天。在每个时间点(第5、9和13天)取2.5×10^8个细胞用于测量。规程对于KRAB DMS是类似的,但在感染后第8天添加多西环素,>12,500×覆盖度,且每个时间点取2×10^8-2.2×10^8个细胞。规程对于叠瓦筛选是类似的,但感染了9.6×10^7个细胞,在感染后第8天添加多西环素,每次传代维持至少2×10^8个细胞达>12,500×覆盖度,且每个时间点取2×10^8-2.7×10^8个细胞。

[0181] 高通量募集以测量转录激活蛋白活性

对于核Pfam结构域激活蛋白筛选,如对于阻抑蛋白筛选那样生成用于在rTetR(SE-G72P)-3XFLAG载体中的核Pfam文库的慢病毒,且3.8×10^7个K562-pDY32 minCMV报道细胞通过旋转感染用慢病毒文库感染2小时,具有两次独立的感染的生物重复实验。感染的细胞生长2天,且然后用杀稻瘟素(10μg/mL,Sigma)选择细胞。每天使用流式细胞术监控感染和选择效率以测量mCherry(BD Accuri C6)。每天通过将细胞浓度稀释回到5×10^5个细胞/mL,将细胞以对数生长条件维持在旋转瓶中,每次重复实验保持总计至少1×10^8个细胞,从而使得最低的维持覆盖度是每个文库元件>18,000×个细胞。在感染后第7天,通过用1000ng/ml多西环素(Fisher Scientific)处理细胞2天来诱导募集,然后将细胞离心失去多西环素和杀稻瘟素,并在未处理的RPMI培养基中再维持4天。在第2天时间点取2×10^8个

细胞用于测量。在多西环素去除后第4天没有激活记忆的证据，如通过流式细胞术由不存在柠檬色阳性细胞所确定的，因此没有收集额外的时间点。

[0182] 报道细胞的磁分离

在每个时间点，细胞以300×g离心5分钟并吸出培养基。然后将细胞重悬浮于相同体积的PBS(Gibco)中并重复离心和吸出，以洗涤细胞并从血清去除任何IgG。Dynabeads™M-280G蛋白(ThermoFisher 10003D)通过涡旋30秒重悬浮。通过将1克无生物素的BSA(Sigma Aldrich)和200μl 0.5M pH 8.0EDTA(ThemoFisher 15575020)添加到DPBS(Gibco)中，每2×10^8个细胞制备50mL封闭缓冲液，用0.22-μm过滤器(Millipore)真空过滤，且然后保持在冰上。通过每200μl珠添加1mL缓冲液、涡旋5秒、放置在磁管支架(Eppendorf)上、等待一分钟、取出上清液并且最后从磁体取出珠且以每最初60μl珠100-600μl封闭缓冲液重悬浮，每1×10^7个细胞制备60μl珠。对于仅仅KRAB DMS，以相同的方式每1×10^7个细胞制备30μl珠。以每100μl重悬浮的珠不多于1×10^7个细胞将珠添加到细胞中，且然后在摇动的情况下于室温温育30分钟。对于具有2×10^8个细胞的样品，使用1.2mL珠，重悬浮于12mL封闭缓冲液中，在15mL Falcon管和大的磁支架中。对于具有<5×10^7个细胞的样品，使用不粘的Ambion 1.5mL管和小的磁支架。温育后，将珠和细胞混合物放置在磁支架上达>2分钟。将未结合的上清液转移到新管中，再次放置在磁体上达>2分钟以去除任何剩余的珠，且然后将上清液转移并作为未结合的级分保存。然后，将珠重悬浮在相同体积的封闭缓冲液中，再次磁分离，弃去上清液，并将具有珠的管作为结合的级分保存。将结合的部分重悬浮在封闭缓冲液或PBS中以稀释细胞(未结合的级分已经是稀释的)。使用每个级分的小部分进行流式细胞术(BD Accuri)，以估计每个级分中的细胞数目(以确保维持文库覆盖度)并基于柠檬色报道基因水平确认分离(结合的级分应是>90％柠檬色阳性的，而未结合的级分取决于报道基因水平的初始分布更加可变)。最后，将样品离心，并将沉淀于-20℃冷冻，直到基因组DNA提取为止。

[0183] 结构域融合蛋白表达水平的高通量测量

在用加入3XFLAG标签的核Pfam结构域文库感染的K562-pDY32细胞(柠檬色OFF)中进行表达水平测量。在杀稻瘟素选择(10μg/mL，Sigma)5天后使用每个生物重复实验1×10^8个细胞，这是感染后7天。将1×10^6个对照K562-JT039细胞(柠檬色ON，无慢病毒感染)掺入每个重复实验中。固定缓冲液I(BD Biosciences，BDB557870)预热至37℃达15分钟，且透化缓冲液III(BD Biosciences，BDB558050)和具有10％FBS(Hyclone)的PBS(Gibco)在冰上冷却。收集表达结构域的细胞文库，并通过流式细胞术(BD Accuri)计数细胞密度。为了固定，于37℃以每100万个细胞20μl将细胞重悬浮于体积对应于沉淀体积的固定缓冲液I(BD Biosciences，BDB557870)中达10-15分钟。用含有10％FBS的1mL冷PBS洗涤细胞，以500×g离心5分钟，且然后吸出上清液。使用冷BD透化缓冲液III(BD Biosciences，BDB558050)以每100万个细胞20μl在冰上将细胞透化30分钟，其缓慢添加并通过涡旋混合。然后如以前那样在1mlPBS+10％FBS中洗涤细胞两次，且然后吸出上清液。使用5μl/1×10^6个细胞的α-FLAG-Alexa647(RNDsystems，IC8529R)于室温遮光进行抗体染色1小时。洗涤细胞并以3×10^7个细胞/ml的浓度重悬浮于PBS+10％FBS中。在对于mCherry阳性活细胞门控后，基于APC-A荧光水平(Sony SH800S)将细胞分选到两个箱中。还在分选仪上分析了少量未染色的对照细胞，以确认染色高于背景。掺入的柠檬色阳性细胞用于评估已知缺乏3XFLAG标签的

细胞中的背景染色水平，并且在所述水平之上得出用于分选的门。分选后，遍及样品的细胞覆盖度范围为每个文库元件336-1,295个细胞。将分选的细胞以500×g离心5分钟，且然后重悬浮于PBS中。在有一项修改的情况下按照制造商的说明书进行基因组DNA提取(QIAgen血液Maxi试剂盒用于具有>1×107个细胞的样品，且具有每最多到5×10^6个细胞一个柱的QIAamp DNA Mini试剂盒用于具有≤1×10^7个细胞的样品)：于56℃进行蛋白酶K+AL缓冲液温育过夜。

[0184] 文库制备和测序

按照制造商的说明书以每个柱最多到1.25×10^8个细胞使用血液和组织(Blood& Tissue)试剂盒(QIAgen)提取基因组DNA。DNA在EB而不是AE中洗脱，以避免后来的PCR抑制。使用含有Illumina衔接子作为延伸的引物通过PCR扩增结构域序列。在50μl(通常大小的一半的)反应中使用5μg基因组DNA进行测试PCR，以验证PCR条件是否将对于每个样品结果产生预期大小的可见条带。然后，在冰上设立12-24x100μl反应(在干净的PCR通风橱中以避免扩增污染DNA)，其中反应的数目取决于每个实验中可用的基因组DNA的量。在每个反应中使用10μg基因组DNA、0.5μl各100μM引物和50μlNEBnext 2x Master Mix(NEB)。热循环规程是将热循环仪预热至98℃，然后于98℃添加样品3分钟，然后是98℃10秒、63℃30秒、72℃30秒的32x循环，且然后是72℃2分钟的最后步骤。所有后来的步骤均在PCR通风橱外进行。合并PCR反应，且≥140μl在2％TBE凝胶的至少三个泳道上与100-bp梯一起运行至少1小时，约395bp的文库条带被切出，且使用QIAquick凝胶提取试剂盒(QIAgen)纯化DNA，其中30ul洗脱到不粘管(Ambion)中。运行确认凝胶以验证小产物被取出。然后使用Qubit HS试剂盒(Thermo Fisher)对这些文库进行定量，与15％PhiX对照(Illumina)合并，并使用单端正向读长(266或300个循环)和8循环索引读长利用高输出(High output)试剂盒在Illumina NextSeq上进行测序。

[0185] 结构域测序分析

使用bcl2fastq(Illumina)对测序读长进行去多重化(demultiplex)。用脚本‘makeIndices.py’使用设计的文库序列生成了Bowtie参考，并使用脚本‘makeCounts.py’以0错配容许量将读长进行比对。使用脚本‘makeRhos.py’计算OFF和ON(或FLAG高和FLAG低)样品之间每个结构域的富集。对于给定的重复实验在两个样品中都具有<5读长的结构域从所述重复实验中除去(分配0计数)，而在一个样品中具有<5读长的结构域将使那些读长调整至5，以避免来自低深度的富集值的膨胀(inflation)。对于所有核内结构域筛选，在给定条件的两个重复实验中具有≤5计数的结构域被从下游分析中过滤掉。对于核内结构域表达筛选，良好表达的结构域是具有超过随机对照中值≥1个标准差的log2(FLAG高:FLAG低)的那些。对于核Pfam结构域阻抑蛋白筛选，命中是具有超过表达不佳的结构域的平均值≥2个标准差的log2(OFF:ON)的结构域。对于核内结构域激活蛋白筛选，命中是具有低于表达不佳的结构域的平均值≤2个标准差的log2(OFF:ON)的结构域。对于沉默基因叠瓦筛选，过滤掉在给定条件的两个重复实验中具有≤20计数的叠瓦，并且命中是具有超过随机和DMD叠瓦对照的平均值≥2个标准差的log2(OFF:ON)的叠瓦。使用PantherDB环球网工具(www.pantherdb.org)计算基因本体论分析富集。背景集是在应用计数过滤器后在实验中良好表达和测量的所有含有结构域的蛋白。使用Fisher氏精确检验计算统计显著性的P-值，计算错误发现率(False Discovery Rate)(FDR)，并且仅显示最显著的结果，均具有FDR

<10%。

[0186] 蛋白印迹和共免疫沉淀

用含有rTetR-融合物-T2A-mCherry-BSD的慢病毒载体转导的细胞用杀稻瘟素(10μg/mL)进行选择,直到mCherry>80%为止。在裂解缓冲液(1%Triton X-100、150mM NaCl、50mM Tris pH7.5、1mM EDTA、蛋白酶抑制剂混合物)中裂解细胞。使用DC蛋白测定试剂盒(Bio-Rad)对蛋白量进行定量。将相等量加载到凝胶上并转移到硝化纤维素或PVDF膜上。使用GATA1抗体(1:1000,兔,Cell Signaling Technologies cat no.3535S)和GAPDH抗体(1:2000,小鼠,ThermoFisher cat no.AM4300)或FLAG M2单克隆抗体(1:1000,小鼠,Sigma-Aldrich,目录号F1804)和组蛋白3抗体(1:1000,小鼠,Abcam cat no.AB1791)作为第一抗体探测膜。驴抗兔IRDye 680LT和山羊抗小鼠IRDye 800CW(1:20,000稀度,分别为LI-COR Biosciences,cat nos.926-68023和926-32210)或山羊抗小鼠IRDye 680RD和山羊抗兔IRDye 800CW(1:20,000稀度,分别为LI-COR Biosciences,cat nos.926-68070和926-32211)分别用作第二抗体。

[0187] 印迹在LiCor Odyssey CLx上成像。使用ImageJ定量条带强度。

[0188] 个别阻抑蛋白募集测定

使用克隆到主链pJT050或pJT126中的GoldenGate将个别效应物结构域作为与T2A-mCherry-BSD标记上游的具有或不具有3XFLAG标签的rTetR或rTetR(SE-G72P)的融合物(参见图例)克隆。K562-pJT039-pEF-柠檬色报道细胞然后用所述慢病毒载体转导,且3天后,用杀稻瘟素(10μg/mL)选择,直到>80%的细胞为mCherry阳性的为止(6-7天)。将细胞分到24-孔平板的单独孔中,并用多西环素(Fisher Scientific)处理或不处理。处理5天后,通过离心细胞去除多西环素,用DPBS(Gibco)置换培养基以稀释任何剩余的多西环素,且然后再次离心细胞并将其转移到新鲜培养基中。通过>7,000个细胞的流式细胞术分析(BD Accuri C6或Beckman Coulter CytoFLEX)每2-3天测量时间点。使用Cytoflow和定制Python脚本分析数据。关于生存力和作为递送标记的mCherry对事件进行门控。为了计算多西环素处理期间的OFF细胞的分数,将二元高斯混合模型拟合到未处理的仅rTetR阴性对照细胞,其适合ON峰和背景沉默的OFF细胞亚群两者,且然后设置低于ON峰的平均值2个标准差的阈值以将已沉默的细胞标记为OFF。使用时间匹配的未处理的对照,计算细胞的背景归一化百分比细胞$_{OFF,归一化的}$=细胞$_{OFF,+dox}$/(1-细胞$_{OFF,未处理的}$)。使用了两个独立转导的生物重复实验。基因沉默模型使用SciPy拟合到归一化的数据,其由多西环素处理阶段期间的指数衰减的渐增形式(例如,从1减去指数衰减)和多西环素去除阶段期间的指数衰减组成,具有关于在沉默和再激活开始之前的滞后时间的额外参数。

[0189] 个别激活蛋白募集测定

使用在主链pJT126中克隆的GoldenGate,将结构域作为与T2A-mCherry-BSD标记上游的rTetR(SE-G72P)的融合物克隆。K562 pDY32 minCMV柠檬色报道细胞然后用各种慢病毒载体转导,且3天后,用杀稻瘟素(10μg/mL)选择,直到>80%的细胞为mCherry阳性的为止(6-7天)。将细胞分到24-孔平板的单独孔中,并用多西环素处理或不处理。通过>15,000个细胞的流式细胞术分析(Biorad ZE5)测量时间点。为了计算多西环素处理期间的ON细胞的分数,将高斯模型拟合到未处理的仅rTetR阴性对照细胞,其适合OFF峰,且然后设置高于OFF峰的平均值2个标准差的阈值以将已激活的细胞标记为ON。使用了两个独立转导的生物

重复实验。

[0190] 加入FLAG标签的蛋白水平的流式细胞术

进行了加入FLAG标签的融合蛋白水平的染色。具体地,K562细胞用慢病毒转导以表达融合蛋白,用杀稻瘟素选择,且然后用固定缓冲液I(BD Biosciences)于37℃固定15分钟。用具有10％FBS的冷PBS洗涤细胞一次,且然后使用透化(Perm)缓冲液III(BD Biosciences)在冰上透化30分钟。将细胞洗涤两次,且然后于4℃用抗-FLAG(XX)染色1小时。最后一轮洗涤后,使用CytoFLEX(Beckman Coulter)流式细胞仪进行流式细胞术。通过根据mCherry表达对细胞进行门控使用CytoFlow对数据进行分析,且然后绘制mCherry+和非转导的细胞中的加入FLAG标签的蛋白水平。由于两个细胞组在相同样品中混合,所以所述方法作为染色效率中的变异性的对照。

[0191] 系统发育和比对分析

从Pfam检索KRAB和同源异型域序列并使用周围的天然序列延伸以达到80AA。选择表达良好的结构域用于比对。系统树和序列比对是使用比对网站Clustal Omega使用默认参数获得的(McWilliam等人,2013;Sievers等人,2011),并且在Jalview中使用默认参数建立了没有距离校正的52系统发育邻接树(Waterhouse等人,2009)。在Jalview中进行比对显示。

[0192] 氨基酸残基保守性的分析

将蛋白序列提交到ConSurf网页服务器(webserver)并使用ConSeq方法进行分析。简而言之,ConSeq通过从具有35-95％序列同一性的同源物列表中取样选择最多到150个同源物用于多信息串比对。然后,重建系统树并使用Rate4Site对保守性进行评分。ConSurf提供归一化的分数,从而使得所有残基的平均分数为零,且标准差为一。由ConSurf计算的保守性分数是蛋白中每个残基处进化保守性的相对度量,且最低分数代表蛋白中最保守的位置。ZNF10 KRABN-末端延伸的独特性通过对所有人蛋白的蛋白BLAST并在BLAST匹配中寻找其他锌指蛋白确定(Johnson等人,2008)。

[0193] ChIP-seq和ChIP-exo分析

从多个来源检索外部ChIP数据集。ENCODE ChIP-seq数据使用ENCODE的统一处理管线(uniform processing pipeline)进行处理(ENCODE Project Consortium等人,2020),并检索低于IDR阈值0.05的窄峰。来自HEK293细胞中加入标签的KRAB ZNF超表达的KRAB ZNF ChIP-exo数据和来自H1 hESCs的KAP1 ChIP-exo数据从GEO登记GSE78099获得(Imbeault等人,2017)。读长被调整为36个碱基对的统一长度,并使用Bowtie(版本1.0.1;(Langmead等人,2009))作图到人基因组的hg38版本,从而允许最多到2个错配且仅保留独特的比对。使用MACS2(版本2.1.0)(Feng等人,2012)以下述设置调入峰:“-ghs-f BAM--keep-dup all--shift-75--extsize 150--nomodel”。使用Python脚本生成浏览器轨迹(browser tracks)。对于一些ChIP-exo数据不可用的KRAB ZNFs,从GEO登记GSE76496(Schmitges等人,2016)和GSE52523(Najafabadi等人,2015)获得来自HEK293细胞中加入标签的KRAB ZNF超表达的ChIP-seq数据。如果数据集中没有其他KRAB ZNF具有离开小于250个碱基对的峰,则KRAB ZNF峰被定义为单独的结合位点。H1细胞的ENCODE H3K27ac ChIP-seq数据集使用ENCODE管线(ENCODE Project Consortium等人,2020)进行处理,使用MACS2调入窄峰,并检索低于IDR阈值0.05的峰。

[0194] 外部数据集

KRAB ZNF、KAP1和H3K27ac的ChIP-seq和ChIP-exo数据(ENCODE Project Consortium等人,2020;Imbeault等人,2017;Najafabadi等人,2015;Schmitges等人,2016)、KRAB ZNF基因进化年龄(Imbeault等人,2017)、KRAB ZNF蛋白共免疫沉淀/质谱分析法数据(Helleboid等人,2019)和KRAB阻抑蛋白活性的CAT测定(Margolin等人,1994;Witzgall等人,1994)是从以前公开的研究中检索的。

[0195] 在像每篇参考文献均被单独且具体指出引入作为参考并在本文中整体陈述的相同程度上,本文引用的所有参考文献,包括出版物、专利申请和专利,均特此引入作为参考。

[0196] 此处描述了本发明的优选实施方案,包括发明人已知的用于实施本发明的最佳模式。在阅读前述描述时,那些优选实施方案的变化形式对于本领域的普通技术人员将变得显而易见。发明人预期熟练的技术人员适当时采用这种变化形式,并且发明人要使本发明以除如本文具体描述的那样以外的其他方式实现。因此,所述发明包括如可适用的法律所允许的所附权利要求中记载的主题的所有修改和等同方案。此外,除非本文另外指出或与另外与上下文明显矛盾,否则本发明包括上述要素以其所有可能变化形式的任何组合。

表 **1: Pfam** 阻抑蛋白

Gene	Pfam 结构域序列	延伸的结构域序列	延伸的结构域 **DNA** 序列	平均阻抑 D5
ZFP28	LVTFGDVAVDFS QEEWEWLNPIQ RNLYRKVMLENY RNLASL (SEQ ID NO: 1)	NKKLEAVGTGIEPKAMSQGLVT FGDVAVDFSQEEWEWLNPIQR NLYRKVMLENYRNLASLGLCVS KPDVISSLEQGKEPW (SEQ ID NO: 2)	AACAAGAAGCTGGAGGCCGTGGGCACCGGCATCGAGCCCAAGGCCATGAGCCA GGGCCTGGTGACCTTCGGCGACGTGGCCGTGGACTTCAGCCAGGAGGAGTGGG AGTGGCTGAACCCCATCCAGAGAAACCTGTACAGAAAGGTGATGCTGGAGAAC TACAGAAACCTGGCCAGCCTGGGCCTGTGCGTGAGCAAGCCCGACGTGATCAG CAGCCTGGAGCAGGGCAAGGAGCCCTGG (SEQ ID NO: 1177)	7.89762464
ZN334	PVSFQDLTVNFT QEEWQQLDPAQ RLLYRDVMLENY SNLVSV (SEQ ID NO: 3)	KMKKFQIPVSFQDLTVNFTQEE WQQLDPAQRLLYRDVMLENYS NLVSVGYHVSKPDVIFKLEQGEE PWIVEEFSNQNYPD (SEQ ID NO: 4)	AAGATGAAGAAGTTCCAGATCCCCGTGAGCTTCCAGGACCTGACCGTGAACTTC ACCCAGGAGGAGTGGCAGCAGCTGGACCCCGCCCAGAGACTGCTGTACAGAGA TGTGATGCTGGAGAACTACAGCAACCTGGTGAGCGTGGGCTACCACGTGAGCA AGCCCGACGTGATCTTCAAGCTGGAGCAGGGCGAGGAGCCCTGGATCGTGGAG GAGTTCAGCAACCAGAACTACCCCGAC (SEQ ID NO: 897)	7.84632686
ZN568	TVTFKDVAVDLT QEEWEQMKPAQ RNLYRDVMLENY SNLVTV (SEQ ID NO: 5)	CSQESALSEEEEDTTRPLETVTF KDVAVDLTQEEWEQMKPAQR NLYRDVMLENYSNLVTVGCQV TKPDVIFKLEQEEEPW (SEQ ID NO: 6)	TGCAGCCAGGAGAGCGCCCTGAGCGAGGAGGAGGAGGACACCACCAGACCCC TGGAGACCGTGACCTTCAAGGACGTGGCCGTGGACCTGACCCAGGAGGAGTGG GAGCAGATGAAGCCCGCCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGAA CTACAGCAACCTGGTGACCGTGGGCTGCCAGGTGACCAAGCCCGACGTGATCTT CAAGCTGGAGCAGGAGGAGGAGCCCTGG (SEQ ID NO: 898)	7.83993427
ZN37A	SVSFRDVTVGFT QEEWQHLDPAQ RTLYRDVMLENY SHLVSV (SEQ ID NO: 7)	ITSQGSVSFRDVTVGFTQEEWQ HLDPAQRTLYRDVMLENYSHLV SVGYCIPKPEVILKLEKGEEPWIL EEKFPSQSHLEL (SEQ ID NO: 8)	ATCACCAGCCAGGGCAGCGTGAGCTTCAGAGATGTGACCGTGGGCTTCACCCA GGAGGAGTGGCAGCACCTGGACCCCGCCCAGAGAACCCTGTACAGAGATGTGA TGCTGGAGAACTACAGCCACCTGGTGAGCGTGGGCTACTGCATCCCCAAGCCCG AGGTGATCCTGAAGCTGGAGAAGGGCGAGGAGCCCTGGATCCTGGAGGAGAA GTTCCCCAGCCAGAGCCACCTGGAGCTG (SEQ ID NO: 899)	7.81563526
ZN181	QVTFNDVAIDFT HEEWGWLSSAQ RDLYKDVMVQN YENLVS (SEQ ID NO: 9)	PQVTFNDVAIDFTHEEWGWLS SAQRDLYKDVMVQNYENLVSV AGLSVTKPYVITLLEDGKEPWM MEKKLSKGMIPDWESR (SEQ ID NO: 10)	CCCCAGGTGACCTTCAACGACGTGGCCATCGACTTCACCCACGAGGAGTGGGGC TGGCTGAGCAGCGCCCAGAGAGACCTGTACAAGGACGTGATGGTGCAGAACTA CGAGAACCTGGTGAGCGTGGCCGGCCTGAGCGTGACCAAGCCCTACGTGATCA CCCTGCTGGAGGACGGCAAGGAGCCCTGGATGATGGAGAAGAAGCTGAGCAA GGGCATGATCCCCGACTGGGAGAGCAGA (SEQ ID NO: 900)	7.80113461
ZN510	SVSFKDVTIEFTQ EEWQQMAPVQ KNLYRDVMLENY SNLVSV (SEQ ID NO: 11)	PLRFSTLFQEQQKMNISQASVS FKDVTIEFTQEEWQQMAPVQK NLYRDVMLENYSNLVSVGYCCF KPEVIFKLEQGEEPW (SEQ ID NO: 12)	CCCCTGAGATTCAGCACCCTGTTCCAGGAGCAGCAGAAGATGAACATCAGCCAG GCCAGCGTGAGCTTCAAGGACGTGACCATCGAGTTCACCCAGGAGGAGTGGCA GCAGATGGCCCCCGTGCAGAAGAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGCAACCTGGTGAGCGTGGGCTACTGCTGCTTCAAGCCCGAGGTGATCTTCA AGCTGGAGCAGGGCGAGGAGCCCTGG (SEQ ID NO: 901)	7.65619264

ZN862	PVVFEDVAVYFT REEWGMLDKRQ KELYRDVMRMN YELLASL (SEQ ID NO: 13)	QDPSAEGLSEEVPVVFEELPVVF EDVAVYFTREEWGMLDKRQKE LYRDVMRMNYELLASLGPAAA KPDLISKLERRAAPW (SEQ ID NO: 14)	CAGGACCCCAGCGCCGAGGGCCTGAGCGAGGAGGTGCCCGTGGTGTTCGAGG AGCTGCCCGTGGTGTTCGAGGACGTGGCCGTGTACTTCACCAGAGAGGAGTGG GGCATGCTGGACAAGAGACAGAAGGAGCTGTACAGAGATGTGATGAGAATGA ACTACGAGCTGCTGGCCAGCCTGGGCCCCGCCGCCGCCAAGCCTGACCTGATCA GCAAGCTGGAGAAGAAGAGCCGCCCCCTGG (SEQ ID NO: 902)	7.64282609
ZN140	SVTFRDVAIDFSQ EEWKWLQPAQR DLYRCVMLENYG HLVSL (SEQ ID NO: 15)	SQGSVTFRDVAIDFSQEEWKW LQPAQRDLYRCVMLENYGHLV SLGLSISKPDVVSLLEQGKEPWL GKREVKRDLFSVSES (SEQ ID NO: 16)	AGCCAGGGCAGCGTGACCTTCAGAGATGTGGCCATCGACTTCAGCCAGGAGGA GTGGAAGTGGCTGCAGCCCGCCCAGAGAGACCTGTACAGATGCGTGATGCTGG AGAACTACGGCCACCTGGTGAGCCTGGGCCTGAGCATCAGCAAGCCCGACGTG GTGAGCCTGCTGGAGCAGGGCAAGGAGCCCTGGCTGGGCAAGAGAGAGGTGA AGAGAGACCTGTTCAGCGTGAGCGAGAGC (SEQ ID NO: 903)	7.59939471
ZN208	SLTFRDVAIEFSLE EWQCLDTAQQN LYRNVMLENYRN LVFL (SEQ ID NO: 17)	GSLTFRDVAIEFSLEEWQCLDTA QQNLYRNVMLENYRNLVFLGIA AFKPDLIIFLEEGKESWNMKRH EMVEESPVICSHF (SEQ ID NO: 18)	GGCAGCCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA GTGCCTGGACACCGCCCAGCAGAACCTGTACAGAAACGTGATGCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCGCCGCCTTCAAGCCCGACCTGATCATCTT CCTGGAGGAGGGCAAGGAGAGCTGGAACATGAAGAGACACGAGATGGTGGAG GAGAGCCCCGTGATCTGCAGCCACTTC (SEQ ID NO: 904)	7.57602814
ZN248	QVSFKDVCVDFT QEEWYLLDPAQK ILYRDVILENYSNL VSV (SEQ ID NO: 19)	NKSQEQVSFKDVCVDFTQEEW YLLDPAQKILYRDVILENYSNLVS VGYCITKPEVIFKIEQGEEPWILE KGFPSQCHPER (SEQ ID NO: 20)	AACAAGAGCCAGGAGCAGGTGAGCTTCAAGGACGTGTGCGTGGGACTTCACCCA GGAGGAGTGGTACCTGCTGGACCCCGCCCAGAAGATCCTGTACAGAGATGTGA TCCTGGAGAACTACAGCAACCTGGTGAGCGTGGGCTACTGCATCACCAAGCCCG AGGTGATCTTCAAGATCGAGCAGGGCGAGGAGCCCTGGATCCTGGAGAAGGGC TTCCCCAGCCAGTGCCACCCCGAGAGA (SEQ ID NO: 905)	7.53353306
ZN571	LVTFRDVAIDFSQ EEWECLDPAQRD LYRDVMLENYSN LISL (SEQ ID NO: 21)	PHLLVTFRDVAIDFSQEEWECL DPAQRDLYRDVMLENYSNLISL DLESSCVTKKLSPEKEIYEMESL QWENMGKRINHHL (SEQ ID NO: 22)	CCCCACCTGCTGGTGACCTTCAGAGATGTGGCCATCGACTTCAGCCAGGAGGAG TGGGAGTGCCTGGACCCCGCCCAGAGAGACCTGTACAGAGATGTGATGCTGGA GAACTACAGCAACCCTGATCAGCCTGGAGAGCAGCTGCGTGACCAAGA AGCTGAGCCCCGAGAAGGAGATCTACGAGATGGAGAGCCTGCAGTGGGAGAA CATGGGCAAGAGAATCAACCACCACCTG (SEQ ID NO: 906)	7.45303805
ZN699	SVVFEDVAVDFT QEEWALLDLAQR NLYRDVMLENFQ NLASL (SEQ ID NO: 23)	EEEERKTAELQKNRIQDSVVFED VAVDFTQEEWALLDLAQRNLY RDVMLENFQNLASLGYPLHTP HLISQWEQEEDLQTVK (SEQ ID NO: 24)	GAGGAGGAGAGAAAGACCGCCGAGCTGCAGAAGAACAGAATCCAGGACAGCG TGGTGTTCGAGGACGTGGCCGTGGACTTCACCCAGGAGGAGTGGGCCCTGCTG GACCTGGCCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGAACTTCCAGAA CCTGGCCAGCCTGGGCTACCCCCTGCACACCCCCCACCTGATCAGCCAGTGGGA GCAGGAGGAGGACCTGCAGACCGTGAAG (SEQ ID NO: 907)	7.44633076

ZN726	LLTFRDVAIEFSLE EWQCLDTAQKN LYRNVMLENYRN LAFL (SEQ ID NO: 25)	GLLTFRDVAIEFSLEEWQCLDTA QKNLYRNVMLENYRNLAFLGIA VSKPDLIICLEKEKEPWNMKRD EMVDEPPGICPHF (SEQ ID NO: 26)	GGCCTGCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA GTGCCTGGACACCGCCCAGAAGAACCTGTACAGAAACGTGATGCTGGAGAACT ACAGAAACCTGGCCTTCCTGGGCATCGCCGTGAGCAAGCCCGACCTGATCATCT GCCTGGAGAAGGAGAAGGAGCCCTGGAACATGAAGAGATGAGATGGTGGA CGAGCCCCCCGGCATCTGCCCCCACTTC (SEQ ID NO: 908)	7.44588981
ZIK1	CVTFEDIAIYFSQ DEWGLLDEAQRL LYLEVMLENFALV ASL (SEQ ID NO: 27)	RAPTQVTVSPETHMDLTKGCV TFEDIAIYFSQDEWGLLDEAQRL LYLEVMLENFALVASLGCGHGT EDEETPSDQNWSVG (SEQ ID NO: 28)	AGAGCCCCCACCCAGGTGACCGTGAGCCCCGAGACCCACATGGACCTGACCAA GGGCTGCGTGACCTTCGAGGACATCGCCATCTACTTCAGCCAGGACGAGTGGG GCCTGCTGGACGAGGCCCAGAGACTGCTGTACCTGGAGGGTGATGCTGGAGAAC TTCGCCCTGGTGGCCAGCCTGGGCTGCGGCCACGGCACCGAGGACGAAGAGAC CCCTAGCGACCAGAACGTGAGCGTGGGC (SEQ ID NO: 909)	7.43302782
ZNF2	SVTFEDVAVVFT DEEWSRLVPIQR DLYKEVMLENYN SIVSL (SEQ ID NO: 29)	AAVSPTTRCQESVTFEDVAVVF TDEEWSRLVPIQRDLYKEVMLE NYNSIVSLGLPVPQPDVIFQLKR GDKPWMVDLHGSE (SEQ ID NO: 30)	GCCGCCGTGAGCCCCACCACCAGATGCCAGGAGAGCGTGACCTTCGAGGACGT GGCCGTGGTGTTCACCGACGAGGAGTGGAGCAGACTGGTGCCCATCCAGAGAG ACCTGTACAAGGAGGTGATGCTGGAGAACTACAACAGCATCGTGAGCCTGGGC CTGCCCGTGCCCCAGCCCGACGTGATCTTCCAGCTGAAGAGGCGACAAGCCC TGGATGGTGGACCTGCACGGCAGCGAG (SEQ ID NO: 910)	7.40745859
Z705F	KVTFEDVAIDFTQ EEWDMMDTSKR KLYRDVMLENIS HLVSL (SEQ ID NO: 31)	HSLEKVTFEDVAIDFTQEEWD MMDTSKRKLYRDVMLENISHL VSLGYQISKSYIILQLEQGKELW REGRVFLQDQNPDRE (SEQ ID NO: 32)	CACAGCCTGGAGAAGGTGACCTTCGAGGACGTGGCCATCGACTTCACCCAGGA GGAGTGGGACATGATGGACACCAGCAAGAGAAAGCTGTACAGAGATGTGATG CTGGAGAACATCAGCCACCTGGTGAGCCTGGGCTACCAGATCAGCAAGAGCTA CATCATCCTGCAGCTGGAGCAGGGCAAGGAGCTGTGGAGAGAGGGCAGAGTG TTCCTGCAGGACCAGAACCCCGACAGAGAG (SEQ ID NO: 911)	7.40598629
ZNF14	SVSFEDVAVNFTL EEWALLDSSQKK LYEDVMQETFKN LVCL (SEQ ID NO: 33)	DSVSFEDVAVNFTLEEWALLDS SQKKLYEDVMQETFKNLVCLGK KWEDQDIEDDHRNQGKNRRC HMVERLCESRRGSKCG (SEQ ID NO: 34)	GACAGCGTGAGCTTCGAGGACGTGGCCGTGAACTTCACCCTGGAGGAGTGGGC CCTGCTGGACAGCAGCCAGAAGAAGCTGTACGAGGACGTGATGCAGGAGACCT TCAAGAACCTGGTGTGCCTGGGCAAGAAGTGGGAGGACCAGGACATCGAGGA CGACCACAGAAACCAGGGCAAGAACAGAAGATGCCACATGGTGGAGAGACTGT GCGAGAGCAGAAGAGGCAGCAAGTGCGGC (SEQ ID NO: 912)	7.3912024
ZN471	LVTFKDVAIDFSQ EEWQWMNPAQ KRLYRSMMLENY QSLVSL (SEQ ID NO: 35)	NVEVVKVMPQDLVTFKDVAID FSQEEWQWMNPAQKRLYRS MMLENYQSLVSLGLCISKPYVIS LLEQGREPWEMTSEMTR (SEQ ID NO: 36)	AACGTGGAGGTGGTGAAGGTGATGCCCCAGGACCTGGTGACCTTCAAGGACGT GGCCATCGACTTCAGCCAGGAGGAGTGGCAGTGGATGAACCCCGCCCAGAAGA GACTGTACAGAAGCATGATGCTGGAGAACTACCAGAGCCTGGTGAGCCTGGGC CTGTGCATCAGCAAGCCCTACGTGATCAGCCTGCTGGAGCAGGGCAGAGAGCC CTGGGAGATGACCAGCGAGATGACCAGA (SEQ ID NO: 913)	7.38691832

ZN624	SVTFKDVAIDFTL EEWRLMDPTQR NLHKDVMLENYR NLVSL (SEQ ID NO: 37)	TQPDEDLHLQAEETQLVKESVT FKDVAIDFTLEEWRLMDPTQR NLHKDVMLENYRNLVSLGLAVS KPDMISHLENGKGPW (SEQ ID NO: 38)	ACCCAGCCCGACGAGGACCTGCACCTGCAGGCCGAGGAGACCCAGCTGGTGAA GGAGAGCGTGACCTTCAAGGACGTGGCCATCGACTTCACCCTGGAGGAGTGGA GACTGATGGACCCCACCCAGAGAAACCTGCACACAAGGACGTGATGCTGGAGAAC TACAGAAACCTGGTGAGCCTGGGCCTGGCCGTGAGCAAGCCCGACATGATCAG CCACCTGGAGAACGGCAAGGGCCCCTGG (SEQ ID NO: 914)	7.37615807
ZNF84	SFSFDDLSVDFTQ KEWQLLDPSQKN LYKDVMLENYSSL VSL (SEQ ID NO: 39)	TMLQESFSFDDLSVDFTQKEW QLLDPSQKNLYKDVMLENYSSL VSLGYEVMKPDVIFKLEQGEEP WVGDGEIPSSDSPEV (SEQ ID NO: 40)	ACCATGCTGCAGGAGAGCTTCAGCTTCGACGACCTGAGCGTGGACTTCACCCAG AAGGAGTGGCAGCTGCTGGACCCCAGCCAGAAGAACCTGTACAAGGACGTGAT GCTGGAGAACTACAGCAGCCTGGTGAGCCTGGGCTACGAGGTGATGAAGCCCG ACGTGATCTTCAAGCTGGAGCAGGGCGAGGAGCCCTGGGTGGGCGACGGCGA GATCCCTAGCAGCGACAGCCCCGAGGTG (SEQ ID NO: 915)	7.37354184
ZNF7	VVTFGDVAVHFS REEWQCLDPGQ RALYREVMLENH SSVAG (SEQ ID NO: 41)	EVVTFGDVAVHFSREEWQCLD PGQRALYREVMLENHSSVAGL AGFLVFKPELISRLEQGEEPWVL DLQGAEGTEAPRTSK (SEQ ID NO: 42)	GAGGTGGTGACCTTCGGCGACGTGGCCGTGCACTTCAGCAGAGAGGAGTGGCA GTGCCTGGACCCCGGCCAGAGAGCCCTGTACAGAGAGGTGATGCTGGAGAACC ACAGCAGCGTGGCCGGCCTGGCCGGCTTCCTGGTGTTCAAGCCCGAGCTGATCA GCAGACTGGAGCAGGGCGAGGAGCCCTGGGTGCTGGACCTGCAGGGCGCCGA AGGCACAGAGGCCCCCAGAACCAGCAAG (SEQ ID NO: 916)	7.35816861
ZN891	PMTFKDVAVEFT QEEWVMMLDSA QRSLYRDVMLEN YRNLTSV (SEQ ID NO: 43)	RNAEEERMIAVFLTTWLQEPM TFKDVAVEFTQEEWVMMLDSA QRSLYRDVMLENYRNLTSVEYQ LYRLTVISPLDQEEIRN (SEQ ID NO: 44)	AGAAACGCCGAGGAGGAGAGAATGATCGCCGTGTTCCTGACCACCTGGCTGCA GGAGCCCATGACCTTCAAGGACGTGGCCGTGGAGTTCACCCAGGAGGAGTGGA TGATGCTGGACAGCGCCCAGAGAAGCCTGTACAGAGATGTGATGCTGGAGAAC TACAGAAACCTGACCAGCGTGGAGTACCAGCTGTACAGACTGACCGTGATCAGC CCCCTGGACCAGGAGGAGATCAGAAAC (SEQ ID NO: 917)	7.35404032
ZN337	FLAFGDVTVDFT QKEWRLLSPAQR ALYREVTLENYSH LVSL (SEQ ID NO: 45)	GPQGARRQAFLAFGDVTVDFT QKEWRLLSPAQRALYREVTLEN YSHLVSLGILHSKPELIRRLEQGE VPWGEERRRRRPGP (SEQ ID NO: 46)	GGCCCCCAGGGCGCCAGAAGACAGGCCTTCCTGGCCTTCGGCGACGTGACCGT GGACTTCACCCAGAAGGAGTGGAGACTGCTGAGCCCCGCCAGAGAGCCCTGT ACAGAGAGGTGACCCTGGAGAACTACAGCCACCTGGTGAGCCTGGGCATCCTG CACAGCAAGCCCGAGCTGATCAGAAGACTGGAGAGCAGGGCGAGGTGCCCTGGG GCGAGGAGAGAAGAAGAAGACCCGGCCCC (SEQ ID NO: 918)	7.3403856
Z705G	KLTFEDVAIDFTQ EEWAMMDTSKR KLYRDVMLENIS HLVSL (SEQ ID NO: 47)	HSLKKLTFEDVAIDFTQEEWAM MDTSKRKLYRDVMLENISHLVS LGYQISKSYIILQLEQGKELWRE GRVFLQDQNPNRE (SEQ ID NO: 48)	CACAGCCTGAAGAAGCTGACCTTCGAGGACGTGGCCATCGACTTCACCCAGGAG GAGTGGGCCATGATGGACACCAGCAAGAGAAAGCTGTACAGAGATGTGATGCT GGAGAACATCAGCCACCTGGTGAGCCTGGGCTACCAGATCAGCAAGAGCTACA TCATCCTGCAGCTGGAGCAGGGCAAGGAGCTGTGGAGAGAGGGCAGAGTGTTC CTGCAGGACCAGAACCCCAACAGAGAG (SEQ ID NO: 919)	7.33888308

ZN529	LVTLRDVVINFSQ EEWEYLDSAQRN LYWDVMMENYS NLLSL (SEQ ID NO: 49)	MPEVEFPDQFFTVLTMDHELV TLRDVVINFSQEEWEYLDSAQR NLYWDVMMENYSNLLSLDLES RNETKHLSVGKDIIQN (SEQ ID NO: 50)	ATGCCCGAGGTGGAGTTCCCCGACCAGTTCTTCACCGTGCTGACCATGGACCAC GAGCTGGTGACCCTGAGAGATGTGGTGATCAACTTCAGCCAGGAGGAGTGGGA GTACCTGGACAGCGCCCAGAGAAACCTGTACTGGGACGTGATGATGGAGAACT ACAGCAACCTGCTGAGCCTGGACCTGGAGAGCAGAAACGAGACCAAGCACCTG AGCGTGGGCAAGGACATCATCCAGAAC (SEQ ID NO: 920)	7.33722191
ZN729	PLTFRDVTIEFSLE EWQCLDTVQQN LYRDVMLENYRN LVFL (SEQ ID NO: 51)	PGAPGSLEMGPLTFRDVTIEFSL EEWQCLDTVQQNLYRDVMLE NYRNLVFLGMAVFKPDLITCLK QGKEPWNMKRHEMVT (SEQ ID NO: 52)	CCCGGCGCCCCCGGCAGCCTGGAGATGGGCCCCCTGACCTTCAGAGATGTGACC ATCGAGTTCAGCCTGGAGGAGTGGCAGTGCCTGGACACCGTGCAGCAGAACCT GTACAGAGATGTGATGCTGGAGAACTACAGAAACCTGGTGTTCCTGGGCATGG CCGTGTTCAAGCCCGACCTGATCACCTGCCTGAAGCAGGGCAAGGAGCCCTGGA ACATGAAGAGACACGAGATGGTGACC (SEQ ID NO: 921)	7.33489189
ZN419	YVTFEDVAVYFS QEEWRLLDDAQ RLLYRNVMLENF TLLASL (SEQ ID NO: 53)	RDPAQVPVAADLLTDHEEGYVT FEDVAVYFSQEEWRLLDDAQRL LYRNVMLENFTLLASLGLASSKT HEITQLESWEEPF (SEQ ID NO: 54)	AGAGACCCCGCCCAGGTGCCCGTGGCCGCCGACCTGCTGACCGACCACGAAGA AGGCTACGTGACCTTCGAGGACGTGGCCGTGTACTTCAGCCAGGAGGAGTGGA GACTGCTGGACGACGCCCAGAGACTGCTGTACAGAAAACGTGATGCTGGAGAAC TTCACCCTGCTGGCCAGCCTGGGCCTGGCCAGCAGCAAGACCCACGAGATCACC CAGCTGGAGAGCTGGGAGGAGCCCTTC (SEQ ID NO: 922)	7.33241867
Z705A	KVTFEDVAIDFTQ EEWAMMDTSKR KLYRDVMLENIS HLVSL (SEQ ID NO: 55)	HSLKKVTFEDVAIDFTQEEWA MMDTSKRKLYRDVMLENISHL VSLGYQISKSYIILQLEQGKELW REGREFLQDQNPDRE (SEQ ID NO: 56)	CACAGCCTGAAGAAGGTGACCTTCGAGGACGTGGCCATCGACTTCACCCAGGA GGAGTGGGCCATGATGGACACCAGCAAGAGAAAGCTGTACAGAGATGTGATGC TGGAGAACATCAGCCACCTGGTGAGCCTGGGCTACCAGATCAGCAAGAGCTAC ATCATCCTGCAGCTGGAGCAGGGCAAGGAGCTGTGGAGAGAGGGCAGAGAGT TCCTGCAGGACCAGAACCCCGACAGAGAG (SEQ ID NO: 923)	7.32024193
ZNF45	AVTFKDVAVVFS EEELQLLDLAQRK LYRDVMLENFRN VVSV (SEQ ID NO: 57)	TKSKEAVTFKDVAVVFSEEELQL LDLAQRKLYRDVMLENFRNVVS VGHQSTPDGLPQLEREEKLWM MKMATQRDNSSGAK (SEQ ID NO: 58)	ACCAAGAGCAAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGGTGTTCAGCGA GGAGGAGCTGCAGCTGCTGGACCTGGCCCAGAGAAAGCTGTACAGAGATGTGA TGCTGGAGAACTTCAGAAACGTGGTGAGCGTGGGCCACCAGAGCACCCCCGAC GGCCTGCCCCAGCTTGAGAGAGAGGAGAAGCTGTGGATGATGAAGATGGCCAC CCAGAGAGACAACAGCAGCGGCGCCAAG (SEQ ID NO: 924)	7.31275735
ZN302	QVTFSDVAIDFSH EEWACLDSAQR DLYKDVMVQNY ENLVSV (SEQ ID NO: 59)	SQVTFSDVAIDFSHEEWACLDS AQRDLYKDVMVQNYENLVSVG LSVTKPYVIMLLEDGKEPWMM EKKLSKAYPFPLSHSV (SEQ ID NO: 60)	AGCCAGGTGACCTTCAGCGACGTGGCCATCGACTTCAGCCACGAGGAGTGGGC CTGCCTGGACAGCGCCCAGAGACCTGTACAAGGACGTGATGGTGCAGAACT ACGAGAACCTGGTGAGCGTGGGCCTGAGCGTGACCAAGCCCTACGTGATCATG CTGCTGGAGGACGGCAAGGAGCCCTGGATGATGGAGAAGAAGCTGAGCAAGG CCTACCCCTTCCCCCTGAGCCACAGCGTG (SEQ ID NO: 925)	7.27433142

ZN486	SLQFRDVAVEFSL EEWHCLDTAQQ NLYRDVMLENYR HLVFL (SEQ ID NO: 61)	PGPLRSLEMESLQFRDVAVEFSL EEWHCLDTAQQNLYRDVMLE NYRHLVFLGIIVSKPDLITCLEQG IKPLTMKRHEMIA (SEQ ID NO: 62)	CCCGGCCCCCTGAGAAGCCTGGAGATGGAGAGCCTGCAGTTCAGAGATGTGGC CGTGGAGTTCAGCCTGGAGGAGTGGCACTGCCTGGACACCGCCCAGCAGAACC TGTACAGAGATGTGATGCTGGAGAACTACAGACACCTGGTGTTCCTGGGCATCA TCGTGAGCAAGCCCGACCTGATCACCTGCCTGGAGCAGGGCATCAAGCCCCTGA CCATGAAGAGACACGAGATGATCGCC (SEQ ID NO: 926)	7.27242434
ZN621	SVTFEDVAVYFT QNQWASLDPAQ RALYGEVMLENY ANVAS (SEQ ID NO: 63)	LQTTWPQESVTFEDVAVYFTQ NQWASLDPAQRALYGEVMLE NYANVASLVAFPFPKPALISHLE RGEAPWGPDPWDTEIL (SEQ ID NO: 64)	CTGCAGACCACCTGGCCCCCAGGAGAGCGTGACCTTCGAGGACGTGGCCGTGTA CTTCACCCAGAACCAGTGGGCCAGCCTGGACCCCGCCCAGAGAGCCCTGTACGG CGAGGTGATGCTGGAGAACTACGCCAACGTGGCCAGCCTGGTGGCCTTCCCCTT CCCCAAGCCCGCCCTGATCAGCCACCTGGAGAGAGGCGAGGCCCCCCTGGGGAC CTGACCCCTGGGACACCGAGATCCTG (SEQ ID NO: 927)	7.25940008
ZN688	TVSFADVAVYFSP EEWGCLRPAQR ALYRDVMQETYG HLGAL (SEQ ID NO: 65)	APLLAPRPGETRPGCRKPGTVS FADVAVYFSPEEWGCLRPAQR ALYRDVMQETYGHLGALGFPG PKPALISWMEQESEAW (SEQ ID NO: 66)	GCCCCCCTGCTGGCCCCCAGACCCGGCGAGACCAGACCCGGCTGCAGAAAGCC CGGCACAGTGAGCTTCGCCGACGTGGCCGTGTACTTCAGCCCCGAGGAGTGGG GCTGCCTGAGACCCGCCCAGAGAGCCCTGTACAGAGATGTGATGCAGGAGACC TACGGCCACCTGGGCGCCCTGGGCTTCCCCGGCCCCAAACCTGCTCTGATCAGCT GGATGGAGCAGGAGAGCGAGGCCTGG (SEQ ID NO: 928)	7.2566174
ZN33A	SVSFKDVTVGFT QEEWQHLDPSQ RALYRDVMLENY SNLVSV (SEQ ID NO: 67)	NKVEQKSQESVSFKDVTVGFTQ EEWQHLDPSQRALYRDVMLEN YSNLVSVGYCVHKPEVIFRLQQ GEEPWKQEEEFPSQS (SEQ ID NO: 68)	AACAAGGTGGAGCAGAAGAGCCAGGAGAGCGTGAGCTTCAAGGACGTGACCG TGGGCTTCACCCAGGAGGAGTGGCAGCACCTGGACCCCAGCCAGAGAGCCCTG TACAGAGATGTGATGCTGGAGAACTACAGCAACCTGGTGAGCGTGGGCTACTG CGTGCACAAGCCCGAGGTGATCTTCAGACTGCAGCAGGGCGAGGAGCCCTGGA AGCAGGAGGAGGAGTTCCCCAGCCAGAGC (SEQ ID NO: 929)	7.23239827
ZN554	LVTFEDVSMDFS QEEWELLEPAQK NLYREVMLENYR NVVSL (SEQ ID NO: 69)	CFSQEERMAAGYLPRWSQELV TFEDVSMDFSQEEWELLEPAQ KNLYREVMLENYRNVVSLEALK NQCTDVGIKEGPLSPA (SEQ ID NO: 70)	TGCTTCAGCCAGGAGGAGAGAATGGCCGCCGGCTACCTGCCCAGATGGAGCCA GGAGCTGGTGACCTTCGAGGACGTGAGCATGGACTTCAGCCAGGAGGAGTGG GAGCTGCTGGAGCCCGCCCAGAAGAACCTGTACAGAGAGGTGATGCTGGAGAA CTACAGAAACGTGGTGAGCCTGGAGGCCCTGAAGAACCAGTGCACCGACGTGG GCATCAAGGAGGGCCCCCTGAGCCCCGCC (SEQ ID NO: 930)	7.22964061
ZN878	SVAFEDVAVNFT QEEWALLDPSQK NLYREVMQETLR N1TSI (SEQ ID NO: 71)	DSVAFEDVAVNFTQEEWALLD PSQKNLYREVMQETLRNLTSIG KKWNNQYIEDEHQNPRRNLRR LIGERLSESKESHQHG (SEQ ID NO: 72)	GACAGCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGGACCCCAGCCAGAAGAACCTGTACAGAGAGGTGATGCAGGAGACCC TGAGAAACCTGACCAGCATCGGCAAGAAGTGGAACAACCAGTACATCGAGGAC GAGCACCAGAACCCCAGAAGAAACCTGAGAAGACTGATCGGCGAGAGACTGAG CGAGAGCAAGGAGAGCCACCAGCACGGC (SEQ ID NO: 931)	7.21922256

ZN772	QVNFEDVFVYFS QEEWVLLDEAQR LLYRDVMLENFA LMASL (SEQ ID NO: 73)	MGPAQVPMNSEVIVDPIQGQ VNFEDVFVYFSQEEWVLLDEAQ RLLYRDVMLENFALMASLGHTS FMSHIVASLVMGSEPW (SEQ ID NO: 74)	ATGGGCCCCGCCCAGGTGCCCATGAACAGCGAGGTGATCGTGGACCCCATCCA GGGCCAGGTGAACTTCGAGGACGTGTTCGTGTACTTCAGCCAGGAGGAGTGGG TGCTGCTGGACGAGGCCCAGAGACTGCTGTACAGAGATGTGATGCTGGAGAAC TTCGCCCTGATGGCCAGCCTGGGCCACACCAGCTTCATGAGCCACATCGTGGCC AGCCTGGTGATGGGCAGCGAGCCCTGG (SEQ ID NO: 932)	7.1961596
ZN224	AMTFKDVAVVFT EEELGLLDLAQRK LYRDVMLENFRN LLSV (SEQ ID NO: 75)	TTFKEAMTFKDVAVVFTEEELG LLDLAQRKLYRDVMLENFRNLL SVGHQAFHRDTFHFLREEKIW MMKTAIQREGNSGDK(SEQ ID NO: 76)	ACCACCTTCAAGGAGGCCATGACCTTCAAGGACGTGGCCGTGGTGTTCACCGAG GAGGAGCTGGGCCTGCTGGACCTGGCCCAGAGAAAGCTGTACAGAGATGTGAT GCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCCACCAGGCCTTCCACAGAG ACACCTTCCACTTCCTGAGAGAGGAGAAGATCTGGATGATGAAGACCGCCATCC AGAGAGAGGGCAACAGCGGCGACAAG (SEQ ID NO: 933)	7.18876477
ZN184	AVTFKDVIVDFTQ EEWKQLDPGQR DLFRDVTLENYTH LVSI (SEQ ID NO: 77)	DSTLLQGGHNLLSSASFQEAVT FKDVIVDFTQEEWKQLDPGQR DLFRDVTLENYTHLVSIGLQVSK PDVISQLEQGTEPW (SEQ ID NO: 78)	GACAGCACCCTGCTGCAGGGCGGCCACAACCTGCTGAGCAGCGCCAGCTTCCA GGAGGCCGTGACCTTCAAGGACGTGATCGTGGACTTCACCCAGGAGGAGTGGA AGCAGCTGGACCCCGGCCAGAGAGACCTGTTCAGAGATGTGACCCTGGAGAAC TACACCCACCTGGTGAGCATCGGCCTGCAGGTGAGCAAGCCCGACGTGATCAGC CAGCTGGAGCAGGGCACCGAGCCCTGG (SEQ ID NO: 934)	7.18783852
ZN544	SVCFEDVAMAFT QEEWEQLDLAQ RTLYREVTLETWE HIVSL (SEQ ID NO: 79)	EARSMLVPPQASVCFEDVAMA FTQEEWEQLDLAQRTLYREVTL ETWEHIVSLGLFLSKSDVISQLE QEEDLCRAEQEAPR (SEQ ID NO: 80)	GAGGCCAGAAGCATGCTGGTGCCCCCCCAGGCCAGCGTGTGCTTCGAGGACGT GGCCATGGCCTTCACCCAGGAGGAGTGGGAGCAGCTGGACCTGGCCCAGAGAA CCCTGTACAGAGAGGTGACCCTGGAGACCTGGGAGCACATCGTGAGCCTGGGC CTGTTCCTGAGCAAGAGCGACGTGATCAGCCAGCTGGAGCAGGAGGAGGACCT GTGCAGAGCCGAGCAGGAGGCCCCCAGA (SEQ ID NO: 935)	7.18695522
ZNF57	SVVFEDVAVDFTL EEWALLDSAQRD LYRDVMLETFRN LASV (SEQ ID NO: 81)	DSVVFEDVAVDFTLEEWALLDS AQRDLYRDVMLETFRNLASVD DGTQFKANGSVSLQDMYGQE KSKEQTIPNFTGNNSCA (SEQ ID NO: 82)	GACAGCGTGGTGTTCGAGGACGTGGCCGTGGACTTCACCCTGGAGGAGTGGGC CCTGCTGGACAGCGCCCAGAGAGACCTGTACAGAGATGTGATGCTGGAGACCT TCAGAAACCTGGCCAGCGTGGACGACGGCACCCAGTTCAAGGCCAACGGCAGC GTGAGCCTGCAGGACATGTACGGCCAGGAGAAGAGCAAGGAGCAGACCATCCC CAACTTCACCGGCAACAACAGCTGCGCC (SEQ ID NO: 936)	7.1854619
ZN283	LVTFRDVAIDFSQ EEWECLDPAQRD LYVDVMLENYSN LVSL (SEQ ID NO: 83)	EESHGALISSCNSRTMTDGLVTF RDVAIDFSQEEWECLDPAQRDL YVDVMLENYSNLVSLDLESKTYE TKKIFSENDIFE (SEQ ID NO: 84)	GAGGAGAGCCACGGCGCCCTGATCAGCAGCTGCAACAGCAGAACCATGACCGA CGGCCTGGTGACCTTCAGAGATGTGGCCATCGACTTCAGCCAGGAGGAGTGGG AGTGCCTGGACCCCGCCCAGAGAGACCTGTACGTGGACGTGATGCTGGAGAAC TACAGCAACCTGGTGAGCCTGGACCTGGAGAGCAAGACCTACGAGACCAAGAA GATCTTCAGCGAGAACGACATCTTCGAG (SEQ ID NO: 937)	7.16688066

ZN549	HVTFEDIAVYFSQ EEWGLLDEAQRC LYHDVMLENFSL MASV (SEQ ID NO: 85)	VITPQIPMVTEEFVKPSQGHVT FEDIAVYFSQEEWGLLDEAQRC LYHDVMLENFSLMASVGCLHGI EAEEAPSEQTLSAQ (SEQ ID NO: 86)	GTGATCACCCCCCAGATCCCCATGGTGACCGAGGAGTTCGTGAAGCCCAGCCAG GGCCACGTGACCTTCGAGGACATCGCCGTGTACTTCAGCCAGGAGGAGTGGGG CCTGCTGGACGAGGCCCAGAGATGCCTGTACCACGACGTGATGCTGGAGAACTT CAGCCTGATGGCCAGCGTGGGCTGCCTGCACGGCATCGAGGCCGAGGAGGCTC CTAGCGAGCAGACCCTGAGCGCCCAG (SEQ ID NO: 938)	7.14938492
ZN211	SVTFEDVAVYFS WEEWDLLDEAQ KHLYFDVMLENF ALTSSL (SEQ ID NO: 87)	VQLRPQTRMATALRDPASGSV TFEDVAVYFSWEEWDLLDEAQ KHLYFDVMLENFALTSSLGCWC GVEHEETPSEQRISGE (SEQ ID NO: 88)	GTGCAGCTGAGACCCCAGACCAGAATGGCCACCGCCCTGAGAGACCCCGCCAG CGGCAGCGTGACCTTCGAGGACGTGGCCGTGTACTTCAGCTGGGAGGAGTGGG ACCTGCTGGACGAGGCCCAGAAGCACCTGTACTTCGACGTGATGCTGGAGAACT TCGCCCTGACCAGCAGCCTGGGCTGCTGGTGCGGCGTGGAGCACGAGGAGACC CCCAGCGAGCAGAGAATCAGCGGCGAG (SEQ ID NO: 939)	7.14721188
ZN506	PLQFRDVAIEFSL EEWHCLDAAQR NLYRDVMLENYR NLIFL (SEQ ID NO: 89)	GPLQFRDVAIEFSLEEWHCLDA AQRNLYRDVMLENYRNLIFLGI VVSKPNLITCLEQGKKPLTMKR HEMIAKPPVMYSHF (SEQ ID NO: 90)	GGCCCCCTGCAGTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA CTGCCTGGACGCCGCCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGATCTTCCTGGGCATCGTGGTGAGCAAGCCCAACCTGATCACCT GCCTGGAGCAGGGCAAGAAGCCCCTGACCATGAAGAGACACGAGATGATCGCC AAGCCCCCCGTGATGTACAGCCACTTC (SEQ ID NO: 940)	7.1466168
ZN615	SLTLEDVAVDFT WEEWQFLSPAQ KDLYRDVMLENY SNLVAV (SEQ ID NO: 91)	MQAQESLTLEDVAVDFTWEE WQFLSPAQKDLYRDVMLENYS NLVAVGYQASKPDALSKLERGE ETCTTEDEIYSRICSEI (SEQ ID NO: 92)	ATGCAGGCCCAGGAGAGCCTGACCCTGGAGGACGTGGCCGTGGACTTCACCTG GGAGGAGTGGCAGTTCCTGAGCCCCGCCCAGAAGGACCTGTACAGAGATGTGA TGCTGGAGAACTACAGCAACCTGGTGGCCGTGGGCTACCAGGCCAGCAAGCCC GACGCCCTGAGCAAGCTGGAGAGAGGCGAGGAGACCTGCACCACCGAGGACG AGATCTACAGCAGAATCTGCAGCGAGATC (SEQ ID NO: 941)	7.13864847
ZN253	PLQFRDVAIEFSL EEWHCLDTAQR NLYRDVMLENYR NLVFL (SEQ ID NO: 93)	GPLQFRDVAIEFSLEEWHCLDT AQRNLYRDVMLENYRNLVFLGI VVSKPDLVTCLEQGKKPLTMER HEMIAKPPVMSSHF (SEQ ID NO: 94)	GGCCCCCTGCAGTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA CTGCCTGGACACCGCCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCGTGGTGAGCAAGCCCGACCTGGTGACCT GCCTGGAGCAGGGCAAGAAGCCCCTGACCATGGAGAGACACGAGATGATCGCC AAGCCCCCCGTGATGAGCAGCCACTTC (SEQ ID NO: 942)	7.12597439
ZN226	AVTFKDVAVAFT EEELGLLGPAQRK LYRDVMVENFRN LLSV (SEQ ID NO: 95)	NMFKEAVTFKDVAVAFTEEELG LLGPAQRKLYRDVMVENFRNLL SVGHPPFKQDVSPIERNEQLWI MTTATRRQGNLGEK (SEQ ID NO: 96)	AACATGTTCAAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGGCCTTCACCGAG GAGGAGCTGGGCCTGCTGGGCCCTGCTCAGAGAAAGCTGTACAGAGATGTGAT GGTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCCACCCCCCCTTCAAGCAGG ACGTGAGCCCCATCGAGAGAAACGAGCAGCTGTGGATCATGACCACCGCCACC AGAAGACAGGGCAACCTGGGCGAGAAG (SEQ ID NO: 943)	7.12032078

ZN730	ALTFRDVAIEFSLE EWQCLDTEQQN LYRNVMLDNYR NLVFL (SEQ ID NO: 97)	GALTFRDVAIEFSLEEWQCLDTE QQNLYRNVMLDNYRNLVFLGI AVSKPDLITCLEQEKEPWNLKT HDMVAKPPVICSHI (SEQ ID NO: 98)	GGCGCCCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA GTGCCTGGACACCGAGCAGCAGAACCTGTACAGAAACGTGATGCTGGACAACT ACAGAAACCTGGTGTTCCTGGGCATCGCCGTGAGCAAGCCCGACCTGATCACCT GCCTGGAGCAGGAGAAGGAGCCCTGGAACCTGAAGACCCACGACATGGTGGCC AAGCCCCCCGTGATCTGCAGCCACATC (SEQ ID NO: 944)	7.1167303
Z585A	SVSFRDVAIDFSR EEWRHLDPSQR NLYRDVMLETYS HLLSV (SEQ ID NO: 99)	SPQKSSALAPEDHGSSYEGSVSF RDVAIDFSREEWRHLDPSQRNL YRDVMLETYSHLLSVGYQVPEA EVVMLEQGKEPWA (SEQ ID NO: 100)	AGCCCCCAGAAGAGCAGCGCCCTGGCCCCCGAGGACCACGGCAGCAGCTACGA GGGAAGCGTGAGCTTCAGAGATGTGGCCATCGACTTCAGCAGAGAGGAGTGGA GACACCTGGACCCCCAGCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGACC TACAGCCACCTGCTGAGCGTGGGCTACCAGGTGCCCGAGGCCGAGGTGGTGAT GCTGGAGCAGGGCAAGGAGCCCTGGGCC (SEQ ID NO: 945)	7.11150182
ZN732	LLTFRDVAIEFSPE EWKCLDPAQQN LYRDVMLENYRN LISL (SEQ ID NO: 101)	ELLTFRDVAIEFSPEEWKCLDPA QQNLYRDVMLENYRNLISLGVA ISNPDLVIYLEQRKEPYKVKIHET VAKHPAVCSHF (SEQ ID NO: 102)	GAGCTGCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCCCGAGGAGTGGAA GTGCCTGGACCCCGCCCAGCAGAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGATCAGCCTGGGCGTGGCCATCAGCAACCCCGACCTGGTGATCT ACCTGGAGCAGAGAAAGGAGCCCTACAAGGTGAAGATCCACGAGACCGTGGCC AAGCACCCCGCCGTGTGCAGCCACTTC (SEQ ID NO: 946)	7.10058289
ZN681	PLKFRDVAIEFSLE EWQCLDTIQQNL YRNVMLENYRNL VFL (SEQ ID NO: 103)	EPLKFRDVAIEFSLEEWQCLDTI QQNLYRNVMLENYRNLVFLGIV VSKPDLITCLEQEKEPWTRKRH RMVAEPPVICSHF (SEQ ID NO: 104)	GAGCCCCTGAAGTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA GTGCCTGGACACCATCCAGCAGAACCTGTACAGAAACGTGATGCTGGAGAACTA CAGAAACCTGGTGTTCCTGGGCATCGTGGTGAGCAAGCCCGACCTGATCACCTG CCTGGAGCAGGAGAAGGAGCCCTGGACCAGAAAGAGACACAGAATGGTGGCC GAGCCCCCCGTGATCTGCAGCCACTTC (SEQ ID NO: 947)	7.09555392
ZN667	PITFGDLAIYFSQE EWEWLSPIQKDL YEDVMLENYRNL VSL (SEQ ID NO: 105)	PSARGKSKSKAPITFGDLAIYFSQ EEWEWLSPIQKDLYEDVMLEN YRNLVSLGLSFRRPNVITLLEKGK APWMVEPVRRR (SEQ ID NO: 106)	CCCAGCGCCAGAGGCAAGAGCAAGAGCAAGGCCCCCATCACCTTCGGCGACCT GGCCATCTACTTCAGCCAGGAGGAGTGGGAGTGGCTGAGCCCCATCCAGAAGG ACCTGTACGAGGACGTGATGCTGGAGAACTACAGAAACCTGGTGGAGCCTGGGC CTGAGCTTCAGAAGACCCAACGTGATCACCCTGCTGGAGAAGGGCAAGGCCCC CTGGATGGTGGAGCCCGTGAGAAGAAGA (SEQ ID NO: 948)	7.08035538
ZN649	SLTLEDVAVDFT WEEWQFLSPAQ KDLYRDVMLENY SNLVSV (SEQ ID NO: 107)	TKAQESLTLEDVAVDFTWEEW QFLSPAQKDLYRDVMLENYSNL VSVGYQAGKPDALTKLEQGEPL WTLEDEIHSPAHPEI (SEQ ID NO: 108)	ACCAAGGCCCAGGAGAGCCTGACCCTGGAGGACGTGGCCGTGGACTTCACCTG GGAGGAGTGGCAGTTCCTGAGCCCCGCCCAGAAGGACCTGTACAGAGATGTGA TGCTGGAGAACTACAGCAACCTGGTGAGCGTGGGCTACCAGGCCGGCAAGCCC GACGCCCTGACCAAGCTGGAGCAGGGCGAGCCTCTGTGGACCCTGGAGGACGA GATCCACAGCCCCGCCCACCCCGAGATC (SEQ ID NO: 949)	7.07364506

ZN470	SVTFTDVAIDFSQ DEWEWLNLAQR SLYKKVMLENYR NLVSV(SEQ ID NO: 109)	SQEEVEVAGIKLCKAMSLGSVT FTDVAIDFSQDEWEWLNLAQR SLYKKVMLENYRNLVSVGLCISK PDVISLLEQEKDPW (SEQ ID NO: 110)	AGCCAGGAGGAGGTGGAGGTGGCCGGCATCAAGCTGTGCAAGGCCATGAGCC TGGGCAGCGTGACCTTCACCGACGTGGCCATCGACTTCAGCCAGGACGAGTGG GAGTGGCTGAACCTGGCCCAGAGAAGCCTGTACAAGAAGGTGATGCTGGAGAA CTACAGAAACCTGGTGAGCGTGGGCCTGTGCATCAGCAAGCCCGACGTGATCA GCCTGCTGGAGCAGGAGAAGGACCCCTGG (SEQ ID NO: 950)	7.07241961
ZN484	SVSFKDVTVDFSR DEWQQLDLAQK SLYREVMLENYF NLISV (SEQ ID NO: 111)	TKSLESVSFKDVTVDFSRDEWQ QLDLAQKSLYREVMLENYFNLIS VGCQVPKPEVIFSLEQEEPCML DGEIPSQSRPDGD (SEQ ID NO: 112)	ACCAAGAGCCTGGAGAGCGTGAGCTTCAAGGACGTGACCGTGGACTTCAGCAG AGATGAGTGGCAGCAGCTGGACCTGGCCCAGAAGAGCCTGTACAGAGAGGTG ATGCTGGAGAACTACTTCAACCTGATCAGCGTGGGCTGCCAGGTGCCCAAGCCC GAGGTGATCTTCAGCCTGGAGCAGGAGGAGCCCTGCATGCTGGACGGCGAGAT CCCCAGCCAGAGCAGACCCGACGGCGAC (SEQ ID NO: 951)	7.07124789
ZN431	TLTFRDVAIEFSLE EWECLNPAQQN LYMNVMLENYK NLVFL (SEQ ID NO: 113)	SGCPGAERNLLVYSYFEKETLTF RDVAIEFSLEEWECLNPAQQNL YMNVMLENYKNLVFLGVAVSK QDPVTCLEQEKEPW (SEQ ID NO: 114)	AGCGGCTGCCCCGGCGCCGAGAGAAACCTGCTGGTGTACAGCTACTTCGAGAA GGAGACCCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGG AGTGCCTGAACCCCGCCCAGCAGAACCTGTACATGAACGTGATGCTGGAGAACT ACAAGAACCTGGTGTTCCTGGGCGTGGCCGTGAGCAAGCAGGACCCCGTGACC TGCCTGGAGCAGGAGAAGGAGCCCTGG (SEQ ID NO: 952)	7.06946125
ZN382	SVSFKDVTVDFT QEEWQQLDPAQ KALYRDVMLENY CHFVSV (SEQ ID NO: 115)	PLQGSVSFKDVTVDFTQEEWQ QLDPAQKALYRDVMLENYCHF VSVGFHMAKPDMIRKLEQGEE LWTQRIFPSYSYLEEDG (SEQ ID NO: 116)	CCCCTGCAGGGCAGCGTGAGCTTCAAGGACGTGACCGTGGACTTCACCCAGGA GGAGTGGCAGCAGCTGGACCCCGCCCAGAAGGCCCTGTACAGAGATGTGATGC TGGAGAACTACTGCCACTTCGTGAGCGTGGGCTTCCACATGGCCAAGCCCGACA TGATCAGAAAGCTGGAGCAGGGCGAGGAGCTGTGGACCCAGAGAATCTTCCCC AGCTACAGCTACCTGGAGGAGGACGGC (SEQ ID NO: 953)	7.06892645
ZN254	LLTFRDVAIEFSLE EWQHLDIAQQN LYRNVMLENYRN LAFL (SEQ ID NO: 117)	PGPPRSLEMGLLTFRDVAIEFSL EEWQHLDIAQQNLYRNVMLE NYRNLAFLGIAVSKPDLITCLEQ GKEPWNMKRHEMVD (SEQ ID NO: 118)	CCCGGCCCCCCCAGAAGCCTGGAGATGGGCCTGCTGACCTTCAGAGATGTGGCC ATCGAGTTCAGCCTGGAGGAGTGGCAGCACCTGGACATCGCCCAGCAGAACCT GTACAGAAACGTGATGCTGGAGAACTACAGAAACCTGGCCTTCCTGGGCATCGC CGTGAGCAAGCCCGACCTGATCACCTGCCTGGAGCAGGGCAAGGAGCCCTGGA ACATGAAGAGACACGAGATGGTGGAC (SEQ ID NO: 954)	7.06718937
ZN124	SVAFEDVAVNFT QEEWALLDPSQK NLYRDVMQETFR NLASI(SEQ ID NO: 119)	SGHPGSWEMNSVAFEDVAVN FTQEEWALLDPSQKNLYRDVM QETFRNLASIGNKGEDQSIEDQ YKNSSRNLRHIISHSGN (SEQ ID NO: 120)	AGCGGCCACCCCGGCAGCTGGGAGATGAACAGCGTGGCCTTCGAGGACGTGGC CGTGAACTTCACCCAGGAGGAGTGGGCCCTGCTGGACCCCAGCCAGAAGAACC TGTACAGAGATGTGATGCAGGAGACCTTCAGAAACCTGGCCAGCATCGGCAAC AAGGGCGAGGACCAGAGCATCGAGGACCAGTACAAGAACAGCAGCAGAAACC TGAGACACATCATCAGCCACAGCGGCAAC (SEQ ID NO: 955)	7.0598763

ZN607	SITFGDVAIDFSH QEWEYLSLVQKT LYQEVMMENYD NLVS (SEQ ID NO: 121)	SYGSITFGDVAIDFSHQEWEYLS LVQKTLYQEVMMENYDNLVSL AGHSVSKPDLITLLEQGKEPWM IVREETRGECTDLD (SEQ ID NO: 122)	AGCTACGGCAGCATCACCTTCGGCGACGTGGCCATCGACTTCAGCCACCAGGAG TGGGAGTACCTGAGCCTGGTGCAGAAGACCCTGTACCAGGAGGTGATGATGGA GAACTACGACAACCTGGTGAGCCTGGCCGGCCACAGCGTGAGCAAGCCCGACC TGATCACCCTGCTGGAGCAGGGCAAGGAGCCCTGGATGATCGTGAGAGAGGAG ACCAGAGGCGAGTGCACCGACCTGGAC (SEQ ID NO: 956)	7.05852729
ZN317	SVTFQDVAVDFT EKEWPILDSSQR KLYKDVMILENYS NITSL (SEQ ID NO: 123)	DLFVCSGLEPHTPSVGSQESVTF QDVAVDFTEKEWPILDSSQRKL YKDVMILENYSNLTSLGYQVGKP SLISHLEQEEEPR (SEQ ID NO: 124)	GACCTGTTCGTGTGCAGCGGCCTGGAGCCCCACACCCCCAGCGTGGGCAGCCA GGAAAGCGTGACCTTCCAGGACGTGGCCGTGGACTTCACCGAGAAGGAGTGGC CCCTGCTGGACAGCAGCCAGAGAAAGCTGTACAAGGACGTGATGCTGGAGAAC TACAGCAACCTGACCAGCCTGGGCTACCAGGTGGGCAAGCCCAGCCTGATCAGC CACCTGGAGCAGGAGGAGGAGCCCAGA (SEQ ID NO: 957)	7.05281313
ZN620	PVTFEDVAVYFT QNEWASLDSVQ RALYREVMILENY ANVAS (SEQ ID NO: 125)	FQTAWRQEPVTFEDVAVYFTQ NEWASLDSVQRALYREVMILEN YANVASLAFPFTTPVLVSQLEQ GELPWGLDPWEPMGRE (SEQ ID NO: 126)	TTCCAGACCGCCTGGAGACAGGAGCCCGTGACCTTCGAGGACGTGGCCGTGTA CTTCACCCAGAACGAGTGGGCCAGCCTGGACAGCGTGCAGAGAGCCCTGTACA GAGAGGTGATGCTGGAGAACTACGCCAACGTGGCCAGCCTGGCCTTCCCCTTCA CCACCCCCGTGCTGGTGAGCCAGCTGGAGCAGGGCGAGCTGCCCTGGGGACTG GACCCCTGGGAGCCTATGGGCAGAGAG (SEQ ID NO: 958)	7.04082891
ZN141	LLTFRDVAIEFSPE EWKCLDPDQQN LYRDVMILENYRN LVSL (SEQ ID NO: 127)	ELLTFRDVAIEFSPEEWKCLDPD QQNLYRDVMILENYRNLVSLGV AISNPDLVTCLEQRKEPYNVKIH KIVARPPAMCSHF (SEQ ID NO: 128)	GAGCTGCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCCCGAGGAGTGGAA GTGCCTGGACCCCGACCAGCAGAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGTGAGCCTGGGCGTGGCCATCAGCAACCCCGACCTGGTGACCT GCCTGGAGCAGAGAAAGGAGCCCTACAACGTGAAGATCCACAAGATCGTGGCC AGACCCCCCGCCATGTGCAGCCACTTC (SEQ ID NO: 959)	7.03997569
ZN584	LVMFEDVTVYFS REEWGLLNVTQK GLYRDVMILENFA LVSSL (SEQ ID NO: 129)	AGEAEAQLDPSLQGLVMFEDV TVYFSREEWGLLNVTQKGLYRD VMILENFALVSSLGLAPSRSPVFT QLEDDEQSWVPSWV (SEQ ID NO: 130)	GCCGGCGAGGCCGAGGCCCAGCTGGACCCCAGCCTGCAGGGCCTGGTTATGTT CGAGGACGTGACCGTGTACTTCAGCAGAGAGGAGTGGGGCCTGCTGAACGTGA CCCAGAAGGGCCTGTACAGAGATGTGATGCTGGAGAACTTCGCCCTGGTGAGC AGCCTGGGCCTGGCCCCCAGCAGAAGCCCCGTGTTCACCCAGCTGGAGGACGA CGAGCAGAGCTGGGTGCCCAGCTGGGTG (SEQ ID NO: 960)	7.03820051
ZN540	LVTFRDVAIDFSQ KEWECLDTTQRK LYRDVMILENYNN LVSL (SEQ ID NO: 131)	AHALVTFRDVAIDFSQKEWECL DTTQRKLYRDVMILENYNNLVSL GYSGSKPDVITLLEQGKEPCVVA RDVTGRQCPGLLS (SEQ ID NO: 132)	GCCCACGCCCTGGTGACCTTCAGAGATGTGGCCATCGACTTCAGCCAGAAGGAG TGGGAGTGCCTGGACACCACCCAGAGAAAGCTGTACAGAGATGTGATGCTGGA GAACTACAACAACCTGGTGAGCCTGGGCTACAGCGGCAGCAAGCCCGACGTGA TCACCCTGCTGGAGCAGGGCAAGGAGCCCTGCGTGGTGGCCAGAGATGTGACC GGCAGACAGTGCCCCGGCCTGCTGAGC (SEQ ID NO: 961)	7.03581318

ZN75D	LLTFEDVAVYFSE EEWQLLNPLEKT LYNDVMQDIYET VISL (SEQ ID NO: 133)	KRIKHWKMASKLILPESLSLLTFE DVAVYFSEEEWQLLNPLEKTLY NDVMQDIYETVISLGLKLKNDT GNDHPISVSTSE (SEQ ID NO: 134)	AAGAGAATCAAGCACTGGAAGATGGCCAGCAAGCTGATCCTGCCCGAGAGCCT GAGCCTGCTGACCTTCGAGGACGTGGCCGTGTACTTCAGCGAGGAGGAGTGGC AGCTGCTGAACCCCCTGGAGAAGACCCTGTACAACGACGTGATGCAGGACATCT ACGAGACCGTGATCAGCCTGGGCCTGAAGCTGAAGAACGACACCGGCAACGAC CACCCCATCAGCGTGAGCACCAGCGAG (SEQ ID NO: 962)	7.02809755
ZN555	SVVFEDVAVDFTL EEWALLDSAQRD LYRDVMLETFQN LASV(SEQ ID NO: 135)	DSVVFEDVAVDFTLEEWALLDS AQRDLYRDVMLETFQNLASVD DETQFKASGSVSQQDIYGEKIPK ESKIATFTRNVSWA (SEQ ID NO: 136)	GACAGCGTGGTGTTCGAGGACGTGGCCGTGGACTTCACCCTGGAGGAGTGGGC CCTGCTGGACAGCGCCCAGAGAGACCTGTACAGAGATGTGATGCTGGAGACCT TCCAGAACCTGGCCAGCGTGGACGACGAGACCCAGTTCAAGGCCAGCGGCAGC GTGAGCCAGCAGGACATCTACGGCGAGAAGATCCCCAAGGAGAGCAAGATCGC CACCTTCACCAGAAACGTGAGCTGGGCC (SEQ ID NO: 963)	7.02680391
ZN658	SVSFQDVTVEFTR EEWQHLGPVERT LYRDVMLENYSH LISV (SEQ ID NO: 137)	NMSQASVSFQDVTVEFTREEW QHLGPVERTLYRDVMLENYSHL ISVGYCITKPKVISKLEKGEEPWS LEDEFLNQRYPGY (SEQ ID NO: 138)	AACATGAGCCAGGCCAGCGTGAGCTTCCAGGACGTGACCGTGGAGTTCACCAG AGAGGAGTGGCAGCACCTGGGCCCCGTGGAGAGAACCCTGTACAGAGATGTGA TGCTGGAGAACTACAGCCACCCTGATCAGCGTGGGCTACTGCATCACCAAGCCCA AGGTGATCAGCAAGCTGGAGAAGGGCGAGGAGCCCTGGAGCCTGGAGGACGA GTTCCTGAACCAGAGATACCCCGGCTAC (SEQ ID NO: 964)	7.01857786
ZN684	SVTFQDVAVDFT AEEWQLLDCAER TLYWDVMLENY RNLISV (SEQ ID NO: 139)	ISFQESVTFQDVAVDFTAEEWQ LLDCAERTLYWDVMLENYRNLI SVGCPITKTKVILKVEQGQEPW MVEGANPHESSPES (SEQ ID NO: 140)	ATCAGCTTCCAGGAGAGCGTGACCTTCCAGGACGTGGCCGTGGACTTCACCGCC GAGGAGTGGCAGCTGCTGGACTGCGCCGAGAGAACCCTGTACTGGGACGTGAT GCTGGAGAACTACAGAAACCTGATCAGCGTGGGCTGCCCCATCACCAAGACCAA GGTGATCCTGAAGGTGGAGCAGGGCCAGGAGCCCTGGATGGTGGAGGGCGCC AACCCCCACGAGAGCAGCCCCGAGAGC (SEQ ID NO: 965)	7.01522838
RBAK	PVSFKDVAVDFT QEEWQQLDPDE KITYRDVMLENYS HLVSV (SEQ ID NO: 141)	NTLQGPVSFKDVAVDFTQEEW QQLDPDEKITYRDVMLENYSHL VSVGYDTTKPNVIIKLEQGEEP WIMGGEFPCQHSPEA (SEQ ID NO: 142)	AACACCCTGCAGGGCCCCGTGAGCTTCAAGGACGTGGCCGTGGACTTCACCCAG GAGGAGTGGCAGCAGCTGGACCCCGACGAGAAGATCACCTACAGAGATGTGAT GCTGGAGAACTACAGCCACCCTGGTGAGCGTGGGCTACGACACCACCAAGCCCA ACGTGATCATCAAGCTGGAGCAGGGCGAGGAGCCCCTGGATCATGGGCGGCGA GTTCCCCTGCCAGCACAGCCCCGAGGCC (SEQ ID NO: 966)	7.01040328
ZN829	PVMFRDVSIDFS QEEWECLDADQ MNLYKEVMLENF SNLVSV (SEQ ID NO: 143)	HPEEEERMHDELLQAVSKGPV MFRDVSIDFSQEEWECLDADQ MNLYKEVMLENFSNLVSVGLS NSKPAVISLLEQGKEPW (SEQ ID NO: 144)	CACCCCGAGGAGGAGGAGAGAATGCACGACGAGCTGCTGCAGGCCGTGAGCA AGGGCCCCGTGATGTTCAGAGATGTGAGCATCGACTTCAGCCAGGAGGAGTGG GAGTGCCTGGACGCCGACCAGATGAACCTGTACAAGGAGGTGATGCTGGAGAA CTTCAGCAACCTGGTGAGCGTGGGCCTGAGCAACAGCAAGCCCGCCGTGATCA GCCTGCTGGAGCAGGGCAAGGAGCCCTGG (SEQ ID NO: 967)	7.0012394

ZN582	LFRDVAIVFSQEE WQWLAPAQRDL YRDVMLETYSNL VSL (SEQ ID NO: 145)	SLGSELFRDVAIVFSQEEWQWL APAQRDLYRDVMLETYSNLVSL GLAVSKPDVISFLEQGKEPWM VERVVSGGLCPVLES (SEQ ID NO: 146)	AGCCTGGGCAGCGAGCTGTTCAGAGATGTGGCCATCGTGTTCAGCCAGGAGGA GTGGCAGTGGCTGGCCCCCGCCCAGAGACCTGTACAGAGATGTGATGCTGG AGACCTACAGCAACCTGGTGAGCCTGGGCCGTGAGCAAGCCCGACGTG ATCAGCTTCCTGGAGCAGGGCAAGGAGCCCTGGATGGTGGAGAGAGTGGTGA GCGGCGGCCTGTGCCCCGTGCTGGAGAGC (SEQ ID NO: 968)	6.98988925
ZN112	MVTFKDVAVVFT EEELGLLDSVQRK LYRDVMLENFRN LLL (SEQ ID NO: 147)	TKFQEMVTFKDVAVVFTEEELG LLDSVQRKLYRDVMLENFRNLL LVAHQPFKPDLISQLEREEKLLM VETETPRDGCSGR (SEQ ID NO: 148)	ACCAAGTTCCAGGAGATGGTGACCTTCAAGGACGTGGCCGTGGTGTTCACCGA GGAGGAGCTGGGCCTGCTGGACAGCGTGCAGAGAAAGCTGTACAGAGATGTG ATGCTGGAGAACTTCAGAAACCTGCTGCTGGTGGCCCACCAGCCCTTCAAGCCC GACCTGATCAGCCAGCTGGAGAGAGGAGAAGCTGCTGATGGTGGAGACCG AGACCCCCAGAGATGGCTGCAGCGGCAGA (SEQ ID NO: 969)	6.98982538
ZN716	LLTFRDIAIEFSLA EWQCLDHAQQN LYRDVMLENYRN LVSL (SEQ ID NO: 149)	AKRPGPPGSREMGLLTFRDIAIE FSLAEWQCLDHAQQNLYRDV MLENYRNLVSLGIAVSKPDLITC LEQNKEPQNIKRNE (SEQ ID NO: 150)	GCCAAGAGACCCGGCCCCCCCGGCAGCAGAGAGATGGGCCTGCTGACCTTCAG AGACATCGCCATCGAGTTCAGCCTGGCCGAGTGGCAGTGCCTGGACCACGCCCA GCAGAACCTGTACAGAGATGTGATGCTGGAGAACTACAGAAACCTGGTGAGCC TGGGCATCGCCGTGAGCAAGCCCGACCTGATCACCTGCCTGGAGCAGAACAAG GAGCCCCAGAACATCAAGAGAAACGAG (SEQ ID NO: 970)	6.98744382
HKR1	FVAFRDVAVYFT QEEWRLLSPAQR TLHREVMLETYN HLVSL (SEQ ID NO: 151)	TCMVHRQTMSCSGAGGITAFV AFRDVAVYFTQEEWRLLSPAQR TLHREVMLETYNHLVSLEIPSSK PKLIAQLERGEAPW (SEQ ID NO: 152)	ACCTGCATGGTGCACAGACAGACCATGAGCTGCAGCGGCGCCGGCGGCATCAC CGCCTTCGTGGCCTTCAGAGATGTGGCCGTGTACTTCACCCAGGAGGAGTGGAG ACTGCTGAGCCCCGCCCAGAGAACCCTGCACAGAGAGGTGATGCTGGAGACCT ACAACCACCTGGTGAGCCTGGAGATCCCCAGCAGCAAGCCCAAGCTGATCGCCC AGCTGGAGAGAGGCGAGGCCCCCTGG (SEQ ID NO: 971)	6.98664414
ZN350	SITLEDVAVDFT WEEWQLLGAAQ KDLYRDVMLENY SNLVAV (SEQ ID NO: 153)	IQAQESITLEDVAVDFTWEEW QLLGAAQKDLYRDVMLENYSN LVAVGYQASKPDAIFKLEQGEQ LWTIEDGIHSGACSDI (SEQ ID NO: 154)	ATCCAGGCCCAGGAGAGCATCACCCTGGAGGACGTGGCCGTGGACTTCACCTG GGAGGAGTGGCAGCTGCTGGGCGCCGCCCAGAAGGACCTGTACAGAGATGTG ATGCTGGAGAACTACAGCAACCTGGTGGCCGTGGGCTACCAGGCCAGCAAGCC CGACGCCCTGTTCAAGCTGGAGCAGGGCGAGCAGCTGTGGACCATCGAGGACG GCATCCACAGCGGCGCCTGCAGCGACATC (SEQ ID NO: 972)	6.98636848
ZN480	HLTFRDVAIEFSQ AEWKCLDPAQR ALYKDVMLENYR NLVSL (SEQ ID NO: 155)	AQKRRKRKAKESGMALPQGHL TFRDVAIEFSQAEWKCLDPAQR ALYKDVMLENYRNLVSLGISLPD LNINSMLEQRREPW (SEQ ID NO: 156)	GCCCAGAAGAGAAGAAAGAGAAAGGCCAAGGAGAGCGGCATGGCCCTGCCCC AGGGCCACCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCAGGCCGAGTGG AAGTGCCTGGACCCCGCCCAGAGAGCCCTGTACAAGGACGTGATGCTGGAGAA CTACAGAAACCTGGTGAGCCTGGGCATCAGCCTGCCCGACCTGAACATCAACAG CATGCTGGAGCAGAGAAGAGAGCCCTGG (SEQ ID NO: 973)	6.98462693

ZN416	CVTFEDVAIYFSQ EEWGLLDEAQRL LYRDVMLENFALI TA (SEQ ID NO: 157)	DSTSVPVTAEAKLMGFTQGCVT FEDVAIYFSQEEWGLLDEAQRL LYRDVMLENFALITALVCWHG MEDEETPEQSVSVEG (SEQ ID NO: 158)	GACAGCACCAGCGTGCCCGTGACCGCCGAGGCCAAGCTGATGGGCTTCACCCA GGGCTGCGTGACCTTCGAGGACGTGGCCATCTACTTCAGCCAGGAGGAGTGGG GCCTGCTGGACGAGGCCCAGAGACTGCTGTACAGAGATGTGATGCTGGAGAAC TTCGCCCTGATCACCGCCCTGGTGTGCTGGCACGGCATGGAGGACGAGGAGAC CCCCGAGCAGAGCGTGAGCGTGGAGGGC (SEQ ID NO: 974)	6.97472813
ZNF92	PLTFRDVKIEFSLE EWQCLDTAQRN LYRDVMLENYRN LVFL (SEQ ID NO: 159)	GPLTFRDVKIEFSLEEWQCLDTA QRNLYRDVMLENYRNLVFLGIA VSKPDLITWLEQGKEPWNLKR HEMVDKTPVMCSHF (SEQ ID NO: 160)	GGCCCCCTGACCTTCAGAGATGTGAAGATCGAGTTCAGCCTGGAGGAGTGGCA GTGCCTGGACACCGCCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCGCCGTGAGCAAGCCCGACCTGATCACCT GGCTGGAGCAGGGCAAGGAGCCCTGGAACCTGAAGAGACACGAGATGGTGGA CAAGACCCCCGTGATGTGCAGCCACTTC (SEQ ID NO: 975)	6.97138149
ZN100	PLTFRDVAIEFSLE EWQCLDSAQQG LYRKVMLENYRN LVF (SEQ ID NO: 161)	SGCPGAERSLLVQSYFEKGPLTF RDVAIEFSLEEWQCLDSAQQGL YRKVMLENYRNLVFLAGIALTKP DLITCLEQGKEP (SEQ ID NO: 162)	AGCGGCTGCCCCGGCGCCGAGAGAAGCCTGCTGGTGCAGAGCTACTTCGAGAA GGGCCCCCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGC AGTGCCTGGACAGCGCCCAGCAGGGCCTGTACAGAAAGGTGATGCTGGAGAAC TACAGAAACCTGGTGTTCCTGGCCGGCATCGCCCTGACCAAGCCCCGACCTGATC ACCTGCCTGGAGCAGGGCAAGGAGCCC (SEQ ID NO: 976)	6.9692141
ZN736	VLTFRDVAVEFSP EEWECLDSAQQR LYRDVMLENYGN LVSL (SEQ ID NO: 163)	GVLTFRDVAVEFSPEEWECLDS AQQRLYRDVMLENYGNLVSLG LAIFKPDLMTCLEQRKEPWKVK RQEAVAKHPAGSFHF (SEQ ID NO: 164)	GGCGTGCTGACCTTCAGAGATGTGGCCGTGGAGTTCAGCCCCGAGGAGTGGGA GTGCCTGGACAGCGCCCAGCAGAGACTGTACAGAGATGTGATGCTGGAGAACT ACGGCAACCTGGTGAGCCTGGGCCTGGCCATCTTCAAGCCCGACCTGATGACCT GCCTGGAGCAGAGAAAGGAGCCCTGGAAGGTGAAGAGACAGGAGGCCGTGGC CAAGCACCCCGCCGGCAGCTTCCACTTC (SEQ ID NO: 977)	6.95843452
ZNF74	SVSFKDVAVDFT QEEWGQLDSPQ RALYRDVMLENY QNLLAL (SEQ ID NO: 165)	KENLEDISGWGLPEARSKESVSF KDVAVDFTQEEWGQLDSPQRA LYRDVMLENYQNLLALGPPLHK PDVISHLERGFEEPW (SEQ ID NO: 166)	AAGGAGAACCTGGAGGACATCAGCGGCTGGGGCCTGCCCGAGGCCAGAAGCA AGGAGAGCGTGAGCTTCAAGGACGTGGCCGTGGACTTCACCCAGGAGGAGTG GGGCCAGCTGGACAGCCCCCCAGAGAGCCCTGTACAGAGATGTGATGCTGGAGA ACTACCAGAACCTGCTGGCCCTGGGCCCCCCCCTGCACAAGCCCGACGTGATCA GCCACCTGGAGAGAGGCGAGGAGCCCTGG (SEQ ID NO: 978)	6.95809395
CBX1	EPERIIGATDSSG ELMFLMKWKNS DEADLVPAKEAN VKCPQVVISFYEE RLT (SEQ ID NO: 167)	EESEKPRGFARGLEPERIIGATDS SGELMFLMKWKNSDEADLVPA KEANVKCPQVVISFYEERLTWH SYPSEDDDKKDDK (SEQ ID NO: 168)	GAGGAGAGCGAGAAGCCCAGAGGCTTCGCCAGAGGCCTGGAGCCCGAGAGAA TCATCGGCGCCACCGACAGCAGCGGCGAGCTGATGTTCCTGATGAAGTGGAAG AACAGCGACGAGGCCGACCTGGTGCCCGCCAAGGAGGCCAACGTGAAGTGCCC CCAGGTGGTGATCAGCTTCTACGAGGAGAGACTGACCTGGCACAGCTACCCCAG CGAGGACGACGACAAGAAGGACGACAAG (SEQ ID NO: 979)	6.95269512

ZN443	SVALEDVAVNFT REEWALLGPCQK NLYKDVMQETIR NLDC (SEQ ID NO: 169)	ASVALEDVAVNFTREEWALLGP CQKNLYKDVMQETIRNLDCVV MKWKDQNIEDQYRYPRKNLRC RMLERFVESKDGTQCG (SEQ ID NO: 170)	GCCAGCGTGGCCCTGGAGGACGTGGCCGTGAACTTCACCAGAGAGGAGTGGGC CCTGCTGGGCCCCTGCCAGAAGAACCTGTACAAGGACGTGATGCAGGAGACCA TCAGAAACCTGGACTGCGTGGTGATGAAGTGGAAGGACCAGAACATCGAGGAC CAGTACAGATACCCCAGAAAGAACCTGAGATGCAGAATGCTGGAGAGATTCGT GGAGAGCAAGGACGGCACCCAGTGCGGC (SEQ ID NO: 980)	6.94561303
ZN195	LLTFRDVAIEFSLE EWKCLDLAQQN LYRDVMLENYRN LFSV (SEQ ID NO: 171)	TLLTFRDVAIEFSLEEWKCLDLA QQNLYRDVMLENYRNLFSVGL TVCKPGLITCLEQRKEPWNVKR QEAADGHPEMGFHH (SEQ ID NO: 172)	ACCCTGCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGAA GTGCCTGGACCTGGCCCAGCAGAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGTTCAGCGTGGGCCTGACCGTGTGCAAGCCCGGCCTGATCACCT GCCTGGAGCAGAGAAAGGAGCCCTGGAACGTGAAGAGACAGGAGGCCGCCGA CGGCCACCCCGAGATGGGCTTCCACCAC (SEQ ID NO: 981)	6.9432522
ZN530	FVAFEDVAIYFSQ EEWELLDEMQRL LYRDVMLENFAV MASL (SEQ ID NO: 173)	AAALRAPTQQVFVAFEDVAIYF SQEEWELLDEMQRLLYRDVML ENFAVMASLGCWCGAVDEGT PSAESVSVEELSQGRTP (SEQ ID NO: 174)	GCCGCCGCCCTGAGAGCCCCCACCCAGCAGGTGTTCGTGGCCTTCGAGGACGTG GCCATCTACTTCAGCCAGGAGGAGTGGGAGCTGCTGGACGAGATGCAGAGACT GCTGTACAGAGATGTGATGCTGGAGAACTTCGCCGTGATGGCCAGCCTGGGCT GCTGGTGCGGCCGCCGTGGACGAGGGAACACCTAGCGCCGAGAGCGTGAGCGT GGAGGAGCTGAGCCAGGGCAGAACCCCC (SEQ ID NO: 982)	6.94292737
ZN782	SVSFQDVTVEFS QEEWQHMGPV ERTLYRDVMLEN YSHLVSV (SEQ ID NO: 175)	NTFQASVSFQDVTVEFSQEEW QHMGPVERTLYRDVMLENYSH LVSVGYCFTKPELIFTLEQGEDP WLLEKEKGFLSRNSP (SEQ ID NO: 176)	AACACCTTCCAGGCCAGCGTGAGCTTCCAGGACGTGACCGTGGAGTTCAGCCAG GAGGAGTGGCAGCACATGGGCCCCGTGGAGAGAACCCTGTACAGAGATGTGAT GCTGGAGAACTACAGCCACCTGGTGAGCGTGGGCTACTGCTTCACCAAGCCCGA GCTGATCTTCACCCTGGAGCAGGGCGAGGACCCCCTGGCTGCTGGAGAAGGAGA AGGGCTTCCTGAGCAGAAACAGCCCC (SEQ ID NO: 983)	6.94217051
ZN791	SVAFEDVSVFSQ EEWALLAPSQKK LYRDVMQETFKN LASI (SEQ ID NO: 177)	DSVAFEDVSVFSQEEWALLAP SQKKLYRDVMQETFKNLASIGE KWEDPNVEDQHKNQGRNLRS HTGERLCEGKEGSQCA (SEQ ID NO: 178)	GACAGCGTGGCCTTCGAGGACGTGAGCGTGAGCTTCAGCCAGGAGGAGTGGG CCCTGCTGGCCCCCAGCCAGAAGAAGCTGTACAGAGATGTGATGCAGGAGACC TTCAAGAACCTGGCCCAGCATCGGCGAGAAGTGGGAGGACCCCAACGTGGAGGA CCAGCACAAGAACCAGGGCAGAAACCTGAGAAGCCACACCGGCGAGAGACTGT GCGAGGGCAAGGAGGGCAGCCAGTGCGCC (SEQ ID NO: 984)	6.93320479
ZN331	LVTFADVAIDFSQ EEWACLNSAQR DLYWDVMLENY SNLVSL (SEQ ID NO: 179)	AQGLVTFADVAIDFSQEEWACL NSAQRDLYWDVMLENYSNLVS LDLESAYENKSLPTEKNIHEIRAS KRNSDRRSKSLGR (SEQ ID NO: 180)	GCCCAGGGCCTGGTGACCTTCGCCGACGTGGCCATCGACTTCAGCCAGGAGGA GTGGGCCTGCCTGAACAGCGCCCAGAGACCTGTACTGGGACGTGATGCTGG AGAACTACAGCAACCTGGTGAGCCTGGACCTGGAGAGCGCCTACGAGAACAAG AGCCTGCCCACCGAGAAGAACATCCACGAGATCAGAGCCAGCAAGAAACAG CGACAGAAGAAGCAAGAGCCTGGGCAGA (SEQ ID NO: 985)	6.92979428

Z354C	PVTFRDVAVFFS QDEWLHLDSAQ RALYREVMLENY SSLVSL (SEQ ID NO: 181)	AVDLLSAQEPVTFRDVAVFFSQ DEWLHLDSAQRALYREVMLEN YSSLVSLGIPFSMPKLIHQLQQG EDPCMVEREVPSDT (SEQ ID NO: 182)	GCCGTGGACCTGCTGAGCGCCCAGGAGCCCGTGACCTTCAGAGATGTGGCCGT GTTCTTCAGCCAGGACGAGTGGCTGCACCTGGACAGCGCCCAGAGAGCCCTGTA CAGAGAGGTGATGCTGGAGAACTACAGCAGCCTGGTGAGCCTGGGCATCCCCT TCAGCATGCCCAAGCTGATCCACCAGCTGCAGCAGGGCGAGGACCCCTGCATG GTGGAGAGAGAGGTGCCCAGCGACACC (SEQ ID NO: 986)	6.92855271
ZN157	SVSFEDVAVDFTR QEWHRLDPAQR TMHKDVMLETYS NLASV (SEQ ID NO: 183)	SPQRFPALIPGEPGRSFEGSVSF EDVAVDFTRQEWHRLDPAQRT MHKDVMLETYSNLASVGLCVA KPEMIFKLERGEELW (SEQ ID NO: 184)	AGCCCCCAGAGATTCCCCGCCCTGATCCCCGGCGAGCCCGGCAGAAGCTTCGAG GGCAGCGTGAGCTTCGAGGACGTGGCCGTGGACTTCACCAGACAGGAGTGGCA CAGACTGGACCCCGCCCAGAGAACCATGCACACAAGGACGTGATGCTGGAGACCT ACAGCAACCTGGCCAGCGTGGGCCTGTGCGTGGCCAAGCCCGAGATGATCTTCA AGCTGGAGAGAGGCGAGGAGCTGTGG (SEQ ID NO: 987)	6.92764017
ZN727	VLTFRDVAVEFSP EEWECLDSAQQR LYRDVMLENYGN LFSL (SEQ ID NO: 185)	RVLTFRDVAVEFSPEEWECLDS AQQRLYRDVMLENYGNLFSLG LAIFKPDLITYLEQRKEPWNARR QKTVAKHPAGSLHF (SEQ ID NO: 186)	AGAGTGCTGACCTTCAGAGATGTGGCCGTGGAGTTCAGCCCCGAGGAGTGGGA GTGCCTGGACAGCGCCCAGCAGAGACTGTACAGAGATGTGATGCTGGAGAACT ACGGCAACCTGTTCAGCCTGGGCCTGGCCATCTTCAAGCCCGACCTGATCACCTA CCTGGAGCAGAGAAAGGAGCCCTGGAACGCCAGAAGACAGAAGACCGTGGCC AAGCACCCCGCCGGCAGCCTGCACTTC (SEQ ID NO: 988)	6.9257026
ZN550	LVTFKDVAVTFTR EEWRQLDLAQRT LYREVMLETCGLL VSL (SEQ ID NO: 187)	AETKDAAQMLVTFKDVAVTFT REEWRQLDLAQRTLYREVMLE TCGLLVSLGHRVPKPELVHLLEH GQELWIVKRGLSHAT (SEQ ID NO: 188)	GCCGAGACCAAGGACGCCGCCCAGATGCTGGTGACCTTCAAGGACGTGGCCGT GACCTTCACCAGAGAGGAGTGGAGACAGCTGGACCTGGCCCAGAGAACCCTGT ACAGAGAGGTGATGCTGGAGACCTGCGGCCTGCTGGTGAGCCTGGGCCACAGA GTGCCCAAGCCCGAGCTGGTGCACCTGCTGGAGCACGGCCAGGAGCTGTGGAT CGTGAAGAGAGGCCTGAGCCACGCCACC (SEQ ID NO: 989)	6.92403295
ZN793	PVSFKDVVVGFT QEEWHRLSPAQ RALYRDVMLETY SNLVSV (SEQ ID NO: 189)	IEYQIPVSFKDVVVGFTQEEWH RLSPAQRALYRDVMLETYSNLV SVGYEGTKPDVILRLEQEEAPWI GEAACPGCHCWED (SEQ ID NO: 190)	ATCGAGTACCAGATCCCCGTGAGCTTCAAGGACGTGGTGGTGGGCTTCACCCAG GAGGAGTGGCACAGACTGAGCCCCGCCCAGAGAGCCCTGTACAGAGATGTGAT GCTGGAGACCTACAGCAACCTGGTGAGCGTGGGCTACGAGGGCACCAAGCCCG ACGTGATCCTGAGACTGGAGCAGGAGGAGGCCCCCTGGATCGGCGAGGCCGCT TGCCCTGGATGCCACTGCTGGGAGGAC (SEQ ID NO: 990)	6.92326085
ZN235	AVTFKDVAVAFT EEELGLLDSAQRK LYRDVMLENFRN LVSV (SEQ ID NO: 191)	TKFQEAVTFKDVAVAFTEEELGL LDSAQRKLYRDVMLENFRNLVS VGHQSFKPDMISQLEREEKLW MKELQTQRGKHSGD (SEQ ID NO: 192)	ACCAAGTTCCAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGGCCTTCACCGAG GAGGAGCTGGGCCTGCTGGACAGCGCCCAGAGAAAGCTGTACAGAGATGTGAT GCTGGAGAACTTCAGAAACCTGGTGAGCGTGGGCCACCAGAGCTTCAAGCCCG ACATGATCAGCCAGCTGGAGAGAGAGGAGAAGCTGTGGATGAAGGAGCTGCA GACCCAGAGAGGCAAGCACAGCGGCGAC (SEQ ID NO: 991)	6.91826902

ZNF8	PVTFRDVAVDFT QEEWGQLDPTQ RILYRDVMLETFG HLLSI (SEQ ID NO: 193)	DEGVAGVMSVGPPAARLQEPV TFRDVAVDFTQEEWGQLDPTQ RILYRDVMLETFGHLLSIGPELPK PEVISQLEQGTELW (SEQ ID NO: 194)	GACGAGGGCGTGGCCGGCGTGATGAGCGTGGGCCCCCCCGCTGCTAGACTTCA GGAGCCCGTGACCTTCAGAGATGTGGCCGTGGACTTCACCCAGGAGGAGTGGG GCCAGCTGGACCCCACCCAGAGAATCCTGTACAGAGATGTGATGCTGGAGACCT TCGGCCACCTGCTGAGCATCGGCCCCGAGCTGCCCAAGCCCGAGGTGATCAGCC AGCTGGAGCAGGGCACCGAGCTGTGG(SEQ ID NO: 992)	6.91722882
ZN724	PLTFMDVAIEFSV EEWQCLDTAQQ NLYRNVMLENYR NLVFL (SEQ ID NO: 195)	GPLTFMDVAIEFSVEEWQCLDT AQQNLYRNVMLENYRNLVFLGI AVSKPDLITCLEQGKEPWNME RHEMVAKPPGMCCYF (SEQ ID NO: 196)	GGCCCCCTGACCTTCATGGACGTGGCCATCGAGTTCAGCGTGGAGGAGTGGCA GTGCCTGGACACCGCCCAGCAGAACCTGTACAGAAACGTGATGCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCGCCGTGAGCAAGCCCGACCTGATCACCT GCCTGGAGCAGGGCAAGGAGCCCTGGAACATGGAGAGACACGAGATGGTGGC CAAGCCCCCCGGCATGTGCTGCTACTTC (SEQ ID NO: 993)	6.89904065
ZN573	LVTFRDVAIDFSR QEWEYLDPNQR DLYRDVMLENYR NLVSL (SEQ ID NO: 197)	HQVGLIRSYNSKTMTCFQELVT FRDVAIDFSRQEWEYLDPNQR DLYRDVMLENYRNLVSLGGHSI SKPVVVDLLERGKEP (SEQ ID NO: 198)	CACCAGGTGGGCCTGATCAGAAGCTACAACAGCAAGACCATGACCTGCTTCCAG GAGCTGGTGACCTTCAGAGATGTGGCCATCGACTTCAGCAGACAGGAGTGGGA GTACCTGGACCCCAACCAGAGAGACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGTGAGCCTGGGCGGCCACAGCATCAGCAAGCCCGTGGTGGTG GACCTGCTGGAGAGAGGCAAGGAGCCC (SEQ ID NO: 994)	6.89366942
ZN577	SLSFEDVAVGFTR EEWQFLDQSQK VLYKEVMLENYIN LVSI (SEQ ID NO: 199)	NATIVMSVRREQGSSSGEGSLS FEDVAVGFTREEWQFLDQSQK VLYKEVMLENYINLVSIGYRGTK PDSLFKLEQGEPPG (SEQ ID NO: 200)	AACGCCACCATCGTGATGAGCGTGAGAAGAGAGCAGGGCAGCAGCAGCGGCG AGGGCAGCCTGAGCTTCGAGGACGTGGCCGTGGGCTTCACCAGAGAGGAGTG GCAGTTCCTGGACCAGAGCCAGAAGGTGCTGTGTACAGAGGCACCAAGCCCGACAGCCTGT ACTACATCAACCTGGTGAGCATCGGCTACAGAGGCACCAAGCCCGACAGCCTGT TCAAGCTGGAGCAGGGCGAGCCCCCCGGC (SEQ ID NO: 995)	6.89093009
ZN789	LLSFEDVAMYFTR EEWGHLNWGQ KDLYRDVMLENY RNMVLL (SEQ ID NO: 201)	FPPARGKELLSFEDVAMYFTREE WGHLNWGQKDLYRDVMLENY RNMVLLGFQFPKPEMICQLEN WDEQWILDLPRTGNRK (SEQ ID NO: 202)	TTCCCCCCCGCCAGAGGCAAGGAGCTGCTGAGCTTCGAGGACGTGGCCATGTAC TTCACCAGAGAGGAGGAGTGGGGCCACCTGAACTGGGGCCAGAAGGACCTGTACAG AGATGTGATGCTGGAGAACTACAGAAACATGGTGCTGCTGGGCTTCCAGTTCCC CAAGCCCGAGATGATCTGCCAGCTGGAGAACTGGGACGAGCAGTGGATCCTGG ACCTGCCCAGAACCGGCAACAGAAAG (SEQ ID NO: 996)	6.88877268
ZN718	LLTFKDVAIEFSPE EWKCLDTSQQNL YRDVMLENYRNL VSL (SEQ ID NO: 203)	ELLTFKDVAIEFSPEEWKCLDTS QQNLYRDVMLENYRNLVSLGV SISNPDLVTSLEQRKEPYNLKIHE TAARPPAVCSHF (SEQ ID NO: 204)	GAGCTGCTGACCTTCAAGGACGTGGCCATCGAGTTCAGCCCCGAGGAGTGGAA GTGCCTGGACACCAGCCAGAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGTGAGCCTGGGCGTGAGCATCAGCAACCCCGACCTGGTGACC AGCCTGGAGCAGAGAAAGGAGCCCTACAACCTGAAGATCCACGAGACCGCCGC CAGACCCCCCGCCGTGTGCAGCCACTTC (SEQ ID NO: 997)	6.87598723

ZN300	LVSFKDVAVDFT QEEWQQLDPSQ RTLYRDVMLENY SHLVSM (SEQ ID NO: 205)	MKSQGLVSFKDVAVDFTQEEW QQLDPSQRTLYRDVMLENYSH LVSMGYPVSKPDVISKLEQGEE PWIIKGDISNWIYPDE (SEQ ID NO: 206)	ATGAAGAGCCAGGGCCTGGTGAGCTTCAAGGACGTGGCCGTGGACTTCACCCA GGAGGAGTGGCAGCAGCTGGACCCCAGCCAGAGAACCCTGTACAGAGATGTGA TGCTGGAGAACTACAGCCACCCTGGTGAGCATGGGCTACCCCGTGAGCAAGCCC GACGTGATCAGCAAGCTGGAGCAGGGCGAGGAGCCCTGGATCATCAAGGGCG ACATCAGCAACTGGATCTACCCCGACGAG (SEQ ID NO: 998)	6.87019452
ZN383	SVMFSDVSIDFS QEEWDCLDPVQ RDLYRDVMLENY GNLVSM (SEQ ID NO: 207)	AEGSVMFSDVSIDFSQEEWDCL DPVQRDLYRDVMLENYGNLVS MGLYTPKPQVISLLEQGKEPW MVGRELTRGLCSDLES (SEQ ID NO: 208)	GCCGAGGGCAGCGTGATGTTCAGCGACGTGAGCATCGACTTCAGCCAGGAGGA GTGGGACTGCCTGGACCCCGTGCAGAGAGACCTGTACAGAGATGTGATGCTGG AGAACTACGGCAACCTGGTGAGCATGGGCCTGTACACCCCCAAGCCCCAGGTG ATCAGCCTGCTGGAGCAGGGCAAGGAGCCCTGGATGGTGGGCAGAGAGCTGA CCAGAGGCCTGTGCAGCGACCTGGAGAGC (SEQ ID NO: 999)	6.86203801
ZN429	PLTFTDVAIEFSLE EWQCLDTAQQN LYRNVMLENYRN LVFL (SEQ ID NO: 209)	GPLTFTDVAIEFSLEEWQCLDTA QQNLYRNVMLENYRNLVFLGIA VSKPDLITCLEKEKEPCKMKRHE MVDEPPVVCSHF (SEQ ID NO: 210)	GGCCCCCTGACCTTCACCGACGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCAG TGCCTGGACACCGCCCAGCAGAACCTGTACAGAAACGTGATGCTGGAGAACTAC AGAAACCTGGTGTTCCTGGGCATCGCCGTGAGCAAGCCCGACCTGATCACCTGC CTGGAGAAGGAGAAGGAGCCCTGCAAGATGAAGAGACACGAGATGGTGGACG AGCCCCCCGTGGTGTGCAGCCACTTC (SEQ ID NO: 1000)	6.85768103
ZN677	LFTFKDVAIEFSQ EEWECLDPAQRA LYRDVMLENYRN LLSL (SEQ ID NO: 211)	ALSQGLFTFKDVAIEFSQEEWE CLDPAQRALYRDVMLENYRNLL SLDEDNIPPEDDISVGFTSKGLS PKENNKEELYHLV (SEQ ID NO: 212)	GCCCTGAGCCAGGGCCTGTTCACCTTCAAGGACGTGGCCATCGAGTTCAGCCAG GAGGAGTGGGAGTGCCTGGACCCCGCCCAGAGAGCCCTGTACAGAGATGTGAT GCTGGAGAACTACAGAAACCTGCTGAGCCTGGACGAGGACAACATCCCCCCCG AGGACGACATCAGCGTGGGCTTCACCAGCAAGGGCCTGAGCCCCAAGGAGAAC AACAAGGAGAGCTGTACCACCTGGTG (SEQ ID NO: 1001)	6.85440091
ZN850	LVMFQDLSIDFS QEEWECLDAAQ KDLYRDVMMEN YSSLVSL (SEQ ID NO: 213)	NMEGLVMFQDLSIDFSQEEWE CLDAAQKDLYRDVMMENYSSL VSLGLSIPKPDVISLLEQGKEPW MVSRDVLGGWCRDSE (SEQ ID NO: 214)	AACATGGAGGGCCTGGTGATGTTCCAGGACCTGAGCATCGACTTCAGCCAGGA GGAGTGGGAGTGCCTGGACGCCGCCCAGAAGGACCTGTACAGAGATGTGATGA TGGAGAACTACAGCAGCCTGGTGAGCCTGGGCCTGAGCATCCCCAAGCCCGAC GTGATCAGCCTGCTGGAGCAGGGCAAGGAGCCCTGGATGGTGAGCAGAGATGT GCTGGGCGGCTGGTGCAGAGACAGCGAG (SEQ ID NO: 1002)	6.85293565
ZN454	SVTFKDVAILFTQ EEWGQLSPAQR ALYRDVMLENYS NLVSL (SEQ ID NO: 215)	AVSHLPTMVQESVTFKDVAILF TQEEWGQLSPAQRALYRDVML ENYSNLVSLGLLGPKPDTFSQLE KREVWMPEDTPGGF (SEQ ID NO: 216)	GCCGTGAGCCACCTGCCCACCATGGTGCAGGAGAGCGTGACCTTCAAGGACGT GGCCATCCTGTTCACCCAGGAGGAGTGGGGCCAGCTGAGCCCCGCCCAGAGAG CCCTGTACAGAGATGTGATGCTGGAGAACTACAGCAACCTGGTGAGCCTGGGC CTGCTGGGCCCCAAGCCCGACACCTTCAGCCAGCTGGAGAAGAGAGAGGTGTG GATGCCCGAGGACACCCCCGGCGGCTTC (SEQ ID NO: 1003)	6.8342036

ZN257	PLTIRDVTVEFSLE EWHCLDTAQQN LYRDVMLENYRN LVFL (SEQ ID NO: 217)	GPLTIRDVTVEFSLEEWHCLDTA QQNLYRDVMLENYRNLVFLGIA VSKPDLITCLEQGKEPCNMKRH EMVAKPPVMCSHI (SEQ ID NO: 218)	GGCCCCCTGACCATCAGAGATGTGACCGTGGAGTTCAGCCTGGAGGAGTGGCA CTGCCTGGACACCGCCCAGCAGAACCTGTACAGAGATGTGATGCTGGAGAACTA CAGAAACCTGGTGTTCCTGGGCATCGCCGTGAGCAAGCCCGACCTGATCACCTG CCTGGAGCAGGGCAAGGAGCCCTGCAACATGAAGAGACACGAGATGGTGGCC AAGCCCCCCGTGATGTGCAGCCACATC (SEQ ID NO: 1004)	6.83044
ZN264	SVTFDDVAVTFTK EEWGQLDLAQR TLYQEVMLENCG LLVSL (SEQ ID NO: 219)	AAAVLTDRAQVSVTFDDVAVTF TKEEWGQLDLAQRTLYQEVML ENCGLLVSLGCPVPKAELICHLE HGQEPWTRKEDLSQ (SEQ ID NO: 220)	GCCGCCGCCGTGCTGACCGACAGAGCCCAGGTGAGCGTGACCTTCGACGACGT GGCCGTGACCTTCACCAAGGAGGAGTGGGGCCAGCTGGACCTGGCCCAGAGAA CCCTGTACCAGGAGGTGATGCTGGAGAACTGCGGCCTGCTGGTGAGCCTGGGC TGCCCCGTGCCCAAGGCCGAGCTGATCTGCCACCTGGAGCACGGCCAGGAGCC CTGGACCAGAAAGGAGGACCTGAGCCAG (SEQ ID NO: 1005)	6.82889596
ZFP82	SVMFSDVSIDFSP EEWEYLDLEQKD LYRDVMLENYSN LVSL (SEQ ID NO: 221)	ALRSVMFSDVSIDFSPEEWEYL DLEQKDLYRDVMLENYSNLVSL GCFISKPDVISSLEQGKEPWKVV RKGRRQYPDLETK (SEQ ID NO: 222)	GCCCTGAGAAGCGTGATGTTCAGCGACGTGAGCATCGACTTCAGCCCCGAGGA GTGGGAGTACCTGGACCTGGAGCAGAAGGACCTGTACAGAGATGTGATGCTGG AGAACTACAGCAACCTGGTGAGCCTGGGCTGCTTCATCAGCAAGCCCGACGTGA TCAGCAGCCTGGAGCAGGGCAAGGAGCCCCTGGAAGGTGGTGAGAAAGGGCAG AAGACAGTACCCCGACCTGGAGACCAAG (SEQ ID NO: 1006)	6.82733193
ZFP14	SVTFRDVAIDFSQ EEWEFLDPAQRD LYRDVMWENYS NFISL (SEQ ID NO: 223)	AHGSVTFRDVAIDFSQEEWEFL DPAQRDLYRDVMWENYSNFIS LGPSISKPDVITLLDEERKEPGM VVREGTRRYCPDLE (SEQ ID NO: 224)	GCCCACGGCAGCGTGACCTTCAGAGATGTGGCCATCGACTTCAGCCAGGAGGA GTGGGAGTTCCTGGACCCCGCCCAGAGAGACCTGTACAGAGATGTGATGTGGG AGAACTACAGCAACTTCATCAGCCTGGGCCCCAGCATCAGCAAGCCCGACGTGA TCACCCTGCTGGACGAGGAGAGAAAGGAGCCCGGCATGGTGGTGAGAGAGGG CACCAGAAGATACTGCCCCGACCTGGAG (SEQ ID NO: 1007)	6.81312035
ZN485	PLTFGDVAVAFT RIEWRHLDAAQR ALYRDVMLENYG NLVSV (SEQ ID NO: 225)	APRAQIQGPLTFGDVAVAFTRI EWRHLDAAQRALYRDVMLENY GNLVSVGLLSSKPKLITQLEQGA EPWTEVREAPSGTH (SEQ ID NO: 226)	GCCCCCAGAGCCCAGATCCAGGGCCCCCCTGACCTTCGGCGACGTGGCCGTGGCC TTCACCAGAATCGAGTGGAGACACCTGGACGCCGCCCAGAGAGCCCTGTACAG AGATGTGATGCTGGAGAACTACGGCAACCTGGTGAGCGTGGGCCTGCTGAGCA GCAAGCCCAAGCTGATCACCCAGCTGGAGCAGGGCGCCGAGCCCTGGACCGAG GTGAGAGAGGCCCCTAGCGGCACACAC (SEQ ID NO: 1008)	6.81172703
ZN737	PLQFRDVAIEFSL EEWHCLDTAQR NLYRNVMLENYR NLVFL (SEQ ID NO: 227)	GPLQFRDVAIEFSLEEWHCLDT AQRNLYRNVMLENYRNLVFLGI VVSKPDLITCLEQGKKPLTMKK HEMVANPSVTCSHF (SEQ ID NO: 228)	GGCCCCCTGCAGTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA CTGCCTGGACACCGCCCAGAGAAACCTGTACAGAAACGTGATGCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCGTGGTGAGCAAGCCCGACCTGATCACCT GCCTGGAGCAGGGCAAGAAGCCCCTGACCATGAAGAAGCACGAGATGGTGGC CAACCCCAGCGTGACCTGCAGCCACTTC (SEQ ID NO: 1009)	6.80882457

ZNF44	SVAFEDVAVNFT HEEWALLGPSQK NLYRDVMRETIR NLNCI (SEQ ID NO: 229)	TLPRGQPEVLEWGLPKDQDSV AFEDVAVNFTHEEWALLGPSQ KNLYRDVMRETIRNLNCIGMK WENQNIDDQHQNLRRNP (SEQ ID NO: 230)	ACCCTGCCCAGAGGCCAGCCCGAGGTGCTGGAGTGGGGCCTGCCCAAGGACCA GGACAGCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCACGAGGAGTGGG CCCTGCTGGGCCCCAGCCAGAAGAACCTGTACAGAGATGTGATGAGAGAGACC ATCAGAAACCTGAACTGCATCGGCATGAAGTGGGAGAACCAGAACATCGACGA CCAGCACCAGAACCTGAGAAGAAACCCC (SEQ ID NO: 1010)	6.80503304
ZN596	SMTFEDIIVDFTQ EEWALLDTSQRK LFQDVMLENISH LVSI (SEQ ID NO: 231)	PSPDSMTFEDIIVDFTQEEWALL DTSQRKLFQDVMLENISHLVSI GKQLCKSVVLSQLEQVEKLSTQ RISLLQGREVGIK (SEQ ID NO: 232)	CCCAGCCCCGACAGCATGACCTTCGAGGACATCATCGTGGACTTCACCCAGGAG GAGTGGGCCCTGCTGGACACCAGCCAGAGAAAGCTGTTCCAGGACGTGATGCT GGAGAACATCAGCCACCTGGTGAGCATCGGCAAGCAGCTGTGCAAGAGCGTGG TGCTGAGCCAGCTGGAGCAGGTGGAGAAGCTGAGCACCCAGAGAATCAGCCTG CTGCAGGGCAGAGAGGTGGGCATCAAG (SEQ ID NO: 1011)	6.80500309
ZN565	LVTFRDVAIEFSLE EWKCLEPAQRDL YREVTLENFGHLA SL (SEQ ID NO: 233)	EESREIRAGQIVLKAMAQGLVT FRDVAIEFSLEEWKCLEPAQRDL YREVTLENFGHLASLGLSISKPD VVSLLEQGKEPW (SEQ ID NO: 234)	GAGGAGAGCAGAGAGATCAGAGCCGGCCAGATCGTGCTGAAGGCCATGGCCC AGGGCCTGGTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGG AAGTGCCTGGAGCCCGCCCAGAGAGACCTGTACAGAGAGGTGACCCTGGAGAA CTTCGGCCACCTGGCCAGCCTGGGCCTGAGCATCAGCAAGCCCGACGTGGTGA GCCTGCTGGAGCAGGGCAAGGAGCCCTGG(SEQ ID NO: 1012)	6.80375161
ZN543	SVTFEDVAVTFT QEEWGQLDAAQ RTLYQEVMLETC GLLMSL (SEQ ID NO: 235)	AASAQVSVTFEDVAVTFTQEE WGQLDAAQRTLYQEVMLETC GLLMSLGCPLFKPELIYQLDHRQ ELWMATKDLSQSSYPG (SEQ ID NO: 236)	GCCGCCAGCGCCCAGGTGAGCGTGACCTTCGAGGACGTGGCCGTGACCTTCACC CAGGAGGAGTGGGGCCAGCTGGACGCCGCCCAGAGAACCCTGTACCAGGAGG TGATGCTGGAGACCTGCGGCCTGCTGATGAGCCTGGGCTGCCCCCTGTTCAAGC CCGAGCTGATCTACCAGCTGGACCACAGACAGGAGCTGTGGATGGCCACCAAG GACCTGAGCCAGAGCAGCTACCCCGGC (SEQ ID NO: 1013)	6.79786357
ZFP69	LLTFKDISIDFTQE EWGQLAPAHQN LYREVMLENYSN LVSV (SEQ ID NO: 237)	RESLEDEVTPGLPTAESQELLTF KDISIDFTQEEWGQLAPAHQNL YREVMLENYSNLVSVGYQLSKP SVISQLEKGEEPW (SEQ ID NO: 238)	AGAGAGAGCCTGGAGGACGAGGTGACCCCCGGCCTGCCCACCGCCGAGAGCCA AGAGCTGCTGACCTTCAAGGACATCAGCATCGACTTCACCCAGGAGGAGTGGG GCCAGCTGGCCCCCGCCCACCAGAACCTGTACAGAGAGGTGATGCTGGAGAAC TACAGCAACCTGGTGAGCGTGGGCTACCAGCTGAGCAAGCCCAGCGTGATCAG CCAGCTGGAGAAGGGCGAGGAGCCCTGG (SEQ ID NO: 1014)	6.79374304
SUMO1	IKLKVIGQDSSEIH FKVKMTTHLKKL KESYCQRQGVP MNSLRFLFEGQRI ADNHTPKELGME EEDVIEVY (SEQ ID NO: 239)	EGEYIKLKVIGQDSSEIHFKVKM TTHLKKLKESYCQRQGVPMNSL RFLFEGQRIADNHTPKELGMEE EDVIEVYQEQTGG (SEQ ID NO: 240)	GAGGGCGAGTACATCAAGCTGAAGGTGATCGGCCAGGACAGCAGCGAGATCC ACTTCAAGGTGAAGATGACCACCCACCTGAAGAAGCTGAAGGAGAGCTACTGC CAGAGACAGGGCGTGCCCATGAACAGCCTGAGATTCCTGTTCGAGGGCCAGAG AATCGCCGACAACCACACCCCCAAGGAGCTGGGCATGGGAGGAGGAGGACGTG ATCGAGGTGTACCAGGAGCAGACCGGCGGC (SEQ ID NO: 1015)	6.77750481

ZNF12	PVSFKDVAVDFT QEEWQQLDPEQ KITYRDVMLENYS NLVSV (SEQ ID NO: 241)	NKSLGPVSFKDVAVDFTQEEW QQLDPEQKITYRDVMLENYSNL VSVGYHIIKPDVISKLEQGEEPW IVEGEFLLQSYPDE (SEQ ID NO: 242)	AACAAGAGCCTGGGCCCCGTGAGCTTCAAGGACGTGGCCGTGGACTTCACCCA GGAGGAGTGGCAGCAGCTGGACCCCGAGCAGAAGATCACCTACAGAGATGTG ATGCTGGAGAACTACAGCAACCTGGTGAGCGTGGGCTACCACATCATCAAGCCC GACGTGATCAGCAAGCTGGAGCAGGGCGAGGAGCCCTGGATCGTGGAGGGCG AGTTCCTGCTGCAGAGCTACCCCGACGAG (SEQ ID NO: 1016)	6.77648818
ZN169	LMAFRDVAVAFT QKEWKLLSSAQR TLYREVMLENYS HLVSL (SEQ ID NO: 243)	SPGLLTTRKEALMAFRDVAVAF TQKEWKLLSSAQRTLYREVMLE NYSHLVSLGIAFSKPKLIEQLEQG DEPWREENEHLL (SEQ ID NO: 244)	AGCCCCGGCCTGCTGACCACCAGAAAGGAGGCCCTGATGGCCTTCAGAGATGT GGCCGTGGCCTTCACCCAGAAGGAGTGGAAGCTGCTGAGCAGCGCCCAGAGAA CCCTGTACAGAGAGGTGATGCTGGAGAACTACAGCCACCTGGTGAGCCTGGGC ATCGCCTTCAGCAAGCCCAAGCTGATCGAGCAGCTGGAGCAGGGCGACGAGCC CTGGAGAGAGGAGAACGAGCACCTGCTG (SEQ ID NO: 1017)	6.77498642
ZN433	SVAFEDVAVTFT QEEWALLDPSQK NLCRDVMQETFR NLASI (SEQ ID NO: 245)	MFQDSVAFEDVAVTFTQEEWA LLDPSQKNLCRDVMQETFRNL ASIGKKWKPQNIYVEYENLRRN LRIVGERLFESKEGHQ (SEQ ID NO: 246)	ATGTTCCAGGACAGCGTGGCCTTCGAGGACGTGGCCGTGACCTTCACCCAGGAG GAGTGGGCCCTGCTGGACCCCAGCCAGAAGAACCTGTGCAGAGATGTGATGCA GGAGACCTTCAGAAAACCTGGCCAGCATCGGCAAGAAGTGGAAGCCCCAGAACA TCTACGTGGAGTACGAGAACCTGAGAAGAAACCTGAGAATCGTGGGCGAGAGA CTGTTCGAGAGCAAGGAGGGCCACCAG (SEQ ID NO: 1018)	6.77303438
SUMO3	INLKVAGQDGSV VQFKIKRHTPLSK LMKAYCERQGLS MRQIRFRFDGQP INETDTPAQLEM EDEDTIDVF (SEQ ID NO: 247)	ENDHINLKVAGQDGSVVQFKIK RHTPLSKLMKAYCERQGLSMR QIRFRFDGQPINETDTPAQLEM EDEDTIDVFQQQTGG (SEQ ID NO: 248)	GAGAACGACCACATCAACCTGAAGGTGGCCGGCCAGGACGGCAGCGTGGTGCA GTTCAAGATCAAGAGACACACCCCCCTGAGCAAGCTGATGAAGGCCTACTGCGA GAGACAGGGCCTGAGCATGAGACAGATCAGATTCAGATTCGACGGCCAGCCCA TCAACGAGACCGACACCCCCGCCCAGCTGGAGATGGAGGACGAGGACACCATC GACGTGTTCCAGCAGCAGACCGGCGGC (SEQ ID NO: 1019)	6.76493545
ZNF98	VLTFRDVALEFSL EEWQCLDTAQQ NLYRNVMLENYR NLVFV (SEQ ID NO: 249)	PGPLGSLEMGVLTFRDVALEFSL EEWQCLDTAQQNLYRNVMLE NYRNLVFVGIAASKPDLITCIEQ GKEPWNVKRHEMVT (SEQ ID NO: 250)	CCCGGCCCCCCTGGGCAGCCTGGAGATGGGCGTGCTGACCTTCAGAGATGTGGC CCTGGAGTTCAGCCTGGAGGAGTGGCAGTGCCTGGACACCGCCCAGCAGAACC TGTACAGAAACGTGATGCTGGAGAACTACAGAAACCTGGTGTTCGTGGGCATC GCCGCCAGCAAGCCCGACCTGATCACCTGCCTGGAGCAGGGCAAGGAGCCCTG GAACGTGAAGAGACACGAGATGGTGACC (SEQ ID NO: 1020)	6.76469777
ZN175	SVSFEDVTVDFSR EEWQQLDPAQR CLYRDVMLELYS HLFAV (SEQ ID NO: 251)	LSQKPQVLGPEKQDGSCEASVS FEDVTVDFSREEWQQLDPAQR CLYRDVMLELYSHLFAVGYHIP NPEVIFRMLKEKEPR (SEQ ID NO: 252)	CTGAGCCAGAAGCCCCAGGTGCTGGGCCCCGAGAAGCAGGACGGCAGCTGCG AGGCTAGCGTGAGCTTCGAGGACGTGACCGTGGACTTCAGCAGAGAGGAGTGG CAGCAGCTGGACCCCGCCCAGAGATGCCTGTACAGAGATGTGATGCTGGAGCT GTACAGCCACCTGTTCGCCGTGGGCTACCACATCCCCAACCCCGAGGTGATCTTC AGAATGCTGAAGGAGAAGGAGCCCAGA (SEQ ID NO: 1021)	6.76307142

ZN347	QVTFRDVAIEFSQ EEWTCLDPAQRT LYRDVMLENYRN LASL (SEQ ID NO: 253)	ALTQGQVTFRDVAIEFSQEEWT CLDPAQRTLYRDVMLENYRNL ASLGISCFDLSIISMLEQGKEPFT LESQVQIAGNPDG (SEQ ID NO: 254)	GCCCTGACCCAGGGCCAGGTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCAG GAGGAGTGGACCTGCCTGGACCCCGCCCAGAGAACCCTGTACAGAGATGTGAT GCTGGAGAACTACAGAAACCTGGCCAGCCTGGGCATCAGCTGCTTCGACCTGAG CATCATCAGCATGCTGGAGCAGGGCAAGGAGCCCTTCACCCTGGAGAGCCAGG TGCAGATCGCCGGCAACCCCGACGGC (SEQ ID NO: 1022)	6.75405678
ZNF25	PVTLKDVIVEFTK EEWKILTPAQRT LYKDVMLENYSH LVSV (SEQ ID NO: 255)	NKFQGPVTLKDVIVEFTKEEWK LLTPAQRTLYKDVMLENYSHLV SVGYHVNKPNAVFKLKQGKEP WILEVEFPHRGFPED (SEQ ID NO: 256)	AACAAGTTCCAGGGCCCCGTGACCCTGAAGGACGTGATCGTGGAGTTCACCAA GGAGGAGTGGAAGCTGCTGACCCCCGCCCAGAGAACCCTGTACAAGGACGTGA TGCTGGAGAACTACAGCCACCCTGGTGAGCGTGGGCTACCACGTGAACAAGCCC AACGCCGTGTTCAAGCTGAAGCAGGGCAAGGAGCCCTGGATCCTGGAGGTGGA GTTCCCCCACAGAGGCTCCCCGAGGAC (SEQ ID NO: 1023)	6.75008459
ZN519	LLTFRDVAIEFSPE EWKCLDPAQQN LYRDVMLENYRN LVS (SEQ ID NO: 257)	ELLTFRDVAIEFSPEEWKCLDPA QQNLYRDVMLENYRNLVSLAV YSYYNQGILPEQGIQDSFKKATL GRYGSCGLENICL (SEQ ID NO: 258)	GAGCTGCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCCCGAGGAGTGGAA GTGCCTGGACCCCGCCCAGCAGAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGTGAGCCTGGCCGTGTACAGCTACTACAACCAGGGCATCCTGC CCGAGCAGGGCATCCAGGACAGCTTCAAGAAGGCCACCCTGGGCAGATACGGC AGCTGCGGCCTGGAGAACATCTGCCTG (SEQ ID NO: 1024)	6.74815071
Z585B	SVSFRDVAIDFSR EEWRHLDLSQRN LYRDVMLETYSH LLSV(SEQ ID NO: 259)	SPQKSSALAPEDHGSSYEGSVSF RDVAIDFSREEWRHLDLSQRNL YRDVMLETYSHLLSVGYQVPKP EVVMLEQGKEPWA (SEQ ID NO: 260)	AGCCCCCAGAAGAGCAGCGCCCTGGCCCCCGAGGACCACGGCAGCAGCTACGA GGGAAGCGTGAGCTTCAGAGATGTGGCCATCGACTTCAGCAGAGAGGAGTGGA GACACCTGGACCTGAGCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGACC TACAGCCACCTGCTGAGCGTGGGCTACCAGGTGCCCAAGCCCGAGGTGGTGAT GCTGGAGCAGGGCAAGGAGCCCTGGGCC (SEQ ID NO: 1025)	6.74700322
ZIM3	RVTFEDVTVNFT QGEWQRLNPEQ RNLYRDVMLENY SNLVSV (SEQ ID NO: 261)	NNSQGRVTFEDVTVNFTQGE WQRLNPEQRNLYRDVMLENYS NLVSVGQGETTKPDVILRLEQG KEPWLEEEEVLGSGRAE (SEQ ID NO: 262)	AACAACAGCCAGGGCAGAGTGACCTTCGAGGACGTGACCGTGAACTTCACCCA GGGCGAGTGGCAGAGACTGAACCCCGAGCAGAGAAACCTGTACAGAGATGTG ATGCTGGAGAACTACAGCAACCTGGTGAGCGTGGGCCAGGGCGAGACCACCAA GCCCGACGTGATCCTGAGACTGGAGCAGGGCAAGGAGCCCTGGCTGGAGGAG GAGGAGGTGCTGGGCAGCGGCAGAGCCGAG (SEQ ID NO: 1026)	6.74462278
ZN517	AVVFEDVAVYFT RIEWSCLAPDQQ ALYRDVMLENYG NLASL (SEQ ID NO: 263)	AMALPMPGPQEAVVFEDVAV YFTRIEWSCLAPDQQALYRDV MLENYGNLASLGFLVAKPALISL LEQGEEPGALILQVAE (SEQ ID NO: 264)	GCCATGGCCCTGCCCATGCCCGGCCCCCAGGAGGCCGTGGTGTTCGAGGACGTT GCCGTGTACTTCACCAGAATCGAGTGGAGCTGCCTGGCCCCCGACCAGCAGGCC CTGTACAGAGATGTGATGCTGGAGAACTACGGCAACCTGGCCAGCCTGGGCTTC CTGGTGGCCAAGCCCGCCCTGATCAGCCTGCTGGAGCAGGGCGAGGAGCCCGG AGCCCTGATCCTGCAGGTGGCCGAG (SEQ ID NO: 1027)	6.71923079

ZN846	LVTFEDVAVDFT QEEWTLLDQAQ RDLYRDVMLENY KNLII (SEQ ID NO: 265)	DSSQHLVTFEDVAVDFTQEEW TLLDQAQRDLYRDVMLENYKN LIILAGSELFKRSLMSGLEQMEE LRTGVTGVLQELDLQ (SEQ ID NO: 266)	GACAGCAGCCAGCACCTGGTGACCTTCGAGGACGTGGCCGTGGACTTCACCCA GGAGGAGTGGACCCTGCTGGACCAGGCCCAGAGGACCTGTACAGAGATGTGA TGCTGGAGAACTACAAGAACCTGATCATCCTGGCCGGCAGCGAGCTGTTCAAGA GAAGCCTGATGAGCGGCCTGGAGCAGATGGAGGAGCTGAGAACCGGCGTGAC CGGCGTGCTGCAGGAGCTGGACCTGCAG (SEQ ID NO: 1028)	6.70970056
ZN230	AVTFKDVAVFFTE EELGLLDPAQRKL YQDVMLENFTNL LSV (SEQ ID NO: 267)	TTFKEAVTFKDVAVFFTEEELGL LDPAQRKLYQDVMLENFTNLLS VGHQPFHPFHFLREEKFWMM ETATQREGNSGGKTI (SEQ ID NO: 268)	ACCACCTTCAAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGTTCTTCACCGAG GAGGAGCTGGGCCTGCTGGACCCCGCCCAGAGAAAGCTGTACCAGGACGTGAT GCTGGAGAACTTCACCAACCTGCTGAGCGTGGGCCACCAGCCCTTCCACCCCTTC CACTTCCTGAGAGAGGAGAAGTTCTGGATGATGGAGACCGCCACCCAGAGAGA GGGCAACAGCGGCGGCAAGACCATC (SEQ ID NO: 1029)	6.70246908
ZNF66	PLQFRDVAIEFSL EEWHCLDMAQR NLYRDVMLENYR NLVFL (SEQ ID NO: 269)	GPLQFRDVAIEFSLEEWHCLDM AQRNLYRDVMLENYRNLVFLGI VVSKPDLITHLEQGKKPSTMQR HEMVANPSVLCSHF (SEQ ID NO: 270)	GGCCCCCTGCAGTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA CTGCCTGGACATGGCCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCGTGGTGAGCAAGCCCGACCTGATCACCC ACCTGGAGCAGGGCAAGAAGCCCAGCACCATGCAGAGACACGAGATGGTGGC CAACCCCAGCGTGCTGTGCAGCCACTTC(SEQ ID NO: 1030)	6.69981008
ZFP1	SVSFTDVTVDFT QEEWEQLDPSQ RILYMDVMLENY SNLLSV (SEQ ID NO: 271)	NKSQGSVSFTDVTVDFTQEEW EQLDPSQRILYMDVMLENYSNL LSVEVWKADDQMERDHRNPD EQARQFLILKNQTPIEE (SEQ ID NO: 272)	AACAAGAGCCAGGGCAGCGTGAGCTTCACCGACGTGACCGTGGACTTCACCCA GGAGGAGTGGGAGCAGCTGGACCCCAGCCAGAGAATCCTGTACATGGACGTGA TGCTGGAGAACTACAGCAACCTGCTGAGCGTGGAGGTGTGGAAGGCCGACGAC CAGATGGAGAGAGACCACACAGAAACCCCGACGAGCCAGGCAGACAGTTCCTGAT CCTGAAGAACCAGACCCCCATCGAGGAG (SEQ ID NO: 1031)	6.69334133
ZN713	SLTFQDVAVDFT REEWDQLYPAQ KNLYRDVMLENY RNLVAL (SEQ ID NO: 273)	EEEEMNDGSQMVRSQESLTFQ DVAVDFTREEWDQLYPAQKNL YRDVMLENYRNLVALGYQLCKP EVIAQLELEEEWVIER (SEQ ID NO: 274)	GAGGAGGAGGAGATGAACGACGGCAGCCAGATGGTGAGAAGCCAGGAGAGC CTGACCTTCCAGGACGTGGCCGTGGACTTCACCAGAGAGGAGTGGGACCAGCT GTACCCCGCCCAGAAGAACCTGTACAGAGATGTGATGCTGGAGAACTACAGAA ACCTGGTGGCCCTGGGCTACCAGCTGTGCAAGCCCGAGGTGATCGCCCAGCTG GAGCTGGAGGAGGAGTGGGTGATCGAGAGA (SEQ ID NO: 1032)	6.68245851
ZN816	RLTFRDVAIEFSLE EWKCLNPAQRAL YRAVMLENYRNL EFV(SEQ ID NO: 275)	EEATKKSKEKEPGMALPQGRLT FRDVAIEFSLEEWKCLNPAQRA LYRAVMLENYRNLEFVDSSLKS MMEFSSTRHSITGE (SEQ ID NO: 276)	GAGGAGGCCACCAAGAAGAGCAAGGAGAAGGAGCCCGGCATGGCCCTGCCCC AGGGCAGACTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGG AAGTGCCTGAACCCCGCCCAGAGAGCCCTGTACAGAGCCGTGATGCTGGAGAA CTACAGAAACCTGGAGTTCGTGGACAGCAGCAGCCTGAAGAGCATGATGGAGTTCA GCAGCACCAGACACAGCATCACCGGCGAG (SEQ ID NO: 1033)	6.67677315

ZN426	SVTFDDVAVDFT QEEWTLLDSTQR SLYSDVMLENYK NLATV (SEQ ID NO: 277)	EKTPAGRIVADCLTDCYQDSVT FDDVAVDFTQEEWTLLDSTQRS LYSDVMLENYKNLATVGGQIIK PSLISWLEQEESRT (SEQ ID NO: 278)	GAGAAGACCCCCGCCGGCAGAATCGTGGCCGACTGCCTGACCGACTGCTACCA GGACAGCGTGACCTTCGACGACGTGGCCGTGGACTTCACCCAGGAGGAGTGGA CCCTGCTGGACAGCACCCAGAGAAGCCTGTACAGCGACGTGATGCTGGAGAAC TACAAGAACCTGGCCACCGTGGGCGGCCAGATCATCAAGCCCAGCCTGATCAGC TGGCTGGAGCAGGAGGAGAGCAGAACC (SEQ ID NO: 1034)	6.67185066
ZN701	LLTFRDVAIEFSQ EEWKCLDPAQRT LYRDVMLENYRN LVSL (SEQ ID NO: 279)	KTRKRKRKAKESGMALLQGLLT FRDVAIEFSQEEWKCLDPAQRT LYRDVMLENYRNLVSLDTSSKC MMKMFSSTGQGNTE (SEQ ID NO: 280)	AAGACCAGAAAGAGAAAGAGAAAGGCCAAGGAGAGCGGCATGGCCCTGCTGC AGGGCCTGCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCAGGAGGAGTGG AAGTGCCTGGACCCCGCCCAGAGAACCCTGTACAGAGATGTGATGCTGGAGAA CTACAGAAACCTGGTGAGCCTGGACACCAGCAGCAAGTGCATGATGAAGATGT TCAGCAGCACCGGCCAGGGCAACACCGAG (SEQ ID NO: 1035)	6.66820921
ZN674	SLTFKDVFVDFTL EEWQQLDSAQK NLYRDVMLENYS HLVSV (SEQ ID NO: 281)	AMSQESLTFKDVFVDFTLEEW QQLDSAQKNLYRDVMLENYSH LVSVGHLVGKPDVIFRLGPGDE SWMADGGTPVRTCAGE (SEQ ID NO: 282)	GCCATGAGCCAGGAGAGAGCCTGACCTTCAAGGACGTGTTCGTGGACTTCACCCTG GAGGAGTGGCAGCAGCTGGACAGCGCCCAGAAGAACCTGTACAGAGATGTGAT GCTGGAGAACTACAGCCACCTGGTGAGCGTGGGCCACCTGGTGGGCAAGCCCG ACGTGATCTTCAGACTGGGCCCCGGCGACGAGAGCTGGATGGCCCGACGGCGGC ACACCTGTGAGAACCTGCGCCGGCGAG (SEQ ID NO: 1036)	6.6636553
ZN627	SVAFEDVAVNFT LEEWALLDPSQK NLYRDVMRETFR NLASV (SEQ ID NO: 283)	DSVAFEDVAVNFTLEEWALLDP SQKNLYRDVMRETFRNLASVG KQWEDQNIEDPFKIPRRNISHIP ERLCESKEGGQGEE (SEQ ID NO: 284)	GACAGCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCTGGAGGAGTGGGC CCTGCTGGACCCCAGCCAGAAGAACCTGTACAGAGATGTGATGAGAGACCT TCAGAAACCTGGCCAGCGTGGGCAAGCAGTGGGAGGACCAGAACATCAGGA CCCCTTCAAGATCCCCAGAAGAAACATCAGCCACATCCCCGAGAGACTGTGCGA GAGCAAGGAGGGCGGCCAGGGCGAGGAG (SEQ ID NO: 1037)	6.66232669
ZNF20	SVAFEDVAVSFT QEEWALLDPSQK NLYRDVMQETFK NLTSV (SEQ ID NO: 285)	MFQDSVAFEDVAVSFTQEEWA LLDPSQKNLYRDVMQETFKNLT SVGKTWKVQNIEDEYKNPRRN LSLMREKLCESKESHH (SEQ ID NO: 286)	ATGTTCCAGGACAGCGTGGCCTTCGAGGACGTGGCCGTGAGCTTCACCCAGGA GGAGTGGGCCCTGCTGGACCCCAGCCAGAAGAACCTGTACAGAGATGTGATGC AGGAGACCTTCAAGAACCTGACCAGCGTGGGCAAGACCTGGAAGGTGCAGAAC ATCGAGGACGAGTACAAGAACCCCAGAAGAACCTGAGCCTGATGAGAGAGAA GCTGTGCGAGAGCAAGGAGAGCCACCAC (SEQ ID NO: 1038)	6.65839711
Z587B	TVTFEDVAVKFT QEEWNLLSEAQR CLYRDVTLENLAL MSSL (SEQ ID NO: 287)	AVVATLRLSAQGTVTFEDVAVK FTQEEWNLLSEAQRCLYRDVTL ENLALMSSLGCWCGVEDEAAP SKQSIYIQRETQVRT (SEQ ID NO: 288)	GCCGTGGTGGCCACCCTGAGACTGAGCGCCCAGGGCACCGTGACCTTCGAGGA CGTGGCCGTGAAGTTCACCCAGGAGGAGTGGAACCTGCTGAGCGAGGCCCAGA GATGCCTGTACAGAGATGTGACCCTGGAGAACCTGGCCCTGATGAGCAGCCTG GGCTGCTGGTGCGGCGTGGAGGACGAGGCCGCCCCTAGCAAGCAGAGCATCTA CATCCAGAGAGAGACCCAGGTGAGAACC (SEQ ID NO: 1039)	6.63154785

ZN316	LVTFEDVAVYFSL EEWERLEADQRG LYQEVMQENYGI LVSL (SEQ ID NO: 289)	EEEEEDEDEDDLLTAGCQELVTF EDVAVYFSLEEWERLEADQRGL YQEVMQENYGILVSLGYPIPKP DLIFRLEQGEEPW (SEQ ID NO: 290)	GAGGAGGAGGAGGAGGACGAGGACGACCTGCTGACCGCCGGCTGCC AGGAGCTGGTGACCTTCGAGGACGTGGCCGTGTACTTCAGCCTGGAGGAGTGG GAGAGACTGGAGGCCGACCAGAGAGGCCTGTACCAGGAGGTGATGCAGGAGA ACTACGGCATCCTGGTGAGCCTGGGCTACCCCATCCCCAAGCCCGACCTGATCTT CAGACTGGAGCAGGGCGAGGAGCCCTGG (SEQ ID NO: 1040)	6.62746569
ZN233	MVTFKDVAVVFT REELGLLDLAQRK LYQDVMLENFRN LLSV (SEQ ID NO: 291)	TKFQEMVTFKDVAVVFTREELG LLDLAQRKLYQDVMLENFRNLL SVGYQPFKLDVILQLGKEDKLR MMETEIQGDGCSGH (SEQ ID NO: 292)	ACCAAGTTCCAGGAGATGGTGACCTTCAAGGACGTGGCCGTGGTGTTCACCAGA GAGGAGCTGGGCCTGCTGGACCTGGCCCAGAGAAAGCTGTACCAGGACGTGAT GCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCTACCAGCCCTTCAAGCTGGA CGTGATCCTGCAGCTGGGCAAGGAGGACAAGCTGAGAATGATGGAGACCGAG ATCCAGGGCGACGGCTGCAGCGGCCAC (SEQ ID NO: 1041)	6.62252575
ZN611	RLTFRDVAIEFSLA EWKCLNPSQRAL YREVMLENYRNL EAV (SEQ ID NO: 293)	EEAAQKRKGKEPGMALPQGRL TFRDVAIEFSLAEWKCLNPSQR ALYREVMLENYRNLEAVDISSKC MMKEVLSTGQGNTE (SEQ ID NO: 294)	GAGGAGGCCGCCCAGAAGAGAAAGGGCAAGGAGCCCGGCATGGCCCTGCCCC AGGGCAGACTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGCCGAGTGG AAGTGCCTGAACCCCAGCCAGAGAGCCCTGTACAGAGAGGTGATGCTGGAGAA CTACAGAAACCTGGAGGCCGTGGACATCAGCAGCAAGTGCATGATGAAGGAGG TGCTGAGCACCGGCCAGGGCAACACCGAG (SEQ ID NO: 1042)	6.61854262
ZN556	TVVFEDVVVDFTL EEWALLNPAQRK LYRDVMLETFKH LASV (SEQ ID NO: 295)	DTVVFEDVVVDFTLEEWALLNP AQRKLYRDVMLETFKHLASVD NEAQLKASGSISQQDTSGEKLSL KQKIEKFTRKNIWA (SEQ ID NO: 296)	GACACCGTGGTGTTCGAGGACGTGTGGTGGACTTCACCCTGGAGGAGTGGGC CCTGCTGAACCCCGCCCAGAGAAAGCTGTACAGAGATGTGATGCTGGAGACCTT CAAGCACCTGGCCAGCGTGGACAACGAGGCCCAGCTGAAGGCCAGCGGCAGCA TCAGCCAGCAGGACACCAGCGGCGAGAAGCTGAGCCTGAAGCAGAAGATCGA GAAGTTCACCAGAAAGAACATCTGGGCC (SEQ ID NO: 1043)	6.61519705
ZN234	GLTFKDVAVVFTE EELGLLDPVQRNL YQDVMLENFRNL LSV (SEQ ID NO: 297)	TTFKEGLTFKDVAVVFTEEELGL LDPVQRNLYQDVMLENFRNLL SVGHHPFKHDVFLLEKEKKLDI MKTATQRKGKSADK (SEQ ID NO: 298)	ACCACCTTCAAGGAGGGCCTGACCTTCAAGGACGTGGCCGTGGTGTTCACCGAG GAGGAGCTGGGCCTGCTGGACCCCGTGCAGAGAAAACCTGTACCAGGACGTGAT GCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCCACCACCCCTTCAAGCACGA CGTGTTCCTGCTGGAGAAGGAGAAGAAGCTGGACATCATGAAGACCGCCACCC AGAGAAAGGGCAAGAGCGCCGACAAG (SEQ ID NO: 1044)	6.60158035
ZN560	LVTFDSVAVEFTQ EEWTLLDPAQRN LYSDVMLENYKN LSSV (SEQ ID NO: 299)	SALQQEFWKIQTSNGIQMDLV TFDSVAVEFTQEEWTLLDPAQR NLYSDVMLENYKNLSSVGYQLF KPSLISWLEEEEELS (SEQ ID NO: 300)	AGCGCCCTGCAGCAGGAGTTCTGGAAGATCCAGACCAGCAACGGCATCCAGAT GGACCTGGTGACCTTCGACAGCGTGGCCGTGGAGTTCACCCAGGAGGAGTGGA CCCTGCTGGACCCCGCCCAGAGAAACCTGTACAGCGACGTGATGCTGGAGAACT ACAAGAACCTGAGCAGCGTGGGCTACCAGCTGTTCAAGCCCAGCCTGATCAGCT GGCTGGAGGAGGAGGAGGAGCTGAGC (SEQ ID NO: 1045)	6.60066711

ZNF77	CVIFEEVAVNFTP EEWALLDHAQRS LYRDVMLETCRN LASL (SEQ ID NO: 301)	DCVIFEEVAVNFTPEEWALLDH AQRSLYRDVMLETCRNLASLDC YIYVRTSGSSSQRDVFGNGISND EEIVKFTGSDSWS (SEQ ID NO: 302)	GACTGCGTGATCTTCGAGGAGGTGGCCGTGAACTTCACCCCCGAGGAGTGGGC CCTGCTGGACCACGCCCAGAGAAGCCTGTACAGAGATGTGATGCTGGAGACCT GCAGAAACCTGGCCAGCCTGGACTGCTACATCTACGTGAGAACCAGCGGCAGC AGCAGCCAGAGAGATGTGTTCGGCAACGGCATCAGCAACGACGAGGAGATCGT GAAGTTCACCGGCAGCGACAGCTGGAGC (SEQ ID NO: 1046)	6.58987943
ZN682	LLTFRDVTIEFSLE EWEFLNPAQQSL YRKVMLENYRNL VSL (SEQ ID NO: 303)	ELLTFRDVTIEFSLEEWEFLNPA QQSLYRKVMLENYRNLVSLGLT VSKPELISRLEQRQEPWNVKRH ETIAKPPAMSSHY (SEQ ID NO: 304)	GAGCTGCTGACCTTCAGAGATGTGACCATCGAGTTCAGCCTGGAGGAGTGGGA GTTCCTGAACCCCGCCCAGCAGAGCCTGTACAGAAAGGTGATGCTGGAGAACTA CAGAAACCTGGTGAGCCTGGGCCTGACCGTGAGCAAGCCCGAGCTGATCAGCA GACTGGAGCAGAGACAGGAGCCCTGGAACGTGAAGAGACACGAGACCATCGC CAAGCCCCCCGCCATGAGCAGCCACTAC (SEQ ID NO: 1047)	6.58030961
ZN614	SLTLEDVAVEFS WEEWQLLDTAQ KNLYRDVMVENY NHLVSL (SEQ ID NO: 305)	IKTQESLTLEDVAVEFSWEEWQ LLDTAQKNLYRDVMVENYNHL VSLGYQTSKPDVLSKLAHGQEP WTTDAKIQNKNCPGI (SEQ ID NO: 306)	ATCAAGACCCAGGAGAGCCTGACCCTGGAGGACGTGGCCGTGGAGTTCAGCTG GGAGGAGTGGCAGCTGCTGCTGGACACCGCCCAGAAGAACCTGTACAGAGATGTGA TGGTGGAGAACTACAACCACCTGGTGAGCCTGGGCTACCAGACCAGCAAGCCC GACGTGCTGAGCAAGCTGGCCCACGGCCAGGAGCCCTGGACCACCGACGCCAA GATCCAGAACAAGAACTGCCCCGGCATC (SEQ ID NO: 1048)	6.57723831
ZN785	AVSFADVAVYFS PEEWECLRPAQR AIYRDVMRETFG HLGAL (SEQ ID NO: 307)	PAHVPGEAGPRRTRESRPGAVS FADVAVYFSPEEWECLRPAQRA LYRDVMRETFGHLGALGFSVPK PAFISWVEGEVEAW (SEQ ID NO: 308)	CCCGCCCACGTGCCCGGCGAGGCCGGCCCCAGAAGAACCAGAGAGAGCAGACC CGGCGCTGTGAGCTTCGCCGACGTGGCCGTGTACTTCAGCCCCGAGGAGTGGG AGTGCCTGAGACCCGCCCAGAGAGCCCTGTACAGAGATGTGATGAGAGAGACC TTCGGCCACCTGGGCGCCCTGGGCTTCAGCGTGCCCAAGCCCGCCTTCATCAGCT GGGTGGAGGGCGAGGTGGAGGCCTGG (SEQ ID NO: 1049)	6.56301724
ZN445	TMTFKDVEVTFS QDEWGWLDSA QRNLYRDVMLE NYRNMAS (SEQ ID NO: 309)	GCPGDQVTPTRSLTAQLQETM TFKDVEVTFSQDEWGWLDSAQ RNLYRDVMLENYRNMASLVGP FTKPALISWLEAREPWG (SEQ ID NO: 310)	GGCTGCCCCGGCGACCAGGTGACCCCACCAGAAGCCTGACCGCCCAGCTGCA GGAGACCATGACCTTCAAGGACGTGGAGGTGACCTTCAGCCAGGACGAGTGGG GCTGGCTGGACAGCGCCCAGAGAAACCTGTACAGAGATGTGATGCTGGAGAAC TACAGAAACATGGCCAGCCTGGTGGGCCCCTTCACCAAGCCCGCCCTGATCAGC TGGCTGGAGGCCAGAGAGCCCTGGGGC (SEQ ID NO: 1050)	6.54429484
ZFP30	LVMFRDVAVDFS QEEWECLNSYQR NLYRDVILENYSN LVS (SEQ ID NO: 311)	ARDLVMFRDVAVDFSQEEWEC LNSYQRNLYRDVILENYSNLVSL AGCSISKPDVITLLEQGKEPWM VVRDEKRRWTLDLE (SEQ ID NO: 312)	GCCAGAGACCTGGTGATGTTCAGAGATGTGGCCGTGGACTTCAGCCAGGAGGA GTGGGAGTGCCTGAACAGCTACCAGAGAAACCTGTACAGAGATGTGATCCTGG AGAACTACAGCAACCTGGTGAGCCTGGCCGGCTGCAGCATCAGCAAGCCCGAC GTGATCACCCTGCTGGAGCAGGGCAAGGAGCCCTGGATGGTGGTGAGAGATGA GAAGAGAAGATGGACCCTGGACCTGGAG (SEQ ID NO: 1051)	6.54105426

ZN225	AVTFKDVAVVFT EEELRLLDLAQRK LYREVMLENFRN LLSV (SEQ ID NO: 313)	TTLKEAVTFKDVAVVFTEEELRL LDLAQRKLYREVMLENFRNLLS VGHQSLHRDTFHFLKEEKFWM METATQREGNLGGK (SEQ ID NO: 314)	ACCACCCTGAAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGGTGTTCACCGA GGAGGAGCTGAGACTGCTGGACCTGGCCCAGAGAAAGCTGTACAGAGAGGTG ATGCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCCACCAGAGCCTGCACAG AGACACCTTCCACTTCCTGAAGGAGGAGAAGTTCTGGATGATGGAGACCGCCAC CCAGAGAGAGGGCAACCTGGGCGGCAAG (SEQ ID NO: 1052)	6.53858149
ZN551	GMTFEDVAIYFS QEEWELLDESQR FLYCDVMLENFA HVTSL (SEQ ID NO: 315)	SPPSPRSSMAAVALRDSAQGM TFEDVAIYFSQEEWELLDESQRF LYCDVMLENFAHVTSLGYCHG MENEAIASEQSVSIQ (SEQ ID NO: 316)	AGCCCCCCCAGCCCCAGAAGCAGCATGGCCGCCGTGGCCCTGAGAGACAGCGC CCAGGGCATGACCTTCGAGGACGTGGCCATCTACTTCAGCCAGGAGGAGTGGG AGCTGCTGGACGAGAGCCAGAGATTCCTGTACTGCGACGTGATGCTGGAGAAC TTCGCCCACGTGACCAGCCTGGGCTACTGCCACGGCATGGAGAACGAGGCCATC GCCAGCGAGCAGAGCGTGAGCATCCAG (SEQ ID NO: 1053)	6.53471613
ZN610	RLTFMDVAIEFSQ EEWKSLDPGQRA LYRDVMLENYRN LVFL (SEQ ID NO: 317)	DEEAQKRKAKESGMALPQGRL TFMDVAIEFSQEEWKSLDPGQ RALYRDVMLENYRNLVFLGICLP DLSIISMLKQRREPL (SEQ ID NO: 318)	GACGAGGAGGCCCAGAAGAGAAAGGCCAAGGAGAAGCGGCATGGCCCTGCCCC AGGGCAGACTGACCTTCATGGACGTGGCCATCGAGTTCAGCCAGGAGGAGTGG AAGAGCCTGGACCCCGGCCAGAGAGCCCTGTACAGAGATGTGATGCTGGAGAA CTACAGAAACCTGGTGTTCCTGGGCATCTGCCTGCCCGACCTGAGCATCATCAGC ATGCTGAAGCAGAGAAGAGAGCCCCTG (SEQ ID NO: 1054)	6.53304307
ZN528	PLKFMDVAIEFSQ EEWKCLDPAQRT LYRDVMLENYRN LVSL (SEQ ID NO: 319)	ALTQGPLKFMDVAIEFSQEEWK CLDPAQRTLYRDVMLENYRNLV SLGICLPDLSVTSMLEQKRDPW TLQSEEKIANDPDG (SEQ ID NO: 320)	GCCCTGACCCAGGGCCCCCTGAAGTTCATGGACGTGGCCATCGAGTTCAGCCAG GAGGAGTGGAAGTGCCTGGACCCCGCCCAGAGAACCCTGTACAGAGATGTGAT GCTGGAGAACTACAGAAACCTGGTGAGCCTGGGCATCTGCCTGCCCGACCTGA GCGTGACCAGCATGCTGGAGCAGAAGAGAGACCCCTGGACCCTGCAGAGCGAG GAGAAGATCGCCAACGACCCCGACGGC (SEQ ID NO: 1055)	6.5320662
ZN284	AVTFKDVAVVFT EEELGLLDVSQRK LYRDVMLENFRN LLSV (SEQ ID NO: 321)	TMFKEAVTFKDVAVVFTEEELG LLDVSQRKLYRDVMLENFRNLL SVGHQLSHRDTFHFQREEKFWI METATQREGNSGGK (SEQ ID NO: 322)	ACCATGTTCAAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGGTGTTCACCGAG GAGGAGCTGGGCCTGCTGGACGTGAGCCAGAGAAAGCTGTACAGAGATGTGAT GCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCCACCAGCTGAGCCACAGAG ACACCTTCCACTTCCAGAGAGAGGAGAAGTTCTGGATCATGGAGACCGCCACCC AGAGAGAGGGCAACAGCGGCGGCAAG (SEQ ID NO: 1056)	6.52062588
ZN418	TVAFEDVAVNFS QEEWSLLSEVQR CLYHDVMLENW VLISSL (SEQ ID NO: 323)	QGTVAFEDVAVNFSQEEWSLLS EVQRCLYHDVMLENWVLISSLG CWCGSEDEEAPSKKSISIQRVSQ VSTPGAGVSPKKA (SEQ ID NO: 324)	CAGGGCACCGTGGCCTTCGAGGACGTGGCCGTGAACTTCAGCCAGGAGGAGTG GAGCCTGCTGAGCGAGGTGCAGAGATGCCTGTACCACGACGTGATGCTGGAGA ACTGGGTGCTGATCAGCAGCCTGGGCTGCTGGTGCGGCAGCGAGGACGAGGA GGCCCCCAGCAAGAAGAGCATCAGCATCCAGAGAGTGAGCCAGGTGAGCACCC CCGGCGCCGGCGTGAGCCCCAAGAAGGCC (SEQ ID NO: 1057)	6.51925026

MPP8	FEVEKILDMKTEGGKVLYKVRWKGYTSDDDTWEPEIHLEDCKEVLLEFRK (SEQ ID NO: 325)	AEAFGDSEEDGEDVFEVEKILDMKTEGGKVLYKVRWKGYTSDDDTWEPEIHLEDCKEVLLEFRKKIAENKAKAVRKDIQR (SEQ ID NO: 326)	GCCGAGGCCTTCGGCGACAGCGAGGAGGACGGCGAGGACGTGTTCGAGGTGGAGAAGATCCTGGACATGAAGACCGAGGGCGGCAAGGTGCTGTACAAGGTGAGATGGAAGGGCTACACCAGCGACGACACCTGGGAGCCCGAGATCCACCTGGAGGCCAAGGCCGTGAGAAAGGACATCCAGAGA (SEQ ID NO: 1058)	6.51334634
ZN490	SISLEDVAVNFTLEEWALLDPGQRNIYRDVMRATFKNLACI (SEQ ID NO: 327)	VLQMQNSEHHGQSIKTQTDSISLEDVAVNFTLEEWALLDPGQRNIYRDVMRATFKNLACIGEKWKDQDIEDEHKNQGRNL (SEQ ID NO: 328)	GTGCTGCAGATGCAGAACAGCGAGCACCACGGCCAGAGCATCAAGACCCAGACCGACAGCATCAGCCTGGAGGACGTGGCCGTGAACTTCACCCTGGAGGAGTGGGCCCTGCTGGACCCCGGCCAGAGAAACATCTACAGAGATGTGATGAGAGCCACCTTCAAGAACCTGGCCTGCATCGGCGAGAAGTGGAAGGACCAGGACATCGAGGAC (SEQ ID NO: 1059)	6.51148602
ZN805	SVTFDDVAVTFTQEEWGQLDLAQRTLYQEVMLENCGLLVSL (SEQ ID NO: 329)	AMALTDPAQVSVTFDDVAVTFTQEEWGQLDLAQRTLYQEVMLENCGLLVSLGCPVPRPELIYHLEHGQEPWTRKEDLSQG (SEQ ID NO: 330)	GCCATGGCCCTGACCGACCCCGCCCAGGTGAGCGTGACCTTCGACGACGTGGCCGTGACCTTCACCCAGGAGGAGTGGGGCCAGCTGGACCTGGCCCAGAGAACCCTGTACCAGGAGGTGATGCTGGAGAACTGCGGCCTGCTGGTGAGCCTGGGCTGCCCCGTGCCCAGACCCGAGCTGATCTACCACCTGGAGCACGGCCAGGAGCCCTGGACCAGAAAGGAGGACCTGAGCCAGGGC (SEQ ID NO: 1060)	6.50974725
Z780B	SVTFRDVAIDFSQEEWECLLYRDVMLENYSHLISL (SEQ ID NO: 331)	VHGSVTFRDVAIDFSQEEWECLQPDQRTLYRDVMLENYSHLISLGSSISKPDVITLLEQEKEPWIVVSKETSRWYPDLES (SEQ ID NO: 332)	GTGCACGGCAGCGTGACCTTCAGAGATGTGGCCATCGACTTCAGCCAGGAGGAGTGGGAGTGCCTGCAGCCCGACCAGAGAACCCTGTACAGAGATGTGATGCTGGAGAACTACAGCCACCTGATCAGCCTGGGCAGCATCAGCAAGCCCGACGTGATCACCCTGCTGGAGCAGGAGAAGGAGCCCTGGATCGTGGTGAGCAAGGAGACCAGCGATGGTACCCCGACCTGGAGCAGGAAGGAGCCCTGGAGCAAGGAGAC (SEQ ID NO: 1061)	6.50607891
ZN763	PVACEDVAVNFTQEEWALLDISQRKLYREVMLETFRNLTSI (SEQ ID NO: 333)	DPVACEDVAVNFTQEEWALLDISQRKLYREVMLETFRNLTSIGKKWKDQNIEYEYQNPRRNFRSLIEGNVNEIKEDSHCG (SEQ ID NO: 334)	GACCCCGTGGCCTGCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGCCCTGCTGGACATCAGCCAGAGAAAGCTGTACAGAGAGGTGATGCTGGAGACCTTCAGAAACCTGACCAGCATCGGCAAGAAGTGGAAGGACCAGAACATCGAGTACCGAGATCAAGGAGGACACAGCCACTGCGGC (SEQ ID NO: 1062)	6.49330748
ZN285	RVTFKDVAVVFTKEELALLYQDVMLENFRNLML (SEQ ID NO: 335)	IKFQERVTFKDVAVVFTKEELALLDKAQINLVRDGIKNNILNLQAKGLSYLSQEVLHCWQIWKQRI (SEQ ID NO: 336)	ATCAAGTTCCAGGAGAGAGTGACCTTCAAGGACGTGGCCGTGGTGTTCACCAAGGAGGAGCTGGCCCTGCTGGACAAGGCCCAGATCAACCTGTACCAGGACGTGATGCTGGAGAACTTCAGAAACCTGATGCTGGTGAGAGATGGCATCAAGAACAACTGCAGGCAGATCTGGAAGCAGAGAATC (SEQ ID NO: 1063)	6.48639829

ZNF85	PLTFRDVAIEFSLK EWQCLDTAQRN LYRNVMLENYRN LVFL (SEQ ID NO: 337)	GPLTFRDVAIEFSLKEWQCLDT AQRNLYRNVMLENYRNLVFLGI TVSKPDLITCLEQGKEAWSMKR HEIMVAKPTVMCSH (SEQ ID NO: 338)	GGCCCCCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGAAGGAGTGGCA GTGCCTGGACACCGCCCAGAGAAACCTGTACAGAAACGTGATGCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCACCGTGAGCAAGCCCGACCTGATCACCT GCCTGGAGCAGGGCAAGGAGGCCTGGAGCATGAAGAGACACGAGATCATGGT GGCCAAGCCCACCGTGATGTGCAGCCAC (SEQ ID NO: 1064)	6.48512557
ZN223	AVTFKDVAVVFT EEEIGLLDLAQRK LYRDVMLENFRN LLSV (SEQ ID NO: 339)	TMSKEAVTFKDVAVVFTEEELG LLDLAQRKLYRDVMLENFRNLL SVGHQPFHRDTFHFLREEKFW MMDIATQREGNSGGK (SEQ ID NO: 340)	ACCATGAGCAAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGGTGTTCACCGA GGAGGAGCTGGGCCTGCTGGACCTGGCCCAGAGAAAGCTGTACAGAGATGTGA TGCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCCACCAGCCCTTCCACAGAG ACACCTTCCACTTCCTGAGAGAGGAGAAGTTCTGGATGATGGACATCGCCACCC AGAGAGAGGGCAACAGCGGCGGCAAG (SEQ ID NO: 1065)	6.48230966
ZNF90	PLEFRDVAIEFSLE EWHCLDTAQQN LYRDVMLENYRH LVFL (SEQ ID NO: 341)	GPLEFRDVAIEFSLEEWHCLDTA QQNLYRDVMLENYRHLVFLGIV VTKPDLITCLEQGKKPFTVKRHE MIAKSPVMCFHF (SEQ ID NO: 342)	GGCCCCCTGGAGTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA CTGCCTGGACACCGCCCAGCAGAACCTGTACAGAGATGTGATGCTGGAGAACTA CAGACACCTGGTGTTCCTGGGCATCGTGGTGACCAAGCCCGACCTGATCACCTG CCTGGAGCAGGGCAAGAAGCCCTTCACCGTGAAGAGACACGAGATGATCGCCA AGAGCCCCGTGATGTGCTTCCACTTC (SEQ ID NO: 1066)	6.47855756
ZN557	LVTFEDVAVEFTQ EEWALLDPAQRT LYRDVMLENCRN LASL (SEQ ID NO: 343)	GHTEGGELVNELLKSWLKGLVT FEDVAVEFTQEEWALLDPAQR TLYRDVMLENCRNLASLGNQV DKPRLISQLEQEDKVM (SEQ ID NO: 344)	GGCCACACCGAGGGCGGCGAGCTGGTGAACGAGCTGCTGAAGAGCTGGCTGA AGGGCCTGGTGACCTTCGAGGACGTGGCCGTGGAGTTCACCCAGGAGGAGTGG GCCCTGCTGGACCCCGCCCAGAGAACCCTGTACAGAGATGTGATGCTGGAGAA CTGCAGAAACCTGGCCAGCCTGGGCAACCAGGTGGACAAGCCCAGACTGATCA GCCAGCTGGAGCAGGAGGACAAGGTGATG (SEQ ID NO: 1067)	6.47397343
ZN425	TVTFDDVALYFSE QEWEILEKWQK QMYKQEMKTNY ETLDSL (SEQ ID NO: 345)	AEPASVTVTFDDVALYFSEQEW EILEKWQKQMYKQEMKTNYET LDSLGYAFSKPDLITWMEQGR MLLISEQGCLDKTRRT (SEQ ID NO: 346)	GCCGAGCCCGCCAGCGTGACCGTGACCTTCGACGACGTGGCCCTGTACTTCAGC GAGCAGGAGTGGGAGATCCTGGAGAAGTGGCAGAAGCAGATGTACAAGCAGG AGATGAAGACCAACTACGAGACCCTGGACAGCCTGGGCTACGCCTTCAGCAAG CCCGACCTGATCACCTGGATGGAGCAGGGCAGAATGCTGCTGATCAGCGAGCA GGGCTGCCTGGACAAGACCAGAAGAACC (SEQ ID NO: 1068)	6.47320582
ZN229	PLSFKDVAVVFTE EELELLDSTQRQL YQDVMQENFRN LLSV (SEQ ID NO: 347)	HSQASAISQDREEKIMSQEPLSF KDVAVVFTEEELELLDSTQRQLY QDVMQENFRNLLSVGERNPLG DKNGKDTEYIQDE (SEQ ID NO: 348)	CACAGCCAGGCCAGCGCCATCAGCCAGGACAGAGAGGAGAAGATCATGAGCCA GGAGCCCCTGAGCTTCAAGGACGTGGCCGTGGTGTTCACCGAGGAGGAGCTGG AGCTGCTGGACAGCACCCAGAGACAGCTGTACCAGGACGTGATGCAGGAGAAC TTCAGAAACCTGCTGAGCGTGGGCGAGAAACCCCCTGGGCGACAAGAACGG CAAGGACACCGAGTACATCCAGGACGAG(SEQ ID NO: 1069)	6.47139743

ZN606	PVTFKDVAVDFT QEEWGQLDLVQ RTLYRDVMLETY GHLLSV (SEQ ID NO: 349)	GSLEEGRRATGLPAAQVQEPVT FKDVAVDFTQEEWGQLDLVQR TLYRDVMLETYGHLLSVGNQIA KPEVISLLEQGEEPW (SEQ ID NO: 350)	GGCAGCCTGGAGGAGGGCAGAAGAGCCACCGGCCTGCCCGCCGCCCAAGTGC AGGAGCCTGTGACCTTCAAGGACGTGGCCGTGGACTTCACCCAGGAGGAGTGG GGCCAGCTGGACCTGGTGCAGAGAACCCTGTACAGAGATGTGATGCTGGAGAC CTACGGCCACACCTGCTGAGCGTGGGCAACCAGATCGCCAAGCCCGAGGTGATCA GCCTGCTGGAGCAGGGCGAGGAGCCCTGG (SEQ ID NO: 1070)	6.46489693
ZN155	AVTFKDVAVVFT EEELGLLDPAQRK LYRDVMLENFRN LLSV (SEQ ID NO: 351)	TTFKEAVTFKDVAVVFTEEELGL LDPAQRKLYRDVMLENFRNLLS VGHQPFHQDTCHFLREEKFW MMGTATQREGNSGGK (SEQ ID NO: 352)	ACCACCTTCAAGGAGGCCGTGACCTTCAAGGACGTGGCCGTGGTGTTCACCGAG GAGGAGCTGGGCCTGCTGGACCCCGCCCAGAGAAAGCTGTACAGAGATGTGAT GCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCCACCAGCCCTTCCACCAGGA CACCTGCCACTTCCTGAGAGAGGAGAAGTTCTGGATGATGGGCACCGCCACCCA GAGAGAGGGCAACAGCGGCGGCAAG (SEQ ID NO: 1071)	6.45744473
ZN222	AVTFKDVAVIFTE EELGLLDPAQRKL YRDVMLENFRNL LSV (SEQ ID NO: 353)	AKLYEAVTFKDVAVIFTEEELGLL DPAQRKLYRDVMLENFRNLLSV GGKIQTEMETVPEAGTHEEFSC KQIWEQIASDLT (SEQ ID NO: 354)	GCCAAGCTGTACGAGGCCGTGACCTTCAAGGACGTGGCCGTGATCTTCACCGAG GAGGAGCTGGGCCTGCTGGACCCCGCCCAGAGAAAGCTGTACAGAGATGTGAT GCTGGAGAACTTCAGAAACCTGCTGAGCGTGGGCGGCAAGATCCAGACCGAGA TGGAGACCGTGCCCGAGGCCGGCACCCACGAGGAGTTCAGCTGCAAGCAGATC TGGGAGCAGATCGCCAGCGACCTGACC (SEQ ID NO: 1072)	6.45544011
ZN442	SVAFEDVAVNFT QEEWALLGPSQK SLYRDVMWETIR NLDCI (SEQ ID NO: 355)	RSDLFLPDSQTNEERKQYDSVA FEDVAVNFTQEEWALLGPSQK SLYRDVMWETIRNLDCIGMKW EDTNIEDQHRNPRRSL (SEQ ID NO: 356)	AGAAGCGACCTGTTCCTGCCCGACAGCCAGACCAACGAGGAGAAGAAAGCAGTA CGACAGCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGG CCCTGCTGGGCCCCAGCCAGAAGAGCCTGTACAGAGATGTGATGTGGGAGACC ATCAGAAACCTGGACTGCATCGGCATGAAGTGGGAGGACACCAACATCGAGGA CCAGCACAGAAACCCCAGAAGAAGCCTG (SEQ ID NO: 1073)	6.44268455
ZNF91	LLTFRDVAIEFSPE EWQCLDTAQQN LYRNVMLENYRN LAFL (SEQ ID NO: 357)	PGTPGSLEMGLLTFRDVAIEFSP EEWQCLDTAQQNLYRNVMLE NYRNLAFLGIALSKPDLITYLEQG KEPWNMKQHEMVD (SEQ ID NO: 358)	CCCGGCACCCCCCGGCAGCCTGGAGATGGGCCTGCTGACCTTCAGAGATGTGGCC ATCGAGTTCAGCCCCGAGGAGTGGCAGTGCCTGGACACCGCCCAGCAGAACCT GTACAGAAACGTGATGCTGGAGAACTACAGAAACCTGGCCTTCCTGGGCATCGC CCTGAGCAAGCCCGACCTGATCACCTACCTGGAGCAGGGCAAGGAGCCCCTGGA ACATGAAGCAGCACGAGATGGTGGAC (SEQ ID NO: 1074)	6.44174437
ZN135	QVTFEDVVVGFS QEEWGQLKPAQ RTLYRDVMLDTF RLLVSV (SEQ ID NO: 359)	TPGVRVSTDPEQVTFEDVVVGF SQEEWGQLKPAQRTLYRDVML DTFRLLVSVGHWLPKPNVISLLE QEAELWAVESRLPQ (SEQ ID NO: 360)	ACCCCCGGCGTGAGAGTGAGCACCGACCCCGAGCAGGTGACCTTCGAGGACGT GGTGGTGGGCTTCAGCCAGGAGGAGTGGGGCCAGCTGAAGCCCGCCCAGAGA ACCCTGTACAGAGATGTGATGCTGGACACCTTCAGACTGCTGGTGAGCGTGGGC CACTGGCTGCCCAAGCCCAACGTGATCAGCCTGCTGGAGCAGGAGGCCGAGCT GTGGGCCGTGGAGAGCAGACTGCCCCAG (SEQ ID NO: 1075)	6.44116741

ZN778	AVTFDDVAVDFT QEEWTLLDPSQR DLYRDVMLENYE NLASV (SEQ ID NO: 361)	EQTQAAGMVAGWLINCYQDA VTFDDVAVDFTQEEWTLLDPS QRDLYRDVMLENYENLASVEW RLKTKGPALRQDRSWFRA (SEQ ID NO: 362)	GAGCAGACCCAGGCCGCCGGCATGGTGGCCGGCTGGCTGATCAACTGCTACCA GGACGCCGTGACCTTCGACGACGTGGCCGTGGACTTCACCCAGGAGGAGTGGA CCCTGCTGGACCCCAGCCAGAGAGACCTGTACAGAGATGTGATGCTGGAGAAC TACGAGAACCTGGCCAGCGTGGAGTGGAGACTGAAGACCAAGGGCCCCGCCCT GAGACAGGACAGAAGCTGGTTCAGAGCC (SEQ ID NO: 1076)	6.43548986
RYBP	RPRLKNVDRSTA QQLAVTVGNVTV IITDFKEK (SEQ ID NO: 363)	PSEANSIQSANATTKTSETNHTS RPRLKNVDRSTAQQLAVTVGN VTVIITDFKEKTRSSSTSSSTVTSS AGSEQQNQSSS (SEQ ID NO: 364)	CCCAGCGAGGCCAACAGCATCCAGAGCGCCAACGCCACCACCAAGACCAGCGA GACCAACCACACCAGCAGACCCAGACTGAAGAACGTGGACAGAAGCACCGCCC AGCAGCTGGCCGTGACCGTGGGCAACGTGACCGTGATCATCACCGACTTCAAG GAGAAGACCAGAAGCAGCAGCACCAGCAGCACCGTGACCAGCAGCGCCG GCAGCGAGCAGCAGAACCAGAGCAGCAGC (SEQ ID NO: 1077)	6.42734946
ZN534	QLSFSDVAIEFSQ EEWKCLDPGQK ALYRDVMLENYR NLVSL (SEQ ID NO: 365)	ALTQGQLSFSDVAIEFSQEEWK CLDPGQKALYRDVMLENYRNL VSLGEDNVRPEACICSGICLPDL SVTSMLEQKRDPWT (SEQ ID NO: 366)	GCCCTGACCCAGGGCCAGCTGAGCTTCAGCGACGTGGCCATCGAGTTCAGCCAG GAGGAGTGGAAGTGCCTGGACCCCGGCCAGAAGGCCCTGTACAGAGATGTGAT GCTGGAGAACTACAGAAACCTGGTGAGCCTGGGCGAGGACAACGTGAGACCCG AGGCCTGCATCTGCAGCGGCATCTGCCTGCCCGACCTGAGCGTGACCAGCATGC TGGAGCAGAAGAGACCCCTGGACC (SEQ ID NO: 1078)	6.42731382
ZN586	SVTFEDVAVNFSL EEWSLLNEAQRC LYRDVMLETLTLI SSL (SEQ ID NO: 367)	AAAAALRAPAQSSVTFEDVAV NFSLEEWSLLNEAQRCLYRDV MLETLTLISSLGCWHGGEDEAA PSKQSTCIHIYKDQGG (SEQ ID NO: 368)	GCCGCCGCCGCCGCCCTGAGAGCCCCCGCCCAGAGCAGCGTTACATTCGAGGAC GTGGCCGTGAACTTCAGCCTGGAGGAGTGGAGCCTGCTGAACGAGGCCCAGAG ATGCCTGTACAGAGATGTGATGCTGGAGACCCTGACCCTGATCAGCAGCCTGGG CTGCTGGCACGGCGGCGAGGACGAGGCCGCTCCTAGCAAGCAGAGCACCTGCA TCCACATCTACAAGGACCAGGGCGGC (SEQ ID NO: 1079)	6.41123861
ZN567	SVSFNDVTVDFT QEEWQHLDHAQ KTLYMDVMLENY CHLISV (SEQ ID NO: 369)	AQGSVSFNDVTVDFTQEEWQ HLDHAQKTLYMDVMLENYCHL ISVGCHMTKPDVILKLERGEEP WTSFAGHTCLEENWKAE (SEQ ID NO: 370)	GCCCAGGGCAGCGTGAGCTTCAACGACGTGACCGTGGACTTCACCCAGGAGGA GTGGCAGCACCTGGACCACGCCCAGAAGACCCTGTACATGGACGTGATGCTGG AGAACTACTGCCACCTGATCAGCGTGGGCTGCCACATGACCAAGCCCGACGTGA TCCTGAAGCTGGAGAGAGGCGAGGAGCCCTGGACCAGCTTCGCCGGCCACACC TGCCTGGAGGAGAACTGGAAGGCCGAG (SEQ ID NO: 1080)	6.40288995
ZN440	PVAFKDVAVNFT QEEWALLDISQR KLYREVMLETFR NLTSL (SEQ ID NO: 371)	DPVAFKDVAVNFTQEEWALLDI SQRKLYREVMLETFRNLTSLGKR WKDQNIEYEHQNPRRNFRSLIE EKVNEIKDDSHCG (SEQ ID NO: 372)	GACCCCGTGGCCTTCAAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGGACATCAGCCAGAGAAAGCTGTACAGAGAGGGTGATGCTGGAGACCT TCAGAAACCTGACCAGCCTGGGCAAGAGATGGAAGGACCAGAACATCGAGTAC GAGCACCAGAACCCCAGAAGAAACTTCAGAAGCCTGATCGAGGAGAAGGTGAA CGAGATCAAGGACGACAGCCACTGCGGC (SEQ ID NO: 1081)	6.40187146

ZN583	LVTFGDVAVNFS QEEWEWLNPAQ RNLYRKVMLENY RSLVSL (SEQ ID NO: 373)	SKDLVTFGDVAVNFSQEEWEW LNPAQRNLYRKVMLENYRSLVS LGVSVSKPDVISLLEQGKEPWM VKKEGTRGPCPDWEY (SEQ ID NO: 374)	AGCAAGGACCTGGTGACCTTCGGCGACGTGGCCGTGAACTTCAGCCAGGAGGA GTGGGAGTGGCTGAACCCCGCCCAGAGAAACCTGTACAGAAAGGTGATGCTGG AGAACTACAGAAGCCTGGTGAGCCTGGGCGTGAGCGTGAGCAAGCCCGACGTG ATCAGCCTGCTGGAGCAGGGCAAGGAGCCCTGGATGGTGAAGAAGGAGGGCA CCAGAGGCCCCTGCCCCGACTGGGAGTAC (SEQ ID NO: 1082)	6.39776145
ZN441	SVAFEDVAINFTC EEWALLGPSQKS LYRDVMQETIRN LDCI (SEQ ID NO: 375)	DSVAFEDVAINFTCEEWALLGP SQKSLYRDVMQETIRNLDCIGM IWQNHDIEEDQYKDLRRNLRC HMVERACEIKDNSQC (SEQ ID NO: 376)	GACAGCGTGGCCTTCGAGGACGTGGCCATCAACTTCACCTGCGAGGAGTGGGC CCTGCTGGGCCCCAGCCAGAAGAGCCTGTACAGAGATGTGATGCAGGAGACCA TCAGAAACCTGGACTGCATCGGCATGATCTGGCAGAACCACGACATCGAGGAG GACCAGTACAAGGACCTGAGAAGAAACCTGAGATGCCACATGGTGGAGAGAGC CTGCGAGATCAAGGACAACAGCCAGTGC (SEQ ID NO: 1083)	6.38715626
ZNF43	PLTFMDVAIEFCL EEWQCLDIAQQ NLYRNVMLENYR NLVFL (SEQ ID NO: 377)	GPLTFMDVAIEFCLEEWQCLDI AQQNLYRNVMLENYRNLVFLGI AVSKPDLITCLEQEKEPWEPMR RHEMVAKPPVMCSH (SEQ ID NO: 378)	GGCCCCCTGACCTTCATGGACGTGGCCATCGAGTTCTGCCTGGAGGAGTGGCAG TGCCTGGACATCGCCCAGCAGAACCTGTACAGAAACGTGATGCTGGAGAACTAC AGAAACCTGGTGTTCCTGGGCATCGCCGTGAGCAAGCCCGACCTGATCACCTGC CTGGAGCAGGAGAAGGAGCCCTGGGAGCCCATGAGAAGACACGAGATGGTGG CCAAGCCCCCCGTGATGTGCAGCCAC (SEQ ID NO: 1084)	6.38246564
CBX5	EPEKIIGATDSCG DLMFLMKWKDT DEADLVLAKEAN VKCPQIVIAFYEE RLT (SEQ ID NO: 379)	QSNDIARGFERGLEPEKIIGATD SCGDLMFLMKWKDTDEADLVL AKEANVKCPQIVIAFYEERLTW HAYPEDAENKEKET (SEQ ID NO: 380)	CAGAGCAACGACATCGCCAGAGGCTTCGAGAGAGGCCTGGAGCCCGAGAAGAT CATCGGCGCCACCGACAGCTGCGGCGACCTGATGTTCCTGATGAAGTGGAAGG ACACCGACGAGGCCGACCTGGTGCTGGCCAAGGAGGCCAACGTGAAGTGCCCC CAGATCGTGATCGCCTTCTACGAGGAGAGACTGACCTGGCACGCCTACCCCGAG GACGCCGAGAACAAGGAGAAGGAGACC (SEQ ID NO: 1085)	6.36905016
ZN589	PVTFEDVAVLFTE AEWKRLSLEQRN LYKEVMLENLRN LVS (SEQ ID NO: 381)	ALPAKDSAWPWEEKPRYLGPV TFEDVAVLFTEAEWKRLSLEQR NLYKEVMLENLRNLVSLAESKPE VHTCPSCPLAFGSQ (SEQ ID NO: 382)	GCCCTGCCCGCCAAGGACAGCGCCTGGCCCTGGGAGGAGAAGCCCAGATACCT GGGCCCCGTGACCTTCGAGGACGTGGCCGTGCTGTTCACCGAGGCCGAGTGGA AGAGACTGAGCCCTGGAGCCAGAGAAAACCTGTACAAGGAGAGGTGATGCTGGAGAA CCTGAGAAACCTGGTGAGCCTGGCCGAGAGCAAGCCCGAGGTGCACACCTGCC CCAGCTGCCCCCTGGCCTTCGGCAGCCAG (SEQ ID NO: 1086)	6.36425087
ZNF10	LVTFKDVFVDFTR EEWKLLDTAQQI VYRNVMLENYKN LVSL (SEQ ID NO: 383)	DAKSLTAWSRTLVTFKDVFVDF TREEWKLLDTAQQIVYRNVML ENYKNLVSLGYQLTKPDVILRLE KGEEPWLVEREIHQ (SEQ ID NO: 384)	GACGCCAAGAGCCTGACCGCCTGGAGCAGAACCCTGGTGACCTTCAAGGACGT GTTCGTGGACTTCACCAGAGAGGAGTGGAAGCTGCTGGACACCGCCCAGCAGA TCGTGTACAGAAACGTGATGCTGGAGAACTACAAGAACCTGGTGAGCCTGGGC TACCAGCTGACCAAGCCCGACGTGATCCTGAGACTGGAGAAGGGCGAGGAGCC CTGGCTGGTGGAGAGAGAGATCCACCAG (SEQ ID NO: 1087)	6.36134473

ZN563	AVAFEDVAVNFT QEEWALLGPSQK NLYRYVMQETIR NLDC (SEQ ID NO: 385)	DAVAFEDVAVNFTQEEWALLG PSQKNLYRYVMQETIRNLDCIR MIWEEQNTEDQYKNPRRNLRC HMVERFSESKDSSQCG (SEQ ID NO: 386)	GACGCCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGGGCCCCAGCCAGAAGAACCTGTACAGATACGTGATGCAGGAGACCA TCAGAAACCTGGACTGCATCAGAATGATCTGGGAGGAGCAGAACACCGAGGAC CAGTACAAGAACCCCAGAAGAAACCTGAGATGCCACACATGGTGGAGAGATTCAG CGAGAGCAAGGACAGCAGCCAGTGCGGC (SEQ ID NO: 1088)	6.3562145
ZN561	SVTFDDVAVDFT PEEWALLDTTEK YLYRDVMLENYM NLASV (SEQ ID NO: 387)	EKTKVERMVEDYLASGYQDSVT FDDVAVDFTPEEWALLDTTEKY LYRDVMLENYMNLASVEWEIQ PRTKRSSLQQGFLKN (SEQ ID NO: 388)	GAGAAGACCAAGGTGGAGAGAATGGTGGAGGACTACCTGGCCAGCGGCTACC AGGACAGCGTGACCTTCGACGACGTGGCCGTGGACTTCACCCCCGAGGAGTGG GCCCTGCTGGACACCACCGAGAAGTACCTGTACAGAGATGTGATGCTGGAGAA CTACATGAACCTGGCCAGCGTGGAGTGGGAGATCCAGCCCAGAACCAAGAGAA GCAGCCTGCAGCAGGGCTTCCTGAAGAAC (SEQ ID NO: 1089)	6.3525504
ZN136	SVAFEDVDVNFT QEEWALLDPSQK NLYRDVMWETM RNLASI (SEQ ID NO: 389)	DSVAFEDVDVNFTQEEWALLD PSQKNLYRDVMWETMRNLASI GKKWKDQNIKDHYKHRGRNLR SHMLERLYQTKDGSQRG (SEQ ID NO: 390)	GACAGCGTGGCCTTCGAGGACGTGGACGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGGACCCCAGCCAGAAGAACCTGTACAGAGATGTGATGTGGGAGACCA TGAGAAACCTGGCCAGCATCGGCAAGAAGTGGAAGGACCAGAACACATCAAGGAC CACTACAAGCACAGAGGCAGAAACCTGAGAAGCCACATGCTGGAGAGACTGTA CCAGACCAAGGACGGCAGCCAGAGAGGC (SEQ ID NO: 1090)	6.35103846
ZN630	PVTFEDVAVDFT QEEWQQLNPAQ KTLHRDVMLETY NHLVSV (SEQ ID NO: 391)	IESQEPVTFEDVAVDFTQEEWQ QLNPAQKTLHRDVMLETYNHL VSVGCSGIKPDVIFKLEHGKDP WIESELSRWIYPDR (SEQ ID NO: 392)	ATCGAGAGCCAGGAGCCCGTGACCTTCGAGGACGTGGCCGTGGACTTCACCCA GGAGGAGTGGCAGCAGCTGAACCCCGCCCAGAAGACCCTGCACAGAGATGTGA TGCTGGAGACCTACAACCACCTGGTGAGCGTGGGCCTGCAGCGGCATCAAGCCC GACGTGATCTTCAAGCTGGAGCACGGCAAGGACCCCTGGATCATCGAGAGCGA GCTGAGCAGATGGATCTACCCCGACAGA (SEQ ID NO: 1091)	6.34648094
ZN527	LVTFRDVALDFSQ EEWEWLKPSQK DLYRDVMLENYR NLVWL (SEQ ID NO: 393)	AVGLCKAMSQGLVTFRDVALD FSQEEWEWLKPSQKDLYRDVM LENYRNLVWLGLSISKPNMISLL EQGKEPWMVERKMSQ (SEQ ID NO: 394)	GCCGTGGGCCTGTGCAAGGCCATGAGCCAGGGCCTGGTGACCTTCAGAGATGT GGCCCTGGACTTCAGCCAGGAGGAGTGGGAGTGGCTGAAGCCCAGCCAGAAG GACCTGTACAGAGATGTGATGCTGGAGAACTACAGAAACCTGGTGTGGCTGGG CCTGAGCATCAGCAAGCCCAACATGATCAGCCTGCTGGAGCAGGGCAAGGAGC CCTGGATGGTGGAGAGAAAGATGAGCCAG (SEQ ID NO: 1092)	6.34024936
ZN333	PVTFADVAVVFT PEEWVFLDSTQR S1YRDVMLENYR NLAS (SEQ ID NO: 395)	DKVEEEAMAPGLPTACSQEPVT FADVAVVFTPEEWVFLDSTQRS LYRDVMLENYRNLASVADQLC KPNALSYLEERGEQW (SEQ ID NO: 396)	GACAAGGTGGAGGAGGAGGCCATGGCCCCCGGCCTGCCCACCGCTTGCAGCCA AGAGCCTGTGACCTTCGCCGACGTGGCCGTGGTGTTCACCCCCGAGGAGTGGGT GTTCCTGGACAGCACCCAGAGAAGCCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGGCCAGCGTGGCCGACCAGCTGTGCAAGCCCAACGCCCTGAGC TACCTGGAGGAGAGAGGCGAGCAGTGG (SEQ ID NO: 1093)	6.33883721

Z324B	MTFEDVAVYFSQ EEWGLLDTAQRA LYRHVMLENFTL VTSL (SEQ ID NO: 397)	TFEDVAVYFSQEEWGLLDTAQR ALYRHVMLENFTLVTSLGLSTSR PRVVIQLERGEEPWVPSGKDM TLARNTYGRLNSGS (SEQ ID NO: 398)	ACCTTCGAGGACGTGGCCGTGTACTTCAGCCAGGAGGAGTGGGGCCTGCTGGA CACCGCCCAGAGAGCCCTGTACAGACACGTGATGCTGGAGAACTTCACCCTGGT GACCAGCCTGGGCCTGAGCACCAGCAGACCCAGAGTGGTGATCCAGCTGGAGA GAGGCGAGGAGCCCTGGGTGCCCAGCGGCAAGGACATGACCCTGGCCAGAAA CACCTACGGCAGACTGAACAGCGGCAGC (SEQ ID NO: 1094)	6.33798774
ZN786	PLTFEDVAIYFSE QEWQDLEAWQ KELYKHVMRSNY ETLVSL (SEQ ID NO: 399)	AEPPRLPLTFEDVAIYFSEQEWQ DLEAWQKELYKHVMRSNYETL VSLDDGLPKPELISWIEHGGEPF RKWRESQKSGNII (SEQ ID NO: 400)	GCCGAGCCCCCCAGACTGCCCCTGACCTTCGAGGACGTGGCCATCTACTTCAGC GAGCAGGAGTGGCAGGACCTGGAGGCCTGGCAGAAGGAGCTGTACAAGCACG TGATGAGAAGCAACTACGAGACCCTGGTGAGCCTGGACGACGGCCTGCCCAAG CCCGAGCTGATCAGCTGGATCGAGCACGGCGGCGAGCCCTTCAGAAAGTGGAG AGAGAGCCAGAAGAGCGGCAACATCATC (SEQ ID NO: 1095)	6.31659272
ZN709	SVVFEDVAVNFT QEEWALLGPSQK KLYRDVMQETFV NLASI (SEQ ID NO: 401)	DSVVFEDVAVNFTQEEWALLG PSQKKLYRDVMQETFVNLASIG ENWEEKNIEDHKNQGRKLRSH MVERLCERKEGSQFGE (SEQ ID NO: 402)	GACAGCGTGGTGTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGGGCCCCAGCCAGAAGAAGCTGTACAGAGATGTGATGCAGGAGACCT TCGTGAACCTGGCCAGCATCGGCGAGAACTGGGAGGAGAAGAACATCGAGGA CCACAAGAACCAGGGCAGAAAGCTGAGAAGCCACATGGTGGAGAGACTGTGC GAGAGAAAGGAGGGCAGCCAGTTCGGCGAG (SEQ ID NO: 1096)	6.31480293
ZN792	CVTFEDVTIYFSQ EEWVLLDEAQRL LYCDVMLENFALI ASL (SEQ ID NO: 403)	AAAALRDPAQGCVTFEDVTIYF SQEEWVLLDEAQRLLYCDVML ENFALIASLGLISFRSHIVSQLEM GKEPWVPDSVDMT(SEQ ID NO: 404)	GCCGCCGCCGCCCTGAGAGACCCCGCCCAGGGCTGCGTGACCTTCGAAGACGT GACCATCTACTTCAGCCAGGAGGAGTGGGTGCTGCTGGACGAGGCCCAGAGAC TGCTGTACTGCGACGTGATGCTGGAGAACTTCGCCCTGATCGCCAGCCTGGGCC TGATCAGCTTCAGAAGCCACATCGTGAGCCAGCTGGAGATGGGCAAGGAGCCC TGGGTGCCCGACAGCGTGGACATGACC (SEQ ID NO: 1097)	6.29907418
ZN599	LVSFEDVVVTFTG EEWGHLDLAQRT LYQEVMLETCRLL VSL (SEQ ID NO: 405)	AAPALALVSFEDVVVTFTGEEW GHLDLAQRTLYQEVMLETCRLL VSLGHPVPKPELIYLLEHGQELW TVKRGLSQSTCAG (SEQ ID NO: 406)	GCCGCCCCCGCCCTGGCCCTGGTGAGCTTCGAGGACGTGGTGGTGACCTTCACC GGCGAGGAGTGGGGCCACCTGGACCTGGCCCAGAGAACCCTGTACCAGGAGGT GATGCTGGAGACCTGCAGACTGCTGGTGAGCCTGGGCCACCCCGTGCCCAAGC CCGAGCTGATCTACCTGCTGGAGCACGGCCAGGAGCTGTGGACCGTGAAGAGA GGCCTGAGCCAGAGCACCTGCGCCGGC (SEQ ID NO: 1098)	6.29676005
ZN613	SLTLEDVAVEFT WEEWQLLGPAQ KDLYRDVMLENY SNLVSV (SEQ ID NO: 407)	IKSQESLTLEDVAVEFTWEEWQ LLGPAQKDLYRDVMLENYSNLV SVGYQASKPDALFKLEQGEPWT VENEIHSQICPEIK (SEQ ID NO: 408)	ATCAAGAGCCAGGAGAGCCTGACCCTGGAGGACGTGGCCGTGGAGTTCACCTG GGAGGAGTGGCAGCTGCTGGGCCCCGCCCAGAAGGACCTGTACAGAGATGTGA TGCTGGAGAACTACAGCAACCTGGTGAGCGTGGGCTACCAGGCCAGCAAGCCC GACGCCCTGTTCAAGCTGGAGCAGGGCGAGCCCTGGACCGTGGAGAACGAGAT CCACAGCCAGATCTGCCCCGAGATCAAG (SEQ ID NO: 1099)	6.28970926

ZF69B	LLTFKDVSVDFTQ EEWGQLAPAHR NLYREVMLENYG NLVSV (SEQ ID NO: 409)	GESLESRVTLGSLTAESQELLTFK DVSVDFTQEEWGQLAPAHRNL YREVMLENYGNLVSVGCQLSKP GVISQLEKGEEPW (SEQ ID NO: 410)	GGCGAGAGCCTGGAGAGCAGAGTGACCCTGGGCAGCCTGACCGCCGAGAGCC AGGAGCTGCTGACCTTCAAGGACGTGAGCGTGGACTTCACCCAGGAGGAGTGG GGCCAGCTGGCCCCCGCCCACAGAAACCTGTACAGAGAGGGTGATGCTGGAGAA CTACGGCAACCTGGTGAGCGTGGGCTGCCAGCTGAGCAAGCCCGGCGTGATCA GCCAGCTGGAGAAGGGCGAGGAGCCCTGG (SEQ ID NO: 1100)	6.28648867
ZN799	SVALEDVAVNFT REEWALLGPCQK NLYKDVMQETIR NLDCV (SEQ ID NO: 411)	ASVALEDVAVNFTREEWALLGP CQKNLYKDVMQETIRNLDCVG MKWKDQNIEDQYRYPRKNLRC RMLERFVESKDGTQCG (SEQ ID NO: 412)	GCCAGCGTGGCCCTGGAGGACGTGGCCGTGAACTTCACCAGAGAGGAGTGGGC CCTGCTGGGCCCCTGCCAGAAGAACCTGTACAAGGACGTGATGCAGGAGACCA TCAGAAACCTGGACTGCGTGGGCATGAAGTGGAAGGACCAGAACATCGAGGAC CAGTACAGATACCCCAGAAAGAACCTGAGATGCAGAATGCTGGAGAGATTCGT GGAGAGCAAGGACGGCACCCAGTGCGGC (SEQ ID NO: 1101)	6.28580406
ZN569	TVTFKDVAIDFTQ EEWKRLDPAQRK LYRNVMLENYNN LITV (SEQ ID NO: 413)	TESQGTVTFKDVAIDFTQEEWK RLDPAQRKLYRNVMLENYNNLI TVGYPFTKPDVIFKLEQEEEPW VMEEEVLRRHWQGE (SEQ ID NO: 414)	ACCGAGAGCCAGGGCACCGTGACCTTCAAGGACGTGGCCATCGACTTCACCCAG GAGGAGTGGAAGAGACTGGACCCCGCCCAGAGAAAGCTGTACAGAAACGTGA TGCTGGAGAACTACAACAACCTGATCACCGTGGGCTACCCCTTCACCAAGCCCG ACGTGATCTTCAAGCTGGAGCAGGAGGAGGCCCTGGGTGATGGAGGAGGA GGTGCTGAGAAGACACTGGCAGGGCGAG (SEQ ID NO: 1102)	6.28572758
ZN564	SVASEDVAVNFT LEEWALLDPSQK KLYRDVMRETFR NLACV (SEQ ID NO: 415)	DSVASEDVAVNFTLEEWALLDP SQKKLYRDVMRETFRNLACVGK KWEDQSIEDWYKNQGRILRNH MEEGLSESKEYDQCG (SEQ ID NO: 416)	GACAGCGTGGCCAGCGAGGACGTGGCCGTGAACTTCACCCTGGAGGAGTGGGC CCTGCTGGACCCCAGCCAGAAGAAGCTGTACAGAGATGTGATGAGAGACCT TCAGAAACCTGGCCTGCGTGGGCAAGAAGTGGGAGGACCAGAGCATCGAGGA CTGGTACAAGAACCAGGGCAGAATCCTGAGAAACCACATGGAGGAGGGCCTGA GCGAGAGCAAGGAGTACGACCAGTGCGGC (SEQ ID NO: 1103)	6.28268424
ZN546	SLAFRDVSIDLSQ EEWECLDAVQR DLYKDVMLENYS NLVSL (SEQ ID NO: 417)	EETQGELTSSCGSKTMANVSLA FRDVSIDLSQEEWECLDAVQRD LYKDVMLENYSNLVSLGYTIPKP DVITLLEQEKEPW (SEQ ID NO: 418)	GAGGAGACCCAGGGCGAGCTGACCAGCAGCTGCGGCAGCAAGACCATGGCCA ACGTGAGCCTGGCCTTCAGAGATGTGAGCATCGACCTGAGCCAGGAGGAGTGG GAGTGCCTGGACGCCGTGCAGAGAGACCTGTACAAGGACGTGATGCTGGAGAA CTACAGCAACCTGGTGAGCCTGGGCTACACCATCCCCAAGCCCGACGTGATCAC CCTGCTGGAGCAGGAGAAGGAGCCCTGG (SEQ ID NO: 1104)	6.27774396
ZFP92	PVSFEDVSVYFTK TEWKLLDLRQKV LYKRVMLENYSH LVSL (SEQ ID NO: 419)	AAILLTTRPKVPVSFEDVSVYFTK TEWKLLDLRQKVLYKRVMLENY SHLVSLGFSFSKPHLISQLERGE GPWVADIPRTW (SEQ ID NO: 420)	GCCGCCATCCTGCTGACCACCAGACCCAAGGTGCCCGTGAGCTTCGAGGACGTG AGCGTGTACTTCACCAAGACCGAGTGGAAGCTGCTGGACCTGAGACAGAAGGT GCTGTACAAGAGAGTGATGCTGGAGAACTACAGCCACCTGGTGAGCCTGGGCT TCAGCTTCAGCAAGCCCCACCTGATCAGCCAGCTGGAGAGAGGCGAGGGCCCC TGGGTGGCCGACATCCCCAGAACCTGG (SEQ ID NO: 1105)	6.273403

YAF2	RPRLKNVDRSSA QHLEVTVGDLTVI ITDFKEK (SEQ ID NO: 421)	KDKVEKEKSEKETTSKKNSHKKT RPRLKNVDRSSAQHLEVTVGDL TVIITDFKEKTKSPPASSAASAD QHSQSGSSSDNT (SEQ ID NO: 422)	AAGGACAAGGTGGAGAAGGAGAAGAGCGAGAAGGAGACCACCAGCAAGAAG AACAGCCACAAGAAGACCAGACCCAGACTGAAGAACGTGGACACAGAAGCAGCGC CCAGCACCTGGAGGTGACCGTGGGCGACCTGACCGTGATCATCACCGACTTCAA GGAGAAGACCAAGAGCCCCCCCGCCAGCAGCGCCGCCAGCGCTGACCAACACA GCCAAAGCGGAAGCAGCAGCGACAACACC(SEQ ID NO: 1106)	6.25768891
ZN723	PLTFTDVAIKFSLE EWQFLDTAQQN LYRDVMLENYRN LVFL (SEQ ID NO: 423)	GPLTFTDVAIKFSLEEWQFLDTA QQNLYRDVMLENYRNLVFLGV GVSKPDLITCLEQGKEPWNMK RHKMVAKPPVVCSHF (SEQ ID NO: 424)	GGCCCCCTGACCTTCACCGACGTGGCCATCAAGTTCAGCCTGGAGGAGTGGCAG TTCCTGGACACCGCCCAGCAGAACCTGTACAGAGATGTGATGCTGGAGAACTAC AGAAACCTGGTGTTCCTGGGCGTGGGCGTGAGCAAGCCCGACCTGATCACCTGC CTGGAGCAGGGCAAGGAGCCCTGGAACATGAAGAGACACAAGATGGTGGCCA AGCCCCCCGTGGTGTGCAGCCACTTC (SEQ ID NO: 1107)	6.25047465
ZNF34	EVTFEDVAVYLSR EEWGRLGPAQR GLYRDVMLETYG NLVSL (SEQ ID NO: 425)	RKPNPQAMAALFLSAPPQAEV TFEDVAVYLSREEWGRLGPAQR GLYRDVMLETYGNLVSLGVGPA GPKPGVISQLERGDE (SEQ ID NO: 426)	AGAAAGCCCAACCCCCAGGCCATGGCCGCCCTGTTCCTGAGCGCCCCCCCTCAA GCCGAAGTGACCTTCGAGGACGTGGCCGTGTACCTGAGCAGAGAGGAGTGGG GCAGACTGGGCCCCGCCCAGAGAGGCCTGTACAGAGATGTGATGCTGGAGACC TACGGCAACCTGGTGAGCCTGGGCGTGGGCCCCGGCGGACCCAAACCTGGAGT GATCAGCCAGCTGGAGAGAGGCGACGAG (SEQ ID NO: 1108)	6.23513709
ZN439	PVAFKDVAVNFT QEEWALLDISQK NLYREVMLETFW NLTSI (SEQ ID NO: 427)	LSLSPILLYTCEMFQDPVAFKDV AVNFTQEEWALLDISQKNLYRE VMLETFWNLTSIGKKWKDQNI EYEYQNPRRNFRSV (SEQ ID NO: 428)	CTGAGCCTGAGCCCCATCCTGCTGTACACCTGCGAGATGTTCCAGGACCCCGTG GCCTTCAAGGACGTGGCCGTGAACTTCACCCAGGAGGAGGTGGGCCCTGCTGGA CATCAGCCAGAAGAACCTGTACAGAGAGGTGATGCTGGAGACCTTCTGGAACCT GACCAGCATCGGCAAGAAGTGGAAGGACCAGAACATCGAGTACGAGTACCAGA ACCCCAGAAGAAACTTCAGAAGCGTG (SEQ ID NO: 1109)	6.22934428
ZFP57	PVTFEDVAVNFT QEEWDCLDASQ RVLYQDVMSETF KNLTS (SEQ ID NO: 429)	AAGEPRSLLFFQKPVTFEDVAV NFTQEEWDCLDASQRVLYQDV MSETFKNLTSVARIFLHKPELITK LEQEEEQWRETRV (SEQ ID NO: 430)	GCCGCCGGCGAGCCCAGAAGCCTGCTGTTCTTCCAGAAGCCCGTGACCTTCGAG GACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGACTGCCTGGACGCCAGCCA GAGAGTGCTGTACCAGGACGTGATGAGCGAGACCTTCAAGAACCTGACCAGCG TGGCCAGAATCTTCCTGCACAAGCCCGAGCTGATCACCAAGCTGGAGCAGGAG GAGGAGCAGTGGAGAGAGAGACCAGAGTG (SEQ ID NO: 1110)	6.2234497
ZNF19	MVTFEDVAVHFT KTEWTGLSPAQR ALYRSVMLENFG NLTAL (SEQ ID NO: 431)	AAMPLKAQYQEMVTFEDVAV HFTKTEWTGLSPAQRALYRSV MLENFGNLTALGYPVPKPALISL LERGDMAWGLEAQDDP (SEQ ID NO: 432)	GCCGCCATGCCCCTGAAGGCCCAGTACCAGGAGATGGTGACCTTCGAGGACGT GGCCGTGCACTTCACCAAGACCGAGTGGACCGGCCTGAGCCCCGCCCAGAGAG CCCTGTACAGAAGCGTGATGCTGGAGAACTTCGGCAACCTGACCGCCCTGGGCT ACCCCGTGCCCAAGCCCGCCCTGATCAGCCTGCTGGAGAGAGGCGACATGGCCT GGGGCCTGGAGGCCCAGGACGACCCC (SEQ ID NO: 1111)	6.21632085

ZN404	PLTFSDVAIDFSQ EEWEYLNSDQRD LYRDVMLENYTN LVSL (SEQ ID NO: 433)	ARVPLTFSDVAIDFSQEEWEYL NSDQRDLYRDVMLENYTNLVSL DFNFTTESNKLSSEKRNYEVNAY HQETWKRNKTFNL (SEQ ID NO: 434)	GCCAGAGTGCCCCTGACCTTCAGCGACGTGGCCATCGACTTCAGCCAGGAGGA GTGGGAGTACCTGAACAGCGACCAGAGAGACCTGTACAGAGATGTGATGCTGG AGAACTACACCAACCTGGTGAGCCTGGACTTCAACTTCACCACCGAGAGCAACA AGCTGAGCAGCGAGAAGAGAACTACGAGGTGAACGCCTACCACCAGGAGACC TGGAAGAGAAACAAGACCTTCAACCTG (SEQ ID NO: 1112)	6.20126205
ZN274	PVTFEDVTLGFTP EEWGLLDLKQKS LYREVMLENYRN LVSV (SEQ ID NO: 435)	ASRLPTAWSCEPVTFEDVTLGF TPEEWGLLDLKQKSLYREVMLE NYRNLVSVEHQLSKPDVVSQLE EAEDFWPVERGIPQ (SEQ ID NO: 436)	GCCAGCAGACTGCCCACCGCCTGGAGCTGCGAGCCCGTGACCTTCGAGGACGT GACCCTGGGCTTCACCCCCGAGGAGTGGGGCCTGCTGGACCTGAAGCAGAAGA GCCTGTACAGAGAGGTGATGCTGGAGAACTACAGAAACCTGGTGAGCGTGGAG CACCAGCTGAGCAAGCCCGACGTGGTGAGCCAGCTGGAGGAGGCCGAGGACTT CTGGCCCGTGGAGAGAGGCATCCCCCAG (SEQ ID NO: 1113)	6.19652061
CBX3	DPERIIGATDSSG ELMFLMKWKDS DEADLVLAKEAN MKCPQIVIAFYEE RLT (SEQ ID NO: 437)	SKKKRDAADKPRGFARGLDPER IIGATDSSGELMFLMKWKDSDE ADLVLAKEANMKCPQIVIAFYE ERLTWHSCPEDEAQ (SEQ ID NO: 438)	AGCAAGAAGAAGAGAGATGCCGCCGACAAGCCCAGAGGCTTCGCCAGAGGCCT GGACCCCGAGAGAATCATCGGCGCCACCGACAGCAGCGGCGAGCTGATGTTCC TGATGAAGTGGAAGGACAGCGACGAGGCCGACCTGGTGCTGGCCAAGGAGGC CAACATGAAGTGCCCCCAGATCGTGATCGCCTTCTACGAGGAGAGACTGACCTG GCACAGCTGCCCCGAGGACGAGGCCCAG (SEQ ID NO: 1114)	6.19641648
ZNF30	SVTFEDVAIAFSQ QEWESLDSSQRG LYRDVMLENYRN LVSM (SEQ ID NO: 439)	AHKYVGLQYHGSVTFEDVAIAF SQQEWESLDSSQRGLYRDVML ENYRNLVSMGHSRSKPHVIALL EQWKEPEVTVRKDGR (SEQ ID NO: 440)	GCCCACAAGTACGTGGGCCTGCAGTACCACGGCAGCGTGACCTTCGAGGACGT GGCCATCGCCTTCAGCCAGCAGGAGTGGGAGAGCCTGGACAGCAGCCAGAGA GGCCTGTACAGAGATGTGATGCTGGAGAACTACAGAAACCTGGTGAGCATGGG CCACAGCAGAAGCAAGCCCCACGTGATCGCCCTGCTGGAGCAGTGGAAGGAGC CCGAGGTGACCGTGAGAAAGGACGGCAGA(SEQ ID NO: 1115)	6.19503476
ZN250	KLTFEDVAVLLSQ DEWDRLCPAQR GLYRNVMMETY GNVVSL (SEQ ID NO: 441)	AAARLLPVPAGPQPLSFQAKLT FEDVAVLLSQDEWDRLCPAQR GLYRNVMMETYGNVVSLGLPG SKPDIISQLERGEDPW (SEQ ID NO: 442)	GCCGCCGCCAGACTGCTGCCCGTGCCCGCCGGCCCCCAACCTCTTAGCTTCCAG GCCAAGCTGACCTTCGAGGACGTGGCCGTGCTGCTGAGCCAGGACGAGTGGGA CAGACTGTGCCCCGCCCAGAGAGGCCTGTACAGAAACGTGATGATGGAGACCT ACGGCAACGTGGTGAGCCTGGGCCTGCCCGGCAGCAAGCCCGACATCATCAGC CAGCTGGAGAGAGAGGCGAGGACCCCTGG (SEQ ID NO: 1116)	6.17058573
ZN570	LVTFRDVAVDFS QEEWDCLDSSQR HLYSNVMLENYR ILVSL (SEQ ID NO: 443)	AVGLLKAMYQELVTFRDVAVD FSQEEWDCLDSSQRHLYSNVM LENYRILVSLGLCFSKPSVILLLEQ GKAPWMVKRELTK (SEQ ID NO: 444)	GCCGTGGGCCTGCTGAAGGCCATGTACCAGGAGCTGGTGACCTTCAGAGATGT GGCCGTGGACTTCAGCCAGGAGGAGTGGGACTGCCTGGACAGCAGCCAGAGA CACCTGTACAGCAACGTGATGCTGGAGAACTACAGAATCCTGGTGAGCCTGGGC CTGTGCTTCAGCAAGCCCAGCGTGATCCTGCTGCTGGAGCAGGGCAAGGCCCCC TGGATGGTGAAGAGAGAGCTGACCAAG (SEQ ID NO: 1117)	6.16932644

ZN675	LLTFRDVAIEFSLE EWQCLDTAQRN LYKNVILENYRNL VFL (SEQ ID NO: 445)	GLLTFRDVAIEFSLEEWQCLDTA QRNLYKNVILENYRNLVFLGIAV SKQDLITCLEQEKEPLTVKRHEM VNEPPVMCSHF (SEQ ID NO: 446)	GGCCTGCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA GTGCCTGGACACCGCCCAGAGAAACCTGTACAAGAACGTGATCCTGGAGAACT ACAGAAACCTGGTGTTCCTGGGCATCGCCGTGAGCAAGCAGGACCTGATCACCT GCCTGGAGCAGGAGAAGGAGCCCCTGACCGTGAAGAGACACGAGATGGTGAA CGAGCCCCCCGTGATGTGCAGCCACTTC (SEQ ID NO: 1118)	6.15995772
ZN695	LLAFRDVALEFSP EEWECLDPAQRS LYRDVMLENYRN LISL (SEQ ID NO: 447)	GLLAFRDVALEFSPEEWECLDP AQRSLYRDVMLENYRNLISLGE DSFNMQFLFHSLAMSKPELIICL EARKEPWNVNTEK (SEQ ID NO: 448)	GGCCTGCTGGCCTTCAGAGATGTGGCCCTGGAGTTCAGCCCCGAGGAGTGGGA GTGCCTGGACCCCGCCCAGAGAAGCCTGTACAGAGATGTGATGCTGGAGAACT ACAGAAACCTGATCAGCCTGGGCGAGGACAGCTTCAACATGCAGTTCCTGTTCC ACAGCCTGGCCATGAGCAAGCCCGAGCTGATCATCTGCCTGGAGGCCAGAAAG GAGCCCTGGAACGTGAACACCGAGAAG (SEQ ID NO: 1119)	6.15609798
ZN548	RVVFEDVAIYFSQ EEWGHLDEAQRL LYRDVMLENLAL LSSL (SEQ ID NO: 449)	NLTEGRVVFEDVAIYFSQEEWG HLDEAQRLLYRDVMLENLALLS SLGSWHGAEDEEAPSQQGFSV GVSEVTASKPCLSSQ (SEQ ID NO: 450)	AACCTGACCGAGGGCAGAGTGGTGTTCGAGGACGTGGCCATCTACTTCAGCCA GGAGGAGTGGGGCCACCTGGACGAGGCCCAGAGACTGCTGTACAGAGATGTG ATGCTGGAGAACCTGGCCCTGCTGAGCAGCCTGGGCAGCTGGCACGGCGCCGA GGACGAAGAGGCCCCTAGCCAGCAGGGCTTCAGCGTGGGCGTGAGCGAGGTG ACCGCCAGCAAGCCCTGCCTGAGCAGCCAG (SEQ ID NO: 1120)	6.14238152
ZN227	AVTFKDVAVVFS REELRLLDLTQRK LYRDVMVENFKN LVAV (SEQ ID NO: 451)	SQNYDLPQKKQEKMTKFQEAV TFKDVAVVFSREELRLLDLTQRK LYRDVMVENFKNLVAVGHLPF QPDMVSQLEAEEKLW (SEQ ID NO: 452)	AGCCAGAACTACGACCTGCCCCAGAAGAAGCAGGAGAAGATGACCAAGTTCCA GGAGGCCGTGACCTTCAAGGACGTGGCCGTGGTGTTCAGCAGAGAGGAGCTGA GACTGCTGGACCTGACCCAGAGAAAGCTGTACAGAGATGTGATGGTGGAGAAC TTCAAGAACCTGGTGGCCGTGGGCCACCTGCCCTTCCAGCCCGACATGGTGAGC CAGCTGGAGGCCGAGGAGAAGCTGTGG (SEQ ID NO: 1121)	6.13508917
ZN132	MVTFEDVAVYFS QEEWELLDAAQR HLYHSVMLENLE LVTSL (SEQ ID NO: 453)	GPAQHTSWPCGSAVPTLKSMV TFEDVAVYFSQEEWELLDAAQR HLYHSVMLENLELVTSLGSWH GVEGEGAHPKQNVSVE (SEQ ID NO: 454)	GGCCCCGCCCAGCACACCAGCTGGCCCTGCGGCAGCGCCGTTCCTACACTGAAG AGCATGGTGACCTTCGAGGACGTGGCCGTGTACTTCAGCCAGGAGGAGTGGGA GCTGCTGGACGCCGCCCAGAGACACCTGTACCACAGCGTGATGCTGGAGAACCT GGAGCTGGTGACCAGCCCTGGGCAGCTGGCACGGCGTGGAGGGCGAGGGAGCT CACCCCAAGCAGAACGTGAGCGTGGAG (SEQ ID NO: 1122)	6.13316124
ZN738	PLTFRDVVIEFSQ EEWQCLDTAQQ DLYRKVMLENFR NLVFL (SEQ ID NO: 455)	SGYPGAERNLLEYSYFEKGPLTF RDVVIEFSQEEWQCLDTAQQD LYRKVMLENFRNLVFLGIDVSKP DLITCLEQGKDPW (SEQ ID NO: 456)	AGCGGCTACCCCGGCGCCGAGAGAAACCTGCTGGAGTACAGCTACTTCGAGAA GGGCCCCCTGACCTTCAGAGATGTGGTGATCGAGTTCAGCCAGGAGGAGTGGC AGTGCCTGGACACCGCCCAGGACCTGTACAGAAAGGTGATGCTGGAGAAC TTCAGAAACCTGGTGTTCCTGGGCATCGACGTGAGCAAGCCCGACCTGATCACC TGCCTGGAGCAGGGCAAGGACCCCTGG (SEQ ID NO: 1123)	6.12742065

ZN420	LVMFRDVAIDFS QEEWECLDSAQR DLYRDVMLENYS NLVSL (SEQ ID NO: 457)	ARKLVMFRDVAIDFSQEEWECL DSAQRDLYRDVMLENYSNLVSL DLPSRCASKDLSPEKNTYETELS QWEMSDRLENCDL (SEQ ID NO: 458)	GCCAGAAAGCTGGTGATGTTCAGAGATGTGGCCATCGACTTCAGCCAGGAGGA GTGGGAGTGCCTGGACAGCGCCCAGAGAGACCTGTACAGAGATGTGATGCTGG AGAACTACAGCAACCTGGTGAGCCTGGACCTGCCCAGCAGATGCGCCAGCAAG GACCTGAGCCCCGAGAAGAACACCTACGAGACCGAGCTGAGCCAGTGGGAGAT GAGCGACAGACTGGAGAACTGCGACCTG (SEQ ID NO: 1124)	6.1074573
ZN514	MTFEDVAVEFSQ WEWGQLNPAQ KDLYREVMLENF RNLAIL (SEQ ID NO: 459)	TFEDVAVEFSQWEWGQLNPA QKDLYREVMLENFRNLAILGLLV SKPYVICQLEEGGEPFMVEREIS TGAHSDWKRRSKSK (SEQ ID NO: 460)	ACCTTCGAGGACGTGGCCGTGGAGTTCAGCCAGTGGGAGTGGGGCCAGCTGAA CCCCGCCCAGAAGGACCTGTACAGAGAGGGTGATGCTGGAGAACTTCAGAAACC TGGCCATCCTGGGCCTGCTGGTGAGCAAGCCCTACGTGATCTGCCAGCTGGAGG AGGGCGGCGACGCCTTCATGGTGGAGAGAGATCAGCACCGGCGCCCACAGC GACTGGAAGAGAAGAAGCAAGAGCAAG (SEQ ID NO: 1125)	6.10685195
ZN626	PLQFRDVAIEFSL EEWHCLDTAQR NLYRNVMLENYS NLVFL (SEQ ID NO: 461)	GPLQFRDVAIEFSLEEWHCLDT AQRNLYRNVMLENYSNLVFLGI TVSKPDLITCLEQGRKPLTMKR NEMIAKPSVMCSHF (SEQ ID NO: 462)	GGCCCCCTGCAGTTCAGAGATGTGGCCATCGAGTTCAGCCTGGAGGAGTGGCA CTGCCTGGACACCGCCCAGAGAAACCTGTACAGAAACGTGATGCTGGAGAACT ACAGCAACCTGGTGTTCCTGGGCATCACCGTGAGCAAGCCCGACCTGATCACCT GCCTGGAGCAGGGCAGAAAGCCCCTGACCATGAAGAGAAACGAGATGATCGCC AAGCCCAGCGTGATGTGCAGCCACTTC (SEQ ID NO: 1126)	6.10541852
ZN806	RVTFKDIAVIFTKE ELAVLDKAQINLY QDVMLENFRNFI SV (SEQ ID NO: 463)	IKFQERVTFKDIAVIFTKEELAVL DKAQINLYQDVMLENFRNFISV DGIKNNILNLQGKGLIYLSQEEL HCWKIWKQRIR (SEQ ID NO: 464)	ATCAAGTTCCAGGAGAGAGTGACCTTCAAGGACATCGCCGTGATCTTCACCAAG GAGGAGCTGGCCGTGCTGGACAAGGCCCAGATCAACCTGTACCAGGACGTGAT GCTGGAGAACTTCAGAAACTTCATCAGCGTGGACGGCATCAAGAACAACATCCT GAACCTGCAGGGCAAGGGCCTGATCTACCTGAGCCAGGAGGAGCTGCACTGCT GGAAGATCTGGAAGCAGAGAATCAGA (SEQ ID NO: 1127)	6.09805184
ZN559	SVTFEDVAVDFT QEEWTLLDQTQR NLYRDVMLENYK NLVAV (SEQ ID NO: 465)	VAGWLTNYSQDSVTFEDVAVD FTQEEWTLLDQTQRNLYRDVM LENYKNLVAVDWESHINTKWS APQQNFLQGKTSSVVEM (SEQ ID NO: 466)	GTGGCCGGCTGGCTGACCAACTACAGCCAGGACAGCGTGACCTTCGAGGACGT GGCCGTGGACTTCACCCAGGAGGAGTGGACCCTGCTGGACCAGACCCAGAGAA ACCTGTACAGAGATGTGATGCTGGAGAACTACAAGAACCTGGTGGCCGTGGAC TGGGAGAGCCACATCAACACCAAGTGGAGCGCCCCCCCAGCAGAACTTCCTGCA GGGCAAGACCAGCAGCGTGGTGGAGATG (SEQ ID NO: 1128)	6.09618421
ZN460	SVTFEDVAVTFT QEEWGQLDVTQ RALYVEVMLETC GLLVAL (SEQ ID NO: 467)	AAAWMAPAQESVTFEDVAVTF TQEEWGQLDVTQRALYVEVML ETCGLLVALGDSTKPETVEPIPS HLALPEEVSLQEQLA (SEQ ID NO: 468)	GCCGCCGCCTGGATGGCCCCCGCCCAGGAGAGCGTGACCTTCGAGGACGTGGC CGTGACCTTCACCCAGGAGGAGTGGGGCCAGCTGGACGTGACCCAGAGAGCCC TGTACGTGGAGGTGATGCTGGAGACCTGCGGCCTGCTGGTGGCCCTGGGCGAC AGCACCAAGCCCGAGACCGTGGAGCCCATCCCCAGCCACCTGGCCCTGCCCGAG GAAGTGAGCCTGCAGGAGCAGCTGGCC (SEQ ID NO: 1129)	6.08494207

ZN268	PLSFMDVFVDFT WEEWQLLDPAQ KCLYRSVMLENY SNLVSL (SEQ ID NO: 469)	VLEWLFISQEQPKITKSWGPLSF MDVFVDFTWEEWQLLDPAQK CLYRSVMLENYSNLVSLGYQHT KPDIIFKLEQGEELC (SEQ ID NO: 470)	GTGCTGGAGTGGCTGTTCATCAGCCAGGAGCAGCCCAAGATCACCAAGAGCTG GGGCCCCCTGAGCTTCATGGACGTGTTCGTGGACTTCACCTGGGAGGAGTGGCA GCTGCTGGACCCCGCCCAGAAGTGCCTGTACAGAAGCGTGATGCTGGAGAACT ACAGCAACCTGGTGAGCCTGGGCTACCAGCACACCAAGCCCGACATCATCTTCA AGCTGGAGCAGGGCGAGGAGCTGTGC (SEQ ID NO: 1130)	6.040812
ZN304	CVTFEDVFVYFSR EEWELLEEAQRFL YRDVMLENFALV ATL (SEQ ID NO: 471)	AAAVLMDRVQSCVTFEDVFVY FSREEWELLEEAQRFLYRDVML ENFALVATLGFWCEAEHEAPSE QSVSVEGVSQVRTAE (SEQ ID NO: 472)	GCCGCCGCCGTGCTGATGGACAGAGTGCAGAGCTGCGTGACCTTCGAGGACGT GTTCGTGTACTTCAGCAGAGAGGAGTGGGAGCTGCTGGAGGAGGCCCAGAGAT TCCTGTACAGAGATGTGATGCTGGAGAACTTCGCCCTGGTGGCCACCCTGGGCT TCTGGTGCGAGGCCGAGCACGAGGCTCCCAGCGAGCAGAGCGTGAGCGTGGA GGGCGTGAGCCAGGTGAGAACCGCCGAG (SEQ ID NO: 1131)	6.03800144
ZIM2	LVTFEDVLVDFSP EELSSLSAAQRNL YREVMLENYRNL VSL (SEQ ID NO: 473)	AGSQFPDFKHLGTFLVFEELVTF EDVLVDFSPEELSSLSAAQRNLY REVMLENYRNLVSLGHQFSKPD IISRLEEEESYA (SEQ ID NO: 474)	GCCGGCAGCCAGTTCCCCGACTTCAAGCACCTGGGCACCTTCCTGGTGTTCGAG GAGCTGGTGACCTTCGAGGACGTGCTGGTGGACTTCAGCCCCGAGGAGCTGAG CAGCCTGAGCGCCGCCCAGAGAAACCTGTACAGAGAGGTGATGCTGGAGAACT ACAGAAACCTGGTGAGCCTGGGCCACCAGTTCAGCAAGCCCGACATCATCAGCA GACTGGAGGAGGAGGAGAGCTACGCC (SEQ ID NO: 1132)	6.03746453
ZN605	QISFEDVAVDFTL EEWQLLNPTQKN LYRDVMLENYSN LVFL (SEQ ID NO: 475)	IQSQISFEDVAVDFTLEEWQLLN PTQKNLYRDVMLENYSNLVFLE VWLDNPKMWLRDNQDNLKS MERGHKYDVFGKIFNS (SEQ ID NO: 476)	ATCCAGAGCCAGATCAGCTTCGAGGACGTGGCCGTGGACTTCACCCTGGAGGA GTGGCAGCTGCTGAACCCCACCCAGAAGAACCTGTACAGAGATGTGATGCTGG AGAACTACAGCAACCTGGTGTTCCTGGAGGTGTGGCTGGACAACCCCAAGATGT GGCTGAGAGACAACCAGGACAACCTGAAGAGCATGGAGAGAGGCCACAAGTA CGACGTGTTCGGCAAGATCTTCAACAGC (SEQ ID NO: 1133)	6.01346476
ZN844	LVAFEDVAVNFT QEEWSLLDPSQK NLYREVMQETLR NLASI (SEQ ID NO: 477)	DLVAFEDVAVNFTQEEWSLLDP SQKNLYREVMQETLRNLASIGE KWKDQNIEDQYKNPRNNLRSL LGERVDENTEENHCG (SEQ ID NO: 478)	GACCTGGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGAG CCTGCTGGACCCCAGCCAGAAGAACCTGTACAGAGAGGTGATGCAGGAGACCC TGAGAAACCTGGCCAGCATCGGCGAGAAGTGGAAGGACCAGAACATCGAGGA CCAGTACAAGAACCCCAGAAACAACCTGAGAAGCCTGCTGGGCGAGAGAGTGG ACGAGAACACCGAGGAGAACCACTGCGGC (SEQ ID NO: 1134)	5.98806163
SUMO5	IKLRVIGQDSSEIH FKVKMTTPLKKLKK KSYCQRQGVPVN SLRFLFEGQRIAD NHTPEELGMEEE DVIEVY (SEQ ID NO: 479)	KDEDIKLRVIGQDSSEIHFKVKM TTPLKKLKKSYCQRQGVPVNSL RFLFEGQRIADNHTPEELGMEE EDVIEVYQEQIGG (SEQ ID NO: 480)	AAGGACGAGGACATCAAGCTGAGAGTGATCGGCCAGGACAGCAGCGAGATCC ACTTCAAGGTGAAGATGACCACCCCCCTGAAGAAGCTGAAGAAGAGCTACTGCC AGAGACAGGGCGTGCCCGTGAACAGCCTGAGATTCCTGTTCGAGGGCCAGAGA ATCGCCGACAACCACACCCCCGAGGAGCTGGGCATGGAGGAGGACGTGAT CGAGGTGTACCAGGAGCAGATCGGCGGC (SEQ ID NO: 1135)	5.96583945

ZN101	SVAFEDVAVNFT QEEWALLSPSQK NLYRDVTLETFRN LASV (SEQ ID NO: 481)	DSVAFEDVAVNFTQEEWALLSP SQKNLYRDVTLETFRNLASVGIQ WKDQDIENLYQNLGIKLRSLVE RLCGRKEGNEHRE (SEQ ID NO: 482)	GACAGCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGAGCCCCAGCCAGAAGAACCTGTACAGAGATGTGACCCTGGAGACCTT CAGAAACCTGGCCAGCGTGGGCATCCAGTGGAAGGACCAGGACATCGAGAACC TGTACCAGAACCTGGGCATCAAGCTGAGAAGCCTGGTGGAGAGACTGTGCGGC AGAAAGGAGGGCAACGAGCACAGAGAG (SEQ ID NO: 1136)	5.90648424
ZN783	PVTFDDVAVYFSE LEWGKLEDWQK ELYKHVMRGNYE TLVSL (SEQ ID NO: 483)	RNFWILRLPPGSKGEAPKVPVT FDDVAVYFSELEWGKLEDWQK ELYKHVMRGNYETLVSLDYAISK PDILTRIERGEEPC (SEQ ID NO: 484)	AGAAACTTCTGGATCCTGAGACTGCCCCCCGGCAGCAAGGGCGAGGCCCCCAA GGTGCCCGTGACATTCGACGACGTGGCCGTGTACTTCAGCGAGCTGGAGTGGG GCAAGCTGGAGGACTGGCAGAAGGAGCTGTACAAGCACGTGATGAGAGGCAA CTACGAGACCCTGGTGAGCCTGGACTACGCCATCAGCAAGCCCGACATCCTGAC CAGAATCGAGAAGAGGCGAGGAGCCCTGC (SEQ ID NO: 1137)	5.87160607
ZN417	TVTFEDVAVNFS QEEWCLLSEAQR CLYRDVMLENLA LISSL (SEQ ID NO: 485)	AAAAPRRPTQQGTVTFEDVAV NFSQEEWCLLSEAQRCLYRDV MLENLALISSLGCWCGSKDEEA PCKQRISVQRESQSRT (SEQ ID NO: 486)	GCCGCCGCCGCCCCCAGAAGACCCACCCAGCAGGGCACCGTGACCTTCGAGGA CGTGGCCGTGAACTTCAGCCAGGAGGAGTGGTGCCTGCTGAGCGAGGCCCAGA GATGCCTGTACAGAGATGTGATGCTGGAGAACCTGGCCCTGATCAGCAGCCTG GGCTGCTGGTGCGGCAGCAAGGACGAGGAGGCCCCCCTGCAAGCAGAGAATCA GCGTGCAGAGAGAGAGCCAGAGCAGAACC (SEQ ID NO: 1138)	5.85910987
ZN182	LVTFEDVAVDFT QEEWQYLNPPQ RTLYRDVMLETYS NLVFV (SEQ ID NO: 487)	SGEDSGSFYSWQKAKREQGLV TFEDVAVDFTQEEWQYLNPPQ RTLYRDVMLETYSNLVFVGQQV TKPNLILKLEVEECPA (SEQ ID NO: 488)	AGCGGCGAGGACAGCGGCAGCTTCTACAGCTGGCAGAAGGCCAAGAGAGAGC AGGGCCTGGTGACCTTCGAGGACGTGGCCGTGGACTTCACCCAGGAGGAGTGG CAGTACCTGAACCCCCCCCCAGAGAACCCTGTACAGAGATGTGATGCTGGAGACC TACAGCAACCTGGTGTTCGTGGGCCAGCAGGTGACCAAGCCCAACCTGATCCTG AAGCTGGAGGTGGAGGAGTGCCCCGCC (SEQ ID NO: 1139)	5.80251318
ZN823	SVAFEDVAVNFT QEEWALLGPSQK SLYRNVMQETIR NLDCI (SEQ ID NO: 489)	DSVAFEDVAVNFTQEEWALLG PSQKSLYRNVMQETIRNLDCIE MKWEDQNIGDQCQNAKRNLR SHTCEIKDDSQCGETFG (SEQ ID NO: 490)	GACAGCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGGGCCCCAGCCAGAAGAGCCTGTACAGAAACGTGATGCAGGAGACCA TCAGAAACCTGGACTGCATCGAGATGAAGTGGGAGGACCAGAACATCGGCGAC CAGTGCCAGAACGCCAAGAGAAACCTGAGAAGCCACACCTGCGAGATCAAGGA CGACAGCCAGTGCGGCGAGACCTTCGGC (SEQ ID NO: 1140)	5.75436578
ZN177	SVTFQEVAVDFS QEEWALLDPAQK NLYKDVMLENFR NLASV (SEQ ID NO: 491)	AAGWLTTWSQNSVTFQEVAV DFSQEEWALLDPAQKNLYKDV MLENFRNLASVGYQLCRHSLIS KVDQEQLKTDERGILQG (SEQ ID NO: 492)	GCCGCCGGCTGGCTGACCACCTGGAGCCAGAACAGCGTGACCTTCCAGGAGGT GGCCGTGGACTTCAGCCAGGAGGAGTGGGCCCTGCTGGACCCCGCCCAGAAGA ACCTGTACAAGGACGTGATGCTGGAGAACTTCAGAAACCTGGCCAGCGTGGGC TACCAGCTGTGCAGACACAGCCTGATCAGCAAGGTGGACCAGGAGCAGCTGAA GACCGACGAGAGAGGCATCCTGCAGGGC (SEQ ID NO: 1141)	5.66150299

ZN197	LVMFEEVSVCFTS EEWACLGPIQRA LYWDVMLENYG NVTSL (SEQ ID NO: 493)	ENPRNQLMALMLLTAQPQELV MFEEVSVCFTSEEWACLGPIQR ALYWDVMLENYGNVTSLEWET MTENEEVTSKPSSSQR (SEQ ID NO: 494)	GAGAACCCCAGAAACCAGCTGATGGCCCTGATGCTGCTGACCGCCCAGCCCCAG GAGCTGGTGATGTTCGAGGAGGTGAGCGTGTGCTTCACCAGCGAGGAGTGGGC CTGCCTGGGCCCCATCCAGAGAGCCCTGTACTGGGACGTGATGCTGGAGAACTA CGGCAACGTGACCAGCCTGGAGTGGGAGACCATGACCGAGAACGAGGAGGTG ACCAGCAAGCCCAGCAGCAGCCAGAGA (SEQ ID NO: 1142)	5.65816459
ZN717	LVSFEEVAVHFT WEEWQDLDDA QRTLYRDVMLET YSSLVSL (SEQ ID NO: 495)	LETYNSLVSLQELVSFEEVAVHF TWEEWQDLDDAQRTLYRDVM LETYSSLVSLGHCITKPEMIFKLE QGAEPWIVEETPN (SEQ ID NO: 496)	CTGGAGACCTACAACAGCCTGGTGAGCCTGCAGGAGCTGGTGAGCTTCGAGGA GGTGGCCGTGCACTTCACCTGGGAGGAGTGGCAGGACCTGGACGACGCCCAGA GAACCCTGTACAGAGATGTGATGCTGGAGAACCTACAGCAGCCTGGTGAGCCTG GGCCACTGCATCACCAAGCCCGAGATGATCTTCAAGCTGGAGCAGGGCGCCGA GCCCTGGATCGTGGAGGAGACCCCCAAC (SEQ ID NO: 1143)	5.64802359
ZN669	SVAFEDVAVNFT QEEWALLDSSQK NLYREVMQETCR NLASV (SEQ ID NO: 497)	RHFRRPEPCREPLASPIQDSVAF EDVAVNFTQEEWALLDSSQKN LYREVMQETCRNLASVGSQWK DQNIEDHFEKPGKDI (SEQ ID NO: 498)	AGACACTTCAGAAGACCCGAGCCCTGCAGAGAGCCCCTGGCCAGCCCCATCCAG GACAGCGTGGCCTTCGAGGACGTGGCCGTGAACTTCACCCAGGAGGAGTGGGC CCTGCTGGACAGCAGCCAGAAGAACCTGTACAGAGAGGTGATGCAGGAGACCT GCAGAAACCTGGCCCAGCGTGGGCAGCCAGTGGAAGGACCAGAACATCGAGGA CCACTTCGAGAAGCCCGGCAAGGACATC (SEQ ID NO: 1144)	5.58623836
ZN256	IVTFEDVAVYFS WKEWGLLDEAQ KCLYHDVMLENL TLTTSL (SEQ ID NO: 499)	AAAELTAPAQGIVTFEDVAVYF SWKEWGLLDEAQKCLYHDVM LENLTLTTSLGGSGAGDEEAPY QQSTSPQRVSQVRIPK (SEQ ID NO: 500)	GCCGCCGCCGAGCTGACCGCCCCCGCCCAGGGCATCGTGACCTTCGAAGACGTG GCCGTGTACTTCAGCTGGAAGGAGTGGGGCCTGCTGGACGAGGCCCAGAAGTG CCTGTACCACGACGTGATGCTGGAGAACCTGACCCTGACCACCAGCCTGGGCGG CAGCGGCGCCGGAGATGAGGAGGCTCCTTACCAGCAGAGCACCAGCCCCCAGA GAGTGAGCCAGGTGAGAATCCCCAAG (SEQ ID NO: 1145)	5.57864488
ZN251	PLTFQDVAVYFS QAEGRQLGPQQ RALYRDVMLENY GNVASL (SEQ ID NO: 501)	AATFQLPGHQEMPLTFQDVAV YFSQAEGRQLGPQQRALYRDV MLENYGNVASLGFPVPKPELIS QLEQGKELWVLNLLGA (SEQ ID NO: 502)	GCCGCCACCTTCCAGCTGCCCGGCCACCAGGAGATGCCCCTGACCTTCCAGGAC GTGGCCGTGTACTTCAGCCAGGCCGAGGGCAGACAGCTGGGCCCCCAGCAGAG AGCCCTGTACAGAGATGTGATGCTGGAGAACTACGGCAACGTGGCCAGCCTGG GCTTCCCCGTGCCCAAGCCCGAGCTGATCAGCCAGGAGCAGGGCAAGGAG CTGTGGGTGCTGAACCTGCTGGGCGCC (SEQ ID NO: 1146)	5.54680119
CBX4	SEFKPFFGNIIITD VTANCLTVTFKEY VT (SEQ ID NO: 503)	RSEAGEPPSSLQVKPETPASAA VAVAAAAAPTTTAEKPPAEAQ DEPAESLSEFKPFFGNIIITDVTA NCLTVTFKEYVTV (SEQ ID NO: 504)	AGAAGCGAGGCCGGCGAGCCCCCCAGCAGCCTGCAGGTGAAGCCCGAGACACC TGCCAGCGCCGCTGTGGCCGTGGCTGCTGCTGCTCCCACCACCACAGCCGA GAAGCCCCCCGCCGAGGCCCAGGACGAACCTGCCGAGAGCCTGAGCGAGTTCA AGCCCTTCTTCGGCAACATCATCATCACCGACGTGACCGCCAACTGCCTGACCGT GACCTTCAAGGAGTACGTGACCGTG (SEQ ID NO: 1147)	5.47206529

PCGF2	MCALCGGYFIDA TTIVECLHSFCKTC IVRYLETNKYCPM (SEQ ID NO: 505)	HRTTRIKITELNPHLMCALCGGY FIDATTIVECLHSFCKTCIVRYLET NKYCPMCDVQVHKTRPLLSIRS DKTLQDIVYK (SEQ ID NO: 506)	CACAGAACCACCAGAATCAAGATCACCGAGCTGAACCCCCACCTGATGTGCGCC CTGTGCGGCGGCTACTTCATCGACGCCACCACCATCGTGGAGTGCCTGCACAGC TTCTGCAAGACCTGCATCGTGAGATACCTGGAGACCAACAAGTACTGCCCCATG TGCGACGTGCAGGTGCACAAGACCAGACCCCTGCTGAGCATCAGAAGCGACAA GACCCTGCAGGACATCGTGTACAAG (SEQ ID NO: 1148)	5.41711547
CDY2	FEVEAIVDKRQD KNGNTQYLVRW KGYDKQDDTWE PEQHLMNCEKCV HDFNR (SEQ ID NO: 507)	ASQEFEVEAIVDKRQDKNGNT QYLVRWKGYDKQDDTWEPEQ HLMNCEKCVHDFNRRQTEKQK KLTWTTTSRIFSNNARRR (SEQ ID NO: 508)	GCCAGCCAGGAGTTCGAGGTGGAGGCCATCGTGGACAAGAGACAGGACAAGA ACGGCAACACCCAGTACCTGGTGAGATGGAAGGGCTACGACAAGCAGGACGAC ACCTGGGAGCCCGAGCAGCACCTGATGAACTGCGAGAAGTGCGTGCACGACTT CAACAGAAGACAGACCGAGAAGCAGAAGAAGCTGACCTGGACCACCACCAGCA GAATCTTCAGCAACAACGCCAGAAGAAGA (SEQ ID NO: 1149)	5.20865573
CDYL2	YEVERIVDKRKNK KGKWEYLIRWKG YGSTEDTWEPEH HLLHCEEFIDEFN (SEQ ID NO: 509)	ASGDLYEVERIVDKRKNKKGKW EYLIRWKGYGSTEDTWEPEHHL LHCEEFIDEFNGLHMSKDKRIKS GKQSSTSKLLRDS (SEQ ID NO: 510)	GCCAGCGGCGACCTGTACGAGGTGGAGAGAATCGTGGACAAGAGAAAGAACA AGAAGGGCAAGTGGGAGTACCTGATCAGATGGAAGGGCTACGGCAGCACCGA GGACACCTGGGAGCCCGAGCACCACCTGCTGCACTGCGAGGAGTTCATCGACG AGTTCAACGGCCTGCACATGAGCAAGGACAAGAGAATCAAGAGCGGCAAGCAG AGCAGCACCAGCAAGCTGCTGAGAGACAGC (SEQ ID NO: 1150)	5.17777542
ZN287	SMTFKDVAVDIT QEDWELMRPVQ KELYKTVTLQNY WNMVSL (SEQ ID NO: 511)	PRGKHAFQTGWLNDLVTKESM TFKDVAVDITQEDWELMRPVQ KELYKTVTLQNYWNMVSLGLT VYRPTVIPILEEPWMVI (SEQ ID NO: 512)	CCCAGAGGCAAGCACGCCTTCCAGACCGGCTGGCTGAACGACCTGGTGACCAA GGAGAGCATGACCTTCAAGGACGTGGCCGTGGACATCACCCAGGAGGACTGGG AGCTGATGAGACCCGTGCAGAAGGAGCTGTACAAGACCGTGACCCTGCAGAAC TACTGGAACATGGTGAGCCTGGGCCTGACCGTGTACAGACCCACCGTGATCCCC ATCCTGGAGGAGCCCTGGATGGTGATC (SEQ ID NO: 1151)	5.15786106
HERC2	RKADLENHNKDG GFWTVIDGKVYD IKDFQTQSLTGNS ILAQFAGEDPVV ALEAALQFEDTRE (SEQ ID NO: 513)	TLIRKADLENHNKDGGFWTVID GKVYDIKDFQTQSLTGNSILAQF AGEDPVVALEAALQFEDTRES MHAFCVGQYLEPDQ (SEQ ID NO: 514)	ACCCTGATCAGAAAGGCCGACCTGGAGAACCACAACAAGGACGGCGGCTTCTG GACCGTGATCGACGGCAAGGTGTACGACATCAAGGACTTCCAGACCCAGAGCC TGACCGGCAACAGCATCCTGGCCCAGTTCGCCGGCGAGGACCCCGTGGTGGCC CTGGAAGCTGCTCTGCAGTTCGAGGACACCAGAGAGAGCATGCACGCCTTCTGC GTGGGCCAGTACCTGGAGCCCGACCAG (SEQ ID NO: 1152)	5.12990133
ZN562	SVTFDDVAVEFTP EEWALLDTTQKY LYRDVMLENYM NLASV (SEQ ID NO: 515)	EKTKIGTMVEDHRSNSYQDSVT FDDVAVEFTPEEWALLDTTQKY LYRDVMLENYMNLASVDFFFCL TSEWEIQPRTKRSS (SEQ ID NO: 516)	GAGAAGACCAAGATCGGCACCATGGTGGAGGACCACAGAAGCAACAGCTACCA GGACAGCGTGACCTTCGACGACGTGGCCGTGGAGTTCACCCCCGAGGAGTGGG CCCTGCTGGACACCACCCAGAAGTACCTGTACAGAGATGTGATGCTGGAGAACT ACATGAACCTGGCCAGCGTGGACTTCTTCTTCTGCCTGACCAGCGAGTGGGAGA TCCAGCCCAGAACCAAGAGAAGCAGC (SEQ ID NO: 1153)	5.08331004

ZN461	LVMFRDVAIDVSQEEWECLNPAQRNLYKEVMLENYSNLVSL (SEQ ID NO: 517)	AHELVMFRDVAIDVSQEEWECLNPAQRNLYKEVMLENYSNLVSLGLSVSKPAVISSLEQGKEPWMVVREETGRWCPGTWK (SEQ ID NO: 518)	GCCCACGAGCTGGTGATGTTCAGAGATGTGGCCATCGACGTGAGCCAGGAGGAGTGGGAGTGCCTGAACCCCGCCCAGAGAAACCTGTACAAGGAGGTGATGCTGGAGAACTACAGCAACCTGGTGAGCCTGGGCCTGAGCGTGAGCAAGCCCGCCGTGATCAGCAGCCTGGAGCAGGGCAAGGAGCCCTGGATGGTGGTGAGAGGAGACCGGCAGATGGTGCCCCGGCACCTGGAAG (SEQ ID NO: 1154)	5.05101639
Z324A	MAFEDVAVYFSQEEWGLLDTAQEEWGLLDTAQRALYRRVMLDNFALVASLGLSTLYRRVMLDNFALVASL (SEQ ID NO: 519)	AFEDVAVYFSQEEWGLLDTAQRALYRRVMLDNFALVASLGLSTSRPRVVIQLERGEEPWVPSGTDTTLSRTTYRRRNPGS (SEQ ID NO: 520)	GCCTTCGAGGACGTGGCCGTGTACTTCAGCCAGGAGGAGTGGGGCCTGCTGGACACCGCCCAGAGAGCCCTGTACAGAAGAGTGATGCTGGACAACTTCGCCCTGGTGGCCAGCCTGGGCCTGAGCACCAGACCCAGAGTGGTGATCCAGCTGGAGAGAGGCGAGGAGCCCTGGGTGCCCAGCGGCACCGACACCACCCTGAGCAGAACCACCTACAGAAGAAGAAACCCCGGCAGC (SEQ ID NO: 1155)	5.01043067
ZN766	HLTFRDVAIEFSQEEWKCLDPVQKALYRDVMLENYRNLVSL (SEQ ID NO: 521)	AQLRRGHLTFRDVAIEFSQEEWKCLDPVQKALYRDVMLENYRNLVSLGICLPDLSIISMMKQRTEPWTVENEMKVAKNPD (SEQ ID NO: 522)	GCCCAGCTGAGAAGAGGCCACCTGACCTTCAGAGATGTGGCCATCGAGTTCAGCCAGGAGGAGTGGAAGTGCCTGGACCCCGTGCAGAAGGCCCTGTACAGAGATGTGATGCTGGAGAACTACAGAAACCTGGTGAGCCTGGGCATCTGCCTGCCCGACCTGAGCATCATCAGCATGATGAAGCAGAGAACCGAGCCCTGGACCGTGGAGAACGAGATGAAGGTGGCCAAGAACCCCGAC (SEQ ID NO: 1156)	4.9926318
ID2	LLYNMNDCYSKLKELVPSIPQNKKVSKMEILQHVIDYILDLQALDSHPLDL (SEQ ID NO: 523)	SDHSLGISRSKTPVDDPMSLLYNMNDCYSKLKELVPSIPQNKKVSKMEILQHVIDYILDLQALDSHPTIVSLHHQRPGQ (SEQ ID NO: 524)	AGCGACCACAGCCTGGGCATCAGCAGAAGCAAGACCCCCGTGGACGACCCCATGAGCCTGCTGTACAACATGAACGACTGCTACAGCAAGCTGAAGGAGCTGGTGCCCAGCATCCCCCAGAACAAGAAGGTGAGCAAGATGGAGATCCTGCAGCACGTGATCGACTACATCCTGGACCTGCAGATCGCCCTGGCACAGCCACCCCACCATCGTGAGCCTGCACCACCAGAGACCCGGCCAG (SEQ ID NO: 1157)	4.86972562
TOX	PQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDGLGEEQKQVYKKKTEAAKKEYLKQLAAY (SEQ ID NO: 525)	KDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWDGLGEEQKQVYKKKTEAAKKEYLKQLAAYRASLVSK (SEQ ID NO: 526)	AAGGACCCCAACGAGCCCCAGAAGCCCGTGAGCGCCTACGCCCTGTTCTTCAGAGACACCCAGGCCGCCATCAAGGGCCAGAACCCCAACGCCACCTTCGGCGAGGTGAGCAAGATCGTGGCCAGCATGTGGGACGGCCTGGGCGAGGAGCAGAAGCAGGTGTACAAGAAGAAGACCGAGGCCGCCAAGAAGGAGTACCTGAAGCAGCTGGCCGCCTACAGAGCCAGCCTGGTGAGCAAG (SEQ ID NO: 1158)	4.84737013
ZN274	PVTFQDVAVDFSREEWGLLGPTQRTEYRDVMLETFGHLVSV (SEQ ID NO: 527)	QEEKQEDAAICPVTVLPEEPVTFQDVAVDFSREEWGLLGPTQRTEYRDVMLETFGHLVSVGWETTLENKELAPNSDIPEE (SEQ ID NO: 528)	CAGGAGGAGAAGCAGGAGGACGCCGCCATCTGCCCCGTGACCGTGCTGCCCGAGGAGCCTGTGACCTTCCAGGACGTGGCCGTGGACTTCAGCAGAGAGGAGTGGGGCCTGCTGCTGGGCCCCACCCAGAGAACCGAGTACAGAGATGTGATGCTGGAGACCTTCGGCCCCAACAGCGACATCCCCGAGGAG (SEQ ID NO: 1159)	4.82395142

ZN75C	LLTFEDMAVYFSE EEWQLLGPLEKTL YNDVMQDIYETA ISL (SEQ ID NO: 529)	KSTKDWKMAPKLIWPESQSLLT FEDMAVYFSEEEWQLLGPLEKT LYNDVMQDIYETAISLGKQRTG KIMGIEMASSFSKE (SEQ ID NO: 530)	AAGAGCACCAAGGACTGGAAGATGGCCCCCAAGCTGATCTGGCCCGAGAGCCA GAGCCTGCTGACCTTCGAGGACATGGCCGTGTACTTCAGCGAGGAGGAGTGGC AGCTGCTGGGCCCCCTGGAGAAGACCCTGTACAACGACGTGATGCAGGACATCT ACGAGACCGCCATCAGCCTGGGCAAGCAGAGAACCGGCAAGATCATGGGCATC GAGATGGCCAGCAGCTTCAGCAAGGAG (SEQ ID NO: 1160)	4.81809368
SCMH1	SSWTVEDVMQF VREADPQLGPHA DLFRKHEIDGKAL LLLRSDMIMMKY MGLKLGPALKLS YHIDRL (SEQ ID NO: 531)	DASRLSGRDPSSWTVEDVMQF VREADPQLGPHADLFRKHEIDG KALLLLRSDMIMMKYMGLKLGP ALKLSYHIDRLKQGKF (SEQ ID NO: 532)	GACGCCAGCAGACTGAGCGGCAGAGACCCCAGCAGCTGGACCGTGGAGGACG TGATGCAGTTCGTGAGAGAGGCCGACCCCCAGCTGGGCCCCCACGCCGACCTGT TCAGAAAGCACGAGATCGACGGCAAGGCCCTGCTGCTGCTGAGAAGCGACATG ATGATGAAGTACATGGGCCTGAAGCTGGGCCCCGCCCTGAAGCTGAGCTACCAC ATCGACAGACTGAAGCAGGGCAAGTTC (SEQ ID NO: 1161)	4.79639316
ZN560	SVTFEDTAVDFT QEEWILLDPVQR NLYRDVMLENYE NVAKV (SEQ ID NO: 533)	AYCLTNCYQYSVTFEDTAVDFT QEEWILLDPVQRNLYRDVMLE NYENVAKVGFQLFKPSVISWLE EEELRTLQQGVLQDW (SEQ ID NO: 534)	GCCTACTGCCTGACCAACTGCTACCAGTACAGCGTGACCTTCGAGGACACCGCC GTGGACTTCACCCAGGAGGAGTGGATCCTGCTGGACCCCGTGCAGAGAAACCT GTACAGAGATGTGATGCTGGAGAACTACGAGAACGTGGCCAAGGTGGGCTTCC AGCTGTTCAAGCCCAGCGTGATCAGCTGGCTGGAGGAGGAGGAGCTGAGAACC CTGCAGCAGGGCGTGCTGCAGGACTGG (SEQ ID NO: 1162)	4.77465441
SCML4	SAWTVEDVVWF VKDADPQALGPH VELFRKHEIDGNA LLLLKSDMVMKY LGLKLGPALKLCY HIDKL (SEQ ID NO: 535)	ARRPRSRNPSAWTVEDVVWFV KDADPQALGPHVELFRKHEIDG NALLLLKSDMVMKYLGLKLGPA LKLCYHIDKLKQAKF (SEQ ID NO: 536)	GCCAGAAGACCCAGAAGCAGAAACCCCAGCGCCTGGACCGTGGAGGACGTGGT GTGGTTCGTGAAGGACGCCGACCCCCAGGCCCTGGGCCCCCACGTGGAGCTGTT CAGAAAGCACGAGATCGACGGCAACGCCCTGCTGCTGCTGAAGAGCGACATGG TGATGAAGTACCTGGGCCTGAAGCTGGGCCCCGCCCTGAAGCTGTGCTACCACA TCGACAAGCTGAAGCAGGCCAAGTTC (SEQ ID NO: 1163)	4.74079704
ZN214	AVTFEDVTIIFTW EEWKFLDSSQKR LYREVMWENYT NVMSV (SEQ ID NO: 537)	AVTFEDVTIIFTWEEWKFLDSSQ KRLYREVMWENYTNVMSVEN WNESYKSQEEKFRYLEYENFSY WQGWWNAGAQMYENQ (SEQ ID NO: 538)	GCCGTGACCTTCGAGGACGTGACCATCATCTTCACCTGGGAGGAGTGGAAGTTC CTGGACAGCAGCCAGAAGAGACTGTACAGAGAGGTGATGTGGGAGAACTACAC CAACGTGATGAGCGTGGAGAACTGGAACGAGAGCTACAAGAGCCAGGAGGAG AAGTTCAGATACCTGGAGTACGAGAACTTCAGCTACTGGCAGGGCTGGTGGAA CGCCGGCGCCCAGATGTACGAGAACCAG (SEQ ID NO: 1164)	4.72989473
CBX7	FAVESIRKKRVRK GKVEYLVKWKG WPPKYSTWEPEE HILDPRLVMAYEE (SEQ ID NO: 539)	ELSAIGEQVFAVESIRKKRVRKG KVEYLVKWKGWPPKYSTWEPE EHILDPRLVMAYEEKEERDRAS GYRKRGPKPKRLLL(SEQ ID NO: 540)	GAGCTGAGCGCCATCGGCGAGCAGGTGTTCGCCGTGGAGAGCATCAGAAAGAA GAGAGTGAGAAAGGGCAAGGTGGAGTACCTGGTGAAGTGGAAGGGCTGGCCC CCCAAGTACAGCACCTGGGAGCCCGAGGAGCACATCCTGGACCCCAGACTGGT GATGGCCTACGAGGAGAAGGAGGAGAGACAGAGCCAGCGGCTACAGAAAG AGAGGCCCCAAGCCCAAGAGACTGCTGCTG (SEQ ID NO: 1165)	4.70199486

ID1	LYDMNGCYSRLK EIVPTLPQNRKVS KVEILQHVIDYIR DL (SEQ ID NO: 541)	GGAGARLPALLDEQQVNVLLY DMNGCYSRLKELVPTLPQNRKV SKVEILQHVIDYIRDLQLELNSES EVGTPGGRGLPVR (SEQ ID NO: 542)	GGCGGCGCCGGCGCCAGACTGCCCGCCCTGCTGGACGAGCAGCAGGTTAACGT GCTGCTGTACGACATGAACGGCTGCTACAGCAGACTGAAGGAGCTGGTGCCCA CCCTGCCCCAGAACAGAAAGGTGAGCAAGGTGGAGATCCTGCAGCACGTGATC GACTACATCAGAGACCTGCAGCTGGAGCTGAACAGCGAGAGCGAGGTGGGCAC CCCCGGCGGCAGAGGCCTGCCTGTGAGA (SEQ ID NO: 1166)	4.66128008
CREM	AKECRRRKKEYVK CLESRVAVLEVQ NKKLIEELETLKDI CSPKTD (SEQ ID NO: 543)	VVMAASPGSLHSPQQLAEEAT RKRELRLMKNREAAKECRRRKK EYVKCLESRVAVLEVQNKKLIEE LETLKDICSPKTDY (SEQ ID NO: 544)	GTGGTGATGGCCGCCAGCCCCGGCAGCCTGCACAGCCCCCAGCAGCTTGCTGA GGAAGCCACCAGAAAGAGAGAGCTGAGACTGATGAAGAACAGAGAGGCCGCC AAGGAGTGCAGAAGAAGAAAGAAGGAGTACGTGAAGTGCCTGGAGAAGCAGAG TGGCCGTGCTGGAGGTGCAGAACAAGAAGCTGATCGAGGAGCTGGAGACCCTG AAGGACATCTGCAGCCCCAAGACCGACTAC (SEQ ID NO: 1167)	4.58757659
FER3L	QRQAANIRERKR MFNLNEAFDQLR RKVPTFAYEKRLS RIETLRLAIVYISF M (SEQ ID NO: 545)	SLLGRPKRKRVITYAQRQAANIR ERKRMFNLNEAFDQLRRKVPTF AYEKRLSRIETLRLAIVYISFMTEL LESCEKKESG (SEQ ID NO: 546)	AGCCTGCTGGGCAGACCCAAGAGAAAGAGAGTGATCACCTACGCCCAGAGACA GGCCGCCAACATCAGAGAGAGAAAGAGAATGTTCAACCTGAACGAGGCCTTCG ACCAGCTGAGAAGAAAGGTGCCCACCTTCGCCTACGAGAAGAGACTGAGCAGA ATCGAGACCCTGAGACTGGCCATCGTGTACATCAGCTTCATGACCGAGCTGCTG GAGAGCTGCGAGAAGAAGGAGAGCGGC (SEQ ID NO: 1168)	4.55608825
SCX	QRHTANARERDR TNSVNTAFTALRT LIPTEPADRKLSKI ETLRLASSYISHL (SEQ ID NO: 547)	GGGPGGRPGREPRQRHTANA RERDRTNSVNTAFTALRTLIPTE PADRKLSKIETLRLASSYISHLGN VLLAGEACGDGQP (SEQ ID NO: 548)	GGCGGCGGCCCCGGCGGCAGACCGGCAGAGAGCCCAGACAAAGACACACCG CCAACGCCAGAGAGAGAGACAGAACCAACAGCGTGAACACCGCCTTCACCGCC CTGAGAACCCTGATCCCCACCGAGCCCGCCGACAGAAAGCTGAGCAAGATCGA GACCCTGAGACTGGCCAGCAGCTACATCAGCCACCTGGGCAACGTGCTGCTGGC CGGCGAGGCCTGCGGCGATGGACAGCCC (SEQ ID NO: 1169)	4.38664628
ASCL1	AVARRNERERNR VKLVNLGFATLRE HVPNGAANKKM SKVETLRSAVEYIR AL (SEQ ID NO: 549)	SGFGYSLPQQQPAAVARRNER ERNRVKLVNLGFATLREHVPNG AANKKMSKVETLRSAVEYIRAL QQLLDEHDAVSAAFQ (SEQ ID NO: 550)	AGCGGCTTCGGCTACAGCCTGCCCCAGCAGCAGCCCGCCGTGGCTAGAAG AAACGAGAGAGAGAGAACAGAGTGAAGCTGGTGAACCTGGGCTTCGCCACCC TGAGAGAGCACGTGCCCAACGGCGCCGCCAACAAGAAGATGAGCAAGGTGGA GACCCTGAGAAGCGCCGTGGAGTACATCAGAGCCCTGCAGCAGCTGCTGGACG AGCACGACGCCGTGAGCGCCGCCTTCCAG (SEQ ID NO: 1170)	4.23952129
ZN764	AVSFADVAVYFC REEWGCLRPAQR ALYRDVMRETYG HLSAL (SEQ ID NO: 551)	APLPPRDPNGAGPEWREPGAV SFADVAVYFCREEWGCLRPAQ RALYRDVMRETYGHLSALGIGG NKPALISWVEEEAELW (SEQ ID NO: 552)	GCCCCCCTGCCCCCAGAGACCCCAACGGCGCCGGCCCTGAATGGAGAGAGCC CGGAGCCGTGAGCTTCGCCGACGTGGCCGTGTACTTCTGCAGAGAGGAGTGGG GCTGCCTGAGACCCGCCCAGAGAGCCCTGTACAGAGATGTGATGAGAGAGACC TACGGCCACCTGAGCGCCCTGGGCATCGGCGGCAACAAGCCCGCCCTGATCAGC TGGGTGGAGGAGGAGGCCGAGCTGTGG (SEQ ID NO: 1171)	4.16413141

SCML2	STWSVDEVIQFM KHTDPQISGPLA DLFRQHEIDGKAL FLLKSDVMMKY MGLKLGPALKLC YYIEKL (SEQ ID NO: 553)	KQGFSKDPSTWSVDEVIQFMK HTDPQISGPLADLFRQHEIDGK ALFLLKSDVMMKYMGLKLGPA LKLCYYIEKLKEGKYS (SEQ ID NO: 554)	AAGCAGGGCTTCAGCAAGGACCCCAGCACCTGGAGCGTGGACGAGGTGATCCA GTTCATGAAGCACACCGACCCCAGATCAGCGGCCCCCTGGCCGACCTGTTCAG ACAGCACGAGATCGACGGCAAGGCCCTGTTCCTGCTGAAGAGCGACGTGATGA TGAAGTACATGGGCCTGAAGCTGGGCCCCGCCCTGAAGCTGTGCTACTACATCG AGAAGCTGAAGGAGGGCAAGTACAGC (SEQ ID NO: 1172)	4.16119992
ASCL5	IQKRNERERQRV KCVNEGYARLRG HLPGALAEKRLSK VETLRAAIRYIKYL (SEQ ID NO: 555)	GAFGVYEYPFEPAFIQKRNERER QRVKCVNEGYARLRGHLPGAL AEKRLSKVETLRAAIRYIKYLQEL LSSAPDGSTPPA (SEQ ID NO: 556)	GGCGCCTTCGGCGTGTACGAGTACCCCTTCGAGCCCGCCTTCATCCAGAAGAGA AACGAGAGAGAGACAGAGAGTGAAGTGCGTGAACGAGGGCTACGCCAGAC TGAGAGGCCACCTGCCCGGCGCCCTGGCCGAGAAGAGACTGAGCAAGGTGGA GACCCTGAGAGCCGCCATCAGATACATCAAGTACCTGCAGGAGCTGCTGAGCA GCGCCCCCGACGGCAGCACCCCCCCTGCT (SEQ ID NO: 1173)	4.14708139
TWST1	QRVMANVRERQ RTQSLNEAFAAL RKIIPTLPSDKLSKI QTLKLAARYIDFL (SEQ ID NO: 557)	SGGGSPQSYEELQTQRVMANV RERQRTQSLNEAFAALRKIIPTL PSDKLSKIQTLKLAARYIDFLYQV LQSDELDSKMAS (SEQ ID NO: 558)	AGCGGCGGCGGCAGCCCCCAGAGCTACGAGGAGCTGCAGACCCAGAGAGTGA TGGCCAACGTGAGAGAGACAGAGAACCCAGAGCCTGAACGAGGCCTTCGCC GCCCTGAGAAAGATCATCCCCACCCTGCCCAGCGACAAGCTGAGCAAGATCCAG ACCCTGAAGCTGGCCGCCAGATACATCGACTTCCTGTACCAGGTGCTGCAGAGC GACGAGCTGGACAGCAAGATGGCCAGC (SEQ ID NO: 1174)	4.09571741
ZN319	DLQRHRRVHTGE KPYKCPNCDKAF K (SEQ ID NO: 559)	FVQHRCDPAREKPLKCPDCEKR FKYASDLQRHRRVHTGEKPYKC PNCDKAFKQREHLNKHQGVHA REQQFKCVWCGERFL (SEQ ID NO: 560)	TTCGTGCAGCACAGATGCGACCCCGCCAGAGAGAAGCCCCTGAAGTGCCCCGA CTGCGAGAAGAGATTCAAGTACGCCAGCGACCTGCAGAGACACAGAAGAGTGC ACACCGGCGAGAAGCCCTACAAGTGCCCCAACTGCGACAAGGCCTTCAAGCAG AGAGAGCACCTGAACAAGCACACCAGGGCGTGCACGCCAGAGAGCAGCAGTTCAA GTGCGTGTGGTGCGGCGAGAGATTCCTG (SEQ ID NO: 1175)	4.08013835
ZN749	CMVFEDVAIYFS QEEWGILNDAQ RHLHSNVMLENF ALLSSV (SEQ ID NO: 561)	NLTEDCMVFEDVAIYFSQEEW GILNDAQRHLHSNVMLENFALL SSVGCWHGAKDEEVPSKQCVS VRVLQVTIPKPALSTL (SEQ ID NO: 562)	AACCTGACCGAGGACTGCATGGTGTTCGAGGACGTGGCCATCTACTTCAGCCAG GAGGAGTGGGGCATCCTGAACGACGCCCAGAGACACCTGCACAGCAACGTGAT GCTGGAGAACTTCGCCCTGCTGAGCAGCGTGGGCTGCTGGCACGGCGCCAAGG ACGAGGAGGTGCCCAGCAAGCAGTGCGTGAGCGTGAGAGTGCTGCAGGTGAC CATCCCCAAGCCCGCCCTGAGCACCCTG (SEQ ID NO: 1176)	4.06508464

表 2: Pfam 激活蛋白

基因	Pfam 结构域序列 sequence	延伸的结构域序列	延伸的结构域 DNA 序列	平均激活
ZN473	FVTLKDVGMDFTL GDWEQLGLEQGD TFWDTALDNCQD LFLL (SEQ ID NO: 563)	AEEFVTLKDVGMDFTLGDWEQ LGLEQGDTFWDTALDNCQDLF LLDPPRPNLTSHPDGSEDLEPLA GGSPEATSPDVTETK (SEQ ID NO: 564)	GCCGAGGAGTTCGTGACCCTGAAGGACGTGGGCATGGACTTCACCCTGGGC GACTGGGAGCAGCTGGGCCTGGAGCAGGGCGACACCTTCTGGGACACCGCC CTGGACAACTGCCAGGACCTGTTCCTGCTGGACCCCCCCAGACCCAACCTGAC CAGCCACCCCGACGGCAGCGAGGACCTGGAGCCCCTGGCTGGAGGAAGCCC TGAGGCCACCAGCCCCGACGTGACCGAGACCAAG (SEQ ID NO: 1178)	-8.6232004
FOXO3	HEKFPSDLDLDMF NGSLECDMESIIRS ELMDADGLDFNFD S (SEQ ID NO: 565)	DSLSGSSLYSTSANLPVMGHEK FPSDLDLDMFNGSLECDMESII RSELMDADGLDFNFDSLISTQN VVGLNVGNFTGAKQ (SEQ ID NO: 566)	GACAGCCTGAGCGGCAGCAGCCTGTACAGCACCAGCGCCAACCTGCCCGTGA TGGGCCACGAGAAGTTCCCCAGCGACCTGGACCTGGACATGTTCAACGGCAG CCTGGAGTGCGACATGGAGAGCATCATCAGAAGCGAGCTGATGGACGCCGA CGGCCTGGACTTCAACTTCGACAGCCTGATCAGCACCCAGAACGTGGTGGGC CTGAACGTGGGCAACTTCACCGGCGCCAAGCAG (SEQ ID NO: 1179)	-8.3891724
FOXO1	QEKLPSDLDGMFIE RLDCDMESIIRNDL MDGDTLDFNFDN (SEQ ID NO: 567)	GGYSSVSSCNGYGRMGLLHQE KLPSDLDGMFIERLDCDMESIIR NDLMDGDTLDFNFDNVLPNQ SFPHSVKTTTHSWVSG (SEQ ID NO: 568)	GGCGGCTACAGCAGCGTGAGCAGCTGCAACGGCTACGGCAGAATGGGCCTG CTGCACCAGGAGAAGCTGCCCAGCGACCTGGACGGCATGTTCATCGAGAGAC TGGACTGCGACATGGAGAGCATCATCAGAAACGACCTGATGGACGGCGACA CCCTGGACTTCAACTTCGACAACGTGCTGCCCAACCAGAGCTTCCCCCACAGC GTGAAGACCACCACCCACAGCTGGGTGAGCGGC (SEQ ID NO: 1180)	-8.3703632
MYBA	GNCIEHVQPTSAFI QQPFIDEDPDKEK KIKELEMLLMSAEN EVRRK (SEQ ID NO: 569)	FYIPVQIPGYQYVSPEGNCIEHV QPTSAFIQQPFIDEDPDKEKKIK ELEMLLMSAENEVRRKRIPSQP GSFSSWSGSFLM (SEQ ID NO: 570)	TTCTACATCCCCGTGCAGATCCCCGGCTACCAGTACGTGAGCCCCGAGGGCA ACTGCATCGAGCACGTGCAGCCCACCAGCGCCTTCATCCAGCAGCCCTTCATC GACGAGGACCCCGACACAAGGAGAAGAAGATCAAGGAGCTGGAGATGCTGCT GATGAGCGCCGAGAACGAGGTGAGAAGAAAGAGAATCCCCAGCCAGCCCG GCAGCTTCAGCAGCTGGAGCGGCAGCTTCCTGATG (SEQ ID NO: 1181)	-8.2096102
MYB	VNIVNVPQPAAAA IQRHYNDEDPEKE KRIKELELLLMSTE NELKGQ(SEQ ID NO: 571)	EAQNVSSHVPYPVALHVNIVNV PQPAAAAIQRHYNDEDPEKEK RIKELELLLMSTENELKGQQVLP TQNHTCSYPGWHST (SEQ ID NO: 572)	GAGGCCCAGAACGTGAGCAGCCACGTGCCCTACCCCGTGGCCCTGCACGTGA ACATCGTGAACGTGCCCCCAGCCCGCCGCCGCCATCCAGAGACACTACAA CGACGAGGACCCCGAGAAGGAGAAGAGAATCAAGGAGCTGGAGCTGCTGCT GATGAGCACCGAGAACGAGCTGAAGGGCCAGCAGGTGCTGCCCACCCAGAA CCACACCTGCAGCTACCCCGGCTGGCACAGCACC (SEQ ID NO: 1182)	-7.2112528

NCOA2	ESPSDEGALLDQLY LALRNFDGLEEIDR ALGIPELVSQSQAV DPEQ (SEQ ID NO: 573)	PFGSSPDDLLCPHPAAESPSDE GALLDQLYLALRNFDGLEEIDRA LGIPELVSQSQAVDPEQFSSQD SNIMLEQKAPVFP (SEQ ID NO: 574)	CCCTTCGGCAGCAGCCCCGACGACCTGCTGTGCCCCCACCCCGCTGCTGAAA GCCCCAGCGACGAGGGCGCTCTGCTGGACCAGCTGTACCTGGCCCTGAGAAA CTTCGACGGCCTGGAGGAGATCGACAGAGCCCTGGGCATCCCCGAGCTGGT GAGCCAGAGCCAGGCCGTGGACCCCGAGCAGTTCAGCAGCCAGGACAGCAA CATCATGCTGGAGCAGAAGGCCCCCGTGTTCCCC (SEQ ID NO: 1183)	-7.1119077
SMCA2	PFSPVQLHQLRAQI LAYKMLARGQPLP ETLQLAV (SEQ ID NO: 575)	SQPGALIPGDPQAMSQPNRGP SPFSPVQLHQLRAQILAYKMLA RGQPLPETLQLAVQGKRTLPGL QQQQQQQQQQQQQQQ (SEQ ID NO: 576)	AGCCAGCCCGGCGCCCTGATCCCCGGCGACCCCCAGGCTATGAGCCAACCTA ACAGAGGCCCCAGCCCCTTCAGCCCCGTGCAGCTGCACCAGCTGAGAGCCCA GATCCTGGCCTACAAGATGCTGGCCAGAGGCCAGCCCCCTGCCCGAGACCCTG CAGCTGGCCGTGCAAGGCAAGAGAACCCTGCCCGGCCTGCAGCAGCAGCAG CAGCAGCAGCAGCAGCAGCAGCAGCAGCAG (SEQ ID NO: 1184)	-6.7916451
KIBRA	LPEGWEEARDFDG KVYYIDHTNRTTS WID (SEQ ID NO: 577)	PRPELPLPEGWEEARDFDGKVY YIDHTNRTTSWIDPRDRYTKPLT FADCISDELPLGWEEAYDPQVG DYFIDHNTKTTQI (SEQ ID NO: 578)	CCCAGACCCGAGCTGCCCCTGCCCGAGGGCTGGGAGGAGGCCAGAGACTTC GACGGCAAGGTGTACTACATCGACCACACCAACAGAACCACCAGCTGGATCG ACCCCAGAGACAGATACACCAAGCCCCTGACCTTCGCCGACTGCATCAGCGA CGAGCTGCCCCTGGGCTGGGAGGAGGCCTACGACCCCCAGGTGGGAGACTA CTTCATCGACCACAACACCAAGACCACCCAGATC (SEQ ID NO: 1185)	-6.707792
NCOA3	EGQSDERALLDQL HTLLSNTDATGLEE IDRALGIPELVNQG QALEPKQ (SEQ ID NO: 579)	LRNSLDDLVGPPSNLEGQSDER ALLDQLHTLLSNTDATGLEEIDR ALGIPELVNQGQALEPKQDAF QGQEAAVMMDQKAG (SEQ ID NO: 580)	CTGAGAAACAGCCTGGACGACCTGGTGGGCCCCCCCAGCAACCTGGAGGGC CAGAGCGACGAGAGAGCCCTGCTGGACCAGCTGCACACCCTGCTGAGCAAC ACCGACGCCACCGGCCTGGAGGAGATCGACAGAGCCCTGGGCATCCCCGAG CTGGTGAACCAGGGCCAGGCCCTGGAGCCCAAGCAGGACGCCTTCCAGGGC CAGGAGGCTGCCGTGATGATGGACCAGAAGGCCGGC (SEQ ID NO: 1186)	-6.4149356
FOXO6	PDRFPADLDLDMF SGSLECDVESIILND FMDSDEMDFNFD S (SEQ ID NO: 581)	LDALPGPYAAAAAGPLGAAPD RFPADLDLDMFSGSLECDVESII LNDFMDSDEMDFNFDSALPPP PPGLAGAPPPNQSWV (SEQ ID NO: 582)	CTGGACGCCCTGCCCGGCCCCTACGCCGCTGCTGCTGGACCTCTTGGAGC TGCCCCTGACAGATTCCCCGCCGACCTGGACCTGGACATGTTCAGCGGCAGC CTGGAGTGCGACGTGGAGAGCATCATCCTGAACGACTTCATGGACAGCGAC GAGATGGACTTCAACTTCGACAGCGCCCTGCCCCCCCCCCCTCCTGGACTTGC TGGAGCTCCTCCTCCCAACCAGAGCTGGGTG (SEQ ID NO: 1187)	-6.0518896
ZN597	PILFEDLAVYFSQEE CVTLHPAQRSLSKD GTKESLEDAALM (SEQ ID NO: 583)	ASMPPTPEAQGPILFEDLAVYF SQEECVTLHPAQRSLSKDGTKE SLEDAALMGEEGKPEIINQQLSL ESMELDELALEKYP (SEQ ID NO: 584)	GCCAGCATGCCCCCCACCCCCGAGGCCCAGGGCCCCATCCTGTTCGAGGACC TTGCTGTGTACTTCAGCCAGGAGGAGTGCGTGACCCTGCACCCCGCCCAGAG AAGCCTGAGCAAGGACGGCACCAAGGAGAGGCCTGGAGGGACGCCGCCCTGAT GGGCGAGGAGGGCAAGCCCGAGATCAACCAGCAGCTGAGCCTGGAGAGCA TGGAGCTGGACGAGCTGGCCCTGGAGAAGTACCCC (SEQ ID NO: 1188)	-5.9555177

APBB1	LPAGWMRVQDTS GTYYWHIPTGTTQ WEP (SEQ ID NO: 585)	GSPSYGSPEDTDSFWNPNAFET DSDLPAGWMRVQDTSGTYYW HIPTGTTQWEPPGRASPSQGSS PQEESQLTWTGFAHGE (SEQ ID NO: 586)	GGCAGCCCCAGCTACGGCAGCCCCGAGGACACCGACAGCTTCTGGAACCCCA ACGCCTTCGAGACCGACAGCGACCTGCCCGCCGGCTGGATGAGAGTGCAGG ACACCAGCGGCACCTACTACTGGCACATCCCCACCGGCACCACCCAGTGGGA GCCCCCGGAAGAGCTAGCCCTAGCCAGGGCAGCAGCCCCCAGGAGGAGAG CCAGCTGACCTGGACCGGCTTCGCCCACGGCGAG (SEQ ID NO: 1189)	-5.8338079
ANM2	VAIADYAATDETQL SFLRGEKILLRQTT ADWWWGERAGC CGYIP (SEQ ID NO: 587)	ECSEAGLLQEGVQPEEFVAIAD YAATDETQLSFLRGEKILLRQTT ADWWWGERAGCCGYIPANH VGKHVDEYDPEDTWQ (SEQ ID NO: 588)	GAGTGCAGCGAGGCCGGCCTGCTGCAGGAGGGCGTGCAGCCCGAGGAGTTC GTGGCTATCGCCGACTACGCCGCCACCGACGAGACCCAGCTGAGCTTCCTGA GAGGCGAGAAGATCCTGATCCTGAGACAGACCACCGCCGACTGGTGGTGGG GCGAGAGAGCCGGCTGCTGCGGCTACATCCCCGCCAACCACGTGGGCAAGC ACGTGGACGAGTACGACCCCGAGGACACCTGGCAG (SEQ ID NO: 1190)	-5.6456716
MED9	SFLPLVHNIIKCMD KDSPEVHQDLNAL KSKFQEMRKLIST MPGIHLSPEQQQ QQLQSLREQVRTK NELLQKYKSL(SEQ ID NO: 589)	NYSFLPLVHNIIKCMDKDSPEV HQDLNALKSKFQEMRKLISTMP GIHLSPEQQQQQLQSLREQVR TKNELLQKYKSLCMF (SEQ ID NO: 590)	AACTACAGCTTCCTGCCCCTGGTGCACAACATCATCAAGTGCATGGACAAGG ACAGCCCCGAGGTGCACCAGGACCTGAACGCCCTGAAGAGCAAGTTCCAGG AGATGAGAAAGCTGATCAGCACCATGCCCGGCATCCACCTGAGCCCCGAGCA GCAGCAGCAGCTGCAGAGCCTGAGAGAGCAGGTGAGAACCAAGAACG AGCTGCTGCAGAAGTACAAGAGCCTGTGCATGTTC (SEQ ID NO: 1191)	-5.5377024
CXXC1	YCICRKPDINCFMI GCDNCNEWFHGD CIRITEKMAKAIRE WYCRECRE (SEQ ID NO: 591)	AGEDSKSENGENAPIYCICRKPD INCFMIGCDNCNEWFHGDCIRI TEKMAKAIREWYCRECREKDPK LEIRYRHKKSRER (SEQ ID NO: 592)	GCCGGCGAGGACAGCAAGAGCGAGAAGAACGGCGAGAACGCCCCCATCTACTGC ATCTGCAGAAAGCCCGACATCAACTGCTTCATGATCGGCTGCGACAACTGCA ACGAGTGGTTCCACGGCGACTGCATCAGAATCACCGAGAAGATGGCCAAGG CCATCAGAGAGTGGTACTGCAGAGAGTGCAGAGAGAAGGACCCCAAGCTGG AGATCAGATACAGACACAAGAAGAGCAGAGAGAGA (SEQ ID NO: 1192)	-5.4566266
CRTC2	PNIILTGDSSPGFSK EIAAALAGVPGFEV SAAGLELGLGLEDE LRMEPLGLEGLNM LSDPCALLPDPAVE ESFRSDR (SEQ ID NO: 593)	GPNIILTGDSSPGFSKEIAAALA GVPGFEVSAAGLELGLGLEDEL RMEPLGLEGLNMLSDPCALLPD PAVEESFRSDRLQ (SEQ ID NO: 594)	GGCCCCAACATCATCCTGACCGGCGACAGCAGCCCCGGCTTCAGCAAGGAGA TCGCCGCCGCCCTGGCTGGCGTGCCTGGATTCGAGGTGAGCGCTGCTGGACT GGAGCTGGGCCTGGGCCTGGAGGACGAGCTGAGAATGGAGCCCCTGGGCCT GGAGGGCCTGAACATGCTGAGCGACCCCTGCGCCCTGCTGCCGACCCCGCT GTGGAGGAGAGCTTCAGAAGCGACAGACTGCAG (SEQ ID NO: 1193)	-5.293256

NOTC2	WANCSSPLPCWD YINNQCDELCNTV ECLFDNFEC (SEQ ID NO: 595)	EACNSHACQWDGGDCSLTME NPWANCSSPLPCWDYINNQC DELCNTVECLFDNFECQGNSKT CKYDKYCADHFKDNHCDQ (SEQ ID NO: 596)	GAGGCCTGCAACAGCCACGCCTGCCAGTGGGACGGCGGCGACTGCAGCCTG ACCATGGAGAACCCCTGGGCCAACTGCAGCAGCCCCCTGCCCTGCTGCTGGGACT ACATCAACAACCAGTGCGACGAGCTGTGCAACACCGTGGAGTGCCTGTTCGA CAACTTCGAGTGCCAGGGCAACAGCAAGACCTGCAAGTACGACAAGTACTGC GCCGACCACTTCAAGGACAACCACTGCGACCAG (SEQ ID NO: 1194)	-5.2584004
CACO1	KECPICKERFPAES DKDALEDHMDGH (SEQ ID NO: 597)	SGGEEANLLLPELGSAFYDMAS GFTVGTLSETSTGGPATPTWKE CPICKERFPAESDKDALEDHMD GHFFFSTQDPFTFE (SEQ ID NO: 598)	AGCGGCGGCGAGGAGGCCAACCTGCTGCTGCCCGAGCTGGGCAGCGCTTTC TACGACATGGCCAGCGGCTTCACCGTGGGCACCCTGAGCGAGACCAGCACCG GCGGACCCGCTACCCCCACATGGAAGGAGTGCCCCATCTGCAAGGAGAGATT CCCCGCCGAGAGCGACAAGGACGCCCTGGAGGACCACATGGACGGCCACTT CTTCTTCAGCACCCAGGACCCCTTCACCTTCGAG (SEQ ID NO: 1195)	-4.6832738
PYGO1	PCGICTNEVNDDQ DAILCEASCQKWF HRICTGMTETAYG LLTAEASAVWGCD TCMA (SEQ ID NO: 599)	RHGHSSSDPVYPCGICTNEVND DQDAILCEASCQKWFHRICTG MTETAYGLLTAEASAVWGCDT CMADKDVQLMRTRETF (SEQ ID NO: 600)	AGACACGGCCACAGCAGCAGCGACCCCGTGTACCCCTGCGGCATCTGCACCA ACGAGGTGAACGACGACCAGGACGCCATCCTGTGCGAGGCCAGCTGCCAGA AGTGGTTCCACAGAATCTGCACCGGCATGACCGAGACCGCCTACGGCCTGCT GACCGCCGAGGCCAGCGCTGTGTGGGGATGCGACACCTGCATGGCCGACAA GGACGTGCAGCTGATGAGAACCAGAGAGACCTTC (SEQ ID NO: 1196)	-4.3430928
IKKA	AQMIEENLNCLGH LSTIIHEANEEQGN SMMNLDWSW (SEQ ID NO: 601)	LVGSSLEGAVTPQTSAWLPPTS AEHDHSLSCVVTPQDGETSAQ MIEENLNCLGHLSTIIHEANEEQ GNSMMNLDWSWLTE (SEQ ID NO: 602)	CTGGTGGGCAGCAGCCTGGAGGGCGCCGTGACCCCCCAGACAAGCGCTTGG CTGCCCCCTACCAGCGCTGAGCACGACCACAGCCTGAGCTGCGTGGTGACCC CCCAGGACGGCGAGACCAGCGCCCAGATGATCGAGGAGAACCTGAACTGCC TGGGCCACCTGAGCACCATCATCCACGAGGCCAACGAGGAGCAGGGCAACA GCATGATGAACCTGGACTGGAGCTGGCTGACCGAG (SEQ ID NO: 1197)	-4.3328612
APC16	DLAPPRKALFTYPK GAGEMLEDGSERF LCESVFSYQVASTL KQVKHDQQVARM EKLAGLVEELEADE WRFKPIEQLLGF (SEQ ID NO: 603)	DLAPPRKALFTYPKGAGEMLED GSERFLCESVFSYQVASTLKQVK HDQQVARMEKLAGLVEELEAD EWRFKPIEQLLGFT (SEQ ID NO: 604)	GACCTGGCCCCCCCCAGAAAGGCCCTGTTCACCTACCCCAAGGGCGCCGGCG AGATGCTGGAGGACGGCAGCGAGAGATTCCTGTGCGAGAGCGTGTTCAGCT ACCAGGTGGCCAGCACCCTGAAGCAGGTGAAGCACGACCAGCAGGTGGCCA GAATGGAGAAGCTGGCCGGCCTGGTGGAGGAGCTGGAGGCCGACGAGTGG AGATTCAAGCCCATCGAGCAGCTGCTGGGCTTCACC (SEQ ID NO: 1198)	-4.1227423
WWP2	LPPGWEKRQDNG RVYYVNHNTRTTQ WED (SEQ ID NO: 605)	AMQHFSQRFLYQSSSASTDHD PLGPLPPGWEKRQDNGRVYYV NHNTRTTQWEDPRTQGMIQE PALPPGWEMKYTSEGVRY (SEQ ID NO: 606)	GCCATGCAGCACTTCAGCCAGAGATTCCTGTACCAGAGCAGCAGCGCCAGCA CCGACCACGACCCCCTGGGCCCCCTTCCTCCTGGATGGGAGAAGAGACAGGA CAACGGCAGAGTGTACTACGTGAACCACAACACCAGAACCACCCAGTGGGA GGACCCCAGAACCCAGGGCATGATCCAGGAGCCCGCCCTGCCCCCCGGATGG GAGATGAAGTACACCAGCGAGGGCGTGAGATAC (SEQ ID NO: 1199)	-4.0489585

RIP	FLVQEVMEEEWN ALQSVENCPEDLA QLEELIDMAVLEEI QQELINQEQSIISEY EKSLQFDEKCLSIM (SEQ ID NO: 607)	NSQNSFLVQEVMEEEWNALQ SVENCPEDLAQLEELIDMAVLE EIQQELINQEQSIISEYEKSLQFD EKCLSIMLAEWEAN (SEQ ID NO: 608)	AACAGCCAGAACAGCTTCCTGGTGCAGGAGGTGATGGAGGAGGAGTGGAAC GCCCTGCAGAGCGTGGAGAACTGCCCCGAGGACCTGGCCCAGCTGGAGGAG CTGATCGACATGGCCGTGCTGGAGGAGATCCAGCAGGAGCTGATCAACCAG GAGCAGAGCATCATCAGCGAGTACGAGAAGAGCCTGCAGTTCGACGAGAAG TGCCTGAGCATCATGCTGGCCGAGTGGGAGGCCAAC (SEQ ID NO: 1200)	-3.97129
AF9	AYLDELVELHRRL MTLRERHILQQIV NLIEETGHFHITNT TFDFDLCSLDKTTV RKLQSYL (SEQ ID NO: 609)	KSDKQIKNGECDKAYLDELVEL HRRLMTLRERHILQQIVNLIEET GHFHITNTTFDFDLCSLDKTTVR KLQSYLETSGTS (SEQ ID NO: 610)	AAGAGCGACAAGCAGATCAAGAACGGCGAGTGCGACAAGGCCTACCTGGAC GAGCTGGTGGAGCTGCACAGAAGACTGATGACCCTGAGAGAGAGACACATC CTGCAGCAGATCGTGAACCTGATCGAGGAGACCGGCCACTTCCACATCACCA ACACCACCTTCGACTTCGACCTGTGCAGCCTGGACAAGACCACCGTGAGAAA GCTGCAGAGCTACCTGGAGACCAGCGGCACCAGC (SEQ ID NO: 1201)	-3.7419986
ZFP28	LVTIKNLAVDFRQQ LHPAQKNFCKNGI WENNSDLGSA (SEQ ID NO: 610)	LEYSLLGEHWDYDALFETQPGL VTIKNLAVDFRQQLHPAQKNFC KNGIWENNSDLGSAGHCVAKP DLVSLLEQEKEPWMV (SEQ ID NO: 612)	CTGGAGTACAGCCTGCTGGGCGAGCACTGGGACTACGACGCCCTGTTCGAGA CCCAGCCCGGCCTGGTGACCATCAAGAACCTGGCCGTGGACTTCAGACAGCA GCTGCACCCCGCCCAGAAGAACTTCTGCAAGAACGGCATCTGGGAGAACAAC AGCGACCTGGGCAGCGCCGGCCACTGCGTGGCCAAGCCCGACCTGGTTAGC CTGCTGGAGCAGGAGAAGGAGCCCTGGATGGTG (SEQ ID NO: 1202)	-3.7291024
WWP1	LPPGWEKRVDSTD RVYFVNHNTKTTQ WED (SEQ ID NO: 613)	AMQQFNQRYLYSASMLAAEN DPYGPLPPGWEKRVDSTDRVY FVNHNTKTTQWEDPRTQGLQ NEEPLPEGWEIRYTREGVR (SEQ ID NO: 614)	GCCATGCAGCAGTTCAACCAGAGATACCTGTACAGCGCCAGCATGCTGGCCG CCGAGAACGACCCCTACGGCCCCCTGCCTCCTGGCTGGGAGAAGAGAGTGG ACAGCACCGACAGAGTGTACTTCGTGAACCACACAACACCAAGACCACCCAGTG GGAGGACCCCAGAACCCAGGGCCTGCAGAACGAGGAGCCCCTGCCCGAGGG CTGGGAGATCAGATACACCAGAGAGGGCGTGAGA(SEQ ID NO: 1203)	-3.728405
DPY30	PTRAYLDQTVVPIL LQGLAVLAKERPP NPIEFLASYLLKNK(SEQ ID NO: 615)	EYGLTDNVERIVENEKINAEKSS KQKVDLQSLPTRAYLDQTVVPIL LQGLAVLAKERPPNPIEFLASYL LKNKAQFEDRN (SEQ ID NO: 616)	GAGTACGGCCTGACCGACAACGTGGAGAGAATCGTGGAGAACGAGAAGATC AACGCCGAGAAGAGCAGCAAGCAGAAGGTGGACCTGCAGAGCCTGCCCACC AGAGCCTACCTGGACCAGACCGTGGTGCCCATCCTGCTGCAGGGCCTGGCCG TGCTGGCCAAGGAGAGACCCCCCAACCCCATCGAGTTCCTGGCCAGCTACCT GCTGAAGAACAAGGCCCAGTTCGAGGACAGAAAC (SEQ ID NO: 1204)	-3.696281
KS6B2	LQSEEDVSQFDTRF TRQTPVDSPDDTA LSESANQAFIGFT (SEQ ID NO: 617)	HMNWDDLLAWRVDPPFRPCL QSEEDVSQFDTRFTRQTPVDSP DDTALSESANQAFLGFTYVAPS VLDSIKEGFSFQPKLR (SEQ ID NO: 618)	CACATGAACTGGGACGACCTGCTGGCCTGGAGAGTGGACCCCCCCTTCAGAC CCTGCCTGCAGAGCGAGGAGGACGTGAGCCAGTTCGACACCAGAATTCACCAG ACAGACCCCCGTGGACAGCCCCGACGACACCGCCCTGAGCGAGAGCGCCAA CCAGGCCTTCCTGGGCTTCACCTACGTGGCCCCCAGCGTGCTGGACAGCATCA AGGAGGGCTTCAGCTTCCAGCCCAAGCTGAGA (SEQ ID NO: 1205)	-3.4939583

PYGO2	PCGACRSEVNDDQ DAILCEASCQKWF HRECTGMTESAYG LLTTEASAVWACD LCLK (SEQ ID NO: 619)	SGPQPPPGLVYPCGACRSEVN DDQDAILCEASCQKWFHRECT GMTESAYGLLTTEASAVWACD LCLKTKEIQSVYIREGM (SEQ ID NO: 620)	AGCGGCCCCCAGCCCCCCCCCGGCCTGGTTTACCCTTGCGGAGCTTGCAGAA GCGAGGTGAACGACGACCAGGACGCCATCCTGTGCGAGGCCAGCTGCCAGA AGTGGTTCCACAGAGAGTGCACCGGCATGACCGAGAGCGCCTACGGCCTGCT GACCACCGAGGCCAGCGCCGTGTGGGCTTGCGACCTGTGCCTGAAGACCAA GGAGATCCAGAGCGTGTACATCAGAGAGGGCATG (SEQ ID NO: 1206)	-3.4423787
U2AF4	FEEVFTELQEKYGEI EEMNVCDNLGDH LVGNVYVKFRREE DGERAVAELSNR WFNGQAV (SEQ ID NO: 621)	VEVQEHYDSFFEEVFTELQEKY GEIEEMNVCDNLGDHLVGNVY VKFRREEDGERAVAELSNRWF NGQAVHGELSPVTDFR (SEQ ID NO: 622)	GTGGAGGTGCAGGAGCACTACGACAGCTTCTTCGAGGAGGTGTTCACCGAG CTGCAGGAGAAGTACGGCGAGATCGAGGAGATGAACGTGTGCGACAACCTG GGCGACCACCTGGTGGGCAACGTGTACGTGAAGTTCAGAAGAGGAGGAGAC GGCGAGAGAGCCGTGGCCGAGCTGAGCAACAGATGGTTCAACGGCCAGGCC GTGCACGGCGAGCTGAGCCCCGTGACCGACTTCAGA (SEQ ID NO: 1207)	-3.3553928
ITCH	LPPGWEQRVDQH GRVYYVDHVEKRT TWDR (SEQ ID NO: 623)	SGLIIPLTISGGSGPRPLNPVTQA PLPPGWEQRVDQHGRVYYVD HVEKRTTWDRPEPLPPGWERR VDNMGRIYYVDHFTR (SEQ ID NO: 624)	AGCGGCCTGATCATCCCCCTGACCATCAGCGGCGGCAGCGGCCCCAGACCCC TGAACCCCGTGACCCAAGCTCCTCTGCCCCCCGGATGGGAGCAGAGAGTGGA CCAGCACGGCAGAGTGTACTACGTGGACCACGTGGAGAAGAGGAACCACCTG GGACAGACCCGAGCCCCTGCCCCCCGGCTGGGAGAGAAGAGTGGACAACAT GGGCAGAATCTACTACGTGGACCACTTCACCAGA (SEQ ID NO: 1208)	-3.3366968
ENL	AYTDELVELHRRL MALRERNVLQQIV NLIEETGHFNVTNT TFDFDLFSLDETTV RKLQSCL (SEQ ID NO: 625)	SKPEKILKKGTYDKAYTDELVEL HRRLMALRERNVLQQIVNLIEE TGHFNVTNTTFDFDLFSLDETT VRKLQSCLEAVAT (SEQ ID NO: 626)	AGCAAGCCCGAGAAGATCCTGAAGAAGGGCACCTACGACAAGGCCTACACC GACGAGCTGGTGGAGCTGCACAGAAGACTGATGGCCCTGAGAGAGAAGAAC GTGCTGCAGCAGATCGTGAACCTGATCGAGGAGACCGGCCACTTCAACGTGA CCAACACCACCTTCGACTTCGACCTGTTCAGCCTGGACGAGACCACCGTGAGA AAGCTGCAGAGCTGCCTGGAGGCCGTGGCCACC (SEQ ID NO: 1209)	-3.3117985
STAT2	SQPVPEPDLPCDLR HLNTEPMEIFRNC VKIEEIMPNGDPLL AGQNTVDEVVVSR P (SEQ ID NO: 627)	SQTVPEPDQGPVSQPVPEPDLP CDLRHLNTEPMEIFRNCVKIEEI MPNGDPLLAGQNTVDEVVVSR PSHFYTDGPLMPSD (SEQ ID NO: 628)	AGCCAGACCGTGCCCGAGCCCGACCAGGGCCCCGTGAGCCAGCCTGTTCCTG AGCCTGACCTGCCCTGCGACCTGAGACACCTGAACACCGAGCCCATGGAGAT CTTCAGAAACTGCGTGAAGATCGAGGAGATCATGCCCAACGGCGACCCCCTG CTGGCCGGCCAGAACACCGTGGACGAGGTGTACGTGAGCAGACCCAGCCAC TTCTACACCGACGGCCCCCTGATGCCCAGCGAC (SEQ ID NO: 1210)	-3.1207026

NOTC1	EEACELPECQEDA GNKVCSLQCNNH ACGWDGGDC (SEQ ID NO: 629)	LCHILDYSFGGGAGRDIPPPLIEE ACELPECQEDAGNKVCSLQCN NHACGWDGGDCSLNFNDPW KNCTQSLQCWKYFSDG (SEQ ID NO: 630)	CTGTGCCACATCCTGGACTACAGCTTCGGCGGCGCCGGCAGAGACATCC CCCCCCCTCTGATCGAGGAGGCTTGCGAGCTGCCCGAGTGCCAGGAGGACGC CGGCAACAAGGTGTGCAGCCTGCAGTGCAACAACCACGCCTGCGGCTGGGA CGGCGGCGACTGCAGCCTGAACTTCAACGACCCCTGGAAGAACTGCACCCAG AGCCTGCAGTGCTGGAAGTACTTCAGCGACGGC (SEQ ID NO: 1211)	-3.1201108
CRTC3	PNTILPEDSSTSLFK DLNSALAGLPEVSL NVDTPFPLEEELQI EPLSLDGLNMLSDS SMGLLDPSVEETFR ADR (SEQ ID NO: 631)	NCGSLPNTILPEDSSTSLFKDLN SALAGLPEVSLNVDTPFPLEEEL QIEPLSLDGLNMLSDSSMGLLD PSVEETFRADRL (SEQ ID NO: 632)	AACTGCGGGCAGCCTGCCCAACACCATCCTGCCCGAGGACAGCAGCACCAGCC TGTTCAAGGACCTGAACAGCGCCCTGGCCGGCCTGCCCGAGGTGAGCCTGAA CGTGGACACCCCCTTCCCCCTGGAGGAGGAGCTGCAGATCGAGCCCCTGAGC CTGGACGGCCTGAACATGCTGAGCGACAGCAGCATGGGCCTGCTGGACCCC AGCGTGGAGGAGACCTTCAGAGCCGACAGACTG (SEQ ID NO: 1212)	-3.0736492
SAV1	LPPGWSVDWTMR GRKYYIDHNTNTT HWSH (SEQ ID NO: 633)	HASGIGRVAATSLGNLTNHGSE DLPLPPGWSVDWTMRGRKYYI DHNTNTTHWSHPLEREGLPPG WERVESSEFGTYYVDH (SEQ ID NO: 634)	CACGCCAGCGGCATCGGCAGAGTGGCCGCCACCAGCCTGGGCAACCTGACC AACCACGGCAGCGAGGACCTGCCCCCTGCCCCCCGGATGGAGCGTGGACTGG ACCATGAGAGGCAGAAAGTACTACATCGACCACAACACCAACACCACCCACT GGAGCCACCCCCTGGAGAGAGAGGGCCTGCCCCCCCGGATGGGAGAGAGTG GAGAGCAGCGAGTTCGGCACCTACTACGTGGACCAC (SEQ ID NO: 1213)	-2.9035402
DPF1	GEDFYREAIEHCRS YNARLCAERSLRLP FLDSQTGVAQNNC YIWMEKTHRGPGL APGQIYTYPARCW RKKR (SEQ ID NO: 635)	PLSLGEDFYREAIEHCRSYNARL CAERSLRLPFLDSQTGVAQNNC YIWMEKTHRGPGLAPGQIYTYP ARCWRKKRRLNIL (SEQ ID NO: 636)	CCCCTGAGCCTGGGCGAGGACTTCTACAGAGAGGCCATCGAGCACTGCAGAA GCTACAACGCCAGACTGTGCGCCGAGAGAAGCCTGAGACTGCCCTTCCTGGA CAGCCAGACCGGCGTGGCCCAGAACACAACTGCTACATCTGGATGGAGAAGAC CCACAGAGGCCCCGGCCTGGCCCCCGGCCAGATCTACACCTACCCCGCTAGA TGCTGGAGAAAGAAGAGAAGACTGAACATCCTG (SEQ ID NO: 1214)	-2.7433919
ABL1	VALYDFVASGDNT LSITKGEKLRVLGY NHNGEWCEAQTK NGQGWVP (SEQ ID NO: 637)	KENLLAGPSENDPNLFVALYDF VASGDNTLSITKGEKLRVLGYN HNGEWCEAQTKNGQGWVPS NYITPVNSLEKHSWYHG (SEQ ID NO: 638)	AAGGAGAACCTGCTGGCCGGCCCCAGCGAGAACGACCCCAACCTGTTCGTGG CCCTGTACGACTTCGTGGCCAGCGGCGACACAACACCCTGAGCATCACCAAGGG CGAGAAGCTGAGAGTGCTGGGCTACAACCACAACGGCGAGTGGTGCGAGGC CCAGACCAAGAACGGCCAGGGCTGGGTGCCCAGCAACTACATCACCCCCGTG AACAGCCTGGAGAAGCACAGCTGGTACCACGGC (SEQ ID NO: 1215)	-2.6728209

WBP4	VKTVWVEGLSEDG FTYYYNTETGESR WEK (SEQ ID NO: 639)	YYDLISGASQWEKPEGFQGDLK KTAVKTVWVEGLSEDGFTYYYN TETGESRWEKPDDFIPHTSDLP SSKVNENSLGTLDE (SEQ ID NO: 64 0)	TACTACGACCTGATCAGCGGCGCCAGCCAGTGGGAGAAGCCCGAGGGCTTC CAGGGCGACCTGAAGAAGACCGCCGTGTGGGTGGAGGGCCT GAGCGAGGACGGCTTCACCTACTACTACAACACCGAGACCGGCGAGAGCAG ATGGGAGAAGCCCGACGACTTCATCCCCCACACCAGCGACCTGCCCAGCAGC AAGGTGAACGAGAACAGCCTGGGCACCCTGGACGAG (SEQ ID NO: 1216)	-2.6121807
BTK	VALYDYMPMNAN DLQLRKGDEYFILE ESNLPWWRARDK NGQEGYIP (SEQ ID NO: 641)	PEPAAAPVSTSELKKVVALYDY MPMNANDLQLRKGDEYFILEE SNLPWWRARDKNGQEGYIPS NYVTEAEDSIEMYEWYS (SEQ ID NO: 642)	CCCGAGCCCGGCCGCCGCCCCCGTGAGCACCAGCGAGCTGAAGAAGGTGGTG GCTCTGTACGACTACATCCAGAAGCAGGGCTACATCCCCAGCAACTACGTGACCGAGG CCGAGGACAGCATCGAGATGTACGAGTGGTACAGC (SEQ ID NO: 1217)	-2.5651252
SMRC2	TREWTEQETLLLLE ALEMYKDDWNKV SEHVGSRTQDECIL HFLRL (SEQ ID NO: 643)	MYTKKNVPSKSKAAASATREW TEQETLLLLEALEMYKDDWNKV SEHVGSRTQDECILHFLRLPIED PYLEDSEASLGPLA (SEQ ID NO: 644)	ATGTACACCAAGAAGAACGTGCCCAGCAAGAGCAAGGCCGCCGCCAGCGCC ACCAGAGAGTGGACCGAGCAGGAGACCCTGCTGCTGCTGGAGGCCCTGGAG ATGTACAAGGACGACTGGAACAAGGTGAGCGAGCACGTGGGCAGCAGAACC CAGGACGAGTGCATCCTGCACTTCCTGAGACTGCCCATCGAGGACCCCTACCT GGAGGACAGCGAGGCCAGCCTGGGCCCCCTGGCC (SEQ ID NO: 1218)	-2.4978538
MTA3	CESCYATQSHQWY SWGPPNMQCRLC AICWLYWKKYGG(SEQ ID NO: 645)	GAVNGAVGTTFQPQNPLLGRA CESCYATQSHQWYSWGPPNM QCRLCAICWLYWKKYGGLKMP TQSEEEKLSPSPTTEDPR (SEQ ID NO: 646)	GGCGCCGTGAACGGCGCCGTGGGCACCACCTTCCAGCCCCAGAACCCCCTGC TGGGCAGAGCCTGCGAGAGCTGCTACGCCACCCAGAGCCACCAGTGGTACA GCTGGGGCCCCCCCAACATGCAGTGCAGACTGTGCGCCATCTGCTGGCTGTA CTGGAAGAAGTACGGCGGCCTGAAGATGCCCACCCAGAGCGAGGAGGAGA AGCTGAGCCCCAGCCCCACCACCGAGGACCCCAGA (SEQ ID NO: 1219)	-2.4098352
WWTR1	LPPGWEMTFTAT GQRYFLNHIEKITT WQD (SEQ ID NO: 647)	GAAGSPAQQHAHLRQQSYDV TDELPLPPGWEMTFTATGQRY FLNHIEKITTWQDPRKAMNQP LNHMNLHPAVSSTPVPQR (SEQ ID NO: 648)	GGCGCCGCCGGCAGCCCCGCCCAGCAGCACGCCCACCTGAGACAACAAAGCT ACGACGTGACCGACGAGCTGCCCCTGCCCCCCGGCTGGGAGATGACCTTCAC CGCCACCGGACAGAGATACTTCCTGAACCACATCGAGAAGATCACCACCTGG CAGGACCCCAGAAAGGCCATGAACCAGCCCCTGAACCACATGAACCTGCACC CCGCCGTGAGCAGCACCCCCGTGCCCCAGAGA (SEQ ID NO: 1220)	-2.3989581
EGR3	GKFAFDSPSNWC QDNIISLMSAGILG VPPASGALSTQTST ASMVQPPQGDVE AMYPALPPYSNCG DLYS (SEQ ID NO: 649)	TVTYLGKFAFDSPSNWCQDNII SLMSAGILGVPPASGALSTQTS TASMVQPPQGDVEAMYPALP PYSNCGDLYSEPVSFH(SEQ ID NO: 650)	ACCGTGACCTACCTGGGCAAGTTCGCCTTCGACAGCCCCAGCAACTGGTGCC AGGACAACATCATCAGCCTGATGAGCGCCGGCATCCTGGGCGTGCCCCCCGC TAGCGGAGCTCTGAGCACACAGACCAGCACCGCCAGCATGGTGCAGCCCCCC CAGGGCGACGTGGAGGCCATGTACCCCGCCCTTCCTCCCTACAGCAACTGCG GCGACCTGTACAGCGAGCCCGTGAGCTTCCAC (SEQ ID NO: 1221)	-2.337045

NFIX	LTQDEFHPFIEALLP HVRAFSYTWFNLQ ARKRKYFKKHE (SEQ ID NO: 651)	YSPYCLTQDEFHPFIEALLPHVR AFSYTWFNLQARKRKYFKKHEK RMSKDEERAVKDELLGEKPEIK QKWASRLLAKLRK (SEQ ID NO: 652)	TACAGCCCCTACTGCCTGACCCAGGACGAGTTCCACCCCTTCATCGAGGCCCT GCTGCCCCACGTGAGAGCCTTCAGCTACACCTGGTTCAACCTGCAGGCCAGA AAGAGAAAGTACTTCAAGAAGCACGAGAAGAGAATGAGCAAGGACGAGGA GAGAGCCGTGAAGGACGAGCTGCTGGGCGAGAAGCCCGAGATCAAGCAGA AGTGGGCCAGCAGACTGCTGGCCAAGCTGAGAAAG (SEQ ID NO: 1222)	-2.289111
KPCI	SGEFGLDNFDSQF TNEPVQLTPDDDD IVRKIDQSEFEGFE (SEQ ID NO: 653)	QGHPFFRNVDWDMMEQKQV VPPFKPNISGEFGLDNFDSQFT NEPVQLTPDDDDIVRKIDQSEF EGFEYINPLLMSAEECV (SEQ ID NO: 654)	CAGGGCCACCCCTTCTTCAGAAACGTGGACTGGGACATGATGGAGCAGAAGC AGGTGGTGCCCCCCTTCAAGCCCAACATCAGCGGCGAGTTCGGCCTGGACAA CTTCGACAGCCAGTTCACCAACGAGCCCGTGCAGCTGACCCCCGACGACGAC GACATCGTGAGAAAGATCGACCAGAGCGAGTTCGAGGGCTTCGAGTACATC AACCCCCTGCTGATGAGCGCCGAGGAGTGCGTG (SEQ ID NO: 1223)	-2.2334296
LMBL1	MKLEAVDRRNPAL IRVASVEDVEDHRI KIHFDGWSHGYDF WIDADHPDIHPAG WCSKTGHPLQPPL GP (SEQ ID NO: 655)	SFLVNMKLEAVDRRNPALIRVA SVEDVEDHRIKIHFDGWSHGY DFWIDADHPDIHPAGWCSKTG HPLQPPLGPREPSSAS (SEQ ID NO: 656)	AGCTTCCTGGTGAACATGAAGCTGGAGGCCGTGGACAGAAGAAACCCCGCC CTGATCAGAGTGGCCAGCGTGGAGGACGTGGAGGACCACAGAATCAAGATC CACTTCGACGGCTGGAGCCACGGCTACGACTTCTGGATCGACGCCGACCACC CCGACATCCACCCCGCCGGCTGGTGCAGCAAGACCGGACACCCCCTGCAGCC TCCTCTGGGCCCCAGAGAGCCCAGCAGCGCTAGC (SEQ ID NO: 1224)	-2.2075416
NOTC1	WKNCTQSLQCWK YFSDGHCDSQCNS AGCLFDGFDC (SEQ ID NO: 657)	LQCNNHACGWDGGDCSLNFN DPWKNCTQSLQCWKYFSDGH CDSQCNSAGCLFDGFDCQRAE GQCNPLYDQYCKDHFSDGH (SEQ ID NO: 658)	CTGCAGTGCAACAACCACGCCTGCGGCTGGGACGGCGGCGACTGCAGCCTG AACTTCAACGACCCCTGGAAGAACTGCACCCAGAGCCTGCAGTGCTGGAAGT ACTTCAGCGACGGCCACTGCGACAGCCAGTGCAACAGCGCCGGCTGCCTGTT CGACGGCTTCGACTGCCAGAGAGCCGAGGGCCAGTGCAACCCCCTGTACGAC CAGTACTGCAAGGACCACTTCAGCGACGGCCAC (SEQ ID NO: 1225)	-2.1840238
FIGN	TDLSAIMPSQLRPV TYQDFENAFCKIQP SISQKELDMYVEW NKMF (SEQ ID NO: 659)	LLVQRTEGFSGLDVAHLCQEAV VGPLHAMPATDLSAIMPSQLR PVTYQDFENAFCKIQPSISQKEL DMYVEWNKMFGCSQ (SEQ ID NO: 660)	CTGCTGGTGCAGAGAACCGAGGGCTTCAGCGGCCTGGACGTGGCCCACCTGT GCCAGGAGGCCGTGGTGGGACCTCTGCACGCCATGCCCGCCACCGACCTGA GCGCCATCATGCCCAGCCAGCTGAGACCCGTGACCTACCAGGACTTCGAGAA CGCCTTCTGCAAGATCCAGCCCAGCATCAGCCAGAAGGAGCTGGACATGTAC GTGGAGTGGAACAAGATGTTCGGCTGCAGCCAG (SEQ ID NO: 1226)	-2.1805996
IMA5	NPYCALIEEAYGLD KIEFLQSHENQEIY QKAFDLIEHYFGTE DEDSSIAPQ (SEQ ID NO: 661)	RLGEQEAKRNGTGINPYCALIEE AYGLDKIEFLQSHENQEIYQKAF DLIEHYFGTEDEDSSIAPQVDLN QQQYIFQQCEA (SEQ ID NO: 662)	AGACTGGGCGAGCAGGAGGCCAAGAGAAACGGCACCGGCATCAACCCCTAC TGCGCCCTGATCGAGGAGGCCTACGGCCTGGACAAGATCGAGTTCCTGCAGA GCCACGAGAACCAGGAGATCTACCAGAAGGCCTTCGACCTGATCGAGCACTA CTTCGGCACCGAGGACGAGCAGCATCGCCCCCCAGGTGGACCTGAA CCAGCAGCAGTACATCTTCCAGCAGTGCGAGGCC (SEQ ID NO: 1227)	-2.15155

ZN496	PFKDMILCFSEED WSLLDPAQTGFYG EFIIGEDY (SEQ ID NO: 663)	QEENVRDTQQVTTLQLPPSRVS PFKDMILCFSEEDWSLLDPAQT GFYGEFIIGEDYGVSMPPNDLA AQPDLSQGEENEPR (SEQ ID NO: 664)	CAGGAGGAGAACGTGAGAGACACCCAGCAGGTGACCACCCTGCAGCTGCCC CCCAGCAGAGTGAGCCCCTTCAAGGACATGATCCTGTGCTTCAGCGAGGAGG ACTGGAGCCTGCTGGACCCCGCCCAGACCGGCTTCTACGGCGAGTTCATCAT CGGCGAGGACTACGGCGTGAGCATGCCCCCCAACGACCTGGCCGCCCAGCCC GACCTTAGCCAGGGCGAGGAGAACGAGCCCAGA (SEQ ID NO: 1228)	-2.1412028

表 **3: KRAB** 阻抑蛋白突变体

变体	突变的开始	氨基酸序列	**DNA** 序列	归一化的平均值 **D13 (0 =** 野生型分数)
谷氨酸;3;5	5	DAKSLEEESRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 665)	GATGCCAAGTCCCTTGAGGAGGAGTCTCGGACCCTGGTTACCTTTAAGGATG TGTTCGTCGACTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1229)	1.28823665
谷氨酸;3;7	7	DAKSLTAEEETLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 666)	GATGCCAAGTCCTTGACCGCTGAGGAGGAGACCCTGGTCACTTTCAAGGACG TTTTTGTGGATTTCACGAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1230)	1.14468005
天冬氨酸;3;7	7	DAKSLTADDDTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 667)	GATGCTAAGTCCCTGACCGCCGATGATGACACCCTTGTGACTTTTAAGGATGT CTTCGTTGACTTCACGAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1231)	1.10622079
异亮氨酸;3;67	67	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKIIIP (SEQ ID NO: 668)	GATGCTAAGTCTCTGACCGCCTGGTCCCGGACTCTTGTTACGTTTAAGGACGT GTTCGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAGAACTACAAAAACCTAGTAAGCCTAG GATACCAACTCACAAAACCAGACGTAATACTAAGACTAGAAAAATAATAAT ACCA (SEQ ID NO: 1232)	1.08345235

脯氨酸:3;6	6	DAKSLTPPPRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 669)	GATGCCAAGTCCCTTACCCCCCCTCCCCGGACTCTGGTGACGTTTAAGGATGT CTTCGTTGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1233)	1.06556067
天冬酰胺:1;1	1	DNKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 670)	GATAATAAGTCCCTTACCGCCTGGTCTCGGACGCTGGTTACTTTCAAGGACGT CTTCGTGGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1234)	1.05496491
脯氨酸:3;9	9	DAKSLTAWSPPPVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 671)	GATGCTAAGTCTTTGACTGCCTGGTCCCCTCCCCCGTTACCTTCAAGGATGT GTTTGTCGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1235)	1.05426168
谷氨酸:2;6	6	DAKSLTEESRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 672)	GACGCCAAGTCCCTGACGGAGGAGTCTCGGACCCTTGTGACTTTTAAGGATG TCTTCGTTGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1236)	1.04111335
丙氨酸:2;7	7	DAKSLTAAARTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 673)	GACGCTAAGTCCCTTACCGCCGCCGCGAGGACCCTGGTTACTTTTAAGGATG TCTTCGTGGATTTCACGCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1237)	1.04035858
缬氨酸:1;4	4	DAKSVTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 674)	GATGCCAAGTCTGTGACCGCTTGGTCCCGGACGCTTGTCACCTTCAAGGACG TTTTTGTGGATTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1238)	0.96430263

脯氨酸;1;7	7	DAKSLTAPSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P(SEQ ID NO: 675)	GATGCTAAGTCTCTTACCGCCCCCTCCCGGACTCTGGTCACCTTCAAGGACGT GTTCGTTGATTTTACGAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1239)	0.96424497
甘氨酸;2;7	7	DAKSLTAGGRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 676)	GACGCCAAGTCCCTTACCGCTGGCGGGCGGACCCTGGTGACTTTCAAGGATG TTTTCGTCGATTTTACGAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1240)	0.94837463
天冬酰胺;3;5	5	DAKSLNNNSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 677)	GATGCCAAGTCTCTTAATAATAATTCCCGGACTCTGGTGACCTTTAAGGACGT TTTCGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1241)	0.92066978
谷氨酰胺;3;7	7	DAKSLTAQQQTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 678)	GATGCCAAGTCCCTTACCGCTCAGCAGCAGACTCTGGTCACGTTCAAGGACG TTTTTGTGGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1242)	0.91183545
脯氨酸;2;10	10	DAKSLTAWSRPPVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 679)	GATGCTAAGTCTCTTACTGCCTGGTCCCGGCCTCCCGTGACCTTCAAGGATGT CTTCGTTGACTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1243)	0.89572995
苏氨酸;2;7	7	DAKSLTATTRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 680)	GACGCTAAGTCCCTGACCGCCACGACCCGGACTCTTGTCACTTTCAAGGATGT GTTCGTTGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1244)	0.88884291

谷氨酸:2;10	10	DAKSLTAWSREEVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 681)	GATGCCAAGTCTCTTACCGCTTGGTCCAGGGAGGAGGTCACTTTCAAGGACG TGTTCGTTGATTTTACCCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1245)	0.86791044
天冬氨酸:2;7	7	DAKSLTADDRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 682)	GATGCCAAGTCCCTTACCGCTGATGACCGGACCCTGGTCACGTTCAAGGATG TGTTTGTTGACTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1246)	0.86480685
谷氨酰胺:3;6	6	DAKSLTQQQRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 683)	GATGCCAAGTCCCTTACGCAGCAGCAGCGGACTCTGGTTACCTTCAAGGACG TGTTTGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1247)	0.86314843
谷氨酸:3;4	4	DAKSEEEWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 684)	GACGCCAAGTCCGAGGAGGAGTGGTCTAGGACCCTTGTCACCTTTAAGGATG TTTTCGTGGATTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1248)	0.84553985
天冬氨酸:2;6	6	DAKSLTDDSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 685)	GATGCCAAGTCCCTTACCGACGACTCTAGGACGCTGGTTACTTTTAAGGATGT GTTCGTCGATTTCACCCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1249)	0.84522896
天冬酰胺:2;7	7	DAKSLTANNRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 686)	GACGCTAAGTCCCTGACGGCCAATAATCGGACCCTTGTTACCTTTAAGGATGT GTTCGTCGATTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1250)	0.84228978

天冬酰胺;3;0	0	NNNSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 687)	AATAATAATTCTCTTACCGCCTGGTCCAGGACGCTGGTGACCTTCAAGGATGT TTTTGTCGATTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1251)	0.83772353
甘氨酸;3;5	5	DAKSLGGGSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 688)	GATGCCAAGTCCCTGGGCGGTGGGTCTCGGACCCTTGTTACTTTTAAGGACG TGTTCGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1252)	0.83010312
天冬氨酸;3;6	6	DAKSLTDDDRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 689)	GATGCCAAGTCCCTTACCGACGACGATAGGACTCTGGTTACCTTCAAGGATG TGTTCGTCGATTTTACGCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1253)	0.82122205
天冬氨酸;3;5	5	DAKSLDDDSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 690)	GATGCCAAGTCTCTGGACGATGATTCCCGGACCCTTGTCACCTTTAAGGATGT GTTCGTTGACTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1254)	0.81720761
赖氨酸;2;62	62	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVIKKLEKGE EP (SEQ ID NO: 691)	GACGCCAAGTCTCTGACTGCTTGGTCCCGGACCCTGGTGACGTTTAAGGATG TCTTCGTTGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATAAAAAAACTAGAAAAAAGGAGAA GAACCA(SEQ ID NO: 1255)	0.79372667
天冬氨酸;1;7	7	DAKSLTADSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P(SEQ ID NO: 692)	GATGCTAAGTCCCTTACCGCCGATTCTAGGACGCTGGTGACCTTCAAGGATG TCTTTGTTGACTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1256)	0.77854784

赖氨酸;3;5	5	DAKSLKKKSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 693)	GATGCCAAGTCTCTGAAGAAGAAGTCCCGGACCCTTGTCACTTTCAAGGACG TTTTTGTGGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1257)	0.76670536
谷氨酸;1;7	7	DAKSLTAESRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 694)	GATGCCAAGTCTCTGACCGCTGAGTCCAGGACTCTTGTTACGTTCAAGGATGT CTTTGTGGACTTCACCCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1258)	0.76316547
脯氨酸;2;7	7	DAKSLTAPPRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 695)	GACGCCAAGTCCCTTACGGCTCCCCCTCGGACCCTGGTCACTTTCAAGGATGT TTTCGTGGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1259)	0.7537889
脯氨酸;3;7	7	DAKSLTAPPPTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 696)	GATGCCAAGTCCCTTACTGCTCCTCCCCCCACGCTGGTTACCTTTAAGGACGT GTTCGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1260)	0.74610592
天冬氨酸;2;8	8	DAKSLTAWDDTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP(SEQ ID NO: 697)	GATGCTAAGTCCCTGACGGCCTGGGACGATACTTTGGTGACCTTTAAGGACG TCTTCGTTGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1261)	0.74252065
天冬酰胺;3;7	7	DAKSLTANNNTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 698)	GACGCTAAGTCCCTTACGGCCAATAATAATACTCTGGTTACCTTTAAGGATGT CTTCGTGGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1262)	0.72618378

甲硫氨酸;1;7	7	DAKSLTAMSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 699)	GATGCCAAGTCTCTGACTGCTATGTCCAGGACCCTTGTCACCTTCAAGGACGT GTTTGTTGATTTCACGCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1263)	0.71061025
天冬酰胺;3;8	8	DAKSLTAWNNLVTFK DVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLV SLGYQLTKPDVILRLEKG EEP (SEQ ID NO: 700)	GATGCCAAGTCCCTTACCGCTTGGAATAATAATCTGGTGACTTTCAAGGATGT TTTTGTCGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1264)	0.70490419
天冬酰胺;2;9	9	DAKSLTAWSNNLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 701)	GATGCTAAGTCCCTGACCGCCTGGTCTAATAATTTGGTTACCTTCAAGGATGT GTTTGTCGACTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1265)	0.7021223
脯氨酸;2;6	6	DAKSLTPPSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 702)	GACGCCAAGTCCCTTACGCCCCCTTCTCGGACTCTGGTGACCTTTAAGGATGT CTTCGTTGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1266)	0.70094498
谷氨酸;3;6	6	DAKSLTEEERTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 703)	GACGCCAAGTCCCTGACGGAGGAGGAGCGGACTCTTGTGACCTTTAAGGAT GTTTTCGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAAC AAATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCT AGGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGA AGAACCA (SEQ ID NO: 1267)	0.69641957
丙氨酸;3;5	5	DAKSLAAASRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 704)	GATGCTAAGTCTTTGGCCGCGGCCTCCAGGACCCTTGTGACTTTCAAGGACG TTTTTGTCGATTTCACCCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1268)	0.69263933

谷氨酰胺;2;7	7	DAKSLTAQQRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 705)	GATGCTAAGTCCCTGACCGCCCAGCAGCGGACGCTTGTTACTTTCAAGGATG TCTTTGTGGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1269)	0.67752105
谷氨酸;3;9	9	DAKSLTAWSEEEVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 706)	GATGCCAAGTCCCTTACCGCTTGGTCTGAGGAGGAGGTTACTTTCAAGGATG TGTTTGTCGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1270)	0.66806615
谷氨酰胺;2;6	6	DAKSLTQQSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 707)	GATGCCAAGTCTCTGACTCAGCAGTCCCGGACCCTTGTGACGTTCAAGGACG TTTTCGTCGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1271)	0.65472821
谷氨酸;1;39	39	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYREVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 708)	GACGCTAAGTCCCTTACGGCCTGGTCTAGGACCCTGGTCACCTTTAAGGATGT TTTCGTGGATTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAGAAGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1272)	0.65227203
脯氨酸;1;10	10	DAKSLTAWSRPLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 709)	GATGCCAAGTCTCTTACCGCTTGGTCCAGGCCCCTGGTGACTTTCAAGGATGT CTTTGTTGACTTCACCCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1273)	0.64837531
苏氨酸;1;4	4	DAKSTTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 710)	GACGCTAAGTCCACCACCGCCTGGTCTAGGACTCTTGTTACGTTCAAGGATGT GTTTGTCGATTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1274)	0.64360588

天冬酰胺;3;9	9	DAKSLTAWSNNNVTFK DVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLV SLGYQLTKPDVILRLEKG EEP (SEQ ID NO: 711)	GACGCCAAGTCTCTTACCGCTTGGTCCAATAATAATGTTACTTTCAAGGATGT GTTTGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1275)	0.63949165
天冬氨酸;3;8	8	DAKSLTAWDDDLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 712)	GATGCTAAGTCCCTTACTGCCTGGGACGATGATCTGGTCACCTTCAAGGACG TTTTTGTGGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1276)	0.63564982
丙氨酸;2;6	6	DAKSLTAASRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 713)	GATGCTAAGTCCCTTACTGCCGCCTCTCGGACGCTGGTGACCTTCAAGGACGT TTTCGTCGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1277)	0.63459967
甘氨酸;1;7	7	DAKSLTAGSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 714)	GACGCCAAGTCTTTGACGGCTGGCTCCCGGACCCTTGTGACTTTCAAGGATG TCTTCGTTGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1278)	0.6338712
丝氨酸;1;7	7	DAKSLTASSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 715)	GACGCTAAGAGTCTGACTGCCTCTTCCCGGACCCTTGTCACGTTCAAGGATGT TTTCGTGGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1279)	0.6299573
丝氨酸;2;6	6	DAKSLTSSSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 716)	GATGCCAAGTCCCTGACGTCTAGTTCAAGGACTCTTGTGACCTTTAAGGACGT CTTCGTTGATTTCACCCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1280)	0.61708486

天冬氨酸;2;3	3	DAKDDTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 717)	GATGCTAAGGATGACACTGCCTGGTCCCGGACGCTTGTGACCTTTAAGGATG TCTTCGTTGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1281)	0.61493283
组氨酸;1;55	55	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQHTKPDVILRLEKGE EP (SEQ ID NO: 718)	GATGCCAAGTCTCTGACCGCTTGGTCCAGGACGCTTGTGACTTTCAAGGATG TCTTCGTTGACTTTACCCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACACACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1282)	0.59443523
赖氨酸;2;6	6	DAKSLTKKSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 719)	GATGCCAAGTCCCTTACCAAGAAGTCTAGGACTCTGGTTACCTTCAAGGACG TCTTTGTGGATTTCACGCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1283)	0.59389075
谷氨酸;2;9	9	DAKSLTAWSEELVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 720)	GATGCCAAGTCCCTTACTGCTTGGTCTGAGGAGCTGGTGACCTTCAAGGATG TCTTTGTTGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1284)	0.58357862
谷氨酰胺;3;8	8	DAKSLTAWQQQLVTFK DVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLV SLGYQLTKPDVILRLEKG EEP (SEQ ID NO: 721)	GATGCCAAGTCCCTTACCGCTTGGCAGCAGCTGGTCACTTTCAAGGACG TGTTCGTTGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1285)	0.58274443
苏氨酸;1;11	11	DAKSLTAWSRTTVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 722)	GATGCCAAGTCCCTTACGGCTTGGTCTAGGACCACCGTGACTTTCAAGGACG TTTTCGTCGATTTTACTCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1286)	0.57209112

天冬酰胺;1;7	7	DAKSLTANSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 723)	GATGCCAAGTCTCTTACCGCTAATTCCAGGACGCTGGTCACCTTCAAGGATGT TTTTGTGGACTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1287)	0.57143202
谷氨酰胺;3;5	5	DAKSLQQQSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 724)	GACGCCAAGTCTCTGCAGCAGCAGTCCCGGACTCTTGTTACCTTCAAGGATGT GTTTGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1288)	0.57133084
脯氨酸;2;8	8	DAKSLTAWPPTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 725)	GATGCTAAGTCCCTTACCGCCTGGCCTCCCACCCTGGTTACTTTTAAGGATGT GTTCGTCGACTTCACGAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1289)	0.56714292
谷氨酸;3;3	3	DAKEEEAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 726)	GACGCTAAGGAGGAGGAGGCCTGGTCCAGGACCCTTGTTACCTTTAAGGAT GTGTTCGTCGATTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAAC AAATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCT AGGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGA AGAACCA (SEQ ID NO: 1290)	0.55531398
天冬酰胺;2;6	6	DAKSLTNNSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 727)	GATGCCAAGTCCCTGACCAATAATTCTCGGACCCTTGTCACTTTTAAGGACGT GTTCGTTGATTTCACGAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1291)	0.55430182
天冬氨酸;1;11	11	DAKSLTAWSRTDVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 728)	GATGCTAAGTCTCTTACGGCCTGGTCCAGGACCGATGTGACTTTCAAGGACG TTTTCGTCGATTTTACCCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1292)	0.55416963

苏氨酸:3;7	7	DAKSLTATTTTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 729)	GACGCCAAGTCCCTTACTGCTACTACCACCACGCTGGTTACCTTTAAGGATGT CTTCGTGGATTTCACAAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1293)	0.54915125
天冬氨酸:2;9	9	DAKSLTAWSDDLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 730)	GATGCCAAGTCCCTTACCGCTTGGTCTGATGACCTGGTTACCTTCAAGGACGT CTTCGTGGATTTTACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1294)	0.54802537
赖氨酸:1;7	7	DAKSLTAKSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 731)	GATGCTAAGTCCCTCACCGCCAAGTCTAGGACCTTGGTCACGTTCAAGGACG TTTTCGTGGATTTTACTCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1295)	0.54460274
谷氨酸:1;15	15	DAKSLTAWSRTLVTFED VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 732)	GATGCCAAGTCTCTGACCGCTTGGTCCCGGACCCTTGTCACGTTCGAGGATGT TTTCGTGGACTTTACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1296)	0.54206866
天冬氨酸:2;15	15	DAKSLTAWSRTLVTFDD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 733)	GATGCCAAGTCCCTGACTGCTTGGTCTAGGACCCTTGTGACCTTTGATGATGT TTTCGTCGACTTCACGCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1297)	0.526979
天冬氨酸:1;15	15	DAKSLTAWSRTLVTFDD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 734)	GACGCTAAGTCCCTTACCGCCTGGTCTCGGACGCTGGTGACCTTTGATGATGT CTTCGTTGATTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1298)	0.52579234

谷氨酰胺:3,9	9	DAKSLTAWSQQQVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP (SEQ ID NO: 735)	GACGCTAAGTCCCTTACCGCCTGGTCTCAGCAGCAGGTTACTTTCAAGGATGTCTTTGTGGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAAATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAGGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAGAACCA (SEQ ID NO: 1299)	0.52493949
谷氨酸:1,50	50	DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVELGYQLTKPDVILRLEKGEEP (SEQ ID NO: 736)	GATGCTAAGAGCCTTACCGCCTGGTCCCGGACCCTGGTGACTTTCAAGGACGTCTTCGTTGATTTTACGAGGGAAGAATGGAAACTACTAGACACAGCACAACAAATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGAACTAGGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAGAACCA (SEQ ID NO: 1300)	0.52368253
谷氨酸:3,8	8	DAKSLTAWEEELVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP (SEQ ID NO: 737)	GACGCCAAGTCCCTTACCGCTTGGGAGGAGGAGCTGGTCACTTTCAAGGATGTTTTTGTGGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAAATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAGGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAGAACCA (SEQ ID NO: 1301)	0.51955266
谷氨酸:1,35	35	DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQEVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP (SEQ ID NO: 738)	GATGCTAAGTCCCTGACCGCCTGGTCTCGGACCCTTGTGACGTTCAAGGACGTCTTGTTGATTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAAGAAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAGGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAGAACCA (SEQ ID NO: 1302)	0.5179103
脯氨酸:1,11	11	DAKSLTAWSRTPVTFKDVFVDFTREEWKLLDTAQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP (SEQ ID NO: 739)	GATGCCAAGTCTCTTACGGCTTGGTCCCGGACCCCCGTTACCTTTAAGGACGTCTTCGTGGATTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAAATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAGGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAGAACCA (SEQ ID NO: 1303)	0.51721922
谷氨酸:3,54	54	DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYEEEKPDVILRLEKGEEP (SEQ ID NO: 740)	GACGCCAAGTCCCTTACGGCTTGGTCTCGGACTCTGGTCACCTTTAAGGATGTGTTCGTTGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAAATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAGGATACGAAGAAGAAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAGAACCA (SEQ ID NO: 1304)	0.51510696

天冬氨酸;2;35	35	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QDDYRNVMLENYKNLV SLGYQLTKPDVILRLEKG EEP (SEQ ID NO: 741)	GATGCTAAGTCCCTGACTGCCTGGTCTAGGACCCTTGTCACCTTTAAGGATGT GTTCGTTGATTTCACGCGGGAAGAATGGAAACTACTAGACACAGCACAACAA GACGACTACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1305)	0.51046316
丙氨酸;3;6	6	DAKSLTAAARTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 742)	GACGCGAAGTCCCTGACGGCCGCTGCCCGGACTCTTGTTACCTTCAAGGATG TGTTCGTCGATTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1306)	0.50902755
谷氨酰胺;1;7	7	DAKSLTAQSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 743)	GATGCTAAGTCCCTCACGGCGGCCGCTGCCACCCTTGTCACCTTCAAGGATGT CTTCGTGGACTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1307)	0.50677669
丙氨酸;3;7	7	DAKSLTAAAATLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 744)	GATGCTAAGTCCCTCACGGCGGCCGCTGCCACCCTTGTCACCTTCAAGGACGT TTTCGTGGATTTTACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1308)	0.50591148
赖氨酸;2;5	5	DAKSLKKWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 745)	GATGCCAAGTCCCTGAAGAAGTGGTCTCGGACCCTTGTGACTTTTAAGGATG TTTTCGTCGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1309)	0.50454904
甘氨酸;3;7	7	DAKSLTAGGGTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 746)	GATGCTAAGTCCCTGACGGCCGGGGGTGGCACCCTTGTTACCTTCAAGGACG TCTTTGTGGATTTCACTAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1310)	0.49495873

天冬氨酸;3;0	0	DDDSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 747)	GATGATGATTCTCTGACGGCCTGGTCCCGGACTCTTGTCACCTTTAAGGACGT TTTCGTGGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1311)	0.48978651
谷氨酰胺;1;32	32	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTQQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 748)	GACGCCAAGTCCCTGACCGCTTGGTCTCGGACGCTTGTGACCTTTAAGGATG TTTTCGTCGATTTCACTAGGGAAGAATGGAAACTACTAGACACACAACACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1312)	0.48972298
酪氨酸;2;69	69	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE YY (SEQ ID NO: 749)	GATGCTAAGTCCCTTACCGCCGAGGAGAGGACTCTGGTCACCTTCAAGGACGT GTTCGTTGATTTTACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAT ACTAC (SEQ ID NO: 1313)	0.48765802
谷氨酸;2;7	7	DAKSLTAEERTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 750)	GATGCTAAGTCCCTTACGGCCGAGGAGAGGACTCTGGTCACCTTCAAGGACG TGTTCGTTGATTTTACCCGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1314)	0.48459406
组氨酸;1;7	7	DAKSLTAHSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 751)	GATGCCAAGTCTCTGACGGCTCACTCCAGGACTCTTGTCACCTTCAAGGATGT GTTCGTTGACTTTACCCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1315)	0.48125191
精氨酸;2;5	5	DAKSLRRWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 752)	GATGCCAAGTCCCTTCGCCGGTGGTCTCGAACTCTGGTCACCTTTAAGGATGT GTTCGTTGACTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1316)	0.4801544

丝氨酸:3;9	9	DAKSLTAWSSSSVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 753)	GATGCTAAGTCTCTTACTGCCTGGTCCAGCTCAAGTGTGACCTTTAAGGACGT TTTCGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1317)	0.47920968
色氨酸:1;53	53	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGWQLTKPDVILRLEKG EEP (SEQ ID NO: 754)	GATGCCAAGTCTCTTACCGCTTGGTCCCGGACTCTGGTCACGTTTAAGGACGT TTTCGTGGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATGGCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1318)	0.47320791
丝氨酸:3;1	1	DSSSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 755)	GATTCATCTTCCCTGACGGCCTGGAGTAGGACCCTTGTTACCTTCAAGGACGT CTTTGTGGATTTCACTCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1319)	0.47295365
赖氨酸:2;34	34	DAKSLTAWSRTLVTFKD VFVDFTREEWKLLDTAQ KKVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 756)	GATGCCAAGTCTCTGACCGCTTGGTCCCGGACTCTTGTGACGTTCAAGGACG TCTTTGTTGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAAAA AAAAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1320)	0.47168537
赖氨酸:3;7	7	DAKSLTAKKKTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 757)	GACGCTAAGTCCCTGACCGCCAAGAAGAAGACTCTTGTCACCTTCAAGGATG TTTTCGTGGATTTTACGAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1321)	0.47075243
甘氨酸:3;9	9	DAKSLTAWSGGGVTFK DVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLV SLGYQLTKPDVILRLEKG EEP (SEQ ID NO: 758)	GATGCTAAGTCCCTTACCGCCTGGTCTGGGGGTGGCGTGACCTTCAAGGATG TTTTCGTCGACTTTACTAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1322)	0.46950237

天冬氨酸;3;9	9	DAKSLTAWSDDDVTFK DVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLV SLGYQLTKPDVILRLEKG EEP (SEQ ID NO: 759)	GATGCCAAGTCCCTTACTGCTTGGTCTGACGATGATGTGACCTTCAAGGATGT TTTCGTCGACTTTACCAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1323)	0.4685333
苏氨酸;2;3	3	DAKTTTAWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 760)	GATGCTAAGACTACCACGGCCTGGTCCAGGACTCTTGTTACCTTTAAGGATGT CTTCGTGGACTTCACCCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1324)	0.46756614
甘氨酸;3;8	8	DAKSLTAWGGGLVTFK DVFVDFTREEWKLLDTA QQIVYRNVMLENYKNLV SLGYQLTKPDVILRLEKG EEP (SEQ ID NO: 761)	GATGCCAAGTCCCTTACTGCTTGGGGCGGTGGGCTGGTGACCTTCAAGGACG TTTTTGTCGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1325)	0.46569305
谷氨酸;1;10	10	DAKSLTAWSRELVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 762)	GACGCCAAGTCTCTGACTGCTTGGTCCCGGGAGCTTGTGACCTTTAAGGATG TCTTCGTTGATTTCACCAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1326)	0.46430578
天冬氨酸;3;4	4	DAKSDDDWSRTLVTFKD VFVDFTREEWKLLDTAQ QIVYRNVMLENYKNLVS LGYQLTKPDVILRLEKGE EP (SEQ ID NO: 763)	GACGCCAAGTCTGATGATGACTGGTCCCGGACCCTTGTCACCTTCAAGGATG TTTTCGTGGATTTTTACTAGGGAAGAATGGAAACTACTAGACACAGCACAACA AATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTA GGATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAA GAACCA (SEQ ID NO: 1327)	0.4639914
丝氨酸;3;5	5	DAKSLSSSRTLVTFKDV FVDFTREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 764)	GACGCCAAGTCACTTAGCAGTTCCTCTAGGACTCTGGTGACCTTTAAGGATGT CTTCGTTGATTTCACCCGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1328)	0.46366844

丝氨酸;3;7	7	DAKSLTASSSTLVTFKDV FVDFTIREEWKLLDTAQQ IVYRNVMLENYKNLVSL GYQLTKPDVILRLEKGEE P (SEQ ID NO: 765)	GACGCTAAGTCCCTTACGGCCTCTAGTTCAACCCTGGTGACCTTCAAGGATGT CTTCGTTGATTTTACTAGGGAAGAATGGAAACTACTAGACACAGCACAACAA ATAGTATACAGAAACGTAATGCTAGAAAACTACAAAAACCTAGTAAGCCTAG GATACCAACTAACAAAACCAGACGTAATACTAAGACTAGAAAAAGGAGAAG AACCA (SEQ ID NO: 1329)	0.45082917

表 4: 叠瓦式阻抑蛋白

基因	序列	平均第 5 天分数
KRBOX1	MMTAVSLTTRPQESVAFEDVAVYFTTKEWAIMVPAERALYRDVMLENYEAVAFVVPPTSKPALVSHLEQGKESCFTQPQG (SEQ ID NO: 766)	7.8278029
ZNF461	MAHELVMFRDVAIDVSQEEWECLNPAQRNLYKEVMLENYSNLVSLGLSVSKPAVISSLEQGKEPWMVVREETGRWCPGTW (SEQ ID NO: 767)	7.7007641
ZNF875	MSCSGAGGITAFVAFRDVAVYFTQEEWRLLSPAQRTLHREVMLETYNHLVSLEIPSSKPKLIAQLERGEAPWREERKCPL (SEQ ID NO: 768)	7.41648619
ZNF57	MDSVVFEDVAVDFTLEEWALLDSAQRDLYRDVMLETFRNLASVDDGTQFKANGSVSLQDMYGQEKSKEQTIPNFTGNNSC (SEQ ID NO: 769)	7.33652783
CBX5	NDIARGFERGLEPEKIIGATDSCGDLMFLMKWKDTDEADLVLAKEANVKCPQIVIAFYEERLTWHAYPEDAENKEKETAK (SEQ ID NO: 770)	6.80089192
CBX3	SKKKRDAADKPRGFARGLDPERIIGATDSSGELMFLMKWKDSDEADLVLAKEANMKCPQIVIAFYEERLTWHSCPEDEAQ (SEQ ID NO: 771)	6.7879261
CBX1	KKKKEESEKPRGFARGLEPERIIGATDSSGELMFLMKWKNSDEADLVPAKEANVKCPQVVISFYEERLTWHSYPSEDDDK (SEQ ID NO: 772)	6.68230107
CTCF	PVTVPVATTSVEELQGAYENEVSKEGLAESEPMICHTLPLPEGFQVVKVGANGEVETLEQGELPPQEDPSWQKDPDYQPP (SEQ ID NO: 773)	6.25722327
RYBP	PPSEANSIQSANATTKTSETNHTSRPRLKNVDRSTAQQLAVTVGNVTVIITDFKEKTRSSSTSSSTVTSSAGSEQQNQSS (SEQ ID NO: 774)	6.18591064
IRF2BP1	MASVQASRRQWCYLCDLPKMPWAMVWDFSEAVCRGCVNFEGADRIELLIDAARQLKRSHVLPEGRSPGPPALKHPATKDL (SEQ ID NO: 775)	5.6588422
MGA	PHTSANLVMTPQGQLITLKGPLFSGPVVAVSPDLLESDLKPQVAGSAVALPENDDLFMMPRIVNVTSLATEGGLVDMGGS (SEQ ID NO: 776)	5.6156366

CBX7	ELFLQEPPAPDVLQAAGEWEPAAQPPEEEADADLAEGPPPWTPALPSSEVTVTDITANSITVTFREAQAAEGFFRDRSGK (SEQ ID NO: 777)	5.00371758
IKZF5	TPSIGNSQPSTPAPAIPVQDPQLLHHCQHCDMYFADNILYTIHMGCHGYENPFQCNICGCKCNKYDFACHFARGQHNQH (SEQ ID NO: 778)	4.98433253
REST	IHSHEGSDLSDNMSEGSDDSGLHGARPVPQESSRKNAKEALAVKAAKGDFVCIFCDRSFRKGKDYSKHLNRHLVNVYYLE (SEQ ID NO: 779)	4.52776703
CBX4	RSEAGEPPSSLQVKPETPASAAVAVAAAAAPTTTAEKPPAEAQDEPAESLSEFKPFFGNIIITDVTANCLTVTFKEYVTV (SEQ ID NO: 780)	4.52247518
KLF10	APFKEEEKSPVSAPKLPKAQATSVIRHTADAQLCNHQTCPMKAASILNYQNNSFRRRTHLNVEAARKNIPCAAVSPNRSK (SEQ ID NO: 781)	4.49441205
SCMH1	DASRLSGRDPSSWTVEDVMQFVREADPQLGPHADLFRKHEIDGKALLLLRSDMMMKYMGLKLGPALKLSYHIDRLKQGKF (SEQ ID NO: 782)	4.38546257
SCML2	KQGFSKDPSTWSVDEVIQFMKHTDPQISGPLADLFRQHEIDGKALFLLKSDVMMKYMGLKLGPALKLCYYIEKLKEGKYS (SEQ ID NO: 783)	4.3468225
HIVEP3	GSGSESGKERRTTSKEISVIQHTSSFEKSDSLEQPSGLEGEDKPLAQFPSPPPAPHGRSAHSLQPKLVRQPNIQVPEILV (SEQ ID NO: 784)	4.34368367
HSF2	SSAQKVQIKQETIESRLSELKSENESLWKEVSELRAKHAQQQQVIRKIVQFIVTLVQNNQLVSLKRKRPLLLNTNGAQKK (SEQ ID NO: 785)	4.31444003
MBD1	TSPVLVPGCPSKAVDPGLPSVKQEPPDPEEDKEENKDDSASKLAPEEEAGGAGTPVITEIFSLGGTRFRDTAVWLPRSKD (SEQ ID NO: 786)	4.25873478
BAZ2A	GQDSEQPQAQLQPEAQLHAPAQPQPQLQLQLQSHKGFLEQEGSPLSLGQSQHDLSQSAFLSWLSQTQSHSSLLSSSVLTP (SEQ ID NO: 787)	4.17482075
CHD4	EIILCDTCPRAYHMVCLDPDMEKAPEGKWSCPHCEKEGIQWEAKEDNSEGEEILEEVGGDLEEEDDHHMEFCRVCKDGGE (SEQ ID NO: 788)	4.15410157
ATF7IP	DSSGVIDLTMDDEESGASQDPKKLNHTPVSTMSSSQPVSRPLQPIQPAPPLQPSGVPTSGPSQTTIHLLPTAPTTVNVTH (SEQ ID NO: 789)	4.13988404
PCGF2	MHRTTRIKITELNPHLMCALCGGYFIDATTIVECLHSFCKTCIVRYLETNKYCPMCDVQVHKTRPLLSIRSDKTLQDIVY (SEQ ID NO: 790)	4.08140835
HSF1	KIRQDSVTKLLTDVQLMKGKQECMDSKLLAMKHENEALWREVASLRQKHAQQQKVVNKLIQFLISLVQSNRILGVKRKIP (SEQ ID NO: 791)	3.89748786

WIZ	LQQKLEEVRQPPPRVRPVPSLVPRPPQTSLVKFVGNIYTLKCRFCEVEFQGPLSIQEEWVRHLQRHILEMNFSKADPPPE (SEQ ID NO: 792)	3.88045025
CBX7	MELSAIGEQVFAVESIRKKRVRKGKVEYLVKWKGWPPKYSTWEPEEHILDPRLVMAYEEKEERDRASGYRKRGPKPKRILL (SEQ ID NO: 793)	3.86396312
UHRF1	CGGRQDPDKQLMCDECDMAFHIYCLDPPLSSVPSEDEWYCPECRNDASEVVLAGERLRESKKKAKMASATSSSQRDWGKG (SEQ ID NO: 794)	3.85377779
MGA	ALSEVPQLKQEISECLIASSFEDDSRVASPLDQNGSFNVVIKEEPLDDYDYELGECPEGVTVKQEETDEETDVYSNSDDD (SEQ ID NO: 795)	3.82825762
SIN3B	SAGHEKLPVHVEDALTYLDQVKIRFGSDPATYNGFLEIMKEFKSQSIDTPGVIRRVSQLFHEHPDLIVGFNAFLPLGYRI (SEQ ID NO: 796)	3.67861396
MECP2	TSTQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVV (SEQ ID NO: 797)	3.54786112
ATF7IP	MDSLEEPQKKVFKARKTMRVSDRQQLEAVYKVKEELLKTDVKLLNGNHENGDLDPTSPLENMDYIKDKEEVNGIEEICFD (SEQ ID NO: 798)	3.46711815
INSM2	CHKKFRRQAYLRKHLSTHEAGSARALAPGFGSERGAPLAFACPLCGAHFPTADIREKHRLWHAVREELLLPALAGAPPET (SEQ ID NO: 799)	3.41894392
KDM5B	LLDSSNSSASASQAMNIKIEPEETTEARTHNLRRRMGCPTPKCENEKEMKSSIKQEPIERKDYIVENEKEKPKSRSKKAT (SEQ ID NO: 800)	3.41515767
TRIM24	PVSGSSPFATQVGVIRCPVCSQECAERHIIDNFFVKDTTEVPSSTVEKSNQVCTSCEDNAEANGFCVECVEWLCKTCIRA (SEQ ID NO: 801)	3.34747947
MBD1	MAEDWLDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFKQGILCYPAPKAHPVA (SEQ ID NO: 802)	3.31964355
SIN3A	PVQGQQQFQRLKVEDALSYLDQVKLQFGSQPQVYNDFLDIMKEFKSQSIDTPGVISRVSQLFKGHPDLIMGFNTFLPPGY (SEQ ID NO: 803)	3.30578618
DMD	RAKQPDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWTELTDWLSLLDQVIKSQRVMV (SEQ ID NO: 804)	3.29433336
ZNF827	HMRCHQHFLRTEAKVKEEIPDPDVKGSPHLSDSACLGQQREGGGTELVGTMMTSNTPERTSQGGAGVSPLLVKEEPKEDN (SEQ ID NO: 805)	3.26257067
HIPK2	QRQTIVIPDTPSPTVSVITISSDTDEEEEQKHAPTSTVSKQRKNVISCVTVHDSPYSDSSSNTSPYSVQQRAGHNNANAF (SEQ ID NO: 806)	3.2515196

MBD1	SGDGTQRQRLKTLCKDCRAQRIAFNREQRMFKRVGCGECAACQVTEDCGACSTCLLQLPHDVASGLFCKCERRRCLRIVE (SEQ ID NO: 807)	3.18634467
TET2	KMAEKAREKEEECEKYGPDYVPQKSHGKKVKREPAEPHETSEPTYLRFIKSLAERTMSVTTDSTVTTSPYAFTRVTGPYN (SEQ ID NO: 808)	3.17106197
CBX8	GSGPPSSGGGLYRDMGAQGGRPSLIARIPVARILGDPEEESWSPSLTNLEKVVVTDVTSNFLTVTIKESNTDQGFFKEKR (SEQ ID NO: 809)	3.14200153
HES3	GSTMDSAGLGQEAPALFRPCTPAVWAPAPAAGGPRSPPPLLLIPESLPGSSASVPPPQPASSRCAESPGLGLRVWRPWGS (SEQ ID NO: 810)	3.13473999
TRPS1	EDGHAISTIKEEPKIDFRVVNILLTPDSKMGEPVSESVVVKREKLEEKDGLKEKVWTESSSDDLRNVTWRGADILRGSPSYT (SEQ ID NO: 811)	3.11964253
TRIM24	VGSRGSSGSSSKPAGADSTHKVPVVMLEPIRIKQENSGPPENYDFPVVIVKQESDEESRPQNANYPRSILTSLLLNSSQS (SEQ ID NO: 812)	3.10950731
TET1	NEGDQPKTPENIPSKEPKDGSPVQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSENSSPSKSEKDEESEQRTASLLNSC (SEQ ID NO: 813)	3.01568183
BAZ2B	CRKGDNEELLLLCDGCDKGCHTYCHRPKITTIPDGDWFCPACIAKASGQTLKIKKLHVKGKKTNESKKGKKVTLTGDTED (SEQ ID NO: 814)	2.9949946
ATRX	SDEQKIKPVTENLVLSSHTGFCQSSGDEALSKSVPVTVDDDDDDNDPENRIAKKMLLEEIKANLSSDEDGSSDDEPEEGK (SEQ ID NO: 815)	2.99299407
DMD	QQHKVLQEDLEQEQVRVNSLTHMVVVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRLTEEQCL (SEQ ID NO: 816)	2.95216006
TRIM24	QVCTSCEDNAEANGFCVECVEWLCKTCIRAHQRVKFTKDHTVRQKEEVSPEAVGVTSQRPVFCPFHKKEQLKLYCETCDK (SEQ ID NO: 817)	2.94621718
AHRR	LDVSIKMEKDSGCEGAADGCVPSQVWLGASDRSHPATFPTRMHLKTEPDSRQQVYISHLGHGVRGAQPHGRATAGRSREL (SEQ ID NO: 818)	2.91519283
CBX4	MELPAVGEHVFAVESIEKKRIRKGRVEYLVKWRGWSPKYNTWEPEENILDPRLLIAFQNRERQEQLMGYRKRGPKPKPLV (SEQ ID NO: 819)	2.8451435
AUTS2	GSLQGAFQPKTSNPIDVAARPGTVPHTLLQKDPRLTDPFRPMLRKPGKWCAMHVHIAWQIYHHQQKVKKQMQSDPHKLD F (SEQ ID NO: 820)	2.80116399
PLAG1	LKVKTEPVDFLDPFTCNVSVPIKDELLPVMSLPSSELLSKPFTNTLQLNLYNTPFQSMQSSGSAHQMITTLPLGMTCPID (SEQ ID NO: 821)	2.78382024

ZNF827	LVKEEPKEDNGLPTSFTLNAADRPANHTKLKDPSEYVANSASALFSQDISVKMASDFLMKLSAANQKEPMNLNFKVKEEP (SEQ ID NO: 822)	2.76221951
TET3	EPQNHFSSFKYSGNAVVESYSVLGNCRPSDPYSMNSVYSYHSYYAQPSLTSVNGFHSKYALPSFSYYGFPSSNPVFPSQF (SEQ ID NO: 823)	2.67090027
RCOR3	IPEFDPGATKYTDKDNGGMLVWSPYHSIPDAKLDEYIAIAKEKHGYNVEQALGMLFWHKHNIEKSLADLPNFTPFPDEWT (SEQ ID NO: 824)	2.6595258
DNMT3B	KNPVSFHPLFEGGLCQTCRDRFLELFYMYDDDGYQSYCTVCCEGRELLCSNTSCCRCFCVECLEVLVGTGTAAEAKLQE (SEQ ID NO: 825)	2.65664419
IKZF2	DTTKAPKGSLKDIYKVFNGEGEQIRAFKCEHCRVLFLDHVMYTIHMGCHGYRDPLECNICGYRSQDRYEFSSHIVRGEHT (SEQ ID NO: 826)	2.65607702
DNMT3B	PAVRTRNNNSVSSRERHRPSPRSTRGRQGRNHVDESPVEFPATRSLRRRATASAGTPWPSPPSSYLTIDLTDDTEDTHGT (SEQ ID NO: 827)	2.64875126
ZNF827	AESYKETQMVKIKEEPMEVDIQDSHVSISPSRNVGYSTLIGREKTEPLQKMPEGRVPPERNLFSQDISVKMASELLFQLS (SEQ ID NO: 828)	2.6349517
ATF7IP	NQKLQKVIQWLLEEKLCALQCAVFDKTLAELKTRVEKIECNKRHKTVLTELQAKIARLTKRFEAAKEDLKKRHEHPPNPP (SEQ ID NO: 829)	2.62333283
KDM5C	VVVKEELGGDVKVESTSPKTFLESKEELSHSPEPCTKMTMRLRRNHSNAQFIESYVCRMCSRGDEDDKLLLCDGCDDNYH (SEQ ID NO: 830)	2.6222632 9
SUV39H1	MAENLKGCSVCCKSSWNQLQDLCRLAKLSCPALGISKRNLYDFEVEYLCDYKKIREQEYYLVKWRGYPDSESTWEPRQNL (SEQ ID NO: 831)	2.61048387
HES1	APCKLGSQAGEAAKVFGGFQVVPAPDGQFAFLIPNGAFAHSGPVIPVYTSNSGTSVGPNAVSPSSGPSLTADSMWRPWRN (SEQ ID NO: 832)	2.54719284
HIVEP3	VKKEDSKEQPDLPSLAPPSSLPLSETSSRPAKSQEGTDSKKVLQFPSLHTTTNVSWCYLNYIKPNHIQHADRRSSVYAGW (SEQ ID NO: 833)	2.54212172
ZNF446	QLSCSVKEEPNVDGQEVAPSSPPLAAQSPEGNHGHQEPASTSFHPPRIQEEWGLLDRSQKELYWDAMLEKYGTVVSLGLP (SEQ ID NO: 834)	2.53378337
HEY2	TSSVIRLNSPTTTSQIMARKKRRGIIEKRRRRDRINNSLSELRRLVPTAFEKQGSAKLEKAEILQMTVDHLKMLQATGGKG (SEQ ID NO: 835)	2.52262483
PRDM11	TVVKTEVCSPLRDQEYGQPCSRRPDSSAMEVEPKKLKGKRDLIVPKSFQQVDFWFCESCQEYFVDECPNHGPPVFVSDTP (SEQ ID NO: 836)	2.50622782

PHF19	EGQYVLCRWTDGLYYLGKIKRVSSSKQSCLVTFEDNSKYWVLWKDIQHAGVPGEEPKCNICLGKTSGPLNEILICGKCGL (SEQ ID NO: 837)	2.50329194
CBX3	KVEEAEPEEFVVEKVLDRRVVNGKVEYFLKWKGFTDADNTWEPEENLDCPELIEAFLNSQKAGKEKDGTKRKSLSDSESD (SEQ ID NO: 838)	2.46766682
FBRS	NPELPPRLGPVPSGLSQKGTQIPDHFRPPLRKPGKWCAMHVRVAYMILRHQEKMKGDSHKLDFRNDLLPCLPGPYGALPP (SEQ ID NO: 839)	2.4558407
CIC	FPSKVCLQLKIREVRQKIMQAATPTEQPPGAEAPLPVPPPTGTAAAPAPTPSPAGGPDPTSPSSDSGTAQAAPPLPPPPE (SEQ ID NO: 840)	2.42893202
PCGF6	EEEEEEEEEDMSHFSLRLEGGRQDSEDEEERLINLSELTPYILCSICKGYLIDATTITECLHTFCKSCIVRHFYYSNRCP (SEQ ID NO: 841)	2.4085226
MNT	DDKKTSNLSVLRTALRYIQSLKRKEKEYEHEMERLAREKIATQQRLAELKHELSQWMDVLEIDRVLRQTGQPEDDQASTS (SEQ ID NO: 842)	2.39586222
HSF1	LEHVHGSGPYSAPSPAYSSSSLYAPDAVASSGPIISDITELAPASPMASPGGSIDERPLSSSPLVRVKEEPPSPPQSPRV (SEQ ID NO: 843)	2.39135328
SERTAD2	ITTSTGFLTDLTLDDILFADIDTSMYDFDPCTSSSGTASKMAPVSADDLLKTLAPYSSQPVTPSQPFKMDLTELDHIMEV (SEQ ID NO: 844)	2.35712893
BMI1	MHRTTRIKITELNPHLMCVLCGGYFIDATTIIECLHSFCKTCIVRYLETSKYCPICDVQVHKTRPLLNIRSDKTLQDIVY (SEQ ID NO: 845)	2.34528179
TET3	RPRLPGPLPPGEAGLPAPSTRPLLSSEVPQISPQEGLPLSQSALSIAKEKNISLQTAIAIEALTQLSSALPQPSHSTPQA (SEQ ID NO: 846)	2.33879669
USP7	MYDPKTRSLNYCGHIYTPISCKIRDLLPVMCDRAGFIQDTSLILYEEVKPNLTERIQDYDVSLDKALDELMDGDIIVFQK (SEQ ID NO: 847)	2.32632143
TRIM28	SQPPVFKVFPGSTTEDYNLIVIERGAAAAATGQPGTAPAGTPGAPPLAGMAIVKEEETEAAIGAPPTATEGPETKPVLMA (SEQ ID NO: 848)	2.30375349
L3MBTL3	TVAGIPASKVSKWSTDEVSEFIQSLPGCEEHGKVFKDEQIDGEAFLLMTQTDIVKIMSIKLGPALKIFNSILMFKAAEKN (SEQ ID NO: 849)	2.28373726
IKZF4	PQEGLLRGTPGPSKEVLRVVGESGEPVKAFKCEHCRILFLDHVMFTIHMGCHGFRDPFECNICGYHSQDRYEFSSHIVRG(SEQ ID NO: 850)	2.28172595
KDM5B	CGSGNDEDRLLLCDGCDDSYHTFCLIPPLHDVPKGDWRCPKCLAQECSKPQEAFGFEQAARDYTLRTFGEMADAFKSDYF (SEQ ID NO: 851)	2.26045013

GATAD2A	QVVMPPLVRGAQQIHSIRQHSSTGPPPLLLAPRASVPSVQIQGQRIIQQGLIRVANVPNTSLLVNIPQPTPASLKGTTAT (SEQ ID NO: 852)	2.24549867
BRMS1	QERIQRLEEDRQSLDLSSEWWDDKLHARGSSRSWDSLPPSKRKKAPLVSGPYIVYMLQEIDILEDWTAIKKARAAVSPQK (SEQ ID NO: 853)	2.23400084
DNMT1	PLSKPRTPRRSKSDGEAKPEPSPSPRITRKSTRQTTITSHFAKGPAKRKPQEESERAKSDESIKEEDKDQDEKRRRVTSR (SEQ ID NO: 854)	2.18589467
ZNF366	MGHMHLHSDSKPFKCLYCPSKFTLKGNLTRHMKVKHGVMERGLHSQGLGRGRIALAQTAGVLRSLEQEEPFDLSQKRRAK (SEQ ID NO: 855)	2.17275454
NAB2	GELQLYRVLQRANLLSYYETFIQQGGDDVQQLCEAGEEEFLEIMALVGMATKPLHVRRLQKALREWATNPGLFSQPVPAV (SEQ ID NO: 856)	2.17138822
KDM1A	QLQEKHVKDEQIEHWKKIVKTQEELKELLNKMVNLKEKIKELHQQYKEASEVKPPRDITAEFLVKSKHRDLTALCKEYDE(SEQ ID NO: 857)	2.13679618
TGS1	LQSKKDTETENPPVENTLSPKLEITEKWEKYWNEYGGGLLWQSWQEKHPGQALSSEPWNFPDTKEEWEQHYSQLYWYYLE (SEQ ID NO: 858)	2.11925487
PRDM11	VSKEPGQLEDEEEEPSSFKADSPAEASLASDPHELPTTSFCPNCIRLKKKVRELQAELDMLKSGKLPEPPVLPPQVIELP (SEQ ID NO: 859)	2.07885469
HDAC6	PGEENLLGEAAGGQDMADSMLMQGSRGLTDQAIFYAVTPLPWCPHLVAVCPIPAAGLDVTQPCGDCGTIQENWVCLSCYQ (SEQ ID NO: 860)	2.03482535
CBX1	MGKKQNKKKVEEVLEEEEEEEYVVEKVLDRRVVKGKVEYLLKWKGFSDEDNTWEPEENLDCPDLIAEFLQSQKTAHETDKS (SEQ ID NO: 861)	2.01737869
TRIM8	VEEKVNQLKEEVRLQYEKLHQLLDEDLRQTVEVLDKAQAKFCSENAAQALHLGERMQEAKKLLGSLQLLFDKTEDVSFMK (SEQ ID NO: 862)	1.99485782
MDFI	QTHPSLASQGSKKSKSSSKSTTSQIPLQAQEDCCVHCILSCLFCEFLTLCNIVLDCATCGSCSSEDSCLCCCCCGSGECA (SEQ ID NO: 863)	1.98953136
NCOR1	HRGSTAGEVYRSHLPTHLDPAMPFHRALDPAAAAYLFQRQLSPTPGYPSQYQLYAMENTRQTILNDYITSQQMQVNLRPD (SEQ ID NO: 864)	1.98515002
HIVEP3	KKERKPQKPGKYICQYCSRPCAKPSVLQKHIRSHTGERPYPCGPCGFSFKTKSNLYKHRKSHAHRIKAGLASGMGGEMYP (SEQ ID NO: 865)	1.97820999
IKZF1	AASENSQDALRVVSTSGEQMKVYKCEHCRVLFLDHVMYTIHMGCHGFRDPFECNMCGYHSQDRYEFSSHITRGEHRFHMS (SEQ ID NO: 866)	1.97447058

CDK2AP1	TSSQYRQLLSDYGPPSLGYTQGTGNSQVPQSKYAELLAIIEELGKEIRPTYAGSKSAMERLKRGIIHARGLVRECLAETE (SEQ ID NO: 867)	1.96228244
ERF	SEGESEEVEVTDISDEDEEDGEVFKTPRAPPAPPKPEPGEAPGASQCMPLKLRFKRRWSEDCRLEGGGGPAGGFEDEGED (SEQ ID NO: 868)	1.94839175
KDM5C	IPKGVWRCPKCVMAECKRPPEAFGFEQATREYTLQSFGEMADSFKADYFNMPVHMVPTELVEKEFWRLVNSIEEDVTVEY (SEQ ID NO: 869)	1.94183352
ZNF446	LLGWITAHVLKQEVLPAAQKTEEPLGSPHPSGTVESPGEGPQDTRIEGSVQLSCSVKEEPNVDGQEVAPSSPPLAAQSPE (SEQ ID NO: 870)	1.92348688
CBFA2T3	DPRELRERHRPLVVPGSRQEEVIDHKLTEREWAEEWKHLNNLLNCIMDMVEKTRRSLTVLRRCQEADREELNHWARRYSD (SEQ ID NO: 871)	1.92105376
USP7	SDRREDYYDIQLSIKGKKNIFESFVDYVAVEQLDGDNKYDAGEHGLQEAEKGVKFLTLPPVLHLQLMRFMYDPQTDQNIK (SEQ ID NO: 872)	1.90807263
SETDB1	EIAELQQAVVEELGISMEELRHFIDEELEKMDCVQQRKKQLAELETWVIQKESEVAHVDQLFDDASRAVTNCESLVKDFY (SEQ ID NO: 873)	1.89374023
PHF1	FPTDIPKSAPHSMTASSSSVSSPSPGLPRRSAPPSPLCRSLSPGTGGGVRGGVGYLSRGDPVRVLARRVRPDGSVQYLVE (SEQ ID NO: 874)	1.88511381
IKZF3	LKPPPICPRDSVKVINKEGEVMDVYRCDHCRVLFLDYVMFTIHMGCHGFRDPFECNMCGYRSHDRYEFSSHIARGEHRAL (SEQ ID NO: 875)	1.87983916
HDAC4	EHQALLDEPYLDRLPGQKEAHAQAGVQVKQEPIESDEEEAEPPREVEPGQRQPSEQELLFRQQALLLEQQRIHQLRNYQA (SEQ ID NO: 876)	1.86554151
DNMT3L	AEPSMDVILVGSSELSSSVSPGTGRDLIAYEVKANQRNIEDICICCGSLQVHTQHPLFEGGICAPCKDKFLDALFLYDDD (SEQ ID NO: 877)	1.85705086
HDAC9	SSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQPFLEPTHTRALSVRQAPLAAVGMDGLEKHRLVSRTH (SEQ ID NO: 878)	1.85605728
ZSCAN22	VKVEEEEEEASLSQGGESSHDHIAHSEAARLRFRHFRYEEASGPHEALAHLRALCCQWLQPEAHSKEQILELLVLEQFLGA (SEQ ID NO: 879)	1.83277386
E2F6	QDIHSIQAFHEQIVIAVKAPAETRLDVPAPREDSITVHIRSTNGPIDVYLCEVEQGQTSNKRSEGVGTSSSESTHPEGPE (SEQ ID NO: 880)	1.82276829
KDM5A	AAQNCQRPCKDKVDWVQCDGGCDEWFHQVCVGVSPEMAENEDYICINCAKKQGPVSPGPAPPPSFIMSYKLPMEDIKETS (SEQ ID NO: 881)	1.78690918

RBBP7	KEGKIVDAKAIFTGHSAVVEDVAWHLLHESLFGSVADDQKLMIWDTRSNTTSKPSHLVDAHTAEVNCLSFNPYSEFILAT (SEQ ID NO: 882)	1.77403489
ATRX	ERERERKLREVIEIEDASPTKCPITTKLVLDEDEETKEPLVQVHRNMVIKLKPHQVDGVQFMWDCCCESVKKTKKSPGSGC (SEQ ID NO: 883)	1.77379044
ZNF446	EPETARLRFRGFCYQEVAGPREALARLRELCCQWLQPEAHSKEQMLEMLVLEQFLGTLPPEIQAWVRGQRPGSPEEAAAL (SEQ ID NO: 884)	1.7514371
GATAD2B	PVVQNAASIVQPSPAHVGQQGLSKLPSRPGAQGVEPQNLRTLQGHSVIRSATNTTLPHMLMSQRVIAPNPAQLQGQRGPP (SEQ ID NO: 885)	1.71132166
TET2	TESCHSQMHRPIKVEPGCKPHACMHTAPPENKTWKKVTKQENPPASCDNVQQKSIIETMEQHLKQFHAKSLFDHKALTLK (SEQ ID NO: 886)	1.67625849
NCOR2	VVPKEEKEEETAAAPPVEEGEEQKPPAAEELAVDTGKAEEPVKSECTEEAEEGPAKGKDAEAAEATAEGALKAEKKEGGS (SEQ ID NO: 887)	1.65967202
BCOR	PPMKSLSSTSAGGKKQAQPSCAPASRPPAKQQKIKENQKTDVLCADEEEDCQAASLLQKYTDNSEKPSGKRLCKTKHLIP (SEQ ID NO: 889)	1.65199942
RBL2	TSRDSSPVMRSSSTLPVPQPSSAPPTPTRLTGANSDMEEEERGDLIQFYNNIYIKQIKTFAMKYSQANMDAPPLSPYPFV (SEQ ID NO: 890)	1.65141032
KDM5D	IPRGIWRCPKCILAECKQPPEAFGFEQATQEYSLQSFGEMADSFKSDYFNMPVHMVPTELVEKEFWRLVSSIEEDVTVEY (SEQ ID NO: 891)	1.59348979
TRIM8	STQHLVALPGGAQPVHSSPVFPPSQYPNGSAAQQPMLPQYGGRKILVCSVDNCYCSSVANHGGHQPYPRSGHFPWTVPSQ (SEQ ID NO: 892)	1.56673764
KLF3	ILPVIQPVVVQPVPFMYTSHLQQPLMVSLSEEMENSSSSMQVPVIESYEKPISQKKIKIEPGIEPQRTDYYPEEMSPPLM (SEQ ID NO: 893)	1.56352551
MTA2	YGGLKTPTQLEGATRGTTEPHSRGHLSRPEAQSLSPYTTSANRAKLLAKNRQTFLLQTTKLTRLARRMCRDLLQPRRAAR (SEQ ID NO: 894)	1.53904229
PHF1	MAQPPRLSRSGASSLWDPASPAPTSGPRPRLWEGQDVLARWTDGLLYLGTIKKVDSAREVCLVQFEDDSQFLVLWKDISP (SEQ ID NO: 895)	1.53480472
RCOR1	AKLARRSQERDNLGMLVWSPNQNLSEAKLDEYIAIAKEKHGYNMEQALGMLFWHKHNIEKSLADLPNFTPFPDEWTVEDK (SEQ ID NO: 896)	1.39637997

参考文献

Al Chiblak,M.,Steinbeck,F.,Thiesen,H.-J.,and Lorenz,P.(2019).DUF3669, a “domain or unknown funetion”within ZNF746 and ZN F777,oligomerizes and

contributes to transcriptional repression.BMC Mol Cell Biol 20,60.

Amabile,A.,Migliara,A.,Capasso,P.,Biffi,M.,Cittaro,D.,Naldini,L.,and Lombardo,A.(2016).Inheritable Sileneing of Endogerous Genes by Hit-and-Run Targeted Epigenetic Editing.Cell 167,219-232.e14.

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 trarscriptior factor sequerces.EMBO J.e98896.

Ashkenazy,H.,Erez,E.,Martz,E.,Pupko,T.,and Ben-Tal,N.(2010).ConSurf2010:calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.Nucleic Acids Res.38,W529-W533.

Bakan,A.,Meireles,L.M.,and Bahar,I.(2011).ProDy:protein dynamics inferredfrom theory and experiments.Bioinformatics 27,1575-1577.

Ballas,N.,Battaglioli,E.,Atouf,F.,Andres,M.E.,Chenoweth,J.,Anderrson,M.E.,Burger,C.,Moniwa,M.,Davie,J.R.,Bowers,W.J.,et al.(2001).Regulation of neuronal traits by a novel transcriptional complex.Neuron 31,353-365.

Berezin,C.,Glaser,F.,Rosenberg,J.,Paz,I.,Pupko,T.,Fariselli,P.,Casadio,R.,and Ben-Tal,N.(2004).ConSeq:the identification or functionally and structurally important residues in protein sequences.Bioinformatics 20,1322-1324.

Bersaglieri,C.,and Santoro,R(20019).Genome Organization in and around the Nucleolus.Cells 8.

Bintu,L.,Yong,J.,Antebi,Y.E.,McCue,K.,Kazuki,Y.,Uno,N.,Oshimura,M.,and Elowitz,M.B.(2016).Dynamics of epigenetic regulation at the single-cell level.Science 351,720-724.

Birbach,A.,Bailey,S.T.,Ghosh,S.,and Schmid,J.A.(2004).Cytosolic,nuclear and nucleolar localization signals determine subcellular distribution and activity of the NF-kappaB inducing kinase NIK.J.Cell Sci.117,3615-3624.

Birtle,Z.,and Ponting,C.P.(2006).Meisetz and the birth of the KRAB motir.Bioinformatics 22,2841-2845.

Blackledge,N.P.,Farcas,A.M.,Kondo,T.,King,H.W.,McGouran,J.F.,Hanssen,L.L.P.,Ito,S.,Cooper,S.,Kondo,K.,Koseki,Y.,et al.(2014).Variant PRC complex-dedendent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation.Cell 157,1445-1459.

Chang,Y.,Sun,L.,Kokura,K.,Horton,J.R.,Fukuda,M.,Espejo,A.,Izumi,V.,Koomen,J.M.,Bedford,M.T.,Zhang,X.,et al.(2011).MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a.Nat.Commun.2,533.

Cheng,C.-T.,Kuo,C.-Y.,and Ann,D.K.(2014).KAPtain in charge of

multiple missions:Emerging roles of KAP1.World J.Biol.Chem.5,308-320.

Chevron,M.P.,Girard,F,Claustres,M.,and Debaille,J.(1994).Expression and subcellular localizattion of dystrophin in skeletal,cardiac and smooth museles during the human development.Neuromuscul.Disord.4,419-432.

Childs,K.S.,and Goodbourn,S.(2003).Identification of novelco-repressot molecules for Interferon Regulatory Factor-2.Nucleic Acids Res.31,3016-3026.

Chittock,E.C.,Latwiel,S.,Miller,T.C.R.,and Müller,C.W.(2017).Molecular architecture of polycomb repressive complexes.Biochem.Soc.Trans.45,193-205.

Chupreta,S.,Holmstrom,S.,Subramanian,L.,and Iñiguez-Lluhí,J.A.(2005).A small conserved surface in SUMO is the critical structural determinant of its transcriptionalinlibitory properties.Mol.Cell.Biol.25,4272-4282.

Copley,R.R.(2005).The EH1 motifin metazoan transcription factors.BMC Genomics 6,169.

Corsetti,M.T.,Levi,G.,Lancia,F.,Sanseverino,L.,Ferini,S.,Boncinelli,E.,and Cotte,G.(1995).Nucleolar localisation of three Hox homeoproteins.J.Cell Sci.108(Pt 1),187-193.

Dorris,D.R.,and Struhl,K.(2000).Artificial recruitment of TFIIDbut not RNA polymerase II holoenzyme,activatestranscription in mammalian cells.Mol.Cell.Biol.20,4350-4358.

Drueppel,L.,Pfleiderer,K.,Schmidt,A.,Hillen,W.,and Berens,C.(2004)A short autonomous repression motif is located within the N-terminal domain of CTCF.FEB S Lett.572,154-158.

Duboule,D.,and Morata,G.(1994).Colinearity and functional hierarchy among genes of the homeotic conplexes.Trends in Genetics 10,358-364.

El-Gebali,S.,Mistry,J.,Bateman,A.,Eddy,S.R.,Luciani,A.,Potter,S.C.,Qureshi,M.,Richardson,L.J.,Salazar,G.A.,Smart,A.,et al.(2019).The Pfam protein fammilies database in 2019.Nucleic Acids Res.47,D427-D432.

Elliott,K.,Sakamuro,D.,Basu,A.,Du,W.,Wunner,W.,Staller,P.,Gaubatz,S.,Zhang,H.,Prochownik,E.,Eilers,M.,et al.(1999).Bin 1 functionally interacts with Myc and inhibits cell proliferation via multiple mechanisms.Oncogene 18,3564-3573.

ENCODE Project Consortium,Moore,J.E.,Purcaro,M.J,Pratt,H.E.,Epstein,C.B.,Shoresh,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.

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 Transciption Activation Domains Reveals Their Sequence Features and permits Prediction by Deep Learning.Mol.Cell 78,890-902.e6.

Feng,J.,Liu,T.,Qin,B.,Zhang,Y.,and Liu,X.S.(2012).IdentifyingChIP-seq enrichment using MACS.Nat.Protoc.7.1728-1740.

Fowler,D.M.,and Fields,S.(2014)Deep mnutational scanning:a new style of protein science.Nat.Methods 11,801-807.

Frey,F.,Sheahan,T.,Finkl,K.,Stoehr,G.,Mann,M.,Benda,C.,and Müller,J.(2016).Molecular basis of PRC1targeting to polycomb response elements by PhoRC.Genes Dev.30,1116-1127.

García,E.,Marcos-Gutiérrez,C.,del Mar Lorente,M.,Moreno,J.C.,annd Vidal,M.(1999)RYBP,a new repressor protein that interacts with conponents of the mammalian Polycomb complex,and with the tanscription factor YY1.EMBO J.18,3404-3418.

Gaudreau,L.,Keaveney,M.,Nevado,J.,Zaman,Z.,Bryant,G.O.,Struhl,K.,and Ptashne,M.(1999).Transcriptional activation by artificial recruitment in yeast is influenced by promoter architecture ad downstream sequences.Proc.Natl.Acad.Sci.U.S.A.96,2668-2673.

Gilbert,S.F.(1971).Developmental Biology(Sinauer Associates).

Gilbert,L.A,Horlbeck,M.A.,Adamson,B.,Villalta,J.E.,Chen,Y.,Whitehead,E.H.,Guimaraes,C.,Panning,B.,Ploegh,H.L.,Bassik,M.C.,et al.(2014),Genome-Scale CRISPR-Mediated Control of Gene Repression and Actiyati on.Cell 159,647-661.

Guetg,C.,Lienemann,P.,Sirri,V.,Grummt,L,Hernandez-Verdun,D.,Hottiger,M.O.,Fussenegger,M.,and Santoro,R.(2010).The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats.EMBO J.29,2135-2146.

Haney,M.S.,Bohlen,C.J.,Morgens,D.W.,Ousey,J.A.,Barkal,A.A.,Tsui,C.K.,Ego,B.K.,Levin,R.,Kamber,R.A.,Collins,H.,et al.(2018)Identification of phagocytosis regulators using magnetic genome-wide CRISPR screens.Nat.Genet.50,1716-1727.

Helleboid,P.-Y.,Heusel,M.,Duc,J.,Piot,C.,Thorball.CW.,Coluccio,A.,Pontis,J.,Imbeault,M.,Turelli,P.,Aebersold,R.,et al.(2019).The interactome of KRAB zinc finger proteins reveals the evolutimary history of their functional diversification.EMBO J.38,e101220.

Heredia,J.D.,Park,J.,Brubaker,R.J.,Szymanski,S.K.,Gill,K.S.,and Procko,E.(2018).Mapping Interaction Sites on uman Chemokine Receptors by Deep Mutational Scanning.J.Immunol.200,3825-3839.

Holland,P.W.H.,Booth,H A.F.,and Bruford,E,A.(2007)Classification and nomenclature of all human homeobox genes.BMC Bi ol.5,47.

Hueber,S.D.,Weiller,G.F., Djordjevic,M.A.,and Frickey.T.(2010) Improving Hox protein classification across the major model organisms.PLoS One 5,e10820.

Hyun,K., Jeo,J.,Park,K.,and Kim,J.(2017),Writing,erasing and reading histone lysine methylations.Exp.Mol.Med.49,e324.

Imbeault,M.,Helleboid,P.-Y.,and Trono,D.(2017).KRAB zinc-finger proteins contribute to the evolutio ofgene regulatory networks.Nature 543, 550-554.

Johnson,M.,Zaretskaya,I.,RaytselisY.,Merezhuk,Y.,McGinnis,S.,and Madden,T.L.(2008).NCBI BLAST:a beiter web interface.Nucleic Acids Res.36,W5-W9。

Jolma,A.,Yan,J.,Whitington,T.,Toivonen,J.,Nitta,K.R.,Rastas,P., Morgurnova,E.,Enge,M.,Taipale,M.,Wei,G.,et al.(2013)。DNA-binding specificities of human tanscription factors.Cell 152,327-339.

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

Kim, G-D.,Ni,J.,Kelesoglu,N.,Roberts,R.J.,and Pradhan,S.(2002)Co-operation and cogmmunication between the human maintenance and de ovo DNA (cytosine-5)methyltrarnsferases.EMBO J.21,4183-4195.

Kinsclla,R.J.,Kähäri,A.Haidcr,S.,Zamora,J.,Proctor,G.,Spudich,G., Almcida-King,J.,Staines,D.,Derwent,P.,Kerhornou,A.,et al.(2011).Ensembl BioMarts:a hub for data retrieval across taxonomie space.Database 2011, bar030.

Konermann,S.,Brigham,M.D.,Trevino,A.E.,Joung,J.,Abudayyeh,O.O., Barcena,C.,Hsu,P.D.,Habib,N.,Gootenberg,J.S.,Nishimasu,H.,et al.(2014) .Genome-scale tanscriptional activation by an engneered CRISPR-Cas9 complex.Nature 517,583-588.

Kotler,E.,Shani,O.,Goldfeld,G.,Lotan-Pompan,M.,Tarcie,O.,Gershoni,A., Hopf,T.A.,Marks,D.S.,Oren,M.,ad Segal,E.(2018).A Systematie p53 Mutatio Library Links Differential Functional Impact to Cancer Mutation Pattern and Evolutionary Conservation.Mol.Cell 71,873.

Iambert,S.A.,Jolma,A.,Carnpitelli,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 175,598-599.

Langmead,B.,Trapnell,C.,Pop,M.,and Salzberg,S.L.(2009).Ultrafast and memory-efficient alignment of short DNA sequences to te human genome.Genome Biol.10,R25.

Lechner,M.S,Begg,G.E.,Speicher,D.W.,and Rauscher,F.J.(2000).Molecular

Determinants for Targeting Heterochromatin Protein 1-Mediated Gene Silencing: Direct Chromoshadow Doman-KAP-1 Corepressor Interaction Is Essential.Mol.Cell.Biol.20,6449-6465.

Li,Q.,Wang,X.,Lu, Z.,Zhang,B.,Guan,Z.,Iiu,Z.,Zhorg,Q.,Gu,L.,Zhou,J.,Zhu,B.,et al.(2010).Polycomb CBX7 directly controls trimethylation of histone H3 at lysine 9at the p16 locus.PLoS One 5,e13732.

Liu,J.,etrumal,N.B.,Oldfield,C.J.,Su,E.W.,Uversky,V.N.,and Dunker,A.K.(2006).Intrinsjc disorder in transcription factors.Biochemistry 4.5,6873-6888.

Losson,R.,and Nielsen,A.L.(2010).The NIZP1 KRAB and C2HR domains cross-talk for transcriptional regulation.Biochim.Biophys.Acta 1799,463-468.

Lynch,V.J.Roth,J.J.,and Wagner,G.P.(2006).Adaptive evolutio of Hox-gene homeodomains after cluster duplications.BMC Evol.Biol.6,86.

Mallo,M.,and Alonso,C.R.(2013).The regulation of Hox gene expression durng animal development.Development 140,3951-3963.

Mann,R.S.,and Hogness,D.S.(1990).Functional dissection of Ultrabithorax proteis in D.melanogaster.Cell 60,597-610.

Margolin,J.F.,Friedman,J.R.,Meyer,W.K.,Vissing,H.,Thjesen,H.J.,and Rauscher,F.J.(1994).Kruppel-associated boxes are potent transcriptional ℃ pression domains.Proc.Natl.Acad.Sci,U.S.A.91,4509-4513.

Martin,R.M.,Ter-Avetisyan,G,Herce,H.D,Ludwig,A.K.,Lättig-Tünnemann,G.,and Cardoso,M.C.(2015).Principles of protein targeting to te nucleolus.Nucleus 6,314-325.

McWillia,H.,Li,W.,Uludag,M.,Squizzato,S.,Park,Y.M.,Buso,NCowley,A.P,and Lopez,R.(2013)Analysis Tool Web Services from the EMBL-EBI.Nucleic Acids Res.41,W597-W600.

Mifsud,W.,and Bateman,A.(2002).Membrane-bound progesterone receptos contain a cytochrome b5-like ligand-binding domain.Genome Biol.3,RESEARCH0068.

Mitchell,P.J,andTjian,R.(1989)Transcriptional regulation in mammalia cells by sequence-specific DNA binding proteins.Science 245,371-378.

Mitrea,D.M.,Cika,J.A.,Guy,C.S.,Ban,D.,Banerjee,P.R,Stanley,C.B.,Nourse,A.,Deniz,A.A.,and Kriwacki,R.W.(2016).Nueleophosmin integrates within the nuclelus via m lti-modal interactions with proteins displaying R-rich linear motifs and rRNA.Elie 5.

Moussa,H.F.,Bsteh,D.,Yelagandula,R.,Pribitzer,C.,Stecher,K.,Bartalska,K.,Michetti,L.,Wang,J.,Zepeda-Martinez,J.A.,Elling,U.,et al,(2019) Canonical PRC 1 controls sequence-independent propagation of Polycomb-mediated gene silencing.Nat.Commun.10,1931.

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.

Nevado,J.,Gaudreau,L.,Adam,M.,and Ptashne,M.(1999).Transcriptional activation by artificial recruitment in mammalian cells.Proceedings of the National Academy of Sciences 96,2674-2677.

Nott,A.,Holtman,L.R.,Coufal,N.G..,Schlachetzki,J.C.M.,Yu,M.,Hu,R.,Han,C.Z.,Pena,M.,Xiao,J.,Wu,Y.,et al.(2019).Brain cell type-specific enhancer-promoter interactome maps and disease-isk association.Science 366,1134-1139,

Partridge,E.C.,Christopher Partridge,E.,Chhetri,S.B.,Prokop,J.W.,Ramaker,R.C.,Jansen,C.S.,Goh,S.-T.,Mackiewicz,M.,Newberry,K.M.,Brandsmeier,L.A.,et al.(2020).Occupancy maps of 208chromatin-associated proteins in one human cell type.Nature 583,720-728.

Peng,H.,Gibson,L.C.,Capili,A.D.,Borden,K.L.B.,Osborne,M.J.,Harper,S.L.,Speicher,D.W.,Zhao,K.,Marmorstein,R.,Rock,T.A.,et al.(2007).The Structurally Disordered KRAB Repression Domain Is Incorporated into a Protease Resistant Core upon Binding to KAP-1-RBCC Domain.Journal of Molecular Biology 370,269-289.

Peng,H.,Ivanov,A.V.,Oh,H.J.,Lau,Y.-F.C.,and Rauscher,F.J,3rd(2009).Epigenetic gene silencing by the SRY protein is mediated by a KRAB-O protein that recruits the KAP1corepressor machiery.J.Biol.Chem.284,35670-35680.

Pupko,T.,Bell,R.E.,Mayrose,I.,Glaser,F.,and Ben-Tal,N.(2002).Rate4Site:an algorithmic tool for the idetification of functional regions in proteits by surface mapping of evolutionary determinants within their homologues.Bioinformatics18Suppl1,S71-S77.

Ramalingam,A.,and Prendergast,G.C.(2007).Binl homolog hob1 supports a Rad6-Setl pathway of transcriptional repression in fission yeast.Cell Cycle 6,1655-1662.

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

Roney,I.J.,Rudner,A.D.,Couture,J.-F,and Kærn,M.(2016).Improvement of the reverse tetracycline transactivator by single amino acid substitutions that reduce leaky target gene expressioi to undetectable levels.Sci.Rep.6,27697.

Rueden,C.T.,Schindelin,J.,Hiner,M.C.,DeZonia,B.E.,Walter,A.E.,Arena,E.T.,and Eliceiri,K.W.(2017).ImageJ2:ImageJ for the next geeration of

scientifie image data.BMC Bioinformatics 18,529.

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

Schmitges,F.W.,Radovani,E.,Najafabadi,H.S.,Barazandeh,M.,Campitelli,L.F.,Yin,Y.,Jolma,A.,Zhong,G.,Guo,H.,Kanagalingam,T.,et al.(2016).Multiparameter fuetional diversity of human C2H2 zinc fiaer proteins.Genome Res.26,1742-1752.

Schnabel,C.A.,and Abate-Shen,C.(1996).Repression by HoxA7 is mediated by the homeodomain and the modulatory action of its N-terminal-arm residues.Mol.Cell Biol.16,2678-2688.

Sidore,A.M.,Plesa,C.,Samson,J.A.,Lubock,N.B.,and Kosuri,S,(2020).DropSynth 2.0:high-fidelity multiplexed gene synthesis in emulsions.Nucleic Acids Res.

Sievers,F.,Wilm,A.,Dineen,D.,Gibson,T.J.,Karplus,K.,Li,W.,Lopez,R.,McWilliam,H.,Remmert,M.,Soding,J.,et al.(2011)Fast,scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.Mol.Syst.Biol.7,539.

Sievers,Q.L.,Petzold,G.,Bunker,R.D.,Renneville,A.,Sabicki,M.,Liddicoat,B.J.,Abdulrahman,W.,Mikkelsen,T.,Ebert,B.L.,and Thomä,N.H.(2018).Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN.Science 362.

Sins,R.J.,3rd,Nishioka,K.,and Reinberg,D.(2003).Histone lysine methylation:a signature for chromatin function.Trends Genet.19,629-639.

SirriV.,Urcuqui-Incbima,S.,Roussel,P.,and Hernandez-Verdun,D,(2008).Nueleolus:the fascinating nuclear body.Histochem.Cell Biol.129,13-31.

Staller,M.V.,Holehouse,A.S.,Swain-Lenz,D.,Das,R.K.,Pappu,R.V.,and Cohen,B.A.(2018).A High-Throughput Mutational Scan of an Intrinsically isordered Acidic Transcriptional Activation Domain.Cell Syst 6,444-455.e6.

Stielow,B.,Finkernagel,F.,Stiewe,T.,Nist,A.,and Suske,G.(2018).MGA,L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polyomb repressive complex PRC1.6.PLoS Genet.14,e1007193.

Takahashi,H,Kasahara,K.,and Kokubo,T.(2009).Saccharomyces cerevisiae Med9 comprises two functionally distinct domains that play different roles in transcriptional regulation.Genes Cells 14,53-67.

UniProt Consortium(2015).UniProt:a hub for protein information.Nucleic Acids Res.43,D204-D212.

Wang,R.,Taylor,A.B.,and Kim,C.A.(2010).Ringl B C-terminal domain/RYBPC-terminal domain Complex.

Waterhorse,A,M.,Procter,J.B.,Martin,D.M.A.,Clamp,M.,and Barton,G.J.

(2009).Jalview Version 2-a multiple sequence alignment editor and analysis workbench.Bioinformatics 25,1189-1191.

Wellik,D.M.,and Capecchi,M.R.(2003).Hox10and Hox11 genes are required to globally pattern the mammalian skeleton.Science 301,363-367.

Witzgall,R.,O'Leary,E.,Leaf,A.,Onaldi,D.,ard Bonventre,J.V.(1994).The Krüppel-associated box-A(KRAB-A)domain of zinc finger proteins mediares transcriptional rpression.Proc.Natl.Acad.Sci.U.S.A.91,4514-4518.

Zhao,Y.,and Potter,S.S.(2001).Functional specificity of the oxa13 homeobox.Development 128,3197-3207.

Zhao,J.,Wang,M.,Chang,L.,Yu,J.,Song,A.,Liu,C.,Huang,W.,Zhang,T.,Wu,X.,Shen,X.,et al.(2020).RYBP/YAF2-PRC1 complexes and histone H1-dependent chromatin compaction mediate propagation of H2AK119ubl dring cell divisin.Nat.Cell Biol.22,439-452.

Zulkower,V.,and Rosser,S.(2020).DNA Chisel,a versatile sequence optimizer.Bioinformatics.

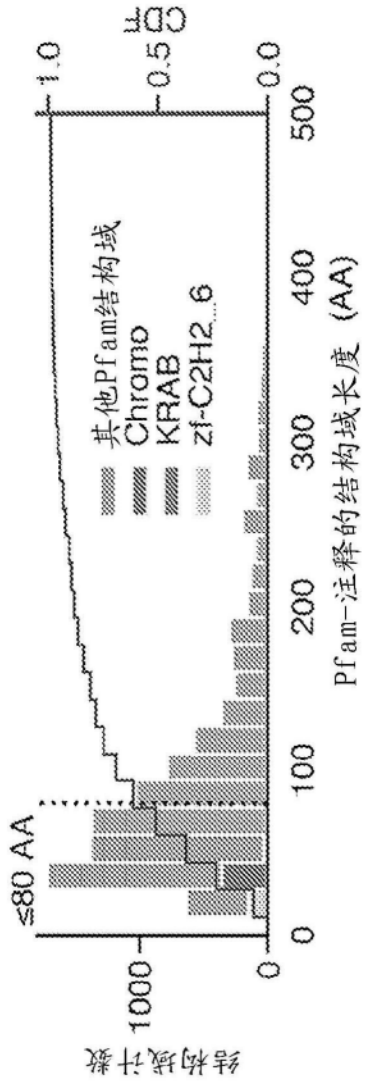

结构域计数
1000
0
≤80 AA
0
100
200
300
400
500
Pfam-注释的结构域长度 (AA)
1.0
0.5
0.0
CDF
其他Pfam结构域
Chromo
KRAB
zf-C2H2_6

图1A

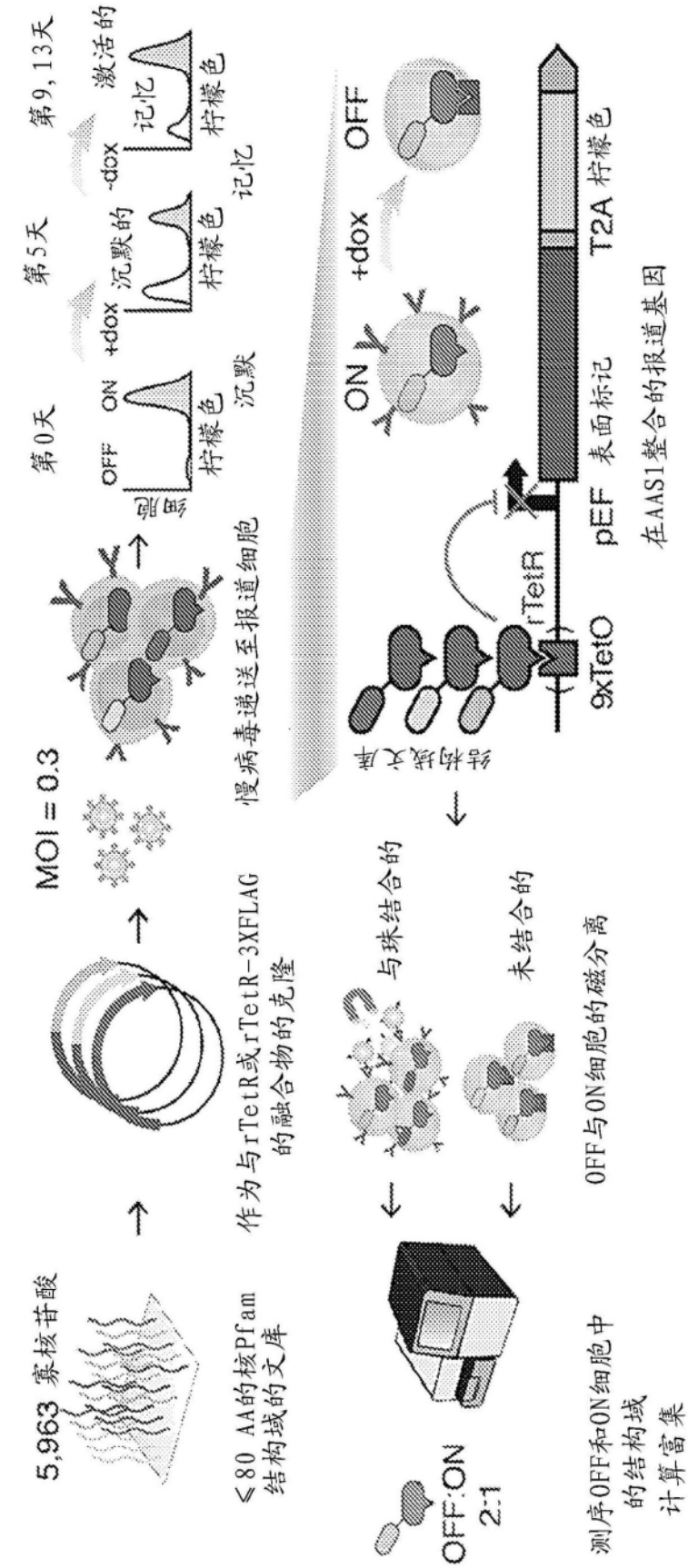

5,963 寡核苷酸
≤80 AA的核Pfam
结构域的文库
作为与rTetR或rTetR-3XFLAG
的融合物的克隆
MOI = 0.3
慢病毒递送至报道细胞
第0天
OFF
ON
+dox
沉默的
柠檬色
沉默
第5天
-dox
激活的
记忆
柠檬色
记忆
第9,13天
OFF:ON
2:1
测序OFF和ON细胞中
的结构域
计算富集
与珠结合的
未结合的
OFF与ON细胞的磁分离
结构域文库
ON
+dox
OFF
9xTetO
rTetR
pEF
表面标记
T2A
柠檬色
在AAS1整合的报道基因

图1B

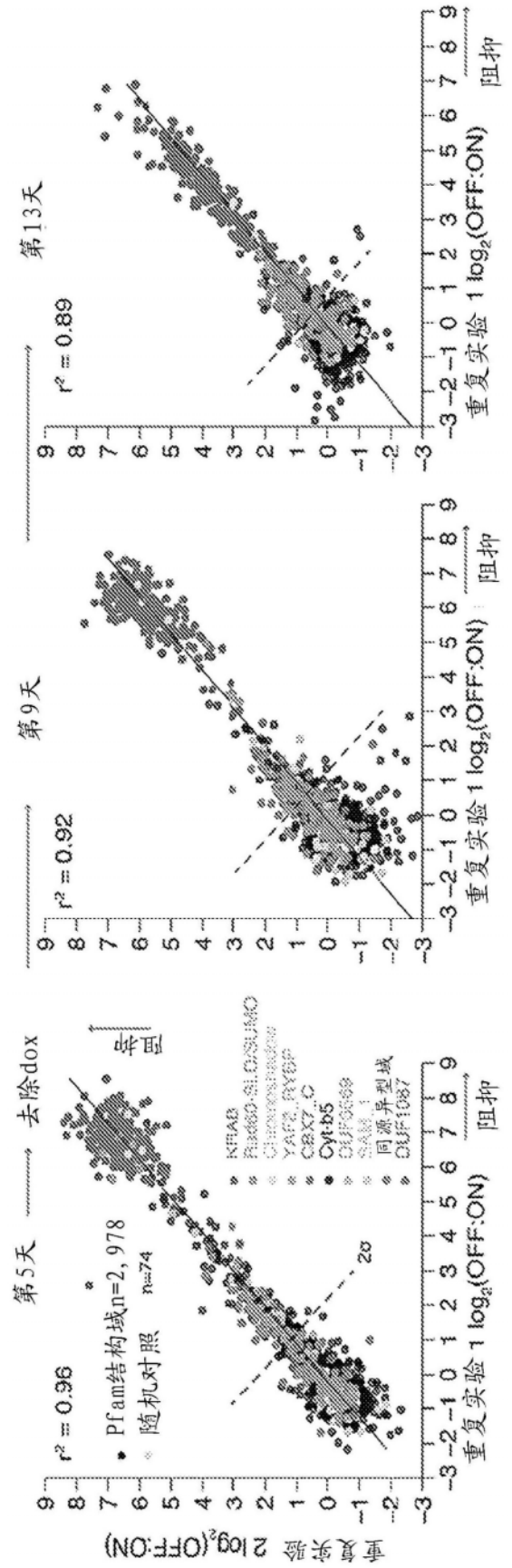

第5天
去除dox
第9天
第13天
r^2 = 0.96
r^2 = 0.92
r^2 = 0.89
Pfam结构域n=2,978
随机对照 n=74
2σ
KRAB
Rad60-SLD/SUMO
Chromoshadow
YAF2_RYBP
CBX7_C
Cyt-b5
DUF3669
SAM_1
同源异型域
DUF1087
重复实验 2 $\log_2$(OFF:ON)
重复实验 1 $\log_2$(OFF:ON)
阻抑
富集

图1C

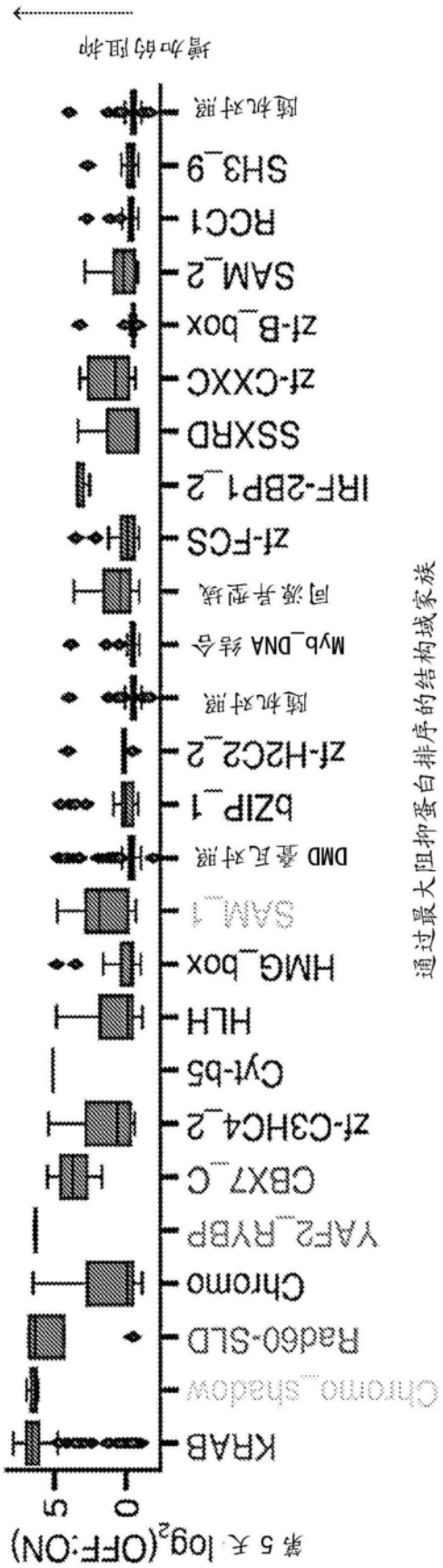

增加的阻抑
随机对照
SH3_9
RCC1
SAM_2
zf-B_box
zf-CXXC
SSXRD
IRF-2BP1_2
zf-FCS
同源异型域
Myb_DNA 结合
随机对照
zf-H2C2_2
bZIP_1
DMD 叠瓦对照
SAM_1
HMG_box
HLH
Cyt-b5
zf-C3HC4_2
CBX7_C
YAF2_RYBP
Chromo
Rad60-SLD
Chromo_shadow
KRAB
0
5
第 5 天 log2(OFF:ON)
通过最大阻抑蛋白排序的结构域家族

图1D

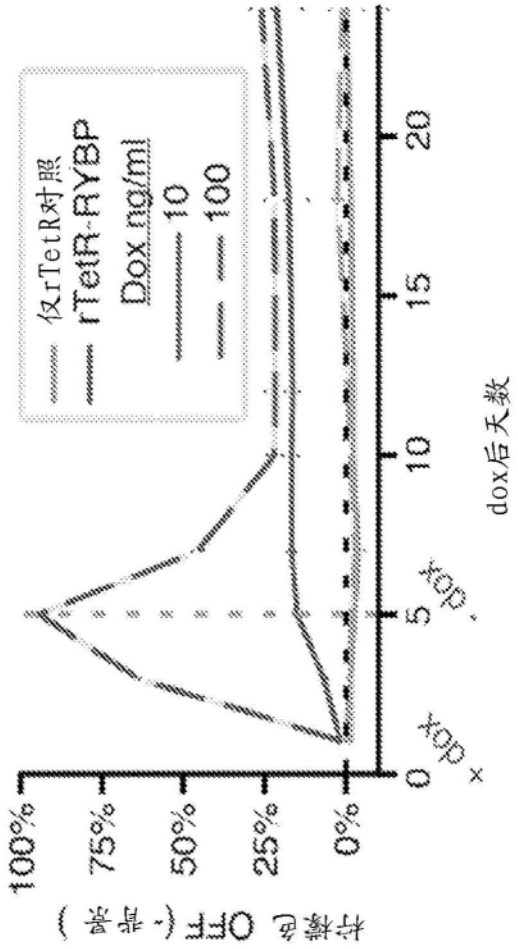

仅rTetR对照
rTetR-RYBP
Dox ng/ml
10
100
0
5
10
15
20
dox后天数
100%
75%
50%
25%
0%
荧光色 OFF (-背景)

图1E

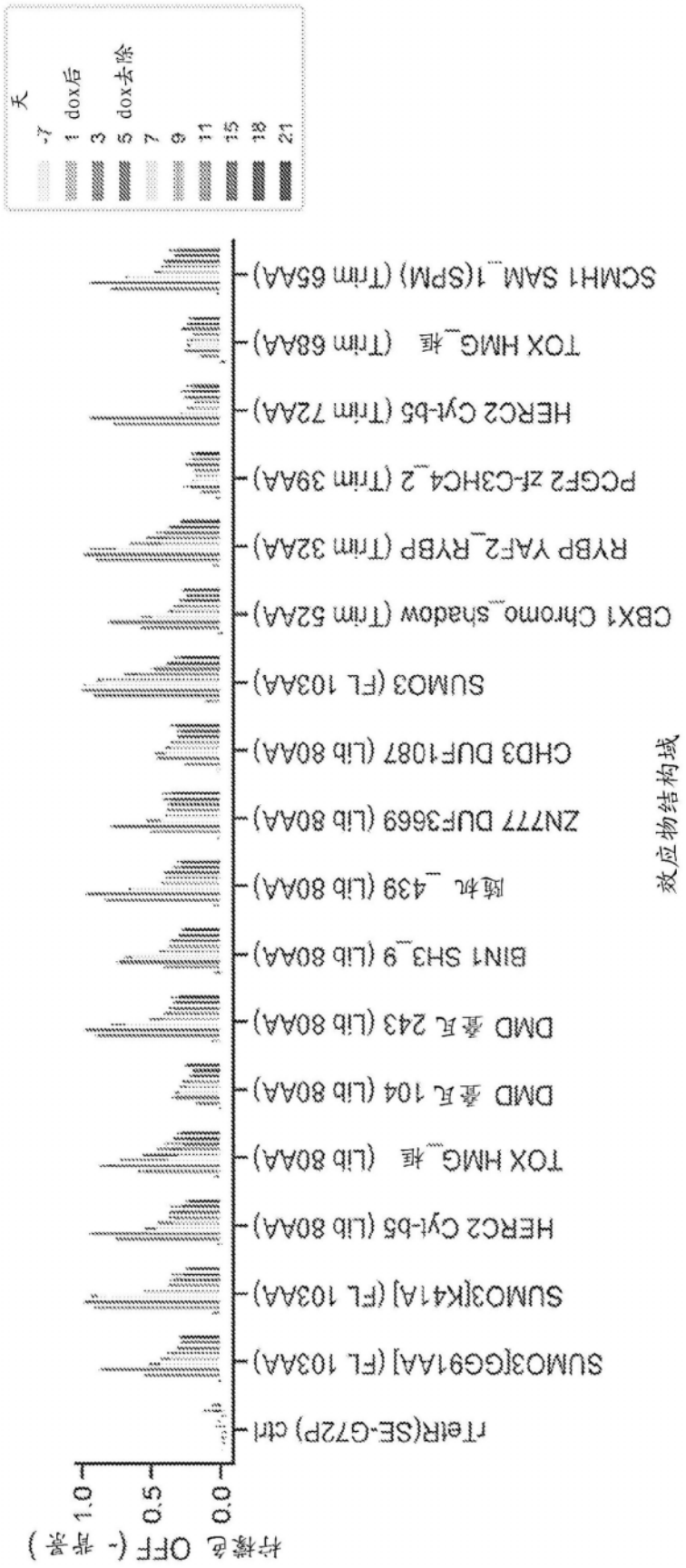

天
-7
1 dox后
3
5 dox去除
7
9
11
15
18
21
柠檬色 OFF (- 背景)
0.0
0.5
1.0
rTetR(SE-G72P) ctrl
SUMO3[G91AA] (FL 103AA)
SUMO3[K41A] (FL 103AA)
HERC2 Cyt-b5 (Lib 80AA)
TOX HMG_框 (Lib 80AA)
DMD 叠瓦 104 (Lib 80AA)
DMD 叠瓦 243 (Lib 80AA)
BIN1 SH3_9 (Lib 80AA)
随机 _439 (Lib 80AA)
ZN777 DUF3669 (Lib 80AA)
CHD3 DUF1087 (Lib 80AA)
SUMO3 (FL 103AA)
CBX1 Chromo_shadow (Trim 52AA)
RYBP YAF2_RYBP (Trim 32AA)
PCGF2 zf-C3HC4_2 (Trim 39AA)
HERC2 Cyt-b5 (Trim 72AA)
TOX HMG_框 (Trim 68AA)
SCMH1 SAM_1(SPM) (Trim 65AA)
效应物结构域

图1F

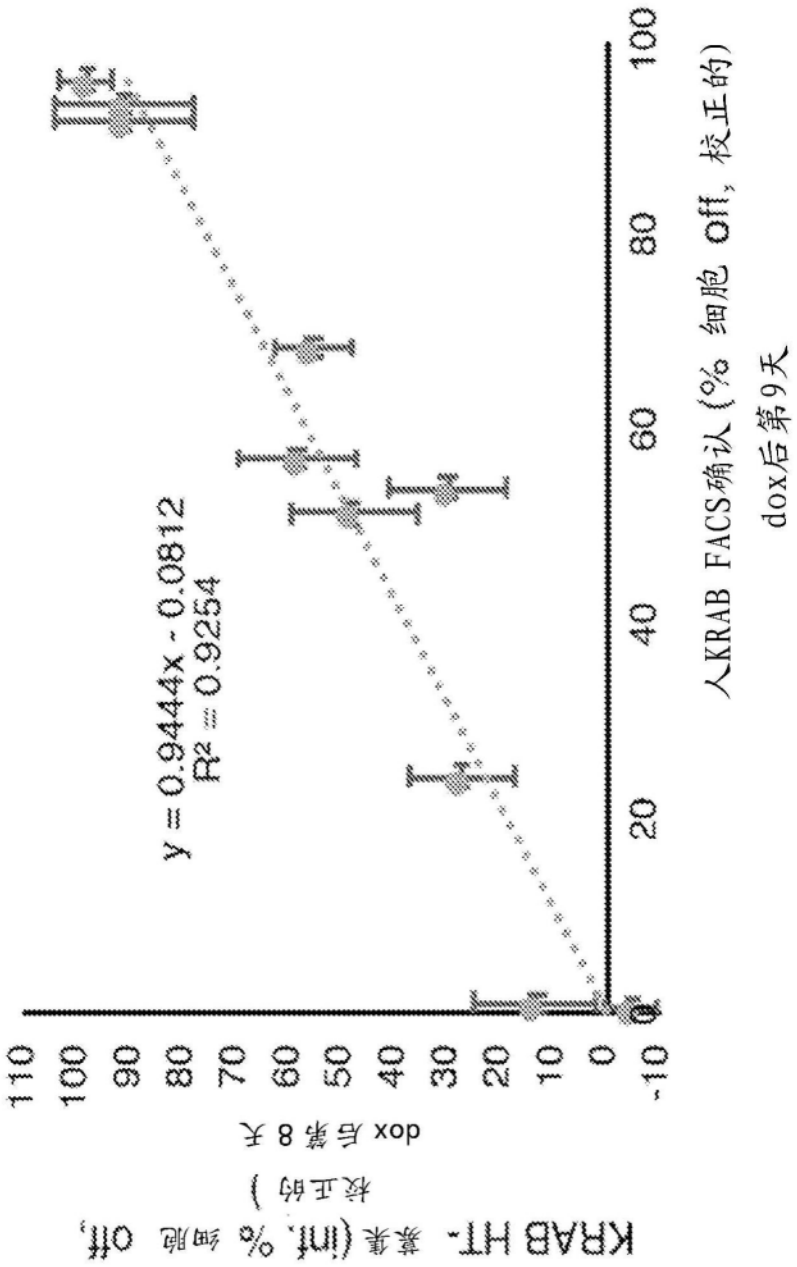

y = 0.9444x - 0.0812
R² = 0.9254
人KRAB FACS确认(% 细胞 off，校正的)
dox后第9天
KRAB HT- 募集(int. % 细胞 off，校正的)
dox 后第 8 天

图1G

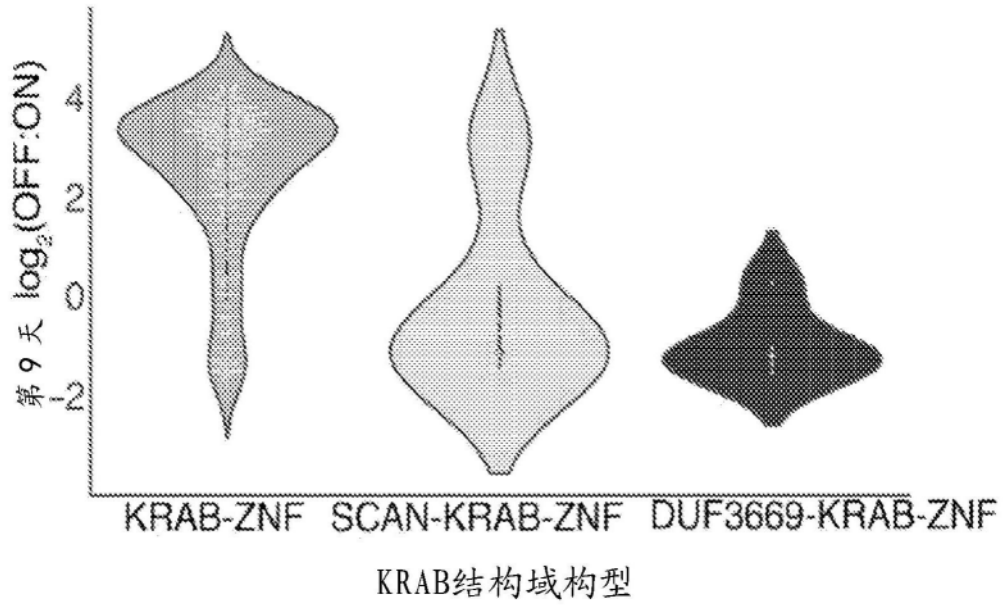

第9天 log2(OFF:ON)
4
2
0
-2
KRAB-ZNF
SCAN-KRAB-ZNF
DUF3669-KRAB-ZNF
KRAB结构域构型

图2A

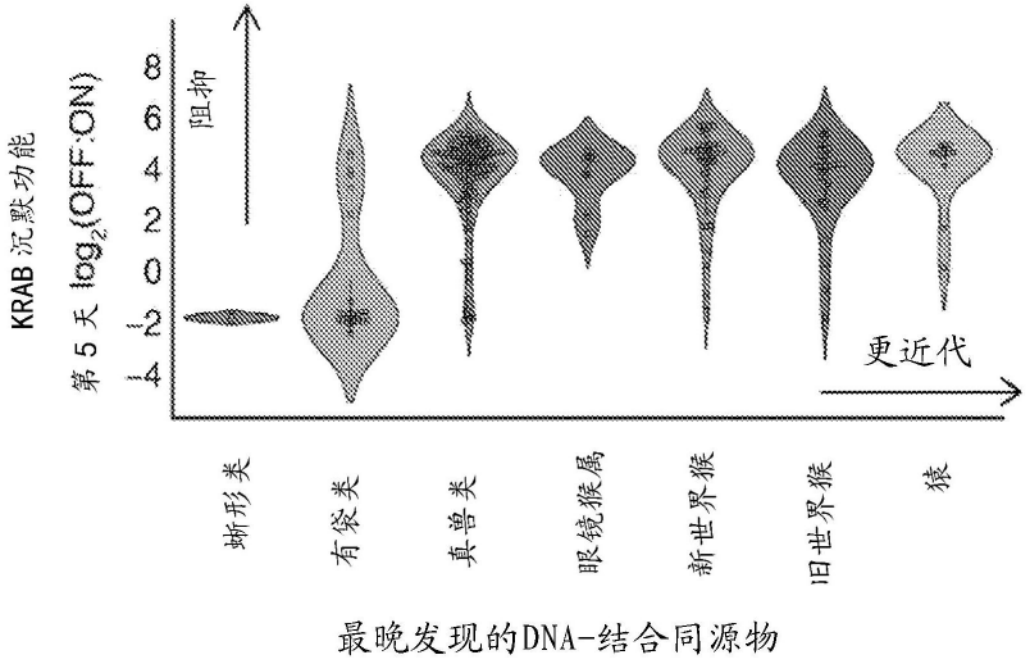

KRAB 沉默功能
第5天 log2(OFF:ON)
8
6
4
2
0
-2
-4
阻抑
更近代
蜥形类
有袋类
真兽类
眼镜猴属
新世界猴
旧世界猴
猿
最晚发现的DNA-结合同源物

图2B

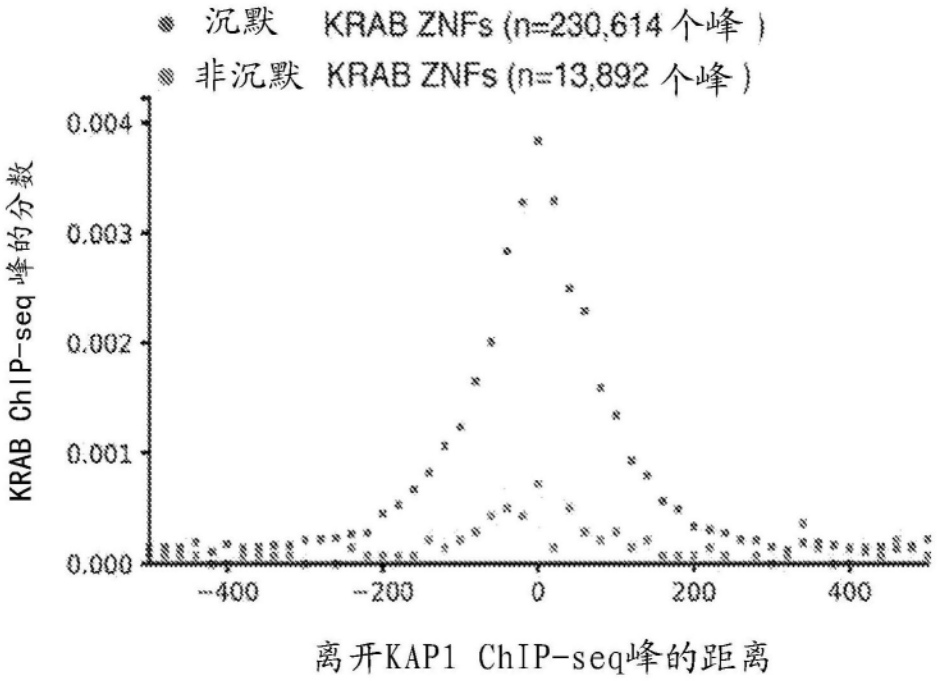

沉默 KRAB ZNFs (n=230,614 个峰)
非沉默 KRAB ZNFs (n=13,892 个峰)
KRAB ChIP-seq 峰的分数
0.004
0.003
0.002
0.001
0.000
-400
-200
0
200
400
离开KAP1 ChIP-seq峰的距离

图2C

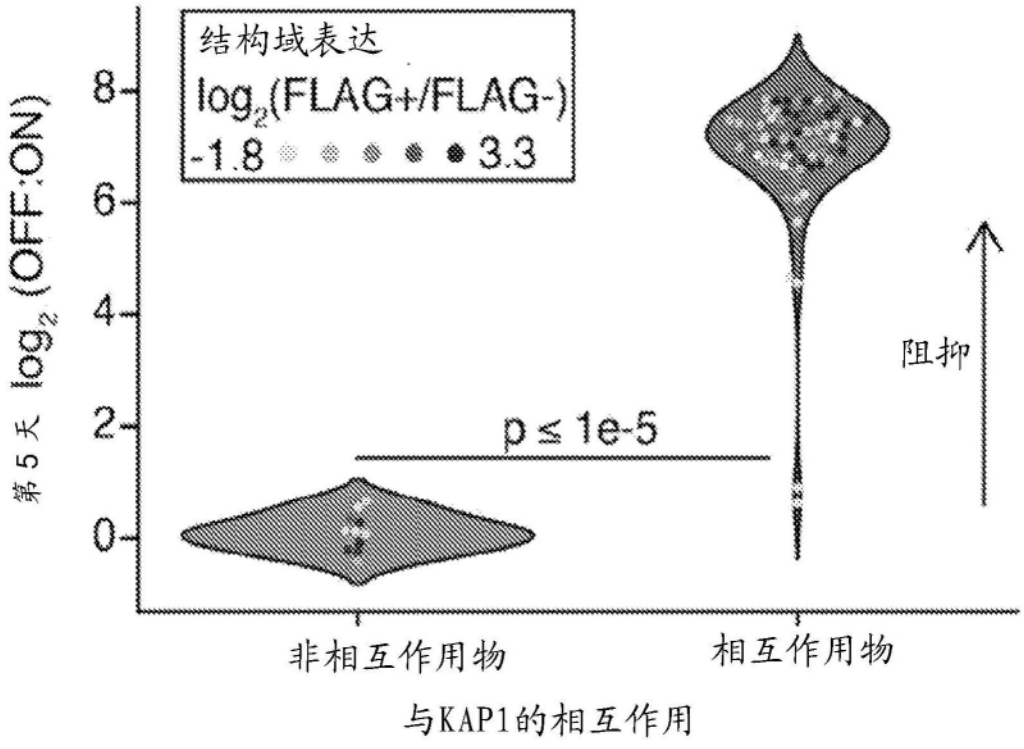

结构域表达
log2(FLAG+/FLAG-)
-1.8
3.3
第5天 log2 (OFF:ON)
8
6
4
2
0
p ≤ 1e-5
阻抑
非相互作用物
相互作用物
与KAP1的相互作用

图2D

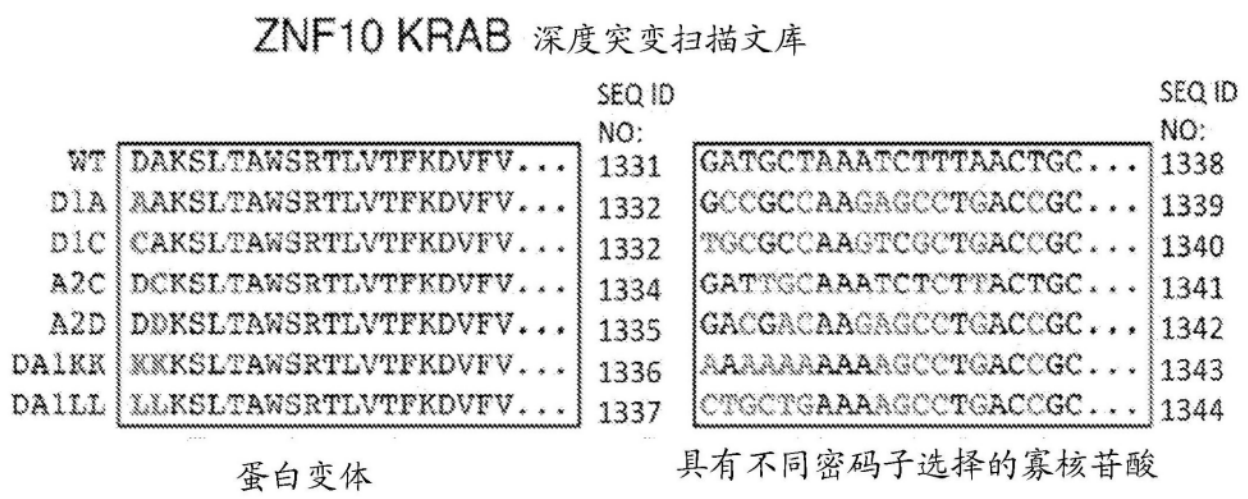

ZNF10 KRAB 深度突变扫描文库
SEQ ID NO:
SEQ ID NO:
WT DAKSLTAWSRTLVTFKDVFV... 1331 GATGCTAAATCTTTAACTGC... 1338
D1A AAKSLTAWSRTLVTFKDVFV... 1332 GCCGCCAAGAGCCTGACCGC... 1339
D1C CAKSLTAWSRTLVTFKDVFV... 1332 TGCGCCAAGTCGCTGACCGC... 1340
A2C DCKSLTAWSRTLVTFKDVFV... 1334 GATTGCAAATCTCTTACTGC... 1341
A2D DDKSLTAWSRTLVTFKDVFV... 1335 GACGACAAGAGCCTGACCGC... 1342
DA1KK KKKSLTAWSRTLVTFKDVFV... 1336 AAAAAAAAAAGCCTGACCGC... 1343
DA1LL LLKSLTAWSRTLVTFKDVFV... 1337 CTGCTGAAAAGCCTGACCGC... 1344
蛋白变体
具有不同密码子选择的寡核苷酸

图3A

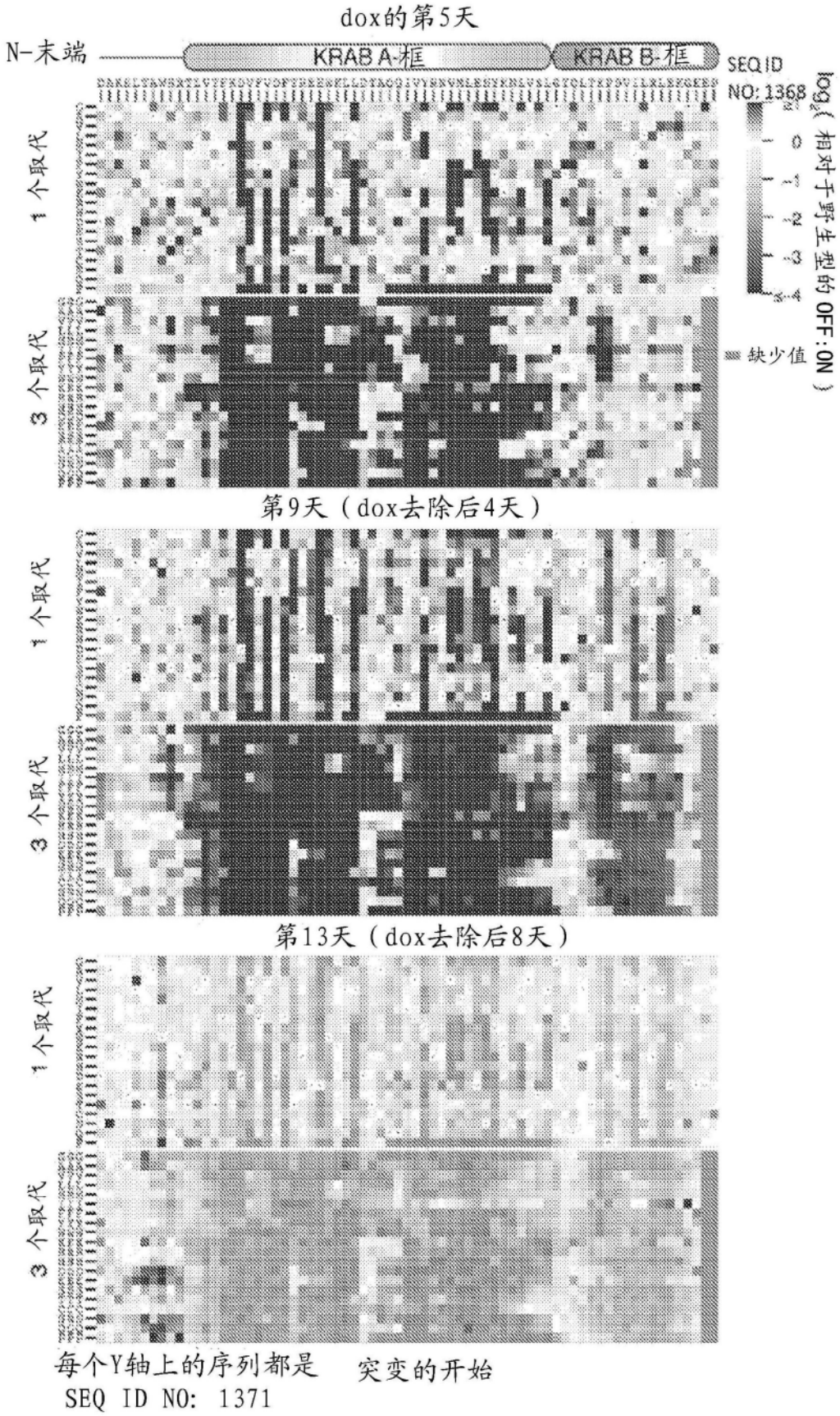

dox的第5天
N-末端
KRAB A-框
KRAB B-框
SEQ ID NO: 1368
log2(相对于野生型的OFF:ON)
缺少值
1个取代
3个取代
第9天（dox去除后4天）
第13天（dox去除后8天）
每个Y轴上的序列都是 SEQ ID NO: 1371
突变的开始

图3B

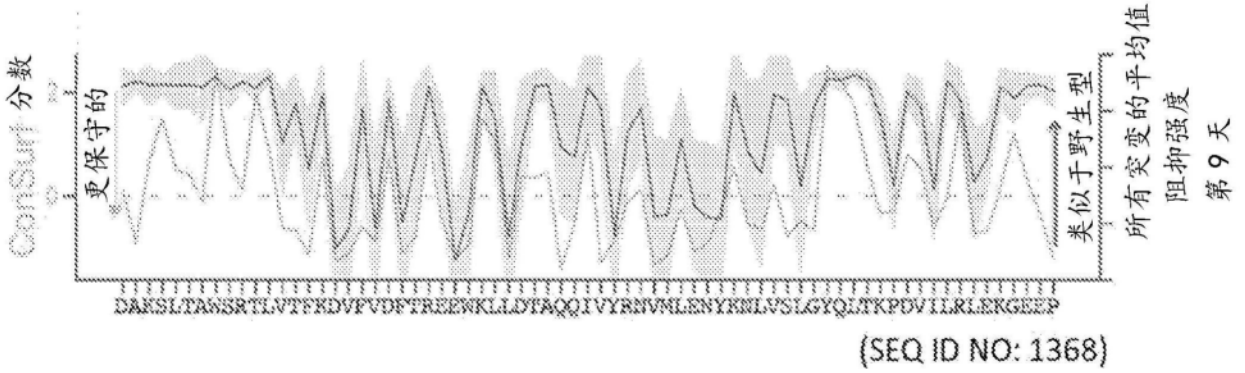

ConSurf 分数
更保守的
类似于野生型
所有突变的平均值
阻抑强度
第 9 天
DAKSLTANSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP
(SEQ ID NO: 1368)

图3C

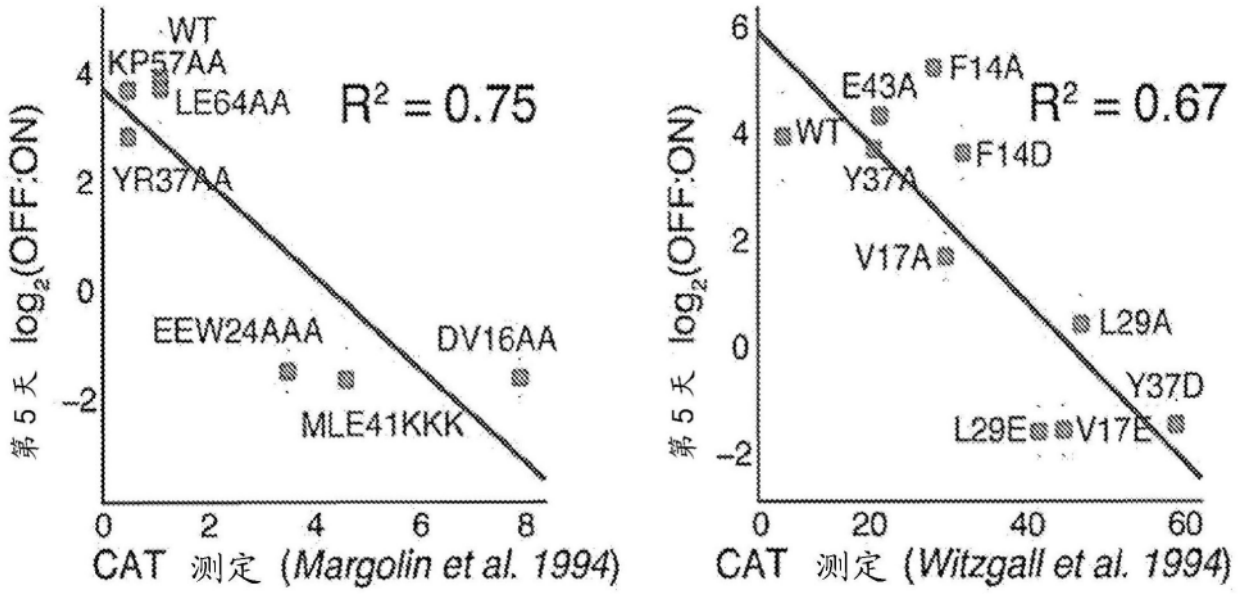

WT
KP57AA
LE64AA
YR37AA
EEW24AAA
DV16AA
MLE41KKK
R² = 0.75
第 5 天 log₂(OFF:ON)
CAT 测定 (Margolin et al. 1994)
E43A
F14A
WT
F14D
Y37A
V17A
L29A
Y37D
L29E
V17E
R² = 0.67
第 5 天 log₂(OFF:ON)
CAT 测定 (Witzgall et al. 1994)

图3D

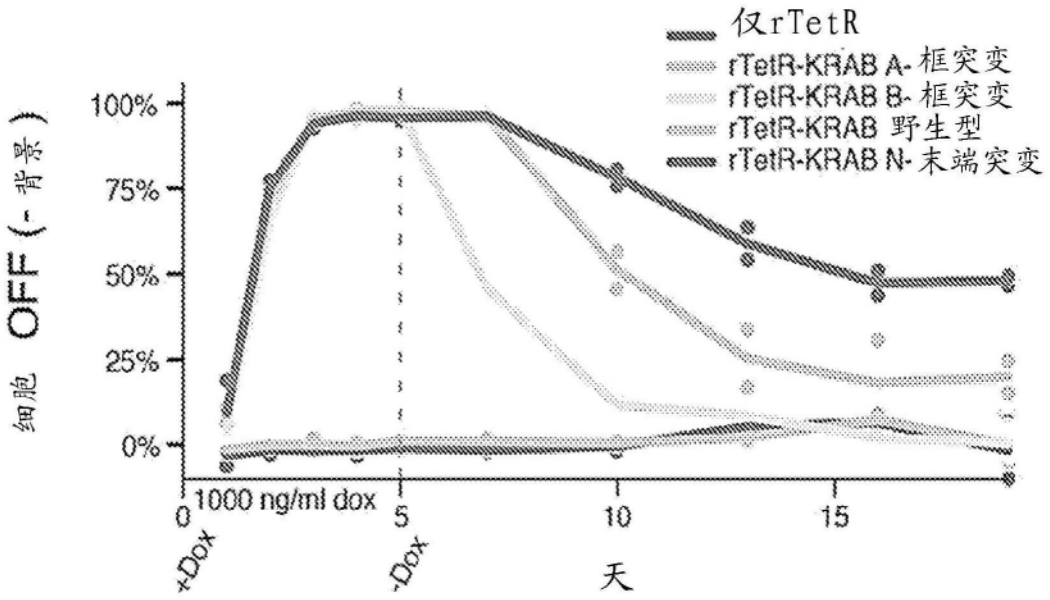

仅rTetR
rTetR-KRAB A-框突变
rTetR-KRAB B-框突变
rTetR-KRAB 野生型
rTetR-KRAB N-末端突变
细胞 OFF（-背景）
100%
75%
50%
25%
0%
1000 ng/ml dox
0
5
10
15
+Dox
-Dox
天

图3E

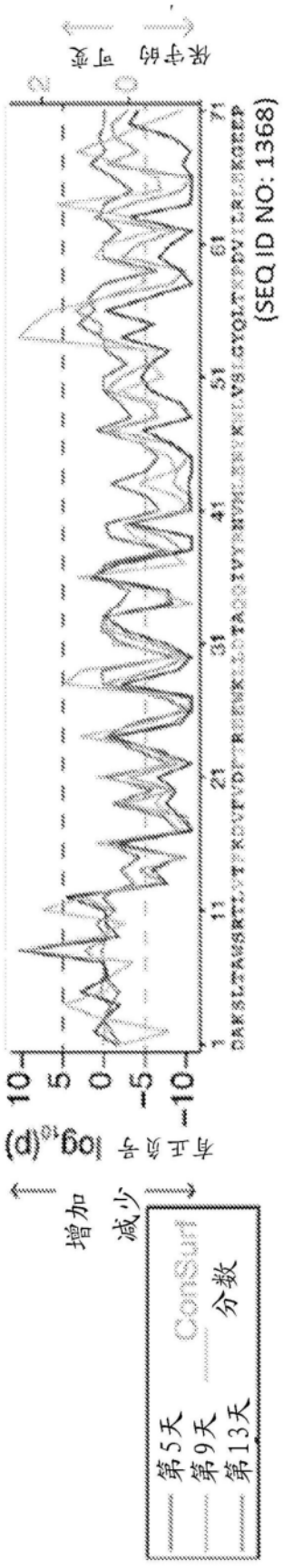


图3F

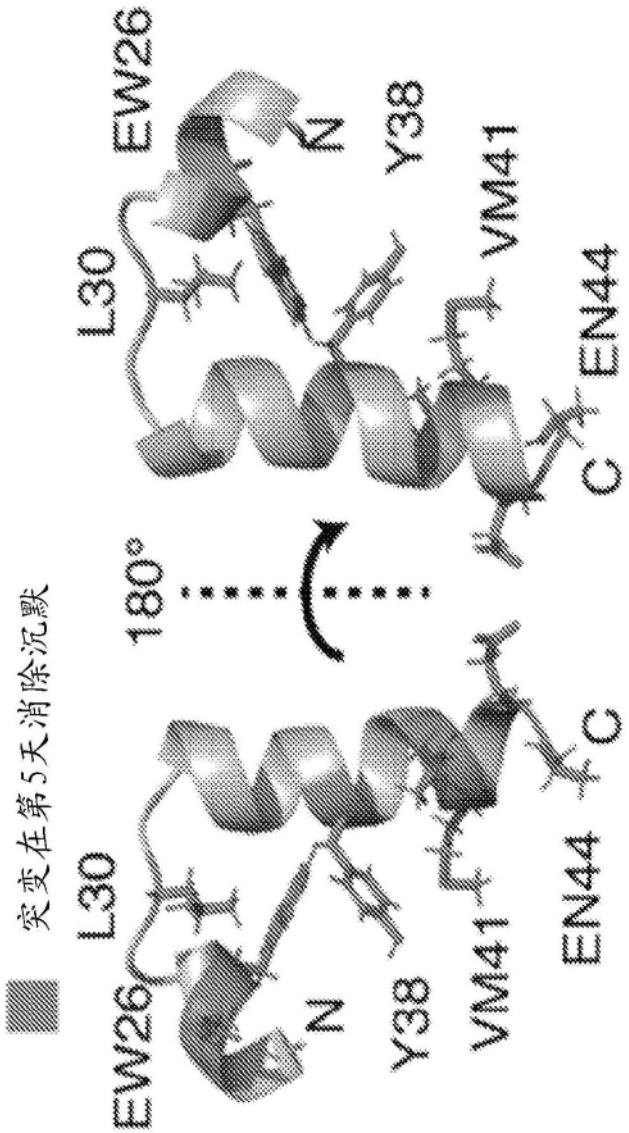


图3G

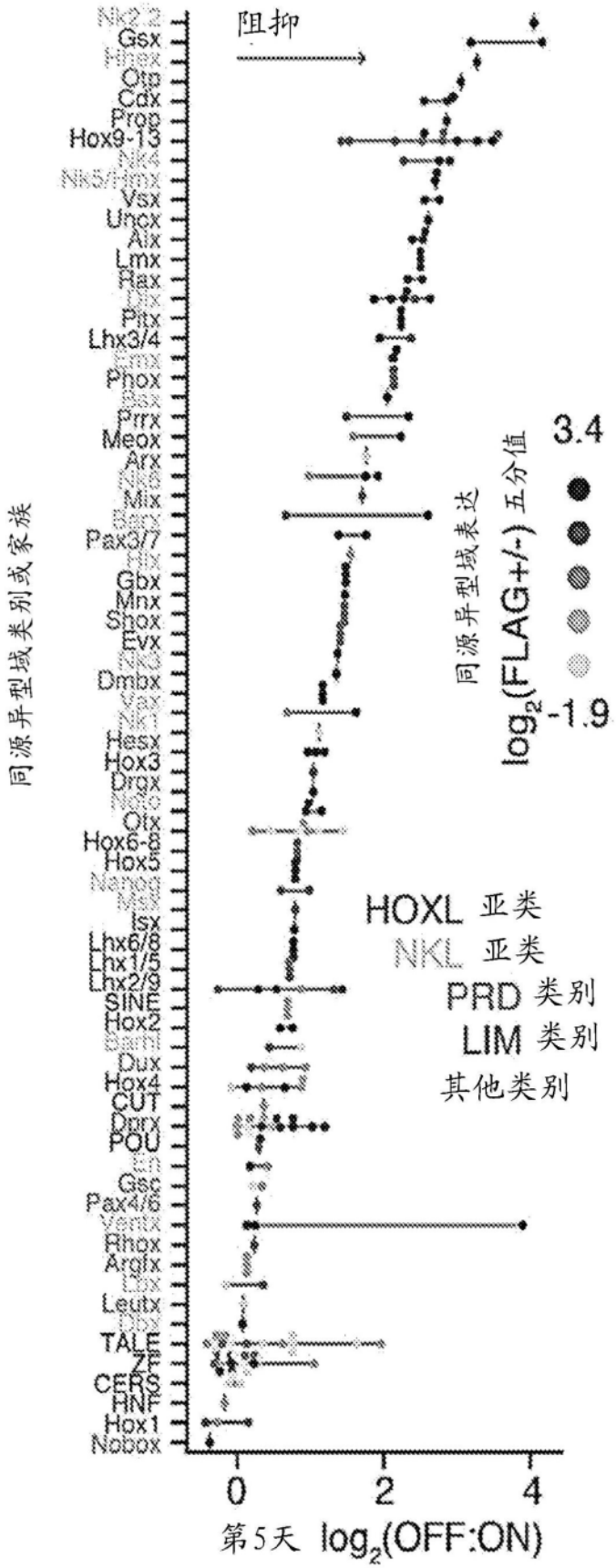

阻抑
同源异型域类别或家族
同源异型域表达
log2(FLAG+/-) 五分值
3.4
-1.9
HOXL 亚类
NKL 亚类
PRD 类别
LIM 类别
其他类别
Nk2.2
Gsx
Hmx
Otp
Cdx
Prop
Hox9-13
Nk4
Nk5/Hmx
Vsx
Uncx
Alx
Lmx
Hax
Pitx
Lhx3/4
Emx
Phox
Esx
Prrx
Meox
Arx
Mix
Bsx
Pax3/7
Tlx
Gbx
Mnx
Shox
Evx
Dmbx
Vax
Hesx
Hox3
Drgx
Nkx
Otx
Hox6-8
Hox5
Nelog
Msx
Isx
Lhx6/8
Lhx1/5
Lhx2/9
SINE
Hox2
Barhl
Dux
Hox4
CUT
Dprx
POU
En
Gsc
Pax4/6
Ventx
Rhox
Arglx
Cbx
Leutx
Dbx
TALE
Zf
CERS
HNF
Hox1
Nobox
0
2
4
第5天 log2(OFF:ON)

图4A

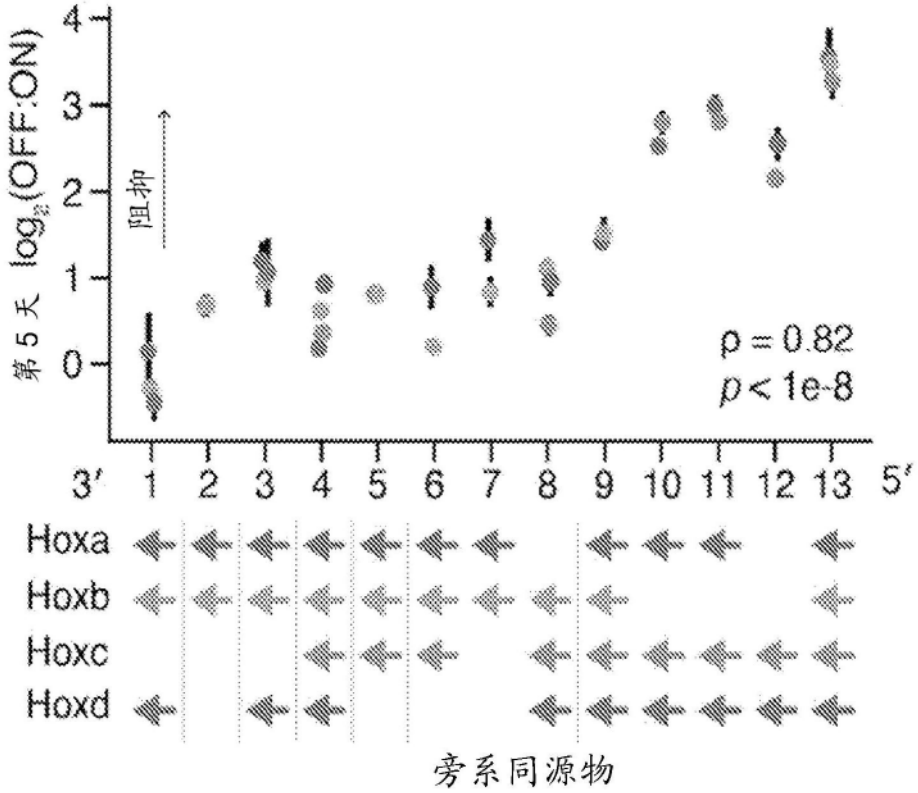

第5天 log2(OFF:ON)
阻抑
ρ = 0.82
p < 1e-8
3′ 1 2 3 4 5 6 7 8 9 10 11 12 13 5′
Hoxa
Hoxb
Hoxc
Hoxd
旁系同源物

图4B

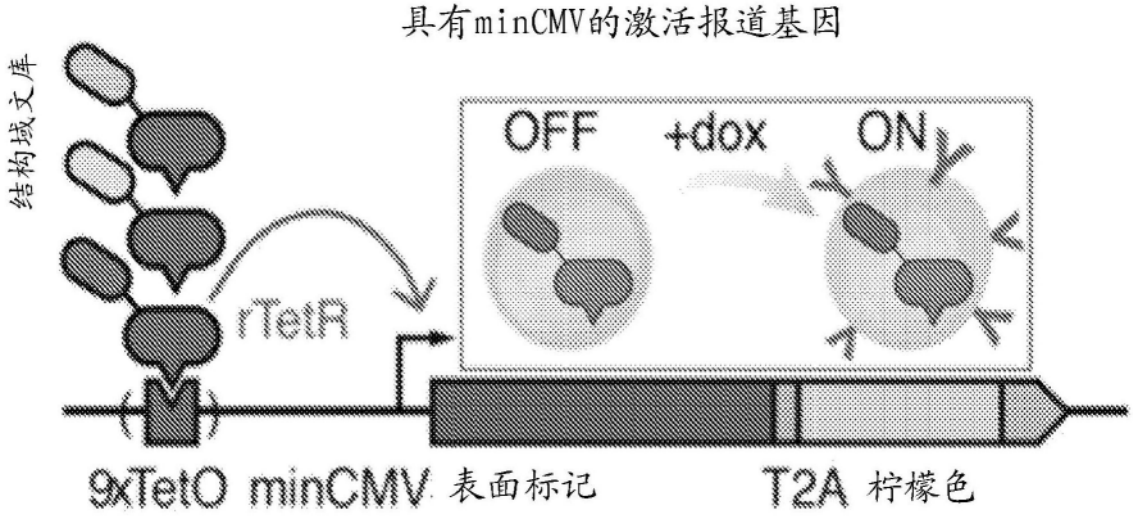

具有minCMV的激活报道基因
结构域文库
OFF
+dox
ON
rTetR
9xTetO minCMV
表面标记
T2A
柠檬色

图5A

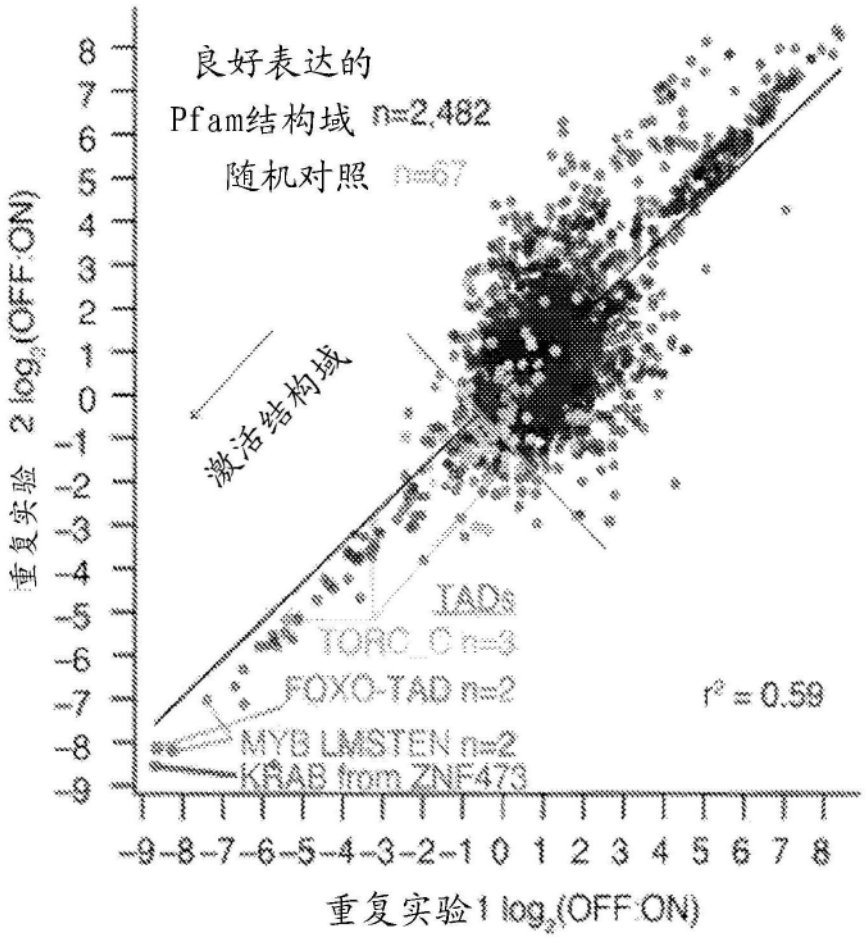

良好表达的
Pfam结构域 n=2,482
随机对照 n=67
激活结构域
TADs
TORC_C n=3
FOXO-TAD n=2
MYB LMSTEN n=2
KRAB from ZNF473
r² = 0.59
重复实验 2 log2(OFF:ON)
重复实验1 log2(OFF:ON)

图5B

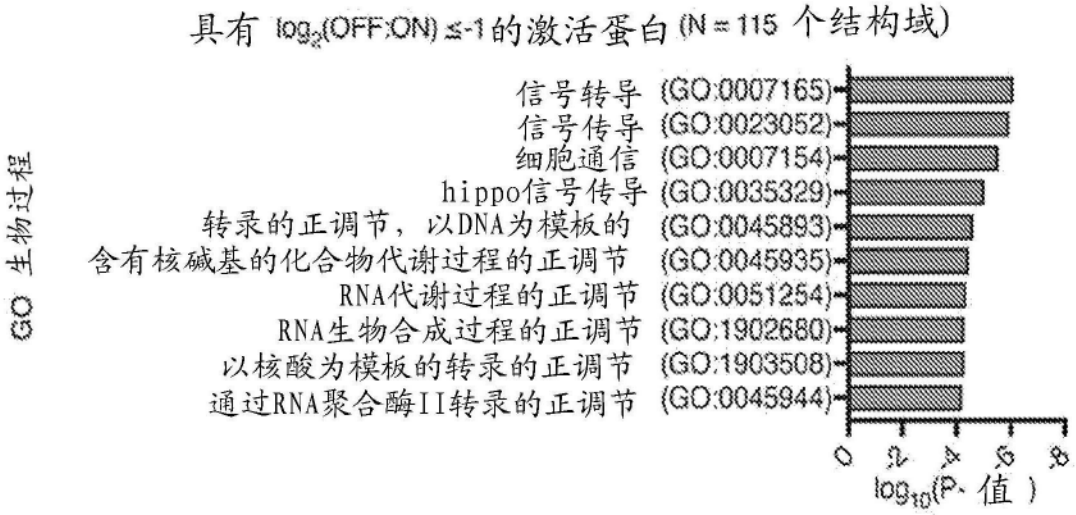

具有 log2(OFF:ON) ≤-1的激活蛋白 (N = 115 个结构域)
GO 生物过程
信号转导 (GO:0007165)
信号传导 (GO:0023052)
细胞通信 (GO:0007154)
hippo信号传导 (GO:0035329)
转录的正调节，以DNA为模板的 (GO:0045893)
含有核碱基的化合物代谢过程的正调节 (GO:0045935)
RNA代谢过程的正调节 (GO:0051254)
RNA生物合成过程的正调节 (GO:1902680)
以核酸为模板的转录的正调节 (GO:1903508)
通过RNA聚合酶II转录的正调节 (GO:0045944)
log10(P-值)

图5C

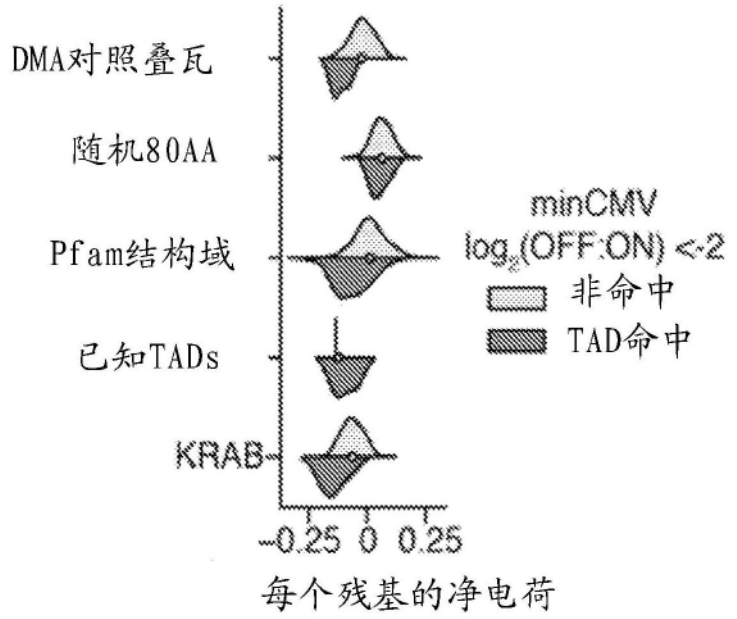

DMA对照叠瓦
随机80AA
Pfam结构域
已知TADs
KRAB
minCMV
log2(OFF:ON) <-2
非命中
TAD命中
-0.25 0 0.25
每个残基的净电荷

图5D

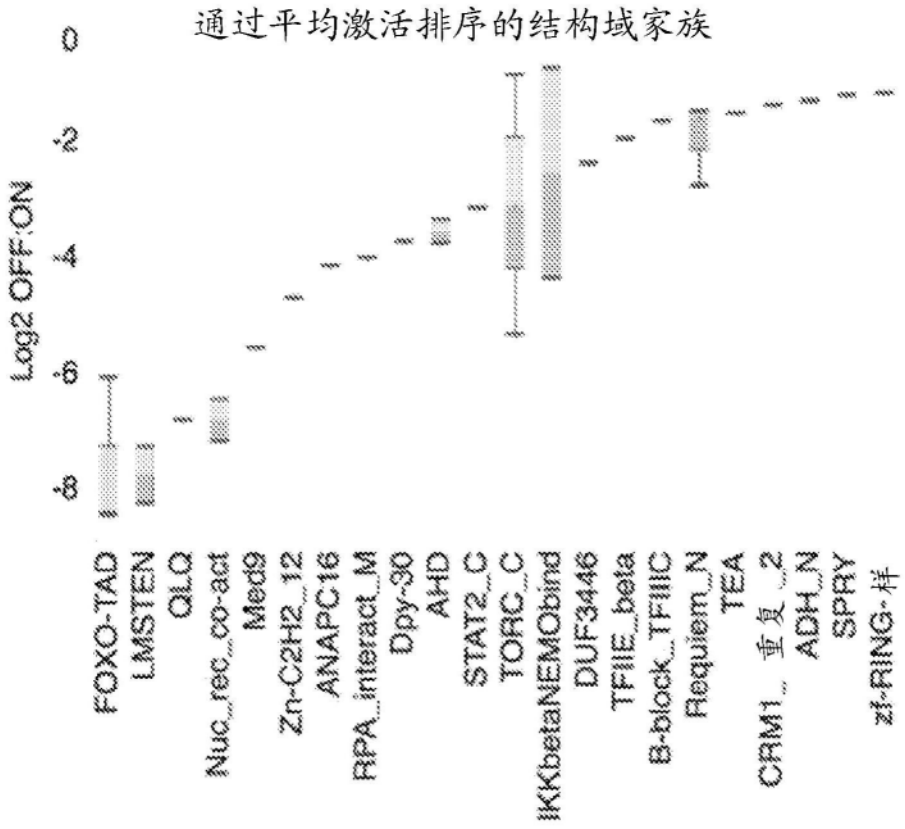

通过平均激活排序的结构域家族
Log2 OFF:ON
0
-2
-4
-6
-8
FOXO-TAD
LMSTEN
QLQ
Nuc_rec_co-act
Med9
Zn-C2H2_12
ANAPC16
RPA_interact_M
Dpy-30
AHD
STAT2_C
TORC_C
IKKbetaNEMObind
DUF3446
TFIIE_beta
B-block_TFIIIC
Requiem_N
TEA
CRM1_重复_2
ADH_N
SPRY
zf-RING-样

图5E

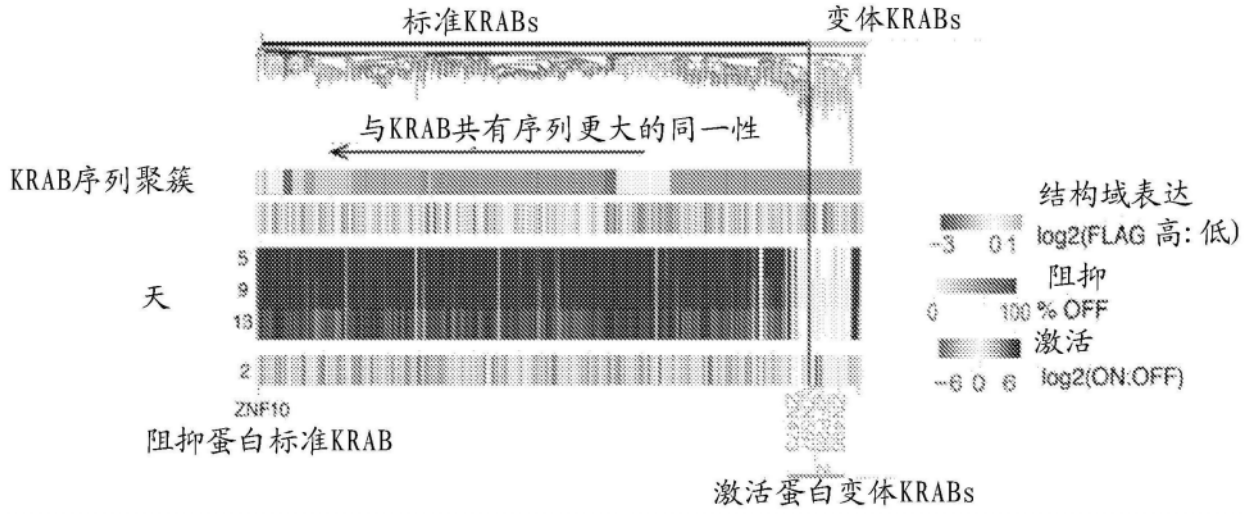


图5F

叠瓦来自不同阻抑复合物的238种蛋白，例如，组蛋白调节物、polycomb、NoRC、DNA甲基化酶、CTCF、KRAB-ZNFs、阻抑TFs等

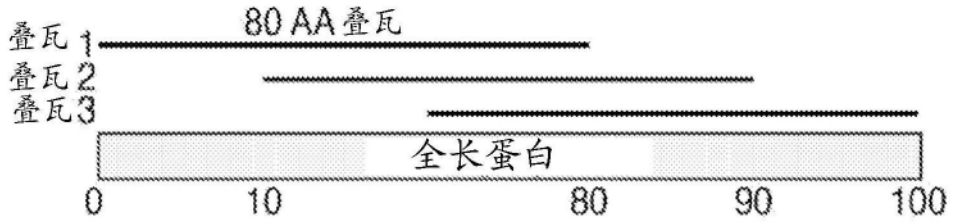


图6A

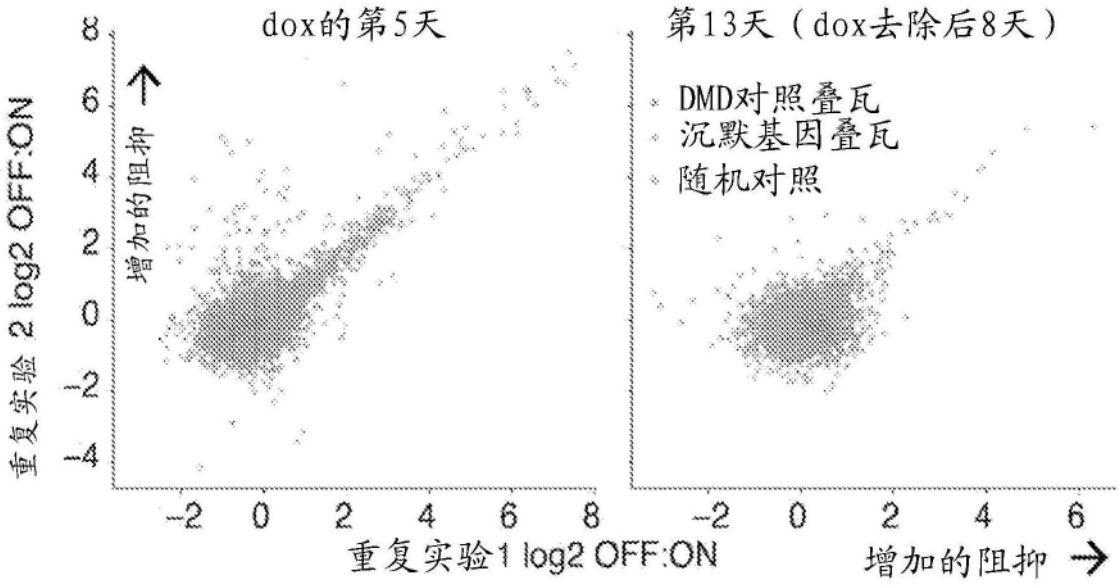

dox的第5天
第13天（dox去除后8天）
DMD对照叠瓦
沉默基因叠瓦
随机对照
重复实验 2 log2 OFF:ON
增加的阻抑
重复实验1 log2 OFF:ON
增加的阻抑

图6B

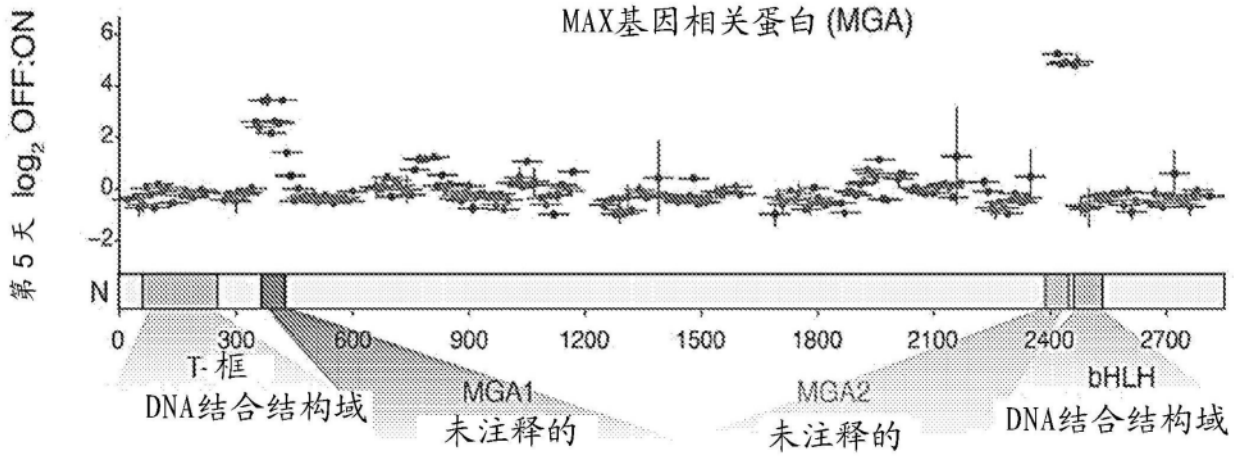

MAX基因相关蛋白 (MGA)
第 5 天 log2 OFF:ON
T-框
DNA结合结构域
MGA1
未注释的
MGA2
未注释的
bHLH
DNA结合结构域

图6C

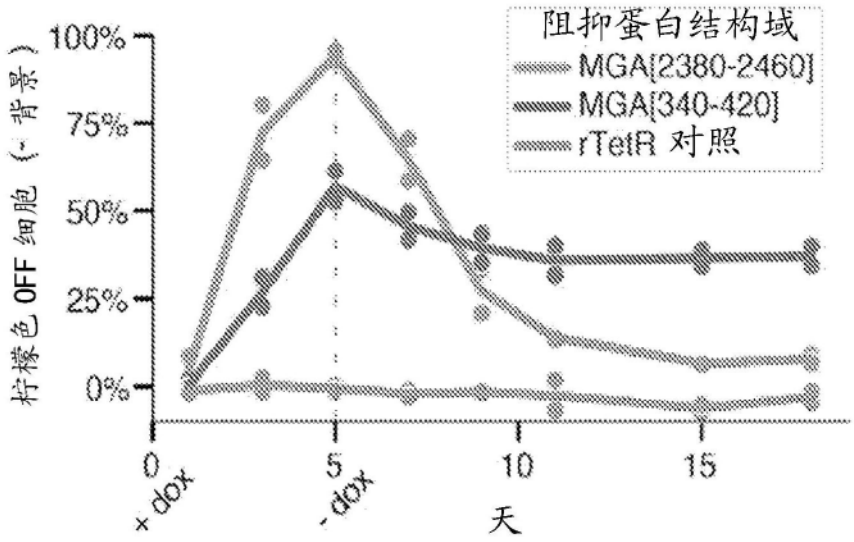

阻抑蛋白结构域
MGA[2380-2460]
MGA[340-420]
rTetR 对照
柠檬色 OFF 细胞（- 背景）
100%
75%
50%
25%
0%
0
5
10
15
+ dox
- dox
天

图6D

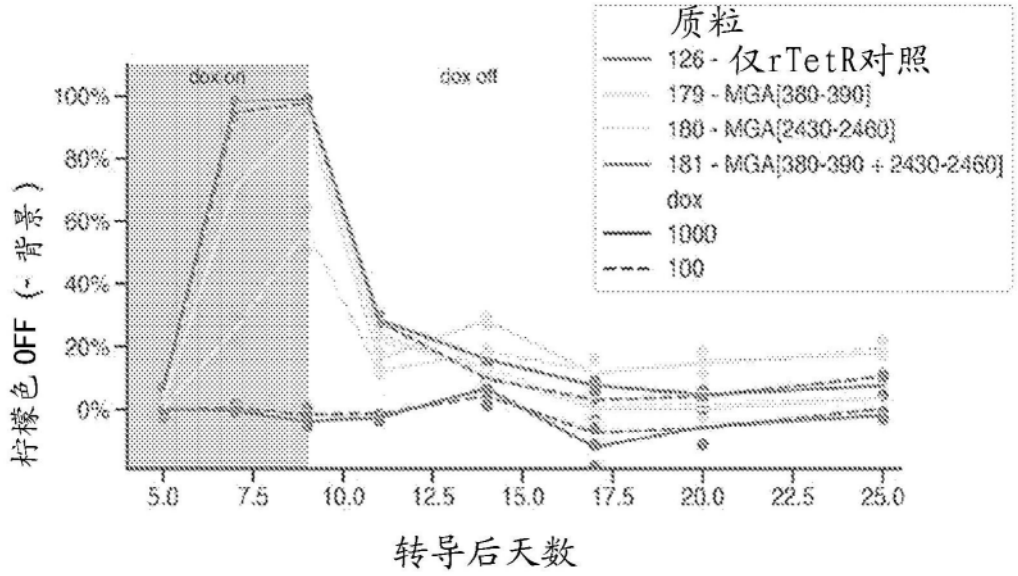

质粒
126 - 仅rTetR对照
179 - MGA[380-390]
180 - MGA[2430-2460]
181 - MGA[380-390 + 2430-2460]
dox
1000
100
dox on
dox off
柠檬色 OFF（- 背景）
100%
80%
60%
40%
20%
0%
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
转导后天数

图6E

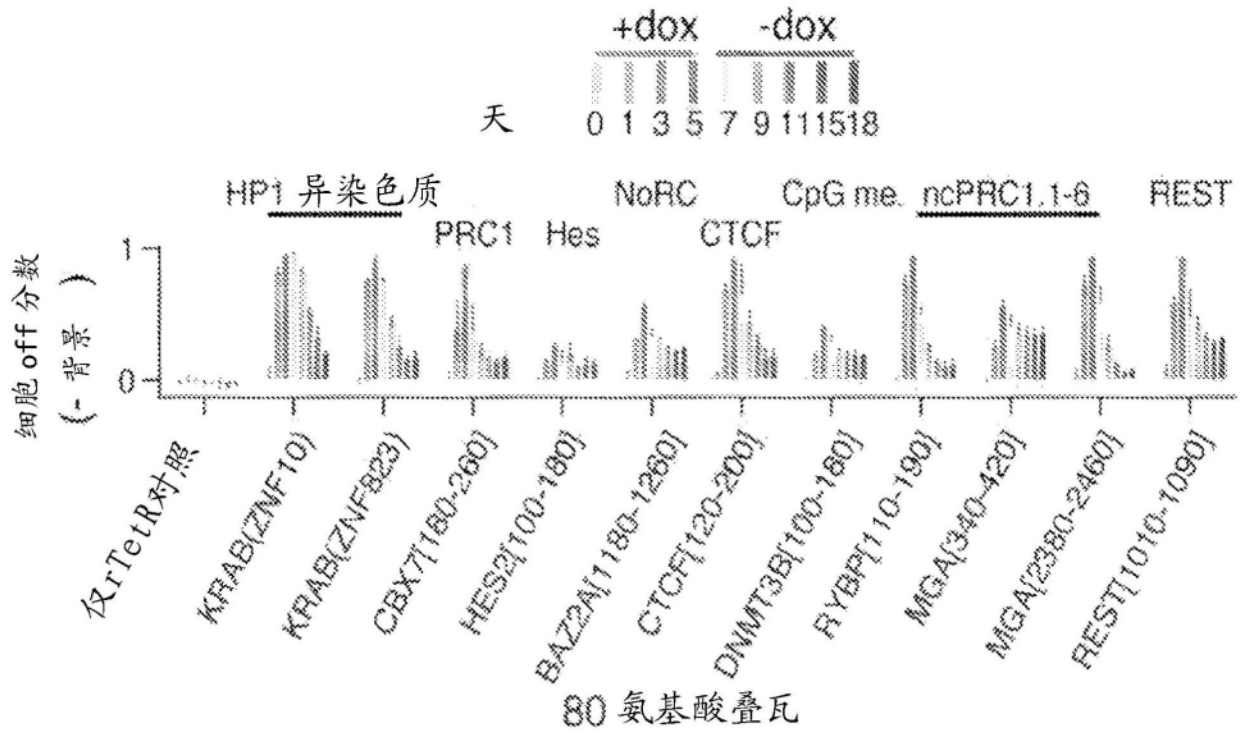

+dox
-dox
天
0 1 3 5 7 9 11 15 18
HP1 异染色质
NoRC
CpG me.
ncPRC1.1-6
REST
PRC1
Hes
CTCF
细胞 off 分数
(- 背景)
1
0
仅rTetR对照
KRAB(ZNF10)
KRAB(ZNF823)
CBX7[180-260]
HES2[100-180]
BAZ2A[1180-1260]
CTCF[120-200]
DNMT3B[100-160]
RYBP[110-190]
MGA[340-420]
MGA[2380-2460]
REST[1010-1090]
80 氨基酸叠瓦

图6F

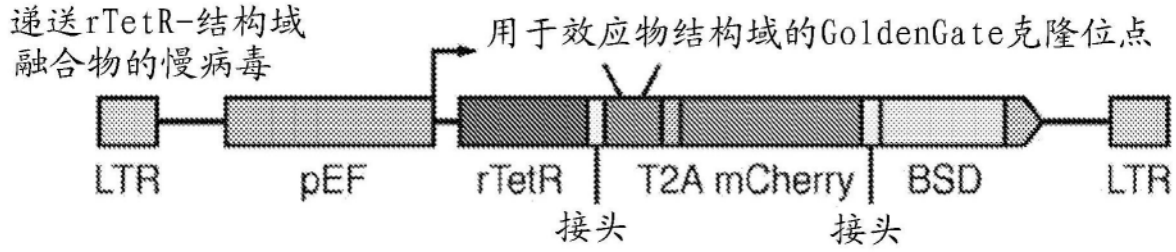

递送rTetR-结构域
融合物的慢病毒
用于效应物结构域的GoldenGate克隆位点
LTR
pEF
rTetR
T2A mCherry
BSD
LTR
接头
接头

图7A

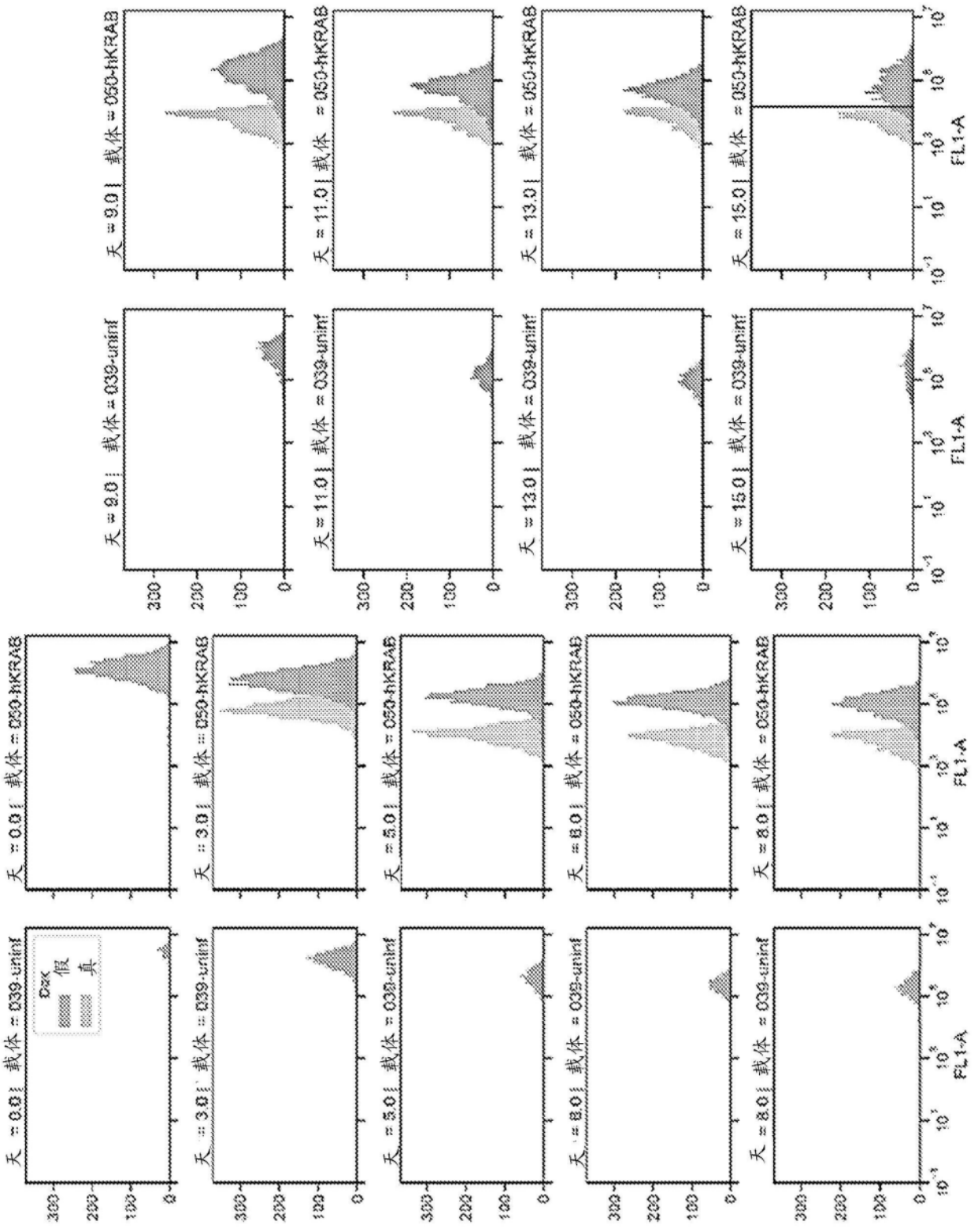

天 = 0.0 | 载体 = 039-uninf
天 = 0.0 | 载体 = 050-hKRAB
天 = 3.0 | 载体 = 039-uninf
天 = 3.0 | 载体 = 050-hKRAB
天 = 5.0 | 载体 = 039-uninf
天 = 5.0 | 载体 = 050-hKRAB
天 = 8.0 | 载体 = 039-uninf
天 = 8.0 | 载体 = 050-hKRAB
天 = 9.0 | 载体 = 039-uninf
天 = 9.0 | 载体 = 050-hKRAB
天 = 11.0 | 载体 = 039-uninf
天 = 11.0 | 载体 = 050-hKRAB
天 = 13.0 | 载体 = 039-uninf
天 = 13.0 | 载体 = 050-hKRAB
天 = 15.0 | 载体 = 039-uninf
天 = 15.0 | 载体 = 050-hKRAB
Dox
假
真
FL1-A
300
200
100
0

图7B

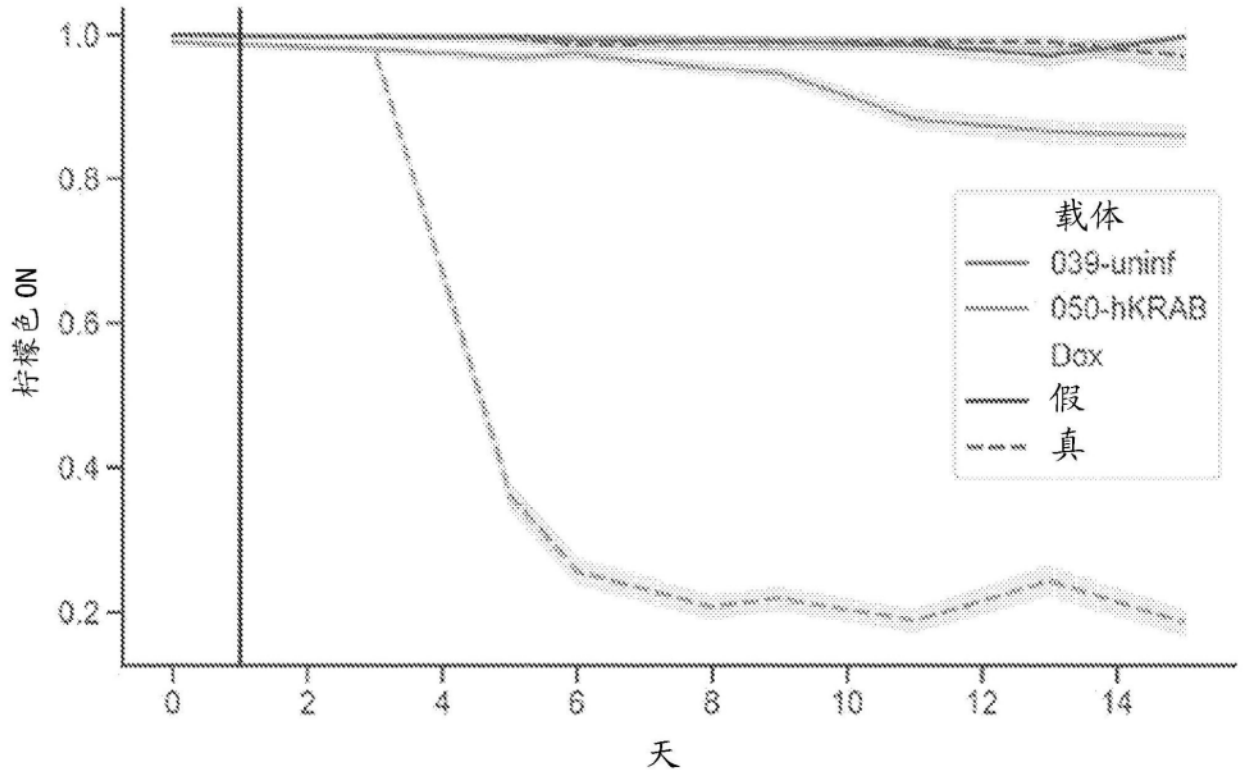

1.0
0.8
0.6
0.4
0.2
柠檬色 ON
0
2
4
6
8
10
12
14
天
载体
039-uninf
050-hKRAB
Dox
假
真

图7C

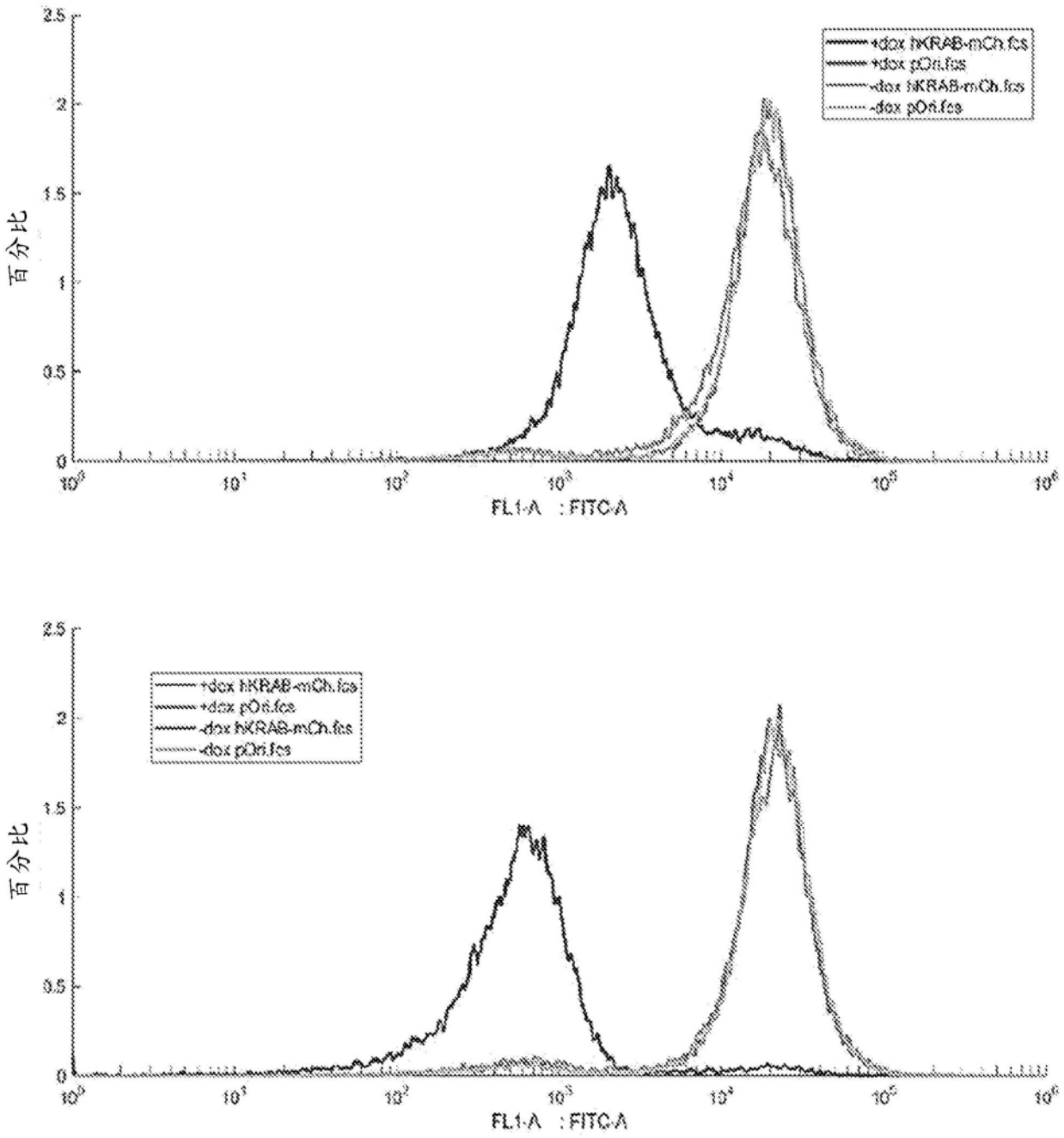

+dox hKRAB-mCh.fcs
+dox pOri.fcs
-dox hKRAB-mCh.fcs
-dox pOri.fcs
百分比
FL1-A :FITC-A
+dox hKRAB-mCh.fcs
+dox pOri.fcs
-dox hKRAB-mCh.fcs
-dox pOri.fcs
百分比
FL1-A :FITC-A

图7D

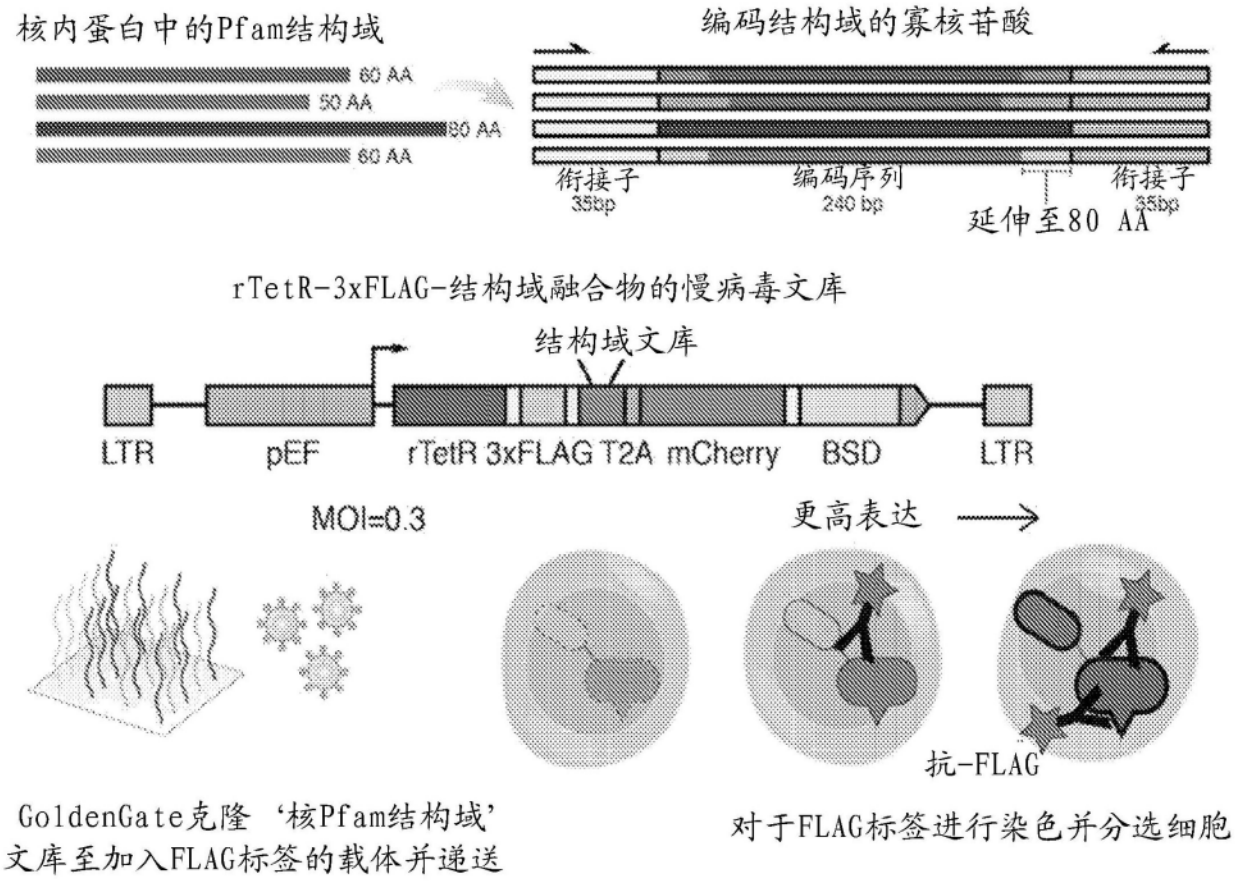

核内蛋白中的Pfam结构域
60 AA
50 AA
80 AA
60 AA
编码结构域的寡核苷酸
衔接子
35bp
编码序列
240 bp
延伸至80 AA
衔接子
35bp
rTetR-3xFLAG-结构域融合物的慢病毒文库
结构域文库
LTR
pEF
rTetR
3xFLAG
T2A
mCherry
BSD
LTR
MOI=0.3
更高表达
抗-FLAG
GoldenGate克隆‘核Pfam结构域’
文库至加入FLAG标签的载体并递送
对于FLAG标签进行染色并分选细胞

图8A

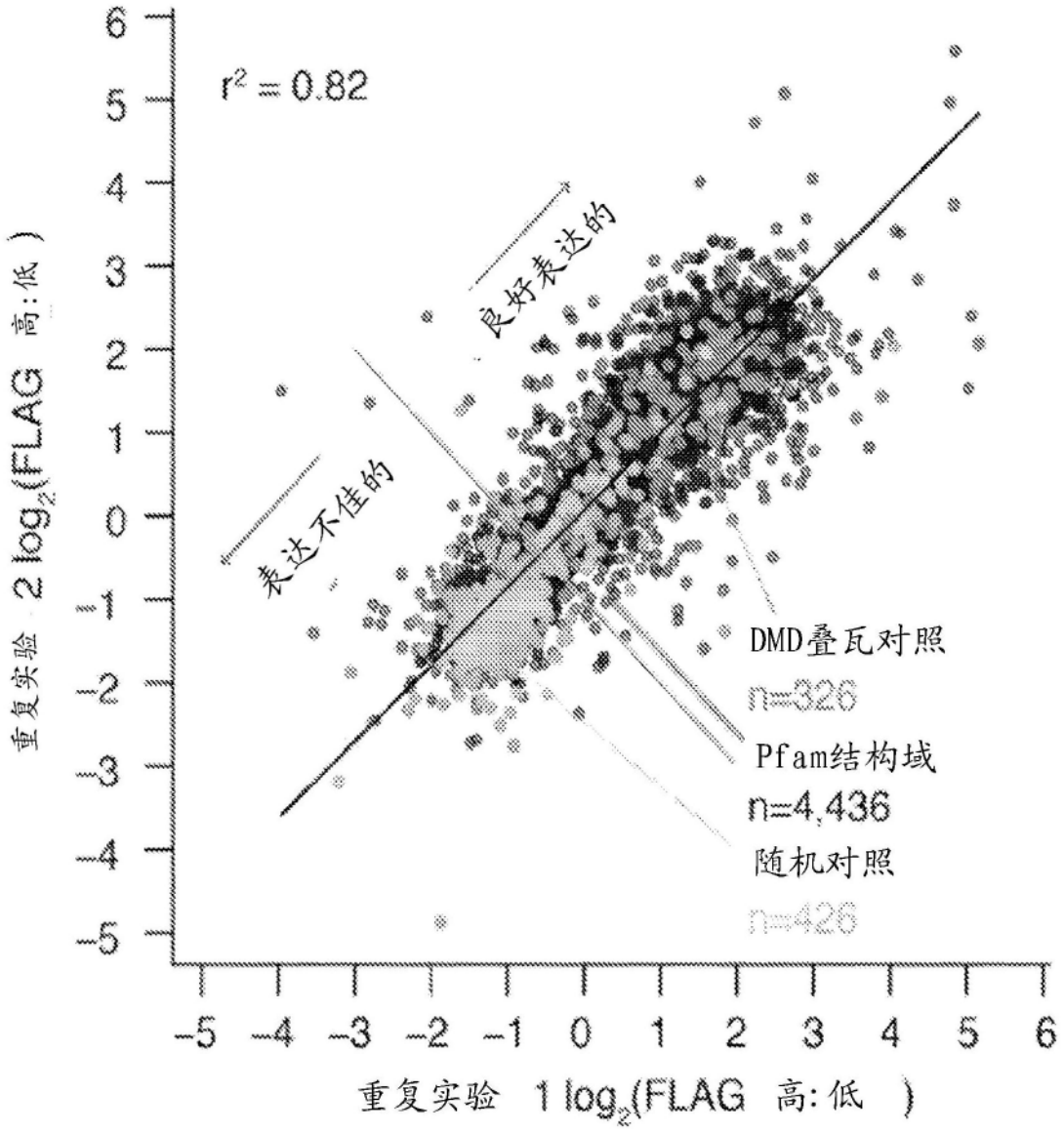

$r^2 = 0.82$
良好表达的
表达不佳的
DMD叠瓦对照
n=326
Pfam结构域
n=4,436
随机对照
n=426
重复实验 2 $\log_2$(FLAG 高:低)
重复实验 1 $\log_2$(FLAG 高:低)
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5 -4 -3 -2 -1 0 1 2 3 4 5 6

图8B

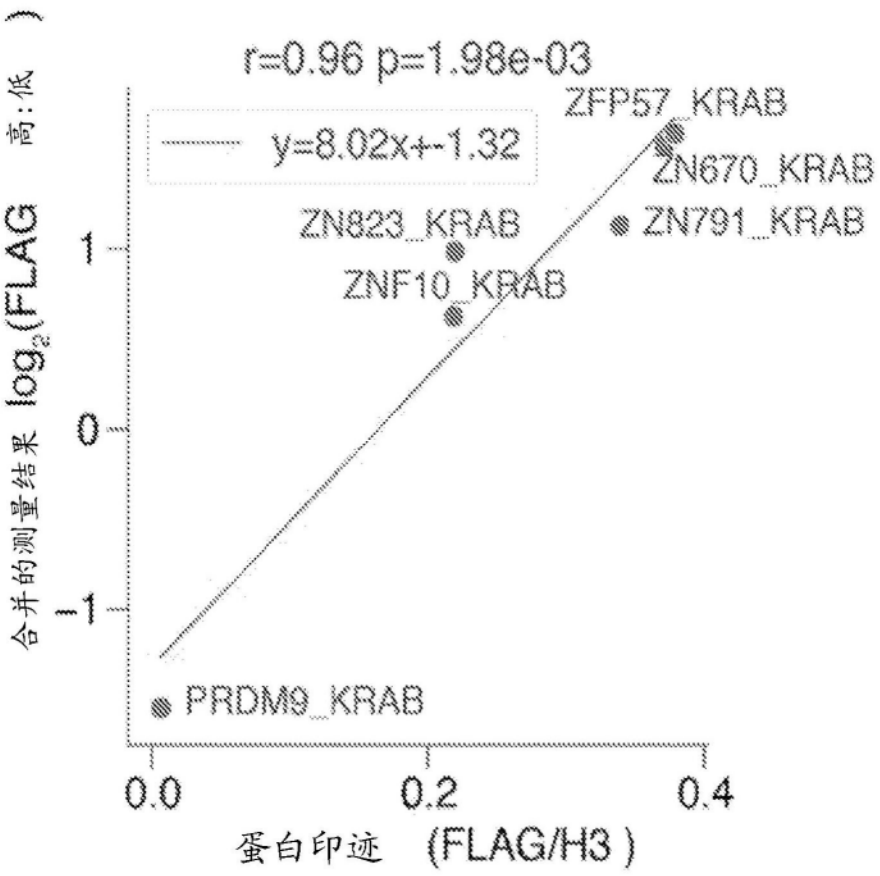

r=0.96 p=1.98e-03
y=8.02x+-1.32
ZFP57_KRAB
ZN670_KRAB
ZN791_KRAB
ZN823_KRAB
ZNF10_KRAB
PRDM9_KRAB
合并的测量结果 log2(FLAG 高:低)
1
0
-1
0.0
0.2
0.4
蛋白印迹 (FLAG/H3)

图8C

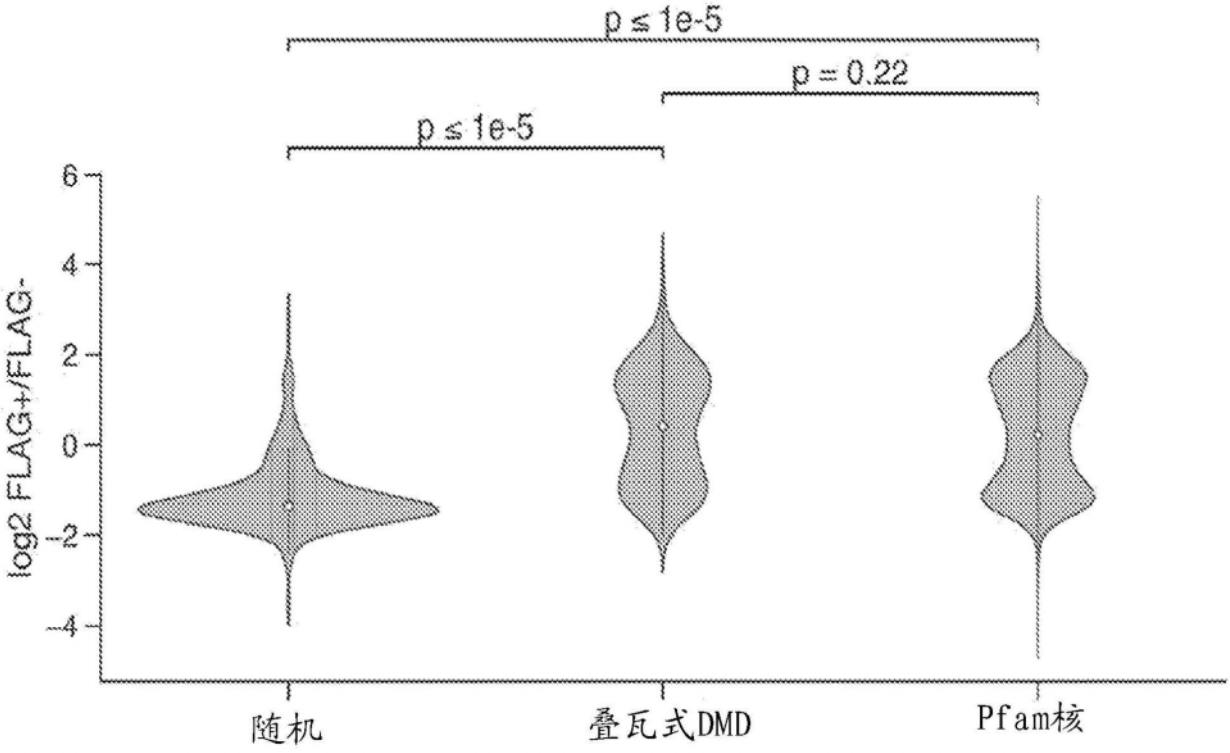


图8D

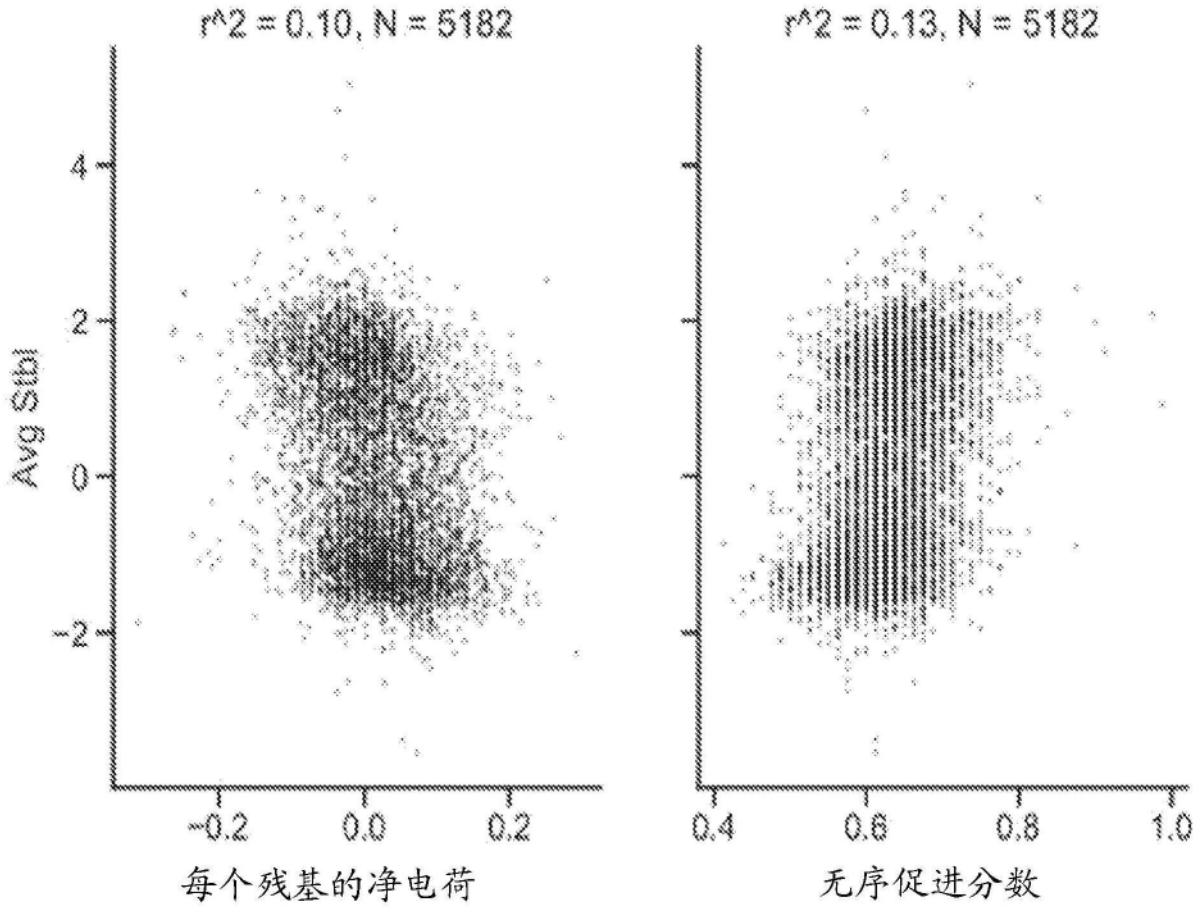

r^2 = 0.10, N = 5182
r^2 = 0.13, N = 5182
Avg Stbl
4
2
0
-2
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
每个残基的净电荷
无序促进分数

图8E

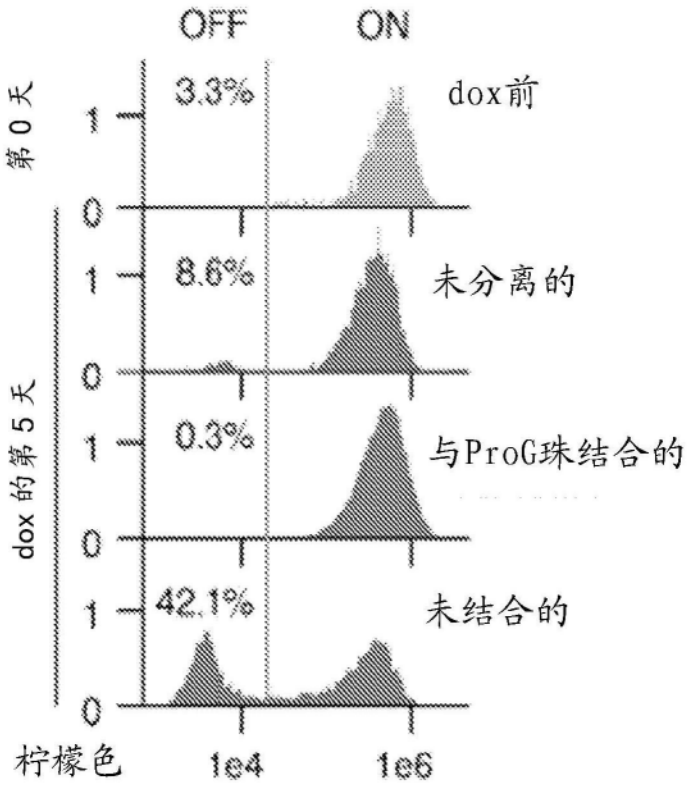

OFF
ON
第0天
3.3%
dox前
8.6%
未分离的
dox的第5天
0.3%
与ProG珠结合的
42.1%
未结合的
柠檬色
1e4
1e6

图9A

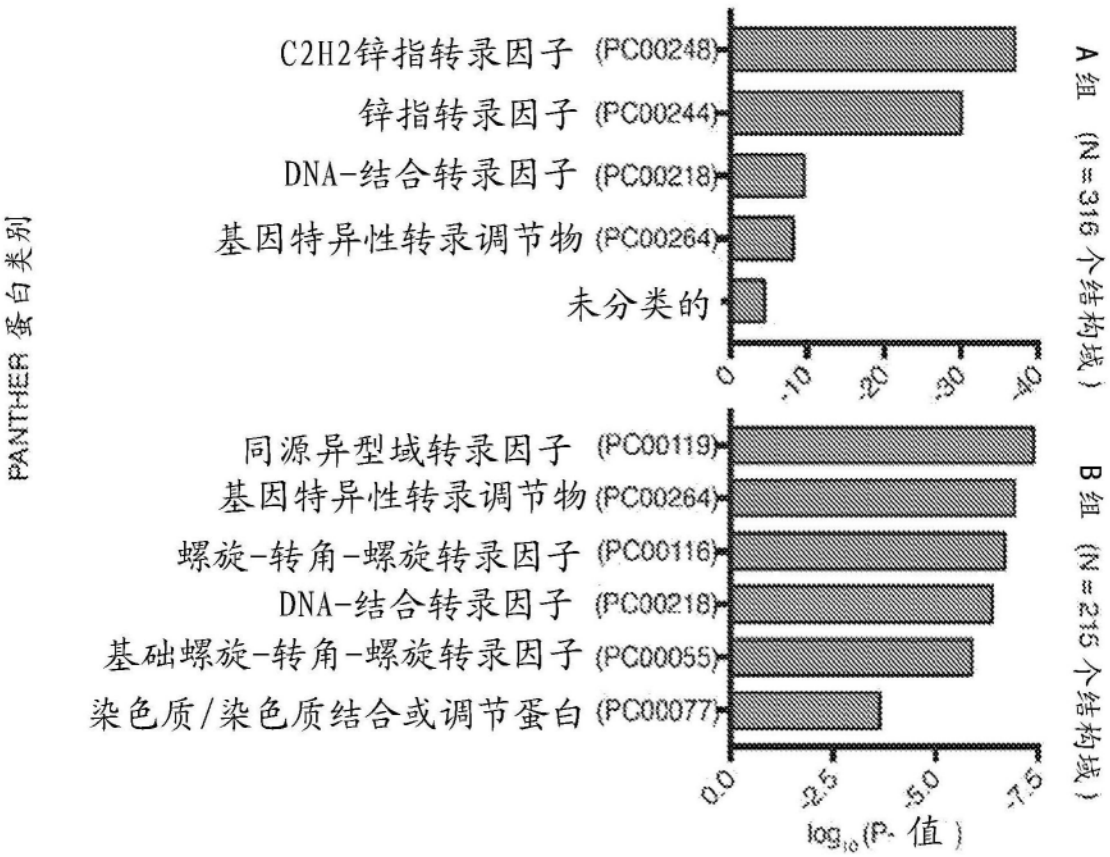

PANTHER 蛋白类别
C2H2锌指转录因子 (PC00248)
锌指转录因子 (PC00244)
DNA-结合转录因子 (PC00218)
基因特异性转录调节物 (PC00264)
未分类的
A 组 (N = 316 个结构域)
同源异型域转录因子 (PC00119)
基因特异性转录调节物 (PC00264)
螺旋-转角-螺旋转录因子 (PC00116)
DNA-结合转录因子 (PC00218)
基础螺旋-转角-螺旋转录因子 (PC00055)
染色质/染色质结合或调节蛋白 (PC00077)
B 组 (N = 215 个结构域)
0 -10 -20 -30 -40
0.0 -2.5 -5.0 -7.5
log10(P-值)

图9B

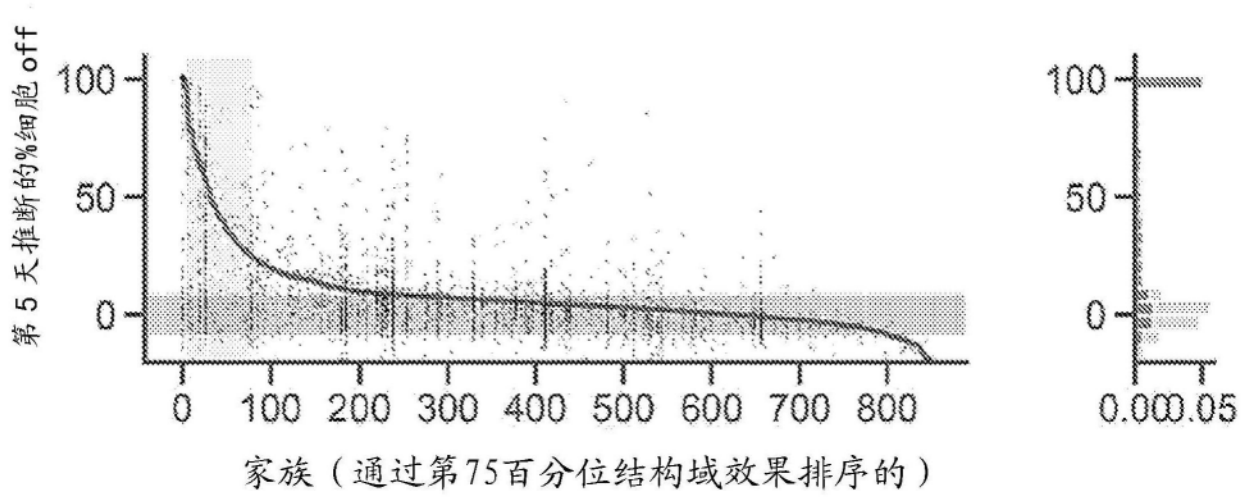

第5天推断的%细胞off
100
50
0
0 100 200 300 400 500 600 700 800
家族（通过第75百分位结构域效果排序的）
100
50
0
0.00 0.05

图9C

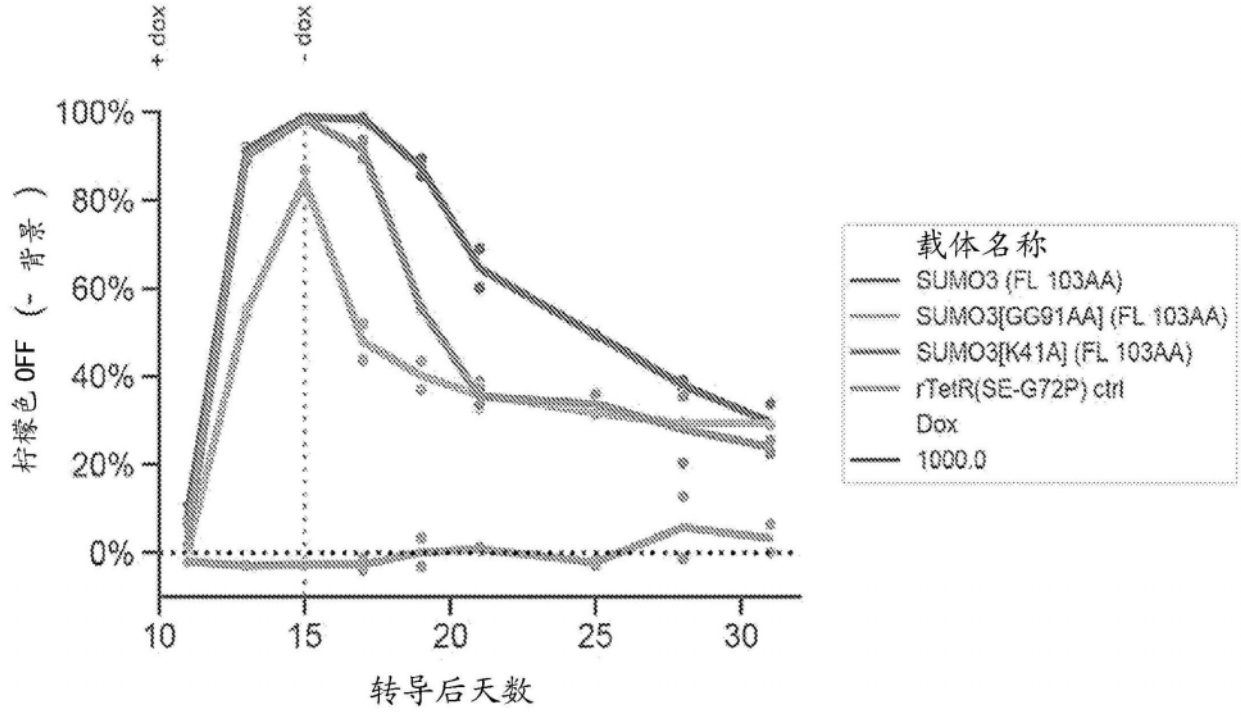

+ dox
- dox
100%
80%
60%
40%
20%
0%
柠檬色 OFF（- 背景）
10
15
20
25
30
转导后天数
载体名称
SUMO3 (FL 103AA)
SUMO3[GG91AA] (FL 103AA)
SUMO3[K41A] (FL 103AA)
rTetR(SE-G72P) ctrl
Dox
1000.0

图9D

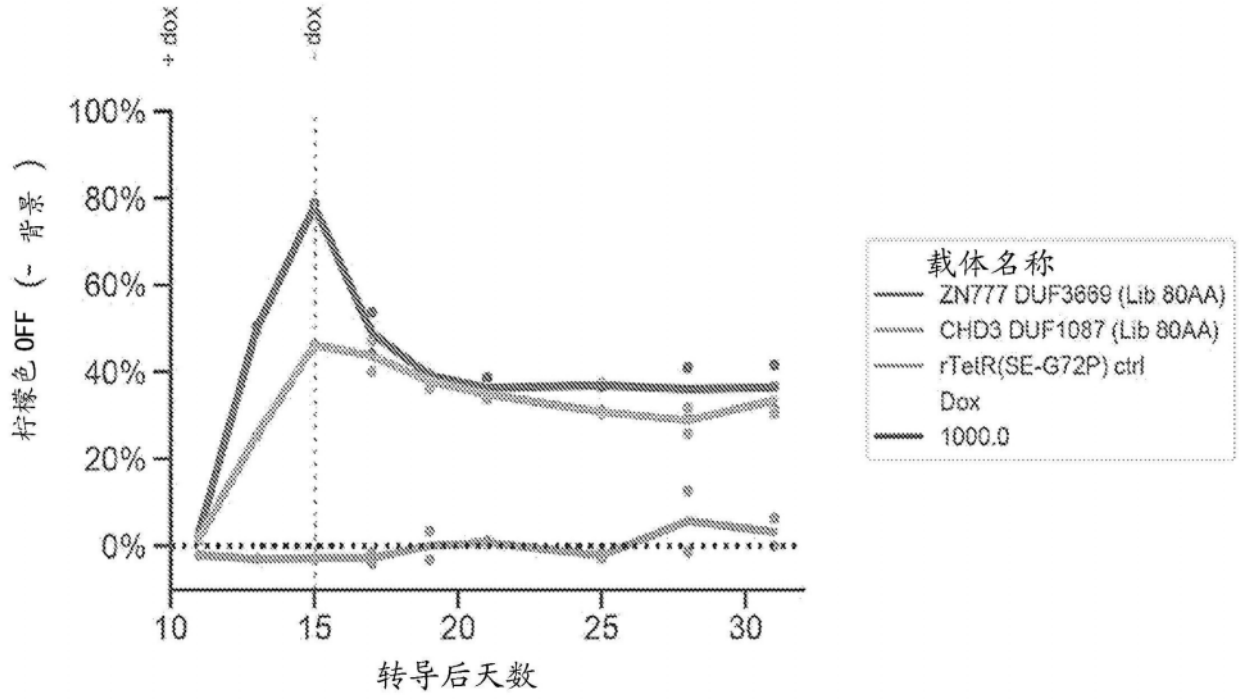

+ dox
- dox
100%
80%
60%
40%
20%
0%
柠檬色 OFF（- 背景）
10
15
20
25
30
转导后天数
载体名称
ZN777 DUF3669 (Lib 80AA)
CHD3 DUF1087 (Lib 80AA)
rTetR(SE-G72P) ctrl
Dox
1000.0

图9E

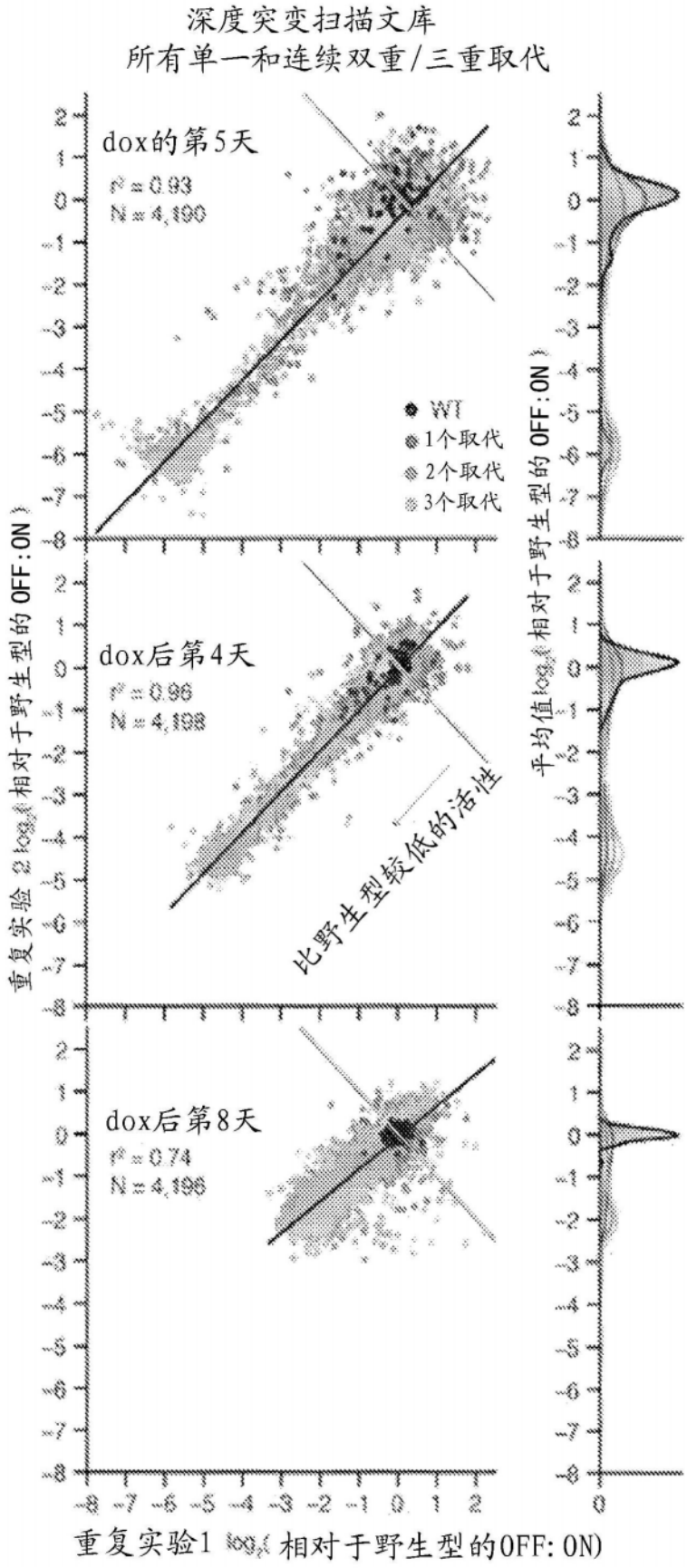

深度突变扫描文库
所有单一和连续双重/三重取代
dox的第5天
r² = 0.93
N = 4,190
WT
1个取代
2个取代
3个取代
dox后第4天
r² = 0.96
N = 4,188
比野生型较低的活性
dox后第8天
r² = 0.74
N = 4,196
重复实验2 log2(相对于野生型的OFF:ON)
平均值log2(相对于野生型的OFF:ON)
重复实验1 log2(相对于野生型的OFF:ON)

图10A

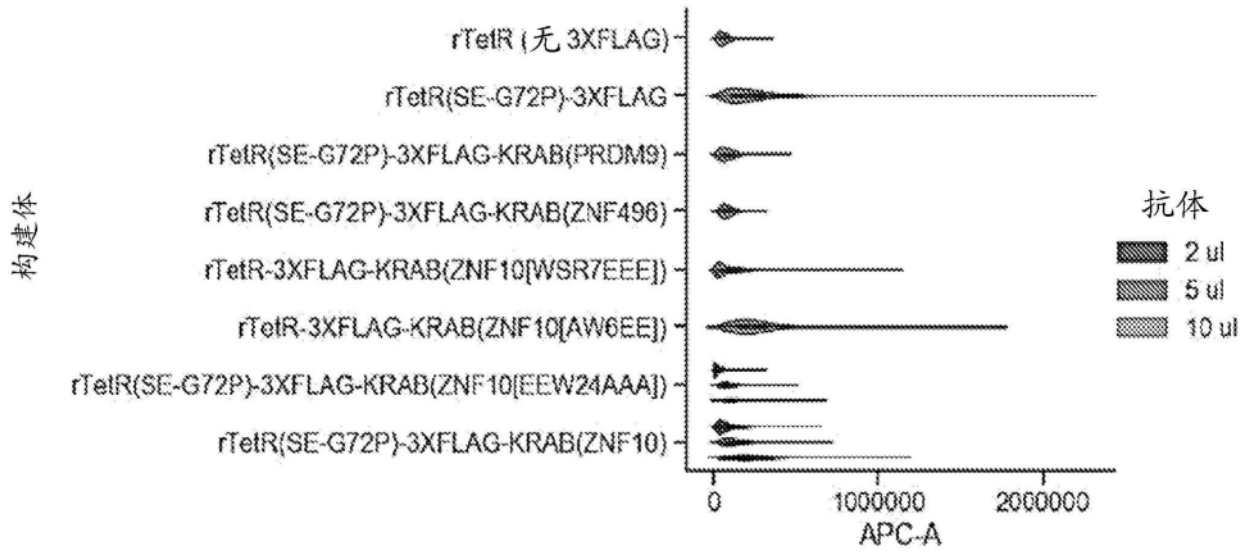

rTetR (无 3XFLAG)
rTetR(SE-G72P)-3XFLAG
rTetR(SE-G72P)-3XFLAG-KRAB(PRDM9)
rTetR(SE-G72P)-3XFLAG-KRAB(ZNF496)
rTetR-3XFLAG-KRAB(ZNF10[WSR7EEE])
rTetR-3XFLAG-KRAB(ZNF10[AW6EE])
rTetR(SE-G72P)-3XFLAG-KRAB(ZNF10[EEW24AAA])
rTetR(SE-G72P)-3XFLAG-KRAB(ZNF10)
构建体
抗体
2 ul
5 ul
10 ul
0
1000000
2000000
APC-A

图10B

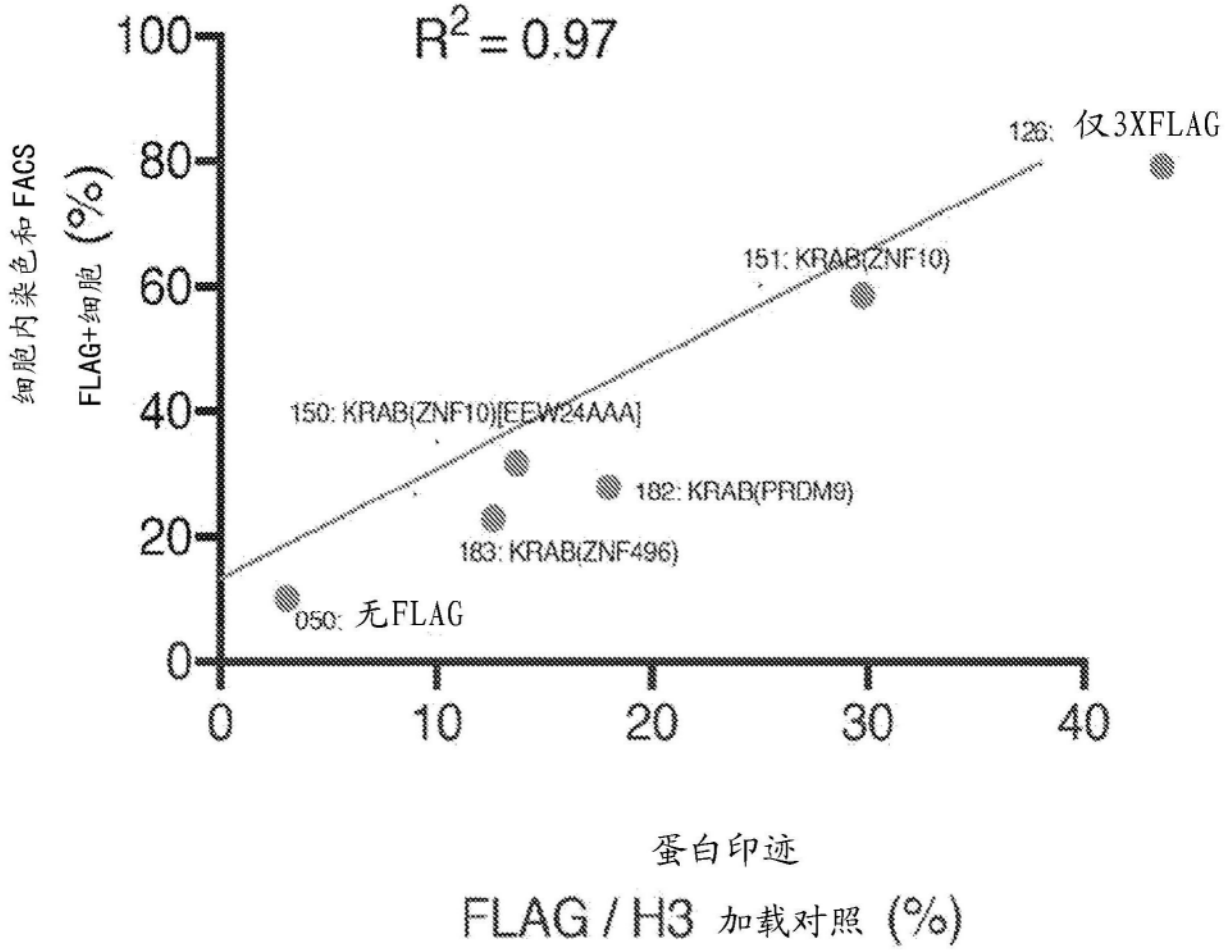

$R^2 = 0.97$
细胞内染色和 FACS
FLAG+细胞 (%)
126: 仅3XFLAG
151: KRAB(ZNF10)
150: KRAB(ZNF10)[EEW24AAA]
182: KRAB(PRDM9)
183: KRAB(ZNF496)
050: 无FLAG
0
20
40
60
80
100
0
10
20
30
40
蛋白印迹
FLAG / H3 加载对照 (%)

图10C

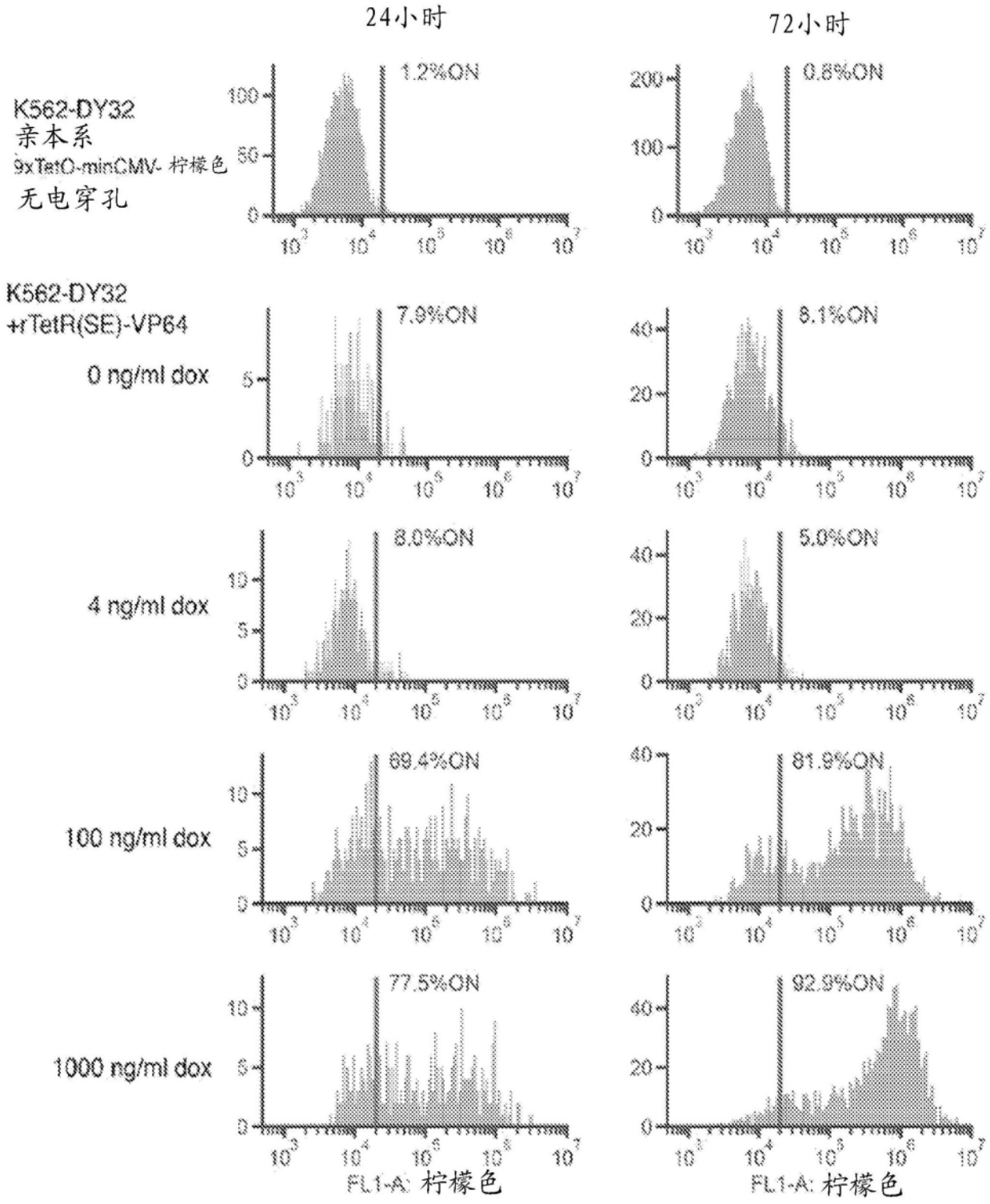

24小时
72小时
K562-DY32
亲本系
9xTetO-minCMV- 柠檬色
无电穿孔
1.2%ON
0.8%ON
K562-DY32
+rTetR(SE)-VP64
0 ng/ml dox
7.9%ON
8.1%ON
4 ng/ml dox
8.0%ON
5.0%ON
100 ng/ml dox
69.4%ON
81.9%ON
1000 ng/ml dox
77.5%ON
92.9%ON
FL1-A: 柠檬色
FL1-A: 柠檬色

图11A

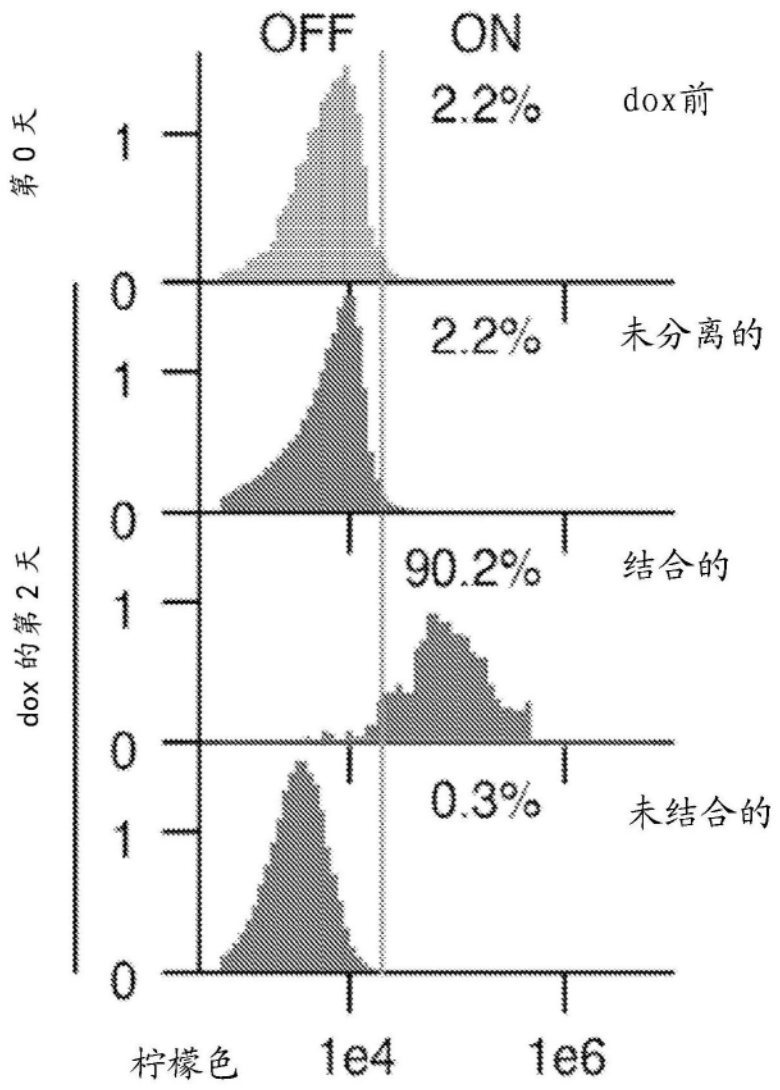

OFF
ON
2.2%
dox前
第0天
1
0
2.2%
未分离的
1
0
90.2%
结合的
dox的第2天
1
0
0.3%
未结合的
1
0
柠檬色
1e4
1e6

图11B

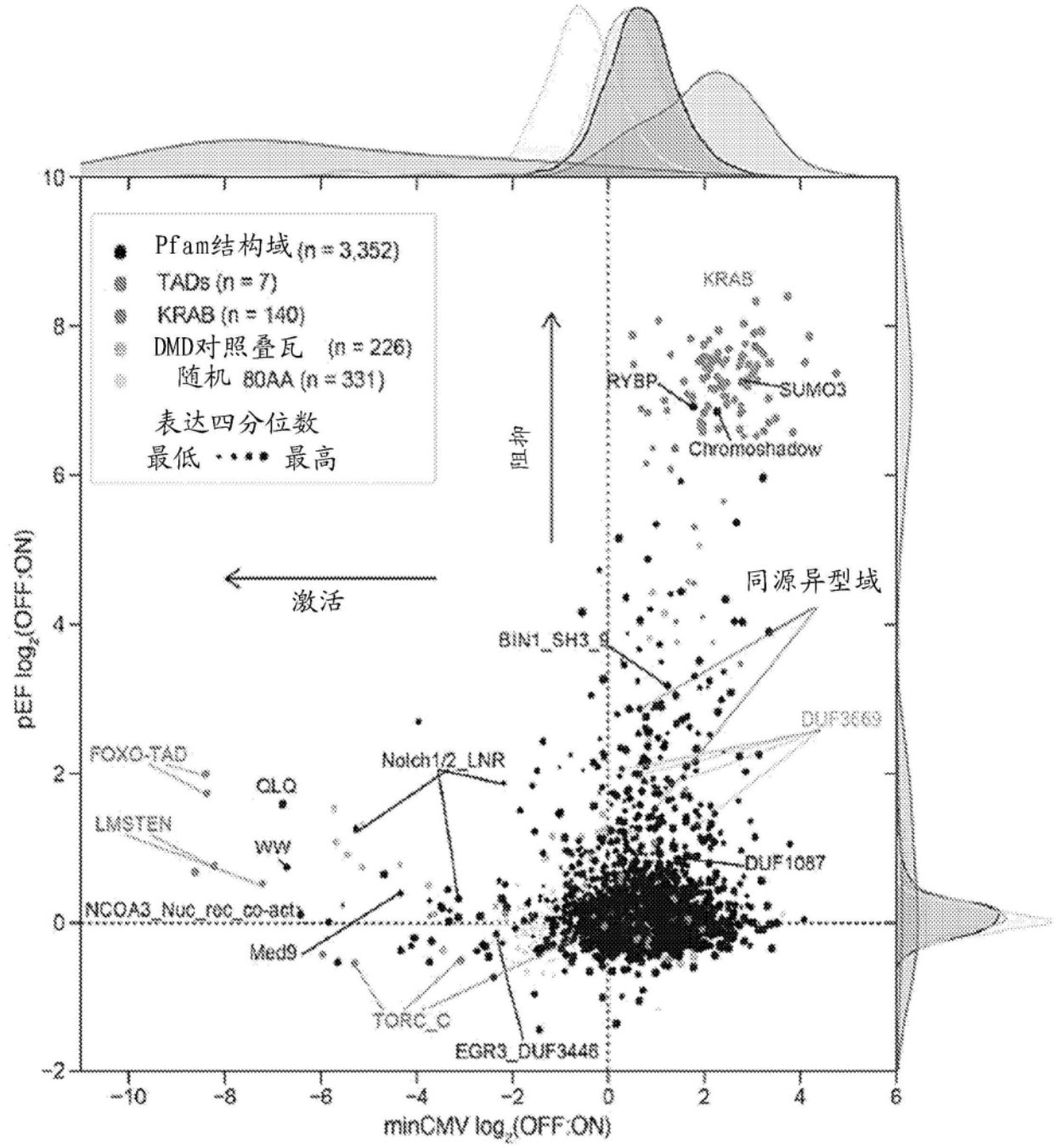

Pfam结构域 (n = 3,352)
TADs (n = 7)
KRAB (n = 140)
DMD对照叠瓦 (n = 226)
随机 80AA (n = 331)
表达四分位数
最低 最高
激活
抑制
同源异型域
KRAB
RYBP
SUMO3
Chromoshadow
BIN1_SH3_0
DUF3669
Notch1/2_LNR
FOXO-TAD
QLQ
LMSTEN
WW
DUF1087
NCOA3_Nuc_rec_co-act
Med9
TORC_C
EGR3_DUF3448
pEF log2(OFF:ON)
minCMV log2(OFF:ON)

图11C

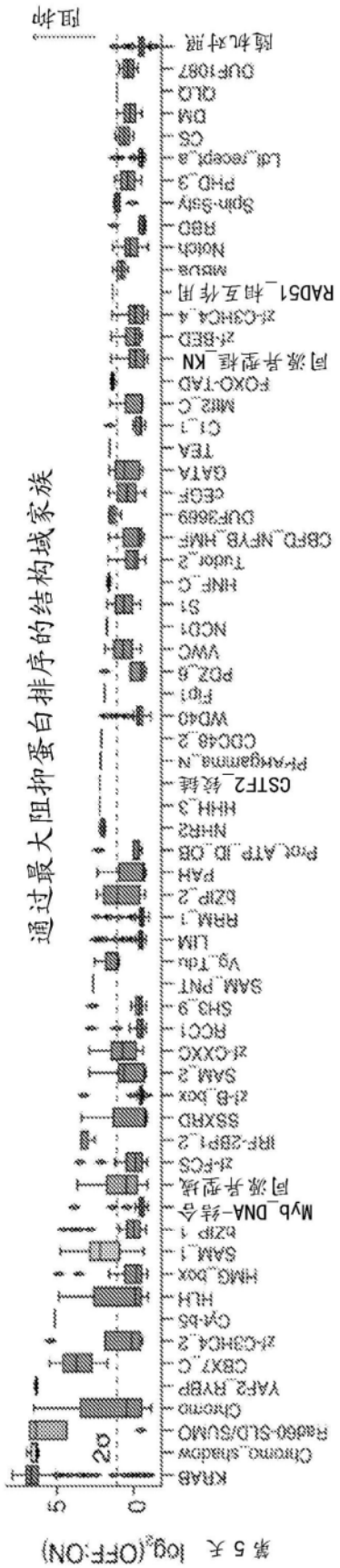

通过最大阻抑蛋白排序的结构域家族
第 5 天 log2(OFF:ON)
5
0
2σ
阻抗
随机对照
OUF1087
CLD
DM
CS
Lbl_Receptor_a
Spin-Sstv
PHD_3
RBD
Notch
mbus
RAD51_相互作用
zf-C3HC4_4
zf-BED
同源异型框_KN
FOXO-TAD
MH2_C
C1_1
TEA
GATA
cECF
OUF3669
CBFD_NFYB_HMF
Tudor_2
HNF_C
S1
NCD1
VWC
PDZ_6
Fip1
WD40
CDC48_2
PF-AMgamma_N
GSTF2_铰链
HHH_3
NHR2
Prot_ATP_ID_OB
PAH
bZIP_2
RRM_1
LIM
Vg_Tdu
SAM_PNT
SH3_9
RCC1
zf-CXXC
SAM_2
SSXRD
IRF-2BP1_2
zf-FCS
同源异型域
Myb_DNA-结合
bZIP_1
SAM_1
HMG_box
HLH
Cyt-b5
zf-C3HC4_2
CBX7_C
YAF2_RYBP
Chromo
Ras60-SLD/SUMO
Chromo_shadow
KRAB

图12A

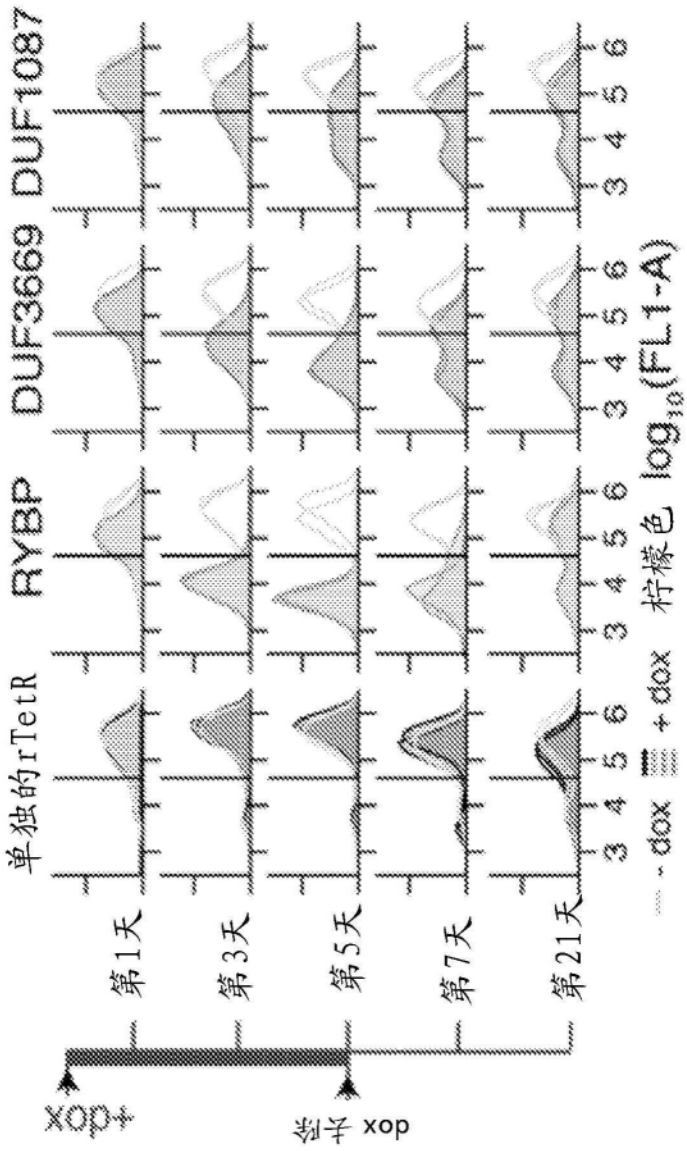

单独的rTetR
RYBP
DUF3669
DUF1087
第1天
第3天
第5天
第7天
第21天
- dox
+ dox
柠檬色 log10(FL1-A)
3 4 5 6
+dox 存在
dox 不存在

图12B

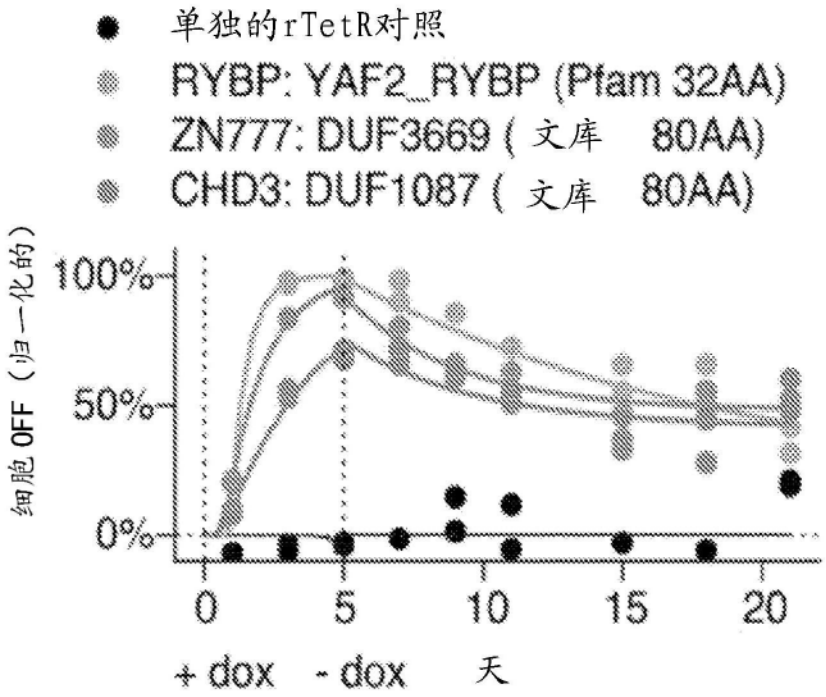

单独的rTetR对照
RYBP: YAF2_RYBP (Pfam 32AA)
ZN777: DUF3669 (文库 80AA)
CHD3: DUF1087 (文库 80AA)
细胞 OFF（归一化的）
100%
50%
0%
0
5
10
15
20
+ dox
- dox
天

图12C

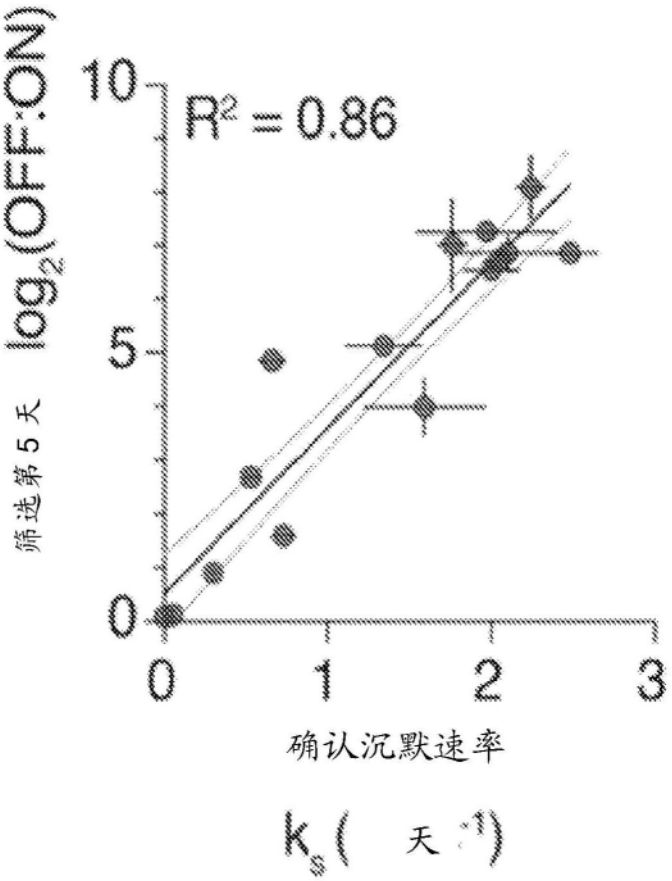

筛选第5天 log2(OFF:ON)
10
5
0
R² = 0.86
0
1
2
3
确认沉默速率
ks(天-1)

图12D

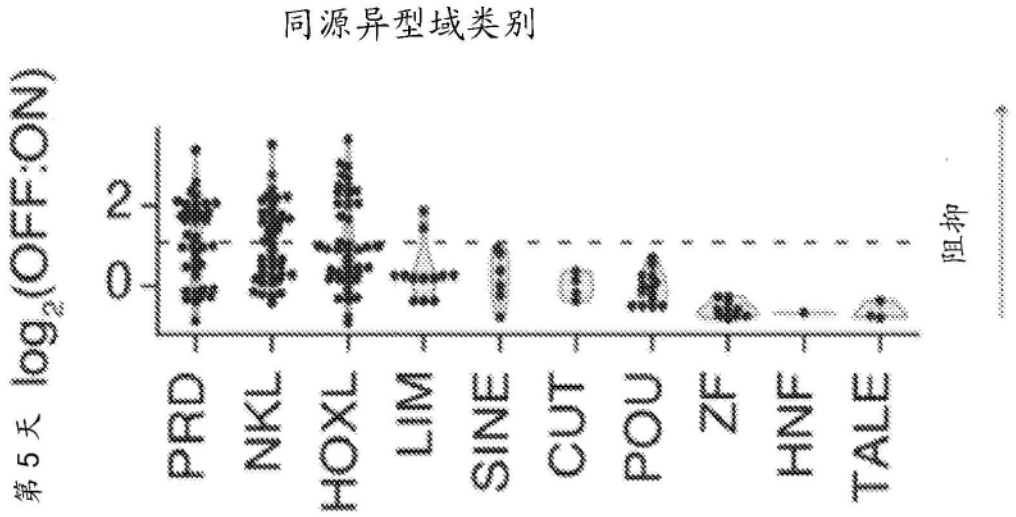

同源异型域类别
第5天 log2(OFF:ON)
2
0
阻抑
PRD
NKL
HOXL
LIM
SINE
CUT
POU
ZF
HNF
TALE

图13A

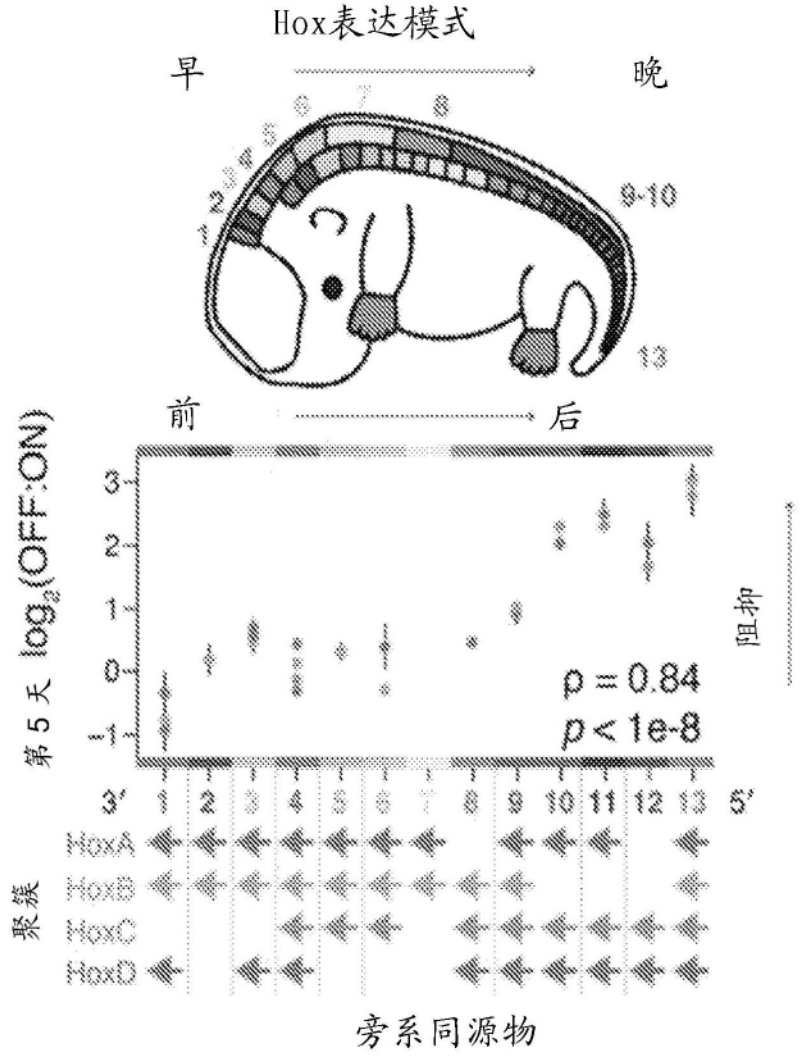

Hox表达模式
早
晚
1
2
3
4
5
6
7
8
9-10
13
前
后
第5天 log2(OFF:ON)
3
2
1
0
-1
阻抑
p = 0.84
p < 1e-8
3'
1
2
3
4
5
6
7
8
9
10
11
12
13
5'
聚簇
HoxA
HoxB
HoxC
HoxD
旁系同源物

图13B

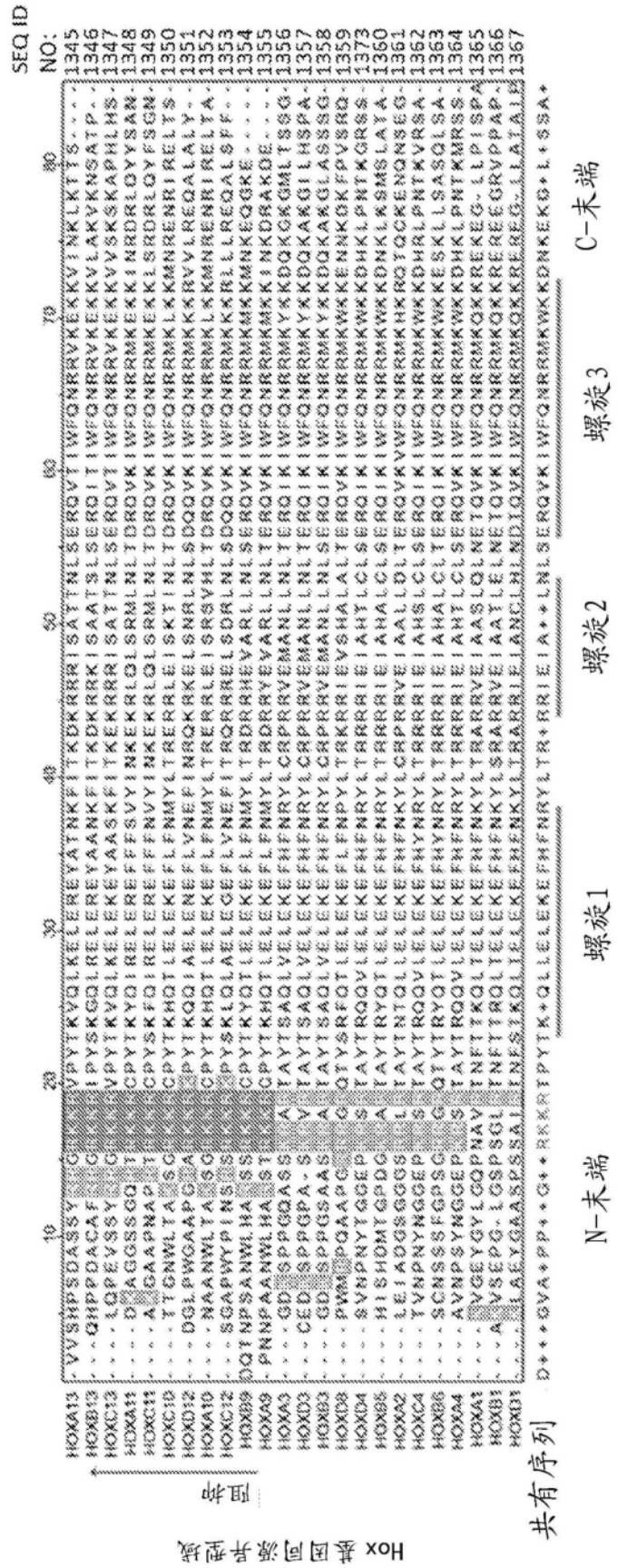


图13C

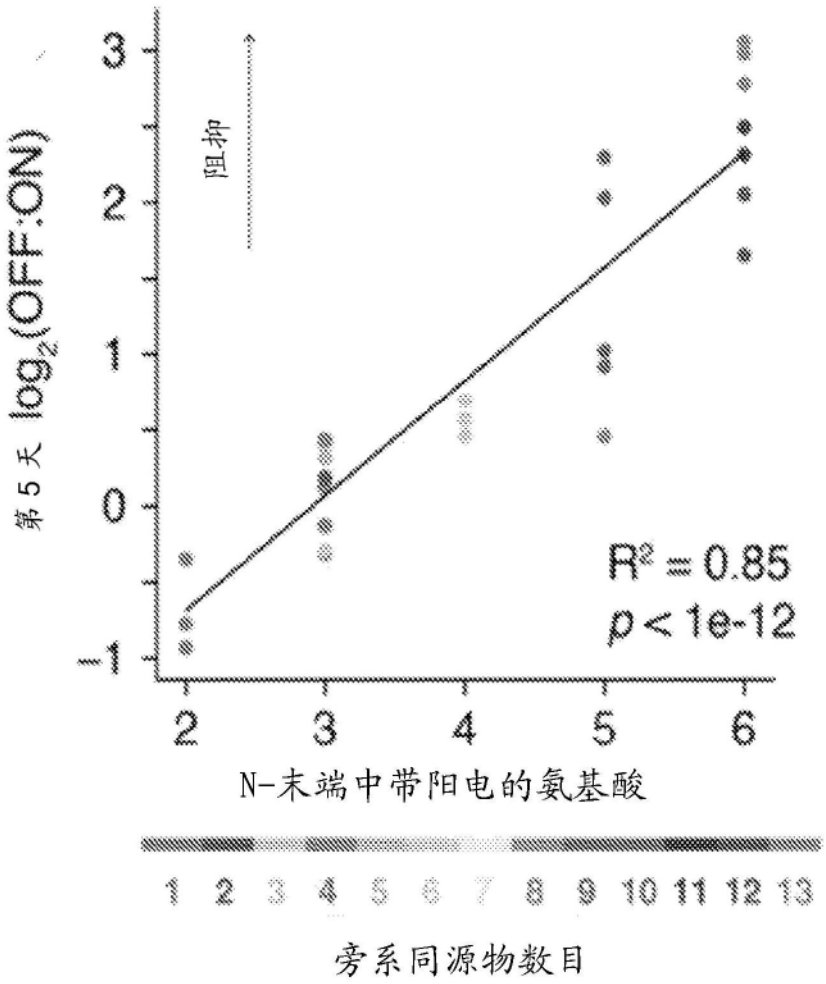

阻抑
第5天 log2(OFF:ON)
3
2
1
0
-1
R² = 0.85
p < 1e-12
2
3
4
5
6
N-末端中带阳电的氨基酸
1 2 3 4 5 6 7 8 9 10 11 12 13
旁系同源物数目

图13D

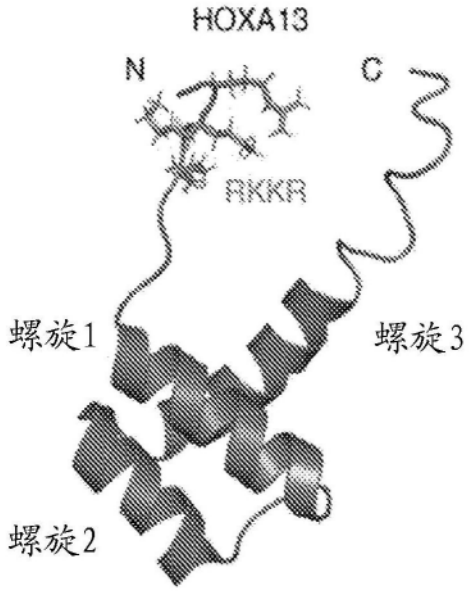

HOXA13
N
C
RKKR
螺旋1
螺旋3
螺旋2

图13E

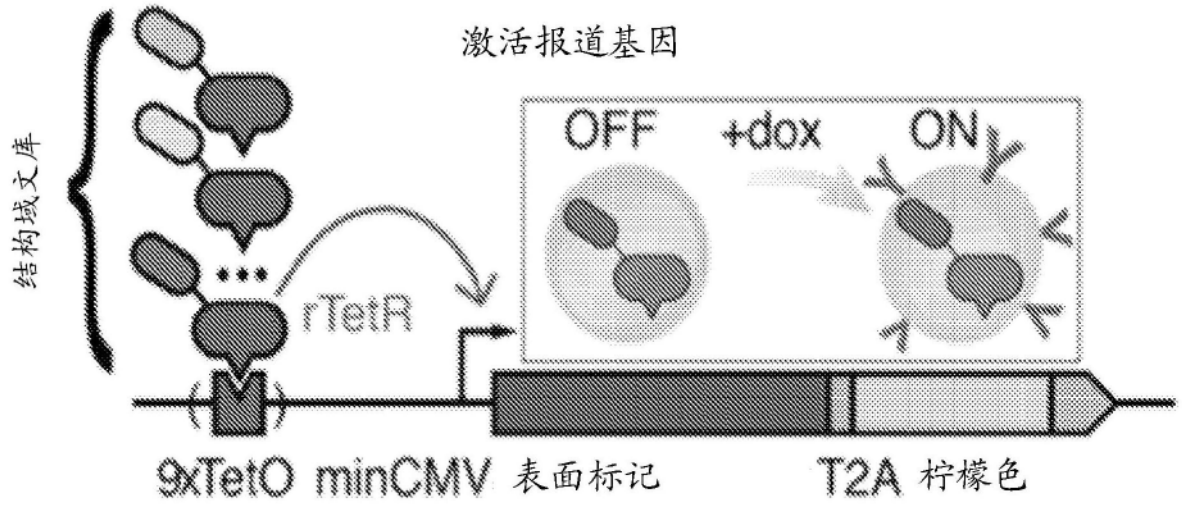

激活报道基因
结构域文库
OFF
+dox
ON
rTetR
9xTetO minCMV
表面标记
T2A
柠檬色

图14A

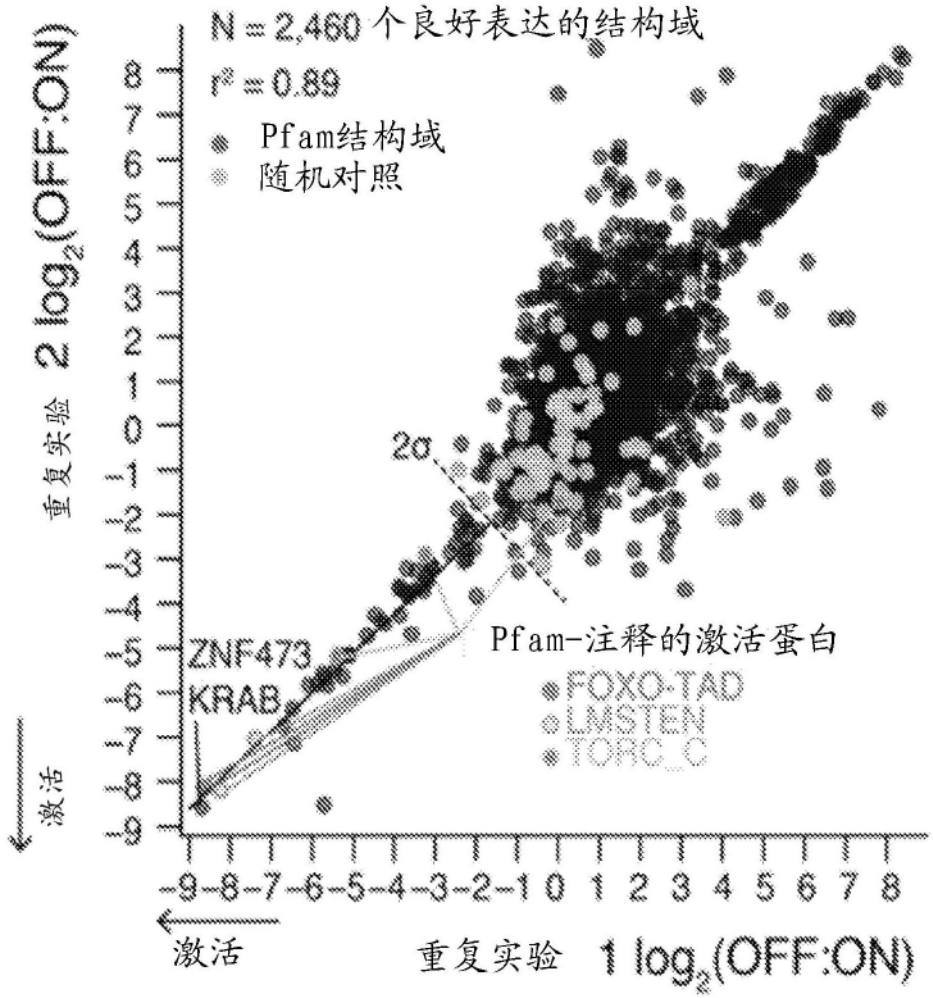

N = 2,460 个良好表达的结构域
r² = 0.89
Pfam结构域
随机对照
2σ
ZNF473
KRAB
Pfam-注释的激活蛋白
FOXO-TAD
LMSTEN
TORC_C
重复实验 2 log₂(OFF:ON)
激活
重复实验 1 log₂(OFF:ON)
激活

图14B

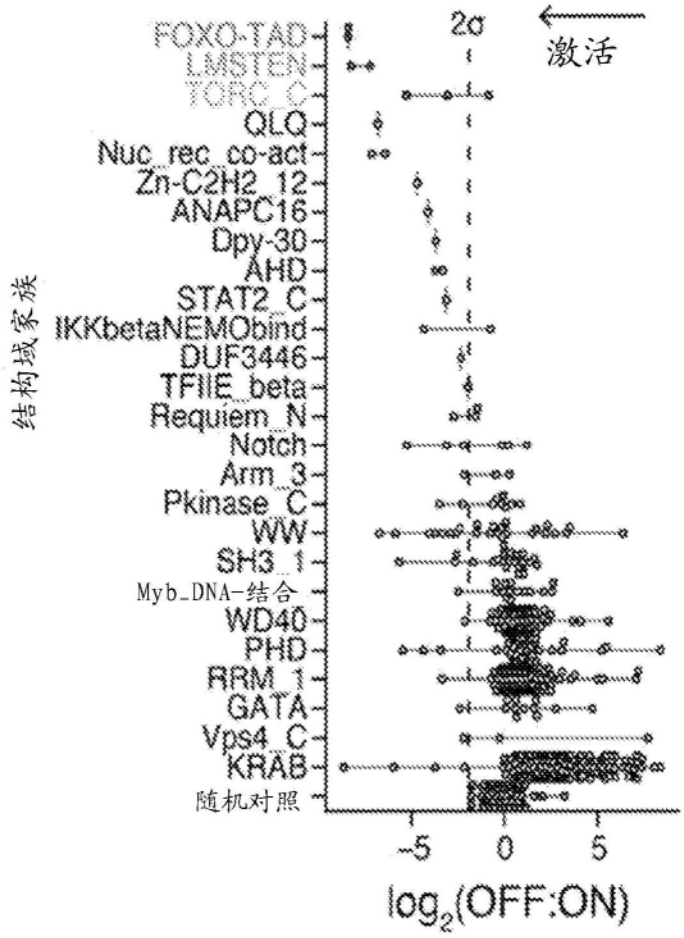

2σ
激活
FOXO-TAD
LMSTEN
TORC_C
QLQ
Nuc_rec_co-act
Zn-C2H2_12
ANAPC16
Dpy-30
AHD
STAT2_C
IKKbetaNEMObind
DUF3446
TFIIE_beta
Requiem_N
Notch
Arm_3
Pkinase_C
WW
SH3_1
Myb_DNA-结合
WD40
PHD
RRM_1
GATA
Vps4_C
KRAB
随机对照
结构域家族
-5
0
5
log2(OFF:ON)

图14C

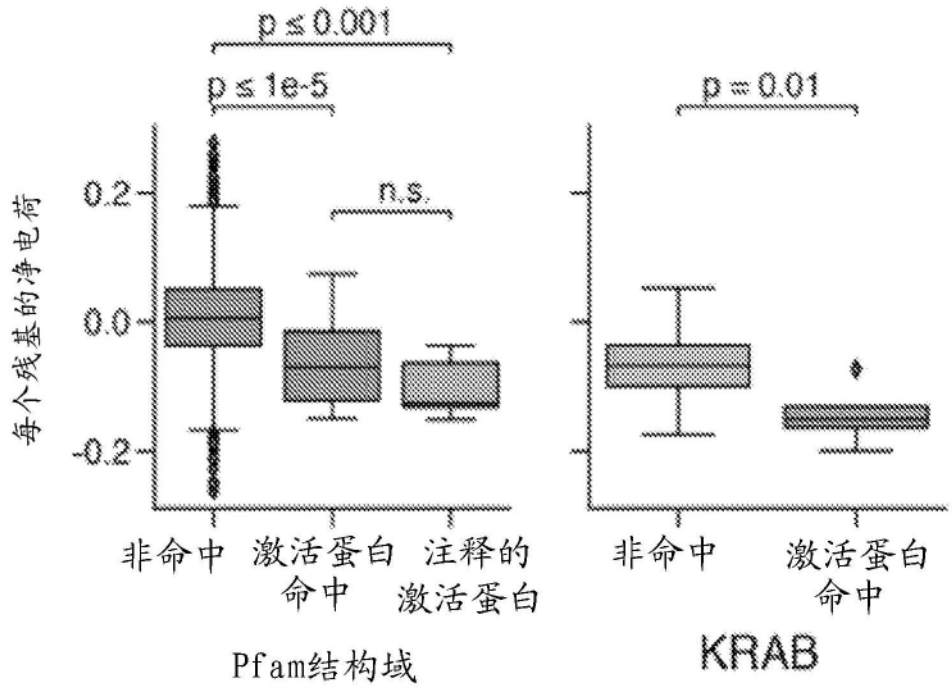

p ≤ 0.001
p ≤ 1e-5
n.s.
p = 0.01
每个残基的净电荷
0.2
0.0
-0.2
非命中
激活蛋白命中
注释的激活蛋白
非命中
激活蛋白命中
Pfam结构域
KRAB

图14D

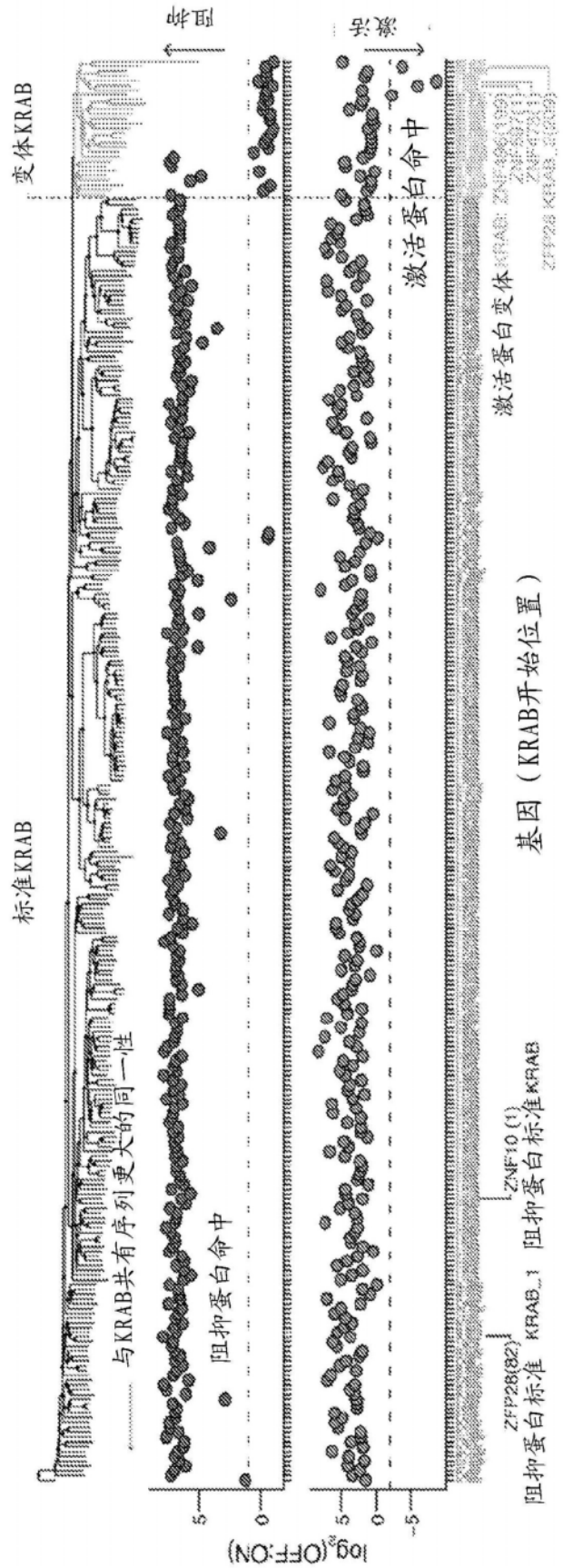

阻抑
激活
变体KRAB
标准KRAB
与KRAB共有序列更大的同一性
阻抑蛋白命中
激活蛋白命中
log2(OFF:ON)
5
0
-5
ZFP28(82)
ZNF10 (1)
阻抑蛋白标准 KRAB_1 阻抑蛋白标准KRAB
激活蛋白变体
基因（KRAB开始位置）

图14E

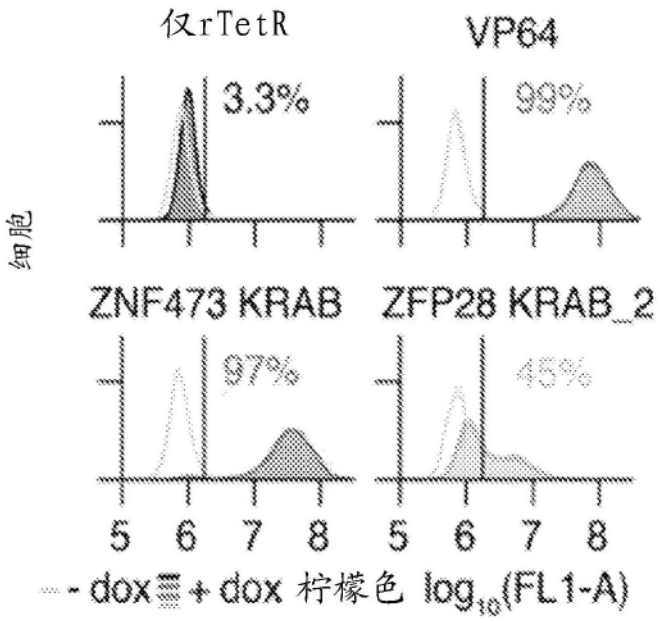

仅rTetR
VP64
3.3%
99%
细胞
ZNF473 KRAB
ZFP28 KRAB_2
97%
45%
5 6 7 8
5 6 7 8
- dox
+ dox
柠檬色 log10(FL1-A)

图14F

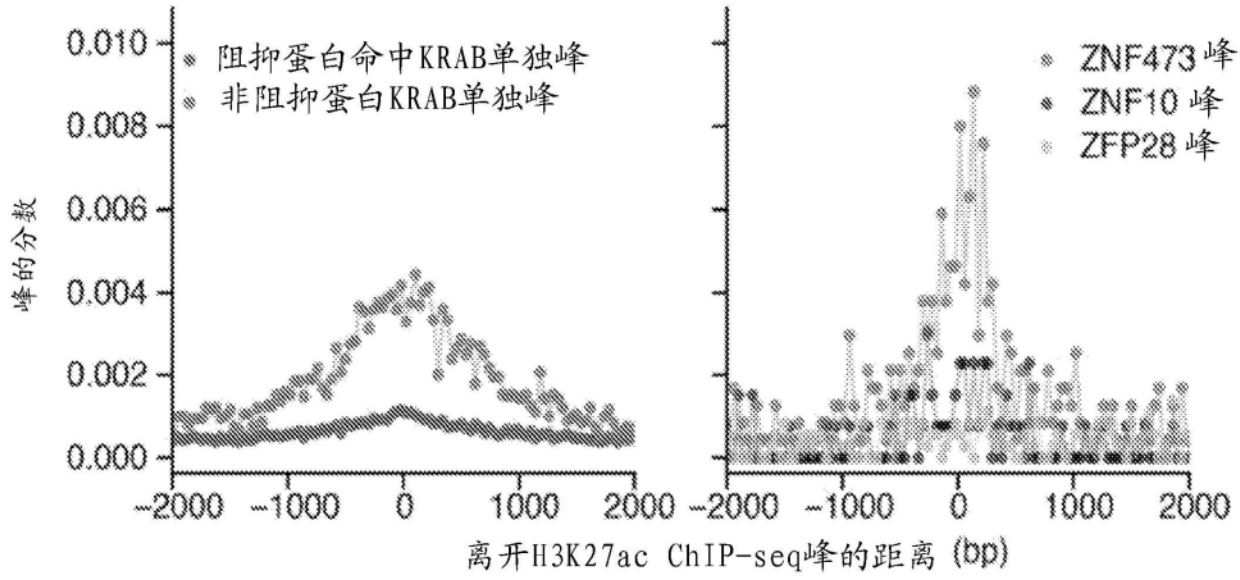

0.010
0.008
0.006
0.004
0.002
0.000
峰的分数
阻抑蛋白命中KRAB单独峰
非阻抑蛋白KRAB单独峰
ZNF473峰
ZNF10峰
ZFP28峰
-2000 -1000 0 1000 2000
-2000 -1000 0 1000 2000
离开H3K27ac ChIP-seq峰的距离 (bp)

图14G

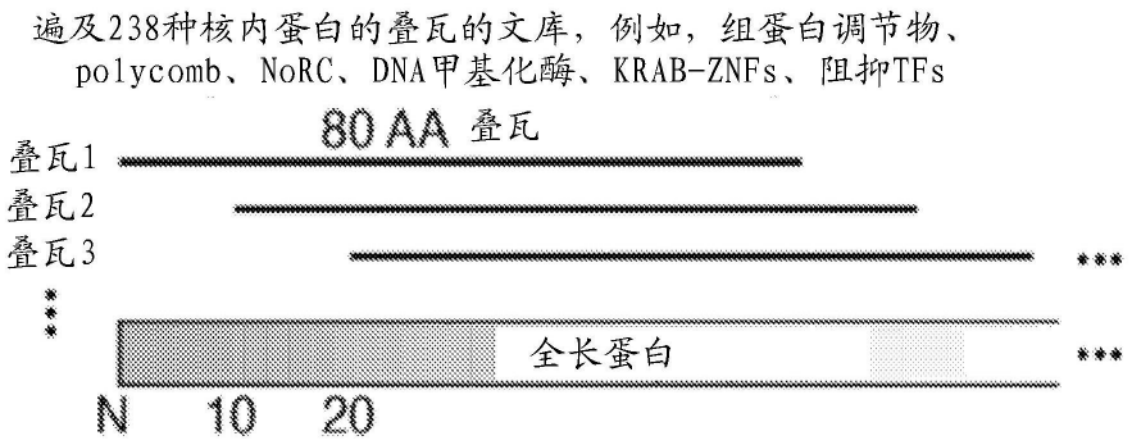

遍及238种核内蛋白的叠瓦的文库，例如，组蛋白调节物、
polycomb、NoRC、DNA甲基化酶、KRAB-ZNFs、阻抑TFs
80 AA 叠瓦
叠瓦1
叠瓦2
叠瓦3
全长蛋白
N
10
20

图15A

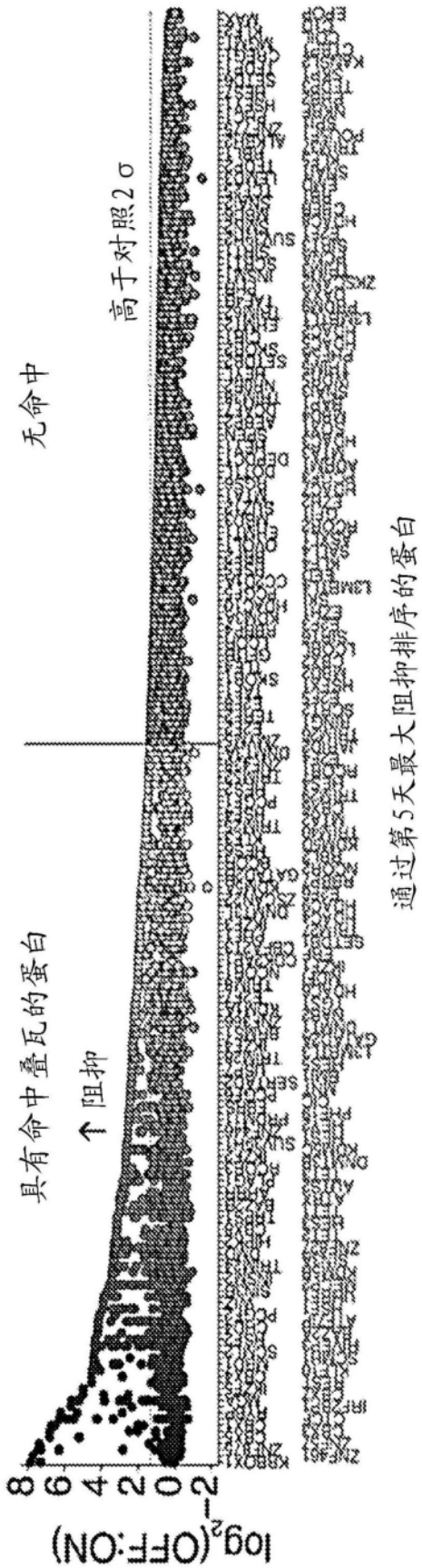


图15B

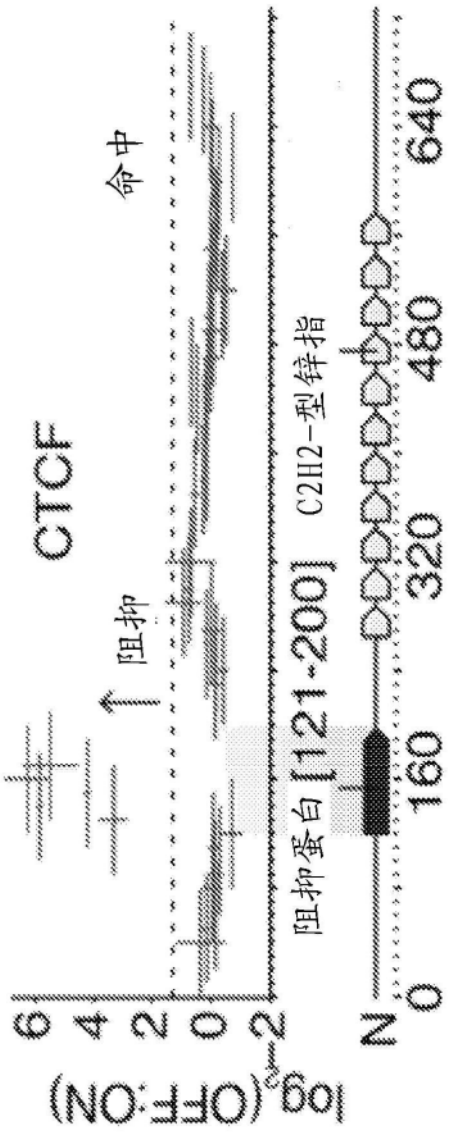

CTCF
命中
阻抑
阻抑蛋白[1121-200] C2H2-型锌指
N
0
160
320
480
640
6
4
2
0
2
log2(OFF:ON)

图15C

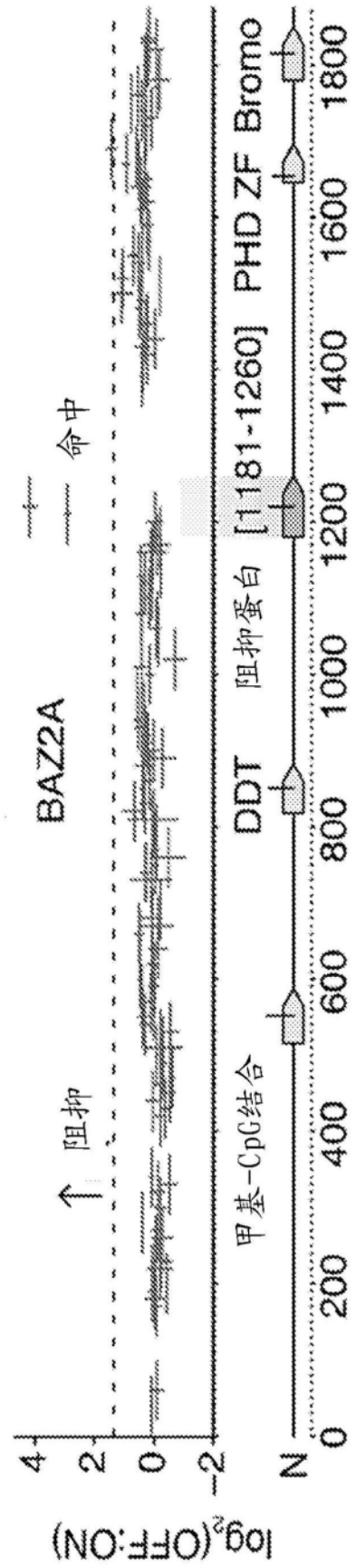

log2(OFF:ON)
4
2
0
−2
BAZ2A
命中
阻抑
甲基-CpG结合
DDT
阻抑蛋白
[1181-1260]
PHD ZF
Bromo
N
0
200
400
600
800
1000
1200
1400
1600
1800

图15D

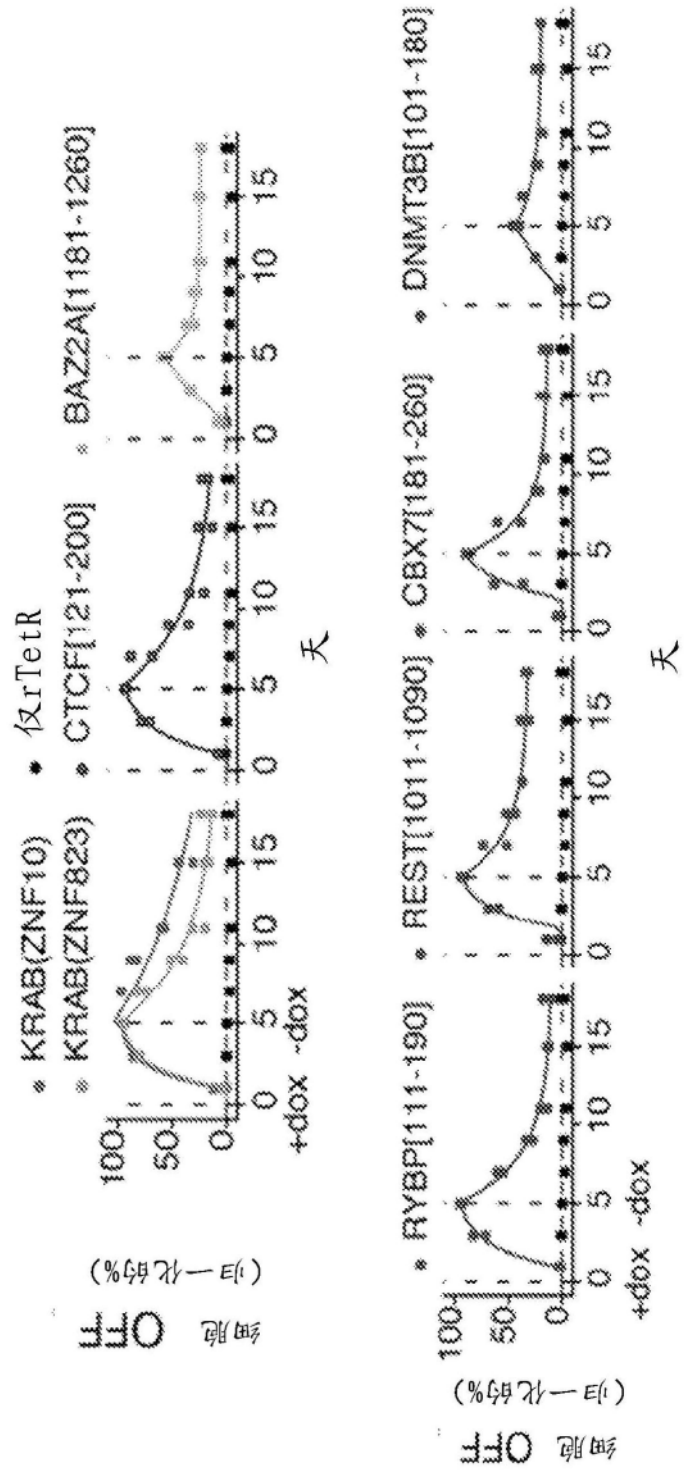

KRAB(ZNF10)
KRAB(ZNF823)
仅rTetR
CTCF[121-200]
BAZ2A[1181-1260]
RYBP[111-190]
REST[1011-1090]
CBX7[181-260]
DNMT3B[101-180]
+dox
-dox
天
细胞 OFF (归一化的%)
100
50
0
0 5 10 15

图15E

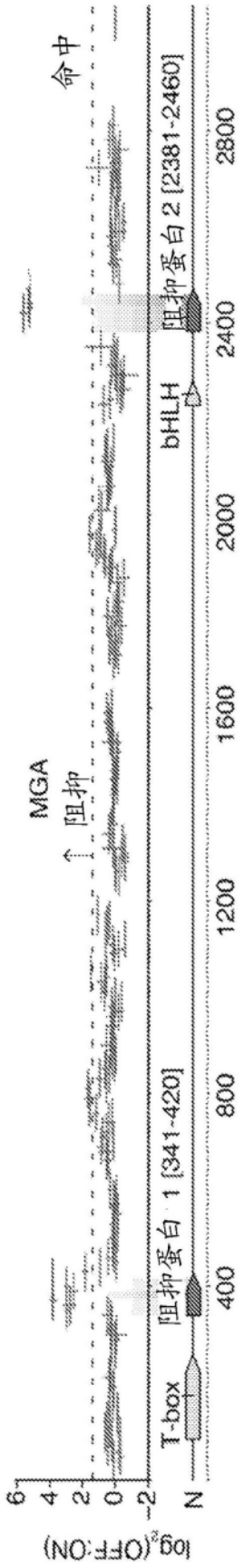

命中
MGA
阻抑
T-box
阻抑蛋白 1 [341-420]
bHLH
阻抑蛋白 2 [2381-2460]
N
400
800
1200
1600
2000
2400
2800
-2
0
2
4
6
log2(OFF:ON)

图15F

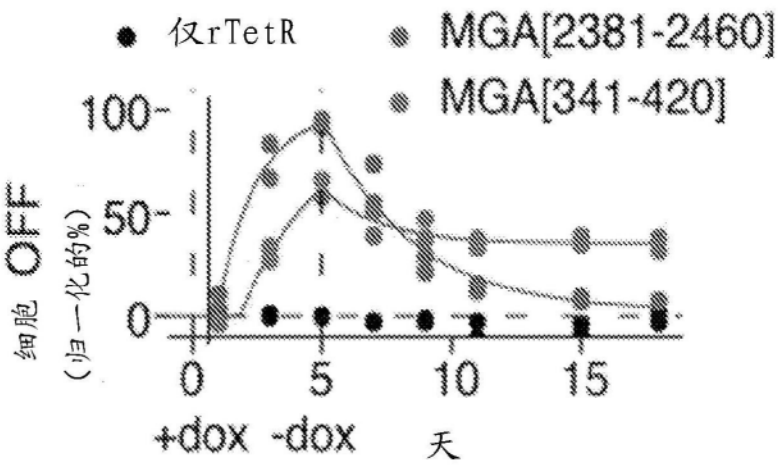

仅rTetR
MGA[2381-2460]
MGA[341-420]
100
50
0
细胞 OFF (归一化的%)
0
5
10
15
+dox -dox
天

图15G

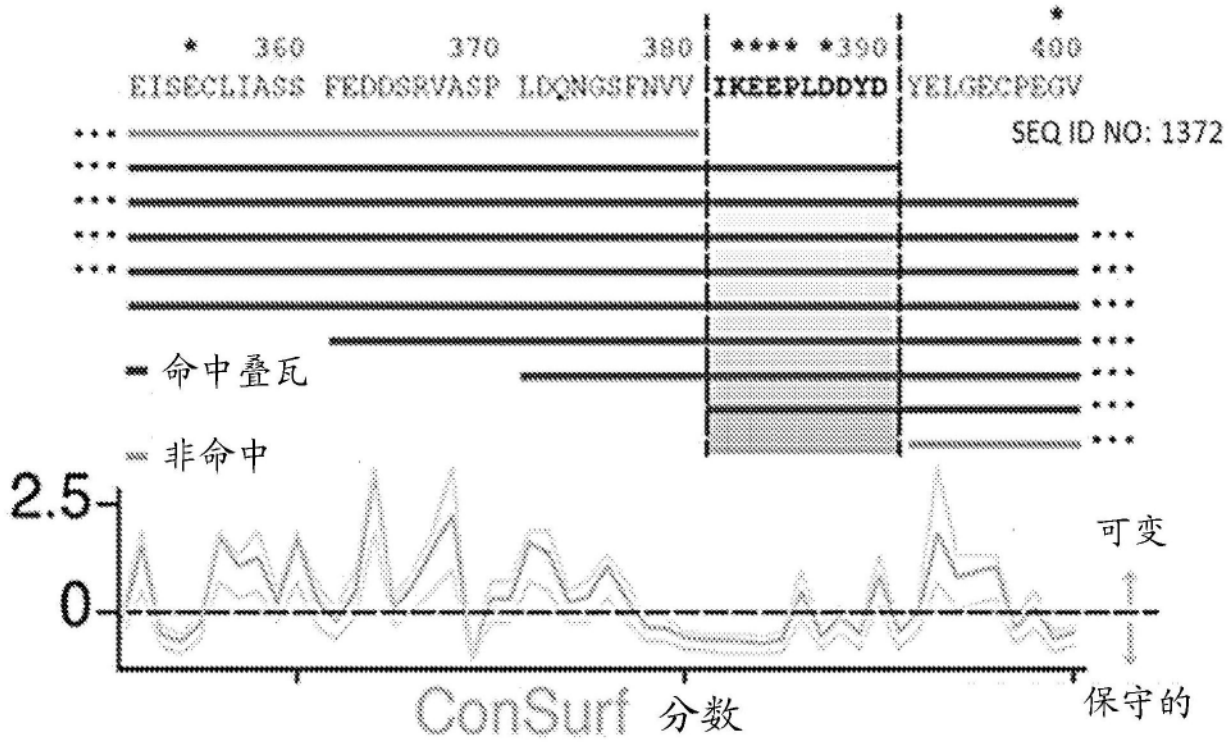

360
370
380
390
400
EISECLIASS FEDQSRVASP LDQNGSFNVV IKEEPLDDYD YELGECPEGV
SEQ ID NO: 1372
命中叠瓦
非命中
2.5
0
可变
保守的
ConSurf 分数

图15H

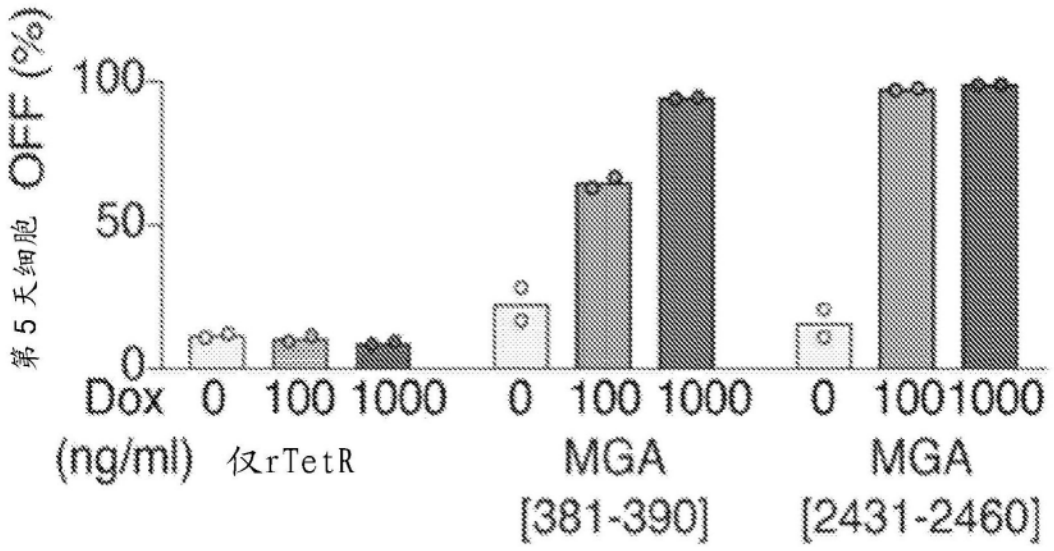

第5天细胞 OFF (%)
100
50
0
Dox 0 100 1000 0 100 1000 0 100 1000
(ng/ml)
仅rTetR
MGA [381-390]
MGA [2431-2460]

图15I

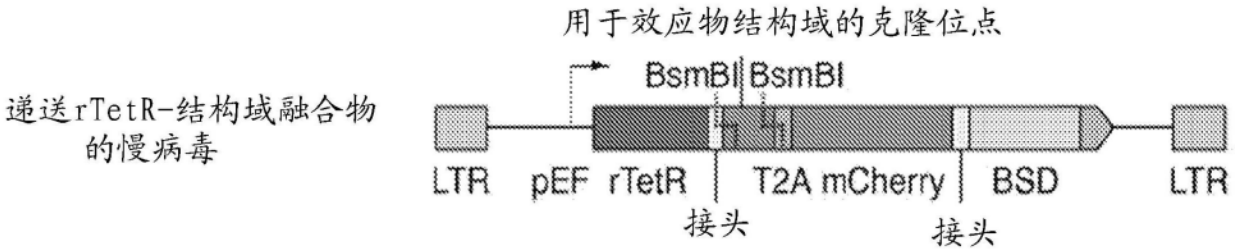

用于效应物结构域的克隆位点
递送rTetR-结构域融合物的慢病毒
BsmBI BsmBI
LTR pEF rTetR T2A mCherry BSD LTR
接头
接头

图16A

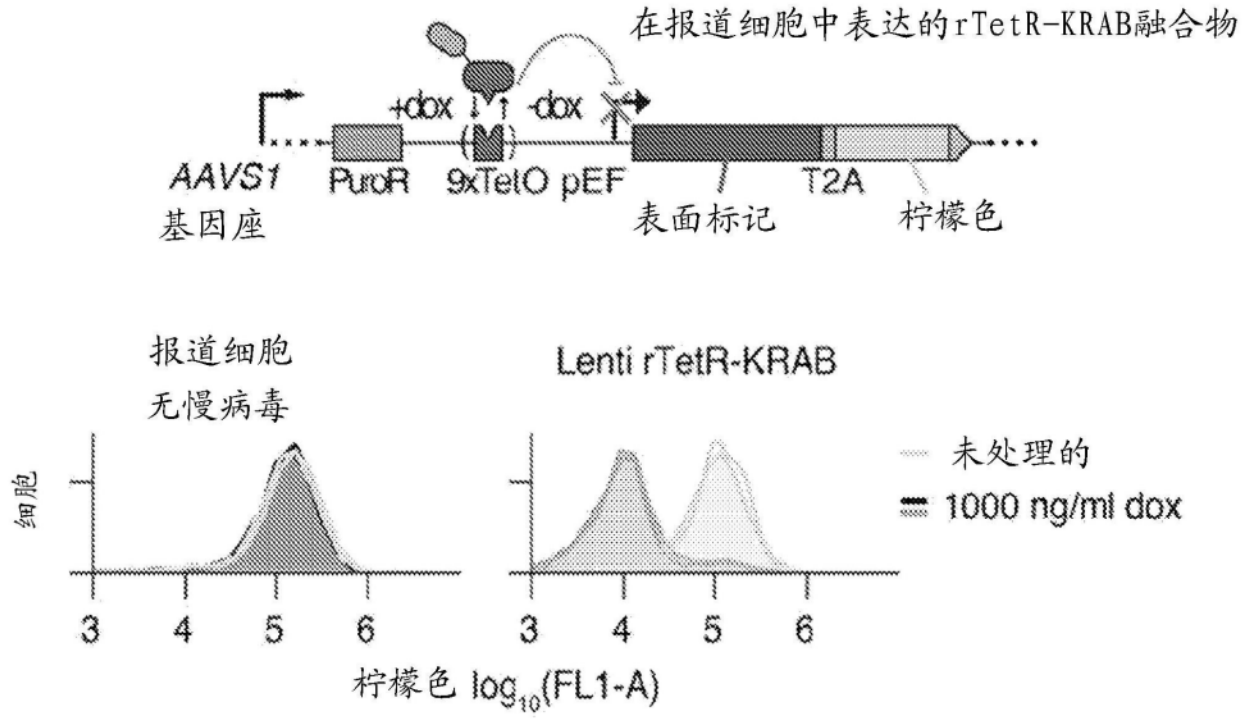

在报道细胞中表达的rTetR-KRAB融合物
+dox
-dox
AAVS1
基因座
PuroR
9xTetO
pEF
表面标记
T2A
柠檬色
报道细胞
无慢病毒
Lenti rTetR-KRAB
细胞
未处理的
1000 ng/ml dox
3
4
5
6
柠檬色 log10(FL1-A)

图16B

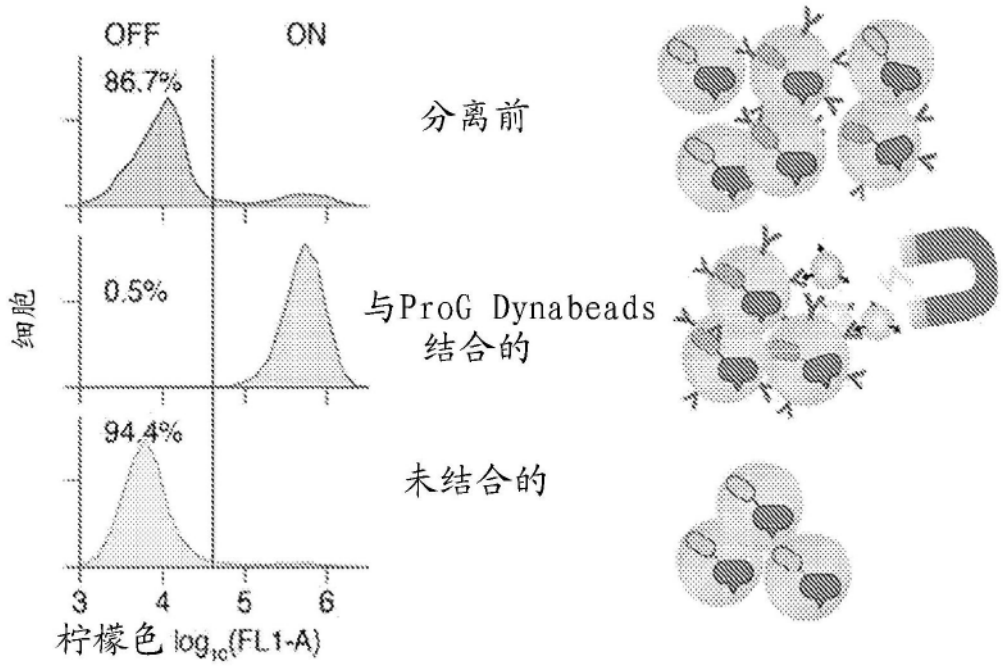

OFF
ON
86.7%
0.5%
94.4%
细胞
分离前
与ProG Dynabeads
结合的
未结合的
3
4
5
6
柠檬色 log10(FL1-A)

图16C

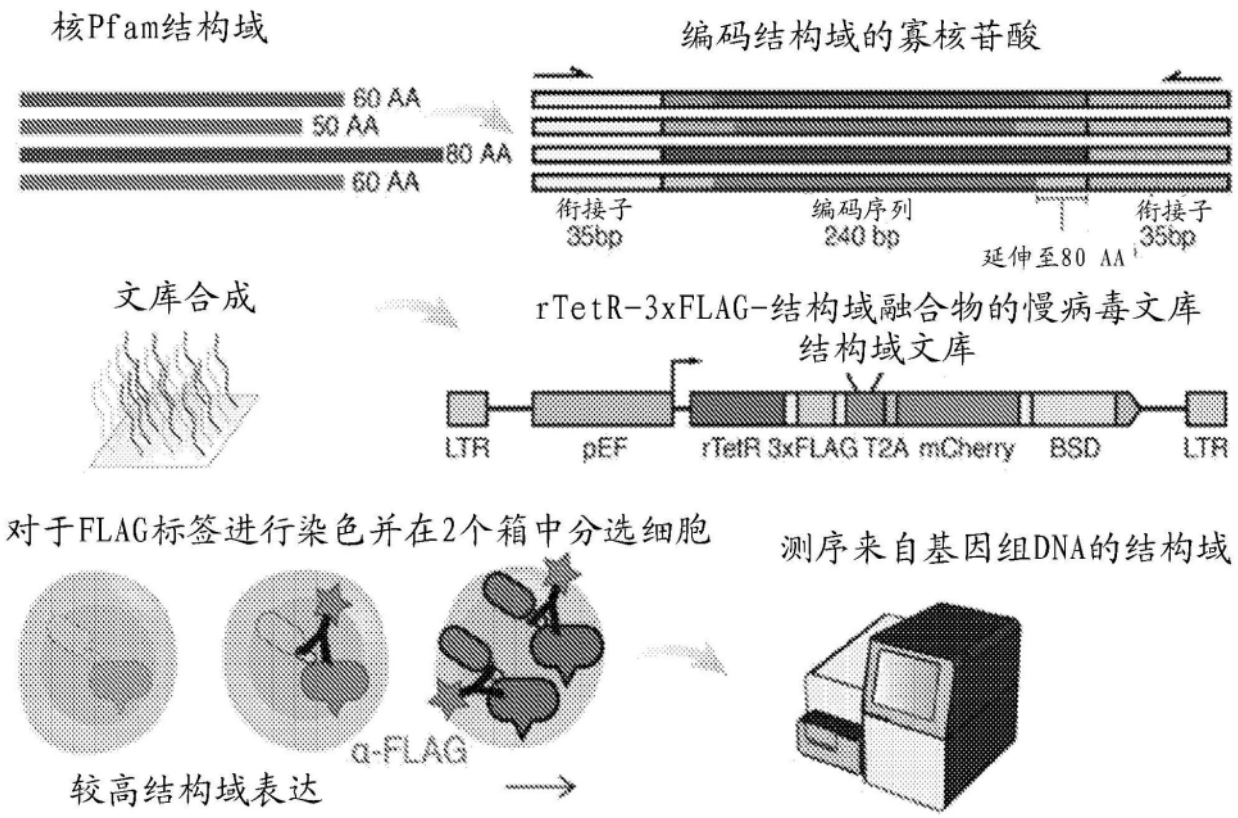

核Pfam结构域
60 AA
50 AA
80 AA
60 AA
编码结构域的寡核苷酸
衔接子
35bp
编码序列
240 bp
延伸至80 AA
衔接子
35bp
文库合成
rTetR-3xFLAG-结构域融合物的慢病毒文库
结构域文库
LTR
pEF
rTetR 3xFLAG T2A mCherry
BSD
LTR
对于FLAG标签进行染色并在2个箱中分选细胞
α-FLAG
较高结构域表达
测序来自基因组DNA的结构域

图17A

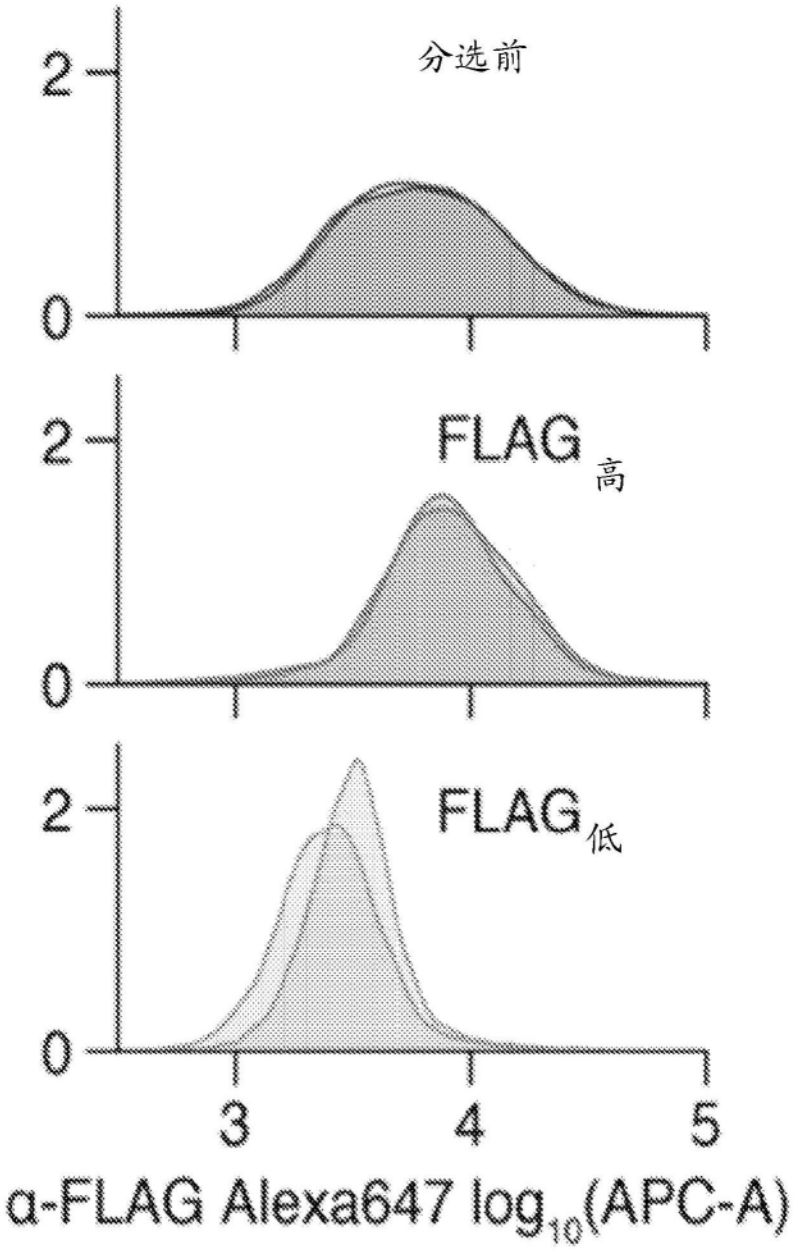


图17B

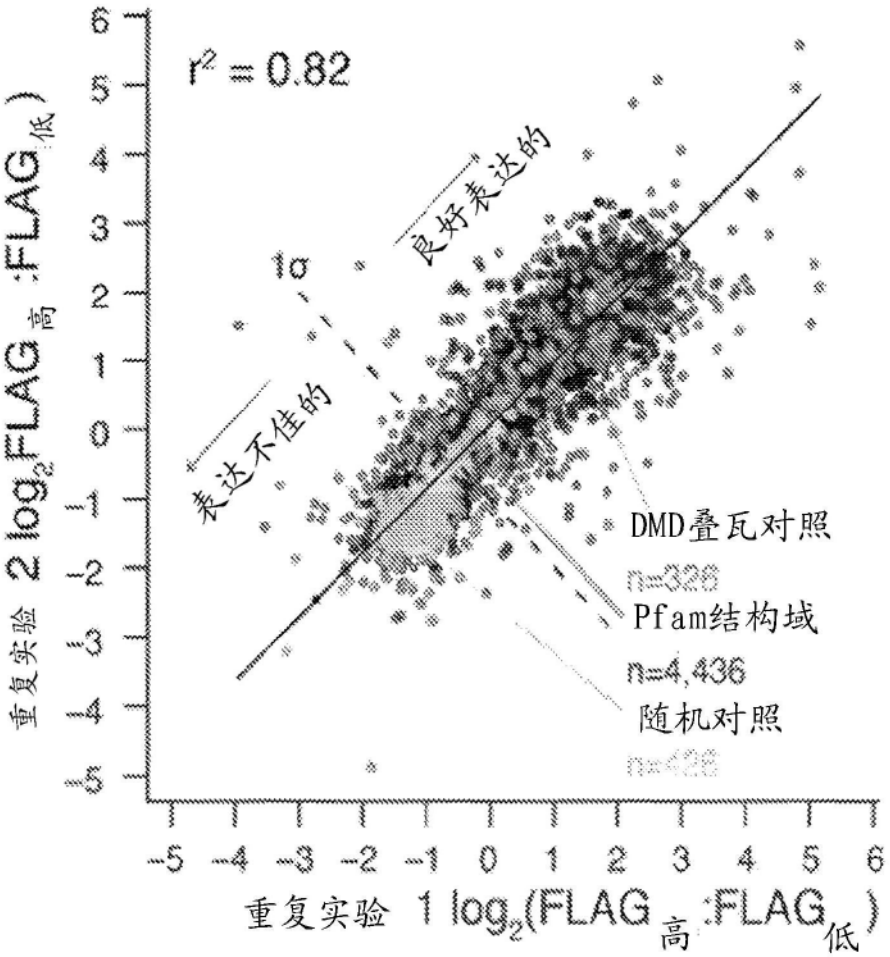

$r^2 = 0.82$
1σ
良好表达的
表达不佳的
DMD叠瓦对照
n=328
Pfam结构域
n=4,436
随机对照
n=426
重复实验 2 $\log_2$(FLAG高:FLAG低)
重复实验 1 $\log_2$(FLAG高:FLAG低)

图17C

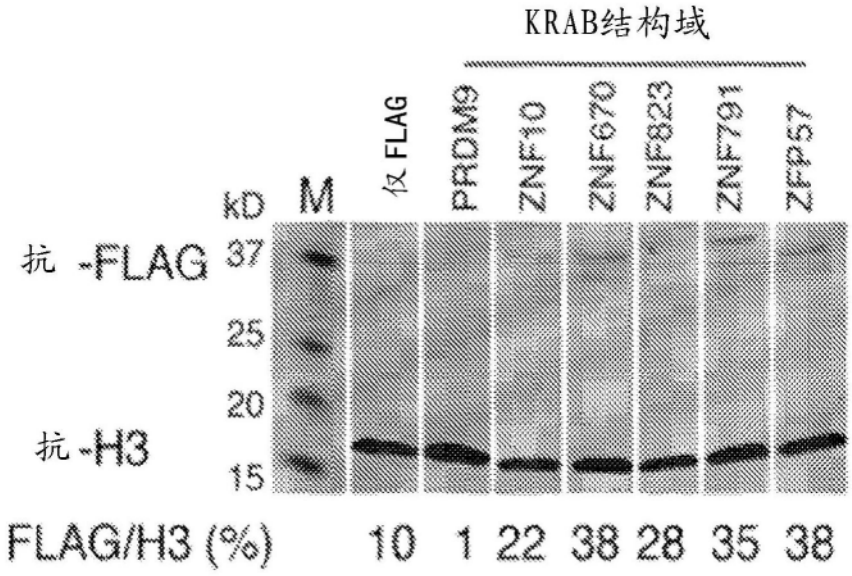

KRAB结构域
kD
M
仅 FLAG
PRDM9
ZNF10
ZNF670
ZNF823
ZNF761
ZFP57
抗 -FLAG
37
25
20
抗 -H3
15
FLAG/H3 (%)
10 1 22 38 28 35 38

图17D

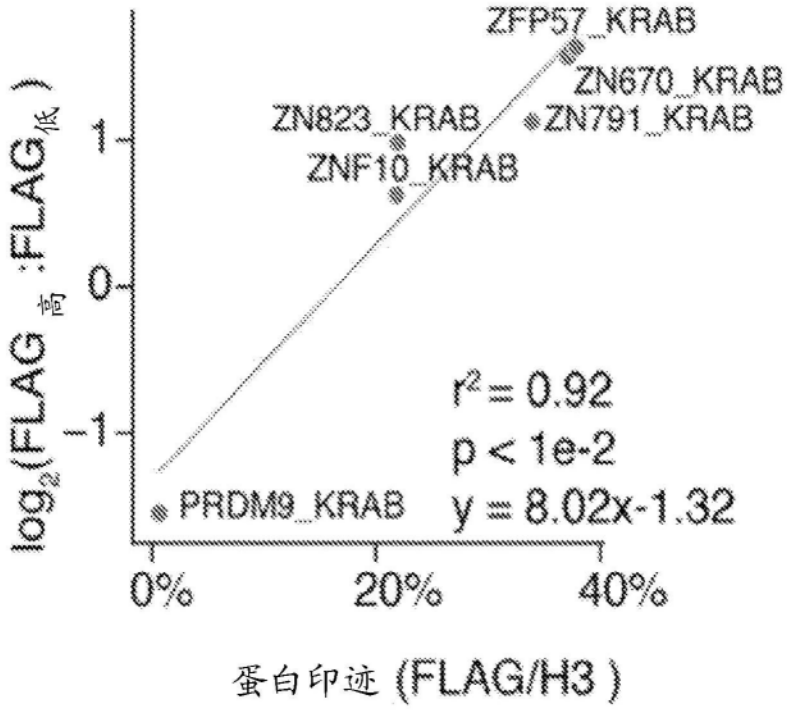

ZFP57_KRAB
ZN670_KRAB
ZN823_KRAB
ZN791_KRAB
ZNF10_KRAB
PRDM9_KRAB
r² = 0.92
p < 1e-2
y = 8.02x-1.32
log₂(FLAG高:FLAG低)
1
0
-1
0%
20%
40%
蛋白印迹 (FLAG/H3)

图17E

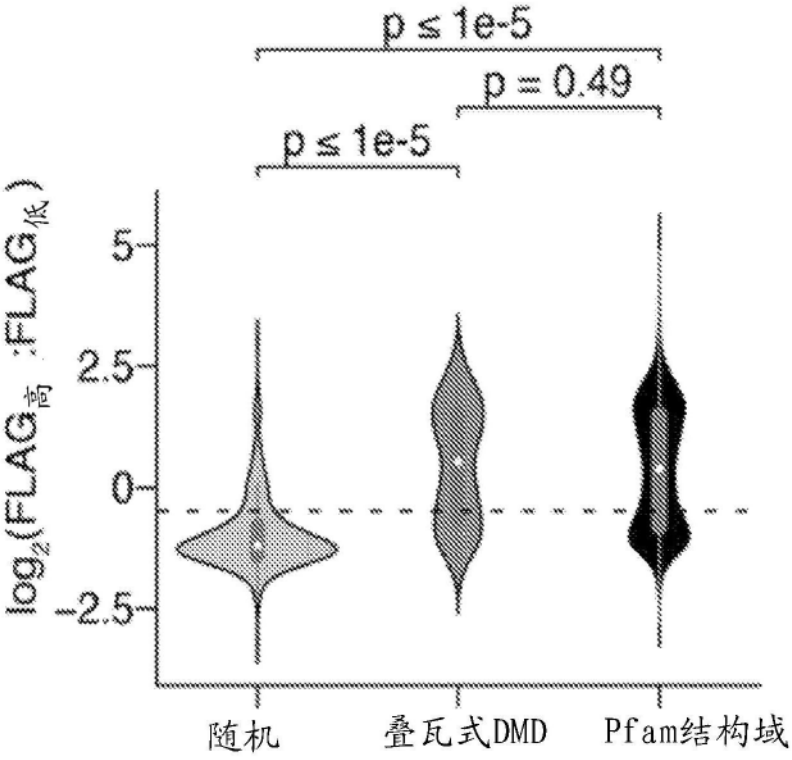

p ≤ 1e-5
p = 0.49
p ≤ 1e-5
log₂(FLAG高:FLAG低)
5
2.5
0
-2.5
随机
叠瓦式DMD
Pfam结构域

图17F

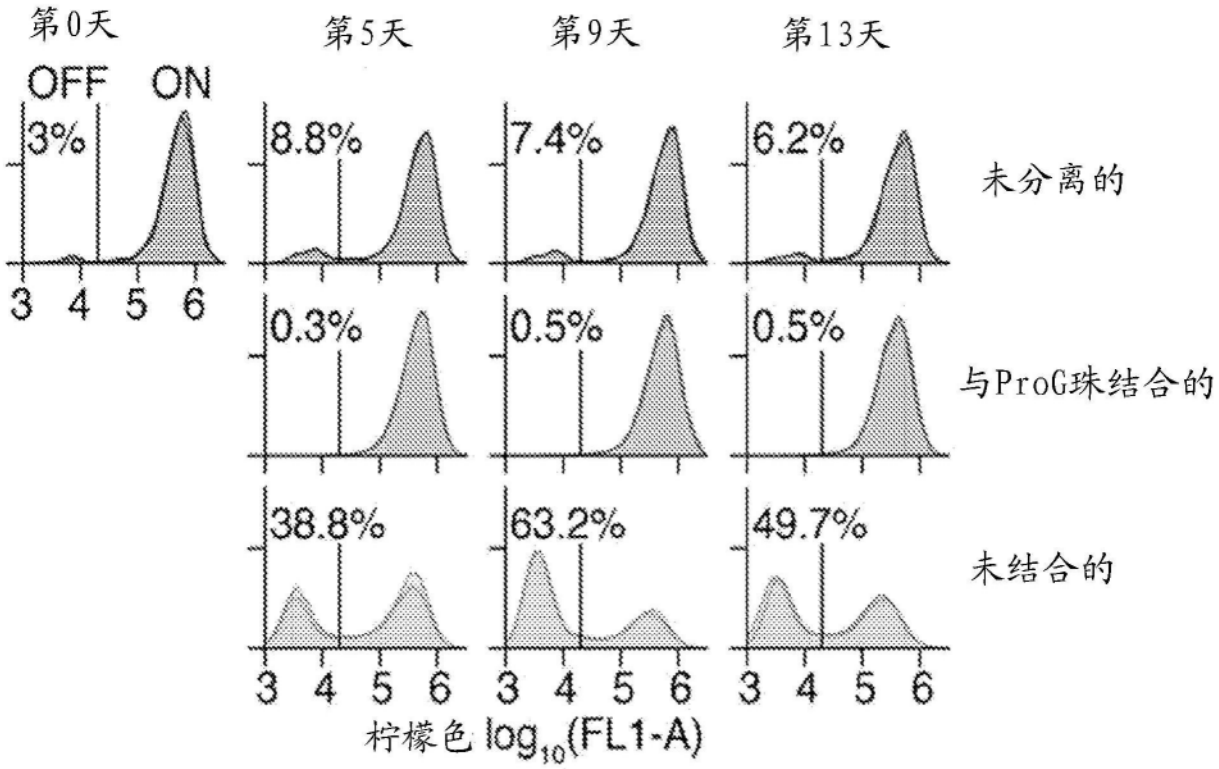

第0天
第5天
第9天
第13天
OFF
ON
3%
8.8%
7.4%
6.2%
未分离的
0.3%
0.5%
0.5%
与ProG珠结合的
38.8%
63.2%
49.7%
未结合的
3 4 5 6
柠檬色 log10(FL1-A)

图18A

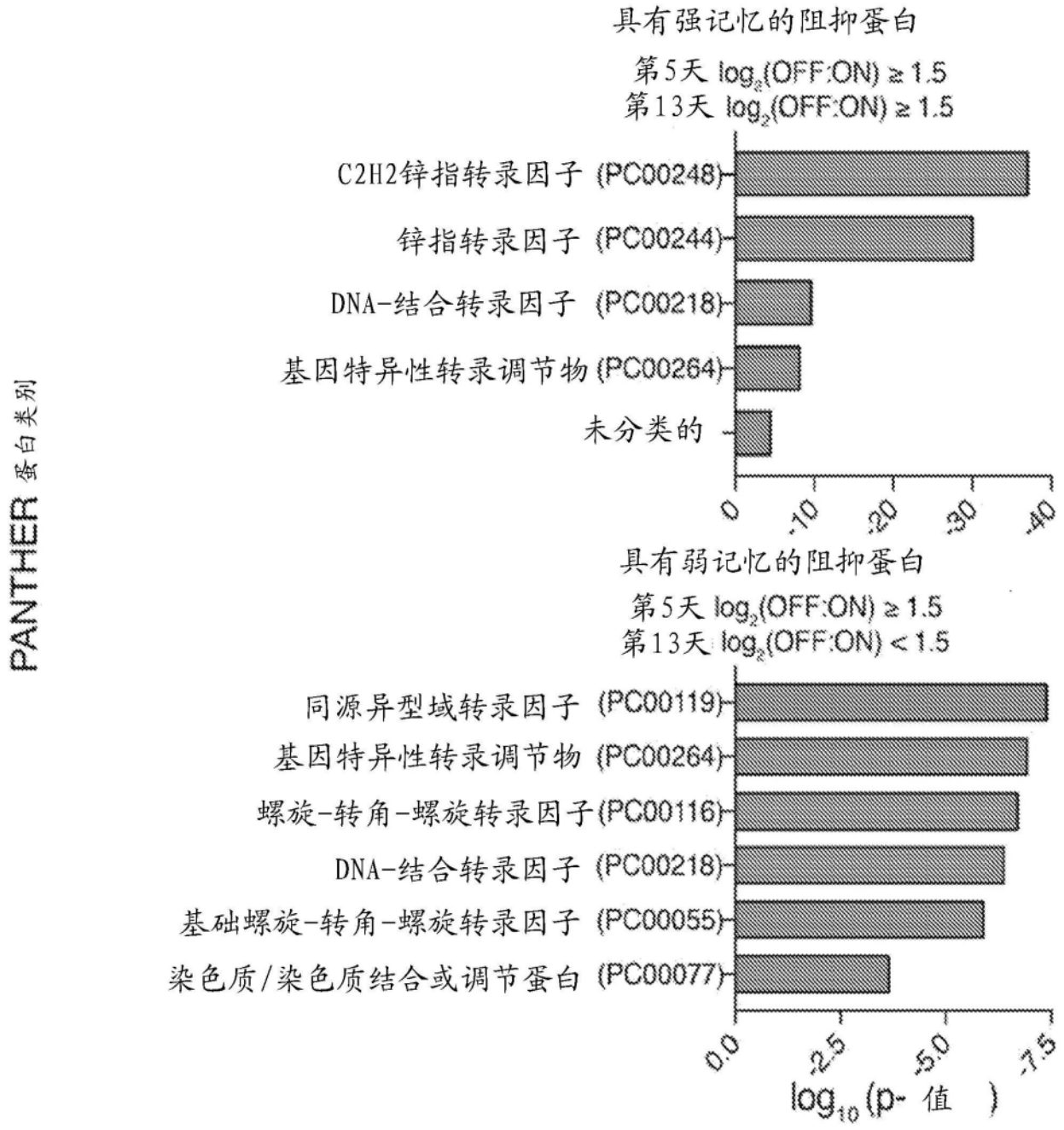

具有强记忆的阻抑蛋白
第5天 log2(OFF:ON) ≥ 1.5
第13天 log2(OFF:ON) ≥ 1.5
C2H2锌指转录因子 (PC00248)
锌指转录因子 (PC00244)
DNA-结合转录因子 (PC00218)
基因特异性转录调节物 (PC00264)
未分类的
0
10
20
30
40
具有弱记忆的阻抑蛋白
第5天 log2(OFF:ON) ≥ 1.5
第13天 log2(OFF:ON) < 1.5
同源异型域转录因子 (PC00119)
基因特异性转录调节物 (PC00264)
螺旋-转角-螺旋转录因子(PC00116)
DNA-结合转录因子 (PC00218)
基础螺旋-转角-螺旋转录因子 (PC00055)
染色质/染色质结合或调节蛋白 (PC00077)
0.0
2.5
5.0
7.5
log10(p-值)
PANTHER 蛋白类别

图18B

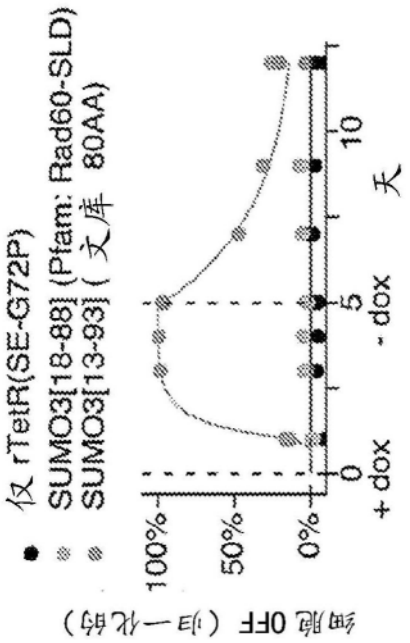

仅 rTetR(SE-G72P)
SUMO3[18-88] (Pfam: Rad60-SLD)
SUMO3[13-93] (文库 80AA)
100%
50%
0%
0
+ dox
5
- dox
10
天
细胞 OFF （归一化的）

图18C

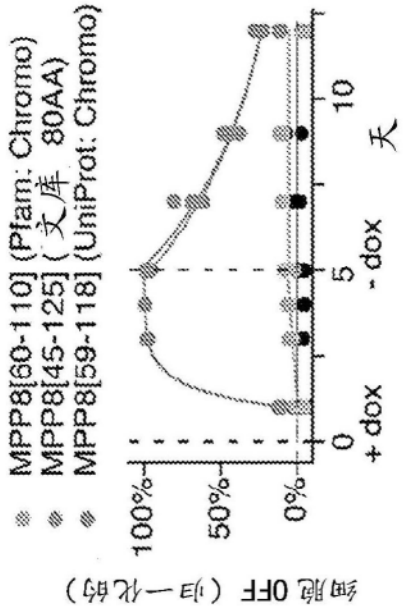

MPP8[60-110] (Pfam: Chromo)
MPP8[45-125] (文库 80AA)
MPP8[59-118] (UniProt: Chromo)
100%
50%
0%
0
+ dox
5
- dox
10
天
细胞 OFF （归一化的）

图18D

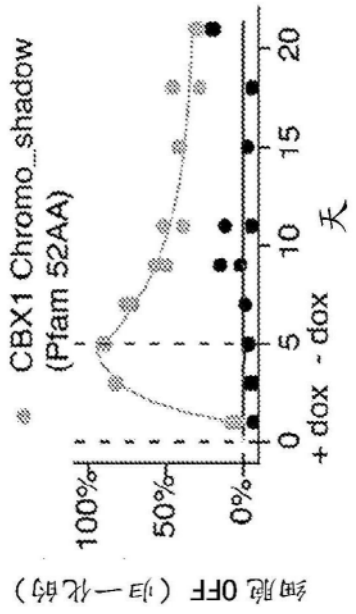

CBX1 Chromo_shadow
(Pfam 52AA)
100%
50%
0%
0
5
10
15
20
+ dox - dox
天
细胞 OFF（归一化的）

图18E

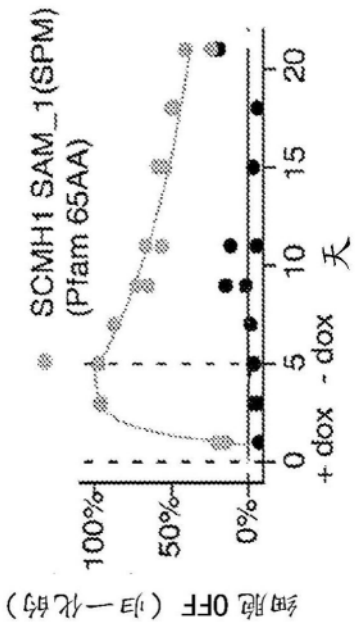

SCMH1 SAM_1(SPM)
(Pfam 65AA)
100%
50%
0%
0
5
10
15
20
+ dox - dox
天
细胞 OFF（归一化的）

图18F

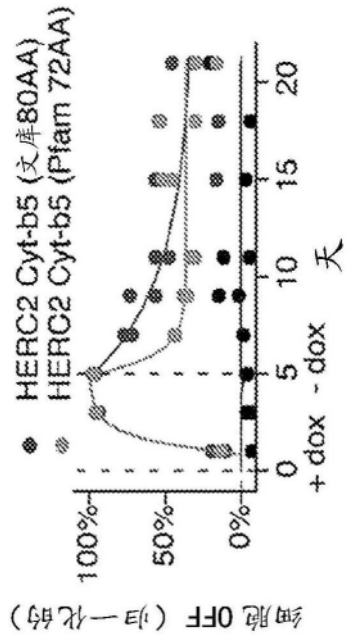

HERC2 Cyt-b5 (文库80AA)
HERC2 Cyt-b5 (Pfam 72AA)
100%
50%
0%
0
5
10
15
20
+ dox - dox
天
细胞 OFF （归一化的）

图18G

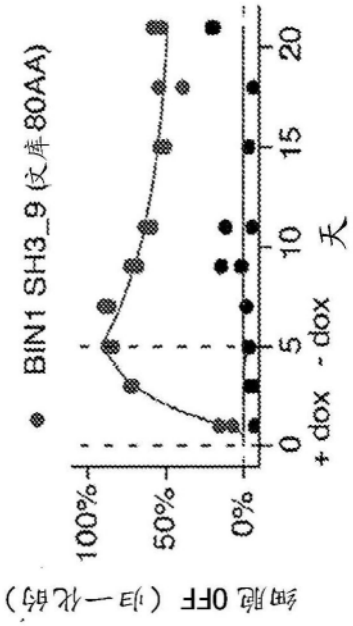

BIN1 SH3_9 (文库80AA)
100%
50%
0%
0
5
10
15
20
+ dox - dox
天
细胞 OFF （归一化的）

图18H

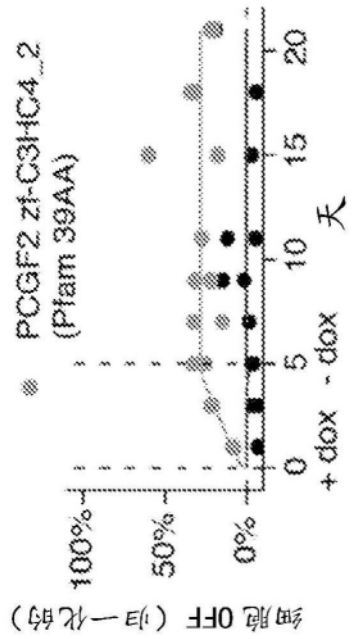

PCGF2 zf-C3HC4_2
(Pfam 39AA)
细胞 OFF（归一化的）
100%
50%
0%
+ dox - dox
0
5
10
15
20
天

图18I

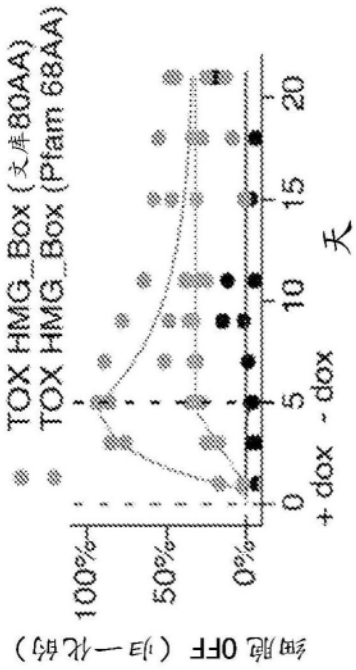

TOX HMG_Box (文库80AA)
TOX HMG_Box (Pfam 68AA)
细胞 OFF（归一化的）
100%
50%
0%
+ dox - dox
0
5
10
15
20
天

图18J

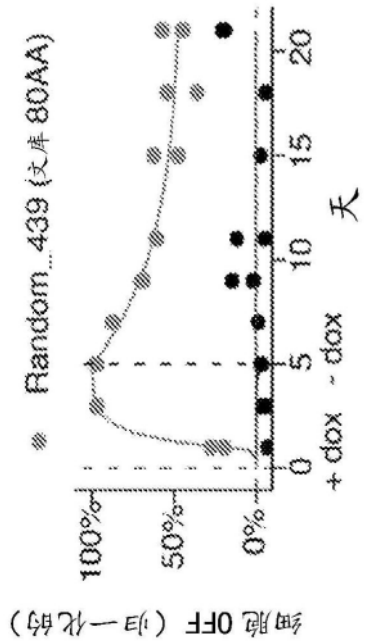


图18K

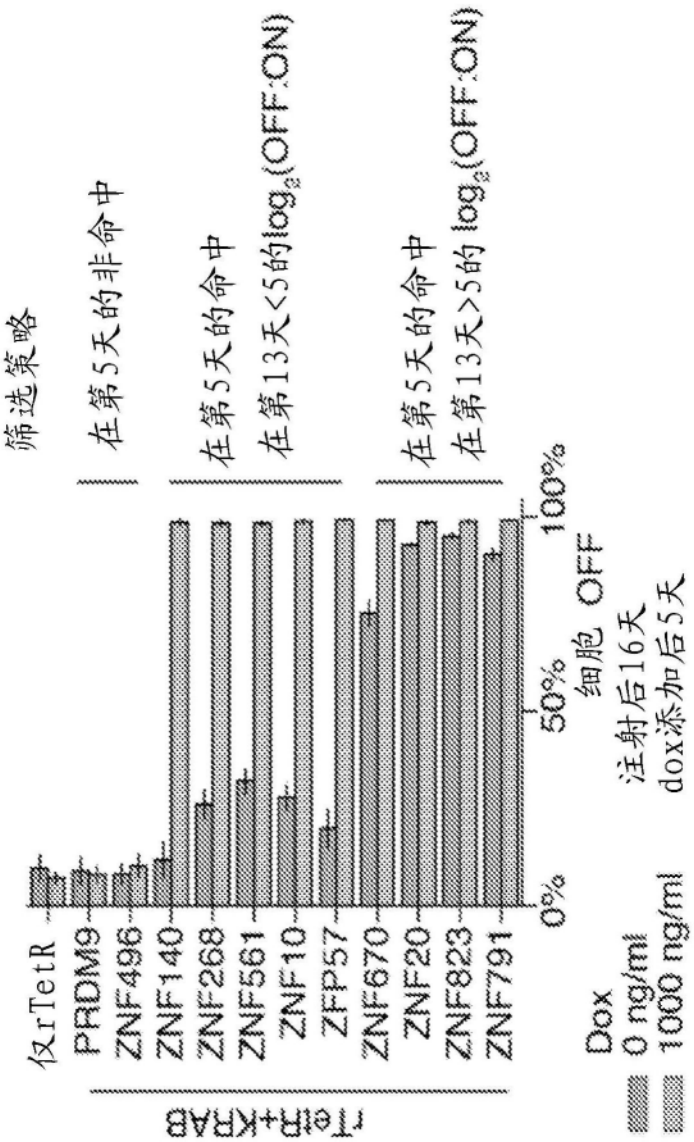

筛选策略
在第5天的非命中
在第5天的命中
在第13天<5的log2(OFF:ON)
在第5天的命中
在第13天>5的 log2(OFF:ON)
仅rTetR
PRDM9
ZNF496
ZNF140
ZNF268
ZNF561
ZNF10
ZFP57
ZNF670
ZNF20
ZNF823
ZNF791
ADD
KRAB+KD
rTetR
0%
50%
100%
细胞 OFF
注射后16天
dox添加后5天
Dox
0 ng/ml
1000 ng/ml

图19A

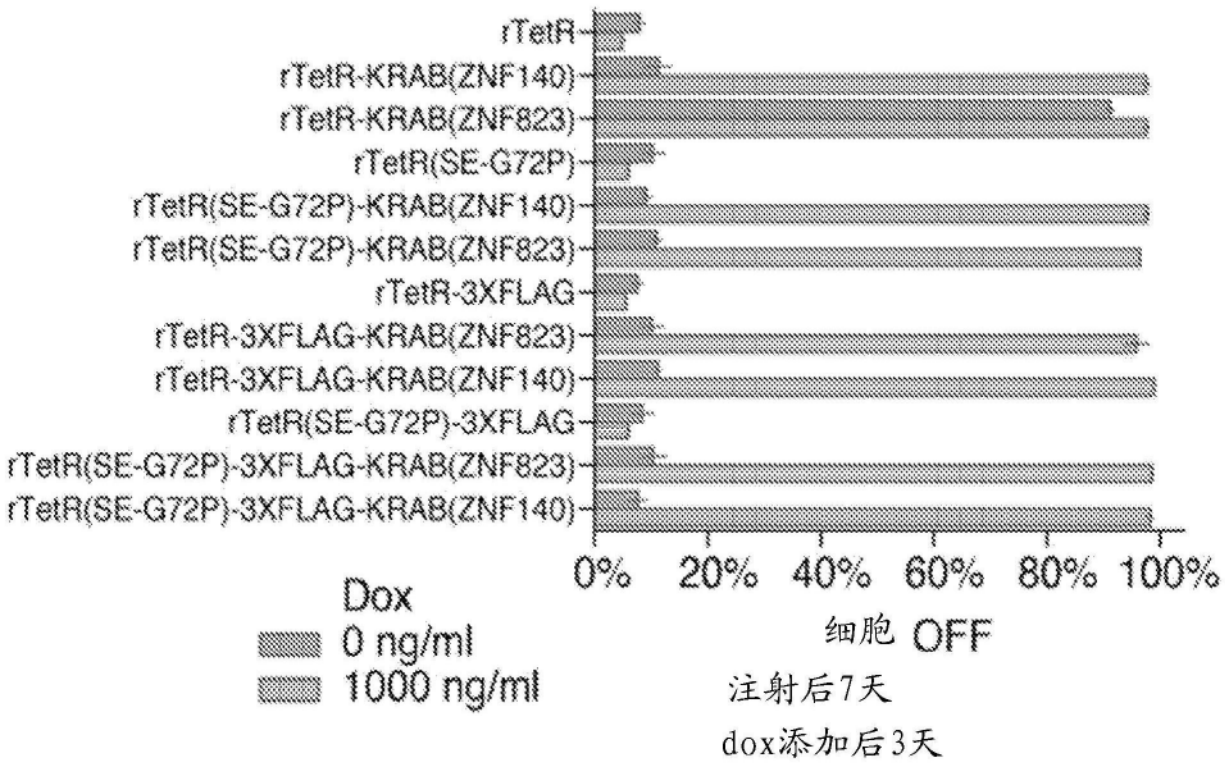

rTetR
rTetR-KRAB(ZNF140)
rTetR-KRAB(ZNF823)
rTetR(SE-G72P)
rTetR(SE-G72P)-KRAB(ZNF140)
rTetR(SE-G72P)-KRAB(ZNF823)
rTetR-3XFLAG
rTetR-3XFLAG-KRAB(ZNF823)
rTetR-3XFLAG-KRAB(ZNF140)
rTetR(SE-G72P)-3XFLAG
rTetR(SE-G72P)-3XFLAG-KRAB(ZNF823)
rTetR(SE-G72P)-3XFLAG-KRAB(ZNF140)
0%
20%
40%
60%
80%
100%
Dox
0 ng/ml
1000 ng/ml
细胞 OFF
注射后7天
dox添加后3天

图19B

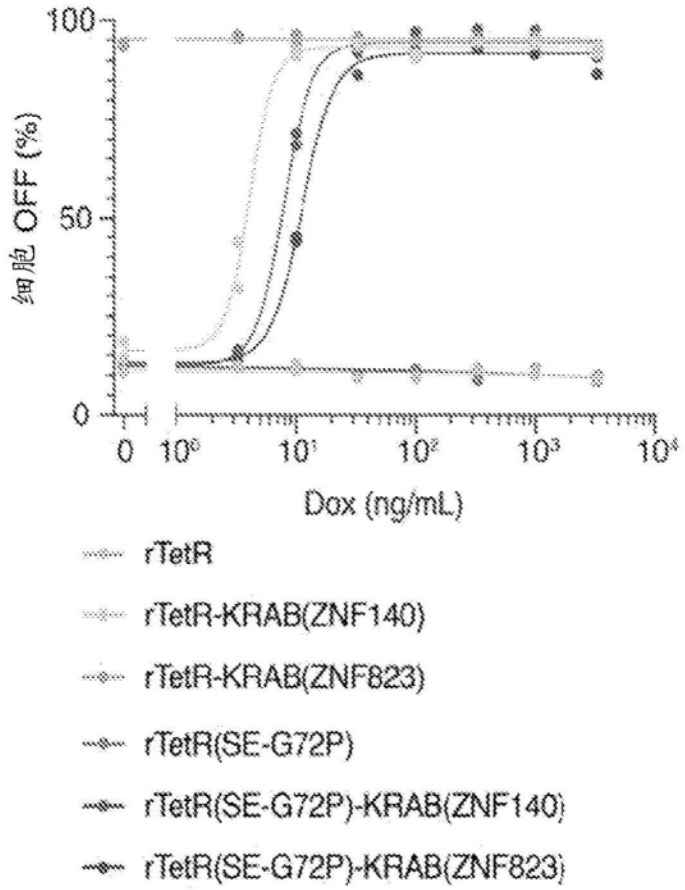

100
50
0
细胞 OFF (%)
0
10^0
10^1
10^2
10^3
10^4
Dox (ng/mL)
rTetR
rTetR-KRAB(ZNF140)
rTetR-KRAB(ZNF823)
rTetR(SE-G72P)
rTetR(SE-G72P)-KRAB(ZNF140)
rTetR(SE-G72P)-KRAB(ZNF823)

图19C

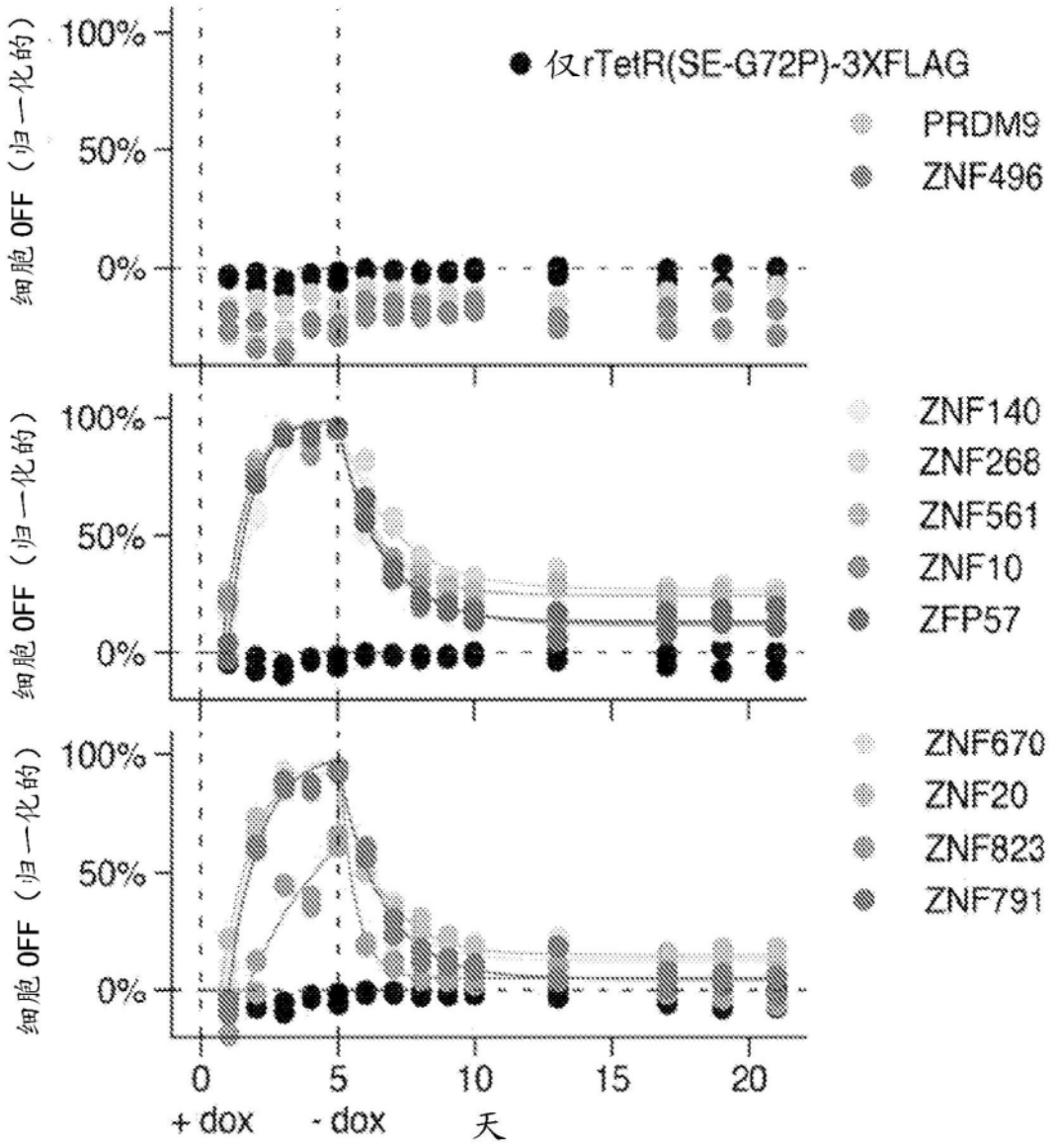

细胞 OFF（归一化的）
100%
50%
0%
仅rTetR(SE-G72P)-3XFLAG
PRDM9
ZNF496
ZNF140
ZNF268
ZNF561
ZNF10
ZFP57
ZNF670
ZNF20
ZNF823
ZNF791
0
5
10
15
20
+ dox
- dox
天

图19D

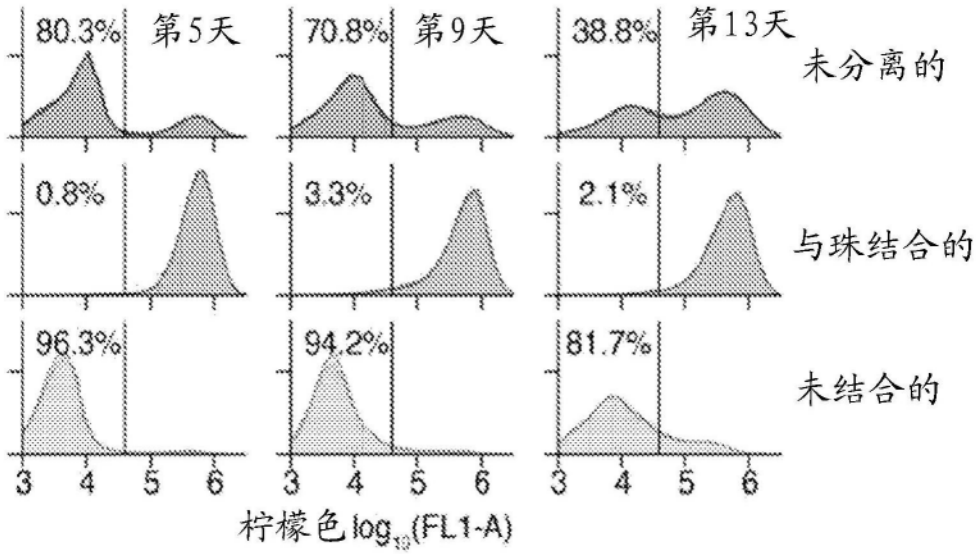

80.3%
第5天
70.8%
第9天
38.8%
第13天
未分离的
0.8%
3.3%
2.1%
与珠结合的
96.3%
94.2%
81.7%
未结合的
3
4
5
6
柠檬色log10(FL1-A)

图20A

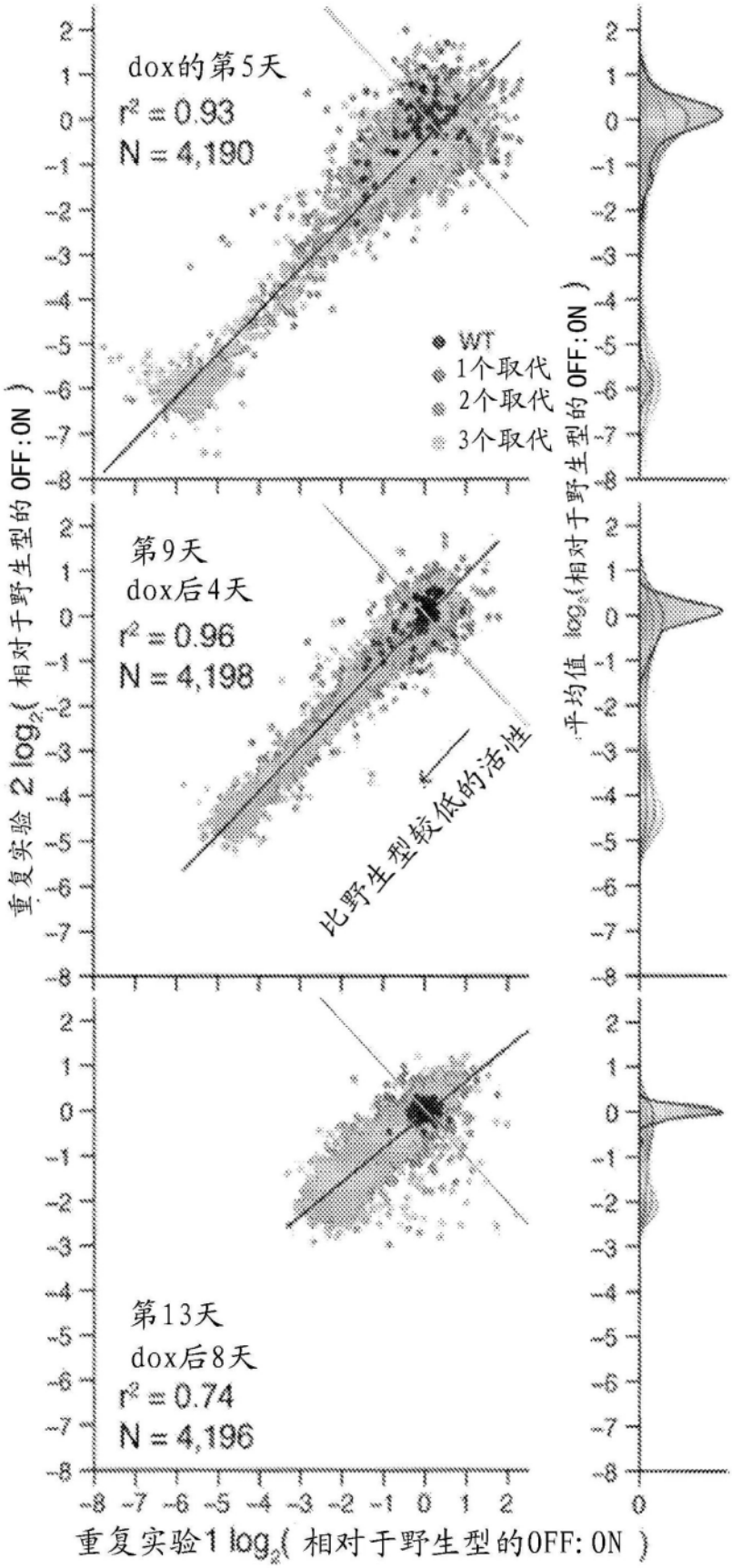

dox的第5天
r^2 = 0.93
N = 4,190
WT
1个取代
2个取代
3个取代
第9天
dox后4天
r^2 = 0.96
N = 4,198
比野生型较低的活性
第13天
dox后8天
r^2 = 0.74
N = 4,196
重复实验 2 $\log_2$(相对于野生型的 OFF:ON)
平均值 $\log_2$(相对于野生型的 OFF:ON)
重复实验1 $\log_2$(相对于野生型的OFF: ON)

图20B

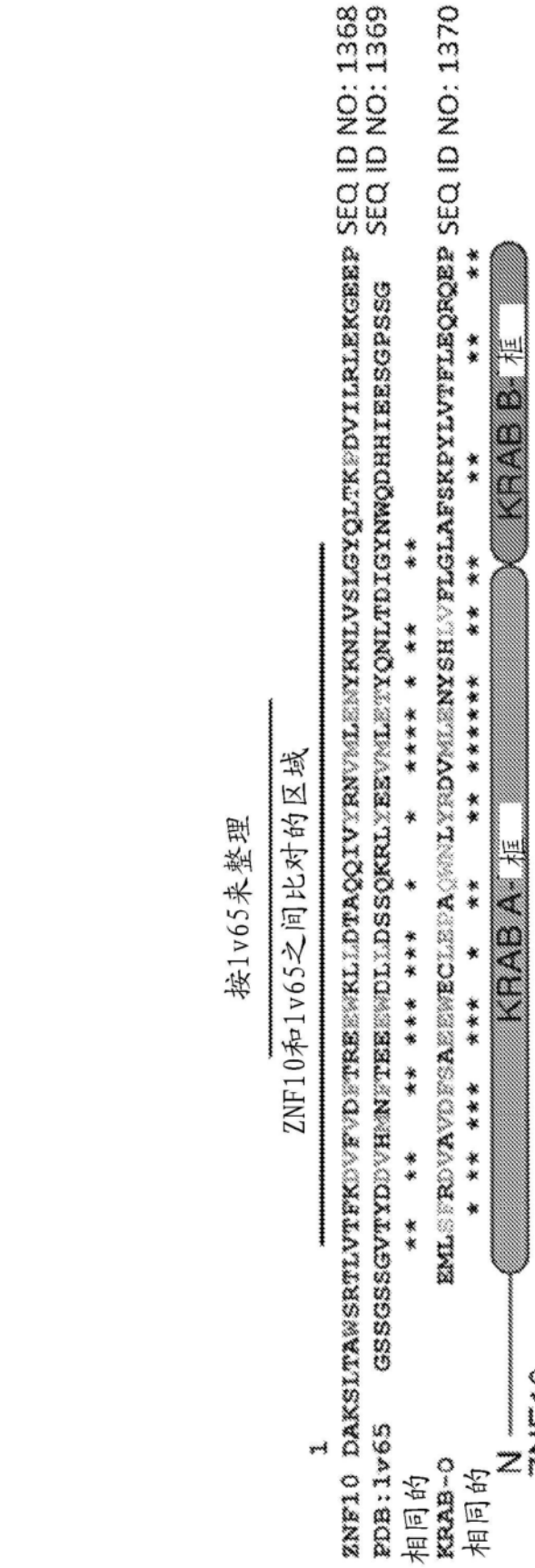

按1v65来整理
ZNF10和1v65之间比对的区域
1
ZNF10 DAKSLTAWSRTLVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEEP SEQ ID NO: 1368
PDB:1v65 GSSGSSGVTYDDVHVNFTEEEWDLLDSSQKRLYEEVMLETYQNLTDIGYNWQDHHIEESGPSSG SEQ ID NO: 1369
相同的
KRAB-O EMLSFRDVAVDFSAEEWECLDPAQRNLYRDVMLENYSHLVFLGLAFSKPYLVTFLEQRQEP SEQ ID NO: 1370
相同的
KRAB A 框
KRAB B 框
N
ZNF10

图20C

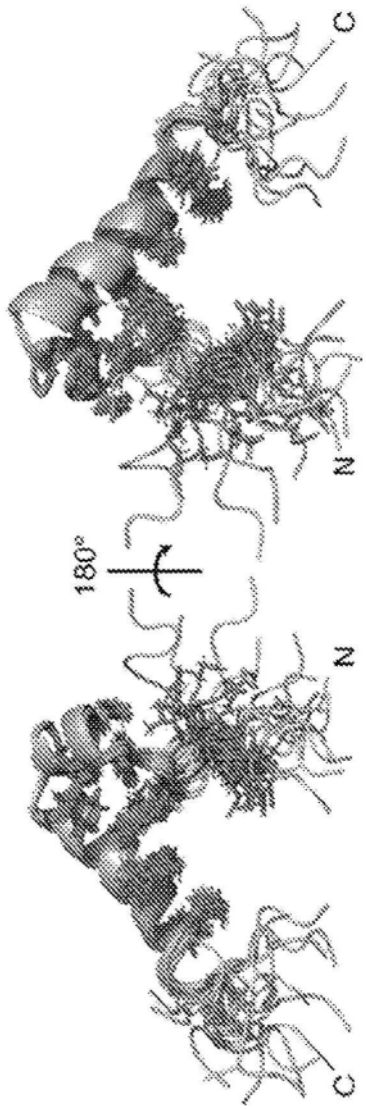


图20D

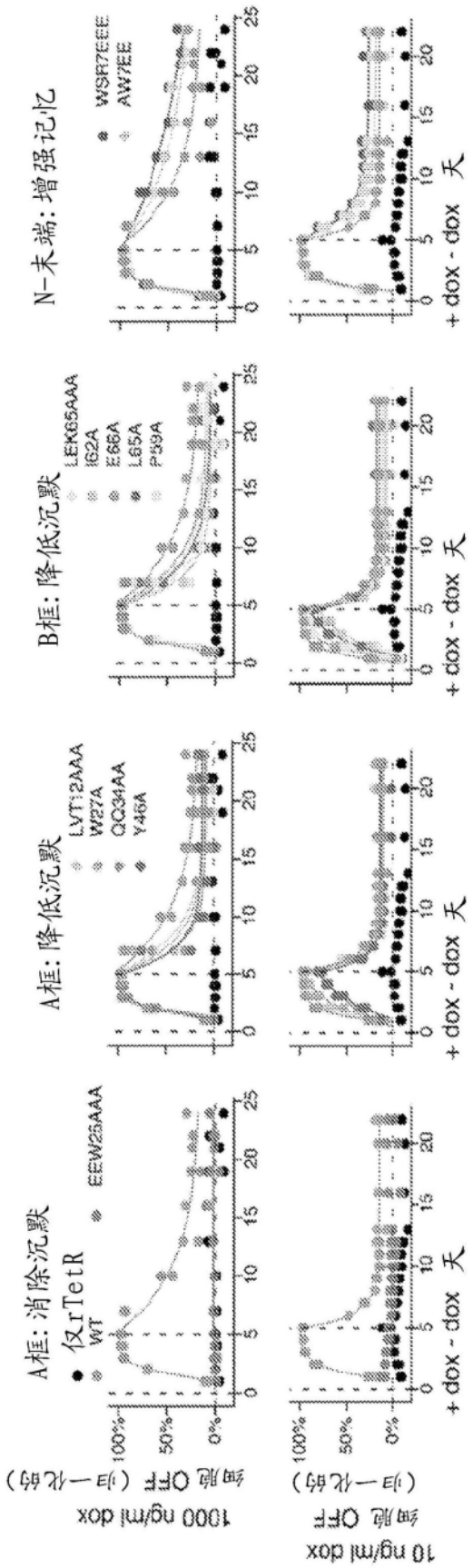

A框：消除沉默
仅rTetR
WT
EEW25AAA
A框：降低沉默
LVT12AAA
W27A
QQ34AA
Y46A
B框：降低沉默
LEK65AAA
I62A
E66A
L65A
P59A
N-末端：增强记忆
WSR7EEE
AW7EE
1000 ng/ml dox
10 ng/ml dox
细胞 OFF（归一化的）
100%
50%
0%
0
5
10
15
20
25
+ dox - dox 天

图20E

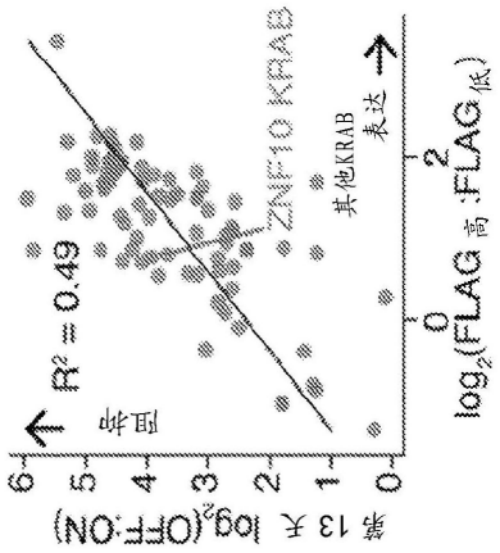

第13天 log2(OFF:ON)
阻抑
R² = 0.49
ZNF10 KRAB
其他KRAB
表达
log2(FLAG高:FLAG低)

图20F

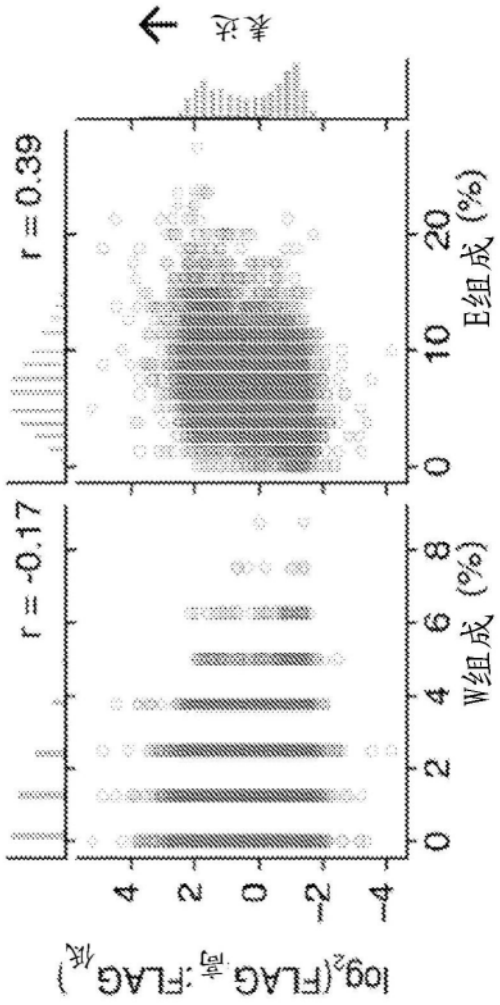

表达
r = 0.39
E组成 (%)
r = -0.17
W组成 (%)
log2(FLAG高:FLAG低)

图20G

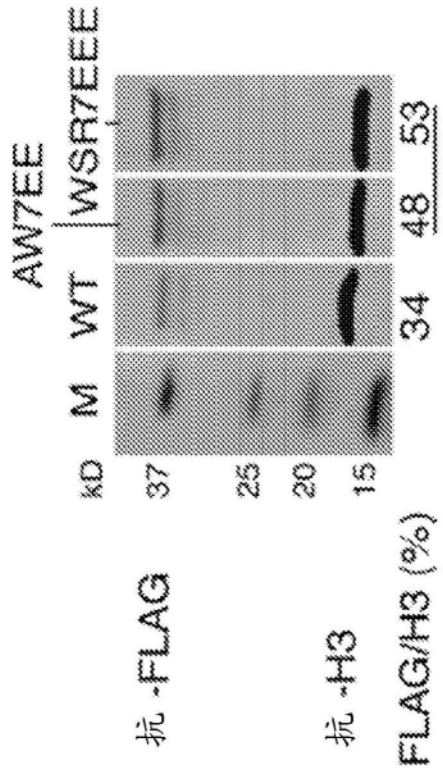


图20H

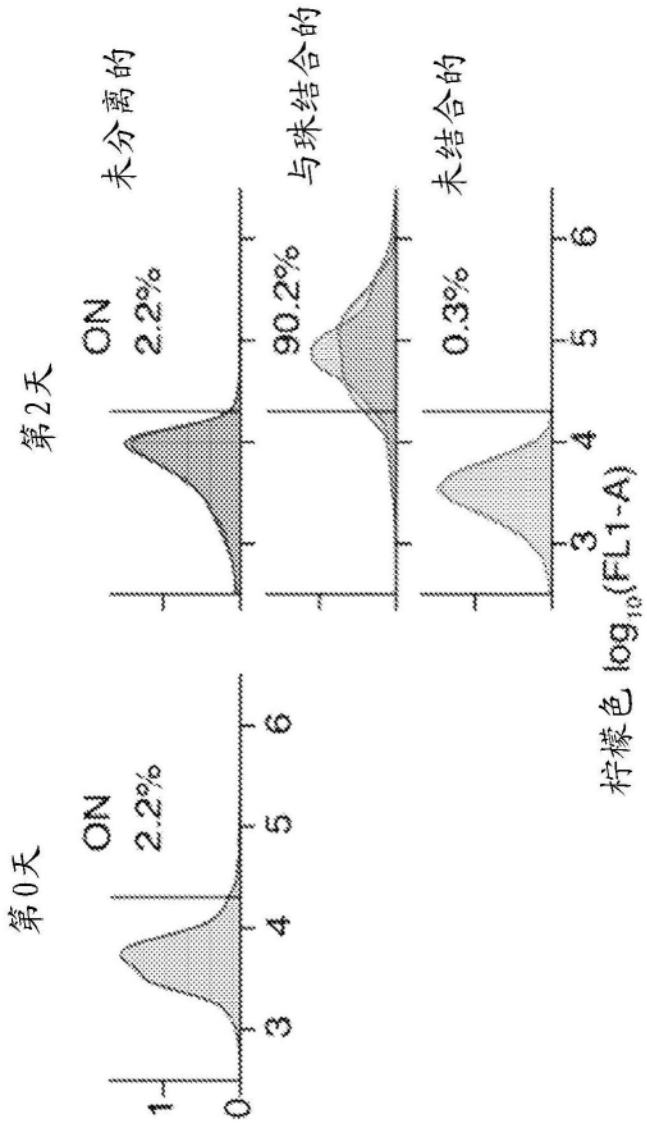

第0天
ON
2.2%
第2天
ON
2.2%
未分离的
90.2%
与珠结合的
0.3%
未结合的
0
1
3
4
5
6
柠檬色 log10(FL1-A)

图21A

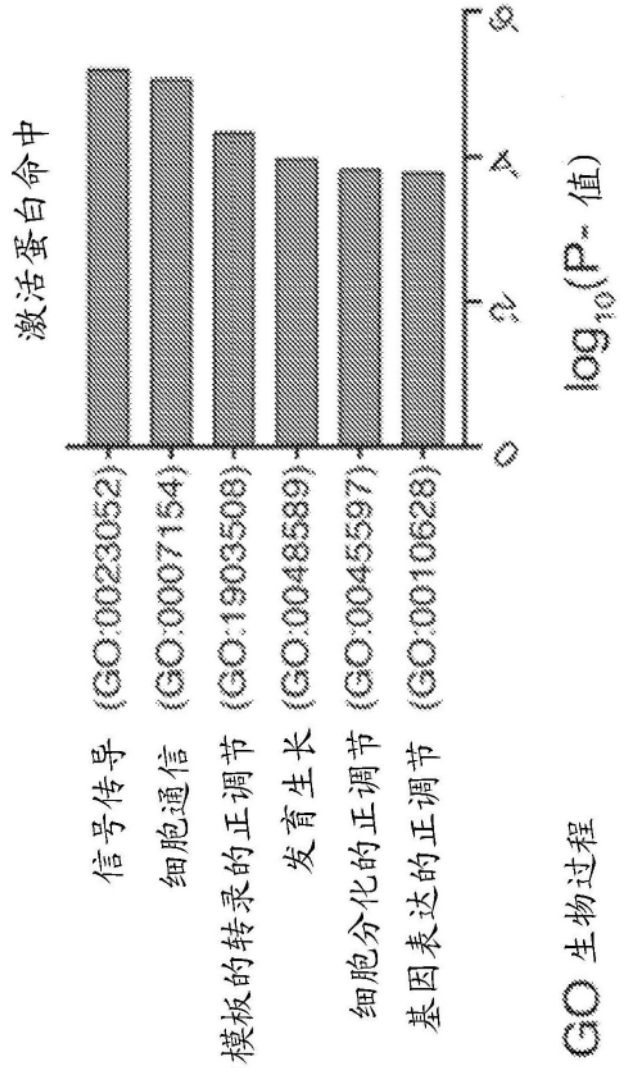

激活蛋白命中
信号传导 (GO:0023052)
细胞通信 (GO:0007154)
以核酸为模板的转录的正调节 (GO:1903508)
发育生长 (GO:0048589)
细胞分化的正调节 (GO:0045597)
基因表达的正调节 (GO:0010628)
GO 生物过程
0
2
4
6
log10(P-值)

图21B

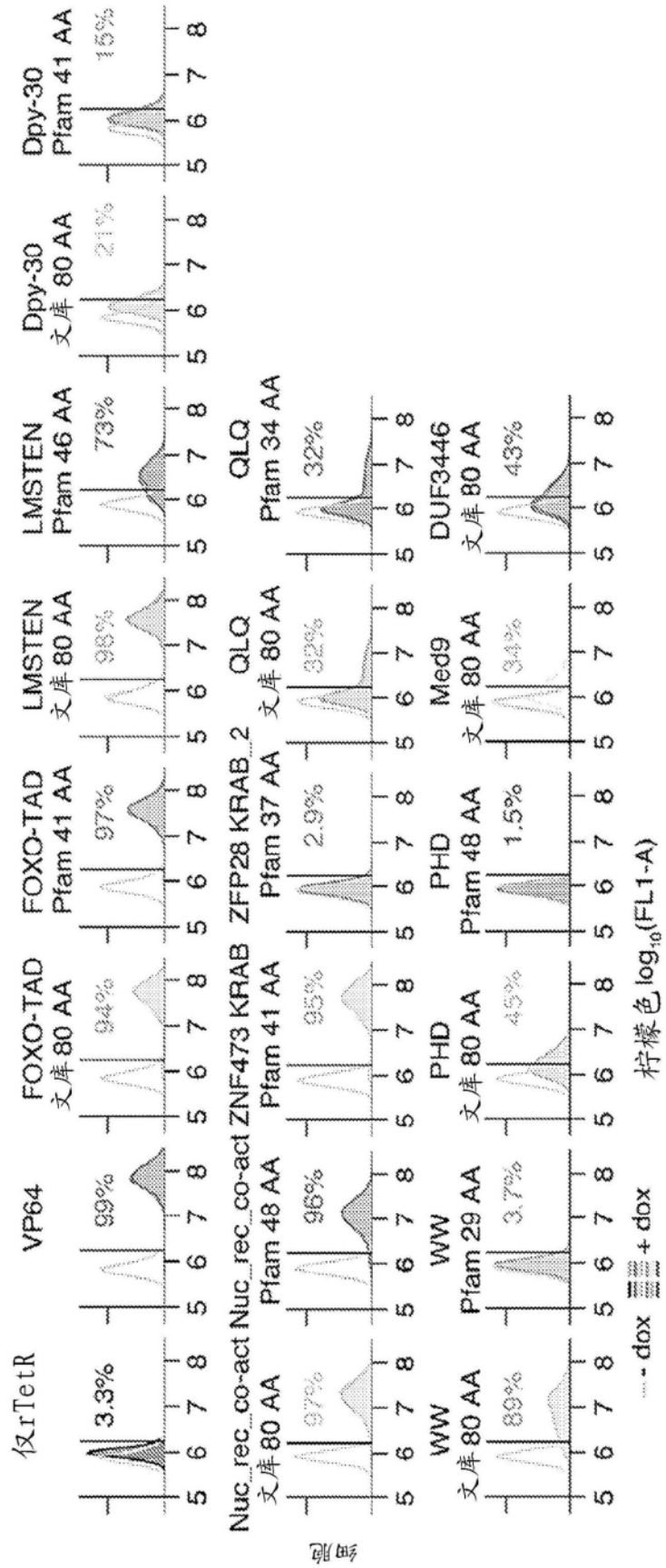

仅rTetR
VP64
FOXO-TAD 文库80 AA
FOXO-TAD Pfam 41 AA
LMSTEN 文库80 AA
LMSTEN Pfam 46 AA
Dpy-30 文库80 AA
Dpy-30 Pfam 41 AA
Nuc_rec_co-act
Nuc_rec_co-act
ZNF473 KRAB
ZFP28 KRAB_2
QLQ
QLQ
WW
WW
PHD
PHD
Med9
DUF3446
文库80 AA
Pfam 48 AA
Pfam 41 AA
Pfam 37 AA
文库80 AA
Pfam 34 AA
Pfam 29 AA
文库80 AA
Pfam 48 AA
文库80 AA
3.3%
99%
94%
97%
98%
73%
21%
15%
97%
96%
93%
2.9%
82%
32%
89%
3.7%
45%
1.5%
84%
43%
细胞
- dox
+ dox
柠檬色 log10(FL1-A)

图21C

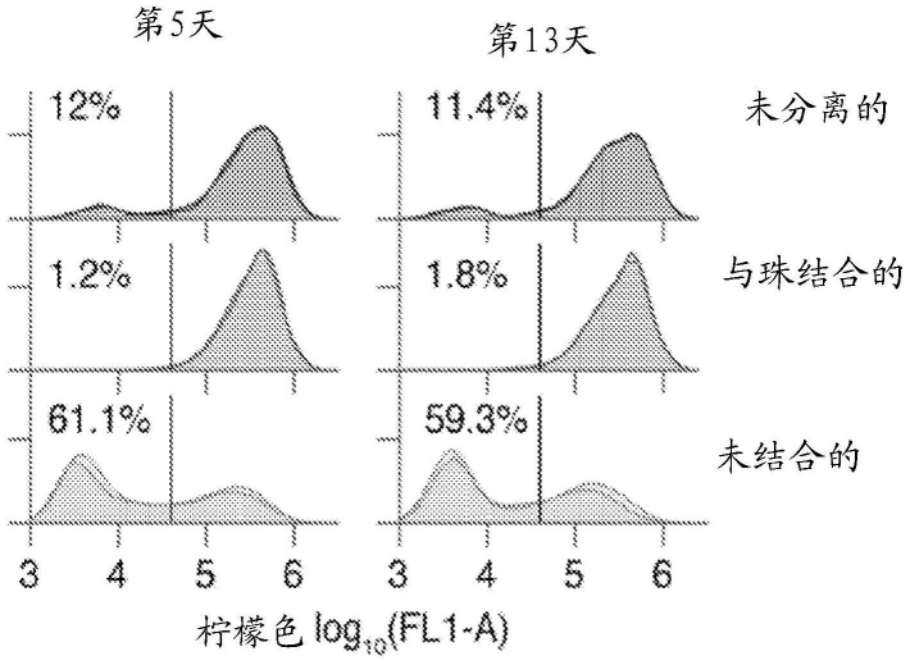

第5天
第13天
12%
11.4%
未分离的
1.2%
1.8%
与珠结合的
61.1%
59.3%
未结合的
3 4 5 6
3 4 5 6
柠檬色 log10(FL1-A)

图22A

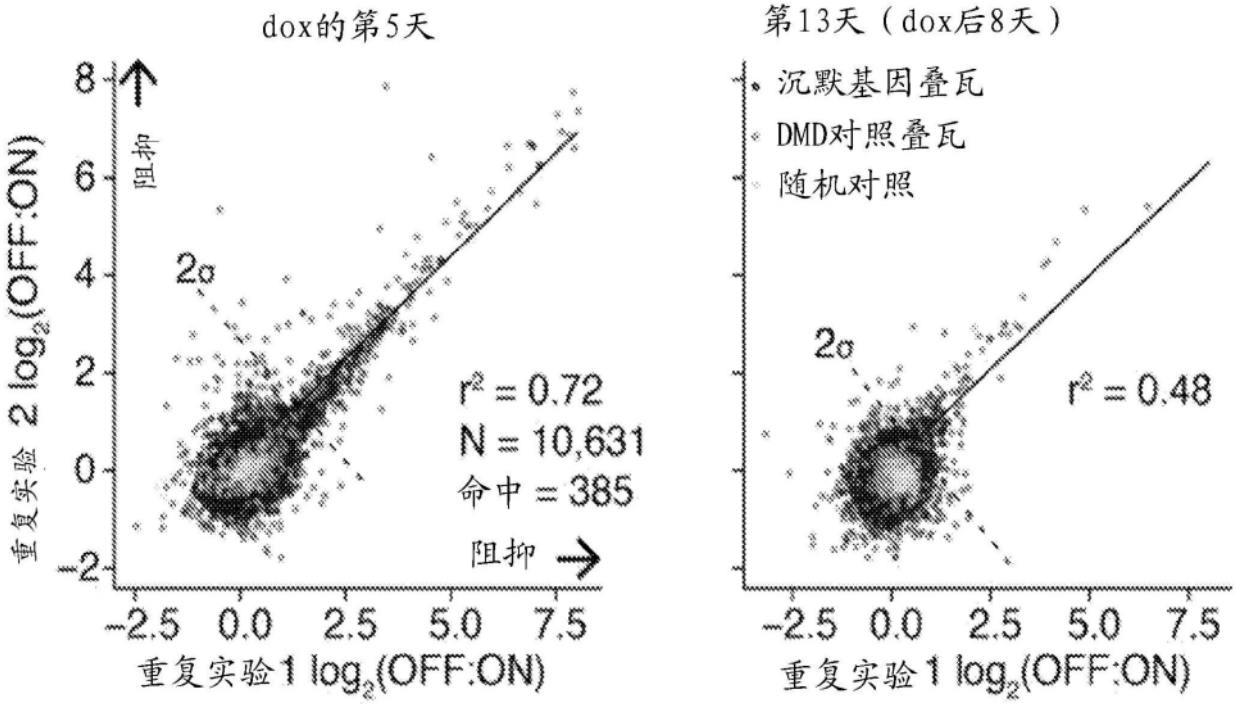

dox的第5天
第13天（dox后8天）
沉默基因叠瓦
DMD对照叠瓦
随机对照
阻抑
2σ
r² = 0.72
N = 10,631
命中 = 385
阻抑
r² = 0.48
重复实验 2 log2(OFF:ON)
重复实验1 log2(OFF:ON)
8 6 4 2 0 -2
-2.5 0.0 2.5 5.0 7.5

图22B

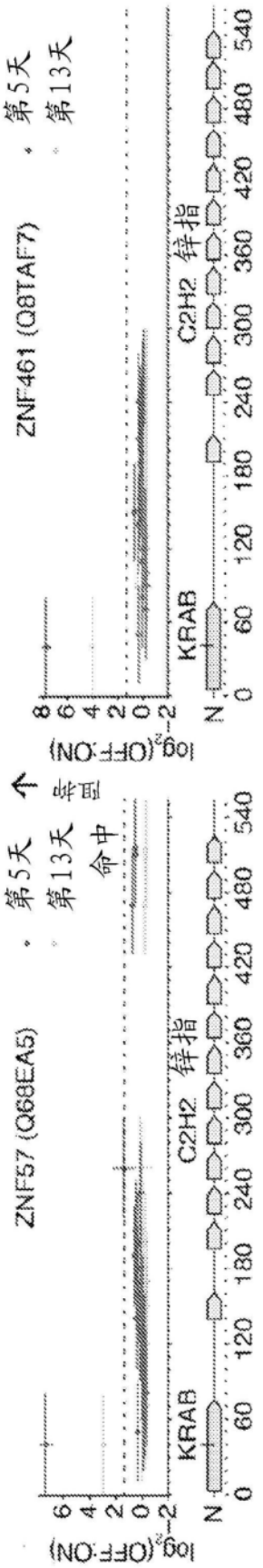

ZNF461 (Q8TAF7)
第5天
第13天
log2(OFF:ON)
KRAB
C2H2 锌指
N
0
60
120
180
240
300
360
420
480
540
ZNF57 (Q69EA5)
第5天
第13天
命中
log2(OFF:ON)
KRAB
C2H2 锌指
N

图22C

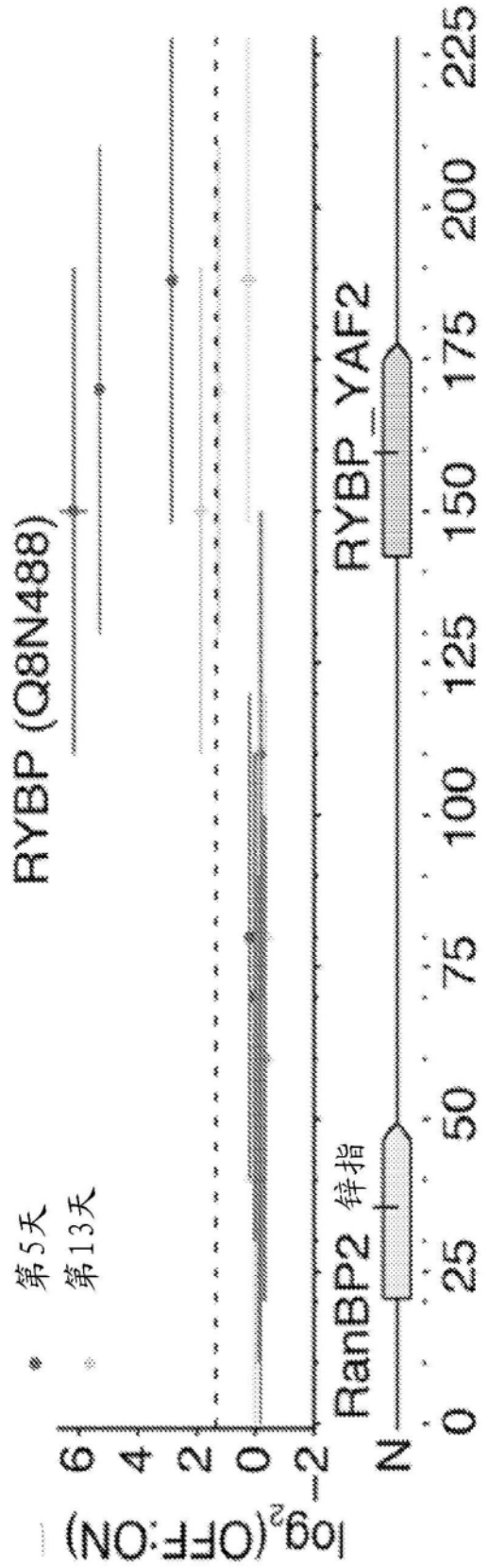

RYBP (Q8N488)
第5天
第13天
log2(OFF:ON)
6
4
2
0
-2
N
RanBP2
锌指
RYBP_YAF2
0
25
50
75
100
125
150
175
200
225

图22D

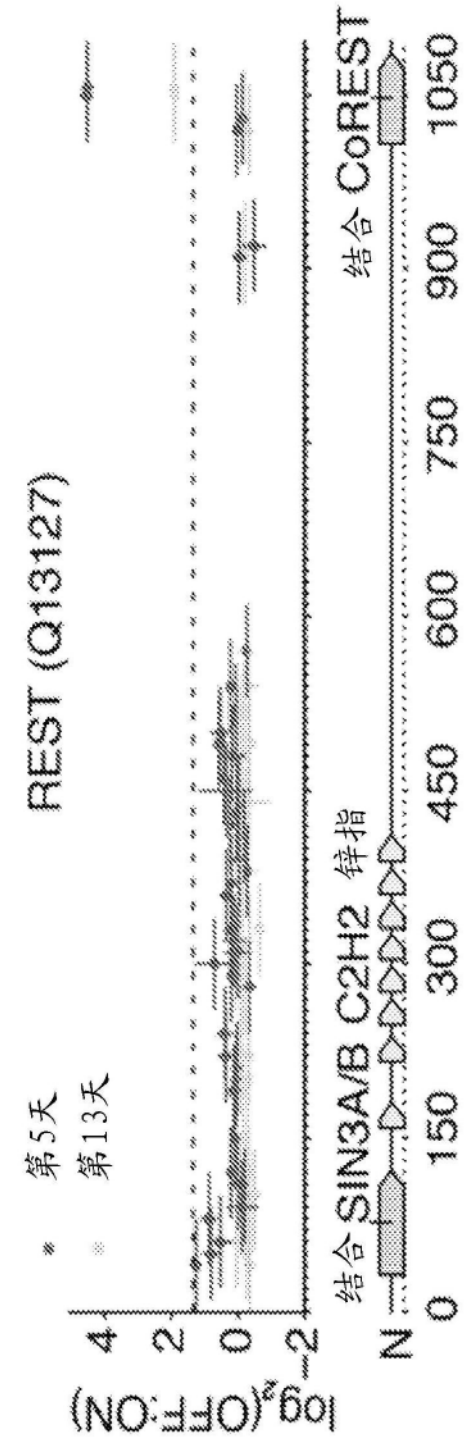

REST (Q13127)
第5天
第13天
log2(OFF:ON)
4
2
0
-2
N
结合 SIN3A/B
C2H2 锌指
结合 CoREST
0
150
300
450
600
750
900
1050

图22E

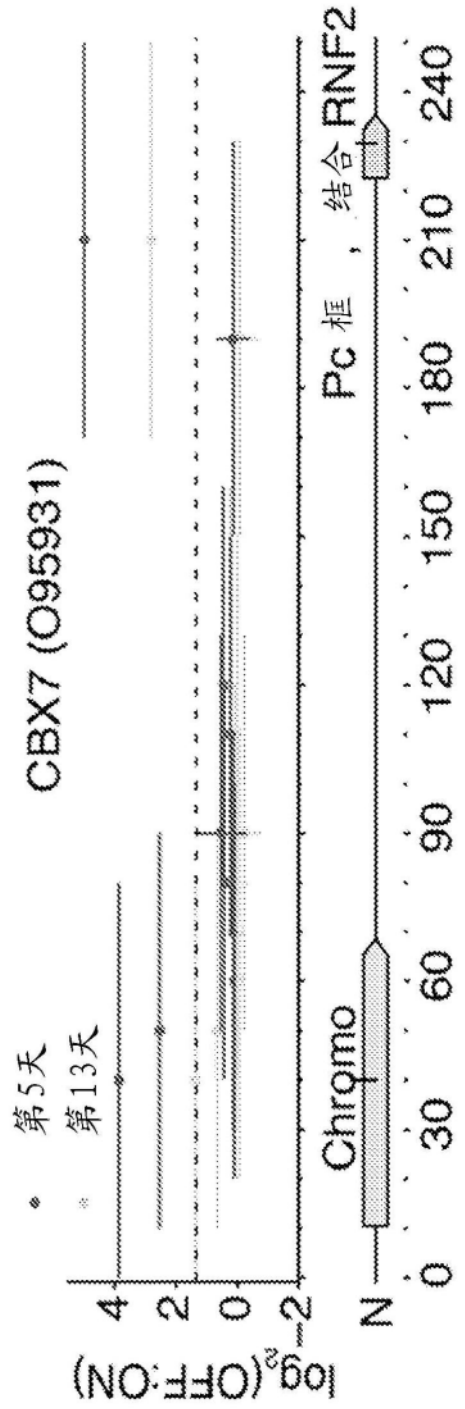

CBX7 (O95931)
第5天
第13天
log2(OFF:ON)
4
2
0
−2
Chromo
Pc 框 ，结合 RNF2
N
0
30
60
90
120
150
180
210
240

图22F

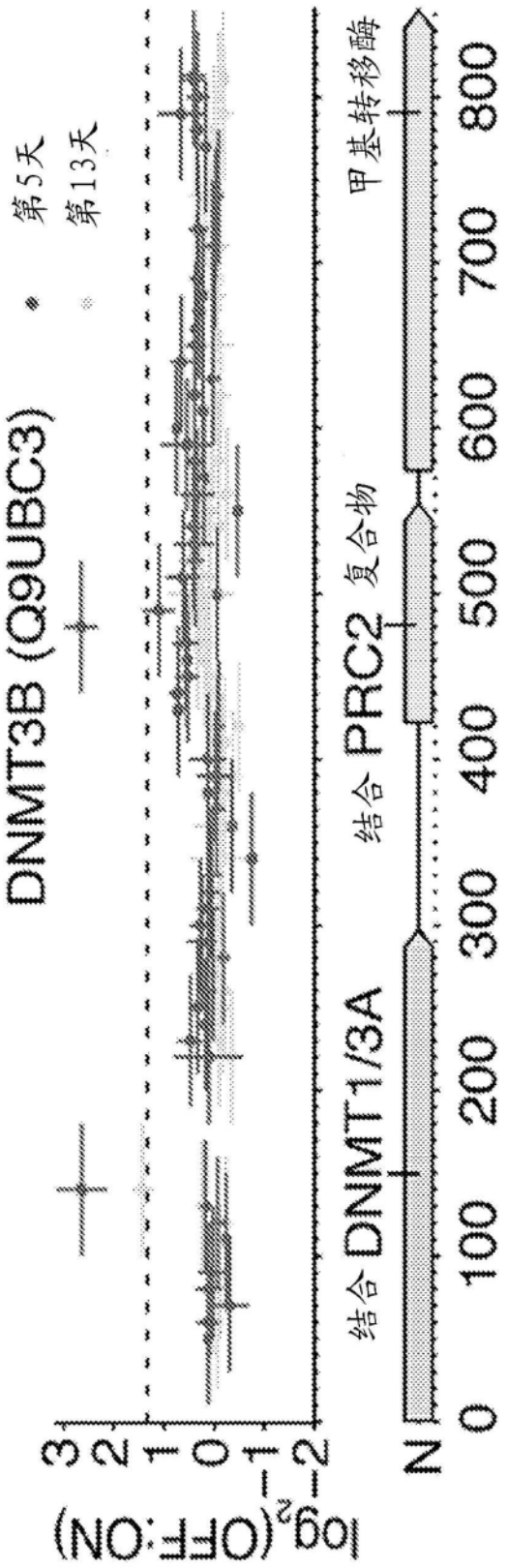

DNMT3B (Q9UBC3)
第5天
第13天
log2(OFF:ON)
3
2
1
0
−1
−2
结合 DNMT1/3A
结合 PRC2 复合物
甲基转移酶
N
0
100
200
300
400
500
600
700
800

图22G

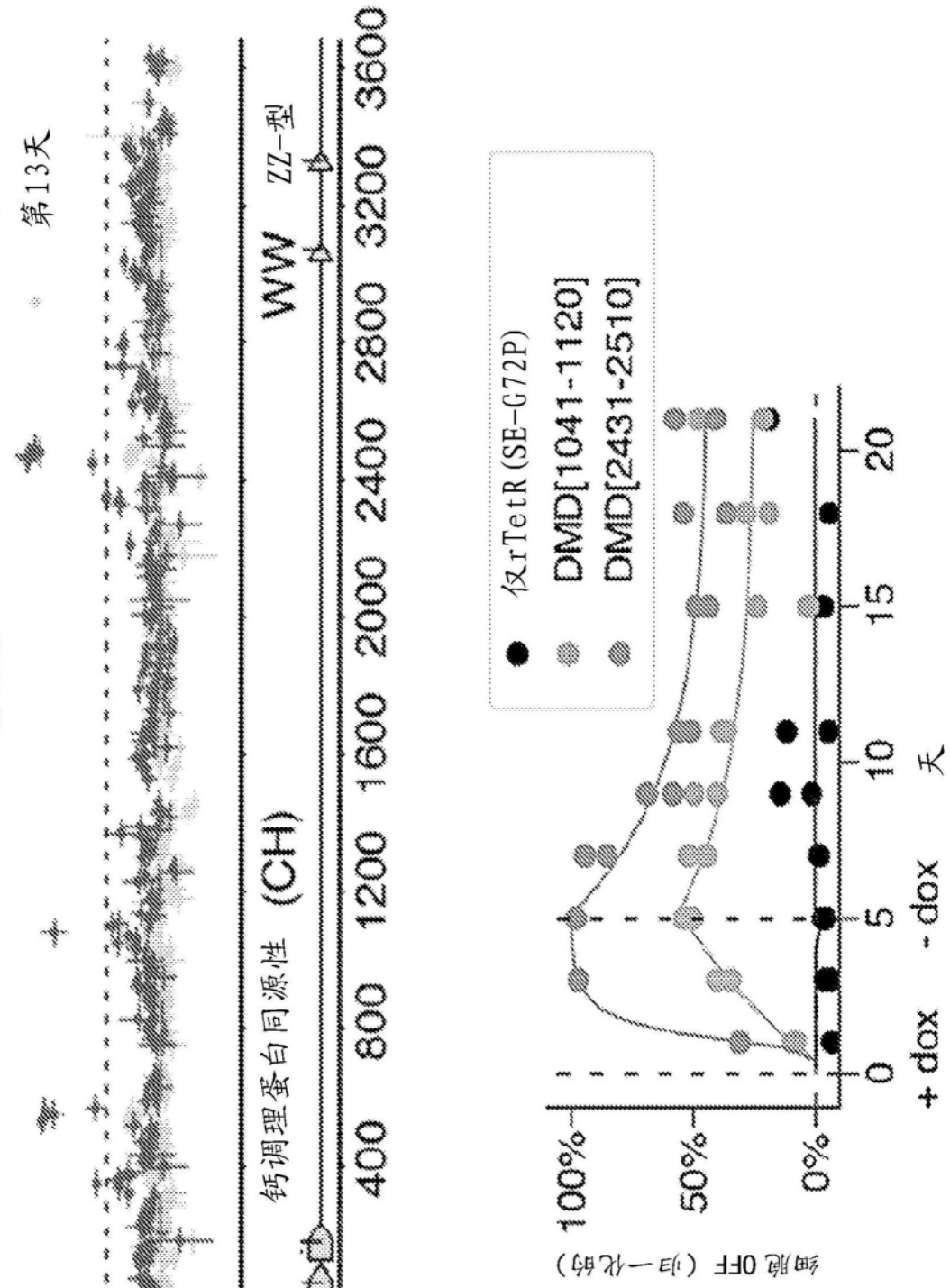

DMD
第5天
第13天
log2(OFF:ON)
钙调理蛋白同源性 (CH)
WW ZZ-型
N
C
0 400 800 1200 1600 2000 2400 2800 3200 3600
仅rTetR(SE-G72P)
DMD[1041-1120]
DMD[2431-2510]
细胞 OFF (归一化的)
100%
50%
0%
+ dox - dox
0 5 10 15 20
天

图22H