

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号
特表2024-524696
(P2024-524696A)

(43)公表日 令和6年7月5日(2024.7.5)

(51)国際特許分類		F I		テーマコード (参考)	
C 1 2 N	15/62 (2006.01)	C 1 2 N	15/62	Z	4 B 0 6 3
C 1 2 N	15/861 (2006.01)	C 1 2 N	15/861	Z Z N A	4 B 0 6 5
C 1 2 N	15/867 (2006.01)	C 1 2 N	15/867	Z	
C 1 2 N	5/10 (2006.01)	C 1 2 N	5/10		
C 1 2 N	15/12 (2006.01)	C 1 2 N	15/12		
		審査請求	未請求	予備審査請求	未請求 (全139頁) 最終頁に続く
(21)出願番号	特願2024-501992(P2024-501992)		(71)出願人	524004847	
(86)(22)出願日	令和4年7月14日(2022.7.14)			ザ ガバニング カウンシル オブ ザ ユ	
(85)翻訳文提出日	令和6年2月19日(2024.2.19)			ニバーシティ オブ トロント	
(86)国際出願番号	PCT/CA2022/051100			カナダ国, オンタリオ州 エム5ジー 0	
(87)国際公開番号	WO2023/000081			シー6, トロント, スイート ダブリュ	
(87)国際公開日	令和5年1月26日(2023.1.26)			ー540, 5 フロアー, 108 カレッ	
(31)優先権主張番号	63/221,611			ジ ストリート, シュワルツ リーズマン	
(32)優先日	令和3年7月14日(2021.7.14)			イノベーション キャンパス, イノベー	
(33)優先権主張国・地域又は機関	米国(US)			ションズ アンド パートナーシップス	
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA		(74)代理人	100114775	
	,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(			弁理士 高岡 亮一	
	AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A		(74)代理人	100121511	
	T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR			弁理士 小田 直	
	,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,			最終頁に続く	
	最終頁に続く				

(54)【発明の名称】 トランス活性化因子ならびにその方法及び使用

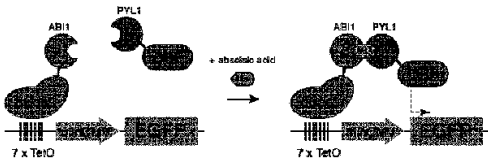
(57)【要約】

DNA標的化ドメイン、好ましくは触媒的に不活性なDNA標的化タンパク質、例えばCRISPR-Casタンパク質と、本明細書に記載の少なくとも1つのトランス活性化ドメインまたはその機能的バリエーションを含むエフェクタードメインとを含む、異種転写活性化因子。また、本明細書において、発現構築物、ベクター、及び前記転写活性化因子をコードまたは発現する細胞、ならびに標的遺伝子の転写活性化のためのシステム及び方法、ならびにその作製及び使用において使用される組成物、キット、及び試薬も提供される。

【選択図】 図1 A

Fig. 1

A



【特許請求の範囲】

【請求項 1】

異種転写活性化因子であって、

DNA 標的化ドメイン、任意選択で酵素的に不活性な CRISPR-CAS タンパク質、ジंकフィンガー DNA 結合ドメイン、tet-リプレッサー、または転写活性化因子様エフェクター (TALE) DNA 結合ドメインと、

表 2 または表 6 に列挙されるトランス活性化ドメイン (TAD) もしくはその機能的バリエーション、任意選択で表 6 から選択される少なくとも 1 つの TAD、または表 1 もしくは表 3 に列挙される TAD、もしくはその機能的バリエーションから選択される少なくとも 2 つの TAD、好ましくは表 4 もしくは表 5 もしくは表 6 に列挙される TAD、及び／またはその機能的バリエーションから選択される少なくとも 1 つの TAD を含む、エフェクタードメインと、を含み、

前記 DNA 標的化ドメインと前記エフェクタードメインが、作動可能に連結される、前記異種転写活性化因子。

【請求項 2】

前記エフェクタードメインが、表 1 もしくは表 3 に列挙される TAD またはその機能的バリエーションから選択される少なくとも 3 つ、または少なくとも 4 つのトランス活性化ドメインを含む、請求項 1 に記載の転写活性化因子。

【請求項 3】

少なくとも 1 つの相互作用成分をさらに含む、請求項 1 または請求項 2 に記載の転写活性化因子。

【請求項 4】

前記 DNA 標的化ドメイン及びエフェクタードメインが、単一のポリペプチドのドメインである、請求項 1 ～ 3 のいずれか 1 項に記載の転写活性化因子。

【請求項 5】

前記 DNA 標的化ドメイン及び第 1 の相互作用成分を含む第 1 のポリペプチドと、エフェクタードメイン及び第 2 の相互作用成分を含む第 2 のポリペプチドと、を含み、ここで、前記第 1 及び第 2 の相互作用成分が、適切な条件下で相互作用する、請求項 3 に記載の転写活性化因子。

【請求項 6】

前記第 1 及び第 2 の相互作用成分が、誘導条件下で相互作用する誘導性ヘテロ二量体対、任意選択で ABI1 と PYL1 を形成する、請求項 5 に記載の転写活性化因子。

【請求項 7】

前記 DNA 標的化ドメインが、ジंकフィンガードメインを含む、請求項 1 ～ 6 のいずれか 1 項に記載の転写活性化因子。

【請求項 8】

前記エフェクタードメインが、配列番号 103、104、105、106、167、及び 185 の TAD のうちの任意の 1 つから選択される少なくとも 1 つの TAD、任意選択で、配列番号 90、91、46、47、101 ～ 106、110、116 ～ 119、156、157、159、162、165、166、167、及び 172 の TAD のうちの任意の 1 つから選択される少なくとも 1 つの TAD を含む、請求項 1 ～ 7 のいずれか 1 項に記載の転写活性化因子。

【請求項 9】

1 つ以上の核局在化シグナル (NLS)、任意選択で SV40 NLS をさらに含む、請求項 1 ～ 8 のいずれか 1 項に記載の転写活性化因子。

【請求項 10】

前記エフェクタードメインが、配列番号 121、123、125、127、129、131、133、135、174、176、178、180、または 182 のアミノ酸配列、またはその中の TAD に対して少なくとも 80%、85%、90%、95%、または 99% の配列同一性を含む、請求項 1 ～ 9 のいずれか 1 項に記載の転写活性化因子。

【請求項 1 1】

請求項 1 ～ 1 0 のいずれか 1 項に記載の転写活性化因子または請求項 1 ～ 1 0 のいずれか 1 項に記載のエフェクタードメインをコードする単離された核酸。

【請求項 1 2】

1 つ以上のプロモーター及び 1 つ以上の転写終結部位に作動可能に連結された、請求項 1 1 に記載の核酸を含む発現構築物。

【請求項 1 3】

請求項 1 1 に記載の核酸または請求項 1 2 に記載の発現構築物を含むベクターであって、任意選択で、前記ベクターがアデノウイルスまたはレンチウイルスベクターである、前記ベクター。

10

【請求項 1 4】

請求項 1 ～ 1 0 のいずれか 1 項に記載の転写活性化因子、請求項 1 1 に記載の核酸、請求項 1 2 に記載の発現構築物、または請求項 1 3 に記載のベクターを含む細胞。

【請求項 1 5】

転写活性化系であって、

- a) 請求項 1 ～ 1 0 のいずれか 1 項に記載の異種転写活性化因子を含み、前記 DNA 標的化ドメインが、CRISPR-Cas タンパク質及び
- b) 少なくとも 1 つの gRNA を含む、前記転写活性化系。

【請求項 1 6】

前記少なくとも 1 つの gRNA が、遺伝子の調節エレメントを標的とし、場合により、前記調節エレメントが、プロモーター領域、エンハンサー領域、または遠位調節部位である、請求項 1 5 に記載の転写活性化系。

20

【請求項 1 7】

細胞における標的遺伝子の転写を活性化する方法であって、前記方法が：

- a) 請求項 1 ～ 1 0 のいずれか 1 項に記載の転写活性化因子、請求項 1 1 に記載の核酸、請求項 1 2 に記載の発現構築物、または請求項 1 3 に記載のベクターを前記細胞に導入することと、
- b) 前記エフェクタードメインが前記標的遺伝子の転写を活性化するような適切な条件下で前記細胞を培養することと、を含む、前記方法。

【請求項 1 8】

前記 DNA 標的化ドメインが CRISPR-Cas タンパク質を含み、前記方法が、少なくとも 1 つの gRNA を前記細胞に導入することと、前記少なくとも 1 つの gRNA が前記 CRISPR-Cas タンパク質と結合して前記転写活性化因子を CRISPR 標的部位に誘導するような適切な条件下で前記細胞を培養することとをさらに含む、請求項 1 7 に記載の方法。

30

【請求項 1 9】

スクリーニングアッセイであって、前記アッセイが、

- a) 請求項 1 ～ 1 0 のいずれか 1 項に記載の転写活性化因子、請求項 1 1 に記載の 1 つ以上の核酸、請求項 1 2 に記載の 1 つ以上の発現構築物、または請求項 1 3 に記載の 1 つ以上のベクターを複数の細胞に導入することであって、前記 DNA 標的化ドメインが CRISPR-Cas タンパク質；複数の gRNA を含む、前記導入することか、または複数の gRNA を請求項 1 4 に記載の細胞集団に導入することであって、前記 DNA 標的化ドメインが CRISPR-Cas タンパク質を含む、前記導入することと；
- b) 前記 1 つ以上の gRNA が CRISPR-Cas タンパク質と結合し、エフェクタードメインが標的遺伝子の転写を活性化するように、前記転写活性化因子を CRISPR 標的部位に誘導するように、前記複数の細胞を培養することと；
- c) 任意選択で、ある量の試験薬物または毒素で処理することと；
- d) 任意選択で、前記複数の細胞を、gRNA の脱落または濃縮を可能にする期間にわたって培養することと；
- e) 前記複数の細胞またはそのサブセットを収集することと、を含む、前記スクリーニ

40

50

ングアッセイ。

【請求項 20】

前記方法が、前記複数の細胞またはそのサブセットにおいて過剰または過少に発現している 1 つ以上の gRNA を同定することをさらに含む、請求項 19 に記載のアッセイ。

【請求項 21】

請求項 1 ～ 10 のいずれか 1 項に記載の転写活性化因子、請求項 11 に記載の核酸、請求項 12 に記載の発現構築物、請求項 13 に記載のベクター、または請求項 14 に記載の細胞を含む組成物。

【請求項 22】

バイアルと、請求項 1 ～ 9 のいずれか 1 項に記載の異種転写活性化因子、請求項 11 に記載の核酸、請求項 12 に記載の発現構築物、請求項 13 に記載のベクター、請求項 14 に記載の細胞、または請求項 21 に記載の組成物と、任意選択で、誘導剤、gRNA、または gRNA 発現構築物のうちの 1 つ以上とを含む、キット。

10

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本出願は、2021 年 7 月 14 日に出願された米国特許出願第 63 / 221, 611 号の優先権の権益を主張し、この開示は、参照することによりその全体が本明細書に組み込まれる。

20

【0002】

本開示は、転写活性化のための試薬及び方法に関し、特に、標的化された転写活性化のための転写活性化因子における異種トランス活性化因子ドメインの使用に関する。

【0003】

導入

タンパク質コード遺伝子の転写は、配列特異的な様式で DNA に結合する転写因子 (TF)、プロモーターからの転写を開始する RNA ポリメラーゼ II 機構、ならびにクロマチン構造を調節し、TF と RNA pol II との間の架橋として作用する多様なクロマチン関連因子及び複合体の協調的な相互作用によって組織化される (Cramer, 2019)。ヒトゲノムは、転写調節の様々な段階に関与する数千のタンパク質をコードし、ChIP-seq などの方法の容易な利用可能性は、多様な条件において何百もの因子のゲノム結合部位を明らかにした (The ENCODE Project Consortium et al., 2020)。同時に、体系的な研究は、ヒト TF の約 3 / 4 の DNA 結合特異性を特徴付けまたは推測した (Badis et al., 2009; Jolma et al., 2013; Lambert et al., 2018; Najafabadi et al., 2015; Weirauch et al., 2014)。同様に、相互作用プロテオミクスアプローチは、多くのクロマチン関連タンパク質を明らかにし、ヒト細胞における転写調節複合体の組成を特徴づけた (Gao et al., 2012; Huttlin et al., 2020; Lambert et al., 2019; Li et al., 2015; Marcon et al., 2014; Mashitalir et al., 2018)。

30

40

【0004】

しかしながら、TF 及びクロマチン関連因子が転写の活性化または抑制を促進するか (または他の手段によってクロマチン状態を調節するか) どうか、及びどのように促進するかは、これらの方法によって与えられる限られた因果的洞察 (causal insight) に起因して、ほとんど不明なままであった。例えば、ChIP-seq 実験で観察されたゲノム結合部位の大多数は、転写事象と因果的に関連しない。すなわち、所与の転写調節因子のノックダウンまたはノックアウトは、調節因子が結合する多くの遺伝子の転写に影響を及ぼさない。一方、転写制御因子の配列ベースの注釈は、ほとんどの転写エフェクター機能が、フォールディングされ、保存されたタンパク質ドメインではなく、縮重

50

線状モチーフによってコードされるので、困難であった (Arnold et al., 2018; Erijman et al., 2020; Sigler, 1988; Staller et al., 2021)。

【0005】

活性化因子バイパスとしても知られる人工的な動員は、定義された文脈における多様なタンパク質の転写効果の特徴付けるための強力な方法である (Ptashne and Gann, 1997; Sadowski et al., 1988)。このアプローチでは、タンパク質またはその断片は、タンパク質を Gal4 または TetR などの十分に特徴付けられた TF の DNA 結合ドメインに融合することによって、レポーター遺伝子に異所的に動員される。定義された文脈は、複数の因子が協調して調節エレメントに結合し、因果的推論を妨害する、内在性遺伝子調節によって提起される課題を軽減する。人工的な動員は、従来、個々の転写調節因子における転写活性化因子またはトランス活性化ドメイン (TAD) を同定するために使用されてきた (Ptashne and Gann, 1997)。しかしながら、最近の研究は、個々にレポーター遺伝子に連結することによって、ショウジョウバエまたは酵母における調節因子の大きなコレクションの転写効果の特徴付けている (Keung et al., 2014; Stampfel et al., 2015)。アレイ型の拡張性が限られているため、これらの研究は、潜在的に新規な因子ではなく、既知の調節因子に焦点を当てた。さらに、古典的モデル生物は、エンハンサー (酵母) または DNA メチル化 (酵母及びショウジョウバエ) などの、哺乳動物における遺伝子発現に重要ないくつかの制御機構及び層を欠く。より最近では、Tyckoらは、注釈を付けたヒトタンパク質ドメインの転写活性化能を特徴付けるための偏りのないプールされたスクリーニング戦略を実施した (Tycko et al., 2020)。この研究は、新規の転写調節因子を同定する際の偏りのないアプローチの値、例えば、抑制の代わりに転写活性化におけるバリエーション K R A B ドメインの予想外の役割を強調した。しかし、ほとんどの転写活性化ドメインは無秩序領域によってコードされるので (Arnold et al., 2018; Dyson and Wright, 2005)、ドメインに集中したスクリーニングはトランス活性化因子のかなりの割合を失う可能性が高い。

【0006】

したがって、転写制御因子複合体の組成及びそれらのゲノム結合パターンの知識の増加にもかかわらず、TF 及び多様なクロマチン関連因子によって誘発される下流の効果の完全な理解は欠けている。

【発明の概要】

【0007】

本明細書に記載されるように、本発明者らは、ヒトタンパク質の転写調節の可能性を偏りのない様式で系統的に同定し、特徴付けるためのプラットフォームを確立した。プールされた形式で 13,000 個のタンパク質を超えてスクリーニングすることによって、数百の強力な活性化因子が同定され、その多くは以前に十分に注釈が付けられていないものであった。トランス活性化ドメインは、活性化ドメインの標準的な「酸性 blob」モデルに付着しない一部を含む、ヒットの間で系統的に明らかにされた。さらに、相互作用プロテオミクスを化学的阻害剤と組み合わせて、新規の転写活性化因子と公知の転写活性化因子の両方の補因子特異性を線引きし、実質的に同一の DNA 結合特異性を有する高度に関連する TF が、異なる補因子複合体を介して転写を活性化し得る方法を強調した。

【0008】

本発明者らは、本明細書において、ヒト細胞における転写活性化因子の第 1 の系統的スクリーニングを記載する。予想どおり、配列特異的転写因子及び他のクロマチン関連タンパク質が濃縮された数百の転写活性化因子が同定された。

【0009】

本明細書に示される結果は、非常に限られた数の TF のみが強力な転写活性化因子であることを示唆する。これは、ヒト TF の既知の DNA 結合特異性及びクロマチン占有に関

10

20

30

40

50

連して道理にかなうものである (Jolma et al., 2013; The ENCODE Project Consortium et al., 2020; Yan et al., 2013)。ほとんどのTFは、短い(約6~10bp)モチーフを認識し、ゲノム全体の数千の部位と結合する。しかし、ほとんどの結合事象は転写の結果に関連しない。DNA結合ドメインの限定された配列特異性に結合した強力な転写活性化ドメインは、広範囲の遺伝子の偽活性化を引き起こし、細胞適合性を妨害する可能性が高い。本明細書で同定された最も強い活性化因子の多くは、DNA結合転写因子ではなく、他のクロマチン関連タンパク質であった。

【0010】

新規で非常に強力なヒトトランス活性化ドメインの発見は、治療上の意味も有する。V P 1 6トランス活性化ドメインまたは三要素のV P R活性化因子などのウイルスタンパク質に由来する成分は、インビボで免疫応答を誘発し、診療所で副作用を引き起こし得る。完全ヒト成分から合成転写調節因子を設計することは、治療用途において有利であると予想される (Israni et al., 2021)。さらに、本明細書に示すように、複数の異なるヒトタンパク質由来の活性化ドメインを組み合わせることにより、ゲノムの高度に圧縮された領域においてさえ遺伝子を強固に上方制御することができる「スーパー活性化因子」を生成することができる。

【0011】

一態様は、異種転写活性化因子であって、DNA標的化ドメイン、任意選択で酵素的に不活性なCRISPR-CASタンパク質、ジンクフィンガーDNA結合ドメイン、tet-リプレッサー、または転写活性化因子様エフェクター (TALE) DNA結合ドメインと、

表1~6、任意選択で表2もしくは表6のいずれか1つに列挙されるトランス活性化ドメイン (TAD)、もしくはその機能的バリエーションから選択される少なくとも1つのTAD、または表1~6、任意選択で表1もしくは表3のいずれか1つに列挙されるTAD、もしくはそれらの任意の機能的バリエーションから選択される少なくとも2つのTAD、好ましくは表4もしくは表5もしくは表6に列挙されるTAD、もしくはそれらの機能的バリエーションから選択される少なくとも1つのTAD、を含む、エフェクタードメインと、を含む、

DNA標的化ドメインとエフェクタードメインが、作動可能に連結される、異種転写活性化因子を含む。

【0012】

一態様は、本明細書に記載のエフェクタードメインをコードする単離された核酸を含む。

【0013】

一態様は、本明細書に記載の異種転写活性化因子をコードする単離された核酸を含む。

【0014】

一態様は、1つ以上のプロモーターと1つ以上の転写終結部位に作動可能に連結された本明細書に記載の核酸を含む発現構築物を含む。

【0015】

一態様は、本明細書に記載の核酸または発現構築物を含むベクターを含み、任意選択で、ベクターはアデノウイルスまたはレンチウイルスベクターである。

【0016】

一態様は、本明細書に記載の転写活性化因子、核酸、発現構築物、またはベクターを含む細胞を含む。

【0017】

一態様は、本明細書に記載の異種転写活性化因子を含む転写活性化システムを含み、DNA標的化ドメインは、CRISPR-Casタンパク質と少なくとも1つのgRNAを含む。

【0018】

10

20

30

40

50

一態様は、細胞における標的遺伝子の転写を活性化する方法であって、該方法は、a) 細胞に、本明細書に記載の転写活性化因子、核酸、発現構築物、またはベクターを導入することと；b) エフェクタードメインが標的遺伝子の転写を活性化するような適切な条件下で、細胞を培養することと、を含む、方法を含む。

【0019】

一態様は、スクリーニング方法であって、該方法は、a) 複数の細胞に、本明細書に記載の転写活性化因子、1つ以上の核酸、1つ以上の発現構築物、または1つ以上のベクターを導入することであって、DNA標的化ドメインが、CRISPR-Casタンパク質と複数のgRNAを含む、導入することと；または、本明細書に記載の細胞の集団に複数のgRNAを導入することであって、DNA標的化ドメインが、CRISPR-Casタンパク質を含む、導入することと；b) 1つ以上のgRNAがCRISPR-Casタンパク質と結合し、転写活性化因子をCRISPR標的部位に誘導してエフェクタードメインが標的遺伝子の転写を活性化するように複数の細胞を培養することと、c) 任意選択で、ある量の試験薬物または毒素で処理することと；d) 任意選択で、gRNAの脱落または濃縮を可能にする期間、複数の細胞を培養することと；e) 複数の細胞またはそのサブセットを回収することと、を含む、方法を含む。

10

【0020】

一態様は、本明細書に記載の転写活性化因子、核酸、発現構築物、ベクター、または細胞を含む組成物を含む。

20

【0021】

一態様は、バイアルと、本明細書に記載の異種転写活性化因子、核酸、発現構築物、ベクター、細胞、または組成物と、任意選択で、誘導剤、gRNA、またはgRNA発現構築物のうちの1つ以上とを含むキットを含む。

【0022】

前項は、単なる例として提供され、本開示及び添付の特許請求の範囲の範囲を限定することを意図しない。本開示の組成物及び方法に関連する追加の目的及び利点は、特許請求の範囲、説明、及び実施例に照らして当業者によって理解されるであろう。例えば、本開示の様々な態様及び実施形態は、多くの組み合わせで利用されてもよく、そのすべては、本明細書によって明確に企図される。これらの追加の利点オブジェクト及び実施形態は、本開示の範囲内に明確に含まれる。本開示の背景を明らかにするために、特に場合によっては実施に関する追加の詳細を提供するために本明細書で使用する刊行物及び他の資料は、参照により組み込まれ、便宜上、添付の参照セクションに列挙される。

30

【0023】

本開示の特徴及び利点は、本開示の例示的な実施形態を示す添付図面と関連してなされる以下の詳細な説明を参照することにより明らかになるであろう。

【図面の簡単な説明】

【0024】

【図1A】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。ヒト細胞中の転写活性化因子を特徴付けるための化学的に誘導された二量体化系の概略図を示す。

40

【図1B】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。示された構築物をHEK293Tレポーター細胞にトランスフェクトし、アブシジン酸またはDMSOで48時間処理した場合の高GFP細胞のパーセントを示す。

【図1C】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。転写活性化因子のためのプールされたORFeomeスクリーニングの概略。

【図1D】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。48時間のABA処理後のプールされたORFeomeスクリーニングにおける高GFP細胞の濃縮。

【図1E】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。非選別ORFeomeと比較した高GFPプールにおけるORFの濃縮。

50

【図 1 F】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。陽性スクリーニングのヒットでの遺伝子オントロジーカテゴリーの濃縮。

【図 1 G】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。陽性スクリーニングのヒットでのInterPro domainの濃縮。

【図 1 H】転写活性化因子のためのプールされたORFeomeのスクリーニングを示す。スクリーニングのヒットでのCORUM複合体の濃縮。

【図 2 A】転写因子ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA結合特異性（配列ロゴとして示される）は、CisB由来である（Weirauch et al., 2014）。アスタリスクは、統計的に有意な活性化因子を示す（ $FDR < 0.05$ ）。ホメオボックスファミリータンパク質。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率（FDR）アプローチを用いて複数の仮説について補正した（Benjamini et al., 2006）。

10

【図 2 B】転写因子ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA結合特異性（配列ロゴとして示される）は、CisB由来である（Weirauch et al., 2014）。アスタリスクは、統計的に有意な活性化因子を示す（ $FDR < 0.05$ ）。フォークヘッドBoxタンパク質。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率（FDR）アプローチを用いて複数の仮説について補正した（Benjamini et al., 2006）。

20

【図 2 C】転写因子ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA結合特異性（配列ロゴとして示される）は、CisB由来である（Weirauch et al., 2014）。アスタリスクは、統計的に有意な活性化因子を示す（ $FDR < 0.05$ ）。Kruppel様因子。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率（FDR）アプローチを用いて複数の仮説について補正した（Benjamini et al., 2006）。

【図 2 D】転写因子ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA結合特異性（配列ロゴとして示される）は、CisB由来である（Weirauch et al., 2014）。アスタリスクは、統計的に有意な活性化因子を示す（ $FDR < 0.05$ ）。SRY-関連HMG-box（SOX）タンパク質。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率（FDR）アプローチを用いて複数の仮説について補正した（Benjamini et al., 2006）。

30

【図 2 E】転写因子ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA結合特異性（配列ロゴとして示される）は、CisB由来である（Weirauch et al., 2014）。アスタリスクは、統計的に有意な活性化因子を示す（ $FDR < 0.05$ ）。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率（FDR）アプローチを用いて複数の仮説について補正した（Benjamini et al., 2006）。ポリコム群RING Finger（PCGF）タンパク質。カノニカル（cPRC1）及び非カノニカル（ncPRC1）の複合体の組成を右に示す。

40

【図 2 F】転写因子ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA結合特異性（配列ロゴとして示される）は、CisB由来である（Weirauch et al., 2014）。アスタリスクは、統計的に有意な活性化因子を示す（ $FDR < 0.05$ ）。Spy1/RINGOファミリータンパク質。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率（FDR）アプロ

50

ーチを用いて複数の仮説について補正した (Benjamin et al., 2006)。

【図3A】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。TAD-seqプールアッセイの概略。

【図3B】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。既知のトランス活性化ドメインの例。ドメイン組織を上を示す。TAD-seqプロットは、高GFP集団または中GFP集団におけるRNAseqリードの濃縮倍率を示す。それぞれの円は、60-aaタイルの中心(30番目のアミノ酸)を示す。黒丸は、統計的に有意なヒットを示す。灰色のボックスは、以前に記載されたトランス活性化ドメインを示す。

【図3C】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。新規のトランス活性化ドメインの例。標識はパネルBと同様である。

【図3D】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。HOXA2のトランス活性化ドメインの位置。

【図3E】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。HOXA2で同定された活性化断片の配列。活性化断片は太字である。中GFP集団において濃縮された3つの断片すべてに共通する領域は、活性化断片の重複として示される。アンテナペディア様ヘキサペプチド配列の位置は、ヘキサペプチドとして示される。

【図3F】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。YAF2中のトランス活性化ドメインの位置。

【図3G】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。RYBP(PDB 3IXS)のYAF2_RYBPドメイン及びCBX7(PDB 3GS2)のCBX_Cドメインに結合したRING1B RAWULドメインの結晶構造。

【図3H】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。示されたタンパク質由来のYAF2_RYBPドメイン及びCBX_Cドメインを、転写活性について個々に試験した。アスタリスクは、統計的に有意な活性化因子を示す(FDR<0.05)。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamin、Krieger、及びYekutieliの偽発見率(FDR)アプローチを用いて複数の仮説について補正した(Benjamin et al., 2006)。右、RING1Bに対するドメインのインビトロ親和性(Wang et al., 2008, 2010)。統計学的有意性を、独立両側t検定を用いて計算した。

【図4A】転写活性化因子の補因子特異性を示す。示された転写調節因子の近接パートナーをBioID2で同定した。選択した補因子複合体の濃縮をヒートマップとして示す。平均スペクトル数(n+1)を、EGFP及びNanoLucベイトのバックグラウンドスペクトル数(n+1)に対して正規化した。

【図4B】転写活性化因子の補因子特異性を示す。(Li et al., 2015)のAP-MS研究に基づくフォークヘッド転写因子の活性化の相互作用パターン。

【図4C】転写活性化因子の補因子特異性を示す。左は、83個の転写調節因子の活性に対するA-485によるp300/CBP阻害の効果。公知のp300相互作用因子を黒い黒丸として示し、公知のNuA4相互作用因子を白抜き丸として示す。右に、p300相互作用因子は、A-485処理による他の転写調節因子よりも有意に影響を受ける。統計的有意性を、ダネットの多重検定補正を用いて一元配置ANOVAで計算した。

【図4D】転写活性化因子の補因子特異性を示す。左は、83個の転写調節因子の活性に対するJQ1によるBETプロモドメインタンパク質阻害の効果。公知のp300相互作用因子を黒い黒丸として示し、公知のNuA4相互作用因子を白抜き丸として示す。右に、p300相互作用因子は、A-485処理による他の転写調節因子よりも有意に影響を受ける。統計的有意性を、ダネットの多重検定補正を用いて一元配置ANOVAで計算し

10

20

30

40

50

た。

【図 4 E】転写活性化因子の補因子特異性を示す。翻訳活性化因子のクラスター化は、多様な阻害剤に対するそれらの感受性に基づく。クラスター化は、平均連鎖を使用したユークリッド距離を使用して実行された。既知の C B P / p 3 0 0 相互作用因子（黒丸）と N u A 4 相互作用因子の交差に濃縮されたクラスターが示される。

【図 5 A】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。低悪性度の子宮内膜間質肉腫に見られる J A Z F 1 - S U Z 1 2 融合体の概略図。 T A D - s e q により同定されるトランス活性化ドメインを示す。

【図 5 B】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。筋線維腫／筋周皮腫に見られる S R F - C 3 o r f 6 2 融合体の概略図。 T A D - s e q により同定されるトランス活性化ドメインを示す。

【図 5 C】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。 J A Z F 1 、 S U Z 1 2 、及び J A Z F 1 - S U Z 1 2 近接相互作用因子を B i o I D 2 で同定した。 J A Z F 1 - S U Z 1 2 は、 P R C 2 成分と N u A 4 成分の両方と相互作用し、超複合体を生成する。

【図 5 D】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。 S R F 、 C 3 o r f 6 2 、 C 3 o r f 6 2 - C t e r m 、及び S R F - C 3 o r f 6 2 近接相互作用因子を B i o I D で同定した。 S R F - C 3 o r f 6 2 融合体は、 C B P 及び p 3 0 0 と強固に相互作用する。

【図 5 E】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。 S R F - C 3 o r f 6 2 は強力な転写活性化因子である。示された P Y L 1 融合体を、ゲノムに組み込まれたレポーターの活性化について個々に試験した。

【図 5 F】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。 S R F - C 3 o r f 6 2 は、補因子の非存在下で血清応答エレメントレポーターの発現を活性化する。 3 × F L A G - V 5 で C 末端がタグ化された示された構築物を、 S R E - ホタルルシフェラーゼレポーター及び構成的 N a n o l u c レポーターと共に N I H 3 T 3 細胞に共トランスフェクトした。相対ルシフェラーゼ活性を測定して、それぞれの構築物の活性を評価した。

【図 5 G】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。ドキシサイクリン誘導性 S R F - C 3 o r f 6 2 - G F P または N a n o l u c - G F P を安定に発現する N I H 3 T 3 細胞を、ドキシサイクリンで 46 時間（最後の 22 時間は低血清条件（0.5 % F C S）である）処理した。遺伝子発現パターンを R N A - s e q によって分析した。有意に上方制御された（黒丸、上）遺伝子及び下方制御された（中の丸、下）遺伝子（絶対 log 2 倍率変化 > 1、F D R < 0.05）。 S R F / M R T F （上、ラベル付き丸）及び S R F / T C F （下、ラベル付き丸）の十分に特徴付けられた標的を示す。

【図 5 H】 S R F - C 3 o r f 6 2 融合体が、 S R F / M R T F 標的遺伝子の発現を促進する強力な p 3 0 0 依存性転写活性化因子を生成することを示す。 N a n o l u c - G F P 発現細胞と比較した、 S R F - C 3 o r f 6 2 - G F P 発現細胞の遺伝子セット濃縮分析。

【図 6 A】転写活性化因子のための O R F e o m e のスクリーニングを示す。プールされたプラスミド D N A 及び感染細胞における O R F e o m e にわたる配列決定リードの分布。

【図 6 B】転写活性化因子のための O R F e o m e のスクリーニングを示す。プラスミドプール及び感染細胞における O R F サイズの分布。

10

20

30

40

50

【図 6 C】転写活性化因子のための ORFeome のスクリーニングを示す。転写活性は、アブジン酸処理に依存する。ORFeome-PYL1 感染細胞を 100 μ M の ABA で処理し、高 GFP 細胞の画分を経時的にフローサイトメーターで測定した。

【図 6 D】転写活性化因子のための ORFeome のスクリーニングを示す。転写活性に対する ABA 濃度の効果。示された構築物でトランスフェクトされたレポーター細胞を、48 時間、増加する量の ABA で処理した。

【図 6 E】転写活性化因子のための ORFeome のスクリーニングを示す。ORFeome-PYL1 ライブラリを発現しない ABA 処理細胞では、高 GFP 集団は観察されない。

【図 6 F】転写活性化因子のための ORFeome のスクリーニングを示す。活性化スクリーニングのヒットでの相互作用ハブの濃縮。

【図 6 G】転写活性化因子のための ORFeome のスクリーニングを示す。活性化因子スクリーニングのヒットでの酵母ツーハイブリッド自己活性化因子の濃縮。

【図 6 H】転写活性化因子のための ORFeome のスクリーニングを示す。活性化スクリーニングで同定された転写活性化因子の個別の検証。示された構築物をレポーター細胞株にトランスフェクトし、高 GFP 細胞分画を、ABA による 48 時間の処理後にフローサイトメーターによって測定した。星印は、高 GFP 集団における統計的に有意な ABA 依存性増加を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。

【図 7 A】転写因子及びクロマチン関連タンパク質ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA 結合特異性 (配列ロゴとして示される) は、CisBP 由来である (Weirauch et al., 2014)。アスタリスクは、統計的に有意な活性化因子を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。ETS ファミリー TF。

【図 7 B】転写因子及びクロマチン関連タンパク質ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA 結合特異性 (配列ロゴとして示される) は、CisBP 由来である (Weirauch et al., 2014)。アスタリスクは、統計的に有意な活性化因子を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。Atonal 関連 bHLH 因子。

【図 7 C】転写因子及びクロマチン関連タンパク質ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA 結合特異性 (配列ロゴとして示される) は、CisBP 由来である (Weirauch et al., 2014)。アスタリスクは、統計的に有意な活性化因子を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。筋原性因子。

【図 7 D】転写因子及びクロマチン関連タンパク質ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA 結合特異性 (配列ロゴとして示される) は、CisBP 由来である (Weirauch et al., 2014)。アスタリスクは、統計的に有意な活性化因子を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の

10

20

30

40

50

仮説について補正した (Benjamini et al., 2006)。Twist / Hand bHLH 因子。

【図 7 E】転写因子及びクロマチン関連タンパク質ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA 結合特異性 (配列ロゴとして示される) は、CisBP 由来である (Weirauch et al., 2014)。アスタリスクは、統計的に有意な活性化因子を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。カゼインキナーゼ。

10

【図 7 F】転写因子及びクロマチン関連タンパク質ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA 結合特異性 (配列ロゴとして示される) は、CisBP 由来である (Weirauch et al., 2014)。アスタリスクは、統計的に有意な活性化因子を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。メディエーター成分。

【図 7 G】転写因子及びクロマチン関連タンパク質ファミリーの転写活性を示す。転写調節因子を、配列化された様式でレポーターの活性化について個々に試験した。DNA 結合特異性 (配列ロゴとして示される) は、CisBP 由来である (Weirauch et al., 2014)。アスタリスクは、統計的に有意な活性化因子を示す ($FDR < 5\%$)。統計学的有意性を、等分散前提の独立両側 t 検定で計算し、Benjamini、Krieger、及び Yekutieli の偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。一次活性化スクリーニングで同定された転写因子は、異所的に発現された場合にヒト iPS 細胞 (右) またはマウス ES 細胞 (左) をリプログラムできる因子が濃縮されている (Ng et al., 2021; Theodorou et al., 2009)。統計的有意性を、ウィルコクソンの順位和検定 (左) またはフィッシャーの直接検定 (右) を用いて計算した。

20

30

【図 8 A】転写因子の差次的活性が発現レベルによって説明されないことを示す。レポーター遺伝子発現とエフェクタータンパク質発現の両方を測定する転写活性化アッセイの概略図。

【図 8 B】転写因子の差次的活性が発現レベルによって説明されないことを示す。Kruppel 様因子のトランス活性化 (左) を、RFP 蛍光によって測定したそれぞれの因子の発現レベル (右) と比較した。バックグラウンド RFP 強度は、破線として示される。

【図 8 C】転写因子の差次的活性が発現レベルによって説明されないことを示す。フォークヘッド TF 活性。

【図 8 D】転写因子の差次的活性が発現レベルによって説明されないことを示す。ホメオドメイン TF 活性。

40

【図 8 E】転写因子の差次的活性が発現レベルによって説明されないことを示す。SOX TF 活性。

【図 9 A】TAD-seq がトランス活性化ドメインを同定することを示す。高 GFP 集団及び中 GFP 集団を、ABA を用いて 60 aa 断片をレポーターに 48 時間動員した後のフローサイトメトリーによって評価した。高 GFP 細胞及び中 GFP 細胞を FACS によって選別し、プールに濃縮された ORF を次世代シーケンシングによって同定した。

【図 9 B】TAD-seq がトランス活性化ドメインを同定することを示す。同定されたトランス活性化因子断片でのアミノ酸の濃縮及び枯渇を、ライブラリ中の不活性断片と比較した。太字 (bold) で示されるアミノ酸は、統計的に有意に濃縮または枯渇された。

50

【図9C】TAD-seqがトランス活性化ドメインを同定することを示す。活性断片での予測されるトランス活性化ドメインの濃縮。9aaTADは、「中程度にストリンジェントなパターン」を用いて9aaTAD予測ツール (<https://www.med.muni.cz/9aaTAD/>) で予測した。100%の確信的な一致のみが分析において考慮された。ADpredアルゴリズムは (Erijman et al., 2020) に記載されている。統計的有意性を、フィッシャーの直接検定を用いて計算した。

【図9D】TAD-seqがトランス活性化ドメインを同定することを示す。活性断片での予測されたTADは、不活性断片中のものよりも長い。統計学的有意性を、等分散前提の両側t検定を用いて計算した。

【図9E】TAD-seqがトランス活性化ドメインを同定することを示す。中GFPプールに対する高GFPプール中の断片の濃縮。有意なヒットを白抜き丸として示す。

【図9F】TAD-seqがトランス活性化ドメインを同定することを示す。トランス活性化断片の個別の検証。示されたTADをPYL1に融合させ、配列化された形式でレポーター細胞にトランスフェクトした。アスタリスクは、統計的に有意な活性化因子を示す (FDR < 5%)。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率 (FDR) アプローチを用いて複数の仮説について補正した (Benjamini et al., 2006)。

【図10A】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。既知のトランス活性化ドメインの例。ドメイン組織を上を示す。TAD-seqプロットは、高GFP集団 (黒丸) または中GFP集団 (白丸) におけるRNA-seqリードの濃縮倍率を示す。それぞれの円は、60-aaタイルの midpoint (30番目のアミノ酸) を示す。黒丸は、統計的に有意なヒットを示す。灰色のボックスは、以前に記載されたトランス活性化ドメインを示す。

【図10B】TAD-seqによるヒトタンパク質における、トランス活性化ドメインの系統的発見を示す。新規のトランス活性化ドメインの例。標識はパネルAと同様である。

【図11A】SPDYE4及びYAF2のトランス活性化ドメインを示す。5つのSpy1/RINGOファミリータンパク質の整列: SPDYE4 (配列番号142); SPDYE1 (配列番号143); SPDYE7P (配列番号144); SPDYE2 (配列番号145); 及びSPDYC (配列番号146)。TAD-seqによってスクリーニングされた4つの断片 (SPDYE4-2 (配列番号154); SPDYE4-3 (配列番号90); SPDYE4-4 (配列番号91); 及びSPDYE4-5 (配列番号155)) をアラインメントの上を示し、活性断片を太字で示す。破線のボックスは、推論された最小活性化ドメインを示す。最小活性化ドメインの領域は、転写を活性化しなかった唯一のSpy1/RINGOファミリーメンバーであるSPDYCにおいて保存されていないことに留意されたい。

【図11B】SPDYE4及びYAF2のトランス活性化ドメインを示す。YAF2 (配列番号147) 及びRYBP (配列番号148) のYAF2_RYBPドメイン、及びCBXファミリータンパク質 (配列番号149~153) のCBX_Cドメインのアラインメント。2つのβシートの位置を矢印で示す。

【図12A】転写活性化因子の相互作用ネットワークを示す。転写活性化因子のBioID2ネットワーク。ベイトタンパク質 (例えば、FAM90A1、SPDYE4、SS18L2、SOX7など) は、薄い灰色の長方形として示される。BAF複合体メンバー、p300/CBP、NuA4複合体メンバー、メディエーター成分、及びTFIID成分を示す。エッジの幅は、2つの複製の平均スペクトル数を示す。明確にするために、2つの高度に結合したブレイタンパク質 (ZNF518A及びZNF518B) を、この可視化から除去した。

【図12B】転写活性化因子の相互作用ネットワークを示す。転写活性化因子のAP-M Sネットワーク。パネルAと同様に標識。明確にするために、9つの高度に連結されたブレイタンパク質 (APEH、ACTC1、ALDH1L1、PSDM4、LSS、FLII、PACSIIN2、RPS3A、QPCTL) をこの可視化から除去した。

10

20

30

40

50

【図13A】83個の転写調節因子の転写活性に対する小分子阻害剤の効果を示す。フラボピリドールによるCDK9阻害の効果。公知のp300相互作用因子を黒い黒丸として示し、公知のNuA4相互作用因子を白抜き丸として示す。

【図13B】83個の転写調節因子の転写活性に対する小分子阻害剤の効果を示す。CX4945によるカゼインキナーゼ2阻害の効果。公知のp300相互作用因子を黒い黒丸として示し、公知のNuA4相互作用因子を白抜き丸として示す。

【図13C】83個の転写調節因子の転写活性に対する小分子阻害剤の効果を示す。AZ191によるDYRK1A/DYRK1B阻害の効果。公知のp300相互作用因子を黒い黒丸として示し、公知のNuA4相互作用因子を白抜き丸として示す。DYRK1A/DYRK1B相互作用因子DCAF7及びDCAF7相互作用因子NCKIPDが強調されている。

10

【図13D】83個の転写調節因子の転写活性に対する小分子阻害剤の効果を示す。AP-MS及びBioIDによって特徴付けられる転写調節因子に対するA-485によるp300阻害の効果。アスタリスクは、統計的に有意な活性化因子を示す(FDR<5%)。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率(FDR)アプローチを用いて複数の仮説について補正した(Benjamini et al., 2006)。

【図13E】83個の転写調節因子の転写活性に対する小分子阻害剤の効果を示す。AP-MS及びBioIDにより特徴付けられる転写調節因子に対するJQ1によるBETプロモドメイン阻害の効果。アスタリスクは、統計的に有意な活性化因子を示す(FDR<5%)。統計学的有意性を、等分散前提の独立両側t検定で計算し、Benjamini、Krieger、及びYekutieliの偽発見率(FDR)アプローチを用いて複数の仮説について補正した(Benjamini et al., 2006)。

20

【図14A】SRF-C3orf62による転写活性化を示す。SRF、C3orf62、C3orf62-Cterm、及びSRF-C3orf62相互作用因子をAP-MSで特徴付けた。SRF-C3orf62融合体は、CBP及びp300と強固に相互作用する。

【図14B】SRF-C3orf62による転写活性化を示す。Nanoluc-GFPを発現する細胞と比較した、SRF-GFP、C3orf62-GFP、またはSRF-C3orf62-GFPを発現するNIH3T3細胞における差次的に発現された遺伝子の分析。

30

【図14C】SRF-C3orf62による転写活性化を示す。SRF-C3orf62-GFPを発現するNIH3T3細胞における有意に上方制御された遺伝子及び下方制御された遺伝子の遺伝子オンロジー濃縮分析。

【図14D】SRF-C3orf62による転写活性化を示す。SRF-C3orf62-GFPによる有意に上方制御される遺伝子とSRF/MRTFまたはSRF/TCFの標的遺伝子との重複(以前に公開された)(Esnault et al., 2014; Gualdrini et al., 2016)。統計学的有意性を、超幾何学的分布検定を用いて計算した。

【図15】EGFPレポーターの転写活性に対するSPDYE4-CITED1-P65-HSF1融合体のそれぞれ個々の成分の最小活性化配列の効果を示す。すべての成分をPYL1二量体化ドメインに融合させ、1μMのアブシジン酸(ABA)を添加して48時間レポーターに動員した。それぞれの融合体の配列を配列番号121、123、127、129、131、133、及び135に示す。試験した個々のドメインのそれぞれの配列を、配列番号47、90、101~104、116、及び118に提供する。

40

【図16】2つ以上のトランス活性化ドメインを融合し、それらを同じレポーターに標的化する場合の例を示す。SPDYE4活性化ドメインの付加は、CITED1ドメイン単独の活性に対して顕著な改善を示した。これは、SPDYE4自体がレポーターに連結された場合の弱い活性とは対照的であり、CITED1活性化ドメインに融合した場合の相乗的な活性を示唆する。すべての融合構築物は、活性化ドメインを示し、示されるような

50

「全長CITED1／2」を除いて、全長タンパク質は示さず、これらは、活性をそれらの対応するトランス活性化ドメインと比較するために試験された。P300 coreは触媒ドメインEP300（アミノ酸1048～1664）である。レポーター細胞を1 μMのアブシジン酸（ABA）または同じ体積のDMSOで48時間処理した後、フローサイトメトリーのために回収した。エラーバーは、4回の反復からのS.D.を表す。

【図17】HEK293T細胞における、ヒトSPDYE4、CITED1、p65、及びHSF1タンパク質からの活性化ドメインの異なる組み合わせ及び配向を、EGFPレポーター（左）またはCD133遺伝子のプロモーターに連結する活性を示す。細胞を回収する前に、1 μMのアブシジン酸（ABA）を24時間または48時間添加することによって動員を誘導した。

【図18】多成分SPDYE4－CITED1－P65－HSF1（SCPH）活性化因子のそれぞれの部分を異なる活性化ドメインで置き換える効果を示す。SPDYE4活性化ドメインは、それを他の個々により強い活性化ドメインで置き換えるとSCPH活性化因子の活性が破壊されるので、最も不可欠であると思われる。一方、本発明者らのスクリーニングからCITED1成分を他の強力なヒトトランス活性化ドメインで置き換えることは、活性に対して有害な影響を及ぼさず、ZXDC及びC3orf62活性化ドメインでその効力を改善することさえ見られた。すべての構築物を、1 μMのアブシジン酸を48時間添加することによってEGFPレポーターに連結した。

【図19A】rTetRまたはdCas9ベースの動員システムと組み合わせて使用した場合の、トランス活性化ドメインまたはその断片の異なる組み合わせを含む117個の異なるエフェクタードメインの効果を示す。rTetR DNA標的化ドメインと組み合わせた117個のエフェクタードメインを使用する転写活性化。

【図19B】rTetRまたはdCas9ベースの動員システムと組み合わせて使用した場合の、トランス活性化ドメインまたはその断片の異なる組み合わせを含む117個の異なるエフェクタードメインの効果を示す。dCas9 DNA標的化ドメインと組み合わせた117個のエフェクタードメインを使用する転写活性化。

【図19C】rTetRまたはdCas9ベースの動員システムと組み合わせて使用した場合の、トランス活性化ドメインまたはその断片の異なる組み合わせを含む117個の異なるエフェクタードメインの効果を示す。rTetRとdCas9ベースの動員システム間の転写活性化の相関。

【発明を実施するための形態】

【0025】

以下は、本開示を実施する際に当業者を助けるために提供される詳細な説明である。別途定義されない限り、本明細書において使用される技術用語及び科学用語は、本開示の所属する技術分野における当業者によって一般に理解されるものと同一の意味を有する。本明細書の記載で使用される専門用語は、特定の実施形態を説明する目的のみのためのものであり、本開示を限定するようには意図されていない。本明細書において言及されるすべての刊行物、特許出願、特許、図面、及び他の参考文献は、参照によりその全体が明示的に組み込まれる。

【0026】

I. 定義

本明細書で使用される場合、以下の用語は、特に明記しない限り、以下に記載される意味を有し得る。しかしながら、当業者によって知られるかまたは理解される他の意味も可能であり、本開示の範囲内であることを理解されたい。本明細書において言及されるすべての刊行物、特許出願、特許、及び他の参考文献は、参照によりその全体が組み込まれる。矛盾する場合、定義を含む本明細書が優位となるであろう。さらに、物質、方法、及び実施例は、例示にすぎず、限定を意図しない。

【0027】

本明細書で使用される場合、「核酸」、「オリゴヌクレオチド」、「プライマー」という用語は、2つ以上の共有結合したヌクレオチドを意味する。文脈が明確に別段の指示を

10

20

30

40

50

示さない限り、この用語は一般に、非限定的に、デオキシリボ核酸（DNA）及びリボ核酸（RNA）を含み、これらは一本鎖（ss）または二本鎖（ds）であり得る。例えば、本開示の核酸分子またはポリヌクレオチドは、一本鎖及び二本鎖のDNA、一本鎖及び二本鎖領域の混合物であるDNA、一本鎖及び二本鎖のRNA、ならびに一本鎖領域と二本鎖領域の混合物であるRNA、一本鎖、より典型的には二本鎖、または一本鎖領域と二本鎖領域の混合物であり得るDNA及びRNAを含むハイブリッド分子から構成され得る。さらに、核酸分子は、RNAもしくはDNAまたはRNAとDNAの両方を含む三本鎖領域から構成され得る。本明細書で使用される場合、「オリゴヌクレオチド」という用語は、一般に、200塩基対までの長さの核酸を指し、一本鎖または二本鎖であってもよい。本明細書で提供される配列は、DNA配列またはRNA配列であり得るが、文脈が明確に別段の指示を示さない限り、提供される配列は、DNAとRNAの両方、ならびに相補的なRNA及びDNAの配列を包含することを理解されたい。例えば、配列5' - G A A T C C - 3' は、5' - G A A U C C - 3'、5' - G G A T T C - 3'、及び5' G G A U U C - 3' を含むと理解される。

10

20

30

40

50

【0028】

本明細書で使用される場合、「機能的バリエーション」という用語は、本明細書に開示されるポリペプチド分子と実質的に同じ機能を実質的に同じ方法で果たす、本明細書に開示されるポリペプチド配列の改変を含む。例えば、機能的バリエーションは、本明細書に記載のポリペプチドの活性断片、例えば、転写活性化活性及び／または補活性化因子相互作用を保持するN-及び／またはC-末端切り詰めを含み得る。機能的バリエーションは、1つ以上の置換されたアミノ酸を有するバリエーション、及び／または非改変または非バリエーションの配列に対して少なくとも最小配列同一性を保持するバリエーションを含み得る。例えば、機能的バリエーションは、1、2、3またはそれ以上のアミノ酸の置換を、10個のアミノ酸ごとに含み得る。例えば、機能的バリエーションは、本明細書に開示される配列に対して少なくとも80%、または少なくとも90%、または少なくとも95%の配列同一性を有する配列を含み得る。機能的バリエーションはまた、本明細書に開示される配列の保存的に置換されたアミノ酸配列を含み得る。置換的なアミノ酸バリエーションは、配列の少なくとも1つの残基が除去され、その場所に異なる残基が挿入されているものである。置換的なアミノ酸バリエーションの例は、保存的なアミノ酸置換である。例えば、転写活性化活性及び／または補活性化因子相互作用を保持する、例えば、本明細書に記載されるように同定され得る、最小断片を含む活性断片などの機能的バリエーションは、例えば、本明細書に記載される方法を使用して同定され得る。

【0029】

本明細書で使用される場合、「保存的なアミノ酸置換」は、あるアミノ酸が、タンパク質の所望の特性を失わない別のアミノ酸残基で置き換えられるものである。適切な保存的なアミノ酸置換は、類似の疎水性、極性、及びR鎖長を有するアミノ酸を互いに置換することによって行うことができる。保存的な置換の例としては、1つの非極性（疎水性）残基（例えば、アラニン、イソロイシン、バリン、ロイシン、もしくはメチオニン）の別の残基への置換、1つの極性（親水性）残基の別の残基への置換（例えば、アルギニンとリジンとの間、グルタミンとアスパラギンとの間、グリシンとセリンとの間）、1つの塩基性残基（例えば、リジン、アルギニン、もしくはヒスチジン）の別の残基への置換、または1つの酸性残基（例えば、アスパラギン酸もしくはグルタミン酸）の別の残基への置換が挙げられる。「保存的な置換」という語句はまた、そのようなポリペプチドが必要な活性を示すならば、非誘導残基の代わりに化学的に誘導残基または非天然アミノ酸を使用することを含む。

【0030】

本明細書で使用される場合、「異種転写活性化因子」または「転写活性化因子」という用語は、DNA標的化ドメインに作動可能に連結された、表2に列挙されるトランス活性化ドメイン（TAD）及びその機能的バリエーションから選択される少なくとも1つのTAD、または表1、表2、表3、表4、表5、及び／もしくは表6に列挙されるTAD及びそ

の任意の1つの機能的バリエーションから選択される少なくとも2つのTADを含むエフェクタードメインを含む、操作された融合タンパク質または操作された二量体を意味する。

【0031】

本明細書で使用される場合、「作動可能に連結される」という用語は、意図される様式で機能することを可能にする2つの成分の関係を指す。例えば、第1のポリペプチドは、共有結合によって（例えば、融合タンパク質として）、または1つ以上の相互作用成分を介して、第2のポリペプチドに作動可能に連結され得る。同様に、レポーター遺伝子がプロモーターに作動可能に連結される場合、プロモーターはレポーター遺伝子の発現を作動させる。

【0032】

転写活性化因子は、エフェクタードメイン及び／またはDNA標的化ドメイン及び／または標的DNAの間の機能的相互作用を提供する、相互作用系の1つ以上の相互作用成分をさらに含み得る。「相互作用成分」という用語は、相互作用システムの1つ以上の成分を包含するために本明細書で使用され、これらが一緒になって前記機能的相互作用を提供する。本明細書で使用される場合、「相互作用系」という用語は、共有結合もしくは非共有結合による相互作用、及び／または構成的もしくは誘導可能な相互作用を可能にする相互作用成分を包含することを意図する。そのような相互作用系は、例えば、ペプチドリンカー、任意選択でプロテアーゼ感受性ペプチドリンカー；1つ以上の二量体、三量体、またはより高次の多量体化成分、例えば相互作用ドメイン、任意選択で誘導可能な二量体、三量体、または多量体化成分、任意選択で誘導性相互作用ドメイン；及び／または転写活性化因子の細胞内局在を調節することができる1つ以上の成分を含み得る。相互作用系は、2つ以上の成分を含み得る。

【0033】

DNA標的化ドメイン及びエフェクタードメインは、例えば、単一のポリペプチド（例えば、融合タンパク質）のドメインとして共有結合され得るか、または例えば、特定の条件下で相互作用する相互作用ドメイン（例えば、二量体）などの相互作用成分によって連結され得る。したがって、異種転写活性化因子は、単一のポリペプチドを含み得るか、またはDNA標的化ドメイン及び第1の相互作用成分（例えば、二量体相互作用ドメイン）を含む第1のポリペプチド、ならびにエフェクタードメイン及び第2の相互作用成分（例えば、二量体相互作用ドメイン）を含む第2のポリペプチドを含み得、ここで、第1の二量体相互作用ドメインと第2の二量体相互作用ドメインは、例えば、特定の条件下で相互作用し得る。Sun Tag系（Tenenbaum et al., 2014）などのより高次の多量体化系もまた、本明細書において企図される。

【0034】

エフェクタードメイン及び／またはDNA標的化ドメイン及び／または標的DNAの間の相互作用は、種々の誘導性相互作用系を使用して制御され得る。例えば、エフェクタードメイン及びDNA標的化ドメインは、自己切断NS3プロテアーゼドメインなどのプロテアーゼ感受性リンカーによって連結され得、これは、グラゾプレビルなどのNS3阻害剤の存在下で安定化される。別の例では、DNA標的化ドメイン及び／またはエフェクタードメインの核への局在化は、局在化ドメイン、例えば、エストロゲン受容体リガンド結合ドメインバリエーションを使用するタモキシフェン調節核局在化などの相互作用成分によって制御され得る。さらなる例において、DNA標的化ドメインは、第1の相互作用ドメインなどの第1の相互作用成分に連結され得、エフェクタードメインは、第2の相互作用ドメインなどの第2の相互作用成分に連結され得、その結果、第1の相互作用ドメインと第2の相互作用ドメインが相互作用する。

【0035】

本明細書で使用される場合、「相互作用ドメイン」という用語は、第1のポリペプチド（例えば、第1の二量体相互作用ドメイン）中の配列モチーフを意味し、これは、第2のポリペプチド（例えば、第2の二量体相互作用ドメイン）中の配列モチーフを含む結合パートナーと相互作用して、第1のポリペプチドと第2のポリペプチドとを作動可能に連結

10

20

30

40

50

することができる。特に、この用語は、例えば適切な誘導条件下で二量体化するヘテロ二量体対と一緒に形成する第1または第2の相互作用二量体ドメインを包含することを意図する。他の相互作用ドメインが具体的に企図され、所望の特性に応じて当業者によって同定され得る。適切な誘導性相互作用ドメイン対としては、非限定的に、例えばラパマイシンまたはAP21967で誘導することができるFKBP／FRB（FK506結合タンパク質／FKBPラパマイシン結合）；例えばアブシジン酸で誘導することができるPYL／ABI；例えばギブベレリンまたはギブベレリン酸で誘導することができるGID1／GAI；ならびに例えば青色光及び／または温度で誘導することができるpMag／nMagが挙げられる。

【0036】

本明細書で使用される場合、「DNA標的化ドメイン」は、DNA結合条件下でDNAに結合し、それによってポリペプチドを前記DNAに標的化するポリペプチドドメインを指す。DNA標的化ドメインは、任意の適切なDNA結合ドメインであってよく、例えば、酵素的に不活性な配列特異的DNA標的化タンパク質、例えば、CRISPR-Casタンパク質（例えば、dCas9、dCas12、もしくは他のCasファミリータンパク質）、ジンクフィンガーDNA結合ドメイン、転写活性化因子様エフェクター（TALE）DNA結合ドメイン、プロモドメイン、クロモドメイン、Tudorドメイン、WD40ドメイン、PHDドメイン、PWWPドメイン、または真核生物もしくは原核生物由来の他のDNA結合ドメイン（DBD）（例えば、フォークヘッド、塩基性ヘリックスルーパーヘリックス、ロイシンジッパー、ホメオドメイン、核ホルモンレセプター、もしくはtetリプレッサー）、またはそれらのバリエーションであり得る。DNA標的化ドメインは、配列特異的な様式でDNAに結合し得る（例えば、Casファミリータンパク質、ジンクフィンガーDNA結合ドメイン、TALE DNA結合ドメイン）、または特異的クロマチン修飾（例えば、プロモドメイン（アセチル化ヒストンの場合）またはクロモドメイン、Tudorドメイン、WD40ドメイン、PHDドメイン、PWWPドメインなど（メチル化ヒストンの場合））に結合し得る。DNA標的化ドメインは、天然の（例えば、非操作の）DNA結合ドメイン、例えば、天然に存在する（例えば、内在性の）転写因子に見出されるDNA結合ドメインであり得るか、またはDNA標的化ドメインは、例えば、カスタム配列特異性（例えば、非操作のDNA結合ドメインとは異なる配列特異性）または改変されたDNA結合親和性を提供するように改変され得る。カスタムDNA結合特異性を提供するために、例えば、ジンクフィンガーDNA結合ドメイン及びTALE DNA結合ドメインを操作する方法は、当該技術分野で公知である（例えば、Maeder et al., 2008及びSanjana et al., 2012.）。酵素的に活性なCas9は、抑制をもたらす場合、例えば、ガイドが切り詰められたガイドである場合にも使用することができる（例えば[24]を参照）。DNA標的化ドメインは、例えば、ジンクフィンガーDNA結合ドメイン及びTALE DNA結合ドメインの場合、固有の標的配列特異性を有し得るか、または標的配列特異性は、例えば、CRISPR-Casタンパク質の場合、ガイドRNAなどの追加の配列特異的因子によって媒介され得る。適切なDNA結合条件は、DNA標的化ドメインに依存し、例えば、tetリプレッサーの場合にはテトラサイクリン、またはCasファミリータンパク質の場合にはガイドRNAなどの追加の因子の存在を含み得る。

【0037】

本明細書で使用される場合、「エフェクタードメイン」という用語は、本明細書に記載の少なくとも1つのトランス活性化ドメイン（TAD）、例えば、表1～5に列挙されるTAD及びその機能的バリエーション、例えばその活性断片を含むポリペプチドドメインを指す。任意選択で、エフェクタードメインは、2つ以上、例えば、本明細書に記載の2つ、3つ、4つ、またはそれ以上のトランス活性化ドメインを含み得る。本明細書に記載の活性化因子において、活性断片は、約15アミノ酸、約20アミノ酸、約30アミノ酸、約40アミノ酸、約50アミノ酸、約60アミノ酸、約70アミノ酸、または15～70アミノ酸の間の任意の数、または70を超えるアミノ酸であり得る。例えば、HSF1に関

10

20

30

40

50

して、活性断片は、H S F 1 のアミノ酸 4 0 6 ~ 4 2 0 に対応する G F S V D T S A L L D L F S P (配列番号 1 0 4) を含み得る。したがって、例として、H S F 1 の活性断片は、H S F 1 のアミノ酸 4 0 1 ~ 4 2 7 を含み得る。他の T A D の活性断片は、任意の適切な方法によって、例えば、本明細書に記載の方法を使用して同定され得る。

【0038】

異種転写活性化因子は、エフェクター N 末端融合体または C 末端融合体であり得、例えば、融合体の順序は、エフェクタードメイン-DNA 標的化ドメインまたは DNA 標的化ドメイン-エフェクタードメインであり得る (例えば、[25]、[26]、[27]、及び [28] を参照されたい)。エフェクタードメインは、リンカーを介して DNA 標的化ドメインに融合され得る。同様に、2 つ以上の T A D は、1 つ以上のリンカーによって一緒に融合され得る。例えば、グリシン及びグリシンセリンリンカーを使用することができる。実施例に記載の転写活性化因子は、種々のグリシンセリンリンカー、例えば、S G G S G G S (配列番号 6)、G G S、S G G S (配列番号 7)、及び / または G S G S G S (配列番号 8) を使用した。他のリンカー、例えば I N S R S S G S (配列番号 9) も使用できる。

【0039】

本明細書で使用される場合、「C R I S P R-C a s」または「C a s」という用語は、RNA に結合し、それが結合する RNA によって特異的な DNA 配列に標的化される、C R I S P R クラスター化等間隔短鎖回文リピート-C R I S P R 関連 (C R I S P R-C a s) タンパク質を指す。C R I S P R-C a s は、クラス II 単量体 C a s タンパク質、例えば、C a s 9 などの II 型 C a s である。C a s 9 タンパク質は、S t r e p t o c o c c u s p y o g e n e s、F r a n c i s e l l a n o v i c i d a、A . N a e s u l n d i i、S t a p h y l o c o c c u s a u r e u s、または N e i s s e r i a m e n i n g i t i d i s 由来の C a s 9 であり得る。任意選択で、C a s 9 は S . p y o g e n e s 由来である。C a s タンパク質はまた、例えば、A c i d a m i n o c o c c u s s p .、L a c h n o s p i r a c e a e b a c t e r i u m、または F r a n c i s e l l a t u l a r e n s i s 由来の C a s 1 2 a (例えば、d C a s 1 2 a) (これらは d C a s バリエーションとして機能することが示されている) でもあり得、C a s Φ (C a s 1 2 j) 及び C a s X (C a s 1 2 e) も使用され得る。

【0040】

本明細書で使用される場合、「d C a s 9」という用語は、DNA エンドヌクレアーゼ活性を欠くが標的 DNA 結合活性を保持する、酵素的に不活性な (または死んだ) C a s 9 を指す。例えば、d C a s 9 は、R u v C 1 及び H N H ヌクレアーゼドメイン中の C A S 9 及び D 1 0 A / H 8 4 0 A 変異の配列を含む。任意選択で、d C a s 9 は、配列番号 1 によってコードされるタンパク質と少なくとも 8 0 %、少なくとも 9 0 %、少なくとも 9 5 %、少なくとも 9 9 %、または 1 0 0 % の配列同一性を有するアミノ酸配列を含み、D 1 0 A / H 8 4 0 A 変異を含み、C a s 9 標的 DNA 結合活性 (例えば、g R N A と標的部位との結合) を保持するタンパク質である。同様に、d C a s 1 2 a は、酵素的に不活性な C a s 1 2 a を指す。

【0041】

本明細書で使用される場合、「ガイド RNA」、「ガイド」、または「g R N A」という用語は、特異的な DNA 配列とハイブリダイズし、最小限にスペーサー配列を含む RNA 分子を指す。ガイド RNA は、C R I S P R-C a s タンパク質に結合するタンパク質結合セグメントをさらに含み得る。特異的な DNA 配列とハイブリダイズするガイド RNA の部分は、本明細書では、核酸標的化配列またはスペーサー配列と呼ばれる。ガイドのタンパク質結合セグメントは、例えば、t r a c r R N A 及び / または直接反復を含み得る。「ガイド」または「ガイド RNA」という用語は、文脈に従って、スペーサー配列のみ、またはスペーサー配列とタンパク質結合セグメントを含む RNA 分子を指し得る。ガイド RNA は、対応する DNA 配列によって表すことができる。このガイドは、例えば、酵素が C a s 9 である場合、[24] に記載されるように、標的部位に対して 1 5 以下の

ヌクレオチドの相補性を含む切り詰められたガイドであり得る。例えば、Cas9が切り詰められたガイドと相互作用する場合、Cas9のDNA結合能力は無傷のままであり、その核酸分解活性は除去される。Cas結合能力を維持する任意の長さのガイドを使用することができる。

【0042】

本明細書で使用される場合、「スペーサー」または「スペーサー配列」という用語は、標的配列またはその部分とRNA-DNA二本鎖を形成するか、または形成することができるガイドの部分を目指す。スペーサー配列は、特異的CRISPR標的配列に相補的であるかまたは対応し得る。スペーサー配列のヌクレオチド配列は、CRISPR標的配列を決定し得、所望のCRISPR標的部位を標的化するように設計または構成され得る。

10

【0043】

本明細書で使用される場合、「tracrRNA」という用語は、例えば、Cas9などのCRISPR-Casタンパク質と相互作用し得、ガイドRNAに接続され得るか、またはその部分を形成し得る「トランスコードされたcrRNA」を指す。tracrRNAは、例えばS. pyogenes由来のtracrRNAであってもよい。tracrRNAは、例えば、5'-gttttcagagcctatgctggaaacagcattagcaagttgaaataaggcctagtcctgttatcaacttgaaaagtggtgacccgagtcgggtgc-3'（配列番号2）の配列を有し得る。他のtracrRNAも使用され得る。適切なtracrRNAは、当業者によって同定され得、例えば、5'-GTTTTTAGAGCTAGAAATAGCAAGTTTAAAT
AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCGAGT
CGGTGC-3'（配列番号3）、または5'-GTTTCAGAGCTACAGCAG
AAATGCTGTAGCAAGTTGAAAT-3'（配列番号4）を含む。

20

【0044】

本明細書で使用される場合、「CRISPR標的部位」または「CRISPR-Cas標的部位」という用語は、適切な条件下で活性化CRISPR-Casタンパク質（例えば、ガイドRNAに結合したdCas9などのCRISPR-Casタンパク質）が結合する核酸を意味する。CRISPR標的部位は、プロトスペーサー隣接モチーフ（PAM）及びCRISPR標的配列（すなわち、活性化CRISPR-Casタンパク質が結合されるガイドのスペーサー配列に対応する）を含む。CRISPR標的配列に対するPAMの配列及び相対位置は、CRISPR-Casタンパク質のタイプに依存する。例えば、Cas9またはdCas9のCRISPR標的部位は、5'から3'に向けて、15~25、16~24、17~23、18~22、または19~21ヌクレオチド、任意選択で20ヌクレオチド標的配列、続いて配列NGGを有する3ヌクレオチドPAMを含み得る。したがって、Cas9標的部位は、配列5'-N₁NGG-3'を有し得、ここで、N₁は、15~25、16~24、17~23、18~22、または19~21ヌクレオチド長、任意選択で20ヌクレオチド長である。

30

【0045】

CRISPR標的部位は、任意の適切なゲノム遺伝子座にあり得る。例えば、CRISPR標的部位は、プロモーター、エンハンサー、3'UTR、または他の調節エレメント内に、遺伝子、任意選択でイントロンまたはエクソン内に、非コードRNAに対応する遺伝子座内に、または遺伝子間領域内にあり得る。任意選択で、CRISPR標的部位は、プロモーターまたはエンハンサー内にある。

40

【0046】

細胞の核に位置する標的DNAは、核に入ることができる転写活性化因子を必要とする。したがって、転写活性化因子は、核局在化され得る、及び／または例えば、1つ以上の核局在化シグナル（NLS）、任意選択で1つ以上のSV40 NLSを含み得る。任意選択で、転写活性化因子は、2つ以上のNLSを含む。任意選択で、転写活性化因子は、1つ以上のN末端NLS、1つ以上のC末端NLS、1つ以上の内部NLS、または1つ以上のN末端、1つ以上のC末端NLS、及び／または1つ以上の内部NLSを含み得る

50

。他の構成も具体的に企図される。一実施形態では、NLSは、配列PKKKRKV（配列番号22）を有するSV40 NLSである。一実施形態では、NLSは、INSRSSGS（配列番号9）などのN末端及び／またはC末端リンカーをさらに含み、任意選択で、配列INSRSSGSPKKRKVGSS（配列番号141）を有する。

【0047】

転写活性化因子は、タグで標識することもできる。例えば、適切なタグとしては、Myc、FLAG、HA、V5、ALFA、T7、6×His、VSV-G、S-タグ、AviTag、StreptTag II、CBP、GFP、mCherryが挙げられるが、これらに限定されない。標識は、N-末端で、C-末端で、またはDNA標的化ドメインとエフェクタードメインとの間などの異種転写活性化因子の2つの成分の間で融合され得る。

10

【0048】

一定の範囲の値が提供されるとき、文脈が他を明確に指示しない限り、その範囲の上限と下限との間、及び、その述べられた範囲内の他の述べられた値または中間の値における、それぞれの中間の値は、その下限の単位の10分の1まで本記載に包含される。任意の下限から任意の上限までの範囲が企図される。独立してより小さな範囲に含まれ得るこれらのより小さい範囲の上限及び下限はまた、述べられた範囲内の任意の具体的に除外された限定を有することを条件として、本記載に包含される。述べられる範囲が一方または両方の限界を含む場合、それらの含まれた限界の一方または両方を除外する範囲はまた、本記載に含まれる。

20

【0049】

本明細書及び添付の特許請求の範囲において使用されるとき、単数形「a」、「an」、及び「the」は、文脈上、明らかに別段に解される場合を除き、複数の指示対象を含むことに留意されたい。

【0050】

本明細書の詳細な説明及び特許請求の範囲の中のすべての数値は、示された値の「約」または「およそ」によって修正され、当業者によって予想される実験誤差及び変動を考慮に入れる。

【0051】

本明細書及び特許請求の範囲で本願で使用される場合、「及び／または（and/or）」という語句は、そのように結合された要素、すなわち、ある場合には結合的に存在し、他の場合には分離的に存在する要素の「いずれかまたは両方」を意味すると理解されるべきである。「及び／または」で列挙される複数の要素は、同じ様式、すなわち、そのように結合された要素の「1つ以上」であると解釈されるべきである。「及び／または」節によって具体的に同定される要素以外の他の要素は、具体的に同定される要素に関連するかどうかにかかわらず、任意選択で存在し得る。

30

【0052】

本明細書及び特許請求の範囲で本願で使用されるとき、「または」は、上記で定義された「及び／または」と同じ意味を有すると理解されるべきである。例えば、リスト内の項目を区切る場合、「または」あるいは「及び／または」は包括的であると解釈されるものとする。すなわち、要素の数またはリストの少なくとも1つを含むが、1つを超えるもの、及び、任意選択で、追加のリストされていない項目を含むと解釈される。「～の1つのみ」または「～の正確に1つ」などの反対に明確に示される用語のみ、または特許請求の範囲で使用される場合、「～からなる」は、要素の数またはリストの正確に1つの要素を含むことを指す。一般的に、本明細書で使用される場合、「または」という用語は、「いずれか」、「～のうちの1つ」、「～のうちの1つのみ」、または「～のうちの正確に1つ」のような、排他性の用語が先行するとき、排他的代替物（すなわち、「一方または他方ではなく、両方」）を示すものとしてのみ解釈されるべきである。

40

【0053】

特許請求の範囲及び本明細書の上記において、すべての移行句、例えば、「～を含む（

50

comprising)」、「～を含む(including)」、「～を担持する」、「～を有する」、「～を含む(containing)」、「～を含む(involving)」、「～を保持する」、「～から構成される」などは、オープンエンドである、すなわち、～を含むがそれらに限定されないと理解される。「～からなる」及び「本質的に～からなる」という移行句のみが、それぞれクローズの移行句または半クローズの移行句となる。

【0054】

本明細書の明細書及び特許請求の範囲で使用される場合、1つまたは複数の要素のリストに関連する「少なくとも1つ」という語句は、要素のリストの要素のうちのいずれかの1つ以上から選択される少なくとも1つの要素を意味すると理解されるべきであるが、要素のリスト内に具体的にリストされているそれぞれ及びすべての要素の少なくとも1つを必ずしも含む必要はなく、要素のリスト内の要素の組み合わせを除外するものではない。この定義はまた、「少なくとも1つ」という語句が参照する要素のリスト内で具体的に同定される要素以外の要素が、具体的に同定される要素に関連するかどうかにかかわらず、任意選択で存在し得ることを可能にする。

10

【0055】

同様に、要素の群に関する「1つ以上」という語句は、記載された群の少なくとも1つのメンバーを含むが、必ずしも記載された群のメンバーのそれぞれのうちの1つを含む必要がないということが本明細書において具体的に企図される。例えば、要素がA、B、及び／またはCの群の1つ以上を含む場合、要素は、A；B；C；A及びB；A及びC；B及びC；またはA、B、及びCを含み得る。例えば上記の例を参照して、群に具体的に列挙されていない追加のメンバーも存在し得、要素は、列挙されていないメンバーDをさらに含み得、したがって、A及びD；B及びD；A、C、及びDなどを含み得る。

20

【0056】

本明細書で使用される場合、「約」という用語は、参照されている数のプラスまたはマイナス10%～15%、5～10%、または任意選択で約5%を意味する。

【0057】

1つを超えるステップまたは動作を含む本明細書に記載される特定の方法において、本方法のステップまたは動作の順序は、文脈が他の意味を示さない限り、本方法のステップまたは動作が列挙される順序に必ずしも限定されないことも理解されたい。さらに、特定のセクションに記載される定義及び実施形態は、当業者によって理解されるように、それらが好適である、本明細書に記載される他の実施形態に適用可能であることが意図される。例えば、以下の文章では、本開示のさまざまな態様がより詳細に定義される。そのように定義されたそれぞれの態様は、逆に明確に示されない限り、任意の他の態様または複数の態様と組み合わせられてもよい。特に、本明細書に記載される任意の特徴は、本明細書に記載される任意の他の特徴または複数の特徴と組み合わせられてもよい。

30

【0058】

本明細書に記載の材料及び方法と同様もしくは同等の任意の材料及び方法を本開示の実施または試験において使用することもできるが、以下の材料及び方法がここに記載される。

40

【0059】

II. 材料及び方法

異種転写活性化因子を生成するためにDNA標的化ドメインに作動可能に連結され得、例えば、治療目的のために、内在性遺伝子を含む所望の遺伝子の遺伝子発現を活性化するために使用され得る、1つ以上のトランス活性化ドメイン(TAD)及びそれらの組み合わせを含む、異種転写活性化因子の集合体が本明細書に記載される。本明細書に記載の異種転写活性化因子において、エフェクタードメインは、融合構築物のゲノム中の任意の遺伝子座への結合を指示することができるDNA標的化ドメインに作動可能に連結される。実施例に示されるように、表1に列挙されるTADのいずれか1つ、その活性断片、例えば、配列番号100～104、107～115、118～119、156～160、16

50

2、164、及び166～185のもの、ならびに例えば、表3に列挙されるような2つ以上のTADまたはその活性断片の組み合わせを含むエフェクタードメインと機能的に関連するdCas9またはrTetRを含む異種転写活性化因子を使用して、標的遺伝子の転写を活性化することができる。TADまたはその活性断片は、TADまたは活性断片の末端上に、リンカー及び／または追加の天然配列、例えば、1、2、3、4、5、6、7またはそれ以上、例えば、5まで、10まで、20まで、30まで、または40まで（またはその間の任意の数）のN末端及び／またはC末端のアミノ酸をさらに含み得る。例えば、表3中の「ZXDCショート」と標識されたTAD（配列番号107）は、ZXDC-12とZXDC-13の両方に見出される40アミノ酸断片を含み、ZXDC-12天然配列のさらなる6個のN末端アミノ酸及びZXDC-13天然配列のさらなる5個のC末端アミノ酸を含む。一実施形態では、活性断片は、同定された断片よりも短くてもよい。例えば、活性断片は、例えばZXDC-12とZXDC-13の両方に見出される40アミノ酸断片の20アミノ酸または25アミノ酸、任意選択で例えば配列番号107の内部部分であってもよい。より短い活性断片は、例えば、HSF1-20とHSF1-21の両方に見出される20アミノ酸断片を含む配列番号119によって例示される。活性断片は、本明細書の他の箇所に記載されるように同定することができる。例えば、最小断片は、活性断片を比較することによって同定することができる。例えばATF6について、表5に示される重複断片はHRLDEDWD S A L F A E L G Y F T D T D E T D E L Q L E A N E T Y E N D F D N Lであり、KLF7については、重複断片はY F S A L P S L E P T T W Q Q T T L E R A Y Q T E P R R I S E T F G E D L D Cである。

10

20

【0060】

したがって、本開示の一態様は、DNA標的化ドメインと、表1～6、任意選択で表1または任意選択で表2または表6のいずれか1つに列挙されるTADのいずれか1つから成る群から選択される少なくとも1つのTAD、それらの活性断片、及びそれらの組み合わせ、例えば、表1または表3に列挙されるTADまたはその機能的バリエーションから選択される少なくとも2つのTAD、好ましくは、表4または表5または表6に列挙されるTADから選択される少なくとも1つのTAD、及び／またはそれらの機能的バリエーションを含むエフェクタードメインとを含む異種転写活性化因子を含み、ここで、DNA標的化ドメイン及びエフェクタードメインは、作動可能に連結される。

【0061】

一実施形態では、少なくとも1つのTADは、表1から選択される。

30

【0062】

一実施形態では、少なくとも1つのTADは、表2から選択される。

【0063】

一実施形態では、少なくとも1つのTADは、表3から選択される。

【0064】

一実施形態では、少なくとも1つのTADは、表4から選択される。

【0065】

一実施形態では、少なくとも1つのTADは、表5から選択される。

【0066】

一実施形態では、少なくとも1つのTADは、表6から選択される。

40

【0067】

少なくとも2つのTADまたはその機能的バリエーションは、任意の表、例えば、表3及び表4のそれぞれから1つ、2つ以上、例えば、表3などからの3つまたは4つから選択され得ることが理解される。グループ化は、表1～6のいずれかに記載されるTADの任意のサブ組合せを含むことができ、例えば、1つ以上のTADを除外することができることも企図される。

【0068】

DNA標的化ドメイン及びエフェクタードメインは、例えば、単一のポリペプチドのドメインとして、共有結合によって作動可能に連結され得、及び／または1つ以上の相互作

50

用成分、例えば、相互作用ドメインを介して作動可能に連結され得、及び／または特定の条件下で相互作用し得る。したがって、一実施形態では、異種転写活性化因子は、単一のポリペプチドである。別の実施形態では、異種転写活性化因子は、一对の（すなわち、第一及び第二の）相互作用ドメイン、任意選択で二量体相互作用ドメイン、任意選択で適切な条件下で二量体化する一对の誘導性二量体相互作用ドメインをさらに含む。例えば、異種転写活性化因子は、DNA標的化ドメイン及び第1の二量体相互作用ドメイン、任意選択で誘導性二量体化ドメインを含む第1のポリペプチドと、エフェクタードメイン及び第2の二量体相互作用ドメイン、任意選択で誘導性二量体化ドメインを含む第2のポリペプチドとを含むことができ、ここで、第1の二量体相互作用ドメインと第2の二量体相互作用ドメインは相互作用し、任意選択で、第1の誘導性二量体化ドメインと第2の誘導性二量体化ドメインは、1つ以上の誘導剤の存在下で相互作用する。

10

【0069】

実施例に示すように、ABI1及びPYL1を含む異種転写活性化因子の二量体化は、アブシジン酸の添加により誘導することができる。したがって、一実施形態では、転写活性化因子は、誘導剤の存在下で誘導可能な転写活性化を提供する第1及び第2の誘導性二量体化ドメインを含む。当業者は、一緒に使用され得る適切な誘導性二量体化ドメインを容易に同定及び選択することができる。任意の適切な誘導性二量体化ドメインが使用され得、例えば、ABI1とPYL1の二量体化は、アブシジン酸の添加により誘導され得る。他の誘導性の系としては、ラパマイシン、ジベレリン酸／ジベレリン、及びスプリットdCas9ベースの系による誘導に基づくものが挙げられる。例えば、GID1とGAIの二量体化は、ギベレリンによって誘導することができ、FKBPとFRBの二量体化は、ラパマイシンまたはその類似体、例えばラパログによって誘導することができる。SunTag系（Tenenbaum et al., 2014）などのより高次の多量体化系もまた、本明細書において企図される。

20

【0070】

DNA標的化ドメインとエフェクタードメインとの間の相互作用はまた、他の誘導系を使用して制御され得る。他の系（二量体化に依存しない）は、グラゾプレビル誘導性安定化（Tague et al., 2018）またはエストロゲン受容体リガンド結合ドメインバリエーションを用いるタモキシフェン調節核局在化を含む。グラゾプレビル誘導性安定化の場合、DNA標的化ドメインとエフェクタードメインは、自己切断NS3プロテアーゼドメインによって連結される。グラゾプレビル（NS3活性を阻害する）の存在下でのみ、DNA標的化ドメインとエフェクタードメインは、一緒に留まり、遺伝子発現を調節する。

30

【0071】

DNA標的化ドメインは、種々のDNA結合ドメイン、例えば、ジンクフィンガーDNA結合ドメイン、転写活性化因子様エフェクター（TALE）DNA結合ドメイン、dCas9、dCas12もしくは他のCasファミリータンパク質、または真核生物もしくは原核生物由来の他のDNA結合ドメイン（例えば、フォークヘッド、塩基性ヘリックスルーパーヘリックス、ロイシンジッパー、ホメオドメイン、核ホルモン受容体、もしくはtet-リプレッサー）、またはそれらのバリエーションから選択することができる。DNA標的化ドメインは、天然の（例えば、非操作の）DNA結合ドメイン、例えば、天然に存在する（例えば、内在性の）転写因子に見出されるDNA結合ドメインであり得るか、またはDNA標的化ドメインは、例えば、カスタム配列特異性（例えば、非操作のDNA結合ドメインとは異なる配列特異性）または改変されたDNA結合親和性を提供するように改変され得る。カスタムDNA結合特異性を提供するために、例えば、ジンクフィンガーDNA結合ドメイン及びTALE DNA結合ドメインを操作する方法は、当該技術分野で公知である（例えば、Maeder et al., 2008及びSanjana et al., 2012.）。本明細書に記載の異種転写活性化因子が、天然DNA結合ドメインを含むDNA標的化ドメインを含む場合、エフェクタードメインは、転写因子が内在的に結合するすべての遺伝子座に標的化され、それによって、内在性転写因子の機能

40

50

を増強／置換する。例えば、Oct 4トランス活性化ドメインをVP16で置き換えると、線維芽細胞をiPS細胞にリプログラミングする効率が増加することが知られている。同様に、エフェクタードメインに作動可能に連結された天然DNA結合ドメインを含む異種転写活性化因子は、内在性転写因子によって調節される標的遺伝子の転写を活性化することによって、例えば、創傷治癒、分化転換、または組織再生を促進し得る。操作された（例えば、個別の配列特異性）ジンクフィンガーDNA結合ドメイン、TALE DNA結合ドメイン、またはCasファミリータンパク質の場合、エフェクタードメインは、制御された様式で、1つ以上の特定の遺伝子座、または任意選択でゲノム中の単一の遺伝子座に運ばれ得る。

【0072】

一実施形態では、DNA標的化ドメインは、dCas9などのCRISPR-Casタンパク質を含む。gRNA及び標的DNA結合活性を保持する酵素的に不活性なCRISPR-Casタンパク質を使用することができる。例えば、Cas9におけるD10A/H840Aの変異は、RuvC1及びHNHヌクレアーゼドメインに変異を導入し、不活性化をもたらす。一実施形態では、CRISPR-Casタンパク質は、配列番号1のアミノ酸配列、または配列番号1と少なくとも80%、少なくとも90%、少なくとも95%、もしくは少なくとも99%の配列同一性を有するアミノ酸配列を有するdCas9であり、D10A/H840Aを含み、gRNA及び標的DNA結合活性を保持する。他の酵素的に不活性なCRISPR-Casタンパク質も企図され、当業者によって同定され得る。

【0073】

一実施形態では、DNA標的化ドメインは、ジンクフィンガーDNA結合ドメインを含む。一実施形態では、ジンクフィンガーDNA結合ドメインは、特異的なDNA配列に結合するように操作された操作されたジンクフィンガーDNA結合ドメインである。

【0074】

エフェクタードメインは、本明細書に記載の少なくとも1つのトランス活性化ドメイン（TAD）またはその活性断片を含む。実施例に示すように、本明細書で同定される種々の全長ORF及びTADを単独または組み合わせて使用して、GFレポーター構築物またはCD133などの内在性遺伝子の転写を活性化することができる。また、実施例に示すように、エフェクタードメインは、表1に示す少なくとも1つのTADドメイン、及び／または例えば表3に示すようなその活性断片を含むことができる。したがって、一実施形態では、エフェクタードメインは、表2、表4、及び／または表5及び／または表6に示される少なくとも1つのTAD、及び／またはこれらのいずれか1つの機能的バリエーション、または表1、表2、表3、表4、表5、及び／または表6に示される2つ以上のTAD、及びこれらのいずれか1つの機能的バリエーションを含む。一実施形態では、TADは、表1、表2、表3、表4、表5、及び／または表6のTADドメイン、ならびにそれらのいずれか1つの機能的バリエーションのいずれか1つと少なくとも80%、少なくとも90%、少なくとも95%、または少なくとも99%の配列同一性を有する配列を有し、転写活性化活性及び／または特定の転写補活性化因子、例えば、CBP/p300、NuA4、及び／またはBRD4との相互作用を保持するポリペプチドを含む。

【0075】

これらのバリエーション及び組み合わせも使用され得る。一実施形態では、エフェクタードメインは、2つ以上のタンデムTAD、任意選択で2つのTAD、3つのTAD、4つのTAD、または4つより多いTAD、例えば、5つのTAD、10個のTAD、15個のTAD、20個のTAD、25個のTAD、30個のTAD、または5個のTADと30個のTADとの間の任意の数のTAD、または30個より多いTADを含む。一実施形態では、エフェクタードメインは、表1、表2、表3、表4、表5、及び／または表6に列挙されるものならびにそれらのいずれか1つの機能的バリエーションから選択される2つ以上のTADまたはその機能的バリエーションを含む。一実施形態では、エフェクタードメインは、表3に列挙されるものから選択される3つまたは4つのTADを含む。一実施形態では

、エフェクタードメインは、配列番号185、配列番号103、配列番号167、配列番号105、配列番号106、及び／または配列番号104の1つ以上を含む。一実施形態では、エフェクタードメインは、配列番号185、任意選択で配列番号90、91、102、または157を含む。一実施形態では、エフェクタードメインは、配列番号103、任意選択で配列番号46、47、または162を含む。一実施形態では、エフェクタードメインは、配列番号167、任意選択で配列番号101、110、166、または172を含む。一実施形態では、エフェクタードメインは、配列番号105、任意選択で配列番号116、117、または165を含む。一実施形態では、エフェクタードメインは、配列番号106、任意選択で配列番号116または117を含む。一実施形態では、エフェクタードメインは、配列番号104、任意選択で配列番号118、119、または159を含む。一実施形態では、エフェクタードメインは、SPDYE4-CITED1-RELA-HSF1（配列番号121）；SPDYE4-CITED1-RELA（配列番号123）；HSF1-RELA-SPDYE4-CITED1（配列番号125）；SPDYE4-CITED1-p65-miniHSF1（配列番号127）；miniSPDYE4-CITED1-p65-HSF1（配列番号129）；SPDYE4-miniCITED1-p65-HSF1（配列番号131）；SPDYE4-CITED1-minip65（C）-HSF1（配列番号133）；またはSPDYE4-CITED1-minip65（N）-HSF1（配列番号135）を含む。一実施形態では、エフェクタードメインは、SPDYE4-CITED1-SERTAD2-HSF1；SPDYE4-CITED1-KLF6-HSF1；SPDYE4-CITED1-ZXDC-HSF1；SPDYE4-CITED1-ATF6-HSF1；SPDYE4-CITED1-FOXO1-HSF1；SPDYE4-CITED1-ATMIN-HSF1；SPDYE4-CITED1-p65-SERTAD2；SPDYE4-CITED1-p65-KLF6；SPDYE4-CITED1-p65-ZXDC；SPDYE4-CITED1-p65-ATF6；SPDYE4-CITED1-p65-FOXO1；SPDYE4-CITED1-p65-ATMIN；SPDYE4-C3orf62-p65-HSF1；SPDYE4-DDIT3-p65-HSF1；SPDYE4-FOXO1-p65-HSF1；SPDYE4-ATMIN-p65-HSF1；SPDYE4-ZXDC-p65-HSF1；C3orf62-CITED1-p65-HSF1；C11orf74-CITED1-p65-HSF1；KLF6-CITED1-p65-HSF1；ZXDC-CITED1-p65-HSF1；またはSOX7-CITED1-p65-HSF1を含む。一実施形態では、エフェクタードメインは、SPDYE4-C3orf62.2-P_{AD}-HSF1（配列番号174）；SPDYE4-C3orf62.3-P_{AD}-HSF1（配列番号176）；SPDYE4-C3orf62-MT-P_{AD}-HSF1（配列番号178）；SPDYE4-DDIT3-MT-P_{AD}-HSF1（配列番号180）；SPDYE4-CITED1-P_{AD}-HSF1-MT（配列番号182）；または3xZNF473-KRAB（配列番号184）を含む。他の組み合わせが、本明細書において具体的に企図される。

10

20

30

40

50

【0076】

エフェクタードメインは、異なる転写補活性化因子選好を有する2つ以上のTADを含み得る。例えば、エフェクタードメインは、CBP／p300成分と相互作用するTAD、例えばFOXO TAD、及びBET成分と相互作用するTAD、例えばSPDYE4 TADを含み得る。エフェクタードメインは、同様の転写補活性化因子選好を有する2つ以上のTADを含み得る。例えば、エフェクタードメインは、CBP／p300成分と相互作用する2つのTADを含むことができ、例えば、エフェクタードメインは、FOXO1 TAD及びCITED1 TADを含むことができる。他の組み合わせが、本明細書において具体的に企図される。

【0077】

本明細書で使用される場合、機能的バリエーションに関して、「有効」とは、機能的バリエーションが、非改変または非バリエーションTAD（例えば、野生型または全長TAD）と比較し

て、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、100%、または100%を超える転写活性化活性及び／または補活性化因子相互作用を保持することを意味する。切り詰めなどのバリエーションの転写活性化活性及び／または補活性化因子相互作用は、例えば、本明細書に記載の方法を用いて決定することができる。例えば、転写活性化活性は、実施例に記載のGFレポーター系を用いて決定することができる。バリエーションは、それぞれのDNA標的化ドメイン（例えば、dCas9）の発現レベルを制御しながら、同じレポーターまたは内在性コンテキストに係留され得る。親TADと比較した場合にレポーター遺伝子または標的遺伝子の誘導発現において検出される任意の差異は、バリエーションの効果に寄与し得る。補活性化因子相互作用は、例えば、実施例に示されるように、例えば、AP-MS及び／またはBioIDによって決定され得る。

10

【0078】

例示的なTAD及びエフェクタードメイン核酸及びポリペプチドを、表1～6ならびに配列番号120～135及び173～184に提供する。一実施形態では、エフェクタードメインは、前記核酸によってコードされるアミノ酸配列、または配列番号120～135及び173～184のTADによってコードされるアミノ酸配列と少なくとも80%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、もしくは少なくとも99%の配列同一性を有するアミノ酸配列を含み得る。そのようなポリペプチドのコードされたポリペプチドの活性（融合または発現及び活性化された場合）は、例えば配列番号120～135及び173～184と同じくらい有効である（例えば、有効な転写活性化の少なくとも80%を提供する）。

20

【0079】

一実施形態では、エフェクタードメインは、リンカーを介してDNA標的化ドメインに融合される。一実施形態では、2つ以上のTADは、1つ以上のリンカーによって一緒に融合される。例えば、グリシン及びグリシンセリンリンカーを使用することができる。実施例に記載の転写活性化因子は、種々のグリシンセリンリンカー、例えば、SGGSGGS（配列番号6）、GGS、SGGS（配列番号7）、及び／またはGSGSGS（配列番号8）を使用した。他のリンカー、例えばINSRSSGS（配列番号9）も使用できる。

【0080】

一実施形態では、転写活性化因子は、1つ以上の核局在シグナル（NLS）を含む。任意の適切なNLSを使用することができる。任意選択で、NLSはSV40 NLSである。1つ以上のNLSは、1つ以上のN末端NLS、1つ以上のC末端NLS、1つ以上の内部NLS、及び／またはそれらの組み合わせであり得る。任意選択で、NLSは、配列番号22のNLSを含み得る。一実施形態では、NLSは、INSRSSGS（配列番号9）などのN末端及び／またはC末端リンカーをさらに含み、任意選択で、配列INSRSSGSPKKRKVGS（配列番号141）を有する。

30

【0081】

本明細書に記載されるように、転写活性化因子またはエフェクタードメインは、核酸によってコードされ得、及び／または発現構築物から発現され得る。したがって、本開示の一態様は、本明細書に記載の転写活性化因子をコードする核酸である。本開示の別の態様は、本明細書に記載の転写活性化因子のエフェクタードメインをコードする核酸である。例えば、核酸は、表1～6、任意選択で表2、4、5、及び／または6、任意選択で表2または表4または表6のいずれか1つに提供されるTAD、または表1～6に提供される2つ以上のTADをコードし得る。一実施形態では、核酸は、配列番号120、122、124、126、128、130、132、134、173、175、177、179、及び180のいずれか1つの核酸、または配列番号120、122、124、126、128、130、132、134、173、175、177、179、及び180と少なくとも80%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または100%の配列同一性を有する配列を

40

50

含むことができ、ここで、異種転写活性化因子は、例えば、配列番号120、122、124、126、128、130、132、134、173、175、177、179、及び180によってコードされるエフェクタードメインと同程度に効果的に、例えば、本明細書に記載のアッセイで評価した場合、例えば少なくとも80%効果的に、少なくとも85%効果的に、少なくとも90%効果的に、少なくとも95%効果的に、少なくとも96%効果的に、少なくとも97%効果的に、少なくとも98%効果的に、少なくとも99%効果的に、少なくとも100%効果的に、または100%を超えて効果的に転写を活性化する。配列同一性は、例えば、完全エフェクタードメイン配列に対するもの、またはそこにコードされる1つ以上のTADもしくはTAD断片に対するものである。他の部分、リンカー、NLS等は、完全に異なり得る。エフェクタードメインをコードする核酸は、本明細書に記載の転写活性化因子をコードする核酸を生成するのに適し得る。例えば、エフェクタードメインをコードする核酸は、適切なクロニング部位に隣接され得るか、または前記核酸を含む発現構築物もしくはベクターは、エフェクタードメイン及びDNA標的化ドメインを作動可能に連結するためのDNA標的化ドメインの挿入を容易にするためのクロニング部位を含み得る。

10

20

30

40

50

【0082】

本明細書で使用される場合、「クロニング部位」という用語は、組換えDNA技術（クロニング）を使用して、目的の核酸分子が挿入され得る、または目的の核酸分子が連結され得る核酸分子の一部を指す。発現カセットとの関連において、クロニング部位は、プロモーターとポリアデニル化シグナルとの間に位置し得、その結果、目的の核酸分子は、プロモーター及びポリアデニル化部位と作動可能に連結した状態で、発現カセットにクロニングされ得る。いくつかのクロニング技術が当業者に公知であり、クロニング部位は、クロニング部位での目的の核酸分子の挿入を可能にするために必要な特徴（制限エンドヌクレアーゼ部位（複数可）、リコンビナーゼ認識部位（複数可）、または平滑末端もしくは突出末端（複数可）など）を含む。クロニング部位は、例えば、目的の核酸分子が挿入されることを可能にする複数の固有の制限酵素認識部位を含むマルチクロニング部位（MCS）またはポリリンカー領域であり得る。代替的に、または付加的に、クロニング部位は、DNA挿入を触媒するためにインテグラーゼまたはCreリコンビナーゼなどの部位特異的リコンビナーゼ（複数可）を使用する、組換えクロニングによるDNA挿入を可能にする1つ以上のリコンビナーゼ認識部位を含み得る。組換えクロニング系の例としては、Gateway（登録商標）（インテグラーゼ）、Creator（商標）（Creリコンビナーゼ）、及びEcho Cloning（商標）（Creリコンビナーゼ）が挙げられる。いくつかのクロニング戦略では、発現カセットまたはベクターを線状分子として提供し、目的の核酸分子の平滑末端または突出末端を、例えばライゲーションまたはポリメラーゼ活性によって、発現カセットまたはベクターの平滑末端または突出末端に結合させ、したがって環状分子を形成することができる。この場合、発現カセットまたはベクターの平滑末端または突出末端と一緒にクロニング部位とみなすことができる。そのようなアプローチは、PCR産物をクロニングするために一般的に使用される。

【0083】

関連する態様は、プロモーターに作動可能に連結された転写活性化因子及び転写終結部位をコードする核酸を含む発現構築物である。任意の適切なプロモーターが使用され得る。適切なプロモーターは、当業者によって同定され得、例えば、CMV、EF1A、またはPGKを含み得る。例えば、配列番号25、26、27、及び／または28のプロモーター及び／またはエンハンサー配列は、発現構築物において使用され得る。誘導性プロモーターもまた使用され得る。

【0084】

一実施形態では、構築物はベクターである。任意の適切なベクターが使用され得る。適切なベクターは、当業者によって同定され得、ウイルスベクター、任意選択でレンチウイルスベクターまたはアデノウイルスベクターを含み得る。適切なベクターは、例えば、エ

フェクター構築物を発現するためのプロモーター、ポリAテイル、発現の安定性を増加させるためのWPREのような3'UTRエレメント、インスレーター配列、レンチウイルスパッケージングシグナル、蛍光タンパク質、及び／または抗生物質耐性マーカーを含み得る。追加の適切な成分は、当業者によって同定され得る。

【0085】

別の実施形態では、転写活性化因子、核酸、構築物、またはベクターは、細胞中にある。任意の適切な細胞が使用され得、所望の用途に基づいて当業者によって決定され得る。細胞は、任意の生物に由来してもよい。任意選択で、細胞は、ヒト細胞またはマウス細胞などの哺乳動物細胞である。任意選択で、細胞は細胞株である。細胞株は、任意の適切な細胞株であり得る。

10

【0086】

転写活性化因子、核酸、構築物、またはベクターは、任意の適切な様式で、例えば、トランスフェクションによって細胞に導入され得る。適切なトランスフェクション試薬及び方法は、当該技術分野で日常的に実施され、当業者によって同定され得る。任意選択で、構築物はウイルスベクター、任意選択でレンチウイルスベクターであり、形質導入によって細胞に導入される。適切な形質導入方法は、当該技術分野で日常的に実施され、当業者によって同定され得る。

【0087】

いくつかの実施形態では、細胞は異種転写活性化因子を安定に発現し、任意選択で、細胞は安定に形質導入され、例えば、異種転写活性化因子をコードする核酸を含むウイルスを用いて調製される。

20

【0088】

別の態様は、本明細書に記載の転写活性化因子、転写活性化因子をコードする核酸、または前記核酸を含む構築物もしくはベクター、または転写活性化因子を発現する細胞を含む転写活性化系である。CRISPR-Casに基づく系の場合、系は少なくとも1つのgRNAを含む。誘導性二量体化ドメインに基づく系の場合、系は任意に少なくとも1種の誘導剤を含む。

【0089】

本明細書に記載の異種転写活性化因子、本明細書に記載の核酸、本明細書に記載の構築物、本明細書に記載のベクター、本明細書に記載の細胞、及び／または本明細書に記載の転写活性化系を含む組成物も提供される。組成物は、BSAなどの担体、または組成物成分に応じて好適な希釈剤、任意選択で水または緩衝生理食塩水を含むことができる。組成物は、転写活性化因子、核酸、構築物、ベクター、または同じもしくは異なる要素を含む細胞のような複数の成分を含むことができる。

30

【0090】

また、本明細書に提供されるのは、例えば、標的遺伝子の転写を活性化するための、または本明細書に記載の方法を実施するためのキットであり、該キットは、本明細書に記載の転写活性化因子、本明細書に記載の転写活性化因子をコードする核酸、発現構築物、もしくはベクター、または本明細書に記載の転写活性化因子を発現する細胞と、任意選択で、転写活性化因子、核酸、発現構築物、ベクター、細胞、もしくは組成物を収容するバイアルとを含む。キットは、1つ以上の前述の成分の複数を含み得る。任意選択で、キットは、gRNA発現構築物、誘導剤、及び／または本明細書に記載の方法を実施するための使用説明書を含む。

40

【0091】

細胞における標的遺伝子の転写を活性化する方法も本明細書に記載される。実施例に示すように、本開示の転写活性化因子は、ゲノム遺伝子座、例えばプロモーターを標的にして、細胞中の標的遺伝子の転写を活性化することができる。

【0092】

本明細書で同定される転写エフェクターは、全長タンパク質、その断片（トランス活性化ドメイン）、その機能的バリエーション、またはトランス活性化ドメインもしくは機能的バ

50

リアントの組み合わせであってもよい。それらは、非常に強力から中程度から弱い様に、複数の異なる転写活性化強度をカバーする。これは、特に高すぎる発現が有害な表現型を引き起こし得る場合に、内在性遺伝子の所望の発現レベルを達成するために使用され得る。活性化強度は、エフェクタードメインに含めるための異なるTADまたは機能的バリエーション（例えば、活性断片）を選択することによって調整され得る。エフェクタードメインの活性化強度は、当業者によって、例えば、本明細書に記載の実施例に示される動員アッセイにおいてMFIまたはGFP陽性細胞パーセントを使用して決定され得る。例えば、エフェクタードメインの相対強度は、特定のDNA標的化ドメイン及び特定のDNA標的と組み合わせた特定のエフェクタードメインのMFIまたはGFP陽性細胞パーセントを、同じDNA標的化ドメイン及び特定のDNA標的と組み合わせたRenillaなどの対照と比較することによって決定することができる。高い活性化は、例えば、対照の少なくとも50倍以上、対照の少なくとも75倍以上、対照の少なくとも100倍以上、または対照の少なくとも150倍以上であると考えることができる。中程度の活性化は、例えば、対照の少なくとも10倍以上、少なくとも20倍以上、少なくとも30倍以上、または少なくとも40倍以上、及び対照の50倍まで、75倍まで、100倍まで、または150倍までであると考えることができる。低い活性化は、例えば、対照の少なくとも2倍、少なくとも2.5倍、少なくとも3倍、少なくとも4倍、または少なくとも5倍、及び対照の10倍まで、20倍まで、30倍まで、または40倍までであると考えることができる。例えば、図19A及び19Bに示すように、Renilla対照に対して、高い活性化は50倍を超えると考えられ、中程度は20倍～50倍であり、低い活性化は少なくとも3倍～20倍である。適切な活性化レベルは、所望の用途に応じて選択され得る。

10

20

【0093】

本明細書に記載されるように、転写調節のための別のレベルの対照を、例えば、ラパログまたはアブシジン酸によって化学的に誘導された二量体化と共に添加することができる。この場合、一方の半分（例えば、DNA標的化ドメイン）はFKBPまたはPYL1に融合され、他方の半分（例えば、エフェクタードメイン）はFRBまたはABI1に融合される。ラパログまたはアブシジン酸による処理は、FKBPとFRBまたはPYL1とABI1との間の相互作用をそれぞれ誘導し、一時的に調節された遺伝子発現をもたらす。実施例に示すように、ABI1及びPYL1を含む異種転写活性化因子の二量体化は、アブシジン酸の添加により誘導することができる。当業者は、一緒に使用され得る適切な誘導性二量体化ドメイン及び誘導剤を容易に同定及び選択することができる。任意の適切な誘導性二量体化ドメインと誘導剤の組み合わせが使用され得、例えば、ABI1とPYL1の二量体化は、アブシジン酸の添加により誘導され得る。他の誘導性の系としては、ラパマイシン、ジベレリン酸／ジベレリン、及びスプリットdCas9ベースの系による誘導に基づくものが挙げられる。例えば、GID1とGAIの二量体化は、ギベレリンによって誘導することができ、FKBPとFRBの二量体化は、ラパマイシンまたはその類似体、例えばラパログによって誘導することができる。SunTag系（Tenenbaum et al., 2014）などのより高次の多量体化系もまた、本明細書において企図される。

30

【0094】

DNA標的化ドメインとエフェクタードメインとの間の相互作用はまた、他の誘導系を使用して制御され得る。他の系（二量体化に依存しない）は、グラゾプレビル誘導性安定化（Taguet et al., 2018）またはエストロゲン受容体リガンド結合ドメインバリエーションを用いるタモキシフェン調節核局在化を含む。グラゾプレビル誘導性安定化の場合、DNA標的化ドメインとエフェクタードメインは、自己切断NS3プロテアーゼドメインによって連結される。グラゾプレビル（NS3活性を阻害する）の存在下でのみ、DNA結合ドメインとエフェクタードメインは、一緒に留まり、遺伝子発現を調節する。

40

【0095】

したがって、本開示の一態様は、細胞における標的遺伝子の発現を活性化する方法であ

50

り、該方法は、本明細書に記載の転写活性化因子を細胞に導入することと、DNA標的化ドメインが転写活性化因子を標的部位に誘導し、エフェクタードメインが標的遺伝子の転写を活性化するような適切な条件下で細胞を培養することを含む。一実施形態では、標的遺伝子は内在性遺伝子である。一実施形態では、転写活性化因子がCRISPR-Casを含む場合、該方法は、細胞中の所望のゲノム遺伝子座を標的とする少なくとも1つのgRNAを細胞中に導入することと、エフェクタードメインが標的遺伝子の転写を活性化するように、少なくとも1つのgRNAがCRISPR-Casタンパク質と結合し、CRISPR-Casタンパク質を誘導して転写活性化因子をCRISPR標的部位に誘導するような適切な条件下で細胞を培養することとをさらに含む。転写活性化因子がDNA標的化ドメインのそれぞれ及びエフェクタードメイン中に誘導性二量体化ドメインを含む実施形態では、該方法は、少なくとも1つの誘導剤を細胞に導入することと、少なくとも1つのエフェクタードメインが標的遺伝子の転写を活性化するように、第1及び第2の誘導性二量体化ドメインが結合する適切な条件下で細胞を培養することとをさらに含む。

10

【0096】

本明細書に記載の方法は、標的遺伝子の遺伝子発現を調節するために、例えば、内在性遺伝子の発現を誘導するために、またはゲノムの定義された領域におけるクロマチンオープニングを調節するために使用され得る。例として、一部のTADは、遺伝子間領域（すなわち、プロモーターまたはエンハンサーではない）においてクロマチンオープニングを促進し得、これは、クロマチンオープニング及び染色体折り畳みの再編成をもたらし得る。

20

【0097】

本明細書に記載の方法は、細胞生存度または目的の表現型に重要である1つ以上のゲノム遺伝子座を同定またはスクリーニングするために使用することができる。例として、本明細書に記載の方法を用いて、ジフテリア毒素などの目的の毒素に対する耐性または感受性に重要な遺伝子またはその調節エレメントをスクリーニングすることができる。別の例では、本明細書に記載の方法を用いて、CD81などの目的のタンパク質の発現に重要である調節エレメントを同定することができる。さらなる例において、本明細書に記載の方法は、特定の条件下、例えば、経時的な薬物処理において、細胞集団において過剰または過小に発現されるgRNAをスクリーニングすることによって、細胞型における必須または非必須の遺伝子を同定するためのハイスループットスクリーニング方法において使用され得る。他の用途は、当業者によって決定され得る。

30

【0098】

上記の開示は、本出願を一般的に説明する。より完全な理解は、以下の具体例を参照することによって得ることができる。これらの実施例は、例示の目的のみで記載され、本開示の範囲を限定することは意図されない。環境が慣例を示唆または与え得るので、等価物の形態及び置換の変化が企図される。本明細書では特定の用語が使用されているが、それらの用語は説明的な意味を意図しており、限定を目的とするものではない。

【実施例】

【0099】

以下の非限定的な例は、本開示の例示である：

40

【0100】

III. 実施例

実施例1. ヒト細胞における転写活性化因子を同定するためのプラットフォーム。

転写調節因子を体系的に同定するために、化学誘導二量体化（CID）システムが使用され、ここで触媒的に不活性なCas9（dCas9）は、そのN末端においてタンパク質ホスファターゼABI1でタグ付けされ、そして潜在的な転写活性化因子は、アブシジン酸受容体PYL1に融合されている（図1A）（Gao et al., 2016; Liang et al., 2011）。細胞のアブシジン酸（ABA）による処理は、ABI1とPYL1との間の相互作用を誘導し、それによって、潜在的な転写活性化因子を、ガイドRNA（gRNA）によって定義される特異的なゲノム遺伝子座に動員する。レ

50

ポーターとして、7xTetOアレイとEGFPの発現を駆動する基礎CMVプロモーターとを有する安定に組み込まれた構築物を含むHEK293T細胞株を使用した(Gao et al., 2016)。ABI1-dCas9及びTetOを標的とする単一のgRNAを発現するモノクローナル細胞株を、レンチウイルス感染によって作製した。PYL1に融合した既知の転写活性化因子VPR、VP64、及びp300によるこの細胞株のトランスフェクションは、ABA処理の際にGFPの強力な誘導をもたらした(図1B)、これは以前の報告(Gao et al., 2016)と一致した。ルシフェラーゼまたはRFPのプロモーターへの動員はGFP発現を誘導せず、ABA単独ではレポーターを活性化しないことを実証した(図1B)。

【0101】

個々のクローンからスケールアップするために、C末端PYL1を含むレンチウイルスベクター中のヒトORFeome 8.1及びORFeome協調クローン(ORFeome Collaboration, 2016; Yang et al., 2011)のプールされたライブラリを作製した。一緒に、これらのオープンリーディングフレーム(ORF)ライブラリは、13,571の固有遺伝子に対応する14,821のクローンを含む。レポーター細胞株を、低感染多重度でプールしたライブラリに感染させ、したがって、ほとんどの細胞が、1つのレンチウイルスのみに感染したことを確実にした(図1C)。プールされたライブラリは、感染前と後でそれぞれORFeomeの96%及び93%をカバーし、挿入サイズの有意な多様性にもかかわらず、ライブラリの広いカバー範囲を実証した(図6A及びB)。感染細胞をABAで処理した後、細胞のサブセットにおけるGFPの明確な誘導が観察された(図1D)。レポーター誘導は、処理時間、ABA濃度、及びORFeomeの存在に依存した(図6C~6E)。さらに、ABAの除去は、高いGFP集団の急速な喪失を引き起こし(図6C)、これは、活性化に対して連続プロモーター占有が必要であることを実証する。転写活性化因子を同定するために、ABA処理細胞のGFP陽性細胞の上位1%を二重で選別し、高GFP集団におけるORFを配列決定によって同定した(図1D)。

【0102】

方法は、実施例7に記載される通りである。

【0103】

実施例2. ORFeome-wideスクリーニングは、公知及び新規の転写活性化因子を同定する。

248個の推定転写活性化因子を、5%偽発見率のカットオフ及び上位1%GFP陽性細胞と未選別細胞との間の読み取り数の少なくとも4倍の変化を用いて同定した(図1E)。遺伝子オントロジー(GO)分析は、転写活性化に関連する複数の機能的カテゴリーに対して有意な濃縮を明らかにした(図1F)。ヒットはまた、多くの転写調節因子(図1G)、クロマチン関連タンパク質複合体のサブユニット(図1H)、ならびにRNAポリメラーゼII及びヒストンアセチルトランスフェラーゼCBP及びp300などの転写の中心ハブの相互作用因子(図6F)に見出されるタンパク質ドメインに高度に濃縮された。さらに、ヒットは、酵母ツーハイブリッドアッセイにおいて自己活性化因子として機能するヒトタンパク質と有意に重複し(ヒット中の29%に対するプロテオーム中の11%; $p < 0.0001$ 、フィッシャーの直接検定; 図6G)、これらは、プロモーターに異所的に動員された場合に酵母におけるレポーター遺伝子発現を活性化するタンパク質である(Luck et al., 2020)。スクリーニング結果はまた、90ヒット及び非ヒットの集合体を、それらを同一のレポーター細胞株に個々にトランスフェクトすることによってアッセイすることによって検証した。結果はスクリーニングと非常に一致した: 個々に試験した場合に再現されたほとんどすべてのヒット、逆に大部分の非ヒットはレポーターを活性化せず(図6H)、この環境において低い偽陽性率及び偽陰性率を示唆した。まとめると、これらの結果は、スクリーニングが機能的に関連する転写活性化因子を再現可能な様式で同定したことを示す。

【0104】

個々のスクリーニングヒットは、転写の異なる段階を調節する十分に特徴付けられた因子を含んでいた（図1E）。例えば、配列特異的転写因子ヒットの中で、HSF1、ATF6、及びDDIT3／CHOP（Vihervaara et al., 2018）などのストレス応答を調節するいくつかのマスターTFに加えて、活性化因子RELB及びMYCL（Barrett et al., 1992; Ryseck et al., 1992）が知られていた。DNA自体に結合しないがTFに結合する補活性化因子は、STAT2、CITED1、及びSERTAD1（Bousoik and Montazeri Aliabadi, 2018; Hsu et al., 2001; Yahata et al., 2001）を含んでいた。開始前複合体の構築を調節する一般的な転写因子であるTFIIIE（GTF2E1及びGTF2E2）の両方のサブユニット（Cramer, 2019）は、タンパク質がRNAポリメラーゼII放出（P-TEFb複合体サブユニットCDK9）及び転写伸長（例えば、ELL3及びMLLT1）を促進する場合と同様に、顕著なヒットであった（Chen et al., 2018; Cramer, 2019）。他の顕著な活性化因子は、クロマチン修飾複合体のサブユニット、例えばSAGA（ATXN7L3、TADA3、SGF29）及びNuA4（MGRBP）、ならびにDNA脱メチル化を媒介するGADD45ファミリータンパク質（GADD45A、GADD45B、GADD45G）（Barreto et al., 2007）。最後に、ライブラリ中に存在した15個のメディエーターサブユニットのうち13個を同定した。これらの強調された例は、複数の異なる機構により、転写サイクルの異なる段階で転写を促進する因子をスクリーニングが明らかにしたことを示す。

10

20

【0105】

既知の転写調節因子に加えて、転写に以前に関連付けされていない（例えば、C11orf74／IFTAP、NCKIPSD、DCAF7、HFM1）または完全に特徴付けられていない（例えば、C3orf62、FAM90A1、SPDYE4、SS18L2、FAM9A、C21orf58）タンパク質の大きな集合体がスクリーニングにおいて同定された（図1H）。これらの結果は、ヒトゲノムが多く以前に未知であった転写調節因子をコードすることを示唆する。これらの因子の幾つかの詳細な特徴付けは、以下に記載される。

【0106】

方法は、実施例7に記載される通りである。

30

【0107】

実施例3. TFファミリー内の異なる転写活性。

転写因子は、それらの異なるDNA結合ドメイン及び補助ドメインによって特徴付けられる複数のファミリーを含む（Lambert et al., 2018）。同じファミリーに属するTFは、多くの場合、高度に類似または同一の配列特異性を有する（Jolma et al., 2013; Lambert et al., 2018; Weirauch et al., 2014）。それにもかかわらず、高度に関連するTFでさえ、固有の補助ドメインに起因して、転写及びクロマチンに対して異なる効果を有し得る。これと一致して、転写因子ファミリーの一部のメンバーのみが、プールされたスクリーニングにおいてヒットとして同定された。プールされたアプローチの感受性を原因として除外するために、フォークヘッドボックス（FOX）、SRY関連HMG-ボックス（SOX）、E-twenty-six（ETS）、atona1関連の塩基性ヘリックスループヘリックス（bHLH）、Twist／Hand、Krueppel様因子（KLF）、及びホメオボックスタンパク質（HOX）ファミリーのメンバーを個々にアッセイした（図2A及び7A）。転写及びクロマチンに関連した2つのタンパク質ファミリー（ポリコム群RINGフィンガー（PCGF）及びカゼインキナーゼファミリー）、以前に転写に関連していない1つのファミリー（Spy1／RINGO）（Gastwirth et al., 2007）、及び24のメディエーターサブユニットも含まれ、これらは進化的に関連しないが、同じ保存された複合体の不可欠な成分である。

40

【0108】

50

プールされたスクリーニングと一致して、多くの高度に相同な転写因子の活性化プロファイルは、一度に1回試験した場合でさえ、顕著に異なった。例えば、37個のフォークヘッドTFのうちの5個、14個のSOXのうちの1個、14個のKLFのうちの5個、及び36個のHOXタンパク質のうちの2個のみがレポーター発現を活性化した(図2A~2D及び7)。活性の差が発現に起因するという自明な説明を排除するために、いくつかのTFをRFP-PYL1融合体としてアッセイして、融合タンパク質発現レベルを説明した。結果はPYL1のみの融合体と高度に一致し、RFPレベルと相関せず(図8)、転写活性の差がTFの固有の可能性を反映したことを示した。

【0109】

特に、スクリーニングにおいて同定された活性化TFは、異所的に発現された場合にマウス胚性幹細胞及びヒト誘導多能性細胞(iPSC)の分化を誘導し得る因子について有意に濃縮され(図7G)(Ng et al., 2021; Theodorou et al., 2009)、アッセイにおける活性化差異が生物学的機能における差異に関連することを示唆した。例えば、関連するbHLHファミリー転写因子のうち、NEUROG1、NEUROG2、NEUROD1、及びNEUROD2は、iPSCのニューロン分化を誘導し得るが、NEUROD6は誘導し得ない(Goparaju et al., 2017)。このパターンは、レポーター遺伝子を活性化するそれらの能力と一致する(図7B)。同様に、4つのHOX因子(HOXA1、HOXA2、HOXB1、HOXB2)のみが、Hoxb1遺伝子座に位置するb1-ARE自己調節エレメントを活性化することができ(Di Rocco et al., 1997)、これらの内の3つは、アッセイにおける活性化因子として特徴付けられた(プールされたスクリーニング及びアレイされたアッセイにおけるHOXA2及びHOXB2、スクリーニングにおけるHOXA1)(図2A)。さらに、最近の研究は、HOX遺伝子の抑制可能性、発現パターン、及びゲノム位置の間の著しい共線性を明らかにし(Tycko et al., 2020)、ホメオボックスクラスターの5'末端のHOX転写因子は抑制性であった。これと一致して、アッセイにおけるHOXファミリー活性化因子は、そのHOX遺伝子クラスターにおける最も多いまたは最後の3'から2番目の遺伝子によってコードされる。

【0110】

特に興味深い場合は、カノニカル及び非カノニカルのパリコーム抑制複合体1(PRC1)(Gahan et al., 2020; Gao et al., 2012)(図2E)の互いに排他的な成分であるPCGFファミリータンパク質の場合であった。一般に、PRC1との関連でクロマチン圧縮及び遺伝子サイレンシングにおいて作用すると考えられるが、PCGF3は、元のスクリーニングにおいて活性化因子として同定された。PCGF5(元のプールされたスクリーニングには存在しなかった)と同様に、PCGF3は、個々に試験した場合にレポーターを強く活性化した(図2E)。対照的に、3つの他のPCGFファミリーメンバー(PCGF1、PCGF2、及びPCGF4/BMI1)は、個々に試験した場合、スクリーニングヒットもレポーターを活性化もしなかった(図2E)。PCGF5は、転写活性化を調節することが以前に示されていた(Gao et al., 2014)、しかし、PCGF3についてそのような役割は記載されていなかった。興味深いことに、PCGFファミリーの系統発生において、PCGF3及びPCGF5は、動物の進化中に早期に生じた別個の群を形成し(Gahan et al., 2020)、それらの転写活性化機能が古代に起源を有することを示唆する。

【0111】

一部の研究は、メディエーター複合体のサブユニットが異なる調節機能を有し得、一部のサブユニットが転写活性化を促進し、他のサブユニットが抑制を促進することを示した(Conaway and Conaway, 2011; Stampfel et al., 2015)。アッセイにおいて、アッセイされた24個のメディエーターサブユニットのうち20個は、メディエーターサブモジュール間の差なしに、レポーターを強く活性化した(図7F)。例えば、MED29は、転写リプレッサー活性を有することが示唆されている(Wang et al., 2004)が、それは、プールされたスクリーニン

グにおけるヒットであり、個々に試験した場合に活性化因子として検証された（図 7 F）。したがって、少なくともレポーター系の文脈において、ほとんどすべてのメディエーターサブユニットが転写活性化を促進する。

【0112】

一次活性化スクリーニングは、細胞周期調節因子の *Spy1* / *RINGO*（卵母細胞における G2 / M 進行の急速な誘導因子）ファミリーに属する 2 つのタンパク質（*SPDY E4* 及び *SPDY E7P*）を同定した。*Spy1* / *RINGO* タンパク質は、サイクリン非依存的に *Cdk1* 及び *Cdk2* に結合し、これらを活性化し、それによって細胞周期進行を促進する（*Gonzalez and Nebreda, 2020*）。しかしながら、それらは、転写調節に以前に関係づけられてはいなかった。最近の拡大の結果として（*Chauhan et al., 2012*）、ヒトゲノムは、少なくとも 19 個の *Spy1* / *RINGO* ファミリー遺伝子及び複数の偽遺伝子を含む。5 つの *Spy1* / *RINGO* タンパク質を転写活性化について個々に試験し、そのうちの 4 つがレポーターを強く活性化した（図 2 F）。これらの結果は、*Spy1* / *RINGO* タンパク質が、転写活性化因子として機能することによって部分的に細胞周期の進行を促進し得ることを示唆する。

10

【0113】

方法は、実施例 7 に記載される通りである。

【0114】

実施例 4. *TAD-seq* は、新規のヒトトランス活性化ドメインを明らかにする。

転写因子は、一般に、メディエーター、*CBP* / *p300* アセチルトランスフェラーゼ、または *TFIID* などの補活性化因子と相互作用するトランス活性化ドメイン（*TAD*）を介して転写を活性化する。ほとんどの *TAD* は、酸性残基及び疎水性残基が豊富な短い非構造化配列である（*Sigler, 1988*）。集中的な努力にもかかわらず、明確なコンセンサスモチーフは出現せず、*TAD* の計算予測を困難にする（*Erijman et al., 2020*; *Ravarani et al., 2018*; *Staller et al., 2021*）。いくつかのグループが、既知の転写制御因子またはランダム配列から *TAD* を同定するためのプールされた動員アプローチを最近実施した（*Arnold et al., 2018*; *Erijman et al., 2020*; *Ravarani et al., 2018*; *Sanborn et al., 2020*）。上記のスクリーニングプラットフォームを改変して、ヒット間の転写活性化に関与する領域（複数可）を同定した。このアプローチは、ランダムに断片化された DNA の代わりに合成された断片を使用したことを除いて、以前に公開された *TAD-seq* 法（*Arnold et al., 2018*）と同様であった。

20

30

【0115】

本発明者らのスクリーニングで同定された 75 個の活性化因子の断片ライブラリを、20 アミノ酸ごとに 60 アミノ酸タイルを用いて、すべてのアミノ酸が 3 つの異なる断片によって表されるように作製した（図 3 A）。プールした断片ライブラリを *PYL1* に融合させ、元の *ORFeome* スクリーニングで使用した同じ *GFP* レポーターに動員した（図 3 A）。再び、*ABA* 依存性 *GFP* 発現は、断片ライブラリに感染したレポーター細胞のサブセットにおいて観察された（図 9 A）。*GFP* 陽性集団を 2 つの独立した複製に分類し、それぞれの集団中の断片を配列決定により定量した。しかしながら、*ORFeome* スクリーニングとは対照的に、高 *GFP* 細胞（上位 1 %）と中程度の *GFP* 細胞（上位 2 ~ 5 %）の両方を選別して、異なる強度のトランス活性化因子を潜在的に同定した。

40

【0116】

プールされたアプローチは、39 の異なるタンパク質における 70 の活性化断片を明らかにした。予想通り、これらの断片は、酸性及び疎水性アミノ酸が濃縮され、正に荷電した（塩基性）アミノ酸が枯渇し（図 9 B）、多くが「酸性 *blob*」概念（*Sigler, 1988*）と一致する「カノニカル」*TAD* を表すことが示された。実際、活性化断片は、2 つの異なるアルゴリズムに基づいてより多くの予測された *TAD* を含み（*Erijman et al., 2020*; *Piskacek et al., 2007*）（図 9

50

C)、予測されたTADは、不活性断片よりも活性化断片の方が長かった(図9D)。しかしながら、多くの活性断片は、いかなるアルゴリズムによっても予測されず、実験的手法の必要性を強調した。

【0117】

70個の活性断片のうち、大多数(44=63%)は、高GFPまたは中程度GFP集団においてのみ同定され、断片が異なる活性化能を有することを示唆した(図9E)。例えば、VP64は、未選別の断片ライブラリと比較した場合、高GFP集団において濃縮されたが、中程度のGFP集団において枯渇した(図9E)。これは、転写活性化におけるVP16の異常な効力と一致する(VP64は、タンデムの4つのVP16ペプチドからなる)(Sadowski et al., 1988)、スクリーニングにおいて同定された10個の断片を個々に試験し、それらの9個がレポーターを強固に活性化し(図9F)、このことは、低い偽陽性率を示唆する。

【0118】

断片スクリーニングは、いくつかの既知のTADを回収した。例えば、KLF6、KLF7、KLF15、ATF6、及びCITED2中の既知のTADを同定した(図3B及び10A)。さらに、活性断片間の重複は、活性に必要とされる最小配列を特定し得るので、以前に同定されたTAD(例えば、KLF6、KLF7、及びSERTAD2に対するもの)の活性に必要とされる領域は、狭められた(図3B及び10A)。カノニカル転写調節因子と新規タンパク質の両方における多くの以前に未知のTADが、ORFeome-wideスクリーニングにおいて発見された(図3C及び10B)。TADは、以前に特徴付けられていないタンパク質SPDYE4、C3orf62、FAM22F、及びFAM90A1で明らかにされた(図3C)。興味深いことに、SPDYE4中の新規TADは、Cdk2と相互作用し、これを活性化するそのSpy1ドメインに隣接しており、これは、潜在的な転写活性及びCdk調節活性が、Speedyファミリータンパク質の別個のドメインによって媒介されることを示す(図3C及び11A)(McGrath et al., 2017)。さらに、Spy1/RINGOファミリータンパク質の転写活性化プロファイルと一致して、TAD領域は、動員アッセイにおいて不活性であった唯一のSpy1/RINGOファミリータンパク質であるSPDYCを除くすべてのファミリーメンバーにおいて保存される(図2F及び11A)。

【0119】

興味深いことに、明らかにされていないTADのいくつかは、典型的なトランス活性化ドメインの特徴を有しなかった。例えば、転写を活性化したHOXA2中の3つの重複する断片は、ホメオボックスDNA結合ドメインとアンテナペディア様ヘキサペプチドモチーフとの間にポリアラニンストレッチを広げ(図3D及び3E)、これは、PBX1補活性化因子と相互作用する(Piper et al., 1999)。しかしながら、ヘキサペプチドモチーフを欠く断片は、依然として転写を活性化し、活性がPBX1を介して調節されないことを示唆する(図3E)。ポリアラニンストレッチは、DrosophilaのUltrabithorax(Ubx)タンパク質において機能的に重要であり、最近、HOXD13などの多くのTFの相分離を駆動する役割を有することが示唆された(Basuet al., 2020)。本明細書の結果は、ポリアラニンストレッチが転写活性化に追加の役割を有し得ることを示唆する。

【0120】

別の非カノニカルな活性化領域は、YAF2由来であり、これは、ポリコム抑制複合体1(PRC1)(Gao et al., 2012)の成分であり、元のORFeomeスクリーニングにおける顕著なヒット(図3F)である。この領域はYAF2_RYBPドメインを含み、これはコアPRC1サブユニットであるRING1Bに結合する逆平行βシートに折り畳まれる(Wang et al., 2010)(図3G)。YAF2のYAF2_RYBPドメイン及びその密接な相同体RYBPは、CBXファミリーポリコムタンパク質に存在するCBX-Cドメインと配列及び構造的相同性を共有する(Wang et al., 2010)(図11B)。タンパク質及びYAF2/RYBPを

含むCBX-Cドメインは、RING1A及びRING1Bと相互に排他的な様式で相互作用して、カノニカル（cPRC1）及び非カノニカル（ncPRC1）ポリコム複合体を形成する（図3G）。

【0121】

すべてのCBX-C及びYAF2__RYBPドメインが転写活性化を促進し得るか否かを試験するために、YAF2（配列番号96）及びRYBP（配列番号140）のYAF2__RYBPドメインならびにCBX2（配列番号136）、CBX4、CBX6（配列番号138）、CBX7及びCBX8（配列番号139）のCBX-Cドメインをクローニングし、レポーターシステムにより活性についてアッセイした。YAF2とRYBPの両方からのYAF2__RYBPモチーフは、YAF2のTADseq結果と一致して、レポーターを強力に活性化した（図3H）。一部のCBX-Cドメインも活性化因子であった：CBX2由来のCBX-Cが最も強力な活性化因子であったのに対して、CBX6及びCBX8由来のものはレポーターを弱く活性化した（図3H）。対照的に、CBX4及びCBX7のCBX-Cドメインは、レポーターに対して効果を有しなかった。このパターンは、CBX-Cドメインタンパク質の進化的祖先を反映した（図3H）。興味深いことに、CBX4及びCBX7のCBX-Cドメインは、転写を活性化するCBXタンパク質またはRYBPのYAF2__RYBPドメイン由来の同じドメインよりも約10倍高い親和性でRING1Bに結合する（図3H）（ $p=0.008$ 、両側t検定）（Wang et al., 2008, 2010）。したがって、CBX-C及びYAF2__RYBPドメインの転写活性の差は、RING1Bに対するそれらの結合親和性によって、または他の因子に対する差次的結合によって説明され得る。いずれの場合も、これらの結果は、CBXタンパク質機能の差異が、CBX-Cドメインの固有の差異によって少なくとも部分的に説明され得ることを示唆する（Morey et al., 2012; Vincenz and Kerppola, 2008）。より広く、これらの結果は、非カノニカル及びカノニカルPRC2複合体の集合及び機能における複雑さの別の層を明らかにする。

【0122】

方法は、実施例7に記載される通りである。

【0123】

実施例5. 新規の転写活性化因子は、既知の補因子と相互作用する。

ORFeome-wideスクリーニングは、不十分にまたは完全に特徴付けられなかったいくつかの強力なトランス活性化因子を明らかにした。これらの因子が転写をどのように調節するかを理解するために、安定なテトラサイクリン誘導性HEK293細胞株を確立し、9つの不十分に特徴付けられたスクリーニングヒット（C3orf62、C11orf74/IFTAP、NCKIPSD、DCAF7、SS18L2、SPDYE4、FAM90A1、FAM22F/NUTM2F、JAZF1）、5つの既知の転写調節因子ヒット（SOX7、KLF6、KLF15、CTBP1、HOXA2、及びHOXB2）、2つの合成トランス活性化因子（VP64及びVPR）、ならびに陰性対照（EGFP及びNanoluc）を、Aquifex aeolicus由来のビオチンリガーゼBirA及びFLAGエピトープタグに融合して発現させた。それらのタンパク質相互作用を、質量分析（AP-MS）に結合した親和性精製及び近接依存性ビオチン化を有する近接パートナー（BioID2）で特徴付けた（Kim et al., 2016）。AP-MSは、安定なタンパク質複合体を特徴付けるための理想的な方法であるが、BioIDは、クロマチンに強固に結合したもののような、より弱いまたは難溶性タンパク質を含む相互作用を同定するのに優れている（Lambert et al., 2015）。

【0124】

転写活性化因子の相互作用は、2つのパターンを明らかにした。第1に、彼らは、新規ヒットのトランス活性化能が、テザリングアッセイのアーチファクトではなく、それらの内在性機能を反映する可能性が高いことを示した。第2に、インタラクトームは、特定の補活性化因子複合体に対する活性化因子の著しい優先性を明らかにし、5つの異なる補因

子 (CBP / p300、BAF、NuA4、メディエーター、及びTFIID) に集中した。

【0125】

転写調節における新規活性化因子の天然の役割を裏付けて、9つの不十分に特徴付けられたヒットのうち8つは、AP-Ms、BioID、または両方における既知の転写補因子に関連する (図4A、図12)。C3orf62、DCAF7、及びFAM22F / NUTM2Fは、既知の転写補活性化因子であるp300及び／またはCBPと相互作用した。対照的に、JAZF1は、BioID中のNuA4ヒストンアセチルトランスフェラーゼ複合体の複数のサブユニットと結合し、これは、この高度に保存された複合体の新規のサブユニットであることを示唆する。次に、SS18L2は、コアBAFメンバーであるSMARCA2及びSMARCA4、ならびにカノニカルBAF (cBAF) 及び非カノニカルBAF (ncBAF) に特異的なサブユニットを含むBAFクロマチンリモデリング複合体と相互作用した (図4A及び12) (Centore et al., 2020)。SPDYE4及びFAM90A1は、転写伸長を調節するBETファミリープロモドメインタンパク質BRD2、BRD3、及びBRD4と相互作用した (Fujisawa and Filippakopoulos, 2017)。SPIN90としても知られるNCKIPSDは、リボヌクレオタンパク質複合体の集合を調節するが、転写活性化にも関連している運動ニューロン (SMN) 複合体の生存と相互作用した (Pellizzoni et al., 2001; Singh et al., 2017; Strasswimmer et al., 1999) (図12)。興味深いことに、NCKIPSDもDCAF7と相互作用した (図12B)。DCAF7とNCKIPSDの両方が、細胞質におけるアクチンダイナミクスの調節に関与しており (Cao et al., 2020; Morita et al., 2006)、これらのタンパク質が細胞質及び核において異なる役割を有し得ることを示唆している。

10

20

【0126】

既知の転写調節因子はまた、補活性化因子複合体に結合する (図4A及び12)。KLF6バイトは、NuA4サブユニットEPC1を近接相互作用因子として同定したが、HOBX2、KLF15、CTBP1、及びSOX7バイトは、近接パートナーとしてCBP及びp300を有した。さらに、SOX7はBAFサブユニットとも結合する。SOX7を除くすべての天然活性化因子とは対照的に、強力な合成活性化因子VP64及びVPRは、近接パートナーとして複数の異なる補活性化因子を同定した。CBP / p300ならびにメディエーターサブユニットMED14及びMED15に結合するウイルスVP16トランス活性化モチーフの4つのタンデムコピーからなるVP64は、以前の報告と一致する (Kundu et al., 2000; Yang et al., 2004)。VPRは、VP64、ヒトp65 / RELATrans活性化ドメイン、及びエプスタインバーウイルスRトランス活性化因子 (Chavez et al., 2015) の融合体であり、CBP / p300ならびにメディエーター及びTFIIDの複数のサブユニットを含むさらに多くの補因子と結合する (図4A及び12A)。複数の補因子とのこのような結合は、dCas9に融合した場合のVPRの例外的なトランス活性化効力を説明する可能性が高い (Chavez et al., 2015, 2016)。

30

40

【0127】

これらの結果は、活性化転写因子及び他の転写調節因子が、特定の補因子に対して強い固有の優先性を有することを示唆する。このことをさらに調べるために、フォークヘッドファミリーTFの以前に公開されたAP-Ms相互作用データセット (Li et al., 2015) を分析し、このアッセイにおける転写活性化結果と比較した。特に、レポーターに動員された場合に転写を活性化したフォークヘッドTFのみが、AP-Msにおいて補活性化因子と相互作用した (図4B)。加えて、質量分析の結果と同様に、フォークヘッドTFの活性化は、異なる補因子選好性を有した: FOXO1及びFOXO3は、CBP及びp300と特異的に相互作用し、FOXN1は、p300とBAF複合体のサブユニットの両方と相互作用し、一方、FOX R1及びFOX R2は、NuA4複合体を

50

選好した（図 4 B）。これらのデータは、高度に類似した配列を認識し、ほぼ同じ程度まで転写を活性化する関連転写因子が、異なる補活性化因子複合体を介して転写を促進し得ることを強く示唆する。

【0128】

転写活性化因子と補因子との間の結合を機能的に調べるために、83個の強固な活性化因子のパネルを配列した。それらの活性化能を、レポーターアッセイを用いて試験したが、ここでは、複数の転写補調節因子を標的とする小分子阻害剤の存在下で試験した。転写キナーゼを標的とする3種のキナーゼ阻害剤（CDK9にフラボピリドール、カゼインキナーゼ2にCX-4945、DYRK1A及びDYRK1BにAZ191）、ならびに2種の転写補因子を阻害する化合物（CBP/p300にA-485、BETファミリープロモドメインタンパク質にJQ1）を使用した。

10

【0129】

3つのキナーゼ阻害剤は、ほぼすべての活性化剤に影響を及ぼしたが、異なる程度であった。フラボピリドールによるCDK9の阻害は、すべての活性化因子の活性のほぼ完全な喪失をもたらした、これは、プロモータークリアランスにおけるp-TEFbの重要な役割と一致した（図13A）。転写の伸長を調節するカゼインキナーゼ2の阻害（Basnet et al., 2014）は、中程度ではあるが全般的な転写活性化の減衰を引き起こした（図13B）。DYRK1A/DYRK1B阻害は、他の2つの化合物よりも微妙な効果を有し、大部分の活性化因子の活性を僅かに減衰した（図13C）。興味深いことに、DYRK1A/DYRK1B阻害剤AZ191によって影響を受けなかった2つの活性化因子は、DYRK1Aと保存された複合体を形成するDCAF7（Breitkreutz et al., 2010; Yu et al., 2019）、及びDCAF7と相互作用するNCKIPSDであった（図13C）。

20

【0130】

転写に対して広い効果を有するキナーゼ阻害剤とは対照的に、CBP/p300のアセチルトランスフェラーゼ活性を阻害することは、一部の、しかしすべてではないトランス活性化因子の活性に強く影響を及ぼした（図4C）。重要なことに、これらの効果は、本発明者らのAP-MS及びBioIDの結果と一致した：6つのCBP/p300相互作用因子のうちの4つの活性は有意に減少したが、7つの非相互作用因子のうちの1つだけがA-485によって影響を受けた（図13D）。例えば、A-485はKLF15の活性を阻害したが、関連するKruppel様因子KLF6に対して効果を有しなかった（図13D）。より広範には、CBP/p300と相互作用することが知られているタンパク質の活性は、非相互作用トランス活性化因子の活性よりもA-485によって有意に阻害された（図4C）。比較すると、NuA4複合体の相互作用因子は、CBP/p300阻害によって有意に影響を受けなかった（図4D及び13D）。

30

【0131】

CBP/p300阻害と同様に、JQ1によるBETファミリー阻害は、いくつかのトランス活性化因子に対して異なる効果を有した。興味深いことに、ほとんどの場合、JQ1処理は、レポーター遺伝子活性の増加をもたらした（図4D）。JQ1処理は、多くの場合、BRD4標的遺伝子の迅速な下方制御を引き起こすが（Loven et al., 2013; Muhar et al., 2018）、レポーター遺伝子の転写活性化にも関連しており（Sdelci et al., 2016）、観察される効果を説明する可能性が高い。それにもかかわらず、すべてのトランス活性化因子がJQ1処理に同様に応答したわけではない。特に、NuA4複合体のサブユニットと相互作用する因子は、JQ1処理によって影響を受けなかった（例えば、JAZF1及びKLF6；図4D及び13D）。実際、NuA4相互作用剤は、CBP/p300相互作用因子または他のトランス活性化因子よりもJQ1処理による影響が有意に少なかった（図4D）。NuA4相互作用因子が独特の様式でJQ1処理に応答する機構は、さらなる研究を必要とするが、これらの結果は、単一のプロモーターとの関連で、転写活性化因子が異なる補活性化因子複合体を介して転写をどのように促進するかを実証する。実際、5つの化合物に対するそれ

40

50

らの感受性に基づく活性化剤の階層的クラスター化は、複数の異なる群を明らかにした（図4E）。多くのパラログス因子（例えば、CITED1及びCITED2、またはPCGF3及びPCGF5）は、互いに隣接してクラスター化し、クラスター化が機能的に関連する群を生成したことを示す。さらに、既知のCBP/p300相互作用因子は、主に、NuA4相互作用因子と同様に、2つの異なるクラスター中にあった（図4E）。これらのクラスターの他のメンバーは、補活性化因子としてCBP/p300またはNuA4、例えばp300に対するYAF2またはMYOG、またはNuA4に対するNOM1を同様に使用する可能性が高い。

【0132】

方法は、実施例7に記載される通りである。

【0133】

実施例6. SRF-C3orf62融合体はCBP/p300と相互作用し、SRF/MRTF転写プログラムを促進する。

転写調節因子を含む融合タンパク質は、白血病及び肉腫などの特定のがんの一般的な特徴である。本発明者らのORFeome-wideスクリーニングからのヒットは、多様ながんにおける融合パートナーとしてCOSMICデータベース（cancer.sanger.ac.uk）に記録された遺伝子について有意に濃縮された（ $p = 0.019$ ；超幾何分布検定）。これらには、よく特徴付けられた融合パートナー、例えば、ユイソイド脂肪肉腫ではEWSR1に、前立腺癌ではTMPRSS2に融合されるERG；ミクソイド脂肪肉腫ではEWSR1またはFUSに融合されるDDIT3/CHOP；粘膜表皮癌ではMAML2に融合されるCRTC1；及び混合系統白血病ではMLLに融合されるENL/MLLT1が含まれた。加えて、いくつかのヒットは、融合パートナーとして文献に記載されているが、機能的に特徴付けられていない。例えば、BTBD18及びNCKIPSDは、白血病においてKMT2A/MLL融合パートナーとして同定された（Alonso et al., 2010；Sano et al., 2000）。さらに、MLLに融合されたBTBD18の断片は、TAD-seqによって同定されたトランス活性化ドメインを含み（図10B）、融合産物の発がん能にTADを関連づける。大部分のMLL融合体は、転写伸長を調節する遺伝子、例えば超伸長因子複合体を含む（Winters and Bernnt, 2017）。興味深いことに、BTBD18は、転写伸長を促進することも示されている（Zhou et al., 2017）。

【0134】

活性化スクリーニングヒットが融合パートナーとして腫瘍形成を促進し得る機構に対してより多くの洞察を得るために、2つの不十分に特徴付けられた融合体、JAZF1-SUZ12及びSRF-C3orf62を、さらなる特徴付けのために選択した。JAZF1-SUZ12融合は、低悪性度子宮内膜間質肉腫（LG-ESS）の特徴であり（Hrzénjak, 2016）、ポリコムタンパク質SUZ12とJAZF1を結合させている（図5A）（Piunti et al., 2019）。SRF-C3orf62は、筋線維腫／筋周細胞腫の小児症例に最近記載された（Antonescus et al., 2017）。この融合体において、血清反応因子（SRF）のDNA結合ドメインは、C3orf62のC末端に融合される（図5B）。特に、両方の場合において、TAD-seq（図3C及び10B）によって同定されるトランス活性化ドメインは、融合構築物（図5A及び5B）に保持される

【0135】

SRFに融合したSUZ12、SRF、JAZF1-SUZ12、SRF-C3orf62、及びC3orf62のC末端断片（C3orf62-Cterm）をBirA-FLAGでタグ付けし、それらの相互作用をBioID及びAP-MSで分析して、JAZF1及びC3orf62について得られたデータを補完した。予想通り、SUZ12近接パートナーは、EZH2、MTF2/PCL2、及びC10orf12（Aleksyenko et al., 2014）（図5C）などのPRC2複合体の他の成分を含んだ。衝撃的なことに、JAZF1-SUZ12融合タンパク質は、近接パートナーとして

10

20

30

40

50

PRC2とNuA4サブユニットの両方を有し（図5C）、これは、この融合が、通常、反対の転写活性と関連する2つのクロマチン関連複合体の超複合体に組み立てられることを示す。これと一致して、LG-ESSとも関連するEPC1-PHF1融合体は、同様にPRC2-NuA4超複合体を組み立て、ポリコム標的の異常な発現をもたらす（Sudarshan et al., 2021）。さらに、JAZF1-SUZ12とEPC1-PHF1の両方が強力な転写活性化因子であることが最近示された（Sudarshan et al., 2021）。したがって、発がん性超複合体へのNuA4の組み込みは、PRC2複合体の通常の抑制機能を無効にし得る。

【0136】

SRF近接パートナーは、血清応答エレメント上でSRFと三元複合体を形成するELK1のような複数の転写因子を含んだ（Buchwalter et al., 2004）（図5D）。しかしながら、それは、いかなる転写補活性化因子とも結合しなかった。対照的に、C3orf62-Cterm構築物とSRF-C3orf62融合体の両方は、近接パートナーとしてCBPとp300の両方を強固に同定した（図5D）。同様に、AP-MSは、CBP及びp300を顕著なSRF-C3orf62相互作用因子として同定した（図14A）。これらの結果と一致して、C3orf62-CtermとSRF-C3orf62の両方が、レポーターに連結された場合にGFプレポーターを強く活性化した（図5E）。この活性は、SRF-C3orf62の相互作用パターンと一致して、A-485によるCBP/p300阻害に対して高度に感受性であった（図5F）。

【0137】

SRFは、三元複合因子（TCF；例えば、ELK1）またはミオカルディン関連転写因子（MRTF；例えば、MAL）のいずれかと共に機能して、標的遺伝子発現を調節する（Buchwalter et al., 2004；Olson and Nordheim, 2010）。結果は、SRF-C3orf62がそのような補因子を伴わずにSRF標的遺伝子を活性化できることを示唆した。このことをさらに試験するために、ルシフェラーゼに基づく血清応答エレメントレポーター（Vartiainen et al., 2007）を用いて、NIH3T3線維芽細胞において異なる構築物の活性をアッセイした。SRFまたはC3orf62単独ではレポーターを活性化しなかったが、SRF-C3orf62は強力に活性化し（図5G）、がん関連融合体が転写活性化におけるSRF補因子の要件を回避することがさらに確認された。

【0138】

MAPキナーゼシグナル伝達によって調節されるSRF/TCF経路は、前初期遺伝子の発現を調節し（Gualdrini et al., 2016）、一方、アクチン-Rhoシグナル伝達依存性SRF/MRTF経路は、細胞運動性及び接着に関与する遺伝子を標的とする（Miralles et al., 2003）。SRF-C3orf62が両方の経路の標的遺伝子を調節できるかどうかを試験するために、GFPに融合したSRF、C3orf62、SRF-C3orf62、及びNanolucを安定に発現する安定なドキシサイクリン誘導性NIH3T3細胞株を作製した。導入遺伝子発現をドキシサイクリンで24時間誘導し、遺伝子発現の変化をRNA-seqで分析した。GFPタグ付きC3orf62またはSRFの発現は、Nanolucと比較してトランスクリプトームに対する影響が非常に制限されているか、または全くなかった（図14B）が、SRF-C3orf62-GFPの発現は、564個の遺伝子の上方制御及び471個の遺伝子の下方制御をもたらした（log2倍率変化>1、FDR<0.05；図5H及び14B）。遺伝子セット濃縮分析（GSEA）及び遺伝子オントロジー分析は、上方制御された遺伝子が、DNA複製、細胞周期の進行、及び有糸分裂に関与する遺伝子において高度に濃縮されたことを明らかにし（図5I及び14C）、一方、下方制御された遺伝子は、細胞外マトリックスに関連していた（図14C）。上方制御された遺伝子はまた、細胞増殖の主要な調節因子であるE2F転写因子の標的において有意に濃縮された（図5I）。これらのシグネチャーは、SRF-C3orf62融合体の発がん性と一致する。さらに、最も上方制御された遺伝子の多くは、アクチン動力学及び筋形成に関与した。例え

10

20

30

40

50

ば、最も上方制御された遺伝子の1つは、平滑筋アクチン (ACTA2)、既知のSRF/MTRF標的、及びSRF融合陽性の筋線維腫/筋周細胞腫の特徴 (Antonescuet al., 2017) (図5H) であった。これと一致して、上方制御された遺伝子の別の有意なシグネチャーは筋形成であった (図5I)。対照的に、FOS、EGR1、またはEGR2などのSRF/TCFの多くの十分に特徴付けられた前初期標的遺伝子は、異所性SRF-C3orf62発現によって影響を受けなかった (図5H)。興味深いことに、VP16へのSRFの融合は、FOSを強力に上方制御することができ、これは、異なるTADの標的遺伝子発現を活性化する能力に固有の差があることを示唆する (Schratt et al., 2002)。実際、SRF-C3orf62によって上方制御された遺伝子と、以前に報告されたSRF/MTRFの標的遺伝子との間に有意な重複があったが、SRF/TCFの標的遺伝子とは重複がなかった (図14D) (Esnault et al., 2014; Gualdrini et al., 2016)。総合すると、これらの結果は、SRF-C3orf62発現が、SRF/TCF標的よりもSRF/MTRF標的遺伝子の増殖性及び筋原性遺伝子発現シグネチャー及び優先的発現をもたらすことを示す。

10

【0139】

方法は、実施例7に記載される通りである。

【0140】

実施例7. 方法:

細胞培養。スクリーニングに使用したpTRE3G-EGFPレポーター細胞株 (スタンフォード大学Lei Stanley Qi研究室からのギフト) を含むすべてのHEK293T細胞を、10%ウシ胎仔血清 (FBS) を含むDMEM中で維持した。NIH-3T3細胞をSachdev Sidhu博士の研究室 (トロント大学) から得て、10%ウシ血清 (BCS) を含むダルベッコ改変イーグル培地 (DMEM) で維持した。すべての培養培地に1%ペニシリン-ストレプトマイシンを補充した。細胞を、加湿インキュベーター中、5%CO₂で37℃に維持し、マイコプラズマ汚染について日常的に試験した。

20

【0141】

レンチウイルスの産生。プールされたORFeome及びトランス活性化ライブラリを含むレンチウイルス粒子を、293T細胞を8:6:1の比でpLX301-ORFs/TADs-PYL1、pSPAX2 (Addgene #12260)、及びpVSV-G (Addgene #8454) でトランスフェクトすることによって生成した。トランスフェクションは、XtremeGENE 9 (Roche) を使用して、15cmディッシュ上で、製造業者のプロトコールに従って行った。トランスフェクションの6~8時間後に培地を回収培地 (DMEM+1.1g/100mLのBSA) に交換した。トランスフェクションの72時間後に、上清を濾過 (0.45 µm) し、プールして回収した。Lipofectamine 2000 (Thermo Fisher Scientific, 11668019) 試薬を用いて6ウェルプレート上でトランスフェクションを行って個々の安定な細胞株を確立する場合に、同様のプロトコールを小規模ウイルス産生についても従った。

30

40

【0142】

細胞株の生成。ABI-dCas9 (ブラストサイジン、6 µg/mL) 及びpTRE3Gプロモーターを標的とするgRNA (配列番号10) を発現するEGFPレポーター株のクローン株 (共発現EBFP2) を作製した。単一細胞を選別し (FACS Aria IIiu, BD)、拡大し、強力な転写活性化因子による誘導を示すクローンを、その後の実験のために選択した。ドキシサイクリン誘導性EGFPタグ付きタンパク質を発現するNIH3T3細胞を生成するために、エンタリークローンをhORFeome集合体から採取し、Gateway適合pSTV6-TetO-cdB-EGFPレンチウイルスプラスミド (Payman Samavarchi-Tehraniからの親切なギフト) にサブクローニングした。NIH-3T3細胞を8 µg/mLポリブレンの存在

50

下で感染させ、感染 24 時間後に $2 \mu\text{g}/\text{mL}$ ピューロマイシンで選択した。

【0143】

プールされた ORFeome ライブラリの生成。ヒト ORFeome 集合体 (v8.1) からのエントリークローンを、それぞれが約 384 個の ORF を含む 40 個の標準化サブプールに集め、レンチウイルス Gateway 適合性目的ベクター pLX301-DEST-PYL1 にクローニングした。LR 反応物を、 150 ng のそれぞれのエントリー ORF サブプールと共に二重に設定し、 $1 \mu\text{L}$ の Gateway LR クロナーゼ II と合計 $5 \mu\text{L}$ の反応体積で組み合わせ、室温で TE 緩衝液中で一晩インキュベートした。次の 2 日間、 $1 \mu\text{L}$ の追加の LR 酵素を $4 \mu\text{L}$ の TE 及び 150 ng の目的ベクター中でそれぞれの反応物に添加した。コロニーを化学的にコンピテントな DH5 α E. coli に形質転換し、カルベニシリン ($100 \mu\text{g}/\mu\text{L}$) を含む LB 寒天プレート上に 30°C で一晩広げた。コロニーを計数して 200 倍を超えるカバー範囲を確実にし、氷上で SOC に収集し、ペレット化し、乾燥ペレットの重量に基づいて複数のカラム上でマキシプレッ

10

【0144】

活性化ドメインタイリングライブラリの生成。タイリングライブラリは、ORFeome スクリーニングにおいて活性化因子として同定された 75 個のタンパク質から生成された。5' アダプター (GSQGS GSM) (配列番号 11) 及び 3' アダプター (GGSE REG) (配列番号 12) を生成するための、5' アダプター GGAAGTCAGGGTAGCGGAAGTATG (配列番号 23) 及び 3' アダプター GGAGGTAGTGT TGAACGCGAAGGC (配列番号 24) を含むオリゴヌクレオチドを、プールされライブラリ (Twist Biosciences) として合成した。鑄型としての 5 nM オリゴとともに NEB Next Ultra II アダプター Q5 マスターミックス (New England Biolabs) を用い、 $6 \times 50 \mu\text{L}$ の PCR 反応を設定した。PCR 条件を最適化して、予想される 300 bp 長の清浄な可視生成物を有する最低サイクルを見出した。熱サイクル条件は、 98°C での最初の 30 秒、次いで 98°C 10 秒、 63°C 20 秒、及び 72°C 15 秒の 2 サイクル、続いてさらに 98°C 10 秒及び 72°C 30 秒の 10 サイクルであり、最終伸長は 72°C 5 分であった。プライマーは、Gateway 適合隣接配列を有するように設計された。得られたライブラリを、2% TAE ゲルに 60 V で 2 時間ローディングした後、QIAgen ゲル抽出キットによってゲル抽出し、その後、20 の別個の BP 反応を用いて合計 $5 \mu\text{L}$ の反応で pDONR221 にクローニングした。エントリープラスミドプールを一晩反応後に DH5 α コンピテント E. coli に形質転換し、カナマイシン ($100 \mu\text{g}/\mu\text{L}$) を含む LB 寒天プレート上で一晩インキュベートした。コロニーを収集し、プラスミド DNA を精製した。次いで、20 回の LR 反応を、前セクションに記載のように設定した。それぞれの反応物を NEB 10-ベータ化学的コンピテント E. coli に形質転換し、カルベニシリン ($100 \mu\text{g}/\mu\text{L}$) を含む LB 寒天プレート上で 30°C で一晩増殖させた。コロニーを計数して、クローニングのそれぞれの段階で 200 倍を超えるカバー範囲を確実にし、プールし、マキシプレッ

20

30

【0145】

プールされた活性化スクリーニング。PYL1 で C 末端をタグ化した ORFeome 及びトランス活性化タイリングライブラリをレンチウイルス粒子にパッケージした。ABI-dCas9 を安定に共発現するクローン EGFP レポーター細胞株及びプロモーターを標的とする gRNA を、ピューロマイシン ($1 \mu\text{g}/\text{mL}$) 選択後に約 30% の細胞生存を伴う低感染多重度 (MOI) で形質導入した。同じ条件下での形質導入されていない細胞は、完全に除去された。十分な細胞を形質導入して、ライブラリの 500 倍を超えるカバー範囲を維持した。細胞を $100 \mu\text{M}$ のアブシジン酸 (ABA, Sigma) で 48 時間処理することによって動員を誘導した。並行して、細胞の対照バッチを等しい体積の DMSO で処理した。次いで、細胞を PBS で洗浄し、解離緩衝液 (1 mM の EDTA、 10 mM の KCl、 150 mM の NaCl、 5 mM の重炭酸ナトリウム、 0.1% グル

40

50

コース)で処理し、フロー緩衝液(5mMのEDTA、25mMのHEPES pH7、1%BSA、PBS)に再懸濁した。それぞれのライブラリに対する高GFP集団、ORFeomeに対する上位1%、及びTADスクリーニングに対する上位1%及び次の4%の2つのピンを選別し、それらのゲノムDNAを、QIAmp DNA Blood Mini Kit(QIAGEN)を用いて直接抽出した。

【0146】

ORFeome配列決定。ネステッドPCRを、選別された集団からの精製されたゲノムDNAのすべて、または前集団からのゲノムDNAの少なくとも5µgを使用して行った。標的ORFeome領域を、T7プロモーター及びPYL1を標的とするプライマーを用いてゲノムDNAから増幅した。この反応の生成物をそれぞれの試料に対してプールし、さらに10サイクルの間、Gateway attB部位の外側を標的とするプライマーによって増幅した。その後、アンプリコンを1%アガロースゲル上で分離し、プライマーダイマーを除く任意の可視PCR産物をゲル精製した。Quant-iT 1X dsDNA HSキット(Thermo Fisher Scientific, Q33232)を用いてDNAを定量した後、Illumina DNA Prep, (M) Tagmentationキット(Illumina, 20018705)を用いて、6サイクルの増幅で、試料当たり50ngを処理した。精製したそれぞれの最終ライブラリ2µlをAgilent TapeStation HS D1000 ScreenTape(Agilent Technologies, 5067-5584)で泳動した。ライブラリを、Quant-iT 1X dsDNA HSキット(Thermo Fisher Scientific, Q33232)を用いて定量し、サイズ調整後に等モル比でプールした。最終プールを、Illumina用のNEBNext Library Quantキット(New England Biolabs, E7630L)を用いて定量し、Illumina MiSeq上でペアエンド配列決定した。

【0147】

TAD配列決定。第1のステップにおいて、T7プロモーター及びバックボーンベクターのPYL1を標的とするプライマーを用いて、精製ゲノムDNAに対してネステッドPCRを行い、約470bpの産物を作製した。次いで、第1の反応の産物をプールし、Gateway部位の外側を標的とするプライマーを使用して、さらに10段階で増幅し、約300bpの産物を作製した。ライブラリをQubit dsDNA Broad Rangeキット上で定量し、Illumina MiSeq上でカスタムPAGE精製R1配列決定プライマーを用いてペアエンド配列決定した。

【0148】

プールした活性化スクリーニングからの配列決定データの分析。ORFeome参照配列のインデックスを、STARアライナーv2.7.8aを使用して、より小さい「ゲノム」サイズを説明するために11に設定されたブレインデックス化文字列の長さで作成した。ORFeomeライブラリからの読み取りを、STARアライナーと整列させ、最大3つのミスマッチを可能にした。TAD配列決定リードに関して、クローニングアダプタ配列を、5'アダプターについてCCAGTGTGGTGGGAATCTGCAATACAAGTTTGTACAAAAAAGTTGGCGGAAGTCAGGGGTAGCGGGAAGT(配列番号20)を用い、3'アダプターについてCCGCCACTTGCTGGATACACTTTGTACTAGAAATGAATGGTTTGGTTGGGATTCTCGTTTCAACTACCTCC(配列番号21)でcutadaptを用いて最初に両端から除去した。Bowtie参照を生成し、0ミスマッチを許容してBowtie v1.2.3を使用してリードをマッピングした。活性化因子を同定するために、edgeRパッケージ(Robinson et al., 2010)を使用して、選別された試料から選別されていない細胞へのカウントの変化を比較することによって、それぞれのORFについてlog2倍率変化、p値、及び偽発見率(FDR)を計算した。

【0149】

アレイ型動員アッセイ。ABI-dCas9及びTetO gRNAを安定に共発現す

るレポーター細胞を、トランスフェクションの日に50～70%コンフルエンスに達するように、48ウェルまたは96ウェルプレートに播種した。試験するそれぞれの構築物150ngを、0.6μlの試薬比でポリエチレンジイミン(PEI)でトランスフェクトした。トランスフェクション後の日に、ABA(100μM)での処理によって動員を誘導した。阻害剤の存在下での連結レポーターアッセイのために、動員を、トランスフェクション後の日であるが、阻害剤または同じ容量の任意のさらなるDMSOのいずれかの存在下で、100μMのABAで同様に誘導した。阻害剤をDMSO中に10mMのストック濃度まで溶解した。使用した最終濃度は、フラボピリドールについては100nM、JQ1については300nM、A-485については1μM、CX-4945については2.5μM、AZ191については3μMであった。すべての阻害剤はStructural Genomics Consortium(SGC)からの親切なギフトであった。誘導後48時間で、細胞を解離させ、液体ハンドリングロボット(TECAN)を用いてフロー緩衝液中に再懸濁し、LSR Fortessa(BD)によって分析した。細胞を、それぞれgRNA及びトランスフェクション対照の尺度として、高EBFP2及びRFP上でゲーティングした。フローサイトメトリーデータを、FlowJo(v10)を用いて分析した。

10

【0150】

SRFレポーターアッセイ。24ウェルプレート上の30,000個のNIH-3T3細胞を、8ngのSRFレポーター(p3DA.luc)、20ngの参照レポーター(pCDNA3.1-Nanoluc-3xFLAG-V5)及び50ngの3xFLAGタグ化構築物(Addgene #87063)でトランスフェクトした。トランスフェクションは、Lipofectamine 3000試薬(Thermo Fisher Scientific, L3000001)を用いて、製造業者のプロトコールに従って行った。ルシフェラーゼ構築物は、Maria Vartiainen博士(ヘルシンキ大学)からの親切なギフトであった。細胞を低血清培地(0.5%BCS)中で18時間維持し、7時間刺激し(15%BCS)、その後、ルシフェラーゼ活性を測定した。4つの独立したトランスフェクションからのデータを用いて、Renilla luciferase活性に対してホタルルシフェラーゼを正規化した。

20

【0151】

RNA配列決定及び分析。EGFPでC末端にタグ付けされたSRF、C3orf62、SRF-C3orf62、またはNanolucの安定な組込みを有するNIH-3T3細胞を、1μg/mLドキシサイクリンで24時間誘導した。RNeasy purificationキット(Qiagen)を用いて、低血清条件(0.5%ウシ血清)で22時間維持した細胞からRNAを抽出し、カラム上でDNaseで処理した。試料を誘導し、6ウェルプレートから技術的に二重で収集した。ポリA選択キットとともにNEB Next Ultra II Directional RNA-seqを用いてライブラリを調製し、プールし、100サイクルNovaSeq 6000 SPで配列決定した。リードを、STAR aligner v2.7.8aを用いてGencodeマウス一次アセンブリ(GRCm39)に整列させた。それぞれの遺伝子のカウントを、Gencode vM26転写アノテーションを用いて生成した。Nanoluc-EGFPを発現する細胞と比較した遺伝子発現の変化を、edgeRパッケージを用いて定量した(Robinson et al., 2010)。

30

40

【0152】

質量分析試料。エントリークローンはヒトORFeome集合体から得た(Yang et al., 2011)。Gatewayリコンビナーゼを用いて、クローンをC末端BioID2-FLAGタグを有するpDEST-pcDNA5ベクター(Kim et al., 2016)に導入した。安定なHEK293 Flp-In T-REx細胞株を、以前に報告したように作製した(Piette et al., 2021)。

【0153】

AP-MSについては、細胞を150mmディッシュ上で70%コンフルエンスまで増

50

殖させた後、 $1\text{ }\mu\text{g}/\text{mL}$ テトラサイクリンを用いてベイト発現を24時間誘導した。次いで、細胞を1倍のPBSで1回洗浄し、掻き取り、ペレット化し、フラッシュ凍結し、そしてプロセッシングまで -80°C で保存した。AP-MSは、以前に記載されているように行った(Lambert et al., 2015)。簡潔に述べると、細胞を、1:4ペレットの重量:体積比を用いて低温溶解緩衝液(50 mM のHEPES-NaOH pH 8.0、 100 mM のKCl、 2 mM のEDTA、 0.1% NP40、 10% グリセロール、 1 mM のPMSF、 1 mM のDTT、 15 nM のカリクリンA、及びプロテアーゼ阻害剤カクテル(Sigma-Aldrich P8340))に再懸濁した。細胞を1回の凍結融解によって溶解し、溶解物を、 35% の振幅で2秒の休止を伴う3回の10秒バーストを使用して 4°C で超音波処理した。超音波処理した溶解物を、 4°C で30分間100Uベンゾナーゼで処理した後、 4°C で20分間20,000gの遠心分離により除去した。バッチ内で処理されたすべての試料からの等量の上清を、 $25\text{ }\mu\text{L}$ の予め洗浄された抗フラッグ磁気ビーズ50%スラリー(Sigma, M8823)を含むチューブに移し、 4°C で2時間インキュベートした。ビーズを磁化によって回収し、上清を廃棄した。以前に記載されているように(Taipale et al., 2014)、ビーズを溶解緩衝液中で1回洗浄し、 2 mM の CaCl_2 を含む 20 mM のTris-HCl pH 8.0で1回洗浄し、2段階でトリプシンを用いてビーズ上で消化した($1\text{ }\mu\text{g}$ トリプシンを4時間、続いて $0.5\text{ }\mu\text{g}$ トリプシンを上清に添加し、 37°C で一晩インキュベートした)。最後に、試料を 5% ギ酸(最終濃度)で酸性化し、 -80°C で保存した。

10

20

【0154】
BioIDの場合、細胞を 150 mM ディッシュ中で 70% コンフルエンスまで増殖させた後、 $1\text{ }\mu\text{g}/\text{mL}$ のテトラサイクリンで18時間、遺伝子発現を誘導した。次いで、 $50\text{ }\mu\text{M}$ のビオチンをそれぞれのプレートに6時間加えた。細胞ペレットをAP-MSと同様に回収し、1:10ペレットの重量:体積比を用いて溶解緩衝液(50 mM のTris-HCl pH 7.5、 150 mM のNaCl、 0.1% SDS、 1% IgEPAL CA-630、 1 mM のEDTA、 1 mM の MgCl_2 、プロテアーゼ阻害剤カクテル(Sigma-Aldrich P8340, 1:500)、及び 0.5% デオキシコール酸ナトリウム)に再懸濁した。超音波処理後、それぞれのサンプルを250UのTurbonuclease(BioVision 9207-50KU)及び $1\text{ }\mu\text{L}$ RNase A溶液(Sigma-Aldrich R6148)で処理し、 4°C で30分間インキュベートした。次いで、SDSを添加して最終濃度 0.25% とし、混合後、試料を 4°C でさらに10分間インキュベートし、続いて20分間20,000gで遠心分離した。上清を、 $30\text{ }\mu\text{L}$ の予め洗浄した充填ストレプトアビジンビーズ(GE Healthcare, 17-5113-01)を含むチューブに移した。ストレプトアビジンプルダウンを 4°C で3時間行った。ビーズを 1 mL のSDS緩衝液(2% SDS/ 50 mM のTris-HCl pH 7.5)中で1回、 1 mL の溶解緩衝液中で1回、及びTAP緩衝液(50 mM のHEPES-KOH pH 8.0、 100 mM のKCl、 10% グリセロール、 2 mM のEDTA、 0.1% IgEPAL CA-630)中で1回洗浄し、その後、 50 mM 重炭酸アンモニウム pH 8.0で 1 mL で3回洗浄した。洗浄後、ビーズを、 $1\text{ }\mu\text{g}$ のトリプシンを含むABC緩衝液中に再懸濁し、 37°C で一晩インキュベートした。翌日、上清を回収し、ストレプトアビジンビーズを $50\text{ }\mu\text{L}$ の水で洗浄し、これを第1の上清画分と合わせた。 $0.5\text{ }\mu\text{g}$ のトリプシンを合わせた上清試料に加え、次いで 37°C で4時間インキュベートした。次いで、ビーズを遠心沈殿させ、上清を回収した。ビーズをABC緩衝液を用いて2回リンスし、これらのリンス液を元の上清と組み合わせた。合わせた上清を遠心蒸発により乾燥した。

30

40

【0155】
質量分析データ取得及び分析(analysis)。試料を、TripleTOF 5600機器(AB SCIEX, Concord, Ontario, Canada)で、以前に記載されているように(Piette et al., 2021) Data-Dependent Acquisition (DDA)を用いて分析した。ProHits

50

v4.0 (Liu et al., 2016)、Mascot、及び Comet (Eng et al., 2013) に実装された Proteowizard (Adusumilli and Mallick, 2017) を用いて、データを処理し、前述のように分析した (Piette et al., 2021)。その後、iProphet (Sh teynberg et al., 2011) を用いて Trans-Proteomic Pipeline によって結果を分析し、iProphet 確率 ≥ 0.95 のタンパク質をさらに分析した。

【0156】

有意な相互作用因子を SAINTexpress (Teo et al., 2014) で同定した。EGFP-BioID2-FLAG 及び EGFP-BioID2-FLAG を陰性対照として使用し、先に記載したストリンジェンシーに 2 倍圧縮を使用した (Mellacheruvu et al., 2013)。SAINTexpress 分析はデフォルトパラメータを使用し、prey タンパク質は、Bayesian FDR カットオフ $\leq 5\%$ を計算に合格した場合に有意であると考えられた。ドットプロット図は、ProHit-viz webserver (Knight et al., 2017) を用いて作成した。

10

【0157】

実施例 8. 転写を活性化するために、複数の TAD を組み合わせることができる
複数の TAD を PYL1 と一緒に融合させ、上記の PYL1-ABI EGFP レポーター系を用いて転写活性について試験した。以下の TAD 断片を使用した：CITED1-8 (配列番号 47)、CITED2-12 (配列番号 49)、C3orf62-12 (配列番号 45)、BRD8-25 (配列番号 40)、ZXDC-12 (配列番号 97)、KLF7-1 (配列番号 72)、ATXN7L3-1 (配列番号 35)、FAM90A1-20 (配列番号 60)、SPDYE4-3 (配列番号 90)、YAF2-7 (配列番号 96)。

20

【0158】

図 16 に示すように、SPDYE4-CITED1 TAD 融合タンパク質は、いずれかの TAD 単独と比較して、転写活性化の顕著な増加をもたらした。

【0159】

p65 及び HSF1 から選択される最大 2 つのさらなる TAD を、異なる順序で SPDYE4-CITED1 TAD と組み合わせ、PYL1-ABI EGFP レポーター系 (図 17、左パネル) または内在性 CD133 遺伝子 (図 17、右パネル) の転写を活性化する能力について試験した。SPDYE4-CITED1-P65-HSF1 (SCPH) エフェクターは、試験した両方の系において最も強い活性化をもたらした。

30

【0160】

さらなる TAD の組み合わせを試験するために、多成分 SPDYE4-CITED1-P65-HSF1 (SCPH) エフェクターのそれぞれの部分を、実施例 4 において同定された別の TAD と個別に置き換えた (例えば、表 1 参照)。結果を図 18 に示す。

【0161】

より短い TAD 配列が SCPH エフェクターにおいて使用され得るかどうかを試験するために、それぞれの TAD を独立して、short または mini TAD 成分で置き換えた。結果を図 15 に示す。mini-TAD 配列を以下のように選択した：

40

【0162】

mini SPDYE4 (配列番号 102)：配列は、SPDYE4 全長タンパク質の本発明者らのタイリングスクリーニングからの 2 つの濃縮された断片間の重複に基づいて選択した。最小配列は、酸性リッチ領域または以下の β ヘアピンのいずれかを含むように設計した。

【0163】

mini HSF1：その活性化ドメインを含む HSF1 の 150 アミノ酸の C 末端領域は、2 つの既知の TAD を含む。より長い TAD は、アミノ酸 431~529 の間であり

50

、より短いもの（「mini-HSF1」または「H(mini)」；配列番号119）は、アミノ酸401～420の間である。本発明者らは、疎水性残基が豊富な短いドメインを、そのコンパクトなサイズのために試験するために選択し、以前に報告されたように、長いTADよりも高い効力を有することが報告された（Newton et al., 1996; 10.1128/MCB.16.3.839）。

【0164】

miniP65:RelAまたはp65は、そのC末端内に2つの異なるトランス活性化ドメインを有することが知られている。第1のTAD（「P(mini N-term)」；配列番号105）はアミノ酸428～520を含み、第2の（「P(mini C-term)」；配列番号106）は先行する521～551残基である（Schmitz and Baeuerle, 1991; 10.1002/j.1460-2075.1991.tb04950.x）。

10

【0165】

miniCITED1（「C(mini)」；配列番号103）を、CITED1-7とCITED1-8との間の重複領域内にC末端酸性リッチ領域を含むように選択した。

【0166】

方法：細胞を48-ウェルプレートに播種した。翌日、細胞が70～90%コンフルエントの間になった後、250ngのそれぞれの構築物を、（Thermo Scientific, 11668019）試薬を用いてそれぞれのウェルにトランスフェクトした。それぞれの構築物を、TagRFPと直接融合させるか、それぞれの成分と直接融合させるか、またはトランスフェクションの手段として用いる同じプラスミドから共発現させた。それぞれの実験において、RFP+細胞の同じゲートを用いて、活性に対するそれぞれの構築物の発現レベルの効果を制御した。EGFPレポーター細胞をクローン株から拡大し、ABI-dCas9及びプロモーターの上流のTetO7部位を標的とするgRNAのレベルを制御した。CD133誘導実験のために、ABI-dCas9を安定に発現する293T細胞株及びプロモーターを標的とする5つのgRNAのプールを、すべての活性化因子に使用した。APCにコンジュゲートしたCD133抗体（Miltenyi Biotec, 130-113-668）を使用した。すべての活性化因子構築物を、それらのC末端でPYL1に融合させ、細胞を1μMのアブシジン酸で24時間または48時間処理することによって、それらの標的に動員した。

20

30

【0167】

実施例9.

活性化ドメインまたは個々の活性化ドメインの断片のさらなる117の異なる組み合わせを、PYL1（dCas9-ABI1系で試験するため）またはドキシサイクリンの存在下でDNAに結合する転写因子であるrTetRのいずれかに融合することによってアッセイした。すべての構築物を同じTetOレポーター構築物で試験した。これらの構築物の活性に大きな差が観察され、活性が無い状態から非常に高い効力までの範囲であった（図19A～B）。一般に、rTetR及びdCas9ベースの動員システムによる結果の間に良好な相関が存在し（図19C）、これは、大部分の活性化ドメインが両方の状況において同様に機能することを示唆している。しかしながら、例えば、VPRは、dCas9融合体よりもrTetR融合体としてはるかに活性が低かった。以前の実験と一致して、SCPH及びそのバリエーションは、系において最も活性な構築物であった。例えば、CITED1（SCPHコンストラクト中の「C」）からの活性化ドメインの代わりにC3orf62の小型化された活性化ドメインと、p65活性化ドメインのより短いバージョン（「P」）とを含むコンストラクトは、著しく小さい（1059bp対1611bp）にもかかわらず、元のSCPHコンストラクトよりも強力であった（図19A～B）。

40

【0168】

方法：C末端にrTetRに融合したスーパー活性化因子を発現するプラスミド200ng及びDsRed 50ngを、7xTetO-EGFPレポーターを安定に発現するHEK293T細胞にトランスフェクトした。トランスフェクションの1日後、細胞を1

50

μg/ml ドキシサイクリンで48時間処理した。処理後、細胞をフローサイトメトリーによりEGFP発現について分析した

【0169】

本出願は、好ましい例であると現在考えられているものを参照して説明されてきたが、本出願は、開示された例に限定されないことが理解されるべきである。それとは反対に、本出願は、添付の特許請求の範囲の趣旨及び範囲内に含まれるさまざまな修正及び同等の変形を包含することが意図される。特許請求の範囲は、好ましい実施形態及び実施例によって限定されるべきではなく、全体として説明と一致する最も広い解釈を与えられるべきである。

【0170】

すべての刊行物、特許及び特許出願は、個々の刊行物、特許または特許出願が、その全体が参照により本明細書に組み入れられるように具体的かつ個別に示されているかのよう

10

配列表

グリシンセリンリンカー（配列番号6）SGGSGGS

グリシンセリンリンカー（配列番号7）SGGS

グリシンセリンリンカー（配列番号8）GSGSGS

リンカー（配列番号9）INSRSSGS

NLS（配列番号22）PKKKRKV

リンカー+NLS（配列番号141）INSRSSGSPKKRKVGS

20

>YAF2のYAF2__RYBPDメイン（配列番号147）

RPRLKNVDRSSAQHLEVTVGDLTVIITDFKEKT

>RYBPのYAF2__RYBPDメイン（配列番号148）

RPRLKNVDRSTAQQLA VTVGNVTVIITDFKEKT

>CBX2のCBX-Cドメイン（配列番号149）

QDWKPTRSLIEHVFVTDVTANLITVTVKESPTSV

>CBX6のCBX-Cドメイン（配列番号150）

GDWRPEMSPCSNVVVTDTVTSNLLTVTIKEFCNPE

>CBX8のCBX-Cドメイン（配列番号151）

ESWSPSLTNLEKVVVTDTVTSNFLT VTIKESNTDQ

30

>CBX4のCBX-Cドメイン（配列番号152）

SEFKPFFGNIIITDVTANCLTVTFKEYVTV

>CBX7のCBX-Cドメイン（配列番号153）

PPWTPALPSSEVTVTDITANSITVTFREAAAE

>SPDYE4-2（配列番号154）

VRSPDEVVVDDEVPGPSAPWIDPSPQPQSLGLKRKSEWSDE

SEEELEEELELERAPEPEDT

>SPDYE4-5（配列番号155）

WVVETLCGLKMKLKRKRASSVLP EHH EAFNRLLGDPVVQK

FLAWDKDLRVSDKYLLAMVI

40

【表 1 - 1】

表 1 : TADseqによりTADとして同定された断片。*印を付けた断片を独立して検証した。

遺伝子	断片	配列	配列番号	logFC 高GFP	FDR高G FP	ヒット高G FP	logFC 中程度GFP	FDR中 程度GFP	ヒット 中程度G FP
ATF6	ATF6- 1	MGEPA	29	3.145 27669 8	0.021 21147	ヒット	1.563 29624 1	6.04E -06	ヒット
		GVAGT							
		MESPF							
		SPGLF							
		HRLDE							
		DWDSA							
		LFAEL							
		GYFTD							
		TDELQ							
		LEAAN							
ATF6	ATF6- 2*	ETYEN	30	4.687 56189	2.94E -05	ヒット	1.477 26354 9	4.41E -05	ヒット
		NFDNL							
		DFDLD							
		HRLDE							
		DWDSA							
		LFAEL							
		GYFTD							
		TDELQ							
		LEAAN							
		ETYEN							

10

20

30

40

10

30

【表 1 - 3】

ATMI N	ATMIN -32	ETQTM SSGFE TLGSL FFTSN ETQTA MDDFL LADLA WNTME SQFSS VETQT SAEPH TVSNF	33	0.635 51575	0.305 5137	0.819 75921 4	0.009 05043 3	ヒット
ATOH 1	ATOH1 -1	MSRLL HAEFW AEVKE LGDHH RQPQP HLLPQ PPPPP QPPAT LQARE HPVYP PELSL LDSTD	34	0.721 82245 3	0.305 06874	0.919 35152 2	0.017 13035 9	ヒット
ATXN 7L3	ATXN7 L3-1*	MKME MSLSG LDNSK	35	1.637 90171 8	0.025 73555	1.683 89347 7	1.80E -05	ヒット

10

20

30

40

50

10

20

30

40

【表 1 - 5】

		RSIAA SRFAP HLEKC LGMGR NSSRI ANRRI								
BRD8	BRD8- 23	KEECF RSGVA EAPVG SKAPS IDGKE ELDLA EKMDI AVSYT GEELD FETVG DIIAI IEDKV	38	0. 588 42095 8	0. 357 87782			0. 591 49457 7	0. 032 98888 7	ヒット
BRD8	BRD8- 24	IDGKE ELDLA EKMDI AVSYT GEELD FETVG DIIAI IEDKV DDHPE	39	1. 593 19321 5	0. 046 62626	ヒット		1. 565 64597 1	2. 69E -05	ヒット

10

20

30

40

50

【表 1 - 6】

BRD8	BRD8- 25*	VLDVA AVEAA LSFCE	40	1. 567 47881 2	0. 020 72139	ヒット	1. 833 74780 2	6. 04E -06	ヒット
BTBD 18	BTBD1 8-24	IDCRE PYAFD TALLE QPCEA EYRI TSAAA TSELE EILDF MLCGS DIEPP IGSLE SPGAE	41	1. 246 17675 5	0. 150 33388		0. 856 10788 9	0. 003 21753 8	ヒット

10

20

30

40

50

【表 1 - 7】

BTBD 18	BTBD1 8-25	EEYRI TSAAA TSELE EILDF MLCGS DIEPP IGSLE SPGAE GCRTP TYHLT ETGKN WIEGE	42	1. 470 63567 9	0. 070 02876		1. 413 32846 5	7. 30E -05	ヒット
C11o rf74	C11or f74-1	MSAHM SGLEI MDEDQ LIKDV LDKFL NCHEQ TYDEE FLNTF THLSQ DLLL PGEVE QDVST	43	2. 356 35648 2	0. 012 114	ヒット	1. 551 23447 6	7. 99E -06	ヒット
C3or f62	C3orf 62-11	NHMCg HCQDS PFKEE	44	6. 022 71451	7. 93E -06	ヒット	-0. 52 60786 24	0. 221 47563 3	

10

20

30

40

50

【表 1 - 8】

		AWALL MDKSP QKATD ADPGS LKQAF DDHNI VETVL DLEED YNVMT							
C3orf62	C3orf62-12*	QDSPF KEEAW ALLMD KSPQK ATDAD PGSLK QAFDD HNIVE TVLDL EEDYN VMTSF KYQIE	45	6.492 95084 5	6.25E -06	ヒット	-0.60 66642 08	0.124 70190 7	
CITE D1	CITED 1-7	LQNWD FGAQA GGAES LSPSA GAQSP AIIIS	46	4.313 81114	8.10E -05	ヒット	1.596 63892 5	0.000 50565 1	ヒット

10

20

30

40

10

20

30

40

[illegible]

20

40

【表 1 - 1 1】

DDIT 3	DDIT3 -1	MAAES LPFSF GTLSS WELEA WYEDL QEVLS SDENG GTYSVS PPGNE EEESK IFTTL DPASL	51	2. 372 68470 4	0. 006 46538	ヒット	1. 501 77863 4	6. 04E -06	ヒット
DDIT 3	DDIT3 -2	WYEDL QEVLS SDENG GTYSVS PPGNE EEESK IFTTL DPASL AWLTE EEPEP AEVTS TSQSP	52	1. 138 25374 7	0. 115 02125		1. 279 97930 9	0. 000 33039 2	ヒット
ELF4	ELF4- 1	MAITL QPSDL IFEFA	53	0. 662 55547 4	0. 257 87399		0. 728 70896 8	0. 015 52453 1	ヒット

10

20

30

40

50

10

20

30

40

10

20

30

40

10

20

30

40

【表 1 - 1 5】

FAM9 0A1	FAM90 A1-20 *	HPAAS HDGAQ PLRVL FRRLE NGRWS SSLT APSFH SPEKP GAFLA QSPHV SEKSE GPCVR	60	3. 016 55115 5	0. 002 81867	ヒット	1. 616 16400 6	0. 000 2923	ヒット
HOXA 2	HOXA2 -4	PGSHP RHGAG GRPKP SPAGS RGSPV PAGAL QPPEY PWMKE KKAAG KTALL PAAAA AATAA	61	1. 664 42718 1	0. 234 42147	ヒット	1. 594 81672 8	0. 021 80528 5	ヒット
HOXA 2	HOXA2 -5	RGSPV PAGAL QPPEY	62	1. 684 06493 4	0. 249 06108	ヒット	2. 311 48338	0. 000 38479 4	ヒット

10

20

30

40

50

10

20

30

40

【表 1 - 1 7】

[illegible]

10

20

30

40

【表 1 - 1 8】

		FAEDP TISLL TGSEP								
HSF1	HSF1 - 25	AQPLF LLDPG SVDTG SNDLP VLFEL GEGSY FSEGD GFAED PTISL LTGSE PPKAK DPTVS	67	0. 772 54878 1	0. 220 14105		1. 152 58234 8	0. 000 2923		ヒット
JAZF 1	JAZF1 -1	MTGIA AASFF SNTCR FGGCG LHFPT LADLI EHIED NHIDT DPRVL EKQEL QQPTY VALSY	68	0. 284 75864 7	0. 673 0689		0. 663 87767 4	0. 031 99195 5		ヒット

10

20

30

40

50

【表 1 - 1 9】

KLF1 5	KLF15 -6	GPVAV GPWRR AAPV KGEHF CLPEF PLGDP DDVPR PFQPT LEEIE EFLEE NMEPG VKEVP	69	2.466 55775 5	0.003 12371	ヒット	1.201 85423	0.000 20370 1	ヒット
KLF6	KLF6- 1	MDVLP MCSIF QELQI VHETG YFSAL PSLEE YWQQT CLELE RYLQS EPCYV SASEI KFDSQ	70	5.453 10764 9	6.25E -06	ヒット	1.867 75368 9	6.47E -06	ヒット
KLF7	KLF7- 1*	MDVLA SYSIF QELQL	72	5.997 66143 7	6.25E -06	ヒット	1.289 60310 1	0.002 20682 5	ヒット

10

20

30

40

50

10

20

30

40

10

20

30

40

【表 1 - 2 2】

		GLGPG AGDPA PGIGP								
MYOD 1	MYOD1 -2	CSFAT TDDFY DDPCF DSPDL RFFED LDPRL MHVGA LLKPE EHSHF PAAVH PAPGA REDEH	77	0.702 53999 8	0.301 42347		0.848 60205 4	0.002 49624 2	ヒット	
NEUR OD1	NEURO D1-6	LRRMK ANARE RNRMH GLNAA LDNLR KVVPC YSKTQ KLSKI ETLRL AKNYI WALSE ILRSG	78	0.670 79442 6	0.295 41863		0.953 33861 9	0.001 35422 7	ヒット	

10

20

30

40

50

【表 1 - 2 3】

NEUR OG 3	NEURO G 3-5	RRSRR KKAND RERNR MHNLN SALDA LRGVL PTFPD DAKLT KIETL RFAHN YIWAL TQTLR	7 9	0. 0 6 0 3 8 7 0 5 9	0. 9 1 7 9 3 7 1 9		0. 6 7 1 3 3 1 7 8 2	0. 0 3 8 3 9 6 9 0 8	ヒット
NPAS 4	NPAS 4 - 3 1	DVPLV PEGLL TPEAS PVKQS FFHYS EKEQN EIDRL IQQIS QLAQQ MDRPF SAEAG TGGL E	8 0	1. 2 2 0 0 8 4 5 7 5	0. 1 4 2 0 8 5 2 2		0. 8 5 6 3 3 9 0 3	0. 0 3 4 7 1 3 1 4 7	ヒット
NPAS 4	NPAS 4 - 3 4	PLGGL EPLDS NLSLS	8 1	0. 9 4 5 9 7 6 8 7 6	0. 1 4 2 8 9 4 8 4		0. 9 1 8 6 4 2 8 2 8	0. 0 0 3 1 9 0 4 8 6	ヒット

10

20

30

40

50

10

40

【表 1 - 2 5】

[illegible]

10

20

30

40

【表 1 - 2 6】

REL B	REL B - 2	SQER I I QFQA TPCPK	8 6	2. 1 4 6 5 4 0 2 9 7	0. 0 0 2 8 1 8 6 7	ヒット	1. 7 1 0 8 6 7 8 3 5	6. 9 9 E - 0 6	ヒット
REL B	REL B - 3	SSLSL AVSRS TDELE I IDEY IKENG FGLDG GQPGP GGLP	8 7	1. 7 6 8 3 9 2 7 8 3	0. 0 1 1 2 6 9 4 2	ヒット	1. 8 0 2 9 0 1 1 7 4	6. 7 6 E - 0 7	ヒット

10

20

30

40

50

【表 1 - 2 7】

SERT AD1	SERTA D1-10	IDTSM YDNEL WAPAS EGLKP GPEDG PGKEE APELD EAELD YLM DV LVGTQ ALERP PGPGR	88	0. 234 94619 6	0. 841 61029		0. 838 50788 2	0. 032 98888 7	ヒット
SERT AD2	SERTA D2-11	SEAGT QKLDG PQESR ADDSK LMDSL PGNFE ITTS GFLTD LTLDD ILFAD IDTSM YDFDP	89	2. 480 48985 5	0. 006 65441	ヒット	1. 821 76653 1	4. 96E -06	ヒット
SPDY E4	SPDYE 4-3*	DPSPQ PQSLG LKRKS	90	4. 820 49587 8	2. 26E -05	ヒット	2. 409 36057	1. 30E -07	ヒット

10

20

30

40

50

10

20

30

40

10

20

30

40

10

20

30

40

【表 1 - 3 1】

ZXDC	ZXDC- 1 2*	SPA EQ HGA QD TELSA GTGNF YLESG GSART DYRAI QLAKE KKQRG AGSNA GASQS TQRKI	9 7	3. 9 6 5 7 4 8 2 5 5	0. 0 1 9 6 7 7 0 5	ヒット	2. 4 9 6 0 6 1 0 7 8	8. 0 8 E - 0 8	ヒット
ZXDC	ZXDC- 1 3	YLESG GSART DYRAI QLAKE KKQRG AGSNA GASQS TQRKI KEGKM SPPHF HASQN SWLCG	9 8	2. 3 7 7 2 8 1 8 7 2	0. 0 0 3 0 7 6 4 6	ヒット	1. 7 3 4 2 3 1 1 1 9	8. 5 7 E - 0 6	ヒット
ZXDC	ZXDC- 1 6	SLVVP SGGRP GPAPA	9 9	0. 4 7 8 4 9 6 0 5 7	0. 4 1 7 2 8 7 6 1		0. 7 2 9 2 5 6 4 5 9	0. 0 1 7 9 2 5 8 4 3	ヒット

10

20

30

40

50

【表 1－3 2】

AGVQC GAQGV QVQLV QDDPS GEGVL PSARG PATFL PFLTV DLPVY

10

20

30

40

50

【表 2 - 1】

表 2 : 表 1 の TAD seq により TAD として同定された断片のサブセット。

遺伝子	断片	配列	配列番号
ATMIN	ATMIN-24	NPGPDTQLPSGPAQNPGIDFDIEE FFSASNIQTQTEESELSTMTTEPV LESLDIETQTDF	32
ATMIN	ATMIN-32	ETQTMSSGFETLGSLFFTSNETQT AMDDFLFLADLAWNTMESQFSSVET QTS AEPHTVSNF	33
ATOH1	ATOH1-1	MSRLLHAE EWAEVKELGDHHRQPQ PHHLPQPPPPPPQPPATLQAREHPV YPPELSLLDSTD	34
ATXN7L3	ATXN7L3-1	MKMEEMSLSGLDNSKLEAIAQE IY ADLVEDSCLGFCFEVHRAVKCGYF FLDDTDPDSMKD	35
ATXN7L3	ATXN7L3-2	QE IYADLVEDSCLGFCFEVHRAVK CGYFFLDDTDPDSMKDFE IVDQPG LDIFGQVFNQWK	36
ATXN7L3	ATXN7L3-4	FE IVDQPGLDIFGQVFNQWKSKEC VCPNCSRSIAASRFAPHLEKCLGM GRNSSRIANRRI	37
BRD8	BRD8-23	KEECFRSGVAEAPVGSKAPSIDGK EELDLAEKMDIAVSYTGEELDFET VGDI IAI IEDKV	38
BRD8	BRD8-24	IDGKEELDLAEKMDIAVSYTGEEL DFETVGD I IAI IEDKVDDHPEVLD VAAVEAALS FCE	39
BRD8	BRD8-25	GEELDFETVGD I IAI IEDKVDDHP EVL DVAAVEAALS FCEENDDPQSL PGPWEHP IQQER	40
BTBD18	BTBD18-24	IDCREPYAFDTALLEQPCEAE EYR ITSAAATSELEE I LDFMLCGSDIE PPIGSLESPGAE	41
BTBD18	BTBD18-25	E EYRITSAAATSELEE I LDFMLCG SDIEPPIGSLESPGAEGCRTPTYH LTETGKNWIEGE	42

10

20

30

40

【表 2－2】

CS RN P 1	CSRNP 1－28	DNIEAPHFPLPGLSPPGDASSCFL ESLMGFSEPAAEALDPFIDSQFED TVPASLMEPVVPV	50
HO XA 2	HOXA2 －4	PGSHPRHGAGGRPKPSPAGSRGSP VPAGALQPPEYPWMKEKKAACKTA LLPAAAAAATAA	61
HO XA 2	HOXA2 －5	RGSPVPAGALQPPEYPWMKEKKA KKTALLPAAAAAATAAATGPACLS HKESLEIADGSG	62
HO XA 2	HOXA2 －6	KKAACKTALLPAAAAAATAAATGP ACLSHKESLEIADGSGGGSRRRLRT AYTNTQLLELEK	63
JA ZF 1	JAZF1 －1	MTGIAAASFSSNTCRFGGCGLHFP TLADLIEHIEDNHIDTDPRVLEKQ ELQQPTYVALSY	68
LI N9	LIN9－ 25	SRLTAILLQIKCLAEGGDLNSFEF KSLTDSLNDIKSTIDASNISCFQN NVEIHVAHIQSG	74
MY CB P	MYCBP －1	MAHYKAADSKREQFRRYLEKSGVL DTLTKVLVALYEEPEKPNSALDFL KHHLGAATPENP	75
MY CL 1	MYCL1 －1	MDYDSYQHIFYDYDCGEDFYRSTA PSEDIWKKFELVPSPPTSPPWGLG PGAGDPAPGIGP	76
NE UR OD 1	NEURO D1－6	LRRMKANARERNRMHGLNAALDNL RKVVPCYSKTQKLSKIETLRLAKN YIWALSEILRSG	78
NE UR OG 3	NEURO G3－5	RRSRRKKANDRERNRMHNLNSALD ALRGVLPTFPDDAKLTKIETLRFA HNYIWALTQTLR	79
RB PJ	RBPJ－ 13	GMALPRLIIRKVDKQTALLDADDP VSQLHKCAFYLKDTERMYLCLSQE RIIQFQATPCPK	85

10

20

30

40

50

【表 3－1】

表 3：TAD及びその機能的バリエーション及び活性断片の例示的な配列。

TAD	配列番号	配列
YAF2 short	100	SHKKTRPRLKNVDRSSAQHLEVTVG DLTVIITDFKEKTKSPPASSA
C3orf 62 short	101	MDKSPQKATDADPGSLKQAFDDHNI VETVLDLEEDYNVMTSFKYQIE
「mini SPDYE 4」または 「S (mi ni)」	102	SDESEEELEEELELERAPEPEDTWV VETL
「mini CITED 1」または 「C (mi ni)」	103	DSDPVDEEVLMSLVVELGLDRANEL
HSF1 short	104	GFSVDTSALLDLFSP
「p65 short (N-term)」； 「P (mi ni N- term)」； 「またはP __Nt」	105	PTQAGEGTLSEALLQLQFDDDEDLGA LLGNSTDPVFTDLASVDNSEFQQ
「p65 short (C-term)」； 「P (mi ni C- term)」；ま たは「P__ Ct」	106	APGLPNGLLSGDEDFSSIADMDFSA LL
ZXDC	107	AGTGNFYLESGGSARTDYRAIQLAK

10

20

30

40

50

【表 3－2】

short		EKKQRGAGSNAGASQSTQRKIKEGKM
RELBshort	108	SSLSLAVSRSTDELEIIDEYIKENG FGLDGGQPGPGEGLP
FAM22 Fshort	109	LSYIDKLCSQEDFVTKVEAVIHPRF LEELLS PDPQMDFLA
KLF6－ TAD	71	DVLPMC SIFQELQIVHETGYFSALP SLEEYWQQTCLELERYLQSEPCYVS ASEIKFDSQ
「C3or f62－T AD」；ま たは「mi niC3o rf62」	110	QDSPFKEEAWALLMDKSPQKATDAD PGSLKQAFDDHNI VETVLDLEEDYN VMTSFKYQIEGSS
SOX7 C－term	111	LGQLSPPEHPGFDALDQLSQVELL GDMDRNEFDQYLNTPGHPDSATGAM ALSGHVPVSQV
FOXMI (697～ 762)	112	DVPKPGSPEPQVSGLAANRSLTEGL VLDTMNDSLSKILLDISFPGLDEDP LGPDNINWSQFIPELQ
ATF6	113	HRLEDWD S ALFAELGYFTDTDELQ LEAANETYENNFDNLDFDLVLPWE SDIWDINNQIGSS
「SERT AD2 _ NAHL」	89	SEAGTQKLDGPQESRADD SKLMDSL PGNFEIT TSTGFLTDLTLD DILFAD IDTSMYDFDP
「DDIT 3－1」； または「m iniDD IT3」	51	MAAESLPFSFGTLSSWELEAWYEDL QEVLS SDENG GTYVSPPGNEEEESK IFTTLDPASL
ELF4	54	DDIHQLEDPSVFP AVIVEQVPYPDL LHLYSGLELDDVHNGIITDGTLCMT QDQILEGSFL
C11or f74	114	SAHMSGLEIMDEDQLIKDVLDKFLN CHEQTYDEEFLNTFTHLSQDLLLLP GEVEQDVST
「mini ATMI N」	32	NPGPDTQLPSGPAQNP GIDFDIEEF FSASNIQTQTEESELSTMTTEPVLE SLDIETQTDF

10

20

30

40

50

【表 3 - 3】

「mini ZXDC」	115	YLESGGSARTDYRAIQLAKEKKQRG AGSNAGASQSTQRKI
「SPDY E4」；ま たは「S」	90	DPSPQPQSLGLKRRKSEWSDESEEL EELELERAPEPEDTWVVETLCGLK MKLKRKRASS
「CITE D1」また は「C」	47	ESLSPSAGAQSPAIIIDSDPVDEEVL MSLVVELGLDRANELPELWLQGNEF DFTADFPSSC
「RELA (p6 5)」； 「REL A」または 「p65」	116	QYLPDTDDRHRIEEKRRKRTYETFKS IMKKSPFSGPTDPRPPRRRIAVPSR SSASVPKPAPQPYPFTSSLSTINYD EFPTMVFPSPGQISQASALAPAPPQV LPQAPAPAPAPAMVSALAQAPAPVP VLAPGPPQAVAPPAPKPTQAGEGTL SEALLQLQFDDDEDLGALLGNSTDPA VFTDLASVDNSEFQQLLNQGI PVAP HTTEPMLMEYPEAITRLVTGAQRPP DPAPAPPLGAPGLPNGLLSGDEDFSS IADMDFSALL
「p65- AD」； 「P__sh ort」； または 「P」	117	PTQAGEGTLSEALLQLQFDDDEDLGA LLGNSTDPAVFTDLASVDNSEFQQLL NQGI PVAPHTTEPMLMEYPEAITR LVTGAQRPPDPAPAPPLGAPGLPNGL LSGDEDFSS IADMDFSALL
「HSF1 -TA D」；「H _TA D」；「H SF1」ま たは「H」	118	GFSVDTSALLDLFSPSVTPDMSLP DLDSSLASIQELLSPQEPPRPPEAE NSSPD SGKQLVHYTAQPLFLLDPGS VDTG SNDLPVLFELGEGSYFSEGDG FAEDPTISLLTGSEPPKAKDPTVS
「mini -HSF1 (401~ 420)」 または「H (min i)」	119	MLSSHGFSVDTSALLDLFSP
CBX2 (CBX- C)	136	SASPPSTGQNPSVSVQTSQDWKPTR SLIEHVFTDVTANLITVTVKESPT SVGFFNLRHY

10

20

30

40

50

【表 3－4】

CBX6 (CBX－ C)	138	GASSEPEAGDWRPEMSPCSNVVVT DVTSNLLTVTIKEFCNPEDFEKVAAG VAGAAGGGGSIGASK
CBX8 (CBX－ C)	139	RPSLIARIPVARILGDPEEESWSPS LTNLEKVVVTDVTSNFLTVTIKESN TDQGFFKEKR
RYBP－ C (YAF 2_RYB P)	140	IQSANATTKTSETNHTSRPRLKNVD RSTAQQLAVTVGNVTVIITDFKEKT RSSSTSSSTVTSSAGSEQNQS
DDIT3 _MT	156	SFGTLSSWELEAWYEDLQEVLSSE DENGTTYVSPPGNEEEESKIFTTLDPA SLAWLT
SPDYE 4_MT	157	ERAPEPEDTWVVETLCGLKMKLKRK RASS
ZXDC_ MT	158	TGNFYLES GGSARTDYRAIQLAKEK KQRGAGSNAGASQSTQRKI
HSF1_ MT	159	GFSVDT SALLDLFSPSVTVPDMSLP DLDSSLASIQELLSPQEPPR
SERTA D2_MT	160	LMDSLPGNF EITTSTGFLTDLTLDD ILFADIDTSMYDFDP
ZNF47 3	161	MAEEFVTLKDVGMDFTLGDWEQLGL EQGDTFWD TALDNCQDLFLLDPPRP NLTSHPDGSEDL EPLAGGSPEATS
CITED 1_MT	162	AIIDSDPVDEEVLM SLVVELGLDRA NELPELWLGQNEFDFTADF
ZNF49 6	163	MQQVTTLQLPPSRVSPFKDMILCFS EEDWSLLDPAQTGFYGEFIIGEDYG VSMPPNDLAAQPDL SQGEENEPRVP ELQDLQGKE
CITED 2_MT	164	VIDTDFIDEEVLM SLVIEMGLDRIK ELPELWLGQNEFD FMTDFV
mini－ p65_M T	165	PTQAGEGTLSEALLQLQFDD EDLGA LLGNSTDP AVFTDLASVDNSEFQQL LNQG
C3orf 62－2	166	ATDADPGSLKQAFDDHNIVETVLDL EEDYNVMT
C3orf 62－3	167	DDHNIVETVLDLEEDYNVMT
miniC 3orf6 2_1	168	DADPGSLKQAFDDHNIVETVLDLEE D

10

20

30

40

50

【表 3－5】

m i n i C 3 o r f 6 2 _ 2	1 6 9	D S P F K E E A W A L L M D K S P Q K A T D A D P G S L
m i n i D D I T 3 _ 1	1 7 0	D L Q E V L S S D E N G G T Y V S P P G N E E E E S K I F T T L D
m i n i D D I T 3 _ 2	1 7 1	L S S W E L E A W Y E D L Q E V L S S D E
m i n i C 3 o r f 6 2 _ M T	1 7 2	A T D A D P G S L K Q A F D D H N I V E T V L D L E E D Y N V M T S F K Y Q I E G S S
m i n i S P D Y E 4 _ 1	1 8 5	E R A P E P E D T W V V E T L

10

20

30

40

50

【表 4－1】

表 4： 表 1 の TAD seq により TAD として同定された断片のサブセット。

遺伝子	断片	配列	配列番号
ATM IN	ATMI N-24	NPGPDTQLPSGPAQNPGIDFD IEEFFSASN IQTQTEESELS TMTTEPVLES LD IETQTDF	32
ATM IN	ATMI N-32	ETQTMSSGFETLGSLFFTSNE TQTAMDDFL LADLAWNTMESQ FSSVETQTS AEPHTVSNF	33
ATO H1	ATOH 1-1	MSRLLHAEWEAEVKELGDHHR QPQPHHLPQPPPPPPQPPATLQ AREHPVYPPELS LLDSTD	34
ATX N7L 3	ATXN 7L3- 1	MKMEEMSLSGLDNSKLEAIAQ E IYADLVEDSCLGFCFEVHRA VKCGYFFLDDTDPDSMKD	35
ATX N7L 3	ATXN 7L3- 2	QE IYADLVEDSCLGFCFEVHR AVKCGYFFLDDTDPDSMKDFE I V D Q P G L D I F G Q V F N Q W K	36
ATX N7L 3	ATXN 7L3- 4	FE I V D Q P G L D I F G Q V F N Q W K S KECVCPCNCSRS I AASRFAPHL EKCLGMGRNSSRI ANRRI	37
BRD 8	BRD8 -23	KEECFRSGVAEAPVGSKAPSI DGKEELDLAEKMDI AVSYTGE ELDFETVGD I I A I I E D K V	38
BRD 8	BRD8 -24	IDGKEELDLAEKMDI AVSYTG EELDFETVGD I I A I I E D K V D D HPEVLDVA AVEAALS FCE	39
BRD 8	BRD8 -25	GEELDFETVGD I I A I I E D K V D DHPEVLDVA AVEAALS FCEEN DDPQSLPGPWEHP I Q Q E R	40
BTB D18	BTBD 18-2 4	IDCREPYAFDTALLEQPCEAE EYRITSAAATSELEE I L D F M L CGSDIEPP I G S L E S P G A E	41
BTB D18	BTBD 18-2 5	EEYRITSAAATSELEE I L D F M LCGSDIEPP I G S L E S P G A E G C RTPTYHLTETGKNWIEGE	42
C11 orf 74	C11o rf74 -1	MSAHMSGLEIMDEDQLIKDVL DKFLNCHEQTYDEEFLNTFTH LSQDLLLLPGEVEQDVST	43
C3o	C3or	NHMCGH C Q D S P F K E E A W A L L M	44

10

20

30

40

50

【表 4－2】

r f 6 2	f 6 2－ 1 1	DKSPQKATDADPGSLKQAFDD HNIVETVLDLEEDYNVMT	
C 3 o r f 6 2	C 3 o r f 6 2－ 1 2	QDSPFKEEAWALLMDKSPQKA TDADPGSLKQAFDDHNIVETV LDLEEDYNVMTSFKYQIE	4 5
FAM 2 2 F	FAM2 2 F－1 9	PRPQRPAETNAHLPPRPQRP AETKVPEEIPPEVVQEYVDIM EELLGSHPGDTGEPEGQR	5 5
FAM 2 2 F	FAM2 2 F－2 0	PAETKVPEEIPPEVVQEYVDI MEELLGSHPGDTGEPEGQREK GKVEQPQEEDGITSDPGL	5 6
FAM 2 2 F	FAM2 2 F－2 2	EKGKVEQPQEEDGITSDPGLL SYIDKLCSEQEDFVTKVEAVIH PRFLEELLSPDPQMDFLA	5 7
FAM 2 2 F	FAM2 2 F－2 3	LSYIDKLCSEQEDFVTKVEAVI HPRFLEELLSPDPQMDFLALS QELEQEEGLTLAQLVEKR	5 8
FAM 9 0 A 1	FAM9 0 A 1－ 1 9	RQPPHSRPLPTAQAQCTMSHH PAASHDGAQPLRVLFRRLENG RWSSSLLTAPSFHSPEKP	5 9
FAM 9 0 A 1	FAM9 0 A 1－ 2 0	HPAASHDGAQPLRVLFRRLEN GRWSSSLLTAPSFHSPEKPGA FLAQSPHVSEKSEGPCVR	6 0
HOX A 2	HOXA 2－4	PGSHPRHGAGGRPSPAGSR GSPVPAGALQPPEYPWMKEKK AAKKTALLPAAAAAATAA	6 1
HOX A 2	HOXA 2－5	RGSPVPAGALQPPEYPWMKEK KAAKKTALLPAAAAAATAAAT GPACLSHKESLEIADGSG	6 2
HOX A 2	HOXA 2－6	KKAAKKTALLPAAAAAATAAA TGPACLSHKESLEIADGSGGG SRRRLRTAYTNTQLLELEK	6 3
JAZ F 1	JAZF 1－1	MTGIAAASFNTCRFGGCGL HFPTLADLIEHIEDNHIDTDP RVLEKQELQQPTYVALSY	6 8
LIN 9	LIN9 －2 5	SRLTAILLQIKCLAEGGDLNS FEFKSLTDSLNDIKSTIDASN ISCFQNNVEIHVAHIQSG	7 4
MYC BP	MYCB P－1	MAHYKAADSKREQFRRYLEKS GVLDTLTKVLVALYEEPEKPN SALDFLKHHLGAATPENP	7 5
MYC	MYCL	MDYDSYQHYFYDYDCGEDFYR	7 6

10

20

30

40

50

【表 4 - 3】

L 1	1 - 1	STAPSEDIWKKFELVPSPPTS PPWGLGPGAGDPAPGIGP	
NEU ROD 1	NEUR OD 1 - 6	LRRMKANARERNRMHGLNAAL DNLRKVVPCYSKTQKLSKIET LRLAKNYIWALSEILRSG	7 8
NEU ROG 3	NEUR OG 3 - 5	RRSRKKKANDRERNRMHNLNS ALDALRGVLPPTFPDDAKLTKI ETLRF AHNYIWALTQTLR	7 9
RBP J	RBP J - 1 3	GMALPRLIIRKVDKQTALLDA DDPVSQ LHKCAFY LKDTERMY LCLSQERIIQFQATPCPK	8 5
SPD YE 4	SPDY E 4 - 3	DPSPQPQSLGLK RKSEWSDS EEEELEEELELERAPEPEDTWV VETLCGLKMKLKRKRASS	9 0
SPD YE 4	SPDY E 4 - 4	SEEELEEELELERAPEPEDTW VVETLCGLKMKLKRKRASSVL PEHHEAFNRLLGDPVVQK	9 1
SS 1 8 L 2	SS 1 8 L 2 - 2	QETIQRLLEENDQLIRCI VEY QNKGRGNECVQYQHVLHRNLI YLAT I ADASPTSTSKAME	9 2
YAF 2	YAF 2 - 6	EKKDKVEKEKSEKETTSKKNS HKKTRPRLKNVDRSSAQHLEV TVGDLTVIITDFKEKTKS	9 5
YAF 2	YAF 2 - 7	SHKKTRPRLKNVDRSSAQHLE VTVGDLTVIITDFKEKTKSPP ASSAASADQHSQSGSSSD	9 6
ZXD C	ZXDC - 1 2	SPA EQHGAQDTELSAGTG NFY LESGGSARTDYRAIQ LAKEKK QRGAGSNAGASQSTQRKI	9 7
ZXD C	ZXDC - 1 3	YLESGGSARTDYRAIQ LAKEK KQRGAGSNAGASQSTQRKIKE GKMSPPHFHASQNSWLCG	9 8
ZXD C	ZXDC - 1 6	SLVVPSGGRPGPAPAAGVQCG AQQVQVQLVQDDPSGEGVLPS ARGPATFLPFLTVDLPVY	9 9

10

20

30

40

50

【表 5】

表 5 : TADseq によって TAD として同定された活性断片のサブセット。

遺伝子	断片	配列	配列 番号	以前に同定さ れた TAD の 位置
ATF6	ATF6 - 1	MGE P A G V A G T M E S P F S P G L F H R L D E D W D S A L F A E L G Y F T D T D E L Q L E A A N E T Y E N N F D N L	29	アミノ酸 1 ~ 150
ATF6	ATF6 - 2	H R L D E D W D S A L F A E L G Y F T D T D E L Q L E A A N E T Y E N N F D N L D F D L D L V P W E S D I W D I N N Q I	30	アミノ酸 1 ~ 150
ATF6 B	ATF6 B - 1	M A E L M L L S E I A D P T R F F T D N L L S P E D W G L Q N S T L Y S G L D E V A E E Q T Q L F R C P E Q D V P F D G	31	アミノ酸 2 ~ 86
CSRN P1	CSRN P1 - 2 8	D N I E A P H F P L P G L S P P G D A S S C F L E S L M G F S E P A A E A L D P F I D S Q F E D T V P A S L M E P V P V	50	アミノ酸 49 3 ~ 583
KLF6	KLF6 - 1	M D V L P M C S I F Q E L Q I V H E T G Y F S A L P S L E E Y W Q Q T C L E L E R Y L Q S E P C Y V S A S E I K F D S Q	70	アミノ酸 1 ~ 200
KLF7	KLF7 - 1	M D V L A S Y S I F Q E L Q L V H D T G Y F S A L P S L E E T W Q Q T C L E L E R Y L Q T E P R R I S E T F G E D L D C	72	アミノ酸 1 ~ 75
KLF7	KLF7 - 2	Y F S A L P S L E E T W Q Q T C L E L E R Y L Q T E P R R I S E T F G E D L D C F L H A S P P P C I E E S F R R L D P L	73	アミノ酸 1 ~ 75

10

20

30

40

50

【表 6 - 1】

表 6 : TADとして同定された断片のサブセット。

遺伝子	断片	配列	配列番号
ATMIN	ATMIN-24	NPGPDTQLPSGPAQNPGIDDI EFFSASNIQTQTEESELSTTP LESLDIEQTDF。	32
ATMIN	ATMIN-32	ETQTMSSGFETLGSLFFTSNE TQTAMDDFL LADLAWNTMESQ FSSVETQTS AEPHTVSNF	33
ATOH1	ATOH1-1	MSRLLHAEWEAEVKELGDHHR QPQPHHLPQPPPPPPQPPATLQ AREHPVYPPELSLLDSTD	34
ATXN7L3	ATXN7L3-1	MKMEEMSLSGLDNSKLEAIAQ E IYADLVEDSCLGFCFEVHRA VKCGYFFLDDTDPDSMKD	35
ATXN7L3	ATXN7L3-2	QE IYADLVEDSCLGFCFEVHR AVKCGYFFLDDTDPDSMKDFE I V D Q P G L D I F G Q V F N Q W K	36
ATXN7L3	ATXN7L3-4	FE I V D Q P G L D I F G Q V F N Q W K S KECVCPCNCSRSIAASRFAPHL EKCLGMGRNSSRIANRR I	37
BRD8	BRD8-23	KEECFRSGVAEAPVGSKAPSI DGKEELDLAEKMDIAVSYTGE ELDFETVGD I I A I I E D K V	38
BRD8	BRD8-24	IDGKEELDLAEKMDIAVSYTG EELDFETVGD I I A I I E D K V D D HPEVLDVA AVEAALSFC E	39
BRD8	BRD8-25	GEELDFETVGD I I A I I E D K V D DHPEVLDVA AVEAALSFC EEN DDPQSLPGPWEHP I Q Q E R	40
BTBD18	BTBD18-24	IDCREPYAFDTALLEQPCEAE EYRITSAAATSELEEILDFML CGSDIEPPIGSLES PGAE	41
BTBD18	BTBD18-25	EEYRITSAAATSELEEILDFM LCGSDIEPPIGSLES PGAE GC RTPTYHLTETGKNWIEGE	42
FAM22F	FAM22F-19	PRPQRPAETNAHLPPPRPQRP AETKVPEEIPPEVVQEYVDIM EELLGSHPGDTGEPEGQR	55
FAM2	FAM2	PAETKVPEEIPPEVVQEYVDI	56

10

20

30

40

50

【表 6－2】

2 F	2 F－2 0	MEELLGSHPGDTGEPEGQREK GKVEQPQEEDGITS DPGL	
HOXA 2	HOXA 2－4	PGSHPRHGAGGRPKPSPAGSR GSPVPAGALQPPEYPWMKEKK AAKKTALLPAAAAAATAA	6 1
HOXA 2	HOXA 2－5	RGSPVPAGALQPPEYPWMKEK KAAKKTALLPAAAAAATAAAT GPACLSHKESLEIADGSG	6 2
HOXA 2	HOXA 2－6	KKAAKKTALLPAAAAAATAAA TGPACLSHKESLEIADGSGGG SRRLRTAYTNTQLLELEK	6 3
JAZF 1	JAZF 1－1	MTGIAAASF SNTCRFGGCGL HFPTLADLIEHIEDNHIDTDP RVLEKQELQQPTYVALSY	6 8
LIN9	LIN9 －2 5	SRLTAILLQIKCLAEGGDLNS FEFKSLTDSLNDIKSTIDASN ISCFQNNVEIHVAHIQSG	7 4
MYCB P	MYCB P－1	MAHYKAADSKREQFRRYLEKS GVLDTLTKVLVALYEEPEKPN SALDFLKHHLGAATPENP	7 5
MYCL 1	MYCL 1－1	MDYDSYQHYFYDYDCGEDFYR STAPSEDIWKKFELVPSPTS PPWGLGPGAGDPAPGIGP	7 6
NEUR OD1	NEUR OD1－ 6	LRRMKANARERNRMHGLNAAL DNLRKVVPCYSKTQKLSKIET LRLAKNYIWALSEILRSG	7 8
NEUR OG3	NEUR OG3－ 5	RRSRRKKANDRERNRMHNLNS ALDALRGVLPTFPDDAKLTKI ETLRFAHNYIWALTQTLR	7 9
RBPJ	RBPJ －1 3	GMALPRLIIRKVDKQTALLDA DDPVSQLHKCAFYLKDTERMY LCLSQERIIQFQATPCPK	8 5
SS18 L2	SS18 L2－2	QETIQRLLEENDQLIRCI VEY QNKGRGNECVQYQHVLHRNLI YLATIADASPTSTSKAME	9 2
YAF2	YAF2 －6	EKKDKVEKEKSEKETTSKKNS HKKTRPRLKNVDRSSAQHLEV TVGDLTVIITDFKEKTKS	9 5
YAF2	YAF2 －7	SHKKTRPRLKNVDRSSAQHLE VTVGDLTVIITDFKEKTKSP ASSAASADQHSQSGSSSD	9 6
ZXDC	ZXDC	SPAEQHGAQDTELSAGTGNFY	9 7

10

20

30

40

50

【表 6 - 3】

	- 1 2	LES GGSARTDYRAIQ LAKEKK Q R G A G S N A G A S Q S T Q R K I	
Z X D C	Z X D C - 1 3	Y L E S G G S A R T D Y R A I Q L A K E K K Q R G A G S N A G A S Q S T Q R K I K E G K M S P P H F H A S Q N S W L C G	9 8
Z X D C	Z X D C - 1 6	S L V V P S G G R P G P A P A A G V Q C G A Q G V Q V Q L V Q D D P S G E G V L P S A R G P A T F L P F L T V D L P V Y	9 9

10

S P D Y E 4 - C I T E D 1 - R E L A - H S F 1ヌクレオチド配列（配列番号 1 2 0）
S P D Y E 4 - C I T E D 1 - R E L A - H S F 1タンパク質配列（配列番号 1 2 1）
【化 1】

MDPSPQPQSLGLKRKSEWSDESEEEEELELERAPEPEDTWVVETLCGLKMKLKRKRASSGGSGGSESLSPSAGA
QSPAIDS DPVDEEVLM SLVVELGLDRANELPELWLGQNEFDFTADFPSSCGGSINSRSSGSPKKRKVGSQYLPDT
DDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSG
QISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQF
DDEDLGALLGNSTDPVFTDLASVDNSEFQQLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGA
PGLPNGLLSGDEDFSSIADMDFSALLGSGSGSGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRP
PEAENSSPD SGKQLVHYTAQPLFLDPGSDVTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

20

リンカー及び N L S の配列は、下線で示されている
S P D Y E 4 - C I T E D 1 - R E L Aヌクレオチド配列（配列番号 1 2 2）
S P D Y E 4 - C I T E D 1 - R E L Aタンパク質配列（配列番号 1 2 3）
【化 2】

MDPSPQPQSLGLKRKSEWSDESEEEEELELERAPEPEDTWVVETLCGLKMKLKRKRASSGGSGGSESLSPSAGA
QSPAIDS DPVDEEVLM SLVVELGLDRANELPELWLGQNEFDFTADFPSSCGGSINSRSSGSPKKRKVGSQYLPDT
DDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSG
QISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFD
DEDLGALLGNSTDPVFTDLASVDNSEFQQLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAP
GLPNGLLSGDEDFSSIADMDFSALL

40

リンカー及び N L S の配列は、下線で示されている
H S F 1 - R E L A - S P D Y E 4 - C I T E D 1ヌクレオチド配列（配列番号 1 2 4）
H S F 1 - R E L A - S P D Y E 4 - C I T E D 1タンパク質配列（配列番号 1 2 5）

50

【化 3】

MGFSVDTSAALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDSGKQLVHYTAQPLFLDPGSVD
 TGSNDLPVLFELGEGSYFSEG DGFAEDPTISLLTGSEPPKAKDPTVSINSRSSGSPKKRKVGSYLPDTPDRHRIEEK
 RKRTYETFKSIMKKSPFSGPTDPRPPRRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSSGQISQASALA
 PAPPQVLPQAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDDEDLGALL
 GNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLAPGLPNGLLS
 GDEDFSSIADMDFSALLGSGSGSDPSPQPQSLGLKRKSEWSESEEEEELELERAPEPEDTWVETLCGLKMKLK
 RKRASGSGSGSESLSPSAGAQSPAIDS DPVDEEVLM SLVVELGLDRANELPELWLQGNEFDFTADFPSSCGGS

10

リンカー及びNLSの配列は、下線で示されている

SPDYE4-CITED1-p65-miniHSF1ヌクレオチド配列（配列番号 126）

SPDYE4-CITED1-p65-miniHSF1タンパク質配列（配列番号 127）

【化 4】

MDPSPQPQSLGLKRKSEWSESEEEEELELERAPEPEDTWVETLCGLKMKLKRKRASGSGSGSESLSPSAGA
 QSPAIDS DPVDEEVLM SLVVELGLDRANELPELWLQGNEFDFTADFPSSCGGSINSRSSGSPKKRKVGSPTQAGE
 GTLSEALLQLQFDDDEDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQR
 PPDAPAPLAPGLPNGLLSGDEDFSSIADMDFSALLGSGSGSMLSSHGFSVDTSAALLDLFSP

20

リンカー及びNLSの配列は、下線で示されている

miniSPDYE4-CITED1-p65-HSF1ヌクレオチド配列（配列番号 128）

miniSPDYE4-CITED1-p65-HSF1タンパク質配列（配列番号 129）

30

【化 5】

MSDESEEEEELELERAPEPEDTWVETLGSGSGSESLSPSAGAQSPAIDS DPVDEEVLM SLVVELGLDRANELP
 ELWLQGNEFDFTADFPSSCGGSINSRSSGSPKKRKVGSPTQAGEGTLSEALLQLQFDDDEDLGALLGNSTDPAVFT
 DLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLAPGLPNGLLSGDEDFSSIADM
 DFSALLGSGSGSGFVDTSAALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDSGKQLVHYTAQ
 PLFLDPGSVDTGSNDLPVLFELGEGSYFSEG DGFAEDPTISLLTGSEPPKAKDPTVS

リンカー及びNLSの配列は、下線で示されている

SPDYE4-miniCITED1-p65-HSF1ヌクレオチド配列（配列番号 130）

SPDYE4-miniCITED1-p65-HSF1タンパク質配列（配列番号 131）

40

【化 6】

MDPSPQPQSLGLKRKSEWSDESEEEEELELERAPEPEDTWVVETLCGLKMKLKRKRASSGGSGGS
 DSDPVDEEVLMSLVVELGLDRANELGGSINSRSSGSPKKRKVGSPTQAGEGTLSEALLQLQFDDDEDLGALLGNST
 DPAVFTDLASVDNSEFQQLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLSGDED
 FSSIADMDFSALLGSGSGSGFSVDTSAALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDSGKQL
 VHHTAQPLFLDPGSDVTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

リンカー及び N L S の配列は、下線で示されている

S P D Y E 4 - C I T E D 1 - m i n i p 6 5 (C) - H S F 1 ヌクレオチド配列 (配列
 番号 1 3 2)

S P D Y E 4 - C I T E D 1 - m i n i p 6 5 (C) - H S F 1 タンパク質配列 (配列番
 号 1 3 3)

【化 7】

MDPSPQPQSLGLKRKSEWSDESEEEEELELERAPEPEDTWVVETLCGLKMKLKRKRASSGGSGGSESLSPSAGA
 QSPAIIIDSDPVDEEVLMSLVVELGLDRANELPELWLQNEFDFTADFPSSCGGSINSRSSGSPKKRKVGSAAPGLPN
 GLLSGDEDFSSIADMDFSALLGSGSGSGFSVDTSAALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENS
 SPDSGKQLVHHTAQPLFLDPGSDVTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

リンカー及び N L S の配列は、下線で示されている

S P D Y E 4 - C I T E D 1 - m i n i p 6 5 (N) - H S F 1 ヌクレオチド配列 (配列
 番号 1 3 4)

S P D Y E 4 - C I T E D 1 - m i n i p 6 5 (N) - H S F 1 タンパク質配列 (配列番
 号 1 3 5)

【化 8】

MDPSPQPQSLGLKRKSEWSDESEEEEELELERAPEPEDTWVVETLCGLKMKLKRKRASSGGSGGSESLSPSAGA
 QSPAIIIDSDPVDEEVLMSLVVELGLDRANELPELWLQNEFDFTADFPSSCGGSINSRSSGSPKKRKVGSGPTQAGE
 GTLSEALLQLQFDDDEDLGALLGNSTDPVFTDLASVDNSEFQQGSGSGSGFSVDTSAALLDLFSPSVTPDMSLPDL
 SSLASIQELLSPQEPPRPPEAENSSPDSGKQLVHHTAQPLFLDPGSDVTGSNDLPVLFELGEGSYFSEGDGFAEDPTI
 SLLTGSEPPKAKDPTVS

リンカー及び N L S の配列は、下線で示されている

> S - C 3 o r f 6 2 . 2 - P _ A D - H ヌクレオチド配列 (配列番号 1 7 3)

> S - C 3 o r f 6 2 . 2 - P _ A D - H タンパク質配列 (配列番号 1 7 4)

【化 9】

MDPSPQPQSLGLKRKSEWSDESEEEEELELERAPEPEDTWVVETLCGLKMKLKRKRASSGGGSATDADPGSLKQA
 FDDHNIVETVLDLEEDYNVMTGGSINSRSSGSPKKRKVGSPTQAGEGTLSEALLQLQFDDDEDLGALLGNSTDPV
 FTDLASVDNSEFQQLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLSGDEDFSSIA
 DMDDFSALLGSGSGSGFSVDTSAALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDSGKQLVHYT
 AQPLFLDPGSDVTGSNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTVS

リンカー及び N L S の配列は、下線で示されている

> S - C 3 o r f 6 2 . 3 - P _ A D - Hヌクレオチド配列 (配列番号 175)
 > S - C 3 o r f 6 2 . 3 - P _ A D - Hタンパク質配列 (配列番号 176)
 【化10】

MDPSPQPQSLGLKRRKSEWSDESEEEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSGDDHNIVETVLDL
 EEDYNVMTGGINSRSGSPKKRKVGSPTQAGEGTLSEALLQLQFDDDELGALLGNSTDPVFTDLASVDNSEFQ
 QLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGGSG
SGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPDSGKQLVHYTAQPLFLDPGSVD
 TGSNDLPVLFELGEGSYFSEG DGFAEDPTISLLTGSEPPKAKDPTVS

10

リンカー及びNLSの配列は、下線で示されている

> S - C 3 o r f 6 2 _ M T - P _ A D - Hヌクレオチド配列 (配列番号 177)
 > S - C 3 o r f 6 2 _ M T - P _ A D - Hタンパク質配列 (配列番号 178)
 【化11】

MDPSPQPQSLGLKRRKSEWSDESEEEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSGATDADPGSLKQA
 FDDHNIVETVLDLEEDYNVMTSFYQIEGSSGGINSRSGSPKKRKVGSPTQAGEGTLSEALLQLQFDDDELGAL
 LGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLS
 GDEDFSSIADMDFSALLGGSGSGSGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQELLSPQEPPRPPEAENSSPD
 SGKQLVHYTAQPLFLDPGSVDTGNDLPVLFELGEGSYFSEG DGFAEDPTISLLTGSEPPKAKDPTVS

20

リンカー及びNLSの配列は、下線で示されている

> S - D D I T 3 _ M T - P _ A D - Hヌクレオチド配列 (配列番号 179)
 > S - D D I T 3 _ M T - P _ A D - Hタンパク質配列 (配列番号 180)
 【化12】

MDPSPQPQSLGLKRRKSEWSDESEEEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSGSGSFGTLSSWE
 LEAWYEDLQEVLSSENGGTYVSPPGNEEEESKIFTTLDPASLAWLTGGINSRSGSPKKRKVGSPTQAGEGTL
 EALLQLQFDDDELGALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDP
 APAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGGSGSGSGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQELLS
 PQEPPRPPEAENSSPDSGKQLVHYTAQPLFLDPGSVDTGNDLPVLFELGEGSYFSEG DGFAEDPTISLLTGSEPPK
 AKDPTVS

30

リンカー及びNLSの配列は、下線で示されている

> S - C - P _ A D - H S F 1 _ M Tヌクレオチド配列 (配列番号 181)
 > S - C - P _ A D - H S F 1 _ M Tタンパク質配列 (配列番号 182)
 【化13】

MDPSPQPQSLGLKRRKSEWSDESEEEEELELERAPEPEDTWVETLCGLKMKLKRKRASSGGSGSGESLSPSAGA
 QSPAIDSDPVDEEVLMSLVVELGLDRANELPELWLQNEFDFTADFPSSCGGINSRSGSPKKRKVGSPTQAGE
 GTLSEALLQLQFDDDELGALLGNSTDPVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQR
 PPDAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALLGGSGSGSGFSVDTSALLDLFSPSVTPDMSLPDLSSLASIQ
 ELLSPQEPPR

40

リンカー及びNLSの配列は、下線で示されている

50

> 3 x ZNF473_KRABヌクレオチド配列 (配列番号183)

> 3 x ZNF473_KRABタンパク質配列 (配列番号184)

MAEEFVTLKDVGMDFTLGDWEQLGLEQGDTFWDTALDNCQ
DLFLLDPPRPNLTSHPDGSEDLPLAGGSPEATSMAEEFV
TLKDVGMDFTLGDWEQLGLEQGDTFWDTALDNCQDLFLLD
PPRPNLTSHPDGSEDLPLAGGSPEATSMAEEFVTLKDVGM
MDFTLGDWEQLGLEQGDTFWDTALDNCQDLFLLDPPRPNL
TSHPDGSEDLPLAGGSPEATS

参考文献：

- Adusumilli, R., and Mallick, P. (2017). Data Conversion with ProteoWizard msConvert. *Methods Mol. Biol. Clifton NJ* 1550, 339–368. 10
- Alekseyenko, A. A., Gorchakov, A. A., Kharchenko, P. V., and Kuroda, M. I. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc. Natl. Acad. Sci.* 111, 2488–2493.
- Ali, R. H., and Rouzbahman, M. (2015). Endometrial stromal tumours revisited: an update based on the 2014 WHO classification. *J. Clin. Pathol.* 68, 325–332. 20
- Alonso, C. N., Meyer, C., Gallego, M. S., Rossi, J. G., Mansini, A. P., Rubio, P. L., Medina, A., Marschalek, R., and Felice, M. S. (2010). BTBD18: A novel MLL partner gene in an infant with acute lymphoblastic leukemia and inv(11)(q13;q23). *Leuk. Res.* 34, e294–e296.
- Antonescu, C., Sung, Y.-S., Zhang, L., Agaram, N., and Fletcher, C. (2017). Recurrent SRF-RELA Fusions Define a Novel Subset of Cellular Myofibroma/Myopericytoma: A Potential Diagnostic Pitfall With Sarcomas With Myogenic Differentiation. *Am. J. Surg. Pathol.* 41, 677–684. 30
- Arnold, C. D., Nemcko, F., Woodfin, A. R., Wienerroither, S., Vlasova, A., Schleiffer, A., Pagan, M., Rath, M., and Stark, A. (2018). A high-throughput method to identify trans-activation domains within transcription factor sequences. *EMBO J.* 37. 40
- Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* 324, 1720–1723.
- Barreto, G., Schaefer, A., Marhold, J., Stach, D., Swaminathan, S. K., Handa, V., Doederlein, G., Maltry, N., Wu, W., Lyko, F., et al. (2007). 50

- Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* 445, 671–675.
- Barrett, J., Birrer, M. J., Kato, G. J., Dosaka-Akita, H., and Dang, C. V. (1992). Activation domains of L-Myc and c-Myc determine their transforming potencies in rat embryo cells. *Mol. Cell. Biol.* 12, 3130–3137.
- Basnet, H., Su, X. B., Tan, Y., Meisenhelder, J., Merkurjev, D., Ohgi, K. A., Hunter, T., Pillus, L., and Rosenfeld, M. G. (2014). Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation. *Nature* 516, 267–271. 10
- Basu, S., Mackowiak, S. D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., Geertsema, H., Ali, S., Jerkovic, I., Ewers, H., et al. (2020). Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* 181, 1062–1079. e30. 20
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 491–507.
- Bousoik, E., and Montazeri Aliabadi, H. (2018). “Do We Know Jack” About JAK? A Closer Look at JAK/STAT Signaling Pathway. *Front. Oncol.* 8.
- Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. *Sci. N. Y. NY* 328, 1043–1046. 30
- Brien, G. L., Remillard, D., Shi, J., Hemming, M. L., Chabon, J., Wynne, K., Dillon, E. T., Cagney, G., Van Mierlo, G., Baltissen, M. P., et al. (2018). Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma. *ELife* 7, e41305. 40
- Buchwalter, G., Gross, C., and Wasylyk, B. (2004). Ets ternary complex transcription factors. *Gene* 324, 1–14.
- Cao, L., Yonis, A., Vaghela, M., Barriga, E. H., Chugh, P., Smith, M. B., Maufront, J., Lavoie, G., Meant, A., Ferber, E., et al. (2020). SPIN90 associates with mDia1 and the Arp2/3 complex to regulate cortical actin organization. *Nat. Cell Biol.* 22, 803–814.
- Centore, R. C., Sandoval, G. J., Soares, L. M. M. 50

- , Kadoch, C., and Chan, H. M. (2020). Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends Genet.* 36, 936–950.
- Chauhan, S., Zheng, X., Tan, Y. Y., Tay, B. –H., Lim, S., Venkatesh, B., and Kaldis, P. (2012). Evolution of the Cdk-activator Speedy/RINGO in vertebrates. *Cell. Mol. Life Sci.* 69, 3835–3850.
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., Iyer, E. P. R., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326–328. 10
- Chavez, A., Tuttle, M., Pruitt, B. W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S. J., Cecchi, R. J., Kowal, E. J. K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. *Nat. Methods* 13, 563–567. 20
- Chen, F. X., Smith, E. R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 19, 464–478.
- Conaway, R. C., and Conaway, J. W. (2011). Function and regulation of the Mediator complex. *Curr. Opin. Genet. Dev.* 21, 225–230.
- Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* 33, 960–982. 30
- Cramer, P. (2019). Organization and regulation of gene transcription. *Nature* 573, 45–54.
- Davis, R. L., Weintraub, H., and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987–1000.
- Di Rocco, G., Mavilio, F., and Zappavigna, V. (1997). Functional dissection of a transcriptionally active, target-specific Hox-Pbx complex. *EMBO J.* 16, 3644–3654. 40
- Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208.
- Eng, J. K., Jahan, T. A., and Hoopmann, M. R. (2013). Comet: an open-source MS/MS sequence database search tool. *PROTEOMICS* 13, 22–24. 50

- Erijman, A., Kozlowski, L., Sohrabi-Jahromi, S., Fishburn, J., Warfield, L., Schreiber, J., Noble, W. S., Soeding, J., and Hahn, S. (2020). A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. *Mol. Cell* 78, 890–902. e6.
- Esnault, C., Stewart, A., Gualdrini, F., East, P., Horswell, S., Matthews, N., and Treisman, R. (2014). Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. *Genes Dev.* 28, 943–958. 10
- Fujisawa, T., and Filippakopoulos, P. (2017). Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.* 18, 246–262.
- Gahan, J. M., Rentzsch, F., and Schnitzler, C. E. (2020). The genetic basis for PRC1 complex diversity emerged early in animal evolution. *Proc. Natl. Acad. Sci.* 117, 22880–22889. 20
- Gao, Y., Xiong, X., Wong, S., Charles, E. J., Lim, W. A., and Qi, L. S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat. Methods* 13, 1043–1049.
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* 45, 344–356. 30
- Gao, Z., Lee, P., Stafford, J. M., von Schimmelmann, M., Schaefer, A., and Reinberg, D. (2014). An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature* 516, 349–354.
- Gastwirt, R. F., McAndrew, C. W., and Donoghue, D. J. (2007). Speedy/RINGO Regulation of CDKs in Cell Cycle, Checkpoint Activation and Apoptosis. *Cell Cycle* 6, 1188–1193. 40
- Gerritsen, M. E., Williams, A. J., Neish, A. S., Moore, S., Shi, Y., and Collins, T. (1997). CREB-binding protein/p300 are transcriptional coactivators of p65. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2927–2932.
- Gonzalez, L., and Nebreda, A. R. (2020). RINGO/Speedy proteins, a family of non-canonical activators of CDK1 and CDK2. *Semin. C* 50

ell Dev. Biol. 107, 21–27.

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

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

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

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

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

Horb, M. E., Shen, C.-N., Tosh, D., and Slack, J. M. W. (2003). Experimental Conversion of Liver to Pancreas. Curr. Biol. 13, 105–115. Hrzenjak, A. (2016). JAZF1/SUZ12 gene fusion in endometrial stromal sarcomas. Orphanet J. Rare Dis. 11, 15.

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

Huttlin, E. L., Bruckner, R. J., Navarrete-Perrea, J., Cannon, J. R., Baltier, K., Gebreab, F., Gygi, M. P., Thornock, A., Zarraga, G., Tam, S., et al. (2020). Dual Proteome-scale Networks Reveal Cell-specific Remodeling of the Human Interactome (Systems Biology).

Israni, D. V., Li, H.-S., Gagnon, K. A., Sander, J. D., Roybal, K. T., Joung, J. K., Wong, W. W., and

10

20

30

40

50

- d Khalil, A. S. (2021). Clinically-driven design of synthetic gene regulatory programs in human cells (Synthetic Biology).
- Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339.
- Karanian, M., Pissaloux, D., Gomez-Brouchet, A., Chevenet, C., Le Loarer, F., Fernandez, C., Minard, V., Corradini, N., Castex, M.-P., Duc-Gallet, A., et al. (2020). SRF-FOXO1 and SRF-NCOA1 Fusion Genes Delineate a Distinctive Subset of Well-differentiated Rhabdomyosarcoma. *Am. J. Surg. Pathol.* 44, 607–616. 10
- Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J., and Khalil, A. S. (2014). Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* 158, 110–120. 20
- Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., and Roux, K. J. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* 27, 1188–1196.
- Knight, J. D. R., Choi, H., Gupta, G. D., Pelletier, L., Raught, B., Nesvizhskii, A. I., and Gingras, A.-C. (2017). ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nat. Methods* 14, 645–646. 30
- Kundu, T. K., Palhan, V. B., Wang, Z., An, W., Cole, P. A., and Roeder, R. G. (2000). Activator-dependent transcription from chromatin in vitro involving targeted histone acetylation by p300. *Mol. Cell* 6, 551–561.
- Lambert, J.-P., Tucholska, M., Go, C., Knight, J. D. R., and Gingras, A.-C. (2015). Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J. Proteomics* 118, 81–94. 40
- Lambert, J.-P., Picaud, S., Fujisawa, T., Hou, H., Savitsky, P., Uuskuela-Reimand, L., Gupta, G. D., Abdouni, H., Lin, Z.-Y., Tucholska, M., et al. (2019). Interactome Rewiring Following Pharmacological Targeting of BET Bromodomains. *Mol. Cell* 73, 621–638. e17.
- Lambert, S. A., Jolma, A., Campitelli, L. F., Da 50

- s, P. K. , Yin, Y. , Albu, M. , Chen, X. , Taipale, J. , Hughes, T. R. , and Weirauch, M. T. (2018). The Human Transcription Factors. *Cell* 172, 650–665.
- Lecoq, L. , Raiola, L. , Chabot, P. R. , Cyr, N. , Arseneault, G. , Legault, P. , and Omichinski, J. G. (2017). Structural characterization of interactions between transactivation domain 1 of the p65 subunit of NF- κ B and transcription regulatory factors. *Nucleic Acids Res.* 45, 5564–5576. 10
- Li, X. , Wang, W. , Wang, J. , Malovannaya, A. , Xi, Y. , Li, W. , Guerra, R. , Hawke, D. H. , Qin, J. , and Chen, J. (2015). Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol. Syst. Biol.* 11.
- Liang, F. -S. , Ho, W. Q. , and Crabtree, G. R. (2011). Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci. Signal.* 4, rs2–rs2. 20
- Liu, G. , Knight, J. D. R. , Zhang, J. P. , Tsou, C. -C. , Wang, J. , Lambert, J. -P. , Larsen, B. , Tyers, M. , Raught, B. , Bandeira, N. , et al. (2016). Data Independent Acquisition analysis in ProHits 4.0. *J. Proteomics* 149, 64–68.
- Loven, J. , Hoke, H. A. , Lin, C. Y. , Lau, A. , Orlando, D. A. , Vakoc, C. R. , Bradner, J. E. , Lee, T. I. , and Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334. 30
- Luck, K. , Kim, D. -K. , Lambourne, L. , Spirohn, K. , Begg, B. E. , Bian, W. , Brignall, R. , Cafarelli, T. , Campos-Laborie, F. J. , Charlotiaux, B. , et al. (2020). A reference map of the human binary protein interactome. *Nature* 580, 402–408.
- Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, Eichinger M, Jiang T, Foley JE, Winfrey RJ, Townsend JA, Unger-Wallace E, Sander JD, Mueller-Lerch F, Fu F, Pearlberg J, Goebel C, Dassie JP, Pruett-Miller SM, Porteus MH, Sgroi DC, Iafrate AJ, Dobbbs D, McCray PB Jr, Cathomen T, Voytas DF, Joung JK. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell.* 2008 Jul 25;31(2):294–301. 40
- Marcon, E. , Ni, Z. , Pu, S. , Turinsky, A. L. , Trimble, S. S. , Olsen, J. B. , Silverman-Gavrila, R. 50

- , Silverman-Gavrila, L., Phanse, S., Guo, H., et al. (2014). Human-Chromatin-Related Protein Interactions Identify a Demethylase Complex Required for Chromosome Segregation. *Cell Rep.* 8, 297–310.
- Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., Zullo, H. J., McKenzie, Z. M., Kubiak, R. L., St Pierre, R., et al. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* 175, 1272–1288. e20. 10
- McGrath, D. A., Fifield, B. -A., Marceau, A. H., Tripathi, S., Porter, L. A., and Rubin, S. M. (2017). Structural basis of divergent cyclin-dependent kinase activation by Spyl/RINGO proteins. *EMBO J.* 36, 2251–2262.
- Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. -P., St-Denis, N. A., Li, T., Miteva, Y. V., Hauri, S., Sardi, M. E., Low, T. Y., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* 10, 730–736. 20
- Miralles, F., Posern, G., Zaromytidou, A. -I., and Treisman, R. (2003). Actin Dynamics Control SRF Activity by Regulation of Its Coactivator MAL. *Cell* 113, 329–342.
- Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A., and DiCroce, L. (2012). Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. *Cell Stem Cell* 10, 47–62. 30
- Morita, K., Celso, C. L., Spencer-Dene, B., Zouboulis, C. C., and Watt, F. M. (2006). HAN11 binds mDia1 and controls GLI1 transcriptional activity. *J. Dermatol. Sci.* 44, 11–20.
- Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J. J., Herzog, V. A., Reichholf, B., et al. (2018). SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* 360, 800–805. 40
- 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.
- Narayan, S., Bryant, G., Shah, S., Berrozpe, G., and Ptashne, M. (2017). OCT4 and SOX2 Work as Transcriptional Activators in Repro 50

gramming Human Fibroblasts. *Cell Rep.* 20, 1585–1596.

Nasrin, N., Ogg, S., Cahill, C. M., Biggs, W., Nui, S., Dore, J., Calvo, D., Shi, Y., Ruvkun, G., and Alexander-Bridges, M. C. (2000). DAF-16 recruits the CREB-binding protein coactivator complex to the insulin-like growth factor binding protein 1 promoter in Hep G2 cells. *Proc. Natl. Acad. Sci.* 97, 10412–10417.

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

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

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

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

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

Piper, D. E., Batchelor, A. H., Chang, C.-P., Cleary, M. L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 Heterodimer Bound to DNA: Role of the Hexapeptide and a Fourth Homeodomain Helix in Complex Formation. *Cell* 96, 587–597.

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

Piunti, A., Smith, E. R., Morgan, M. A. J., Ugarengo, M., Khaltyan, N., Helmin, K. A., Ryan, C. A., Murray, D. C., Rickels, R. A., Yilmaz, B. D., et al. (2019). CATACOMB: An endogenous induci

10

20

30

40

50

ble gene that antagonizes H3K27 methylation activity of Polycomb repressive complex 2 via an H3K27M-like mechanism. *Sci. Adv.* 5, eaax2887.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. *Nature* 386, 569–577.

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

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

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

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

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

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

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

10

20

30

40

50

- Sdelci, S., Lardeau, C. -H., Tallant, C., Klepsch, F., Klaiber, B., Bennett, J., Rathert, P., Schuster, M., Penz, T., Fedorov, O., et al. (2016). Mapping the chemical chromatin reactivation landscape identifies BRD4-TAF1 cross-talk. *Nat. Chem. Biol.* 12, 504-510.
- Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R. L., Aebersold, R., and Nesvizhskii, A. I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Span Classnoca sespan Cell. Proteomics MCP* 10, M111.007690. 10
- Sigler, P. B. (1988). Transcriptional activation. Acid blobs and negative noodles. *Nature* 333, 210-212.
- Singh, R. N., Howell, M. D., Ottesen, E. W., and Singh, N. N. (2017). Diverse role of survival motor neuron protein. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1860, 299-315. 20
- Skapek, S. X., Ferrari, A., Gupta, A. A., Lupo, P. J., Butler, E., Shipley, J., Barr, F. G., and Hawkins, D. S. (2019). Rhabdomyosarcoma. *Nat. Rev. Dis. Primer* 5, 1-19.
- Staller, M. V., Ramirez, E., Holehouse, A. S., Pappu, R. V., and Cohen, B. A. (2021). Design principles of acidic transcriptional activation domains. *BioRxiv* 2020.10.28.359026. 30
- Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* 528, 147-151.
- Strasswimmer, J., Lorson, C. L., Breiding, D. E., Chen, J. J., Le, T., Burghes, A. H. M., and Androphy, E. J. (1999). Identification of Survival Motor Neuron as a Transcriptional Activator-Binding Protein. *Hum. Mol. Genet.* 8, 1219-1226. 40
- Sudarshan, D., Avvakumov, N., Lalonde, M. -E., Alerasool, N., Jacquet, K., Mameri, A., Rousseau, J., Lambert, J. -P., Paquet, E., Setty, S. T., et al. (2021). Recurrent chromosomal translocations in sarcomas create a mega-complex that mislocalizes NuA4/TIP60 to Polycomb target loci. *BioRxiv* 2021.03.26.436670. 50

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

Taipale, M. , Tucker, G. , Peng, J. , Krykbaeva, I . , Lin, Z. –Y. , Larsen, B. , Choi, H. , Berger, B. , Gingras, A. –C. , and Lindquist, S. (2014) . A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell* 158, 434–448. 10

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

Teo, G. , Liu, G. , Zhang, J. , Nesvizhskii, A. I. , Gingras, A. –C. , and Choi, H. (2014) . SAINTexpress: Improvements and additional features in Significance Analysis of INTERactome software. *J. Proteomics* 100, 37–43. 20

The ENCODE Project Consortium, Moore, J. E. , Purcaro, M. J. , Pratt, H. E. , Epstein, C. B. , Shores, N. , Adrian, J. , Kawli, T. , Davis, C. A. , Dobin, A. , et al. (2020) . Expanded encyclopedias of DNA elements in the human and mouse genomes. *Nature* 583, 699–710.

Theodorou, E. , Dalember, G. , Heffelfinger, C . , White, E. , Weissman, S. , Corcoran, L. , and Snyder, M. (2009) . A high throughput embryonic stem cell screen identifies Oct–2 as a bifunctional regulator of neuronal differentiation. *Genes Dev.* 23, 575–588. 30

Tycko, J. , DelRosso, N. , Hess, G. T. , Aradhana, Banerjee, A. , Mukund, A. , Van, M. V. , Ego, B. K. , Yao, D. , Spees, K. , et al. (2020) . High–Throughput Discovery and Characterization of Human Transcriptional Effectors. *Cell* 0.

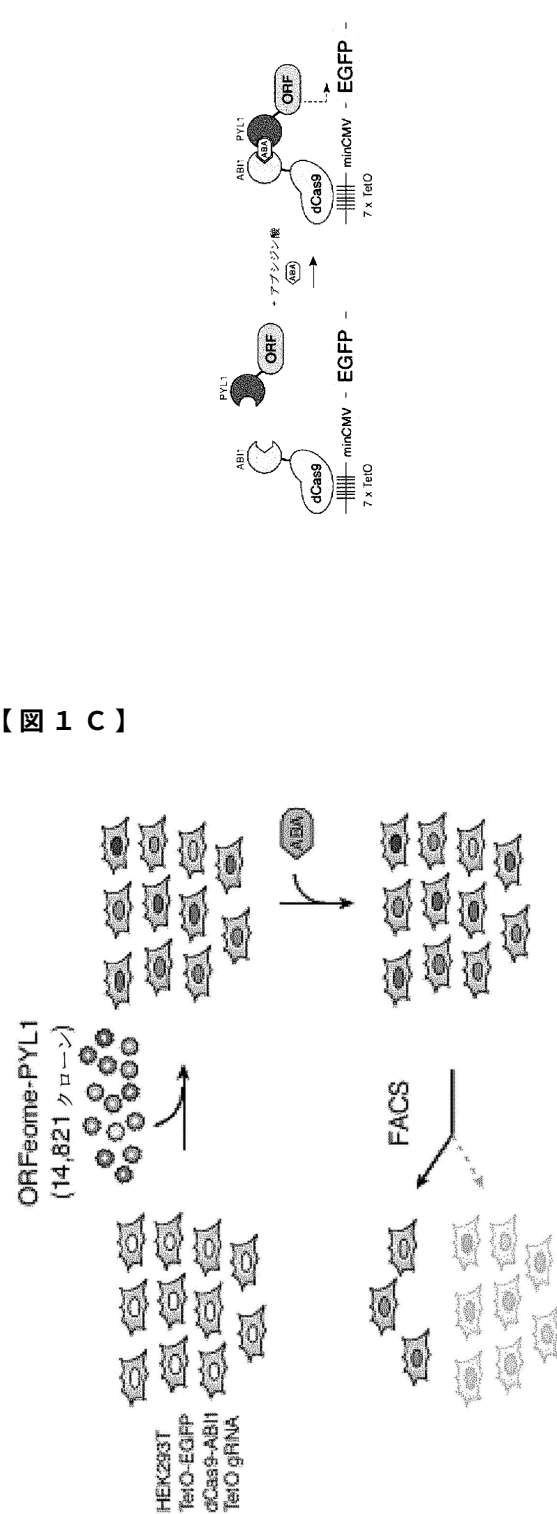
Vannam, R. , Sayilgan, J. , Ojeda, S. , Karakyriakou, B. , Hu, E. , Kreuzer, J. , Morris, R. , Herrera Lopez, X. I. , Rai, S. , Haas, W. , et al. (2021) . Targeted degradation of the enhancer lysine acetyltransferases CBP and p300. *Cell Chem. Biol.* 28, 503–514. e12. 40

Vartiainen, M. K. , Guettler, S. , Larijani, B. , and Treisman, R. (2007) . Nuclear Actin Regulates Dynamic Subcellular Localization and Activity of the SRF Cofactor MAL. *Science* 316, 1749–1752. 50

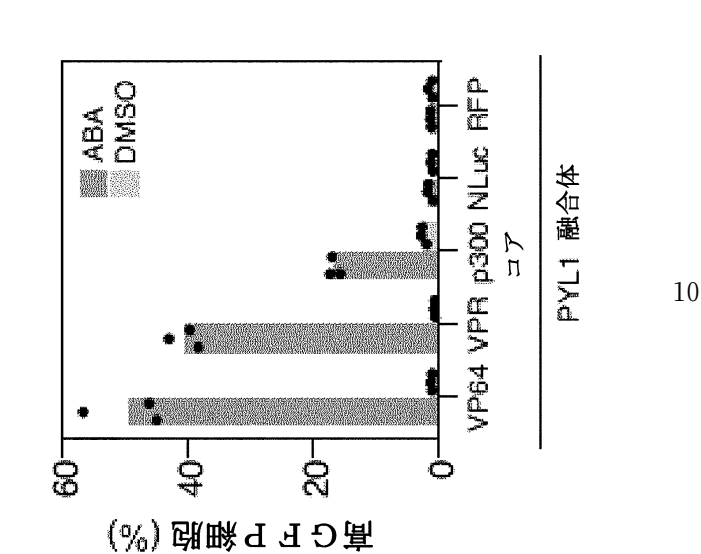
- Vihervaara, A., Duarte, F. M., and Lis, J. T. (2018). Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.* 19, 385–397.
- Vincenz, C., and Kerppola, T. K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc. Natl. Acad. Sci.* 105, 16572–16577. 10
- Wang, R., Ilangovan, U., Robinson, A. K., Schirf, V., Schwarz, P. M., Lafer, E. M., Demeler, B., Hinck, A. P., and Kim, C. A. (2008). Structural transitions of the RING1B C-terminal region upon binding the polycomb cbox domain. *Biochemistry* 47, 8007–8015.
- Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V., Ilangovan, U., Robinson, A. K., Schirf, V., Hart, P. J., Lafer, E. M., Demeler, B., et al. (2010). Polycomb Group Targeting through Different Binding Partners of RING1B C-Terminal Domain. *Structure* 18, 966–975. 20
- Wang, Y., Li, Y., Zeng, W., Zhu, C., Xiao, J., Yuan, W., Wang, Y., Cai, Z., Zhou, J., Liu, M., et al. (2004). IXL, a new subunit of the mammalian Mediator complex, functions as a transcriptional suppressor. *Biochem. Biophys. Res. Commun.* 325, 1330–1338.
- Wang, Y., Chen, J., Hu, J.-L., Wei, X.-X., Qin, D., Gao, J., Zhang, L., Jiang, J., Li, J.-S., Liu, J., et al. (2011). Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep.* 12, 373–378. 30
- Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, H. S., Lambert, S. A., Mann, I., Cook, K., et al. (2014). Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* 158, 1431–1443.
- Winters, A. C., and Bernt, K. M. (2017). MLL-Rearranged Leukemias—An Update on Science and Clinical Approaches. *Front. Pediatr.* 5. 40
- Yahata, T., Shao, W., Endoh, H., Hur, J., Coser, K. R., Sun, H., Ueda, Y., Kato, S., Isselbacher, K. J., Brown, M., et al. (2001). Selective coactivation of estrogen-dependent transcription by CITED1 CBP/p300-binding protein. *Genes Dev.* 15, 2598–2612.
- Yan, J., Enge, M., Whittington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja 50

- , T., Taipale, M., et al. (2013). Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* 154, 801–813.
- Yang, F., DeBeaumont, R., Zhou, S., and Naeaeer, A. M. (2004). The activator–recruited cofactor/Mediator coactivator subunit ARC92 is a functionally important target of the VP16 transcriptional activator. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2339–2344. 10
- Yang, X., Boehm, J. S., Yang, X., Salehi–Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S. R., Alkan, O., Bhimdi, T., et al. (2011). A public genome–scale lentiviral expression library of human ORFs. *Nat. Methods* 8, 659–661.
- Yu, D., Cattoglio, C., Xue, Y., and Zhou, Q. (2019). A complex between DYRK1A and DCAF7 phosphorylates the C-terminal domain of RNA polymerase II to promote myogenesis. *Nucleic Acids Res.* 47, 4462–4475. 20
- Zhou, L., Canagarajah, B., Zhao, Y., Baibakov, B., Tokuhira, K., Maric, D., and Dean, J. (2017). BTBD18 Regulates a Subset of piRNA-Generating Loci through Transcription Elongation in Mice. *Dev. Cell* 40, 453–466. e5.
24. Kiani S et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nature Methods* 12, 1051 (2015).
25. Beerli R. R. et al Toward controlling gene expression at will: Specific regulation of the *erbB-2*/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks *PNAS* 95, 14628 (1995). 30
26. Cong, L., Zhou, R., Kuo, Yc. et al. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun* 3, 968 (2012).
27. Gilbert LA, Larson MH, Morsut L, Liu Z, Barrar GA, Torres SE, Stern–Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS *Cell*. 2013 Jul 9. pii: S0092–8674 (13) 00826–X. 40
28. Sanson, K. R., Hanna, R. E., Hegde, M. et al. Optimized libraries for CRISPR–Cas9 genetic screens with multiple modalities. *Nat Commun* 9, 5416 (2018).

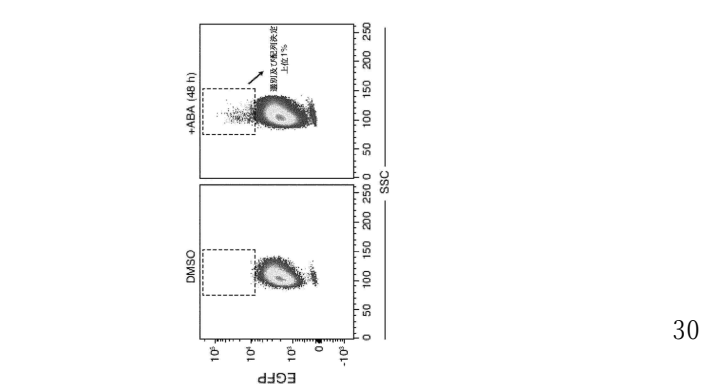
【図面】
【図 1 A】



【図 1 B】



【図 1 D】



10

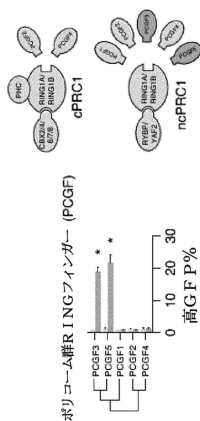
20

30

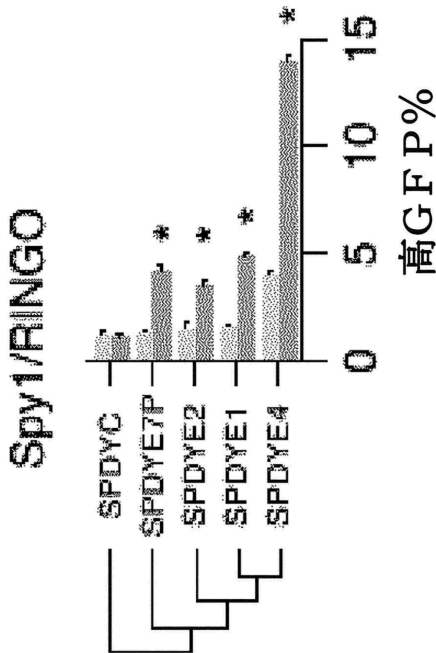
40

50

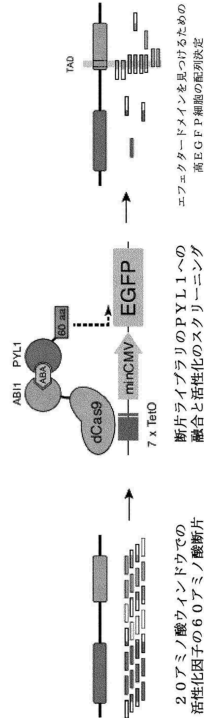
【図 2 E】



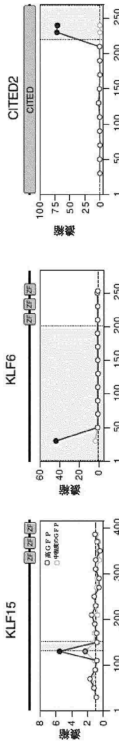
【図 2 F】



【図 3 A】



【図 3 B】



10

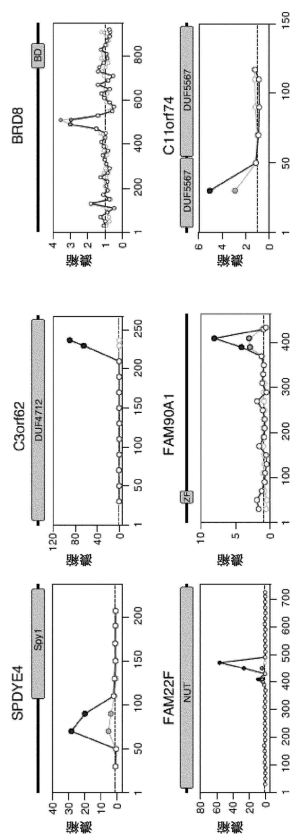
20

30

40

50

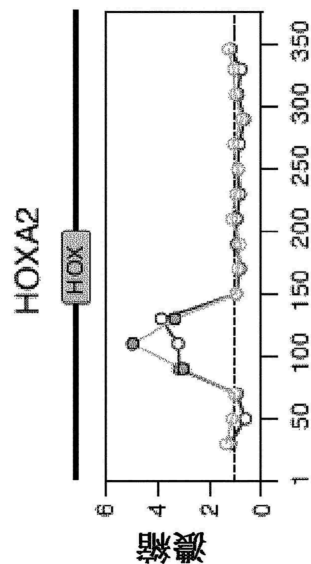
【図 3 C】



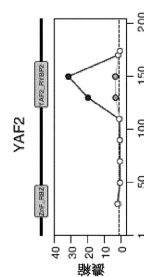
【図 3 E】



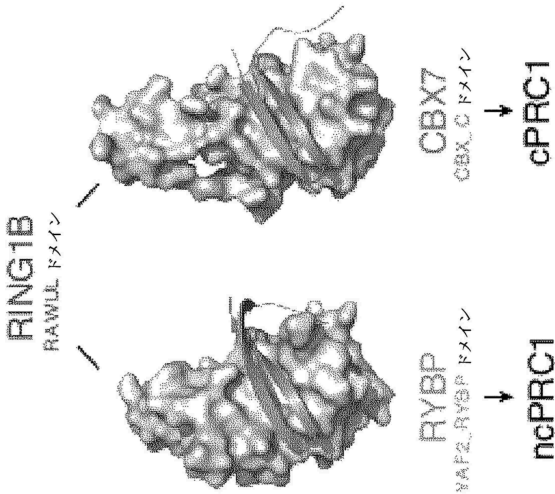
【図 3 D】



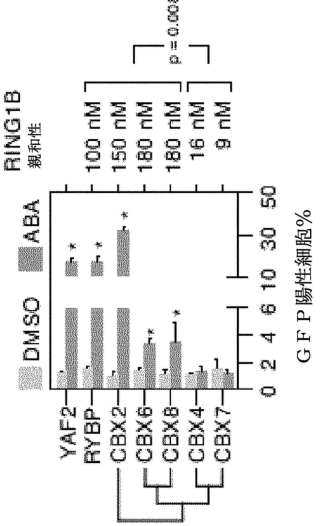
【図 3 F】



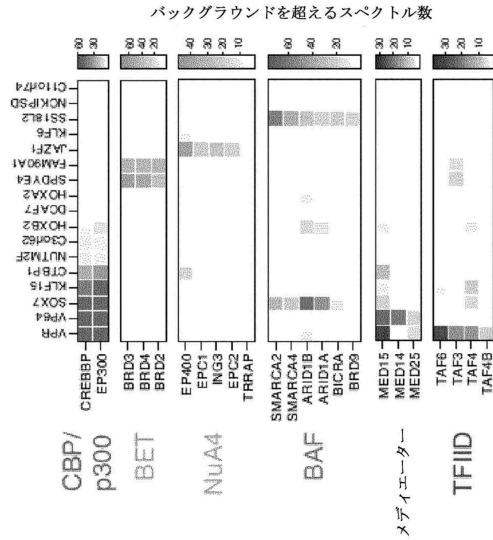
【図 3 G】



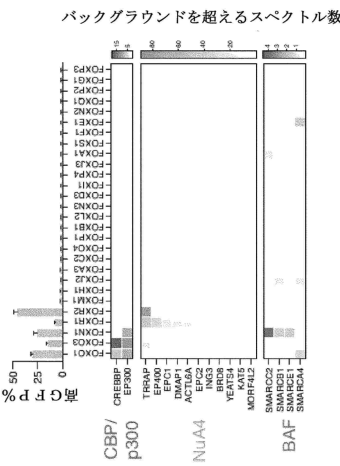
【図 3 H】



【図 4 A】



【図 4 B】



10

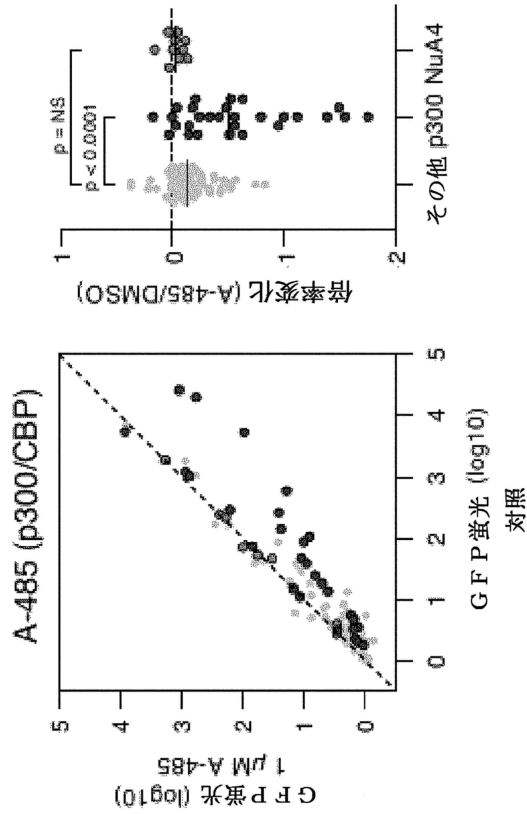
20

30

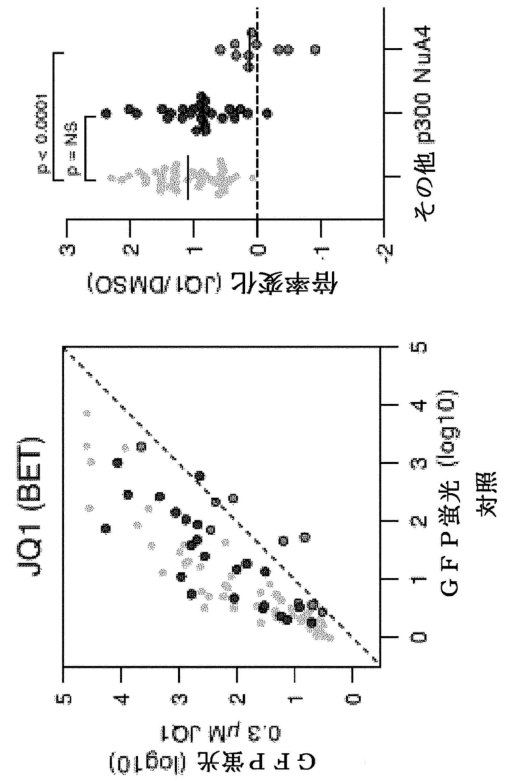
40

50

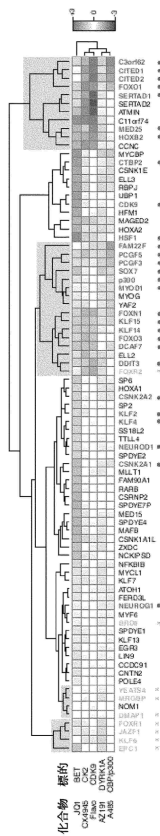
【図 4 C】



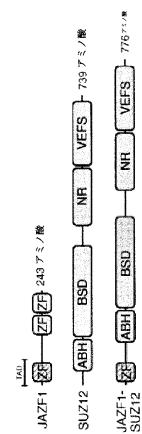
【図 4 D】



【図 4 E】



【図 5 A】



10

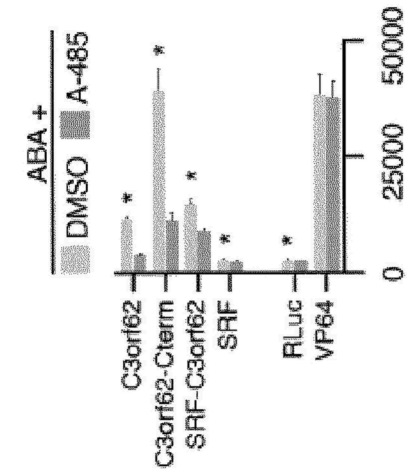
20

30

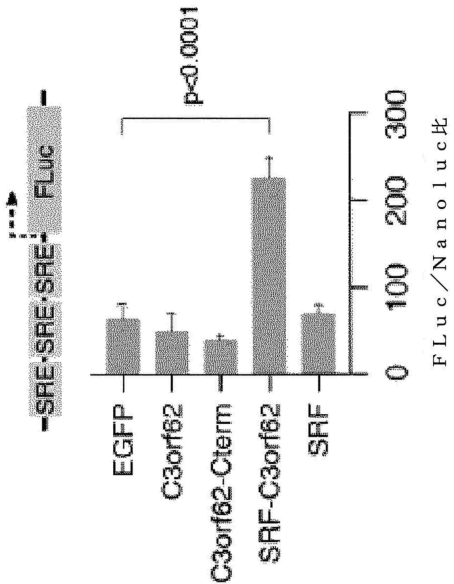
40

50

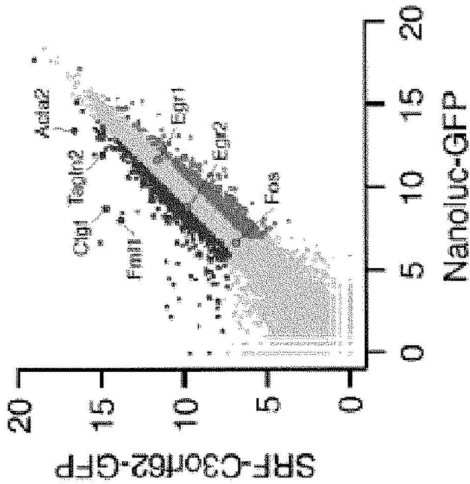
【図 5 F】



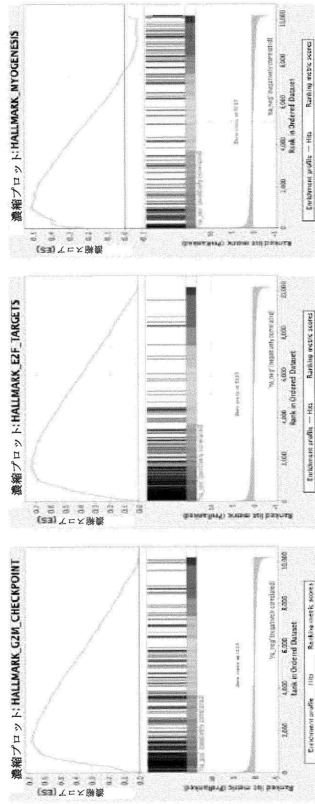
【図 5 G】



【図 5 H】



【図 5 I】



10

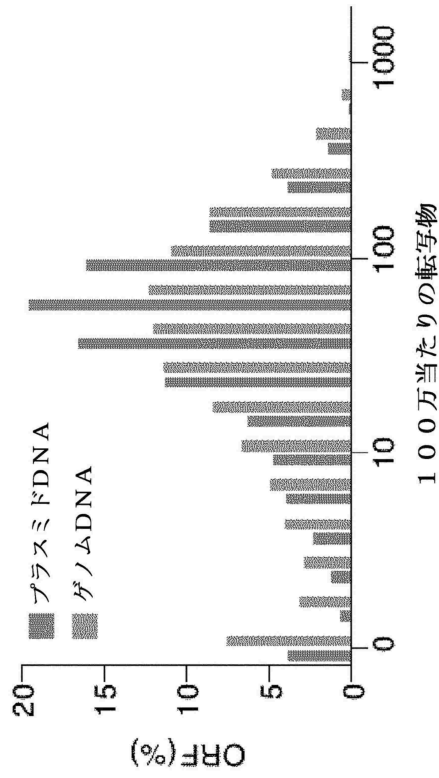
20

30

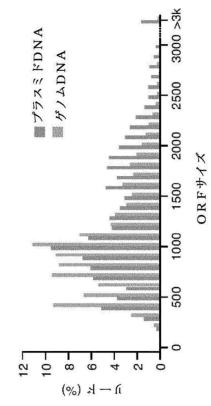
40

50

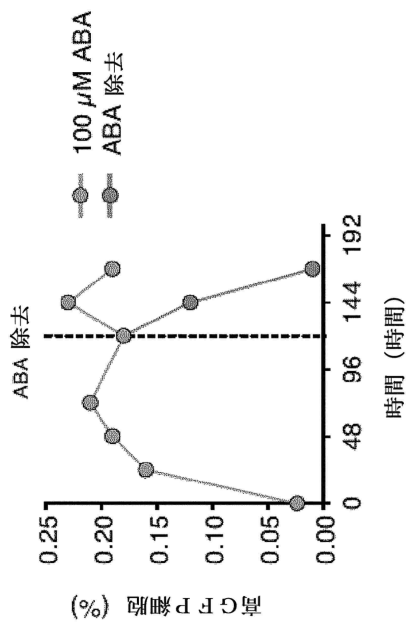
【図 6 A】



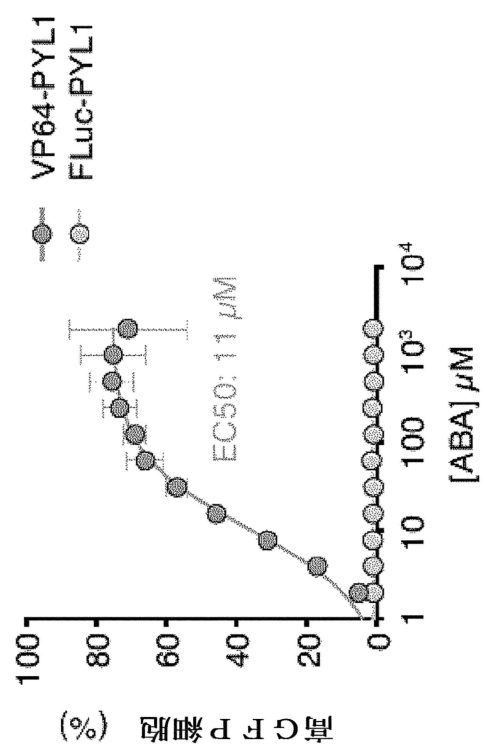
【図 6 B】



【図 6 C】



【図 6 D】



10

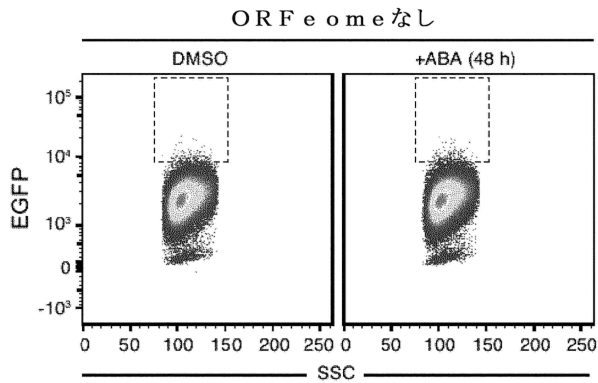
20

30

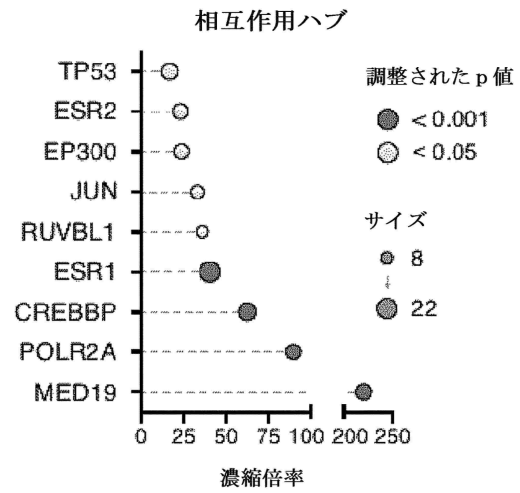
40

50

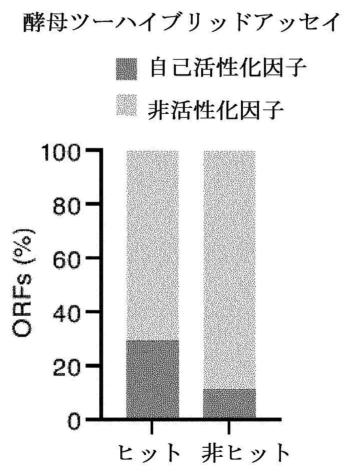
【図 6 E】



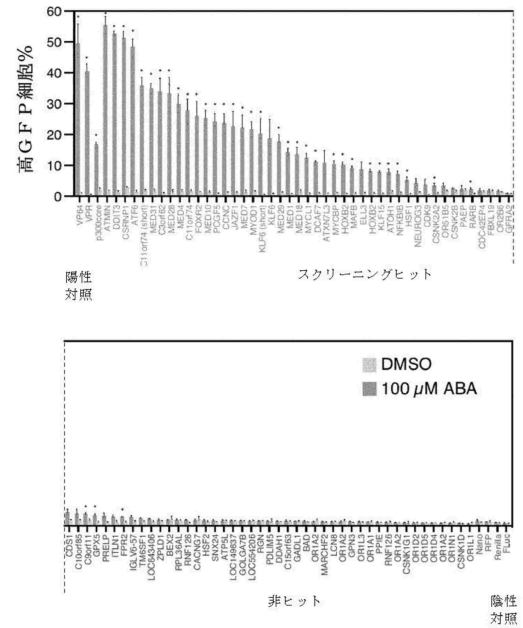
【図 6 F】



【図 6 G】



【図 6 H】



10

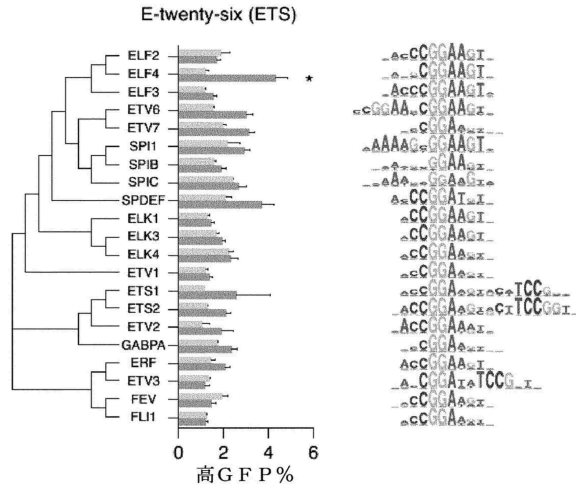
20

30

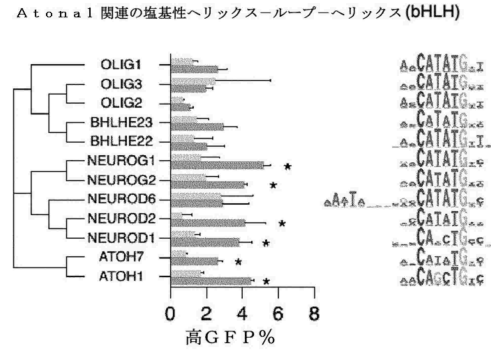
40

50

【図 7 A】

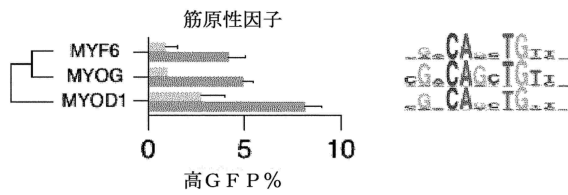


【図 7 B】

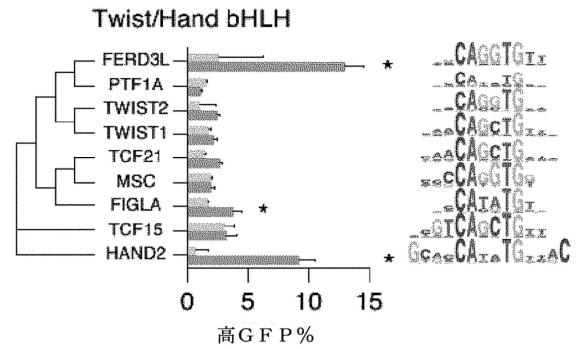


10

【図 7 C】



【図 7 D】



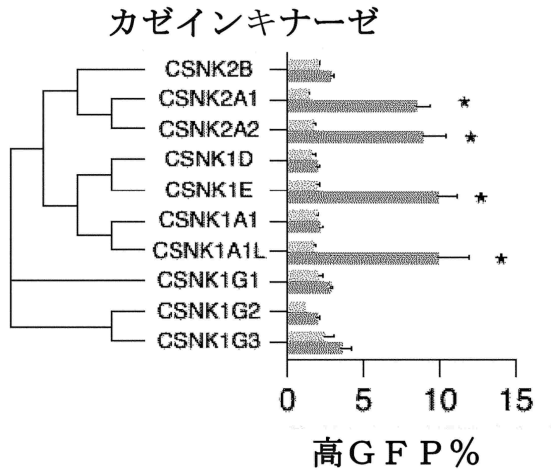
20

30

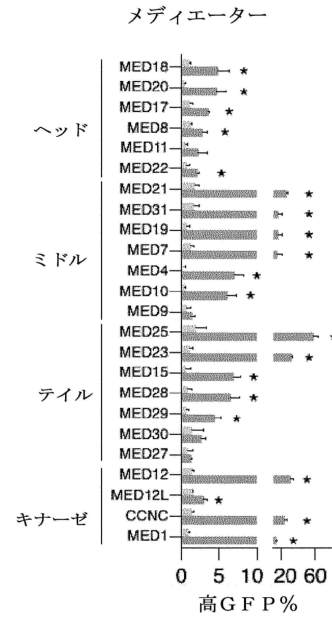
40

50

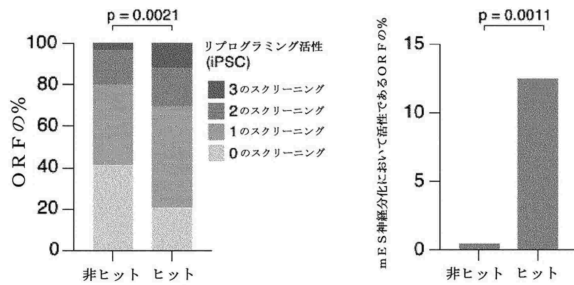
【図 7 E】



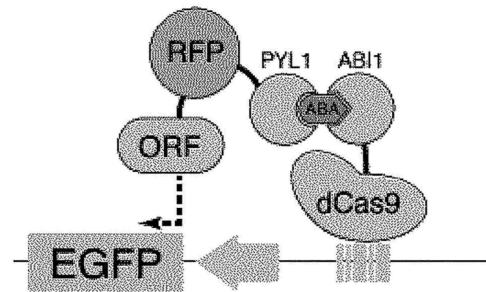
【図 7 F】



【図 7 G】



【図 8 A】



10

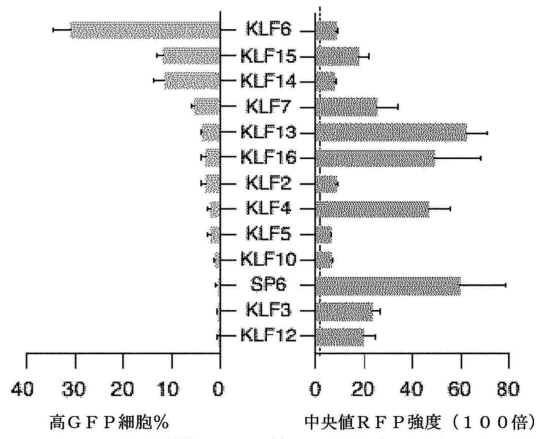
20

30

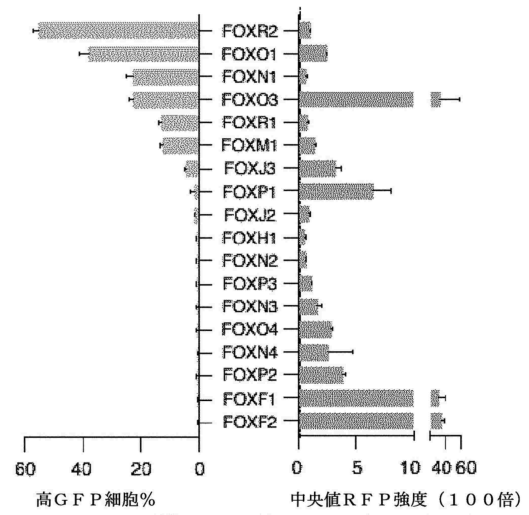
40

50

【図 8 B】

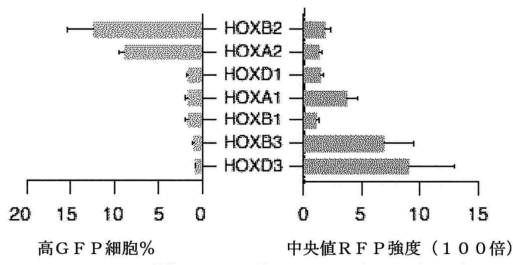


【図 8 C】

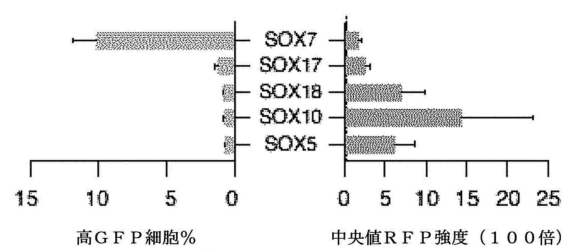


10

【図 8 D】



【図 8 E】



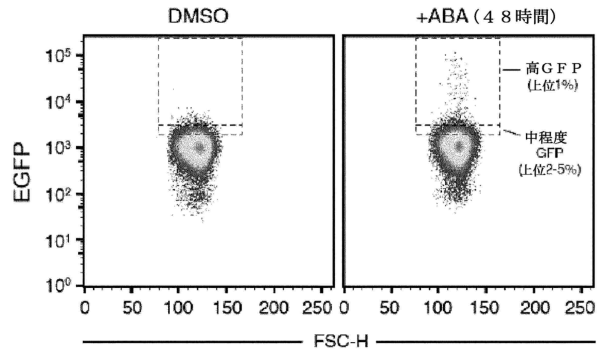
20

30

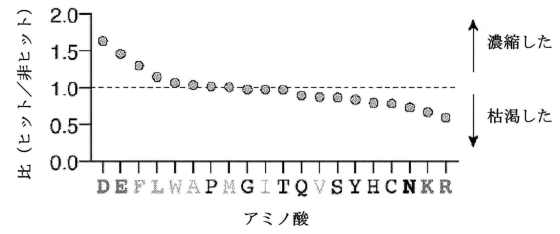
40

50

【図 9 A】

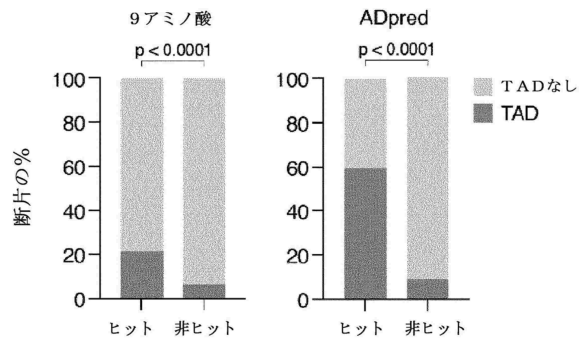


【図 9 B】

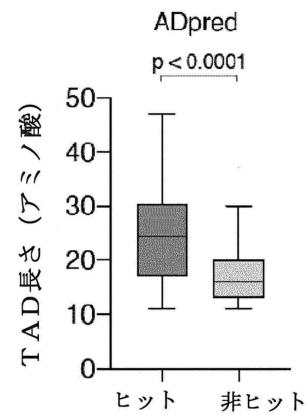


10

【図 9 C】



【図 9 D】



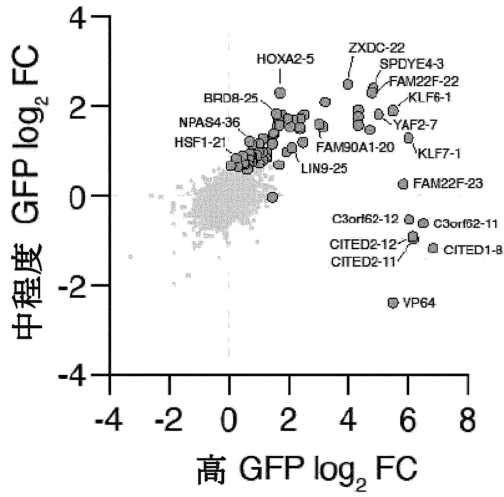
20

30

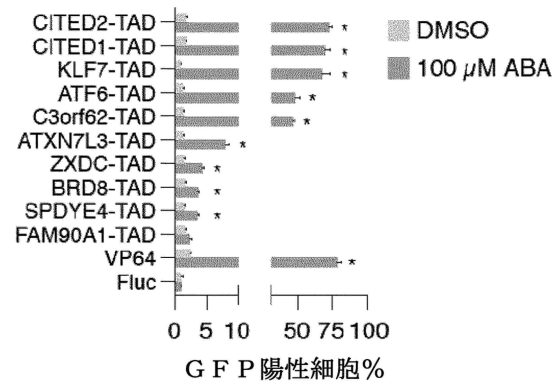
40

50

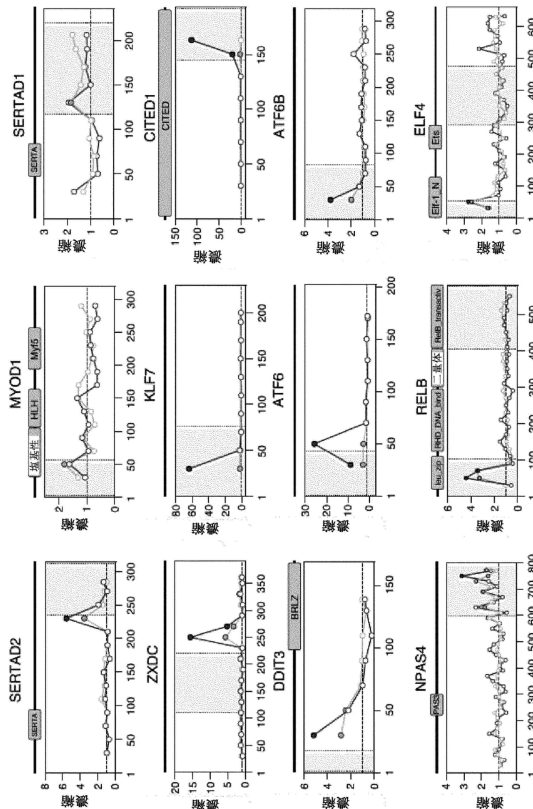
【図 9 E】



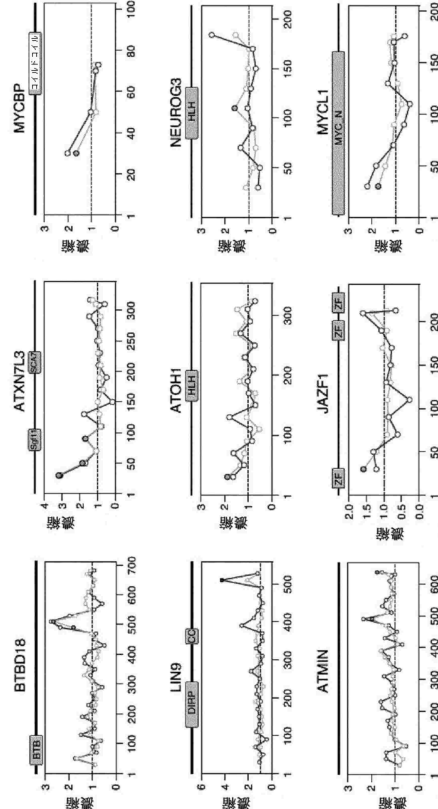
【図 9 F】



【図 10 A】



【図 10 B】



10

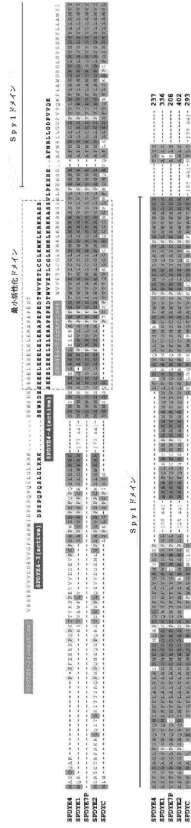
20

30

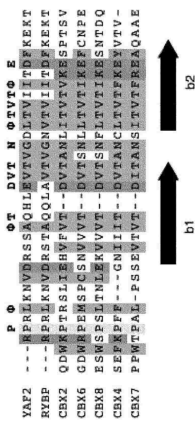
40

50

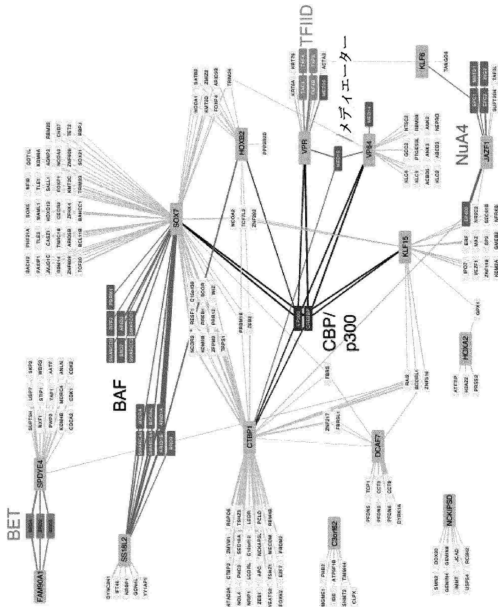
【図 1 1 A】



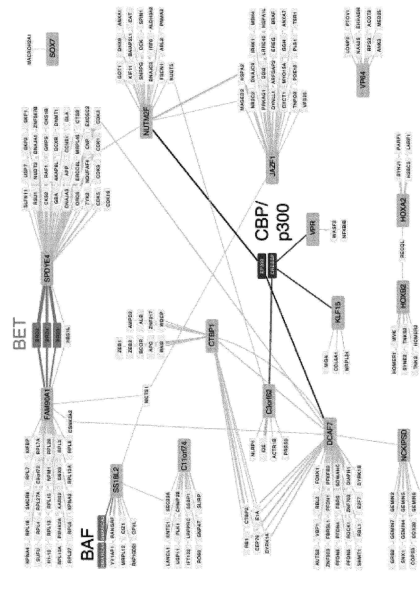
【図 1 1 B】



【図 1 2 A】



【図 1 2 B】



10

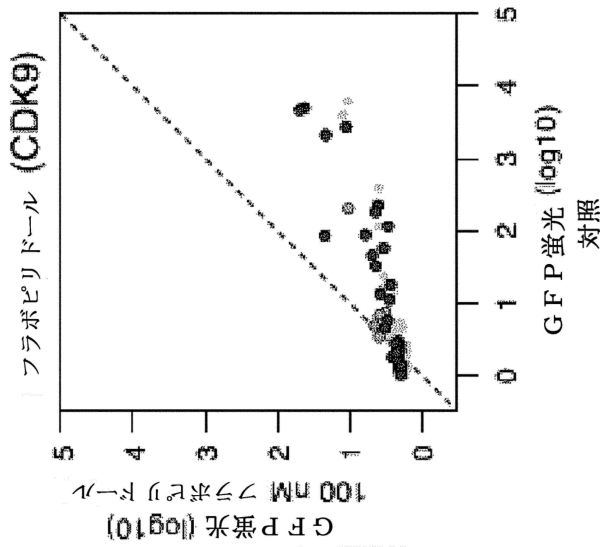
20

30

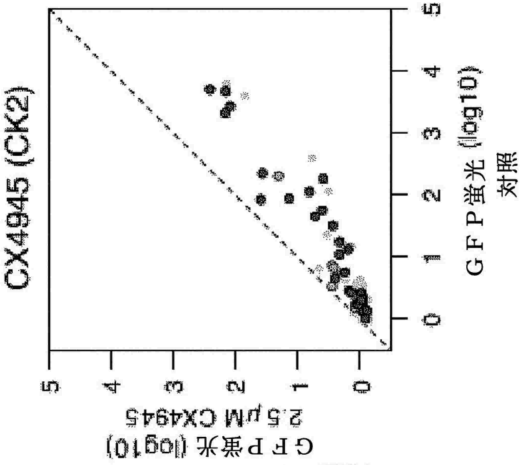
40

50

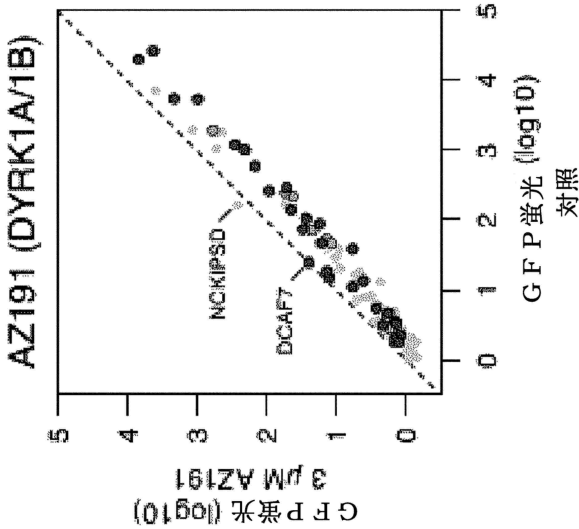
【図 1 3 A】



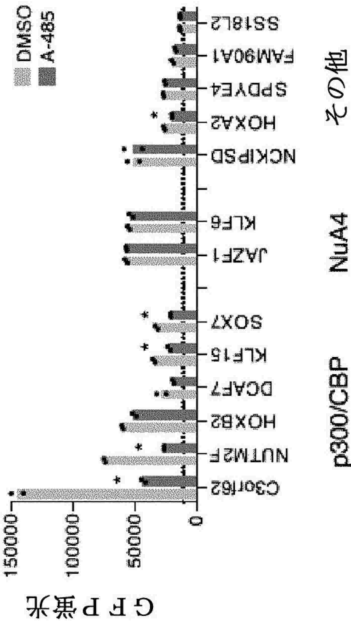
【図 1 3 B】



【図 1 3 C】



【図 1 3 D】



10

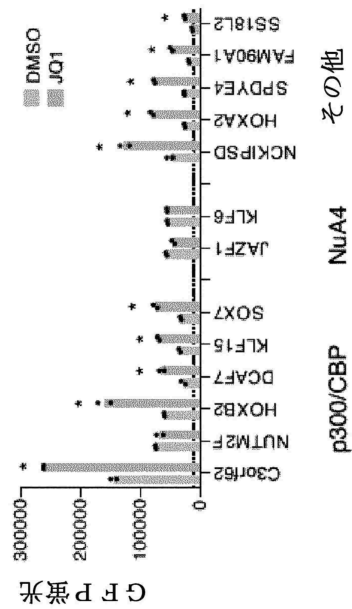
20

30

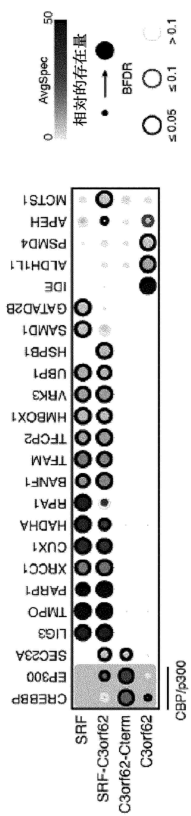
40

50

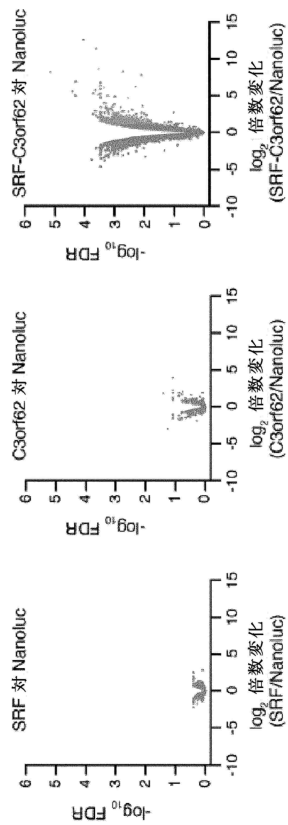
【図 1 3 E】



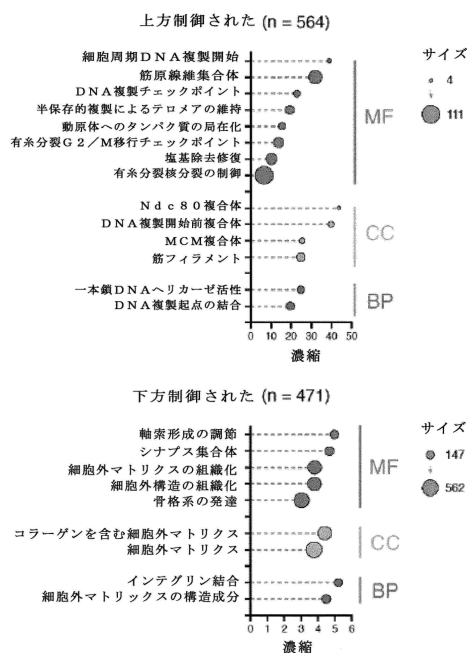
【図 1 4 A】



【図 1 4 B】



【図 1 4 C】



10

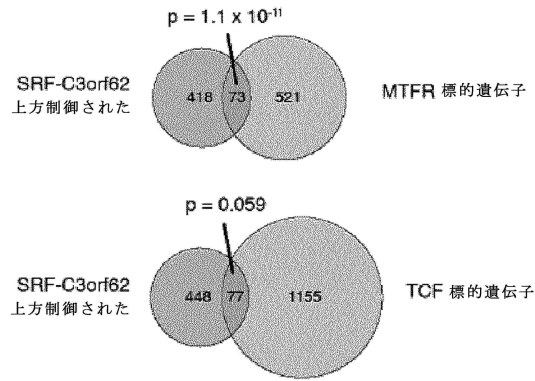
20

30

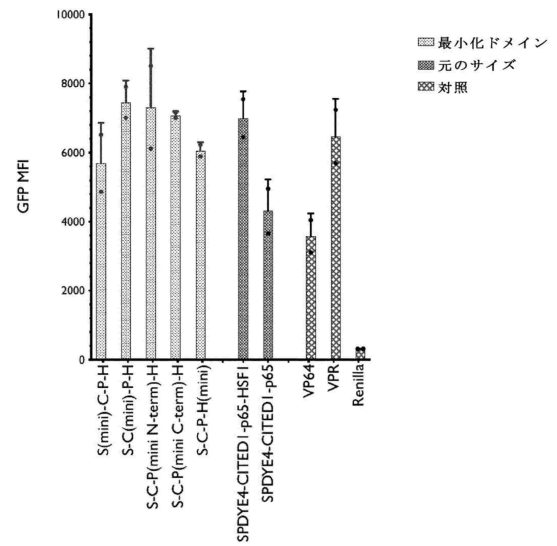
40

50

【図 14 D】

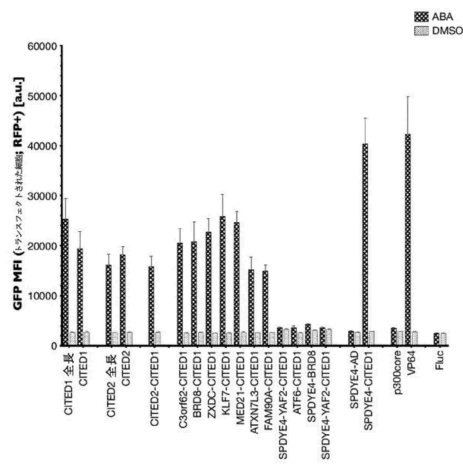


【図 15】

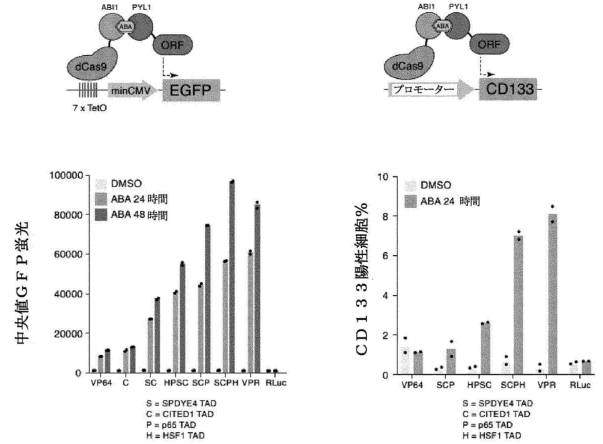


10

【図 16】



【図 17】



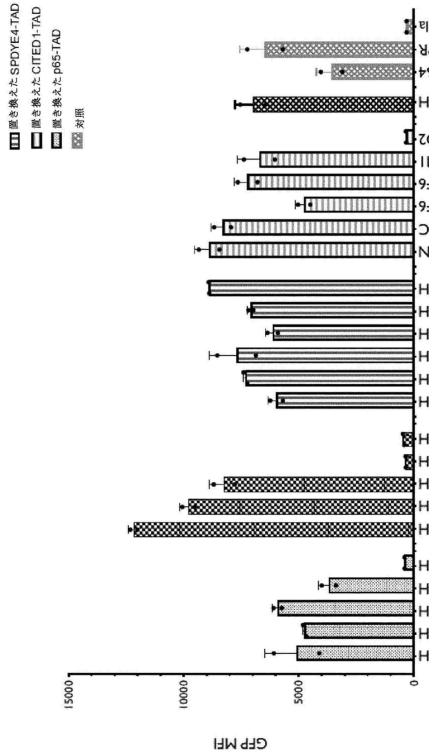
20

30

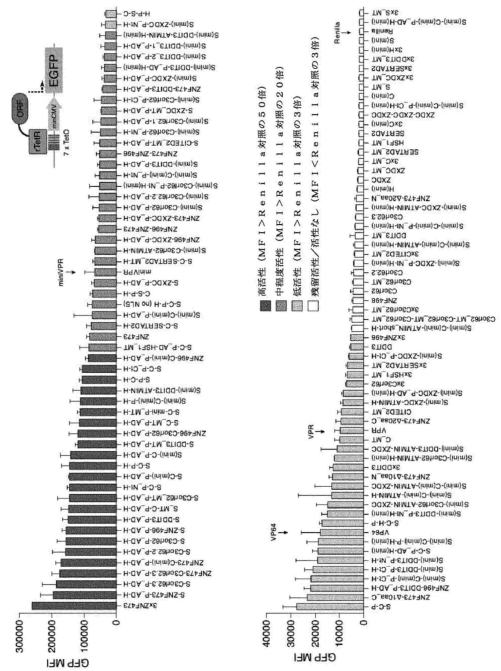
40

50

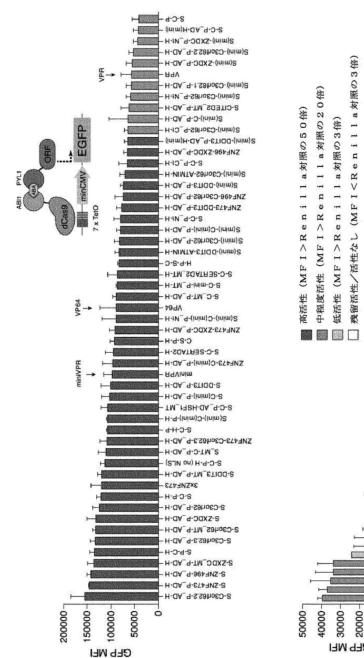
【図 18】



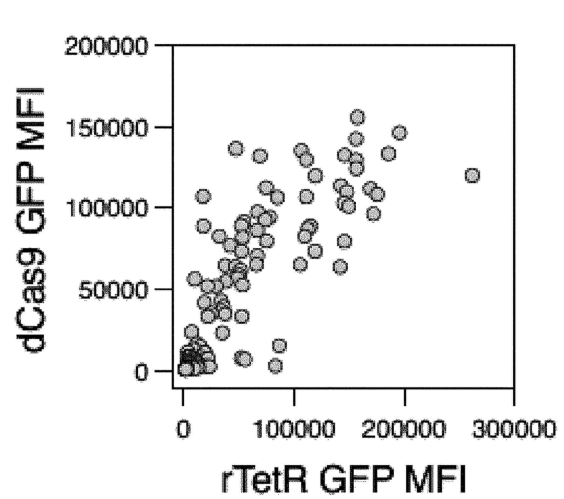
【図 19 A】



【図 19 B】



【図 19 C】



【配列表】

202452469600001.xml

10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/CA2022/051100
A. CLASSIFICATION OF SUBJECT MATTER IPC: <i>C12N 15/62</i> (2006.01), <i>C12N 9/14</i> (2006.01), <i>C12N 15/113</i> (2010.01), <i>C12N 15/861</i> (2006.01) CPC: <i>C12N 15/62</i> (2020.01), <i>C12N 9/14</i> (2020.01), <i>C12N 15/113</i> (2020.01), <i>C12N 15/861</i> (2020.01) According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC: All Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used) Databases: Questel Orbit, Google Patent, Scopus, GenomeQuest Keywords: SEQ ID NO: 32, ATMIN, transactivation domain, transcriptional activator, DNA targeting domain, inducible, heterodimer		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KALLUNKI, T., et AL, "How to Choose the Right Inducible Gene Expression System for Mammalian Studies?". CELLS, 30 July 2019 (30-07-2019), Vol. 8(8), pp. 796, ISSN 2073-4409. Retrieved from the Internet: <doi: 10.3390/cells8080796> * Figures 3A, C; pages 5-7*	3, 5, and 6
X	JURADO, S., ETA L, "Dual Functions of ASCIZ in the DNA Base Damage Response and Pulmonary Organogenesis". PLoS Genetics, 21 October 2010 (21-10-2010), Vol. 6(10), pp. e1001170, ISSN 1553-7404. Retrieved from the Internet: <doi: 10.1371/journal.pgen.1001170> * Figure 8; Page 7, right-hand column-page 8, left-hand column	1, 4, 7, 11-14, 17, 21 and 22
Y		2, 3, 5, 6, 8-10, 15, 16 and 18-20
☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search 20 September 2022 (20-09-2022)		Date of mailing of the international search report 14 October 2022 (14-10-2022)
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 819-953-2476		Authorized officer Urszula Liwak (819) 360-9443

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2022/051100

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CASAS-MOLLANO, J.A., et AL, " <i>CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes</i> ". The CRISPR Journal, October 2020 (10-2020), Vol. 3(5), pp. 350-364, ISSN 2573-1602. Retrieved from the Internet: <doi: 10.1089/crispr.2020.0064> *Figure 1; page 354, first paragraph; pages 350-352; Table 1*	2, 8-10, 15, 16, and 18-20
A	TYCKO, J., et AL, " <i>High-throughput discovery and characterization of human transcriptional effectors</i> ". Cell, 23 December 2020 (23-12-2020), Vol. 183(7), pp. 2020-2035.e16, ISSN 1097-4172. Retrieved from the Internet: <doi:10.1016/j.cell.2020.11.024> *whole document*	
A	ARNOLD, C.D., et AL, " <i>A high-throughput method to identify trans-activation domains within transcription factor sequences</i> ". The EMBO Journal, Epub 13 July 2018 (13-07-2018), Vol. 37(16), e98896, ISSN 1460-2075. Retrieved from the Internet: doi: 10.15252/embj.201798896. *whole document*	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2022/051100

Box No. I **Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

- a. ☒ forming part of the international application as filed.
- b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)),
 ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2022/051100

Box No. II	Observations where certain claims were found unsearchable (Continuation of item 2 of the first sheet)
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons: 1. ☐ Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely: 2. ☐ Claim Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically: 3. ☐ Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).	
Box No. III	Observations where unity of invention is lacking (Continuation of item 3 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows: Groups 1-70 - Claims 1-22 (all partially) are directed to a heterologous transcriptional activator comprising a DNA targeting domain and an effector domain comprising at least one transactivation domain (TAD) selected from the TADs listed in Table 2 or Table 6, or a functional variant thereof, optionally Table 6, or at least two TADs selected from the TADs listed in Table 1 or Table 3, or functional variants thereof, preferably at least one TAD selected from the TADs listed in Table 4, or Table 5 or Table 6, and/or functional variants thereof, wherein the DNA targeting domain and the effector domain are operably linked, and wherein each unique combination of DNA targeting domain and effector domain is considered a separate alleged invention. Continued on extra sheet. 1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims. 2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees. 3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claim Nos.: 4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim Nos.: 1-22 as limited to a heterologous transcriptional activator comprising a DNA targeting domain and an effector domain comprising at least one transactivation domain (TAD), wherein the TAD is ATMIN-24 (SEQ ID NO:32), and wherein the DNA targeting domain and the effector domain are operably linked Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee. ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation. ☐ No protest accompanied the payment of additional search fees.	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA2022/051100

Continuation of: Box No. III

An *a posteriori* analysis has concluded that there is no common inventive concept linking the subject matter of the Groups 1-70. Kallunki, T., et al. "How to Choose the Right Inducible Gene Expression System for Mammalian Studies?", CELLS, vol. 8(8):796, 30 July 2019 (30-07-2019) discloses a heterologous transcriptional activator comprising a DNA targeting domain and an effector domain comprising at least one transactivation domain (TAD) for inducing gene expression.

The claims must be limited to one inventive concept as set out in PCT Rule 13.

フロントページの続き

(51)国際特許分類

F I

テーマコード (参考)

C 1 2 N	15/31 (2006.01)	C 1 2 N	15/31
C 1 2 Q	1/68 (2018.01)	C 1 2 Q	1/68
C 1 2 Q	1/02 (2006.01)	C 1 2 Q	1/02

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,N
E,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,
CV,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,IT,J
M,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY
,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,T
H,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

(74)代理人 100202751

弁理士 岩堀 明代

(74)代理人 100208580

弁理士 三好 玲奈

(74)代理人 100191086

弁理士 高橋 香元

(72)発明者 タイパレ, ミッコ ジューナス オスカリ

カナダ国, オンタリオ州 エム6ジェー 3エイチ3, トロント, 63 ラスホルム ロード

(72)発明者 アレラスール, ナダー

アメリカ合衆国, カリフォルニア州 94110, サンフランシスコ, アpartment 629, 2
823 エイティーンズ ストリート

(72)発明者 レング, ヘ

カナダ国, オンタリオ州 エム5ティー 2エックス5, トロント, 60 セイント パトリック ス
トリート, 530 #

F ターム (参考) 4B063 QA01 QQ52 QR77 QX01

4B065 AA90X AB01 AC14 BA02 CA23 CA24