

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2007-289152

(P2007-289152A)

(43) 公開日 平成19年11月8日(2007.11.8)

(51) Int.Cl.	F I	テーマコード (参考)
C12N 15/09 (2006.01)	C12N 15/00 Z N A A	4 B O 2 4
C40B 40/06 (2006.01)	C40B 40/06	4 B O 6 3
C12Q 1/68 (2006.01)	C12Q 1/68 Z	
GO1N 33/53 (2006.01)	GO1N 33/53 M	
GO1N 33/566 (2006.01)	GO1N 33/566	
審査請求 未請求 請求項の数 36 O L 外国語出願 (全 149 頁) 最終頁に続く		

(21) 出願番号 特願2007-63117 (P2007-63117)
 (22) 出願日 平成19年3月13日 (2007.3.13)
 (31) 優先権主張番号 11/373, 519
 (32) 優先日 平成18年3月13日 (2006.3.13)
 (33) 優先権主張国 米国 (US)

(71) 出願人 504354117
 エイジェンシー・フォー・サイエンス、テクノロジー・アンド・リサーチ
 シンガポール国, 138668, バイオポリス・ウェイ 20, #07-01 セントロス
 (74) 代理人 100099623
 弁理士 奥山 尚一
 (74) 代理人 100096769
 弁理士 有原 幸一
 (74) 代理人 100107319
 弁理士 松島 鉄男

最終頁に続く

(54) 【発明の名称】 核酸相互作用分析

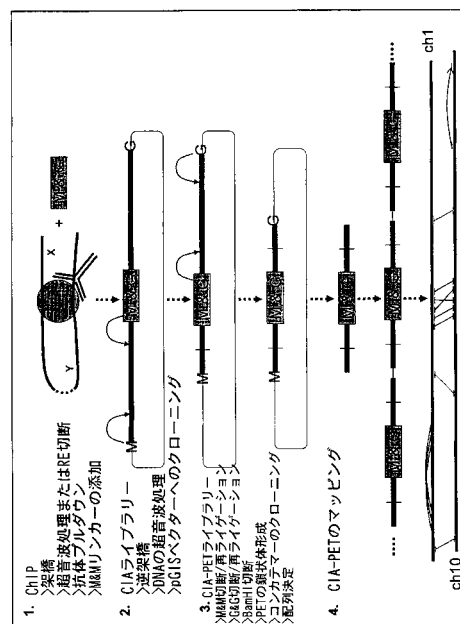
(57) 【要約】 (修正有)

【課題】単離オリゴヌクレオチドおよび核酸 - タンパク質複合体由来の少なくとも2つのポリヌクレオチドを検出および/または同定するための単離オリゴヌクレオチドの使用法を提供する。

【解決手段】DNA / 蛋白質 / DNA複合体の結合されているDNAフラグメントをリンカー配列によって連結する。リンカー配列は制限酵素切断部位を含む。蛋白質複合体を架橋により固定後制限酵素でリンカーを切断する。そして遊離した末端をタグとして塩基配列の解析を行なう。

【選択図】 図2

FIGURE 2



【特許請求の範囲】**【請求項 1】**

少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含む単離オリゴヌクレオチドであって、前記第 1 のタグおよび前記第 2 のタグが核酸 - タンパク質複合体のタグである、単離オリゴヌクレオチド。

【請求項 2】

前記第 1 のタグが第 1 のポリヌクレオチドのタグであり、前記第 2 のタグが第 2 のポリヌクレオチドのタグであり、ここで、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体のポリヌクレオチドである、請求項 1 に記載の単離オリゴヌクレオチド。

10

【請求項 3】

前記オリゴヌクレオチドが、少なくとも 1 つの制限酵素の認識部位をさらに含む、請求項 1 または 2 に記載の単離オリゴヌクレオチド。

【請求項 4】

前記オリゴヌクレオチドが、少なくとも 1 つのリンカーをさらに含む、請求項 1 ~ 3 のいずれかに記載の単離オリゴヌクレオチド。

【請求項 5】

前記リンカーが、少なくとも 1 つの制限酵素の認識部位を含む、請求項 4 に記載の単離オリゴヌクレオチド。

【請求項 6】

前記少なくとも 1 つのリンカーが、タグの間に挿入されているか、第 1 のタグの上流または第 2 のタグの下流に配置されている、請求項 4 または 5 に記載の単離オリゴヌクレオチド。

20

【請求項 7】

前記少なくとも 1 つの制限酵素の認識部位が、I Is 型制限酵素の認識部位である、請求項 3、5 または 6 のいずれかに記載の単離オリゴヌクレオチド。

【請求項 8】

前記少なくとも 1 つの制限酵素の認識部位が、ホーミング制限酵素の認識部位である、請求項 3、5、または 6 に記載の単離オリゴヌクレオチド。

【請求項 9】

前記少なくとも 1 つの第 1 のタグが第 1 のポリヌクレオチドの 5' 末端および 3' 末端を含み、前記少なくとも 1 つの第 2 のタグが第 2 のポリヌクレオチドの 5' 末端および 3' 末端を含み、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体のポリヌクレオチドである、請求項 1 ~ 8 のいずれかに記載の単離オリゴヌクレオチド。

30

【請求項 10】

前記リンカーが、第 1 のポリヌクレオチドを切断して第 1 のタグを得ることができる制限酵素によって認識される第 1 の制限認識部位、および第 2 のポリヌクレオチドを切断して第 2 のタグを得ることができる制限酵素によって認識される第 2 の制限認識部位を含み、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体のポリヌクレオチドである、請求項 5 または 6 に記載の単離オリゴヌクレオチド。

40

【請求項 11】

前記リンカーが、第 1 のポリヌクレオチドを切断して第 1 のポリヌクレオチドの 3' 末端を得ることができる第 1 の制限酵素によって認識される第 1 の制限認識部位、および第 2 のポリヌクレオチドを切断して第 2 のポリヌクレオチドの 5' 末端を得ることができる第 2 の制限酵素によって認識される第 2 の制限認識部位を含み、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体のポリヌクレオチドである、請求項 5 または 6 に記載の単離オリゴヌクレオチド。

【請求項 12】

前記リンカーが、第 1 のポリヌクレオチドを切断して第 1 のポリヌクレオチドの 5' 末

50

端を得ることができる第 1 の制限酵素によって認識される第 1 の制限認識部位、および第 2 のポリヌクレオチドを切断して第 2 のポリヌクレオチドの 3' 末端を得ることができる第 2 の制限酵素によって認識される第 2 の制限認識部位を含み、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体のポリヌクレオチドである、請求項 5 または 6 に記載の単離オリゴヌクレオチド。

【請求項 13】

前記リンカーが、第 1 のポリヌクレオチドを切断して第 1 のポリヌクレオチドの 3' 末端を得ることができる第 1 の制限酵素によって認識される第 1 の制限認識部位、および第 2 のポリヌクレオチドを切断して第 2 のポリヌクレオチドの 5' 末端を得ることができる第 2 の制限酵素によって認識される第 2 の制限認識部位を含み、前記第 1 のポリヌクレオチドが、第 1 のポリヌクレオチドを切断して第 1 のポリヌクレオチドの 5' 末端を得ることができる第 3 の制限酵素によって認識される第 3 の制限認識部位をさらに含み、前記第 2 のポリヌクレオチドが、第 2 のポリヌクレオチドを切断して第 2 のポリヌクレオチドの 3' 末端を得ることができる第 4 の制限酵素によって認識される第 4 の制限認識部位を含み、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体のポリヌクレオチドであり、前記少なくとも 1 つの第 1 のタグが前記第 1 のポリヌクレオチドの 5' 末端および 3' 末端のライゲーションから得られ、前記少なくとも 1 つの第 2 のタグが前記第 2 のポリヌクレオチドの 5' 末端および 3' 末端のライゲーションから得られる、請求項 5、6、または 11 のいずれかに記載の単離ポリヌクレオチド。

10

【請求項 14】

前記核酸 - タンパク質複合体がクロマチン構造の一部である、請求項 1 ~ 13 のいずれかに記載の単離オリゴヌクレオチド。

20

【請求項 15】

前記ポリヌクレオチドが DNA または RNA である、請求項 1 ~ 14 のいずれかに記載の単離オリゴヌクレオチド。

【請求項 16】

請求項 1 ~ 15 のいずれかに記載のオリゴヌクレオチドを含むベクター。

【請求項 17】

少なくとも 2 つの単離オリゴヌクレオチドを含むオリゴヌクレオチドのコンカテマーであって、各単離オリゴヌクレオチドが少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含み、前記第 1 のタグおよび第 2 のタグが核酸 - タンパク質複合体のタグである、コンカテマー。

30

【請求項 18】

前記第 1 のタグが第 1 のポリヌクレオチドのタグであり、前記第 2 のタグが第 2 のポリヌクレオチドのタグであり、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体に由来し、各単離オリゴヌクレオチドが少なくとも 1 つの制限酵素の認識部位を含む、請求項 17 に記載のオリゴヌクレオチドのコンカテマー。

【請求項 19】

前記少なくとも 1 つの制限部位が少なくとも 1 つのリンカーに含まれる、請求項 18 に記載のオリゴヌクレオチドのコンカテマー。

40

【請求項 20】

前記少なくとも 1 つのリンカーが、タグの間に挿入されているか、または第 1 のタグの上流および / または第 2 のタグの下流に配置されている、請求項 19 に記載のオリゴヌクレオチドのコンカテマー。

【請求項 21】

前記少なくとも 1 つの制限酵素の認識部位が、IIs 型制限酵素の認識部位である、請求項 18 ~ 20 のいずれかに記載のオリゴヌクレオチドのコンカテマー。

【請求項 22】

前記第 1 のタグが第 1 のポリヌクレオチド由来の 5' 末端および 3' 末端を含み、前記第 2 のタグが第 2 のポリヌクレオチド由来の 5' 末端および 3' 末端をさらに含む、請求

50

項 17 ~ 21 のいずれかに記載のオリゴヌクレオチドのコンカテマー。

【請求項 23】

請求項 1 ~ 15 のいずれかに記載の少なくとも 1 つのオリゴヌクレオチドを含むオリゴヌクレオチドライブラリー。

【請求項 24】

少なくとも 1 つの単離オリゴヌクレオチドを調製する方法であって、

(d) 核酸 - タンパク質複合体を供するステップと、

(e) 少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含むオリゴヌクレオチドを調製するステップであって、前記第 1 のタグおよび第 2 のタグが前記核酸 - タンパク質複合体から得られる、ステップと、

10

(f) 前記オリゴヌクレオチドを単離するステップと、
を含む、方法。

【請求項 25】

前記第 1 のタグが第 1 のポリヌクレオチドから得られ、前記第 2 のタグが第 2 のポリヌクレオチドから得られ、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体から得られる、請求項 24 に記載の方法。

【請求項 26】

前記ステップ (b) が、

(i) 少なくとも 1 つの制限酵素の認識部位を含む少なくとも 1 つのリンカーを挿入するステップと、

20

(ii) 前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドを、リンカー中の少なくとも 1 つの制限部位を認識する少なくとも 1 つの制限酵素で切断して、第 1 のポリヌクレオチドから得た第 1 のタグ、第 2 のポリヌクレオチドから得た第 2 のタグ、およびリンカーを含むオリゴヌクレオチドを形成するステップと
を含む、請求項 24 または 25 に記載の方法。

【請求項 27】

前記ステップ (b) が、

(i) 少なくとも 1 つの制限酵素の認識部位を含む少なくとも 1 つのリンカーを、複合体の第 1 のポリヌクレオチドと第 2 のポリヌクレオチドとの間に挿入するステップと、

(ii) 少なくとも 1 つの制限酵素の制限部位を前記第 1 のポリヌクレオチドの 5' 末端および前記第 2 のポリヌクレオチドの 3' 末端の各々に付加するステップと、

30

(iii) 前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドを、少なくとも 1 つの認識部位を認識する少なくとも 1 つの制限酵素で切断して切断フラグメントを得るステップと、

(iv) 前記切断フラグメントをライゲーションして、前記第 1 のポリヌクレオチドから得た第 1 のタグ、前記第 2 のポリヌクレオチドから得た第 2 のタグ、各ポリヌクレオチドの 5' 末端および 3' 末端を含むタグ、および前記タグの間に挿入されたリンカーを含むオリゴヌクレオチドを形成するステップと
を含む、請求項 24 または 25 に記載の方法。

【請求項 28】

40

前記ステップ (ii) の少なくとも 1 つの制限酵素の認識部位が、アダプターまたはベクターの一部である、請求項 27 に記載の方法。

【請求項 29】

前記核酸 - タンパク質複合体が、クロマチン免疫沈降によって得られたものである、請求項 24 ~ 28 のいずれかに記載の方法。

【請求項 30】

核酸 - タンパク質複合体由来の少なくとも 2 つのポリヌクレオチドを検出および / または同定する方法であって、

(g) 核酸 - タンパク質複合体を供するステップと、

(h) 少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含むオリゴ

50

ヌクレオチドを調製するステップであって、前記第 1 のタグが第 1 のポリヌクレオチドから得られ、前記第 2 のタグが第 2 のポリヌクレオチドから得られ、前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体から得られるステップと、

(i) 前記オリゴヌクレオチドを配列決定するステップと、

(j) 第 1 のタグおよび第 2 のタグのヌクレオチド配列に基づいて少なくとも 2 つのポリヌクレオチドをマッピングし、それにより、前記少なくとも 2 つのポリヌクレオチドを検出および / または同定するステップとを含む、方法。

【請求項 3 1】

10

前記ステップ (b) が、

(i) 少なくとも 1 つの制限認識部位を含む少なくとも 1 つのリンカーを挿入するステップと、

(i i) 前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドを、リンカー中の少なくとも 1 つの制限部位を認識する少なくとも 1 つの制限酵素で切断して、第 1 のポリヌクレオチドから得た第 1 のタグ、第 2 のポリヌクレオチドから得た第 2 のタグ、および前記タグの間にリンカーを含むオリゴヌクレオチドを形成するステップとを含む、請求項 3 0 に記載の方法。

【請求項 3 2】

20

前記ステップ (b) が、

(i) 少なくとも 1 つの制限酵素の認識部位を含む少なくとも 1 つのリンカーを、複合体の第 1 のヌクレオチドと第 2 のヌクレオチドとの間に挿入するステップと、

(i i) 少なくとも 1 つの制限酵素の制限部位を前記第 1 のポリヌクレオチドの 5 ' 末端および前記第 2 のポリヌクレオチドの 3 ' 末端の各々に付加するステップと、

(i i i) 前記第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドを、少なくとも 1 つの認識部位を認識する少なくとも 1 つの制限酵素で切断して切断フラグメントを得るステップと、

(i v) 前記切断フラグメントをライゲーションして、前記第 1 のポリヌクレオチドから得た第 1 のタグ、前記第 2 のポリヌクレオチドから得た第 2 のタグ、各ポリヌクレオチドの 5 ' 末端および 3 ' 末端を含むタグ、および前記タグの間に挿入されたリンカーを含むオリゴヌクレオチドを形成するステップとを含む、請求項 3 0 に記載の方法。

30

【請求項 3 3】

前記ステップ (i i) の少なくとも 1 つの制限酵素の認識部位が、アダプターまたはベクターの一部である、請求項 3 2 に記載の方法。

【請求項 3 4】

前記核酸 - タンパク質複合体が、クロマチン免疫沈降によって得られたものである、請求項 3 0 ~ 3 3 のいずれかに記載の方法。

【請求項 3 5】

前記オリゴヌクレオチドが、配列決定前にステップ (a) ~ (b) によって得られた少なくとも 1 つのさらなるオリゴヌクレオチドと鎖状体形成される (concatenate)、請求項 3 0 ~ 3 4 のいずれかに記載の方法。

40

【請求項 3 6】

前記ポリヌクレオチドが同一染色体上に存在するか、前記ポリヌクレオチドが異なる染色体上に存在する、請求項 3 0 ~ 3 5 のいずれかに記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般に、遺伝子発現分野に関する。特に、本発明は、核酸相互作用に関する。特に、本発明は、核酸 - タンパク質相互作用事象および成分の分析、検出および同定に

50

関する。

【背景技術】

【0002】

クロマチン相互作用は、遺伝子調節において重要である。最近完了したヒトゲノム配列により、遺伝情報の枠組みが得られる。しかし、ヒトゲノム構造および情報は、しばしば、一次元の直線として示され、細胞系の複雑さおよび連携を説明するには不十分である。どのようにしてゲノムが実際に生きた細胞中の三次元の核内で組織化された系として機能するのかを理解するために完全に異なる観点が必要である。ゲノムDNA（全長2mと推定されている）は、核内の染色体中にわずかに数ミクロンに凝縮されている。染色体はユー

10

【0003】

これらの活動は、順序付けられているようである。巨大な染色体ループが活性遺伝子を含むことが認められている。さらに、遺伝子座調節領域（LCR）、エンハンサー、およびインスレーターなどの遠位調節エレメントが特定の遺伝子座を活性またはサイレントな転写を行う領域に再配置することによって作用することが示唆されている。最近の研究により、 β -グロビン、ごく最近では、サイトカイン遺伝子（IFN- γ ）で、LCRが同一染色体上の離れたプロモーターと直接相互作用することができ、異なる染色体上のプロモーターでさえも相互作用することができることが証明されている。染色体間相互作用および染色体間相互作用が、重要な経路の遺伝子調節を調整する複数の遺伝子座で起こる一

20

【0004】

クロマチン相互作用の研究のために使用されるテクノロジー - 三次元構造およびクロマチン相互作用を研究するために多数のアプローチが使用されているが、これらは全て非常に限定されている。この問題に適用することができるテクノロジーを、視覚化ツール（顕微鏡法、蛍光インサイチューハイブリッド形成法（FISH）、およびRNA-TRAP

30

【0005】

多数の初期の研究で核内のクロマチンの空間的構成を調査するために顕微鏡法を使用していた。しかし、このような細胞遺伝学的アプローチでは、染色体中のクロマチンの大まかなセグメント情報しか得ることができない。FISHは、この傾向を有意に改良しており、特定の遺伝子座をゲノムDNAとハイブリッド形成する蛍光標識DNAまたはRNAプローブを介して染色体上の特定の物理的位置に局在化させる。しかし、分解能は依然として非常に限られている。FISHを修正したRNA-TRAPは、遺伝子プロモーター

40

【0006】

染色体高次構造捕捉法（3C）は、本来、酵母の染色体高次構造を調査するためにデザインされ（Dekker et al, 2002）、長距離および/または異なる染色体に分離した遺伝因子の相互作用を研究するために使用されている。3Cでは、DNA-タンパク質（クロマチン）構造は、インビボでホルムアルデヒド架橋しており、クロマチンは、制限酵素消化によって断片化される。次いで、DNA結合タンパク質によって繫留されたDNAフラグメントを、ライゲーションによって互いに連結し、2つの疑いのある公知のエレメントの連結点（junction）を、PCRによって検出する。3C手順で得られたタンパク質と架橋したキメラDNAフラグメントを抗体プルダウンによって富化させるクロマチン免疫沈降（3

50

C - C h I P) によって、特定の D N A 結合タンパク質または転写因子によって媒介されたクロマチン相互作用の検出をさらに増強することができる。

【 0 0 0 7 】

各技術およびいくつかの技術の組み合わせがいくつかの特定の染色体内相互作用および染色体間相互作用の同定に有用であることが証明されているが、これらのアプローチは、どのような可能な遠位クロマチン相互作用が存在することができるかについての既存の知識または推測と、P C R によって 1 回に 1 つの領域のこのような連結点を検出するためにデザインしたプライマーとに依存する。したがって、クロマチン相互作用を研究するための現在のテクノロジーでは、新規のクロマチン相互作用の同定および全ゲノムレベルでの大規模の同定が非常に制限される。

10

【 0 0 0 8 】

染色体が核内で空間的に組織化される方法およびどのようにして染色体が遠位遺伝子の転写を一斉に調節することができるのかについて非常に興味を持たれているにもかかわらず、現在、バラバラの間接的な情報しか利用できない。この態様の情報の欠如は、主に、染色体相互作用の三次元の問題に有効に取り組むことができる確固たるテクノロジーの欠如に起因する。

【 0 0 0 9 】

染色体相互作用の三次元の問題に有効に取り組むことができ、既存の技術の欠点および制限を克服することができるより有効な方法および確固たるテクノロジーが当分野で必要である。

20

【 発明の開示 】

【 0 0 1 0 】

本発明は、核酸複合体、特に核酸 - タンパク質複合体由来の少なくとも 1 つの核酸配列またはフラグメントを検出、同定および / または調製する新規の方法を提供することによって上記問題を解決する。特に、本発明の方法は、核酸 - タンパク質複合体由来の少なくとも 2 つの核酸配列またはフラグメントを検出、同定および / または調製する方法を提供する。本発明はまた、本発明の任意の実施形態にかかる方法を使用して調製したオリゴヌクレオチドに関する。特定の D N A 結合タンパク質によって媒介される長い距離および異なる染色体間のクロマチン相互作用事象を同定する方法も提供する。

【 0 0 1 1 】

別の態様では、本発明は、少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含む単離オリゴヌクレオチドであって、第 1 のタグおよび第 2 のタグが核酸 - タンパク質複合体のタグである、単離オリゴヌクレオチドの提供によって上記問題を解決する。特に、少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含む単離オリゴヌクレオチドであって、第 1 のタグが第 1 のポリヌクレオチドから得られ、第 2 のタグが第 2 のポリヌクレオチドから得られ、第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体から得られる、単離オリゴヌクレオチドを提供する。第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドは、同一の核酸領域の一部であり得、また、核酸 - タンパク質複合体の異なる核酸領域に由来し得る。

30

【 0 0 1 2 】

単離オリゴヌクレオチドは、さらに、少なくとも 1 つの制限酵素の認識部位、少なくとも 1 つのリンカーを含み得る。特に、少なくとも 1 つの制限酵素の認識部位を、リンカー中に含めることができる。リンカーをタグの間 (between the tags) に挿入することができる、また、リンカーを少なくとも 1 つのタグに隣接させる (すなわち、少なくとも 1 つのタグの上流および / または下流に配置する) ことができる。少なくとも 1 つの制限酵素の認識部位は、非対称であり得る。少なくとも 1 つの制限酵素の認識部位は、例えば、I I s 型制限酵素またはホーミング制限酵素の認識部位であり得る。

40

【 0 0 1 3 】

少なくとも 1 つの第 1 のタグは、第 1 のポリヌクレオチド由来の 5 ' 末端および 3 ' 末端を含むことができ、少なくとも 1 つの第 2 のタグは、第 2 のポリヌクレオチド由来の 5

50

′末端および3′末端を含むことができる。単離ポリヌクレオチドは、さらに、少なくとも1つのリンカーを含むことができる。リンカーをタグの間に挿入することができ、また、リンカーを少なくとも1つのタグの上流および/または下流に配置することができる。リンカーは、少なくとも1つの制限酵素の認識部位を含む。少なくとも1つの制限酵素の認識部位は、非対称であり得る。少なくとも1つの制限酵素の認識部位は、例えば、IIs型制限酵素またはホーミング制限酵素の認識部位であり得る。

【0014】

リンカーは、第1のポリヌクレオチドを切断して第1のタグを得ることができる制限酵素によって認識される第1の制限認識部位、および第2のポリヌクレオチドを切断して第2のタグを得ることができる制限酵素によって認識される第2の制限認識部位を含むことができる。特に、リンカーは、第1のポリヌクレオチドを切断して第1のポリヌクレオチドの3′末端を得ることができる制限酵素によって認識される第1の制限認識部位、および第2のポリヌクレオチドを切断して第2のポリヌクレオチドの5′末端を得ることができる第2の制限酵素によって認識される第2の制限認識部位を含むことができる。第1の制限認識部位および第2の制限認識部位を、同一または異なる制限酵素によって認識することができる。

10

【0015】

別の態様によれば、第1のポリヌクレオチドを、第3の認識部位を認識する第3の制限酵素によってさらに切断して第1のポリヌクレオチドの5′末端を得ることができ、第2のポリヌクレオチドを、第4の認識部位を認識する第4の制限酵素によって切断して第2のポリヌクレオチドの3′末端が得られる。この実施形態によれば、少なくとも1つの第1のタグは、第1のポリヌクレオチドの5′末端および3′末端のライゲーションによって得られ、少なくとも1つの第2のタグは、第2のポリヌクレオチドの5′末端および3′末端のライゲーションによって得られる。第3の認識部位および第4の認識部位を、同一または異なる制限酵素によって認識することができる。さらなる認識部位を、第1のポリヌクレオチドの5′末端および第2のポリヌクレオチドの3′末端にライゲーションしたアダプター中にそれぞれ含めることができる。あるいは、第3の制限部位および第4の制限部位は、第1のポリヌクレオチド-リンカー-第2のポリヌクレオチド構造が挿入されたベクター中に存在し得る。この場合、第3の制限部位は第1のポリヌクレオチドの5′末端に隣接し、第4の制限部位は第2のポリヌクレオチドの3′末端に隣接する。

20

30

【0016】

単離オリゴヌクレオチドの核酸-タンパク質複合体は、クロマチン構造の一部であり得る。目的のタンパク質が結合する核酸フラグメントは、目的のタンパク質が結合する領域（例えば、ヒストン結合部位）を含む任意の核酸フラグメントであり得る。ポリヌクレオチドは、DNAまたはRNAであり得る。

【0017】

別の態様では、本発明は、少なくとも2つの単離オリゴヌクレオチドを含むオリゴヌクレオチドのコンカテマーであって、各単離オリゴヌクレオチドが少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含み、第1のタグおよび第2のタグが核酸-タンパク質複合体のタグである、コンカテマーを提供する。特に、少なくとも2つの単離オリゴヌクレオチドを含むオリゴヌクレオチドのコンカテマーであって、各単離オリゴヌクレオチドが少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含み、第1のタグが第1のポリヌクレオチドのタグであり、第2のタグが第2のポリヌクレオチドのタグであり、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体に由来する、コンカテマーを提供する。コンカテマーは、少なくとも1つのリンカーをさらに含むことができる。リンカーは、少なくとも1つの制限酵素の認識部位を含むことができる。さらに、コンカテマーの各単離オリゴヌクレオチドは、少なくとも1つの制限酵素の認識部位を含み得る。少なくとも1つの制限酵素の認識部位は、少なくとも1つのリンカーおよび/または少なくとも1つのアダプター中に含まれてもよい。

40

【0018】

50

リンカーをタグの間に挿入することができ、また、リンカーを少なくとも1つのタグの上流および/または下流に配置することができる。リンカーは、少なくとも1つの制限酵素の認識部位を含むことができる。少なくとも1つの制限酵素の認識部位は、IIs型制限酵素またはホーミング制限酵素の認識部位であり得る。コンカテマーの各オリゴヌクレオチドの第1のタグは第1のポリヌクレオチド由来の5'末端および3'末端を含むことができ、第2のタグは第2のポリヌクレオチド由来の5'末端および3'末端を含むことができる。

【0019】

本発明のオリゴヌクレオチドまたはコンカテマーを、ベクターおよび/または細胞に挿入することができる。細胞は、細菌細胞であり得る。

10

【0020】

ポリヌクレオチド核酸-タンパク質複合体は、クロマチン構造の一部であり得る。ポリヌクレオチドは同一染色体上に存在することができ、また、異なる染色体上に存在することができる。

【0021】

別の態様では、本発明は、少なくとも1つのオリゴヌクレオチドを含むオリゴヌクレオチドのライブラリーまたはオリゴヌクレオチドのコンカテマーであって、オリゴヌクレオチドが少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含み、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体から得られる、オリゴヌクレオチドのライブラリーまたはオリゴヌクレオチドのコンカテマーを提供する。

20

【0022】

ライブラリーの少なくとも1つのオリゴヌクレオチドは、少なくとも1つのリンカーを含むことができる。リンカーをタグの間に挿入することができ、また、リンカーを少なくとも1つのタグの上流および/または下流に配置することができる。第1のタグは第1のポリヌクレオチド由来の5'末端および3'末端を含むことができ、第2のタグは第2のポリヌクレオチド由来の5'末端および3'末端を含むことができる。

【0023】

別の態様では、本発明は、少なくとも1つの単離オリゴヌクレオチドの調製方法であって、

30

(a) 核酸-タンパク質複合体を供する工程と、

(b) 少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含むオリゴヌクレオチドを調製する工程であって、第1のタグおよび第2のタグが核酸-タンパク質複合体から得られる、工程と、

(c) オリゴヌクレオチドを単離する工程と、
を含む、方法を提供する。

【0024】

特に、本発明は、少なくとも1つの単離オリゴヌクレオチドの調製方法であって、

(a) 核酸-タンパク質複合体を供する工程と、

40

(b) 少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含むオリゴヌクレオチドを調製する工程であって、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体から得られる、工程と、

(c) オリゴヌクレオチドを単離する工程と、
を含む、方法を提供する。

【0025】

1つの実施形態では、本発明のこの態様の工程(b)が、

(i) 少なくとも1つの制限酵素の認識部位を含む少なくとも1つのリンカーを挿入する工程と、

50

(i i) 第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドを、リンカー中の少なくとも 1 つの制限部位を認識する少なくとも 1 つの制限酵素で切断して、第 1 のポリヌクレオチドから得た第 1 のタグ、第 2 のポリヌクレオチドから得た第 2 のタグ、およびタグの間にリンカーを含むオリゴヌクレオチドを形成する工程とを含む。

【 0 0 2 6 】

別の実施形態では、本発明のこの態様の工程 (b) が、

(i) 少なくとも 1 つの制限酵素の認識部位を含む少なくとも 1 つのリンカーを、複合体の第 1 のポリヌクレオチドと第 2 のポリヌクレオチドとの間に挿入する工程と、

(i i) 少なくとも 1 つの制限酵素の各制限部位を第 1 のポリヌクレオチドの 5 ' 末端および第 2 のポリヌクレオチドの 3 ' 末端に付加する工程と、 10

(i i i) 第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドを少なくとも 1 つの認識部位を認識する少なくとも 1 つの制限酵素で切断して切断フラグメントを得る工程と

、
(i v) 切断フラグメントをライゲーションして、第 1 のポリヌクレオチドから得た第 1 のタグ、第 2 のポリヌクレオチドから得た第 2 のタグ、各ポリヌクレオチドの 5 ' 末端および 3 ' 末端を含むタグ、およびタグの間に挿入されたリンカーを含むオリゴヌクレオチドを形成する工程であって、工程 (i i) の少なくとも 1 つの制限酵素の認識部位がアダプターの一部またはベクターの一部である、工程とを含む。 20

【 0 0 2 7 】

核酸 (例えば、DNA) および / または目的のタンパク質への光活性化可能な部分の組み込み、ならび抗体媒介沈降または親和性媒介技術による核酸 / タンパク質複合体の単離によって、核酸 - タンパク質複合体からポリヌクレオチドを単離することができる。このような親和性ベースの技術の例には、ストレプトアビジン / ビオチン、グルタチオン - S - トランスフェラーゼ / グルタチオンマトリクス、マルトース結合タンパク質 / アミロースマトリクスの相互作用が含まれる。

【 0 0 2 8 】

別の態様では、本発明は、核酸 - タンパク質複合体由来の少なくとも 2 つのポリヌクレオチドを検出および / または同定する方法であって、 30

(a) 核酸 - タンパク質複合体を供する工程と、

(b) 少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含むオリゴヌクレオチドを調製する工程であって、第 1 のタグが第 1 のポリヌクレオチドから得られ、第 2 のタグが第 2 のポリヌクレオチドから得られ、第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体から得られる、工程と、

(c) オリゴヌクレオチドを配列決定する工程と、

(d) 第 1 のタグおよび第 2 のタグのヌクレオチド配列に基づいて少なくとも 2 つのポリヌクレオチドをマッピングし、それにより、少なくとも 2 つのポリヌクレオチドを検出および / または同定する工程とを含む方法を提供する。 40

【 0 0 2 9 】

工程 (b) で得たオリゴヌクレオチドを、工程 (c) で配列決定する前に増幅することができる。ポリメラーゼ連鎖反応によって増幅することができる。増幅したオリゴヌクレオチドを、増幅後、工程 (c) で配列決定する前に少なくとも 1 つの精製工程に供することができる。少なくとも 1 つの精製工程は、ゲル電気泳動であり得る。

【 0 0 3 0 】

本発明のオリゴヌクレオチドを、配列決定前に工程 (a) ~ (b) によって得た少なくとも 1 つのさらなるオリゴヌクレオチドと鎖状体形成する (concatenate) ことができる。

【 0 0 3 1 】

Sanger 法またはピロシーケンスなどの多重配列決定によって配列決定することができる 50

。

【 0 0 3 2 】

ポリヌクレオチドのトランスフュージョン(transfusion)または転座について検出および/または同定することができる。したがって、この方法を使用して、核酸 - タンパク質複合体(クロマチン中の核酸 - タンパク質複合体)中で互いに近接したポリヌクレオチドおよび/または遺伝子を検出および/または同定することができる。さらに、目的のタンパク質が結合する核酸フラグメントは、目的のタンパク質が結合する領域(例えば、ヒストン結合部位)を含む任意の核酸フラグメントであり得る。ポリヌクレオチドは、DNAまたはRNAであり得る。

【 0 0 3 3 】

オリゴヌクレオチドを、少なくとも1つの細胞にトランスフェクトすることができる。エレクトロポレーションによってトランスフェクションを行うことができる。細胞は、細菌細胞であり得る。ポリヌクレオチドは、DNAまたはRNAであり得る。

【 0 0 3 4 】

別の態様では、本発明は、オリゴヌクレオチド、オリゴヌクレオチドのコンカテマー、またはオリゴヌクレオチドまたはコンカテマーのライブラリーを含むベクターを提供する。

【 発明を実施するための最良の形態 】

【 0 0 3 5 】

[定義]

制限酵素 - 制限酵素(または制限エンドヌクレアーゼ)は、二本鎖DNAを切断する酵素である。この酵素は、2つの切り込み(塩基を損傷することなく二重らせんの各リン酸骨格に1つ)を作製する。酵素に切断された化学結合は、その末端が相補的であれば、リガーゼとして知られる他の酵素によって再形成することができ、これにより異なる染色体または遺伝子から得た制限フラグメントを互いにつなぎ合わせることができる。II型酵素は、特定の核酸配列を認識し、その認識配列部位の近くまたは内部の定義された位置でDNAを切断する。これらにより、個別の制限フラグメントおよび異なるゲルバンドパターンが得られる。IIs型酵素は、その認識配列の外側の一方を切断する。MmeIにより、ほとんどのIIs型制限酵素と同様に、長さが異なる末端が得られる。Dunn et al, 2002は、MmeIがおよそ1:1の比で18/20または19/21塩基を切断することができることを示した。全図中に示した配列は、それぞれMmeIを使用した1つの共通の変異形(variant)を示す。

【 0 0 3 6 】

PET中間体はまた、ポリヌクレオチドが異なる長さのポリヌクレオチドであり得るので、種々の長さを有する(すなわち、pGIS8プラスミドおよびMPET中間体へのクローニング後のMおよびGアダプターDNA配列)。III型酵素はまたは、巨大な制限酵素と修飾酵素との組み合わせである。これらは、その認識配列の外側を切断し、切断するために同一DNA分子内に反対方向に2つのこのような配列を必要とする。ホーミングエンドヌクレアーゼは、巨大な非対称の認識部位(12~40塩基対)およびイントロン(DNA)またはインティン(タンパク質)のいずれかに通常は包埋されたコード配列を有する稀な二本鎖DNAアーゼである。制限酵素は、平滑末端またはオーバーハングを有する粘着末端のいずれかが残存する切断物を作製することができる。粘着末端フラグメントを、最初に切断されたフラグメントだけでなく、適合可能な付着末端または粘着末端を有する任意の他のフラグメントとライゲーションすることができる。このようなものとして、異なる酵素によって産生された末端も適合可能である。したがって、粘着末端を、ライゲーションすることができる末端ともいうことができる。多数のII型制限酵素は、回文DNA配列を切断する。制限酵素が非縮重回文切断部位を有する場合、産生される全末端は適合可能である。「回文」配列は、一方の鎖上の配列が相補鎖上の配列を反対方向に読む配列である。そのようなものとして、その末端が結合し(mate)、核酸鎖が自己環状化するような回文付着末端が得られるように核酸鎖を処理することが可能である。本文

10

20

30

40

50

脈中での「回文」の意味は、その言葉遣いとは異なる。例えば、配列GTAATGは回文DNA配列ではなく、配列GTATACは回文DNA配列である。付着末端または粘着末端を遊離する制限酵素の例には、BamHI、EcoRI、およびHindIIIが含まれる。平滑末端、非付着末端、または非粘着末端を遊離する制限酵素の例には、BseRIおよびAluIが含まれる。

【0037】

ヌクレオチド - ヌクレオシドのリン酸エステル；核酸（DNAまたはRNA）の基本構造単位。ヌクレオチドは塩基対を形成する - 2つの鎖を有するDNA分子またはRNA分子の相補鎖を連結する水素結合によって結合した化学塩基対の1つ；塩基対は、DNAでは、アデニンとチミンとの対およびグアニンとシトシンの対、RNAではアデニンとウラシルの対およびグアニンとシトシンの対である。ヌクレオチドの短い鎖を、オリゴヌクレオチドといい、より長い鎖をポリヌクレオチドという。ヌクレオチドを、他のヌクレオチドと結合するか鎖状体形成することができる。用語「ヌクレオチド」を、用語「核酸」と交換可能に使用することができる。核酸ストレッチは、5'末端および3'末端を有する。核酸ストレッチの末端領域を、それぞれ5'末端および3'末端ということができる。ポリヌクレオチドの5'末端または3'末端を使用して、ポリヌクレオチドの実際の5'末端または3'末端を含むポリヌクレオチドの任意の領域、フラグメント、または全片が含まれることが理解される。

10

【0038】

コンカテマー - 末端と末端が連結した少なくとも2つのヌクレオチドモノマー配列から構成され、任意選択的に、リンカーまたはスパーサーによって分離している。本発明の目的のために、コンカテマーは、本発明の方法にしたがって調製された少なくとも2つのオリゴヌクレオチドを含む。

20

【0039】

クローン、クローニング - 一方の生物由来のヌクレオチド（遺伝子など）を他方の生物に導入することおよび/または遺伝子操作技術による複製ヌクレオチド。

【0040】

ライブラリー - 通常、1つまたは複数のプラスミド中に含まれる、クローン化した核酸配列、オリゴヌクレオチドまたはポリヌクレオチドの集団。

【0041】

ベクター - バクテリオファージ、プラスミド、または一方の細胞から他方の細胞に遺伝物質を導入する他の因子(agent)。

30

【0042】

得る、由来する - 材料に一定の所望の特徴を付与するために核酸およびタンパク質などの生体材料に対して分子生物学および遺伝子工学ならびに操作技術を使用すること。用語「得る」および「由来する」を、本発明で交換可能に使用することができる。

【0043】

増幅 - 核酸コピー数の増加。一般的に使用されている1つの方法は、ポリメラーゼ連鎖反応(PCR)である。当業者に公知の他の増幅方法も使用することができる。

【0044】

トランスフェクションまたは形質転換 - 外来分子を細胞に移入するための任意の方法。リポフェクション、リン酸カルシウム沈殿、レトロウイルス送達、エレクトロポレーション、および遺伝子銃形質転換は、使用することができる技術のほんの一例である。

40

【0045】

クロマチン - 細胞核において、塩基性色素で容易に染色され、凝縮されて細胞分裂中に染色体を形成する核酸とタンパク質（主に、ヒストン）との複合体。クロマチンは、核酸 - タンパク質複合体の例である。染色体領域は、同一または異なる染色体上のいずれかの他の領域と相互作用することができる。したがって、相互作用事象は、染色体間事象または染色体内事象であってよく、関連領域での遺伝物質の再整列を含み得る。

【0046】

50

トランスフュージョン - 新規のキメラ転写物を形成するRNAプロセッシングレベルでの遺伝情報の再整列。

【0047】

転座 - ゲノムDNAレベルでの遺伝情報の再整列。

【0048】

核酸 - タンパク質複合体 - 遺伝物質とタンパク質（クロマチン中または転写因子が核酸ストレッチに結合した場合に見出されるタンパク質など）との相互作用。DNA - タンパク質 - DNA（DPD）複合体は、タンパク質が目的の核酸（DNA）の2つのストレッチ間に結合するより特異的な構造である。DNAなどの核酸のストレッチを、タグまたは同定可能なDNA配列となるように操作することができる。

10

【0049】

タグ、タグ - リンカー構造 - タグまたはサイン(signature)は、同定可能な核酸配列であり、任意の連続DNA領域由来の5'または3'の最も末端の核酸配列（末端；通常、18~20bp）のいずれかをいうか、タグは、5'または3'の最も末端の核酸配列または任意の連続DNA領域の末端を含むことができる。リンカーは、人工核酸配列である。したがって、タグ - リンカー - タグ構造は、リンカーが2つのタグの間にリンカーが挿入された核酸の整列である。別の可能な整列は、リンカーがタグに隣接した（すなわち、少なくとも1つのタグの上流および/または下流に配置した）リンカー - タグ - タグ - リンカーである。用語「タグ」および「サイン」を、本発明で交換可能に使用することができる。

20

【0050】

ジタグ(ditag) - ポリヌクレオチドの短い（通常、12~60bp）核酸フラグメント由来の末端タグまたはサイン。ジタグを、米国特許出願番号20050255501および/または同第20050059022号（その内容が本明細書中で参照することにより組み込まれる）にしたがって調製することができる。

【0051】

配列決定 - 生体高分子（この場合、核酸）の構成要素の順序を決定するために使用される方法。使用される配列決定技術には、Sanger法およびその修正形態ならびにピロシーケンスまたは配列決定の「454法」が含まれる。

【0052】

以下の説明では、本発明の実施形態を説明するために、詳細、特定の量、およびパラメーターを示す。しかし、このような詳細を用いることなく本発明を実施することができるのが当業者に明らかなはずである。詳細のうちのいくつかは、本発明を曖昧にしないために、長々とは説明していない。

30

【0053】

特定の実施形態のための本発明の方法の実施のために、本発明の他の実施形態を実施する目的のために開示した任意の説明も使用することができ、これは、本明細書中で参照することにより組み込まれる。特に、技術、試薬、実験条件、制限部位、酵素、ベクター、およびプライマーなど。特に、どのようにして他の実施形態のために開示された技術および材料を本発明の実施形態に適用するのかについては、当業者に明らかである。

40

【0054】

当業者は、本明細書中で詳細に教示していない技術を、Molecular Cloning: A Laboratory Manual by Sambrook and Russell, Third Edition, 2001, published by Cold Spring Harbor Laboratory Pressなどの標準的な分子生物学の参考図書で見出すことができることを認識している。

【0055】

[説明]

本発明は、核酸複合体、特に、核酸 - タンパク質複合体由来の少なくとも1つの核酸配列またはフラグメントを検出、同定、および/または調製する新規の方法に関する。特に、本発明の方法は、核酸 - タンパク質複合体由来の少なくとも2つの核酸配列またはフラ

50

グメントを検出、同定、および/または調製する方法を提供する。本発明はまた、オリゴヌクレオチドおよび/またはオリゴヌクレオチドのコンカテマーを提供する。

【0056】

1つの態様によれば、本発明は、クロマチン相互作用分析(CIA)法を提供する。CIAは、*de novo*での遠位調節領域および染色体間相互作用に関する新規の情報を捕捉するためにデザインされている。本方法は、特定のDNA結合タンパク質(ヒストンなど)によって媒介される長距離および異なる染色体間のクロマチン相互作用事象を同定するようにデザインされている。

【0057】

このようなDNA連結点を検出するためのCIA法の2つの実施形態(すなわち、2つのライゲーションした各DNAフラグメントから1つのタグサイン(約20bp)を抽出して「タグ1-リンカー-タグ2」(「第1のタグ-リンカー-第2のタグ」とも呼ばれる)対合末端ジタグ(PET)構造を形成するCIA-PET法(図1)および「PET1-リンカー-PET2」構造(いわゆる、*diPET*)中の2つの関連DNAフラグメントを示すために2つの対合末端ジタグ(PET)を得るCIA-*diPET*法(図2))を本発明で提供する。CIA-PETおよびCIA-*diPET*のタグを、「454」ピロシーケンス法などの多重配列決定技術を使用して直接配列決定することができるか、従来の配列決定法を使用したクローニングおよび配列決定のために鎖状体形成することができる。

10

【0058】

したがって、本発明は、少なくとも1つの単離オリゴヌクレオチドの調製方法であって、

20

(a) 核酸-タンパク質複合体を供する工程と、

(b) 少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含むオリゴヌクレオチドを調製する工程であって、第1のタグおよび第2のタグが核酸-タンパク質複合体から得られる工程と、

(c) オリゴヌクレオチドを単離する工程と、
を含む、方法を提供する。

【0059】

第1のタグを第1のポリヌクレオチドから得ることができ、第2のタグを第2のポリヌクレオチドから得ることができ、第1のポリヌクレオチドおよび第2のポリヌクレオチドは、核酸-タンパク質複合体から得られる。

30

【0060】

工程(b)が、

(i) 少なくとも1つの制限酵素の認識部位を含む少なくとも1つのリンカーを挿入する工程と、

(ii) 第1のポリヌクレオチドおよび第2のポリヌクレオチドを、リンカー中の少なくとも1つの制限部位を認識する少なくとも1つの制限酵素で切断して、第1のポリヌクレオチドから得た第1のタグ、第2のポリヌクレオチドから得た第2のタグ、およびリンカーを含むオリゴヌクレオチドを形成する工程と
を含むことができる。

40

【0061】

特に、工程(b)が、

(i) 少なくとも1つの制限酵素の認識部位を含む少なくとも1つのリンカーを、複合体の第1のポリヌクレオチドと第2のポリヌクレオチドとの間に挿入する工程と、

(ii) 少なくとも1つの制限酵素の各制限部位を第1のポリヌクレオチドの5'末端および第2のポリヌクレオチドの3'末端に付加する工程と、

(iii) 第1のポリヌクレオチドおよび第2のポリヌクレオチドを少なくとも1つの認識部位を認識する少なくとも1つの制限酵素で切断して切断フラグメントを得る工程と、

50

(i v) 切断フラグメントをライゲーションして、第 1 のポリヌクレオチドから得た第 1 のタグ、第 2 のポリヌクレオチドから得た第 2 のタグ、各ポリヌクレオチドの 5 ' 末端および 3 ' 末端を含むタグ、およびタグの間に挿入されたリンカーを含むオリゴヌクレオチドを形成する工程とを含むことができる。

【 0 0 6 2 】

本発明はまた、核酸 - タンパク質複合体由来の少なくとも 2 つのポリヌクレオチドを検出および / または同定する方法であって、

(a) 核酸 - タンパク質複合体を供する工程と、

(b) 少なくとも 1 つの第 1 のタグおよび少なくとも 1 つの第 2 のタグを含むオリゴヌクレオチドを調製する工程であって、第 1 のタグが第 1 のポリヌクレオチドから得られ、第 2 のタグが第 2 のポリヌクレオチドから得られ、第 1 のポリヌクレオチドおよび第 2 のポリヌクレオチドが核酸 - タンパク質複合体から得られる工程と、

(c) オリゴヌクレオチドを配列決定する工程と、

(d) 配列に基づいて少なくとも 2 つのポリヌクレオチドをマッピングし、それにより、少なくとも 2 つのポリヌクレオチドを検出および / または同定する工程とを含む方法を提供する。

【 0 0 6 3 】

リンカー配列は、各末端に 2 つの I I 型制限酵素の認識部位を保有する (図 2 および 4) 。それにより、I I 型制限消化後に、各タグ配列サインの各々 (約 2 0 b p) を、2 つの連結 DNA フラグメントから切り出してタグ - リンカー - タグ構造 (この構造は、一方のタグは染色体の 1 つの DNA 領域を示す一方で、他のタグは同一染色体上の離れた領域の遺伝子座または異なる染色体中の遺伝子座を示す。) を形成することができる。この対合末端ジタグ構造を、「 C I A - P E T 」といい、より長い DNA ストレッチへの鎖状体形成 (concatenation) によって有効に配列決定することができるか、配列決定によって直接分析することができる。

【 0 0 6 4 】

あるいは、リンカー配列を 2 つのタグに隣接させて、リンカー - タグ - タグ - リンカー構造を得ることができる。

【 0 0 6 5 】

この方法では、核酸 - タンパク質複合体 (未変性の DNA - タンパク質 - DNA (D P D) 複合体など) を、ホルムアルデヒド、グルタルアルデヒド、またはメタノールなどの適切な固定液によって架橋する。次いで、架橋 D P D 複合体を、超音波処理、ハイドロシャーリング (hydroshearing) (Hydroshear, Gene Machines) 、皮下注射針による引き抜きの反復 (repeated drawing) 、または制限酵素消化によって断片化することができる。異なる染色体に由来するか、距離の長い DNA フラグメントを、D P D 複合体中の DNA 結合タンパク質によって繫留する。これらの DNA フラグメントのうちのどの DNA フラグメントが P C R プライマーを生成するのかについての既存の知識または推移を必要とする 3 C 技術と異なり、D P D 複合体中に結合した遠位関係の DNA フラグメントの末端を、ライゲーションによって特定のリンカーで連結する。

【 0 0 6 6 】

リンカー (約 2 0 b p) は、各 D P D 複合体中のタンパク質によって繫留された異なる DNA フラグメントの末端を連結するための 2 つの I I 型制限酵素の認識部位を含む。次いで、2 つの関連する DNA フラグメントの DNA 連結点を、対合末端ジタグ化 (P E T) ストラテジー (米国特許出願番号 20050255501 号) によってタグ化することができる。I I s 型制限酵素の認識部位が好ましい。I I 型制限酵素に加えて、任意の他の適切な制限酵素 (I I I 型またはホーミング制限酵素が含まれる) を使用することができる。

【 0 0 6 7 】

したがって、本発明は、1 つの態様では、特定の DNA 結合タンパク質 (ヒストンなど) によって媒介される長距離および異なる染色体間のクロマチン相互作用事象を同定する

10

20

30

40

50

方法を提供する。別の態様では、本発明は、少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含む単離オリゴヌクレオチドであって、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体から得られる、単離オリゴヌクレオチドを提供する。タグは、核酸-タンパク質複合体中のクロマチン領域に対応する。これらのタグを、クロマチン相互作用事象の分析、同定、および/または検出のために配列決定することができる(図3および5)。

【0068】

単離オリゴヌクレオチドは、少なくとも1つのリンカーをさらに含むことができる。リンカーをタグの間に挿入することができ、また、リンカーを少なくとも1つのタグの上流および/または下流に配置することができる。リンカーは、少なくとも1つの制限酵素の認識部位を含むことができ、少なくとも1つの制限酵素の認識部位は非対称であってよく、少なくとも1つの制限酵素の認識部位はIIs型制限酵素またはホーミング制限酵素の認識部位であり得る。

10

【0069】

あるいは、少なくとも1つの第1のタグは第1のポリヌクレオチド由来の5'末端および3'末端を含むことができ、少なくとも1つの第2のタグは第2のポリヌクレオチド由来の5'末端および3'末端を含むことができる。単離オリゴヌクレオチドは、少なくとも1つのリンカーをさらに含むことができる。リンカーをタグの間に挿入することができ、また、リンカーを少なくとも1つのタグの上流および/または下流に配置することができる。リンカーは、少なくとも1つの制限酵素の認識部位を含むことができ、少なくとも1つの制限酵素の認識部位は非対称であってよく、少なくとも1つの制限酵素の認識部位はIIs型制限酵素またはホーミング制限酵素の認識部位であり得る。

20

【0070】

リンカーは、第1のポリヌクレオチドを切断して第1のタグを得ることができる制限酵素によって認識される第1の制限認識部位、および第2のポリヌクレオチドを切断して第2のタグを得ることができる制限酵素によって認識される第2の制限認識部位を含むことができる(図1および4)。

【0071】

リンカーは、第1のポリヌクレオチドを切断して第1のポリヌクレオチドの3'末端を得ることができる制限酵素によって認識される第1の制限認識部位、および第2のポリヌクレオチドを切断して第2のポリヌクレオチドの5'末端を得ることができる第2の制限酵素によって認識される第2の制限認識部位を含むことができる。あるいは、リンカーは、第1のポリヌクレオチドを切断して第1のポリヌクレオチドの5'末端を得ることができる制限酵素によって認識される第1の制限認識部位、および第2のポリヌクレオチドを切断して第2のポリヌクレオチドの3'末端を得ることができる第2の制限酵素によって認識される第2の制限認識部位を含むことができる。

30

【0072】

第1のポリヌクレオチドを、第3の認識部位を認識する第3の制限酵素によってさらに切断して第1のポリヌクレオチドの3'末端を得ることができ、第2のポリヌクレオチドを、第4の認識部位を認識する第4の制限酵素によってさらに切断して第2のポリヌクレオチドの5'末端を得ることができ、少なくとも1つの第1のタグは、第1のポリヌクレオチドの5'末端および3'末端のライゲーションによって得られ、少なくとも1つの第2のタグは、第2のポリヌクレオチドの5'末端および3'末端のライゲーションによって得られる(図2および6)。第1の認識部位および第3の認識部位は同一の配列を有することができ、第2の認識部位および第4の認識部位は同一の配列を有することができる。

40

【0073】

単離オリゴヌクレオチドの核酸-タンパク質複合体は、クロマチン構造の一部であり得る。目的のタンパク質が結合する核酸フラグメントは、目的のタンパク質が、例えば、ヒ

50

ストン結合部位に結合する領域を含む任意の核酸フラグメントであり得る。ポリヌクレオチドは、DNAまたはRNAであり得る。

【0074】

得られたCIA-PET配列を、クロマチン相互作用連結点を局在化するために基準ゲノム配列にマッピングすることができ、連結点は、長距離の染色体内（同一染色体上）または染色体間（異なる染色体上）のいずれかであり得る。この情報を、核内の染色体三次元組織化構造、転写因子によって媒介される長距離および異なる染色体における遺伝子調節の調和、および他の後成的遺伝学に関する問題の研究のために使用することができる（図3および5）。

【0075】

完全なゲノム配列を利用可能であるので、全てのクロマチン相互作用の可能性を同定するための全ゲノムアプローチを開発することが可能である。全ゲノムタイリング(tiling)アレイは、ゲノム調査のための魅力的なアプローチであり、これは、マイクロアレイ中の全ゲノムを対象とするために20~60量体のオリゴヌクレオチドをタイリングし、生体内容物を含むDNAプローブをアレイとハイブリッド形成し、全アレイエレメントのハイブリッド形成シグナル強度のプロフィールが遺伝因子調査のための読み出し情報である。タイリングアレイは、エクソンの同定およびChIP(ChIP-chip)と組み合わせた場合の転写因子結合部位の局在化に有用であることが証明されている。アレイベースのアプローチは、その高度な複雑さおよび並列性(parallel nature)によって有効であるけれど、ハイブリッド形成ベースの検出がクロマチン相互作用における2つのDNAフラグメントの非直線関係を検出することができるとは考えられない。

【0076】

[1つの実施形態のオリゴヌクレオチドの調製]

核酸-タンパク質複合体を、クロマチン免疫沈降(ChIP)によって得ることができる。

【0077】

[クロマチン免疫沈降(ChIP)]

ChIPを使用して富化し、それにより、ヒストンおよび核酸-タンパク質複合体中の核酸に結合することができる他のタンパク質などの特定のタンパク質と会合するゲノム領域を同定可能である(Taverman et al., Genome Biol. 2004.5(3):p.210に概説)。タンパク質をその相互作用部位でDNAと架橋することを目的とする。培養物中の生きた細胞へのホルムアルデヒド、グルタルアルデヒド、またはメタノールなどの適切な固定液の直接添加によってこれを迅速且つ有効に行う。

【0078】

次いで、これらの固定細胞の粗抽出物を調製し、超音波処理、ハイドロシャーリング、皮下注射針による引き抜きの反復、または制限酵素消化によって、クロマチンを、通常、約1kbの平均サイズに切断し、次いで、目的のDNA会合タンパク質（例えば、転写因子またはヒストン）に対して惹起する抗体での免疫沈降反応で使用する。次いで、各免疫沈降で富化したDNAフラグメントを脱連結し、精製し、種々の方法によって同定することが可能である。ChIP使用の利点は、このアプローチがクロマチンおよび他のヒストンタンパク質の迅速な架橋によってインビボ遺伝子調節ネットワークを「フリーズ」し、それにより、例えば、異種発現によって課される人工産物を含まない、調節系の「真の」姿(picture)をいつでも示す理論を得ることができることである。

【0079】

最近、ChIPは、全ゲノム(Lieb et al., Nat Genet, 2001, 28(4):p.327-34)、全染色体(Euskirchen et al., Mol Cell Biol. 2004. 24(9):p.3804-14)、およびCpG島(Weinmann et al., Genes Dev, 2002. 16(2):p.235-44)、または転写因子結合部位(TFBS)などのタンパク質結合部位のゲノムレベルで局在化することができる見込みがある「ChIP-chip」または「ChIP-on-chip」アプローチ(Buck and Lieb, 2004に概説)におけるマイクロアレイと組み合わされている。このアプローチの実用性は酵母などの

10

20

30

40

50

小さなゲノムで証明されている一方で(Lieb et al., Nat Genet, 2001, 28(4):p.327-34)、より複雑な生物の全ゲノムマイクロアレイの産生の費用および複雑さが依然として制限要因である。

【0080】

CpG島マイクロアレイは、CpG含有量が高いヒトゲノムフラグメントを含む。CpG島はしばしばプロモーター領域に対応するので(Antequera and Bird, Proc Natl Acad Sci USA, 1993. 90(24):p.11995-9)、このようなマイクロアレイは妥協案となりえる。しかし、推定タンパク質結合部位の位置は、依然として、アレイ上にスポットされたCpGリッチプローブのゲノムDNAの上流および下流(これが超音波処理されたChIPフラグメントのおおよそのサイズであるので、通常、1~2kb)の試験によって間接的に推測しなければならない。

10

【0081】

代替法として、ChIP富化DNAフラグメントのクローニングおよび配列決定が以前に試みられてきたが、少しの成功しかおさめられていない。全ゲノムの高バックグラウンドに対してChIP富化の標的が得られることが問題である。特異的標的が100倍富化されている場合でさえも、依然としてChIPライブラリー中に小フラクシオンのクローンしか認められず、このことが標準的なDNA配列決定を非常に高額な解決法にしている。したがって、これらの環境下でのChIPクローンの配列決定は、富化標的の同定のための良好なアプローチではない。遺伝子発の連続分析(Serial Analysis of Gene Expression)(SAGE)および大量並行サイン配列決定(Massively Parallel Signature Sequencing(MPSS))(Brenner et al., Nat Biotechnol, 2000. 18(6):p.630-4)により、ChIP富化の検出のための有用な定量ツールとしても示唆されており、基本原理は、ChIP富化DNAフラグメントから生成されたタグが非特異的バックグラウンドと比較して多数存在することである。

20

【0082】

次いで、これらのタグを、目的の一般的領域の同定のために、ゲノム配列にマッピングすることができる(すなわち、1~2kb(超音波処理フラグメントを示す))。20bpのSAGEタグおよびMPSSタグは、ほとんどの場合、特定のゲノムの位置を定義するのに十分に特異的なはずであるにもかかわらず、ゲノムにマッピングする場合、タグの約1~2kb上流および下流の全配列を依然として試験しなければならない。これは、CpG島マイクロアレイアプローチが直面する問題と同一である。さらに、これらの方法を使用して完全に対象とするためには、必須の制限酵素の認識部位(マッピング-酵素部位)の利用可能性に依存し、認識部位が一定のゲノムの位置に存在しない場合、特定のタグが対応するChIPフラグメントから喪失しているので、この位置は、ゲノム内の「盲点」である。

30

【0083】

上記の問題から、ゲノムレベルの転写調節分析を容易にするために必要なのは、全ゲノムアレイの代替法として、タンパク質結合領域に隣接する核酸配列を正確且つ迅速に特定する方法であることが明らかである。これに関して、本発明によって提供された新規のアプローチは、以下のいくつかの利点を有する。(i)各タグが5'および3'サインに含まれる連続DNAセグメントに由来することが既に公知であるので、本発明の1つの方法によって生成されたタグ配列がより高いマッピング特異性を示す。この情報によって目的のゲノム領域の正確な局在化が容易になり、標準的なSAGEタグまたはMPSSタグの上流および下流の配列をおよそ1~2kb毎に繰り返し試験する必要性が回避される。(ii)したがって、本発明の方法は、マッピング-酵素部位の存在のためのいかなる要件とも無関係である。(iii)配列決定前のタグの鎖状体形成は、1回の配列決定の読み取りで数個のタグを同定することができることを意味する。(iv)したがって、このクラスター中のマッピングされた全タグに共通の(すなわち、重複した)領域により、問題の核酸-タンパク質複合体に関与するDNA領域が定義される。

40

【0084】

50

[タグおよびジタグ]

本出願の目的のために、タグは、核酸分子から得られたヌクレオチド配列またはサインであり、タグが得られるか由来するポリヌクレオチドに相当する。収縮するか相当することが意図されるポリヌクレオチドは、RNA、mRNA、ゲノムDNA、全長cDNA、またはcDNAであり得る。

【 0085 】

本発明では、本発明のオリゴヌクレオチド中に存在する2つのタグを、それぞれ、ジタグと呼ぶこともできる。タグと同様に、ジタグは、由来するか相当する元の核酸分子よりも短い。好ましくは、ジタグは、元の核酸分子よりも短くなければならない。「収縮」の結果として、ジタグは、本質的に、元の核酸分子の5'末端領域(5'タグとしても示される)および3'末端領域(3'タグとしても示される)のいずれかまたは両方を含むことができる。したがって、最初は5'タグおよび3'タグの間または内側に存在する元の核酸分子の一部は、ジタグ中に含まれない。本発明のジタグは、元の核酸分子のほとんどの有益な特徴(すなわち、核酸の出発サインおよび終結サイン)を保持している。

10

【 0086 】

ジタグを形成する5'タグおよび3'タグは、同一または異なるサイズを有することができる。好ましくは、これらは、同数のヌクレオチドを有する。ジタグは、任意のサイズであってよいが、由来する親配列のサイズより有意義且つ有利である必要がある。タグまたはジタグの好ましいサイズは、ゲノムの複雑さによって決定される。細菌ゲノムについては、約8bp~約16bpのタグが十分であり得るのに対して、ヒトゲノムのような複雑なゲノムについては、16~20bpのタグ(言い換えれば、32~40bpのジタグ)と見なすことができる。一般に、ジタグのサイズは、約12~60bpである。

20

【 0087 】

本出願の目的のために、用語「5'末端(terminus)」、「5'末端(end)」、および「5'タグ」は、互いに等価であり、交換可能に使用することができる。同様に、用語「3'末端(terminus)」、「3'末端(end)」、および「3'タグ」は、互いに等価であり、交換可能に使用することができる。減少または相当することを意図する核酸分子内の元の核酸分子または一部では、5'末端および3'末端は、それぞれ、先端に最も近い領域または部分および分子の中央領域から最も遠い領域または部分に相当する。

【 0088 】

本発明の1つの態様によれば、ジタグ中に含まれる5'タグおよび3'タグは、それぞれ、減少するか相当することが意図される核酸分子またはその一部の5'末端および3'末端に最も近い制限酵素によって切断された分子領域である。したがって、ジタグのサイズを、使用する制限酵素によって決定することができる。

30

【 0089 】

したがって、本発明は、少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含む単離オリゴヌクレオチドを含む単離オリゴヌクレオチドであって、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体から得られる、単離オリゴヌクレオチドを提供する。オリゴヌクレオチドは、タグの間に挿入された少なくとも1つのリンカーをさらに含むことができる。オリゴヌクレオチドはまた、タグに隣接させるために挿入された少なくとも1つのリンカーをさらに含むことができる(すなわち、少なくとも1つのリンカーを、少なくとも1つのタグの上流および/または下流に配置することができる)。

40

【 0090 】

[リンカー]

特に、各リンカーは、少なくとも1つの第1の制限部位および少なくとも1つの第2の隣接制限部位を少なくとも含むことができる。したがって、各リンカー中に存在する制限部位数は、1つまたは複数、好ましくは、2つであり得る。制限部位は、非対称の制限部位であり得る。非対称の制限部位の例は、ホーミングエンドヌクレアーゼ非対称認識部位

50

およびいくつかのⅡⅡ型(またはクラスⅡⅡ)認識部位である。その認識部位の片側を切断するⅡⅡs型制限酵素が好ましい。

【0091】

しかし、当分野で公知の任意の認識部位を使用することができる。核酸分子内の少なくとも1つの認識部位を認識し、且つ使用することができる制限酵素は、当業者に明らかである(例えば、Current Protocols in Molecular Biology, Vol. 2, 1995, Ed. Ausubel, et al. Greene Publish. Assoc. & Wiley Interscience, Unit 3.1.15; New England Biolabs Catalog, 2005を参照のこと)。可能な制限部位およびこれらを認識する対応制限酵素のリストを以下に報告する。

【0092】

例として、制限部位を認識する制限酵素を、本発明のジタグの調製のために使用することができる。特に、ⅡⅡs型酵素(例えば、MmeI)。MmeIを使用する場合、この酵素は、減少することが意図される核酸分子に隣接する2つのアダプターのそれぞれの内側の配列を認識するが、核酸分子の内側を切断して17~21ヌクレオチドを含むタグを形成する(図1および4中に示すMmeI切断物を参照のこと)。2つのこのようなタグを、平滑末端形成およびライゲーションによってさらに処理して、34~38ヌクレオチドを含むジタグを形成することができる。したがって、互いにスプライシングするか同一の核酸分子の5'末端および3'末端をライゲーションすることによってジタグが得られる。

【0093】

例として、非対称部位を移入することができる。本発明の目的に有用な非対称部位配列は、以下である。i) 2つのホーミングエンドヌクレアーゼ非対称認識部位の配列、またはii) ⅡⅡ型制限酵素によって認識される制限エンドヌクレアーゼ非対称切断部位の配列。

【0094】

ホーミングエンドヌクレアーゼは、市販されており、New England Biolabs, Inc.によって説明されており、非対称部位配列の説明を、New England Biolabs Catalogから利用することもできる。これらのホーミングエンドヌクレアーゼ非対称認識部位の配列は、18~39bpである。しかし、本発明では、認識部位配列は、これらの配列やこれらのサイズに限定されない。好ましくは、非対称部位配列を切断することができる制限ホーミング
30
エンドヌクレアーゼは、以下からなる群から選択される。I-CeuI、PI-SceI、PI-PspI、およびI-SceI。しかし、上記リストは、完全ではない。当分野で公知である他のホーミングエンドヌクレアーゼおよびその後に発見することができる他のホーミングエンドヌクレアーゼは、本発明の範囲に含まれる。

【0095】

ⅡⅡ型制限酵素の例には、以下が含まれる。AarI、AceII、AloI、BaeI、Bbr7I、BbvI、BbvII、BccI、Bce83I、BceAI、BcefiI、BcgI、BciVI、BfiI、BinI、BpII、BsaXI、BscAI、BseMI、BseRI、BsgI、BsmI、BsmAI、BsmFI、Bsp24I、BspCNI、BspMI、BsrI、BsrDI、BstF5I、BtgZI、
40
BtsI、CjeI、CjePI、EciI、Eco31I、Eco57I、Eco57MI、Esp3I、FaiI、FauI、FokI、GsuI、HaeIV、HgaI、Hin4I、HphI、HpyAV、Ksp632I、MboII、MlyI、MmeI、MnII、PleI、PpiI、PsrI、RleAI、SapI、SfaNI、SspD5I、Sth132I、StsI、TaqII、TspDTI、TspGWI、TspRI、およびTth111II(Rebase Enzymes(登録商標)のウェブサイト中のリスト：<http://rebase.neb.com/cgi-bin/outsidelist>; Szybalski, W., 1985, Gene, 40:169も参照のこと)。しかし、上記リストは、完全ではない。当分野で公知である他のⅡⅡ型酵素およびその後に発見することができる他のⅡⅡ型酵素は、本発明の範囲に含まれる。

【0096】

10

20

30

40

50

いくつかのクラス I I 制限酵素の認識部位および切断部位の例は、B b v I (GCAGC 8 / 12)、H g a I (GACGC 5 / 10)、B s m F I (GGGAC 10 / 14)、S f a N I (GCATC 5 / 9)、および B s p I (ACCTGC 4 / 8)である(括弧内は認識部位および切断部位である)。

【0097】

人工制限エンドヌクレアーゼも使用することができる。これらのエンドヌクレアーゼを、タンパク質工学によって調製することができる。例えば、エンドヌクレアーゼ F o k I は、挿入によって D N A 基質の両鎖上の認識部位からさらに離れた1つのヌクレオチドを切断するように操作されている。Li and Chandrasegaran. Proc. Nat. Acad. Sciences USA 90:2764-8, 1993を参照のこと。このような技術を適用して、所望の認識配列および認識部

10

【0098】

本発明では、本発明の単離オリゴヌクレオチドを、他の単離オリゴヌクレオチドと連結または鎖状体形成して、オリゴヌクレオチドのコンカテマーを形成することができる。配列決定または適切なプラスミドまたはベクターへのクローニングのためにかなり多数のオリゴヌクレオチドを互いに連結することができる。

【0099】

したがって、別の態様では、本発明は、少なくとも2つの単離オリゴヌクレオチドを含むオリゴヌクレオチドのコンカテマーであって、各単離オリゴヌクレオチドが少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含み、ここで、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体から得られる、コンカテマーである。

20

【0100】

オリゴヌクレオチドのコンカテマーの単離オリゴヌクレオチドは、少なくとも1つのリンカーをさらに含むことができる。リンカーを、タグの間に挿入することができる。あるいは、リンカーをタグに隣接させることができる。リンカーは、少なくとも1つの制限酵素の認識部位を含むことができ、少なくとも1つの制限酵素の認識部位は、I I s 型制限酵素の認識部位であり得る。

30

【0101】

1つの実施形態では、オリゴヌクレオチドのコンカテマーの第1のタグは、第1のポリヌクレオチド由来の3'末端を含むことができ、第2のタグは、第2のポリヌクレオチド由来の5'末端をさらに含む。鎖状体形成したポリヌクレオチドの単離オリゴヌクレオチドは、それぞれ、タグの間に挿入された少なくとも1つのリンカーをさらに含むことができる。別の実施形態では、オリゴヌクレオチドのコンカテマーの第1のタグは、第1のポリヌクレオチド由来の5'末端を含むことができ、第2のタグは、第2のポリヌクレオチド由来の3'末端をさらに含む。

【0102】

別の実施形態では、オリゴヌクレオチドのコンカテマーの第1のタグは、第1のポリヌクレオチド由来の5'末端および3'末端を含むことができ、第2のタグは、第2のポリヌクレオチド由来の5'末端および3'末端をさらに含む。オリゴヌクレオチドのコンカテマーの単離オリゴヌクレオチドは、それぞれ、タグの間に挿入された少なくとも1つのリンカーをさらに含むことができる。

40

【0103】

これらの実施形態のいずれかのリンカーは、少なくとも1つの制限酵素の認識部位を含むことができ、少なくとも1つの制限酵素の認識部位は、I I s 型制限酵素の認識部位であり得る。

【0104】

オリゴヌクレオチドのコンカテマーを、ベクターまたは細胞に挿入ことができ、細

50

胞は細菌細胞であり得る。

【0105】

オリゴヌクレオチドのコンカテマーは、クロマチン構造に由来し得る。ポリヌクレオチドは、同一染色体上に存在するか、ポリヌクレオチドは、異なる染色体上に存在する。

【0106】

これらは好ましいコンカテマーであるが、鎖状体形成することができる本発明のオリゴヌクレオチド数がオリゴヌクレオチドの長さに依存し、当業者が過度に実験することなく容易に決定することができることが明らかである。コンカテマーの形成後、配列決定のために複数のタグをベクターにクローン化することができるか、ジタグまたはコンカテマーを当業者に公知の方法によってクローニングすることなく直接配列決定することができる。したがって、ジタグの鎖状体形成により、1つのベクターまたはクローン内の複数のジタグの配列決定による連続的様式において全長cDNAのような核酸分子を有効に分析可能である。

10

【0107】

用語「ベクター」または「組換えベクター」は、ジタグ遺伝子配列の挿入または組み込みによって操作されたプラスミド、ウイルス、または当分野で公知の他の送達体(vehicle)も使用することができることが意図される。このようなベクターは、有効な転写を容易にするプロモーター配列を含む。ベクターは、典型的には、複製起点、プロモーター、および形質転換細胞を表現型で選択することが可能な特定の遺伝子を含む。本発明での使用に適切なベクターには、例えば、pBlueScript(Stratagene, La Jolla, CA)、pBC、pZErO-1(Invitrogen, Carlsbad, CA)、およびpGEM3z(Promega, Madison, WI)、またはこれらの修飾ベクター、ならびに当業者に公知の他の類似のベクターが含まれる。特定の用途(realization)として、pGEM3zベクターを修飾し、これを、pGIS8という(図7および11)。pGEMベクターは、米国特許第4,766,072号(本明細書中で参照することにより組み込まれる)にも開示されている。

20

【0108】

タグまたはジタグが由来する親ポリヌクレオチドまたは核酸分子の産生のために、全長ライブラリーとして、適切なベクターを使用することができる。したがって、本発明の範囲内である適切なベクターは、ベクター骨格がポリヌクレオチドの挿入後にポリヌクレオチドまたはジタグに隣接するアダプター中に含まれる同一の制限部位を含まないベクターである。好ましくは、本発明は、ベクター骨格(ポリヌクレオチドの挿入時に除去される多クローニング部位を含むスタッファー(stuffer)領域内以外)は、アダプターに含まれる第1の制限部位および第2またはさらなる制限部位を含まないベクターを提供する。特に、ベクターは、アダプター中に含まれる少なくともII制限部位(例えば、IIs型制限部位)および少なくとも第2またはさらなる制限部位を含まない。より好ましくは、ベクター骨格(ポリヌクレオチドの挿入時に除去されるスタッファー領域内以外)は、MmeIおよびBamHIを含まない。

30

【0109】

スタッファーの外側の任意の領域中にMmeIを含まないこのようなベクターの例は、図7および11に示すベクターpGIS8である。pGIS8では、MmeI認識部位を、変異誘発によって欠失した。

40

【0110】

したがって、本発明は、少なくとも1つのオリゴヌクレオチドを含むオリゴヌクレオチドのライブラリーであって、オリゴヌクレオチドが少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含み、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体から得られる、ライブラリーを提供する。本発明のオリゴヌクレオチドのコンカテマーのライブラリーも提供する。

【0111】

オリゴヌクレオチドライブラリーの少なくとも1つのオリゴヌクレオチドは、タグの間

50

に挿入された少なくとも1つのリンカーを含むことができる。あるいは、リンカーは、その代わりにタグに隣接することができる。第1のタグは、第1のポリヌクレオチド由来の5'末端および3'末端を含むことができ、第2のタグは、第2のポリヌクレオチド由来の5'末端および3'末端をさらに含む。

【0112】

ポリヌクレオチド核酸-タンパク質複合体は、クロマチン構造の一部であり得る。ポリヌクレオチドは、同一染色体上に存在することができるか、ポリヌクレオチドは、異なる染色体上に存在することができる。

【0113】

1つの態様によれば、オリゴヌクレオチドを増幅する。例えば、PCRまたは任意の他の公知の増幅法の使用による。PCRプライマーおよびプローブ配列を、ジタグの配列情報に基づいて調製することができる。したがって、ベクター内の特定の領域に対応する適切なPCRプライマーを使用する。このような領域は、ジタグおよびアダプターを含むオリゴヌクレオチドに隣接する。PCRを、ライゲーション(自己環状化)反応によって直接行い、短い(例えば、200bp)PCR産物を得ることができる。

【0114】

次いで、必要なジタグを含むこれらのPCR産物を、少なくとも第2の制限部位(アダプター内)を認識する酵素で切断して、必要な短い付着ジタグを生成することができる。第2の制限部位またはさらなる制限部位を認識する制限酵素として、例えば、BamHIを使用することができ、50bpの付着ジタグが生成される。この増幅工程の利点は、生成されたジタグがジタグライブラリー増幅を必要とせず、自己環状化タグ化プラスミドを形質転換しないことによって回避することができることである。次に、増幅オリゴヌクレオチドを、ベクターから切り出し(この例では、BamHIでの消化による)、次のクローニングおよび配列決定分析のために長いDNAまたはRNAストレッチ中に鎖状体形成することができる(図1および2)。

【0115】

特定の態様として、本発明は、オリゴヌクレオチドが少なくとも1つのジタグを含み、ジタグが約34~38ヌクレオチドを含み、且つ全長cDNAまたはそのフラグメントの5'末端由来のヌクレオチドおよび3'末端由来のヌクレオチドのスプライシングによって得られる、cDNAライブラリーを開示する。

【0116】

本発明のジタグライブラリーは、元の核酸分子を含むライブラリーの代表である。例えば、核酸分子を含むライブラリーが全長ポリヌクレオチドライブラリーである場合、ジタグライブラリーは、全長ジタグライブラリーの代表である。各ジタグクローンは、特定の全長クローンを特徴づける十分な情報を含む。本発明のより重要なジタグは、核酸-タンパク質複合体由来の元の全長ポリヌクレオチドの5'末端および3'末端を含む。したがって、ジタグは、ポリヌクレオチド構造の代表である。

【0117】

したがって、ジタグライブラリーのジタグクローンを配列決定および分析することで十分である。目的のジタグが見出された場合、対応する全長ポリヌクレオチドを選択し、例えば、PCRによって全長ポリヌクレオチドライブラリーから調製するか、逆転写ポリメラーゼ連鎖反応(RT-PCR)によって標的RNAサンプルから直接調製することができる。

【0118】

本発明のオリゴヌクレオチドを、「454」配列決定(ピロシーケンス)法(Margulies et al, 2005)によって配列決定するか、クローニングのためにより長いDNAストレッチに鎖状体形成してCIA-PEIライブラリーを作製し、その後、Sangerキャピラリー法を使用して配列決定することができる。

【0119】

したがって、本発明は、本発明のオリゴヌクレオチドを調製し、核酸-タンパク質複合

10

20

30

40

50

体由来の少なくとも2つのポリヌクレオチドを検出および/または同定し、本発明のオリゴヌクレオチドおよび鎖状体形成したオリゴヌクレオチドを含むベクターを調製する方法を提供する。

【0120】

したがって、別の態様では、本発明は、少なくとも1つの単離オリゴヌクレオチドの調製方法であって、

(a) 核酸-タンパク質複合体を供する工程と、

(b) 少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含むオリゴヌクレオチドを調製する工程であって、第1のタグおよび第2のタグが核酸-タンパク質複合体から得られる工程と、

10

(c) オリゴヌクレオチドを単離する工程と、
を含む、方法を提供する。

【0121】

特に、少なくとも1つの単離オリゴヌクレオチドの調製方法であって、

(a) 核酸-タンパク質複合体を供する工程と、

(b) 少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含むオリゴヌクレオチドを調製する工程であって、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸-タンパク質複合体から得られる工程と、

20

(c) オリゴヌクレオチドを単離する工程と、
を含む、方法を提供する。

【0122】

1つの実施形態では、本発明のこの態様の工程(b)が、

(i) 少なくとも1つの制限酵素の認識部位を含む少なくとも1つのリンカーを挿入する工程と、

(ii) 第1のポリヌクレオチドおよび第2のポリヌクレオチドを、リンカー中の少なくとも1つの制限部位を認識する少なくとも1つの制限酵素で切断して、第1のポリヌクレオチドから得た第1のタグ、第2のポリヌクレオチドから得た第2のタグ、およびタグの間にリンカーを含むオリゴヌクレオチドを形成する工程と
を含む。

30

【0123】

別の実施形態では、本発明のこの態様の工程(b)が、

(i) 少なくとも1つの制限酵素の認識部位を含む少なくとも1つのリンカーを、複合体の第1のポリヌクレオチドと第2のポリヌクレオチドとの間に挿入する工程と、

(ii) 少なくとも1つの制限酵素の各制限部位を第1のポリヌクレオチドの5'末端および第2のポリヌクレオチドの3'末端に付加する工程と、

(iii) 第1のポリヌクレオチドおよび第2のポリヌクレオチドを少なくとも1つの認識部位を認識する少なくとも1つの制限酵素で切断して切断フラグメントを得る工程と、

(iv) 切断フラグメントをライゲーションして、第1のポリヌクレオチドから得た第1のタグ、第2のポリヌクレオチドから得た第2のタグ、各ポリヌクレオチドの5'末端および3'末端を含むタグ、およびタグの間に挿入されたリンカーを含むオリゴヌクレオチドを形成する工程であって、工程(ii)の少なくとも1つの制限酵素の認識部位がベクターの一部である工程と
を含む。

40

【0124】

核酸-タンパク質複合体を、クロマチン免疫沈降によって得ることができる。

【0125】

別の態様では、本発明は、核酸-タンパク質複合体由来の少なくとも2つのポリヌクレオチドを検出および/または同定する方法であって、

50

(a) 核酸 - タンパク質複合体を供する工程と、

(b) 少なくとも1つの第1のタグおよび少なくとも1つの第2のタグを含むオリゴヌクレオチドを調製する工程であって、第1のタグが第1のポリヌクレオチドから得られ、第2のタグが第2のポリヌクレオチドから得られ、第1のポリヌクレオチドおよび第2のポリヌクレオチドが核酸 - タンパク質複合体から得られる工程と、

(c) オリゴヌクレオチドを配列決定する工程と、

(d) 第1のタグおよび第2のタグのヌクレオチド配列に基づいて少なくとも2つのポリヌクレオチドをマッピングし、それにより、少なくとも2つのポリヌクレオチドを検出および/または同定する工程と

を含む方法を提供する。

10

【0126】

1つの実施形態では、本発明のこの態様の工程(b)が、

(i) 少なくとも1つの制限酵素の認識部位を含む少なくとも1つのリンカーを挿入する工程と、

(ii) 第1のポリヌクレオチドおよび第2のポリヌクレオチドを、リンカー中の少なくとも1つの制限部位を認識する少なくとも1つの制限酵素で切断して、第1のポリヌクレオチドから得た第1のタグ、第2のポリヌクレオチドから得た第2のタグ、およびタグの間にリンカーを含むオリゴヌクレオチドを形成する工程とを含む。

【0127】

20

別の実施形態では、本発明のこの態様の工程(b)が、

(i) 少なくとも1つの制限酵素の認識部位を含む少なくとも1つのリンカーを、複合体の第1のポリヌクレオチドと第2のポリヌクレオチドとの間に挿入する工程と、

(ii) 少なくとも1つの制限酵素の各制限部位を第1のポリヌクレオチドの5'末端および第2のポリヌクレオチドの3'末端に付加する工程と、

(iii) 第1のポリヌクレオチドおよび第2のポリヌクレオチドを少なくとも1つの認識部位を認識する少なくとも1つの制限酵素で切断して切断フラグメントを得る工程と、

(iv) 切断フラグメントをライゲーションして、第1のポリヌクレオチドから得た第1のタグ、第2のポリヌクレオチドから得た第2のタグ、各ポリヌクレオチドの5'末端および3'末端を含むタグ、およびタグの間に挿入されたリンカーを含むオリゴヌクレオチドを形成する工程であって、工程(ii)の少なくとも1つの制限酵素の認識部位がアダプターの一部またはベクターの一部である工程とを含む。

30

【0128】

本発明において、核酸 - タンパク質複合体を、クロマチン免疫沈降によって得ることができる。核酸 - タンパク質複合体を、目的のDNAおよび/またはタンパク質への光活性化可能な部分の組み込み、ならび抗体媒介沈降または親和性媒介技術によるDNA/タンパク質複合体の単離によって得ることができる。このような親和性ベースの技術の例には、ストレプトアビジン/ビオチン、グルタチオン-S-トランスフェラーゼ/グルタチオンマトリクス、マルトース結合タンパク質/アミロースマトリクスの相互作用が含まれる。

40

【0129】

工程(b)で得られたオリゴヌクレオチドを、工程(c)での配列決定する前に増幅することができ、ポリメラーゼ連鎖反応によって増幅することができる。増幅したオリゴヌクレオチドを、増幅後であるが工程(c)で配列決定する前に少なくとも1つの精製工程に供することができる。少なくとも1つの精製工程は、ゲル電気泳動であり得る。

【0130】

本発明のオリゴヌクレオチドを、配列決定前に工程(a)~(b)によって得た少なくとも1つの他のオリゴヌクレオチドと鎖状体形成し、その後にオリゴヌクレオチドのコン

50

カテマーを作製するために配列決定することができる。

【0131】

Sanger法または多重配列決定によって配列決定することができる。多重配列決定は、ピロシーケンスであり得る。Bonetta(2006)によって記載された方法などの任意の適切な配列決定方法を使用することができる。目的のタンパク質が結合する核酸フラグメントは、目的のタンパク質が結合する領域（例えば、ヒストン結合部位）を含む任意の核酸フラグメントであり得る。ポリヌクレオチドは、DNAまたはRNAであり得る。

【0132】

ポリヌクレオチドは、同一染色体上に存在することができ、また、ポリヌクレオチドは、異なる染色体上に存在することができる。

10

【0133】

ポリヌクレオチドのトランスフュージョンまたは転座について検出および/または同定することができる。

【0134】

オリゴヌクレオチドを、細胞にトランスフェクトすることができる。エレクトロポレーションによってトランスフェクションを行うことができる。細胞は、細菌細胞であり得る。

【0135】

本発明のこの態様の実施形態を、以下のCIA-PETおよびCIA-diPET技術として以下により詳細に記載する。

20

【0136】

別の態様では、本発明は、本発明のオリゴヌクレオチド、オリゴヌクレオチドのコンカテマー、またはオリゴヌクレオチドのライブラリーを含むベクターを提供する。

【0137】

[CIA-PET]

CIA-PET法（図1および4）では、DPD複合体中のタンパク質によって繫留されたDNAフラグメントを、ライゲーションを介してリンカー配列によって連結する。リンカー配列は、2つのMmeI部位を含む。逆(reversal)架橋後、ライゲーションしたDNAをMmeIで消化して、対合末端ジタグを放出させる（CIA-PET）。各CIA-PETは、2つの隣接タグ（それぞれ約20bp）を有するリンカーを含む（図1の工程2）。したがって、CIA-PET中に含まれる2つのタグは、線状ゲノム配列中で互いに離れているが、互いに相互作用し、ヒストンタンパク質などの特定のタンパク質によって媒介される2つの遠位遺伝子領域を示す。

30

【0138】

CIA-PETのゲル精製後、配列特異的アダプターを、CIA-PETの各部位に添加し、次いで、PCRによって増幅する（図1の工程3）。増幅CIA-PETを、「454」ピロシーケンス法などの多重配列決定法または任意の他の適切な配列決定法によって直接配列決定するか、クローニングのためにより長いDNAストレッチに鎖状体形成してCIA-PETライブラリーを作製し、その後、Sangerキャピラリー法を使用して配列決定することができる。CIA-PET配列を、基準ゲノム配列にマッピングする。真のクロマチン相互作用部位を、特定の遺伝子座中のCIA-PETクラスターの頻発に基づいて同定することができ、シングルトンとして無作為に散乱したバックグラウンドノイズと区別する（図3）。

40

【0139】

[CIA-diPET]

CIA-diPET法（図2および6）では、2つの各関連DNAフラグメントを示す対合末端ジタグ配列を抽出および配列決定し、配列あたりたった1つのタグと比較する。そのようなものとして、得られたCIA-diPETは、CIA-PET法のタグよりも長いタグ配列を含む。当業者は、したがって、本発明のCIA検出法が当分野の他の方法よりもクロマチン相互作用のマッピングでより特異的であることを認識する。DPD複合

50

体中のDNAフラグメントを、MmeI部位およびGsuI部位を含むリンカー配列によって連結する。逆架橋後、ライゲーションしたDNAを、超音波処理によって無作為に破壊する。

【0140】

次いで、DNAをサイズ分画し、そのクローニング部位(図7)に即時型(immediate)MmeI部位およびGsuI部位を含むpGIS8ベクターにクローン化する(図2の工程2)。細菌細胞中での形質転換および増殖後、ライブラリークローンを、消化、自己ライゲーション、および形質転換の連続操作に供して1つのdiPETライブラリーを作製する。1つのdiPETライブラリーのプラスミドDNAをBamHIによって消化して、diPET構造を遊離し(図2の工程4)、MS-PET配列決定法を使用して直接配列決定するか、クローニングおよび配列決定のためにさらに鎖状体形成することができる。CIA-diPET配列を、基準ゲノム配列にマッピングする。真のクロマチン相互作用部位を、特定の遺伝子座中のCIA-diPETクラスターの頻発に基づいて同定することができ、シングルトンとして無作為に散乱したバックグラウンドノイズと区別する(図5)。

10

【0141】

ここに本発明を一般的に説明してきたが、例示の目的で提供され、本発明を制限することを意図しない以下の実施例を参照して、本発明がより深く理解される。

【0142】

品質管理およびリンカーまたはアダプターを確実に正確に挿入するために、ゲル電気泳動を実施して、過程中的望ましくないフラグメントを除去する(図8~10)。

20

[実施例]

【0143】

[CIAの方法]

マウスES細胞(E14)中のNanog転写因子、NanogについてのChIP-PETデータ、および他の重要なES細胞転写因子(Oct4およびSox2など)を、本発明を実施するための生体系として使用した。ChIP-PETデータにより、これらの転写因子が結合する場所を示す線状マップが得られ、これを、クロマチン相互作用データを立証するための基準として使用する。クロマチン間の相互作用を、2つの方法(CIA-PETおよびCIA-diPET)によって検出した。

30

【0144】

実施例で使用した配列を、以下および配列表に示す。2つの鎖を示す場合、上の鎖はセンス鎖であり、下の鎖はアンチセンス鎖である。全配列を、5'→3'方向で示す。ヌクレオチドを、Nまたはn(図中ではXまたは数字)で示す場合、その位置を任意のヌクレオチド(A、C、G、またはT)を示すことができることを意味する。

【0145】

M & M リンカー

5' GTCGGAGGCCAAGGCGGCCGTACGTCCAACT 3' (31 nt)(配列番号1)

5' GTTGGACGTACGGCCGCCTTGGCCTCCGAC 3' (31 nt)(配列番号2)

5' 末端をリン酸化する。

40

M & G リンカー

5' GTCGGAGGCCAAGGCGGCCGTACGCTGGAGT 3' (31 nt)(配列番号3)

5' CTCCAGCGTACGGCCGCCTTGGCCTCCGACT 3' (31 nt)(配列番号4)

5' 末端をリン酸化する。

P1鎖状体形成アダプター(PMR011)

5' GGATCCCTTAATCGCCTTGCAGCACATC 3' (28 nt)(配列番号5)

5' GATGTGCTGCAAGGCGATTAAGGGATCCNN 3' (30 nt)(配列番号6)

上のアダプターの5'末端をリン酸化する。

50

下のアダプターの5'末端をリン酸化しない。

P 2 鎖状体形成アダプター (P M R 0 1 2)

5' GGATCCCCTGTGTGAAATTGTTATCCGCT 3' (29 nt)(配列番号7)

5' AGCGGATAACAATTTACACAGGGGATCCNN 3' (31 nt)(配列番号8)

上のアダプターの5'末端をリン酸化する。

下のアダプターの5'末端をリン酸化しない。

D 1 d i P E T i n g アダプター (P M R 0 1 1)

5' GGATCCCTTAATCGCCTTGCAGCACATC 3' (28 nt)(配列番号9)

5' GATGTGCTGCAAGGCGATTAAGGGATCCNN 3' (30 nt)(配列番号10)

上のアダプターの5'末端をリン酸化する。

下のアダプターの5'末端をリン酸化しない。

D 2 d i P E T i n g アダプター (P M R 0 1 2)

5' GGATCCAATGCTCCTCCCTGTGTGAAATTGTTATCCGCT 3' (39 nt)(配列番号11)

5' AGCGGATAACAATTTACACAGGGAGGAGCATTGGATCCNN 3' (41 nt)(配列番号12)

上のアダプターの5'末端をリン酸化する。

下のアダプターの5'末端をリン酸化しない。

P M R 0 1 1 プライマー

5' GATGTGCTGCAAGGCGATTAAG 3' (22 nt)(配列番号13)

5'末端をリン酸化する。

P M R 0 1 2 プライマー

5' AGCGGATAACAATTTACACAGG 3' (23 nt)(配列番号14)

5'末端をリン酸化する。

R e c A 選択オリゴ

5' AGTCGGAGGCCAAGGCGGCCGTACGCTGGAGT 3' (配列番号15)

(ビオチニル化)

【0146】

p G I S 8 ベクター

このベクターは、p G E M ベクター (Promega) に由来し、以下の制限酵素のための固有の多クローニング部位をこの順序で含む。

B a m H 1 → M m e 1 ; 切断領域 ; G s u 1 → B a m H 1 → B s e R 1

以下のオリゴヌクレオチドを使用する。

5' AATTGGATCCGACTCGAGGATGAATTCTCCAGGATCCCTCCTC 3' (43 nt)(配列番号16)

5' TCGAGAGGAGGGATCCTGGAGAATTCATCCTCGAGTCGGATCC 3' (43 nt)(配列番号17)

5'末端をリン酸化する。

このベクターの残りは、B s e R 1 部位、B a m H 1 部位、M m e 1 部位、または G s u 1 部位のいずれも含まない。

【0147】

p G I S 8 プラスミドは、使用することができるベクターの一例である (図7)。センス鎖配列を配列番号18に示し、アンチセンス鎖配列は配列番号19である。制限酵素の認識部位を含む多クローニング部位 (配列番号16および17に示す) を、ベクターに挿入する。種々の制限酵素の認識部位を示す p G I S 8 についての配列表を、図11に示す。要件を満たす任意の他のベクターを調製し、当業者が使用することができる。

【0148】

M & M アダプター P E T

10

20

30

40

50

5' NNNNNNNNNNNNNNNNNNAGTTGGACGTACGGCCGCCTTGGCCTCCGACTNNNNNNNNNNNNNNNNNN 3' (69 nt) (配列番号21)

5' AGCGGATAAC AATTTACAC AGGGGATCCNNNNNNNNNNNNNNNNNNNA GTTGGACGTACGGCCGCCTTGGCCT
CCGACTNNNNNNNNNN NNNNNNNNNN GGATCCCTTA ATCGCCTTGCAGCACATC 3' (128 nt)(配列番号23
)

5' GATCCNNNNNNNNNNNNNNNNNNNNNAGTTGGACGTACGGCCGCTTGGCCTCCGACTNNNNNNNNNNNNNNNNNNNN
G 3' (77 nt) (配列番号25)

(種々の n t)

[illegible][illegible]

5' GGATCCTGGAGNNNNNNNNNNNNNNNNNNNNNNNNNACTCAGCGTACGGCCGCCTTGCCCTCCGACTNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNN NNNNNNGTCGGATC 3' (116 nt)(配列番号33)

[illegible]

制限酵素消化後に挿入された部分的 M & G リンカーを有する図 9 のオリゴヌクレオチド (50

[illegible]

5' GATCCGACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCCGACNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCCAG 3' (120 nt)(配列番号 52)

5' GATCCTGGAGNNNG
TCGGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTTCG 3' (120 nt)(配列番号 53)

5' NNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGTCGGAGGCCAAGGCGGCCGTACGCTGGAGTNNNNNNNNNNNNNNNNNNNN
NNNNNNNNN 3' (85 nt) (配列番号54)

5' NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACTCCAGCGTACGGCCGCCTTGGCCTCCGACTNNNNNNNNNNNNNNNNNN
NNNNNNNNN 3' (85 nt) (配列番号55)

[illegible]

5' GGACTGGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTCGGATCC 3' (62 nt)(配列番号58)

5' GGATCCGACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCCAGTCC 3' (62 nt)(配列番号59)

[illegible][illegible]

[マウス胚幹細胞を使用した C I A - P E T のプロトコール]

C I A - P E T 法 (図 1 および 4) は、以下の 5 つの部を含む。(1) C h I P D N A - タンパク質 - D N A 複合体の生成、(2) C I A - P E T の調製、(3) C I A - P E T の増幅、(4) C I A - P E T の配列決定、および (5) C I A - P E T 配列のマッピング (図 3)。

50

(1) ChIP DNA - タンパク質 - DNA 複合体の生成

(a) マウス胚幹 (ES) 細胞を、白血病阻害因子 (Chemicon) の存在下での無支持細胞 (feeder-free) 条件で培養する。

【0151】

(b) 約 $1 \sim 2 \times 10^8$ 細胞を回収し、ホルムアルデヒド (最終濃度 1 % ; Sigma) にて室温で 10 分間架橋する。

【0152】

(c) 細胞溶解物およびクロマチンの調製

c1. 細胞を溶解緩衝液 (50 mM HEPES、1 mM EDTA、0.15 M NaCl、1 % SDS、1 % Triton X-100、0.1 % デオキシコール酸ナトリウム、全て Ambion) に溶解する。 10

c2. クロマチンを、Branson450超音波細胞破壊機 (20 % 負荷出力、30 秒、5 ~ 8 回) を使用した超音波処理によって可溶化する。

c3. クロマチンを 10 倍希釈し、SDS を 0.1 % に低下させる。

c4. 次いで、抽出物を、14,000 rpm にて 4 で 10 分間の遠心分離によって明澄化する。

c5. この抽出物を、使用するまで -80 で保存する。

【0153】

(d) 免疫沈降

d1. 2 μ g のモノクローナル抗体 (F7、Santa Cruz) を、プロテイン G セファロー 20
ス (Pharmacia) に結合させる。

d2. 抗体被覆ビーズを、クロマチン抽出物と 4 で 16 時間インキュベートする。

d3. 次いで、ビーズを、以下の緩衝液を使用して洗浄する (Sigma Chemical Company の試薬) :

緩衝液 1 (50 mM HEPES、1 mM EDTA、0.15 M NaCl、0.1 % SDS、1 % Triton X-100、0.1 % デオキシコール酸ナトリウム) で 2 回洗浄する。

緩衝液 2 (50 mM HEPES、1 mM EDTA、0.5 M NaCl、1 % SDS、1 % Triton X-100、0.1 % デオキシコール酸ナトリウム) で 1 回洗浄 30
する。

緩衝液 3 (20 mM TrisHCl (pH 8.0)、1 mM EDTA、0.25 M LiCl、0.5 % NP40、0.5 % デオキシコール酸ナトリウム) で 1 回洗浄 30
する。

緩衝液 4 (20 mM TrisHCl (pH 8.0)、1 mM EDTA) で 1 回洗浄 30
する。

d4. 次いで、タンパク質 - DNA 複合体を、溶離緩衝液 (50 mM TrisHCl (pH 8.0)、1 mM EDTA、1 % SDS) にて 65 で 20 分間ビーズから溶離 30
する。

d5. 次いで、溶離物を、PBS (Ambion) 中にて 4 で 3 時間透析して、SDS 40
を除去する。

【0154】

(2) CIA - PET の調製

(a) 末端修復

Epicentre End-Itキットおよび以下の化学物質を使用して末端修復を行う。

クロマチン (5 μ g まで)	2.5 μ l
10 $\times$ 末端修復緩衝液	5 μ l
2.5 mM dNTP ミックス	5 μ l
10 mM ATP	5 μ l
末端修復酵素ミックス	1 μ l
無ヌクレアーゼ水	50 μ l にする

短時間ボルテックスして混合し、次いで、室温で45分間インキュベートし、70で10分間の加熱によって反応を停止させる。27 μ lの無ヌクレアーゼ水の添加によって濃度を65 ng / μ lに調整する。

【0155】

(b) Aテーリング

DNA	20 μ l
10 mM dATP (Roche)	0.5 μ l
10 $\times$ ExTaq 緩衝液 (Takara)	2.5 μ l
ExTaq ポリメラーゼ (Takara)	0.5 μ l
無ヌクレアーゼ水	25 μ lにする

10

PCR装置中にて72で30分間インキュベートし、次いで、4にする。一旦72でインキュベートすると、その直後にチューブを取り出してライゲーションを行うのに最良である。

【0156】

(c) M & MリンカーでのDNAのライゲーション (配列番号1および2)

クロマチンDNA (200 μ g)	3.1 μ l
M & Mアダプター - Tテール (38 ng)	3.1 μ l
5 $\times$ PEG 含有リガーゼ緩衝液 (Invitrogen)	6 μ l
T4リガーゼ (5 U / μ l) (Invitrogen)	1 μ l
無ヌクレアーゼ水	12.3 μ l

20

16で一晩ライゲーションして、M & Mリンカーを挿入したオリゴヌクレオチド (配列番号20および21)を得る。

【0157】

(d) プロテイナーゼKでの逆架橋 (reverse cross-link)

DNAサンプルを20 μ lアリコートに分割し、15 μ lの20 mg / ml プロテイナーゼK (Ambion)の存在下での65にて一晩のインキュベーションによって逆架橋する。翌日、1 μ lの10 mg / ml RNAアーゼA (Qiagen)を添加し、RNAを37で45分間分解し、その後、DNAのフェノール抽出およびエタノール沈殿を行う。20 μ lの溶離緩衝液中に再懸濁し、-20で保存する。得られた濃度は、通常、500 ng / μ lである。

30

【0158】

DNAを定量し、Takara広範囲ラダーおよびInvitrogen低分子量ラダーならびに(a)由来の0.5 μ lの材料を含む1%ゲルによる0.5 μ lのDNAの泳動によって品質管理を行う。(d)由来の材料をゲル上により少なく塗布すべきであり、約50 ~ 100 bpのマークでいかなる明るいバンドも示されないはずである。

【0159】

(e) Mme1切断

10 μ gのDNA	20 μ l
10 $\times$ NEB u f f e r 4 (New England Biolabs)	20 μ l
10 $\times$ SAM (New England Biolabs)	20 μ l
Mme1 (2 U / μ l) (New England Biolabs)	20 μ l
無ヌクレアーゼ水	120 μ l

40

2つのチューブに分割し、37で一晩インキュベートする。SAMを新たに調製すべきである。グリコブルーを使用してフェノールクロロホルムおよびエタノール沈殿を行う。

【0160】

DNAを定量し、Takara広範囲ラダーおよびInvitrogen低分子量ラダーならびに工程(d)由来の0.5 μ lの材料を共に含む2%ゲルまたはPAGEゲルによる3 μ lの泳動によって品質管理を行う。

【0161】

(f) ゲル精製

50

得られたDNAを、適切なラダー（例えば、20 μ lのTakara広範囲ラダー）と共に、Scie-Plas中型ユニット（60 μ l / ウェル）中の2%アガロースゲルにロードする。80 Vで約1.5時間泳動し、365 nmのUVで視覚化する。不正確に挿入されたリンカーを有するオリゴヌクレオチドを、ゲル電気泳動によって除去する（不正確に挿入されたリンカーの例は、図8～10を参照のこと）。ジタグバンドを切り出し、製造者の説明書にしたがって使い捨てFermentas Eluta Tubesを使用して電気溶出する。90 Vで1～1.5時間電気溶出し、採取したジタグを、エタノール沈殿する。200 μ lの溶出液毎に、以下を添加する。

3 M NaOAc (pH 5.2) (Amresco)	20 μ l	
1 M MgCl ₂ (Ambion)	4.5 μ l	10
グリコチューブ (Ambion)	2 μ l	
100%エタノール	800 μ l	

沈殿したジタグを、12 μ l溶離緩衝液(Qiagen)中に再懸濁し、純度チェックおよび視覚的定量的のための低分子量ラダー(Invitrogen)と共に、4～20%PAGEミニゲルで2～5 μ lを泳動する。正確に得られたジタグは、配列番号22および23に示す配列を有する。

【0162】

(3) PETの増幅

(a) アダプターライゲーション（配列番号5～8のアダプターの鎖状体形成）

DNA (100 ng)	6 μ l	20
アダプター (10 μ g)	6 μ l	
10×リガーゼ緩衝液（スperlミジン含有）	1.5 μ l	
T4 DNAリガーゼ（5 U / μ l、Invitrogen）	1 μ l	
無ヌクレアーゼ水	0.5 μ l	

総体積は15 μ lである。16 で16時間インキュベートする。

10×スperlミジン含有ライゲーション緩衝液を以下から作製する。

60 mM Tris-HCl (pH 7.5) (Ambion)		
60 mM MgCl ₂ (Ambion)		
50 mM NaCl (Ambion)		
1 mg / ml BSA (New England Biolabs)		30
70 mM β -メルカプトエタノール (Sigma)		
1 mM ATP (Invitrogen)		
20 mM DTT (Invitrogen)		
10 mM スperlミジン (Sigma)		

【0163】

(b) PCR増幅

以下のプライマー（PMR11および12）（それぞれ、配列番号13および14）およびQiagenのHotstartaqキットを使用して増幅する。

DNA	1 μ l	
10×PCR緩衝液 (Qiagen)	10 μ l	40
dNTPミックス（各10 mM） (Invitrogen)	2 μ l	
PMR11	1 μ l (0.2 μ M)	
PMR12	1 μ l (0.2 μ M)	
Hotstartaq DNAポリメラーゼ (Qiagen)	0.5 μ l	
無ヌクレアーゼ水	100 μ lにする	

【0164】

十分に混合し、以下のようにPCR装置中でインキュベートする。

1. 95 で15分間
2. 94 で0.5分間
3. 55 で0.5分間

4. 7. 2 で 1 分間

5. 工程 2 を 2 5 回 繰 り 返 す

6. 7. 2 で 1 0 分間

P C R 精 製 キ ャ ッ ト (Qiagen) を 使 用 し て 精 製 す る。

【 0 1 6 5 】

(4) C I A - P E T の 配 列 決 定

C I A - P E T を、 4 5 4 多 重 配 列 決 定 装 置 (454 life sciences) の プ ロ ト コ ー ル に し た が っ て 直 接 配 列 決 定 す る こ と が で き る。こ の 技 術 は、Margulies et al (2005) お よ び 米 国 特 許 出 願 番 号 20030068629 号 に 教 示 さ れ て い る。こ れ ら の 引 例 は、そ の 全 体 が 本 明 細 書 中 で 参 照 す る こ と に よ り 組 み 込 ま れ る。

10

【 0 1 6 6 】

(5) C I A - P E T の マ ッ ピ ン グ

Compressed Suffix Array を 使 用 し て、マ ッ ピ ン グ を 行 う こ と が で き る。2 つ の 異 な る D N A フ ラ グ メ ン ト ($n > 3$) に わ た る 複 数 の 連 結 を、実 際 の 遠 位 調 節 領 域 を 示 す た め に 利 用 す る べ き で あ る (図 3)。

〔 実 施 例 3 〕

【 0 1 6 7 】

[C I A - d i P E T 法]

C I A - d i P E T 法 (図 2 お よ び 6) は、以 下 の 部 を 含 む。(1) C h I P D N A - タ ン パ ク 質 - D N A 複 合 体 の 生 成、(2) C I A - ラ イ ブ ラ リ ー の 調 製、(3) C I A - P E T ラ イ ブ ラ リ ー の 調 製、(4) 配 列 決 定、お よ び (5) C I A - P E T 配 列 の マ ッ ピ ン グ (図 5)。

20

【 0 1 6 8 】

(1) C h I P D N A - タ ン パ ク 質 - D N A 複 合 体 の 生 成 (上 記)

(a) マ ウ ス 胚 幹 (E S) 細 胞 を、白 血 病 阻 害 因 子 (Chemicon) の 存 在 下 で の 無 支 持 細 胞 条 件 で 培 養 す る。

【 0 1 6 9 】

(b) 約 $1 \sim 2 \times 10^8$ 細 胞 を 回 収 し、ホ ル ム ア ル デ ヒ ド (最 終 濃 度 1 % ; Sigma) に て 室 温 で 1 0 分 間 架 橋 す る。

【 0 1 7 0 】

30

(c) 細 胞 溶 解 物 お よ び ク ロ マ チ ン の 調 製

c 1 . 細 胞 を 溶 解 緩 衝 液 (5 0 m M H E P E S、1 m M E D T A、0 . 1 5 M N a C l、1 % S D S、1 % T r i t o n X - 1 0 0、0 . 1 % デ オ キ シ コ ー ル 酸 ナ ト リ ウ ム、全 て Ambion) に 溶 解 す る。

c 2 . ク ロ マ チ ン を、Branson 450 超 音 波 細 胞 破 壊 機 (2 0 % 負 荷 出 力、3 0 秒、5 ~ 8 回) を 使 用 し た 超 音 波 処 理 に よ っ て 可 溶 化 す る。

c 3 . ク ロ マ チ ン を 1 0 倍 希 釈 し、S D S を 0 . 1 % に 低 下 さ せ る。

c 4 . 次 い で、抽 出 物 を、1 4 , 0 0 0 r p m に て 4 分 で 1 0 分 間 の 遠 心 分 離 に よ っ て 明 澄 化 す る。

c 5 . こ の 抽 出 物 を、使 用 す る ま で - 8 0 ° C で 保 存 す る。

40

【 0 1 7 1 】

(d) 免 疫 沈 降

d 1 . 2 μ g の モ ノ ク ロ ー ナ ル 抗 体 (F 7、Santa Cruz) を、プ ロ テ イ ン G セ フ ァ ロ ー ス (Pharmacia) に 結 合 さ せ る。

d 2 . 抗 体 被 覆 ビ ーズ を、ク ロ マ チ ン 抽 出 物 と 4 分 で 1 6 時 間 イ ン キ ュ ベ ー ト す る。

d 3 . 次 い で、ビ ーズ を、以 下 の 緩 衝 液 を 使 用 し て 洗 浄 す る (Sigma Chemical Company の 試 薬) :

緩 衝 液 1 (5 0 m M H E P E S、1 m M E D T A、0 . 1 5 M N a C l、0 . 1 % S D S、1 % T r i t o n X - 1 0 0、0 . 1 % デ オ キ シ コ ー ル 酸 ナ ト リ ウ ム) で 2 回 洗 浄 す る。

50

緩衝液 2 (50 mM HEPES、1 mM EDTA、0.5 M NaCl、1% SDS、1% Triton X-100、0.1% デオキシコール酸ナトリウム) で 1 回洗浄する。

緩衝液 3 (20 mM TrisHCl (pH 8.0)、1 mM EDTA、0.25 M LiCl、0.5% NP40、0.5% デオキシコール酸ナトリウム) で 1 回洗浄する。

緩衝液 4 (20 mM TrisHCl (pH 8.0)、1 mM EDTA) で 1 回洗浄する。

d 4 . 次いで、タンパク質 - DNA 複合体を、溶離緩衝液 (50 mM TrisHCl (pH 8.0)、1 mM EDTA、1% SDS) にて 65 で 20 分間ビーズから溶離する。 10

d 5 . 次いで、溶離物を、PBS (Ambion) 中にて 4 で 3 時間透析して、SDS を除去する。

【 0 1 7 2 】

(2) CIA - diPET ライブラリーの調製

(a) 末端修復 (CIA - PET の調製に関する)、Epicentre End-It キットを使用して末端修復を行う。

クロマチン (5 μ g まで)	2.5 μ l	
10 $\times$ 末端修復緩衝液 (Epicentre)	5 μ l	
2.5 mM dNTP ミックス (Epicentre)	5 μ l	20
10 mM ATP (Epicentre)	5 μ l	
末端修復酵素ミックス (Epicentre)	1 μ l	
無ヌクレアーゼ水	31.5 μ l	

短時間ボルテックスして混合し、次いで、室温で 45 分間インキュベートし、70 で 10 分間の加熱によって反応を停止させる。27 μ l の無ヌクレアーゼ水の添加によって濃度を 65 ng / μ l に調整する。

【 0 1 7 3 】

(b) A テーリング (上記の CIA - PET と同様)

DNA	20 μ l	
10 mM dATP (Roche)	0.5 μ l	30
10 $\times$ ExTaq 緩衝液 (Takara)	2.5 μ l	
ExTaq ポリメラーゼ (Takara)	0.5 μ l	
無ヌクレアーゼ水	1.5 μ l	

PCR 装置中にて 72 で 30 分間インキュベートし、次いで、4 にする。一旦 72 でインキュベートすると、その直後にチューブを取り出してライゲーションを行うのに最良である。

【 0 1 7 4 】

(c) M & G リンカーでの DNA のライゲーション (配列番号 3 および 4)

クロマチン DNA (200 ng)	3.1 μ l	
M & G アダプター - T テール (38 ng)	7.6 μ l	40
5 $\times$ PEG 含有リガーゼ緩衝液 (Invitrogen)	6 μ l	
T4 リガーゼ (5 U / μ l) (Invitrogen)	1 μ l	
最終体積	12.3 μ l	

16 で一晩ライゲーションして、M & G リンカーを挿入したオリゴヌクレオチド (配列番号 26 および 27) を得る。

【 0 1 7 5 】

(d) プロテイナーゼ K での逆架橋 (上記の CIA - PET と同様)

DNA サンプルを 20 μ l アリコートに分割し、15 μ l の 20 mg / ml プロテイナーゼ K (Ambion) の存在下での 65 にて一晩のインキュベーションによって逆架橋する。翌日、1 μ l の 10 mg / ml RNA アーゼ A (Qiagen) を添加し、RNA を 37 で 4 50

5 分間分解し、その後、DNA のフェノール抽出およびエタノール沈殿を行う。20 μ l の溶離緩衝液中に再懸濁し、-20 で保存する。濃度は、通常、500 ng / μ l である。

【0176】

定量し、Takara広範囲ラダーおよびInvitrogen低分子量ラダーならびに(a)由来の0.5 μ l の材料を含む1%ゲルによる0.5 μ l の泳動によって品質管理を行う。(d)由来の材料をゲル上により少なく塗布すべきであり、約50~100 bpのマークでいかなる明るいバンドも示されない。

【0177】

(e) Nla IIIでの消化

Nla IIIを、-80 (半減期は、-80 で6ヶ月)で保存する。使用直前に氷上に置く。

DNA (約1 μ g)	2 μ l
10 $\times$ NEB buffer 4 (New England Biolabs)	5 μ l
Nla III (New England Biolabs)	1 μ l
100 $\times$ BSA (New England Biolabs)	0.5 μ l
無ヌクレアーゼ水	41.5 μ l

5つの上記反応チューブを調製する。37 で1時間インキュベートする。グリコブルーを使用してフェノールクロロホルムおよびエタノール沈殿を行い、10 μ l の溶離緩衝液(Qiagen)に再懸濁する。

【0178】

(f) Epicentre End-Itキットを使用した末端の研磨

DNA (5 μ g まで)	34 μ l
10 $\times$ 末端修復緩衝液(Epicentre)	5 μ l
2.5 mM dNTP ミックス(Epicentre)	5 μ l
10 mM ATP (Epicentre)	5 μ l
末端修復酵素ミックス(Epicentre)	1 μ l

室温で45分間インキュベートし、70 で10分間の加熱によって反応を停止させる。

【0179】

(i) Mme I および Gsu I 隣接部位を有する pGIS8 ベクター (配列番号18および19、図7および11) へのクローニング

1.7 ml 微量遠心管を使用して (複数のチューブを使用する)、氷上に準備する。

40 ng / μ l プレカット pGIS	1 μ l
DNA	6 μ l
5 $\times$ PEG 含有リガーゼ緩衝液 (Invitrogen)	2 μ l
T4 DNA リガーゼ (5 U / μ l) (Invitrogen)	1 μ l

16 で一晩 (12~16時間) インキュベートして、配列番号28および29に示すオリゴヌクレオチドを得る。ベクター自己ライゲーションコントロールも準備する。

【0180】

エレクトロポレーションによって50 μ l のエレクトロコンピテントTOP10細胞 (Invitrogen) あたり1 μ l のライゲーション反応物を形質転換する。1 ml LB 培地中に各アリコート37 で1時間回収し、次いで、品質管理および滴定のために一連の連続希釈物 (LB 寒天 + アンピシリン) にプレートする。

【0181】

次いで、残りの培養物の巨大な寒天プレート (Q-trays) へのプレーティングによってこの過程を拡大し、Qiagen HiSpeed Plasmid Maxiキットを使用してマキシプレップ (maxiprep) を行う。

【0182】

(3) CIA-diPET ライブラリー

(a) Mme I 切断

10

20

30

40

50

10 μ g の DNA	100 μ l
10 $\times$ NE Buffer 4 (New England Biolabs)	20 μ l
10 $\times$ SAM (New England Biolabs)	20 μ l
Mme 1 (2 U / μ l) (New England Biolabs)	12 μ l
無ヌクレアーゼ水	48 μ l

37 で一晩インキュベートして、配列番号 30 および 31 に示すオリゴヌクレオチドを得る。SAMを、新たに調製すべきである。グリコブルーを使用してフェノールクロロホルムおよびエタノール沈殿を行い、12 μ l の溶離緩衝液に再懸濁する。

【0183】

DNAを定量し、Takara広範囲ラダーおよびInvitrogen低分子量ラダーならびに工程 (e) 由来の1 μ l の材料を共に含む2%ゲルまたはPAGEゲルによる1 μ l の泳動によって品質管理を行う。

【0184】

(b) 環状化

以下(MJ Research)を含む各ウェルを有する96ウェルプレートを用意する。

100 ng の DNA	50 μ l
ライゲーション溶液 1 (Takara Ligation Kit ver2)	50 μ l

プレートをしっかりと密封して蒸発を防ぐ。16 で一晩インキュベートする。3つのカラムを使用してPCR精製(Qiagen)を行い、40 μ l の各溶離緩衝液に再懸濁し、総量が120 μ l になる。

【0185】

(c) Gsu 1 切断

9つのチューブを用意する

環状化 DNA	12 μ l
10 $\times$ TANGO 緩衝液 (Fermentas)	8.6 μ l
10 $\times$ SAM (New England Biolabs)	8.6 μ l
Gsu I (5 U / μ l) (Fermentas)	1 μ l
無ヌクレアーゼ水	55.8 μ l

30 で少なくとも2時間消化させるが、一晩切断しない。配列番号 32 および 33 に示すオリゴヌクレオチドを得る。

【0186】

(d) 環状化

以下を含む9つのチューブを用意する。

約 100 ng の DNA	50 μ l
ライゲーション溶液 1 (Takara Ligation Kit ver2)	50 μ l

16 で一晩インキュベートする。フェノールクロロホルムエタノール沈殿を行い、12 μ l の溶離緩衝液に再懸濁する。

【0187】

(e) ローリングサークル増幅 (Templiphi kit、Amersham) を使用した増幅

氷上で溶液を解凍する。以下を含む3~4つのチューブを調製する。

2.5 ng の DNA	1 μ l
Templiphi kit の変性緩衝液 (Amersham)	10 μ l

95 で3分間加熱し、その後、氷上で短時間冷却する。

【0188】

反応緩衝液に、以下を添加する。

Templiphi kit プレミックス (Amersham)	10 μ l
---------------------------------	------------

軽くたたかか穏やかにボルテックスして十分に混合する。震盪せずに30 で16~18時間インキュベートする。材料を、例えば、1 μ l のマイクロピペッティングによって試験する。材料は、粘性を示すはずである。ピコグリーン蛍光光度法 (Quant-iT DNA アッセイキット、Molecular Probes) によって二本鎖DNAを定量する。

10

20

30

40

50

【 0 1 8 9 】

(f) B a m H 1 切 断

1 0 0 μ g の D N A	1 μ l
1 0 $\times$ 固有の B a m H 1 緩衝液 (New England Biolabs)	1 0 μ l
1 0 0 $\times$ ウシ血清アルブミン (New England Biolabs)	1 μ l
B a m H 1 (2 0 U / μ l ; 2 倍過剰) (New England Biolabs)	2 0 μ l
無ヌクレアーゼ水	7 8 μ l

必要に応じて、D N A を消化するためのさらなるチューブを調製し、3 7 で一晩インキュベートして、配列番号 3 4 および 3 5 に示すオリゴヌクレオチドを得る。

【 0 1 9 0 】

(g) ゲル精製

適切なラダー (例えば、2 0 μ l のTakara広範囲ラダー) と共に、Scie-Plas中型ユニット (6 0 μ l / ウェル) 中の 2 % アガロースゲルにロードする。8 0 V で約 1 . 5 時間泳動し、3 6 5 n m の U V で視覚化する。不正確に挿入されたリンカーを有するオリゴヌクレオチドを、ゲル電気泳動によって除去する (不正確に挿入されたリンカーの例は、図 8 ~ 1 0 を参照のこと)。ジタグバンドを切り出し、製造者の説明書にしたがって使い捨て Fermentas Eluta Tubes を使用して電気溶出する。9 0 V で 1 ~ 1 . 5 時間電気溶出し、採取したジタグを、エタノール沈殿する。2 0 0 μ l の溶出液毎に、以下を添加する。

3 M N a O A c (p H 5 . 2)	2 0 μ l
1 M M g C l 2	4 . 5 μ l
グリコチューブ	2 μ l
1 0 0 % エタノール	8 0 0 μ l

沈殿した d i P E T を、1 2 μ l 溶離緩衝液中に再懸濁し、純度チェックおよび視覚的定量的ための低分子量ラダー (Invitrogen) と共に、4 ~ 2 0 % P A G E ミニゲルで 2 ~ 5 μ l を泳動する。

【 0 1 9 1 】

(4) C I A - d i P E T の配列決定

(a) ゲル精製 B a m H I 付着 d i P E T の鎖状体形成

C I A - d i P E T 2 0 0 ~ 1 0 0 0 n g	6 μ l
1 0 $\times$ リガーゼ緩衝液 (スベルミジン含有)	1 μ l
T 4 D N A リガーゼ (5 U / μ l 、Invitrogen)	1 μ l
無ヌクレアーゼ水	2 μ l

1 6 で 2 時間から一晩インキュベートする。次いで、6 5 で 1 0 分間加熱不活化する。

【 0 1 9 2 】

(b) C I A - d i P E T コンカテマーの部分的 B a m H I 再消化

Qiagen PCR 精製 QuickSpin キットを使用して、コンカテマー D N A を精製する。次いで、Nanodrop (Nanodrop technologies) のための 1 μ l の使用によって D N A を定量し、短い B a m H I の再消化物を定量する。

コンカテマー	2 0 μ l
1 0 $\times$ B a m H I 緩衝液 (New England Biolabs)	3 μ l
B a m H I (1 U / μ l に希釈) (New England Biolabs)	0 . 2 μ l
1 0 0 $\times$ B S A (New England Biolabs)	0 . 5 μ l
無ヌクレアーゼ水	6 . 3 μ l

3 7 で 3 0 分間インキュベートし、それ以上はインキュベートしない。6 μ l のローディング色素を迅速に添加し、6 5 で 1 5 分間加熱し、氷上で冷却し、その後に P A G E ゲルにローディングする。

【 0 1 9 3 】

(c) 鎖状体形成した C I A - d i P E T の P A G E 精製

好ましくは、サイズ測定のためのTakara広範囲ラダーおよびInvitrogen低分子量ラダー

10

20

30

40

50

を隣接させた4～20%の勾配のPAGEミニゲルの1つの穴に全サンプルをローディングする。200Vで約1時間電気泳動する。SYBRグリーンIで15～30分間染色し、ゲル切り出し物についてDark Readerトランスイルミネーター(Clare Chemical)で視覚化する。

【0194】

(d) コンカテマーの切り出し

鎖状体形成DNAを3つの個別の画分(低(400～1000bp)、中(1000～2000bp)、および高(>2000bp))に切り出す。切り出した各サイズ画分のゲルスライスを0.6mlの微量遠心管に入れ、21Gニードルで底を突き刺した。この突き刺したチューブを、1.7ml微量遠心管内に入れ、16110gにて4で5分間遠心分離した。それにより、ゲル片が都合良くバラバラになり、1.7mlの各チューブの底に回収される。200μlのLoTE:NH₄OAc(167:33)(LoTEは以下の配合表に従い、NH₄OAcはAmbionから入手)を、各チューブに添加し、65で2時間の加熱によって溶離する。

10

【0195】

LoTE緩衝液:

3mM Tris-HCl(pH7.5)(Ambion)

0.2mM EDTA(Ambion)

【0196】

上記のマイクロスピニフィルターユニットを用いた16110gでの4で10分間のスピンによって上清(溶離した鎖状体形成DNAを含む)をゲル片から分離する。溶離した各サイズ画分に対してフェノール/クロロホルム抽出を行い、その後にエタノール沈殿を行う。

20

溶離DNA画分 200μl

3M 酢酸ナトリウム(pH5.2) 20μl

GlycoBlue 2.2μl

100%エタノール 800μl

-80で30分間保持し、次いで、16110gにて4で30分間スピンし、75%エタノールで1回洗浄する。6μlのLoTE緩衝液中にペレットを再懸濁した。

【0197】

30

(e) pZErO-1ベクターへのライゲーション

使用前に、pZErO-1クロニングベクターを、37で2時間の10単位のBamHI(New England Biolabs)での2μgのpZErO-1プラスミドDNA(Invitrogen)の消化によって調製する。消化されたプラスミドDNAを、フェノール・クロロホルム抽出し、エタノール沈殿を行い、33ng/μlの濃度の60μlのLoTEに再懸濁する。プラスミドを、コントロールとしてのベクターの自己ライゲーションの準備(わずかなコロニーが存在するはずである)およびアガロースゲルでの泳動によって立証することができる。

【0198】

以下のようにライゲーションを準備する。

40

コンカテマーDNA画分 6μl

BamHI/pZErO-1 1μl

5×リガーゼ緩衝液(PEG含有)(Invitrogen) 2μl

T4 DNAリガーゼ(5U/μl)(Invitrogen) 1μl

16で一晩インキュベートする。熱不活化させない。コントロールとして、並行して、ベクターの自己ライゲーションも準備する。自己ライゲーションのためのベクターを調製する場合、コンカテマーDNA画分を無ヌクレアーゼ水と置換する。

【0199】

エレクトロポレーション前に各ライゲーション反応物を精製して塩を除去する。無ヌクレアーゼ水で体積を200μlに調整し、フェノール/クロロホルム抽出(pH7.9)

50

および - 80 で 30 分間のエタノール沈殿 (GlycoBlueを使用) を行う。スピンし、ペレットを、70 %エタノールで少なくとも 2 回洗浄し、12 μ l の LoTE に再懸濁する。

【0200】

1 μ l の精製ライゲーション反応物を、予め冷却した 1.7 ml 微量遠心管中の 25 μ l のエレクトロコンピテント細胞 (例えば、Lucigen の大腸菌、Invitrogen の Top10) に添加する。ピペットを上下させて混合しないこと。その代わりに、ピペットチップで穏やかに攪拌する。氷上に 5 分間静置し、次いで、予め冷却した Biorad エレクトロポレーションキュベット (光路 0.1 cm) に移す。

【0201】

氷上でさらに 5 分間静置する。Biorad Micropulser ユニット、単一パルス、プログラム EC1 を使用して、エレクトロポレーションを行う。時定数は、通常、4.5 ~ 5 ms である。パルスから 10 秒以内に室温で 1 ml を無添加の LB 培地に添加し、15 ml Falcon チューブに移し、200 rpm にて 37 で 1 時間震盪しながら回収する。

【0202】

20 ~ 50 μ l (1 ml のうち) を、低塩 LB 寒天 + ゼオシン (Immedia Zeocin Agar, Invitrogen) を含む小寒天プレートにプレートし、37 で一晩インキュベートする。

【0203】

(f) ライブラリーの品質管理

コロニー数を計数し、自己ライゲーションバックグラウンドの排除後に有効にライブラリーを測定する。プライマー PMR011 および PMR012 を使用した PCR スクリーニングのために 24 ~ 48 コロニーを選別する。pZerO-1 ベクター自体にコントロール PCR を含める。PCR の結果に基づいて、一晩の培養物 (低塩 LB + ゼオシン、Immedia、Invitrogen) について 1 ~ 4 $\times$ 96 ウェルプレートのコロニーを選別し、プラスミドを精製および配列決定して、インサートあたりのジタグの平均数を決定する。この段階で、ライブラリーを、精製ライゲーションミックスの形態にて、大量形質転換、プラスミド抽出、および配列決定の実施が望まれるまで - 20 で保存することができる。

【0204】

(g) ライブラリープレATING およびコロニー選別の配列決定

形質転換 TOP10 (Invitrogen) 細菌細胞を、ロボット選別を容易にするための 2,000 / プレート未満のコロニー密度で 22 $\times$ 22 cm 寒天プレート (Q-trays, Genetix) にプレートした。各コロニーを選別し、LB + ゼオシン (上記を参照のこと) を含む 384 ウェルプレートにて 37 で一晩培養した。384 ウェルプレートの複数のコロニーを複製し、15 % グリセロール (Sigma) の存在下にて - 80 で保存する。

【0205】

(h) テンプレートの調製

pZerO-1 由来のクローン由来のプラスミド DNA を、Sprintprep 固相キット (Agencourt) を使用して調製する。

【0206】

(i) DNA 配列決定

プラスミドを、両方の方向から配列決定するために、配列決定プライマー PMR011 および PMR012 (それぞれ、配列番号 13 および 14) を使用して配列決定する。

【0207】

(5) CIA - diPET のマッピング

Compressed Suffix Array を使用して、マッピングを行うことができる。2 つの異なる DNA フラグメント ($n > 3$) にわたる複数の連結を、実際の遠位調節領域を示すために利用すべきである (図 5)。3 つを超える PET が、DNA の同一ストレッチの再整列を示さなければならない。すなわち、10 k B を超える距離で分離しているか、異なる染色体上に存在するタグ 1 およびタグ 2 を、ゲノムの異なる位置にマッピングするタグから

10

20

30

40

50

なる任意のキメラの前に、ゲノム再整列の代表として利用する。

〔実施例 4〕

【0208】

使用のための M および G ベクター pGIS8 の調製

(a) pGIS8 ベクターを得る。

【0209】

(b) 従来どおりローリングサークル増幅 (Templiphi kit, Amersham) によって増幅する。氷上で溶液を解凍する。以下を含む 3 ~ 4 つのチューブを調製する。

1 ng の DNA (Maxiprep) 1 μ l

Templiphi kit 変性緩衝液 10 μ l

10

95 で 3 分間加熱し、次いで、氷上で短時間冷却する。

【0210】

反応緩衝液に以下を添加する。

Templiphi kit プレミックス 10 μ l

軽くたたか穏やかにボルテックスして十分に混合する。震盪せずに 30 で 16 ~ 18 時間インキュベートする。材料を試験する。材料は粘性を示すはずである。ピコグリーン蛍光光度法 (Quant-It kit, Molecular Probes) によって DNA を定量する。

【0211】

(c) 制限酵素消化の実施

DNA 25 μ l

20

XhoI (20 U / μ l) (New England Biolabs) 1 μ l

EcoRI (20 U / μ l) (New England Biolabs) 1 μ l

10 $\times$ EcoRI 緩衝液 (New England Biolabs) 5 μ l

100 $\times$ BSA (New England Biolabs) 1 μ l

無ヌクレアーゼ水 17 μ l

37 で 16 時間インキュベートする。PCR 精製キット (Qiagen) を使用して精製する。1% アガロースゲルにて 5 μ l を試験する。

【0212】

(d) 平滑末端化

以下のように 2 つのチューブを準備する。

30

DNA 22.5 μ l

10 $\times$ 末端修復緩衝液 (Epicentre) 5 μ l

2.5 mM dNTP ミックス (Epicentre) 5 μ l

10 mM ATP (Epicentre) 5 μ l

末端修復酵素ミックス 1 μ l

無ヌクレアーゼ水 11.5 μ l

室温で 45 分間インキュベートし、70 で 10 分間の加熱によって反応を停止させる。

【0213】

当業者は、本発明の精神および範囲を逸脱することなく、種々の修正形態および改良を実施することができると認識される。

40

【0214】

例えば、本発明の CIA 法を、哺乳動物細胞の代わりに、酵母細胞などの他の細胞種で使うことができる。または、ChIP を使用する代わりに、本方法を、免疫沈降を必要としない適切な固定液での架橋によって実施することができる。この変形形態では、プロトコールは以下である。

1. 細胞を採取する。

2. ホルムアルデヒドにて 36 で 10 分間架橋する。

3. ビーズビーターでの細胞溶解およびその後の遠心分離によって上清を得る。

4. 超音波処理、ハイドロシャーリング、皮下注射針による引き抜きの反復、または制限酵素消化によって DNA を剪断する。

50

【 0 2 1 5 】

【 0 2 1 6 】

1. DNAを平滑末端化する。

10

3. ライゲーションする。

5. DNAの定量を、Invitrogen/Molecular ProbesのPicoGreenの蛍光定量キットを使用して行うことができる。

6. M m e I 制限酵素を使用して消化し、単離オリゴヌクレオチドを得る。

【 0 2 1 7 】

【 0 2 1 8 】

20

[参考文献]

- Antequera and Bird (1993), Proc Natl Acad Sci U S A. 90(24): 11995-9
- Ausubel (1995) Current Protocols in Molecular Biology, Vol. 2, 1995, Ed. Ausubel
 , et al., Greene Publish. Assoc. & Wiley Interscience, Unit 3.1.15
- Bonetta (2006) Nature Methods 3(2): 141-147.
- Brenner et al (2000)., Nat Biotechnol, 2000. 18(6): p. 630-4
- Buck and Lieb (2004) Genomics 83(3):349-60
- Dekker J, Rippe K, Dekker M and Kleckner N (2002) Science 295 (5558):1306-11.
- Dunn et al (2002) Genome Research 12(11):1756-1765
- Euskirchen et al., Mol Cell Biol, 2004. 24(9): p. 3804-14 30
- Li and Chandrasegaran (1993) Proc. Nat. Acad. Sciences USA 90:2764-8
- Lieb et al., Nat Genet, 2001. 28(4): p. 327-34
- Margulies et al, 2005 Nature 437, 376-380 (15 Sep 2005)
- New England Biolabs Catalog 2005. New England Biolabs (Ipswich, MA)
- Szybalski, W., 1985, Gene, 40:169
- Taverner et al., Genome Biol, 2004. 5(3):210.
- US 20050255501
- US HYPERLINK "[http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetacgi%2FPTO%2Fsearch-adv.html&r=2&p=1&f=G&l=50&d=PG01&S1=%28ruan.IN.+AND+ditag.BIS.%29&OS=IN/ruan+and+spec/ditag&RS=\(IN/ruan+AND+SPEC/ditag\)](http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetacgi%2FPTO%2Fsearch-adv.html&r=2&p=1&f=G&l=50&d=PG01&S1=%28ruan.IN.+AND+ditag.BIS.%29&OS=IN/ruan+and+spec/ditag&RS=(IN/ruan+AND+SPEC/ditag))" 20050059 40
- 022
- US 20030068629
- US 4,766,072
- Weinmann et al., Genes Dev, 2002. 16(2):235-44)

【図面の簡単な説明】

【 0 2 1 9 】

【図 2】C I A - d i P e t 法のための本発明の別の方法の概観を示す図である。

【図3】CIA - PETのマッピングを示す図である。実際の相互作用領域を示すCIA 50

- P E Tにより、2つの異なるゲノム領域にわたると予想され、多重C I A - P E Tにより、クラスター化していると予想される。

【図4】C I A - P E Tの関連制限酵素の認識配列を示す図である。NまたはXを使用して示した各塩基は、任意のヌクレオチド（A、C、G、またはTのいずれか）であり得る。NおよびXを含む領域は、異なるポリヌクレオチドから得た部分を示す。図中の上から下までに認められる配列は以下である。配列番号1、配列番号2、配列番号36、配列番号37、配列番号20、配列番号21、配列番号5、配列番号6、配列番号7、配列番号8、配列番号13、配列番号22、配列番号23、配列番号14、配列番号24、配列番号25（配列番号24および25は、3回反復している）。

【図5】C I A - d i P E Tのマッピングを示す図である。相互作用領域を示すC I A - d i P E Tにより、2つの異なるゲノム領域にわたると予想され、多重C I A - d i P E Tにより、クラスター化していると予想される。C I A - d i P E Tの2つのP E Tによって得られたさらなる情報は、ゲノムへのC I A - d i P E Tのマッピングに有用である。

10

【図6】C I A - d i P E Tの関連制限酵素の認識配列を示す図である。塩基NおよびXは、任意のヌクレオチド（A、C、G、またはTのいずれか）を示し、これらは、異なるポリヌクレオチドから得られる。数字1および2は、2つの領域由来のヌクレオチドまたは1つのポリヌクレオチドから得た末端を示す一方で、数字3および4は、2つの領域由来のヌクレオチドまたは別のポリヌクレオチドから得た末端を示す。数字（1、2、3、および/または4）は、任意のヌクレオチド（A、C、G、またはTのいずれか）を示し得る。図中の上から下までに認められる配列は以下である。配列番号3、配列番号4、配列番号26、配列番号27、配列番号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配列番号34、配列番号35（配列番号24および25は、3回反復している）。

20

【図7】本発明で 사용할 ことができるベクターの例である複数の固有のクローニング部位を含むベクターp G I S 8を示す図である。

【図8】どのようにしてリンカーが挿入されていないオリゴヌクレオチドが制限酵素によって切断されず、それにより、非常に長く、電気泳動によって除去されるのかを示す図である。図中の上から下までに認められる配列は以下である。配列番号3、配列番号4、配列番号38、配列番号39、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、および配列番号45。

30

【図9】リンカーの一部のみがオリゴヌクレオチドに挿入された場合、どのようにして不正確に切断され、それにより、非常に長いので電気泳動によって除去されるオリゴヌクレオチドが得られるのかについて示す図である。図中の上から下までに認められる配列は以下である。配列番号3、配列番号4、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列番号51、配列番号52、および配列番号53。

【図10】どのようにして不正確に挿入されたリンカーが、非常に長く、且つ長さが変化し得るので、電気泳動によって除去することができるオリゴヌクレオチドを生じるのかを示す図である。図中の上から下までに認められる配列は以下である。配列番号3、配列番号4、配列番号54、配列番号55、配列番号56、配列番号57、配列番号58、配列番号59、配列番号60、配列番号61、配列番号62、および配列番号63。

40

【図11A】示した種々の制限酵素の認識部位を含む実施例で使用したベクターp G I S 8の核酸配列を示す図である。この配列は、配列表中に配列番号18および配列番号19として認められる。

【図11B】示した種々の制限酵素の認識部位を含む実施例で使用したベクターp G I S 8の核酸配列を示す図である。この配列は、配列表中に配列番号18および配列番号19として認められる。

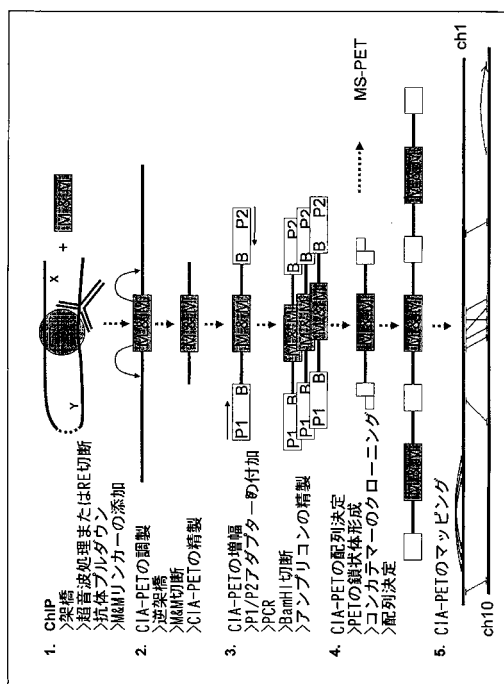
【図11C】示した種々の制限酵素の認識部位を含む実施例で使用したベクターp G I S 8の核酸配列を示す図である。この配列は、配列表中に配列番号18および配列番号19として認められる。

50

【図 1 1 D】示した種々の制限酵素の認識部位を含む実施例で使用したベクター p G I S 8 の核酸配列を示す図である。この配列は、配列表中に配列番号 1 8 および配列番号 1 9 として認められる。

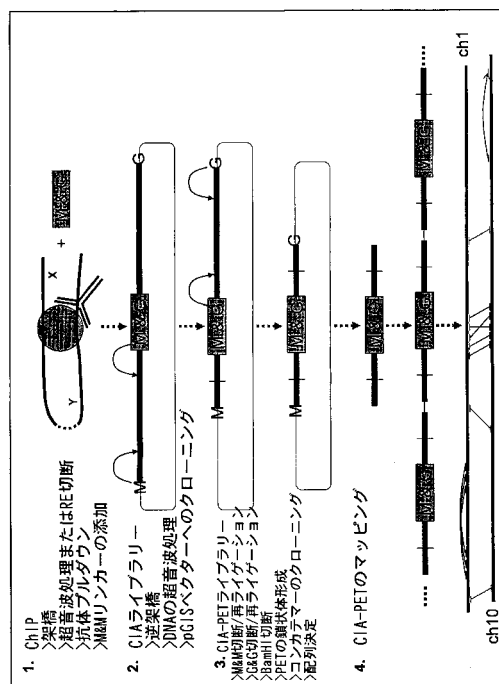
【図 1】

FIGURE 1



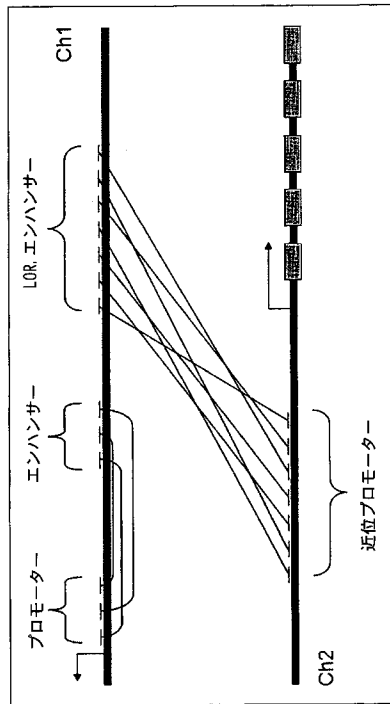
【図 2】

FIGURE 2



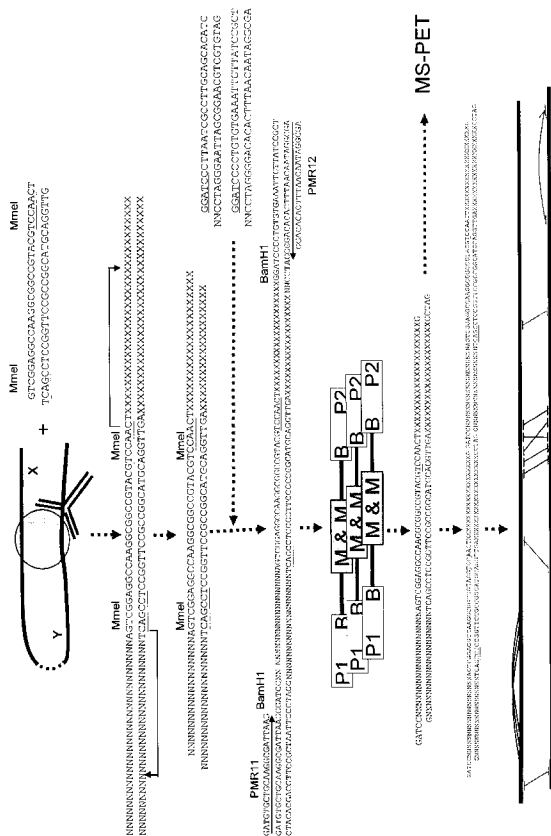
【 図 3 】

FIGURE 3



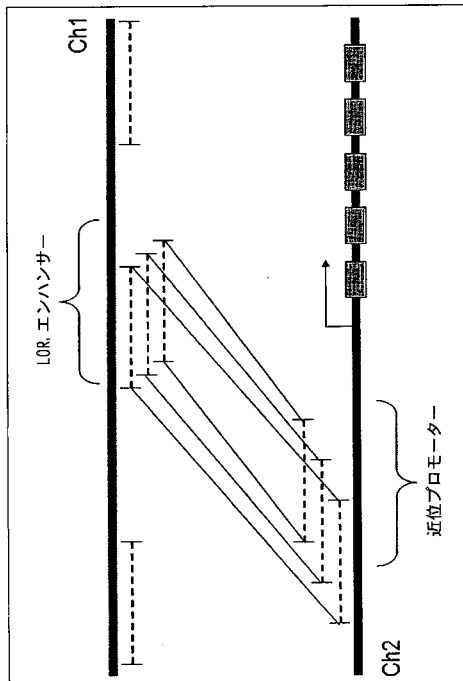
【 図 4 】

FIGURE 4



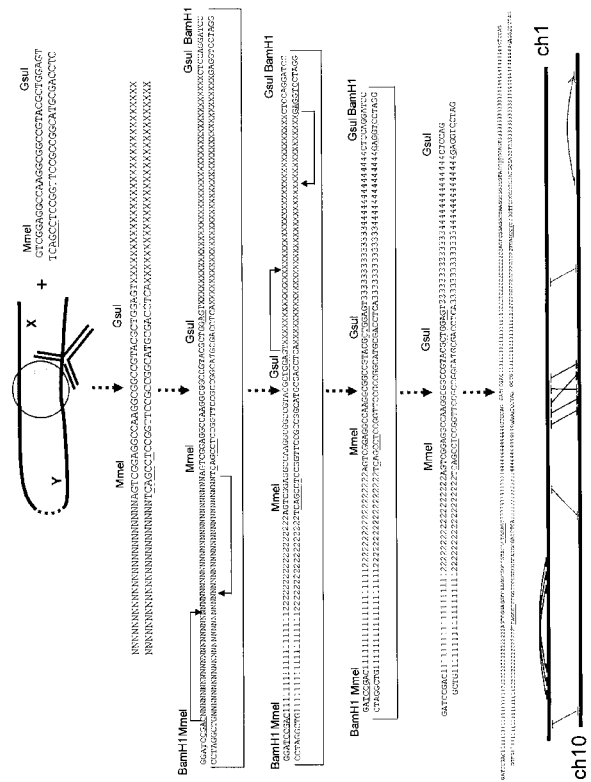
【 図 5 】

FIGURE 5



【 図 6 】

FIGURE 6



【 1 1 A 】

FIGURE 11

EcoRI NotI EcoPI51

 1 GGGGATTC GATATGCG GCGCGCTG ATAAATGAG GACTTCAC GCGAGCTCA GCGAAATG GATTAAGCT AGCGCGCG GCGTCGAGA
 CCGATTAG CTTAGAGCG GCGAGGAC ATTATGAT GTGAGAG CATTGAAAT GTTATTCA CTATCTGA TCGAGGCG GCGAGCTCT

 101 AGCTCAGCG GATCTGCG TAAATCTTC AACCTTATC TGCACACG TATGTAAC ATTGTGTT CTACAGTAA AGACATGCG
 TCGATGCG GTAGAGCG ACTTTCAG TTATATAG TACAGCAT ATTGTGTT TATGATCG GATCTGAT TGTGAGCG

 201 TCTATACG CCGTCTGA GAAAGTAC GAACTGCG TGAAGTGC CAGCGTAT AGCGCTTA AGCGCTTA TGTATGCG CCGAGCAG TCGAGCAT
 AGATATCG GGAAGCAT GTTCGATCG ACTTTCAG CAGCGTAT TGGGATTA TGGGATTA AGCTATAG GCGCTGCT AGCGCTAT

 EcoPI51

 301 GCGACACTA CAAATACAA CAGATGCG ATGACTTGT CTTAGAGCG AAGCAGCA AGCGATCT GAGACATC ASCATATAG TGGAGCTGA
 CCGTGTAT GTTATGTT CCGTTCAG TACAGACA GATGCGCT TACGTATG CTTCTGAG CTTCTGAG TGTATGCT AGCGCTAT

 EcoPI51

 401 TGGCAGTC CATACGTTA GCGCTGGA CAGAGCAT CAGCGACA GTTATCAT GATTCAGT GAAATGCG CTTGTATCG AGCGATTT
 AGCTGTAG GTATGTAAT CCGGCACT CCGTCTGA CTTGCTGT CATATGTA CTTTATGCT GTTATGCG GATCTGCG TCGCTTAA

 501 ACGTACCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 601 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

【 1 1 B 】

FIGURE 11 continued

HaeI XhoI EcoRI BamHI

 701 GCGATTAGA TCTATGAT CTTGCGCG GCGGAGCG GCGGAGCG AGCTGCTG AGTATGTC AGTATGTC AGTATGTC AGTATGTC
 CCGTAATCT AGATGATG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 801 TTGAGCTA AGCGGATG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 AAGCGCAT TACGTTAC TACGTTAC CTTATGTA AGAGCTCA GCGAGGAC GCGAGGAC TACATATA TACATATA TACATATA TACATATA

 901 GCGTATCT GTTCTGCG TATATGTT ATGCTGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 1001 GAGTATCT AATATGTC GTTCTGCG ATGCTGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 CCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 1101 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 TATATGCT ATCAGAGA TACGAGATA AGCGAGGAT GATATGTA GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 1201 ATATGAGA TATGTTCT ATGCTGCG TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT
 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 1301 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 1401 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 1501 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG
 GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG GCGGAGCG

 1601 TACTATCT CTTGAGCA AGCGGATG AGCGGATG TCGGATG TCGGATG TCGGATG TCGGATG TCGGATG TCGGATG
 ATGATAGA GACTGCT TGGGCTG TGGGCTG TGGGCTG TGGGCTG TGGGCTG TGGGCTG TGGGCTG TGGGCTG TGGGCTG

 EcoPI51

 1701 ACGAGCTGT TATGTTCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT
 TCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT

 EcoPI51

 1801 GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT
 GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT

 1901 GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT
 GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT GCTTATCT

 2001 TATATATA ATGATGTT TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA
 TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA TATATATA

【 1 1 C 】

FIGURE 11 continued

 2101 TGTCTATT TGTCTATT TGTCTATT TGTCTATT TGTCTATT TGTCTATT TGTCTATT TGTCTATT TGTCTATT TGTCTATT
 AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA

 2201 GCGAGCTCA GCGAGCTCA GCGAGCTCA GCGAGCTCA GCGAGCTCA GCGAGCTCA GCGAGCTCA GCGAGCTCA GCGAGCTCA
 CCGAGCTCA CCGAGCTCA CCGAGCTCA CCGAGCTCA CCGAGCTCA CCGAGCTCA CCGAGCTCA CCGAGCTCA CCGAGCTCA

 2301 GAGTATTA ATGTTGCG GAGTATTA ATGTTGCG GAGTATTA ATGTTGCG GAGTATTA ATGTTGCG GAGTATTA ATGTTGCG
 GAGTATTA GAGTATTA GAGTATTA GAGTATTA GAGTATTA GAGTATTA GAGTATTA GAGTATTA GAGTATTA GAGTATTA

 2401 GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA
 GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA

 2501 TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT
 TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT TCGATGTT

 2601 TGTCTATG TGTCTATG TGTCTATG TGTCTATG TGTCTATG TGTCTATG TGTCTATG TGTCTATG TGTCTATG TGTCTATG
 AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA

 2701 ATGAGAGC ATGAGAGC ATGAGAGC ATGAGAGC ATGAGAGC ATGAGAGC ATGAGAGC ATGAGAGC ATGAGAGC ATGAGAGC
 TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT TATGTTCT

 2801 GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA
 GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA GCGGAGCTA

 2901 ACGAGGAT GTGAGGAT GTGAGGAT GTGAGGAT GTGAGGAT GTGAGGAT GTGAGGAT GTGAGGAT GTGAGGAT GTGAGGAT
 TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA

 3001 AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA AGAGATTA
 TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA TGTGCTTA

 3101 GCGTATGAG AGCGCTTC GTTCTGCG GTTCTGCG GTTCTGCG GTTCTGCG GTTCTGCG GTTCTGCG GTTCTGCG GTTCTGCG
 CCGATGTC TCGGAGAG CAGAGGCG AAGCAGCTA CCGATGTC GTTCTGCG GTTCTGCG GTTCTGCG GTTCTGCG GTTCTGCG

 3201 GATCGGGA GATCGGGA GATCGGGA GATCGGGA GATCGGGA GATCGGGA GATCGGGA GATCGGGA GATCGGGA GATCGGGA
 CCGGCTCT CCGGCTCT CCGGCTCT CCGGCTCT CCGGCTCT CCGGCTCT CCGGCTCT CCGGCTCT CCGGCTCT CCGGCTCT

【 1 1 D 】

FIGURE 11 continued

 3301 TGCATATG GCGTGTGA ATACCGCA GATCGTGA GAGATATG GCGATGCG GCGATGCG GCGATGCG GCGATGCG GCGATGCG
 AGCTGTGA GCGATATG TATGCTGT CTATGCTT CTTTATG GCGATGCG GCGATGCG GCGATGCG GCGATGCG GCGATGCG

 3401 TATCGAGC GCGAGGAG ATATGCTGT GCGAGGAG ATATGCTGT GCGAGGAG ATATGCTGT GCGAGGAG ATATGCTGT GCGAGGAG
 TATCGAGC GCGAGGAG ATATGCTGT GCGAGGAG ATATGCTGT GCGAGGAG ATATGCTGT GCGAGGAG ATATGCTGT GCGAGGAG

 3501 TGTAAAGA GCGGAGTA ATGATATC GATCTACT A
 AATGTTCT GCGGATCT TATATGCT GATATGTA T

【配列表】

2007289152000001.app

フロントページの続き

(51) Int. Cl. F I テーマコード (参考)
 C 1 2 Q 1/02 (2006.01) C 1 2 Q 1/02

- (72) 発明者 イー・ジュン・ルアン
 シンガポール国, 1 3 8 6 7 2, バイオポリス・ストリート 6 0, ゲノム # 0 2 - 0 1, ゲノ
 ム・インスティテュート・オヴ・シンガポール内
- (72) 発明者 メリッサ・ジェイン・フルウッド
 シンガポール国, 1 3 8 6 7 2, バイオポリス・ストリート 6 0, ゲノム # 0 2 - 0 1, ゲノ
 ム・インスティテュート・オヴ・シンガポール内
- (72) 発明者 チア・リン・ウェイ
 シンガポール国, 1 3 8 6 7 2, バイオポリス・ストリート 6 0, ゲノム # 0 2 - 0 1, ゲノ
 ム・インスティテュート・オヴ・シンガポール内
- (72) 発明者 ハック・ファイ・ワン
 シンガポール国, 1 3 8 6 7 2, バイオポリス・ストリート 6 0, ゲノム # 0 2 - 0 1, ゲノ
 ム・インスティテュート・オヴ・シンガポール内

F ターム(参考) 4B024 AA11 AA20 CA01 CA10 CA11 CA20 HA11 HA20
 4B063 QA01 QA13 QQ02 QQ08 QQ42 QQ52 QQ91 QR08 QR14 QR20
 QR32 QR35 QR42 QR62 QR77 QR80 QS12 QS16 QS25 QS28
 QS31

【 外国語明細書 】

1. Title of Invention

NUCLEIC ACID INTERACTION ANALYSIS

2. Detailed Explanation of the Invention

Field of the Invention

[001] The present invention generally relates to the field of gene expression. More specifically, the present invention relates to nucleic acid interactions. In particular, the invention relates to analysis, detection and identification of nucleic acid-protein interaction events and components

Background of the Invention

[002] Chromatin interaction is important in gene regulation. The recently completed human genome sequences provide frameworks of genetic information. However, the human genome structure and information is often presented as a one-dimensional linearity, which is short to explain the complexity and coordination of a cellular system. An entirely different perspective is required for understanding how a genome actually functions as an orchestrated system in a three-dimensional nucleus in living cells. Genomic DNA (estimated to be two meters long stretched out) is condensed in chromosomes only a few microns across in a nucleus. It is known that chromosomes are unevenly organized into euchromatins and heterochromatins, which are packaged by chromatin proteins and communicated by transcription factors for transcription and replication.

[003] These activities appear to be ordered. It has been observed that large chromosomal loops contain active genes. Furthermore, it has been suggested that distal regulatory elements such as locus control regions (LCR), enhancers, and insulators act by repositioning specific genetic loci to regions with active or silent transcription. Recent work has demonstrated in β -globin, and most recently, in cytokine gene (IFN- γ) that LCRs may directly interact with promoters in long distances on the same chromosome and even interact with promoters on different chromosomes. It is possible that intra- and inter-chromosomal interactions are general phenomena occurring at multiple genetic loci in coordinating gene regulation of important pathways. Inter-chromosomal interactions have also been

implicated in diseases. For example, dysregulation of myc transcript is achieved by chromosomal translocations that juxtapose the c-myc/pvt-1 locus on chromosome 15 with one of the immunoglobulin loci on chromosome 12. Further analysis of interchromosomal interactions at whole genome level is necessary to identify all interactions, and will shed light on high-order gene regulations in cells.

[004] Technologies used for studies of chromatin interactions - A number of approaches have been used to study the three-dimensional structure and chromatin interactions, all with considerable limitations. Technologies applicable to this question may be roughly classified as visualization tools, such as microscopy, Fluorescence In Situ hybridization (FISH), and RNA-TRAP (RNA Tagging and Recovery of Associated Proteins); and molecular methodologies, such as chromosome conformation capture (3C), and 3C followed by chromatin immunoprecipitation (3C-ChIP).

[005] Microscopy was used in many early studies to investigate chromatin spatial organization in nuclei. However, such cytogenetic approach may only provide rough segment information of chromatins in chromosomes. FISH is a significant improvement in this direction, which localizes specific genetic loci to particular physical locations on chromosomes through fluorescence labelled DNA or RNA probes hybridizing to genomic DNA. However, the resolution was very still limited. Modified from FISH, RNA-TRAP is a method that may show distal enhancers in close physical proximity with gene promoters.

[006] Chromosome Conformation Capture (3C) was originally designed to investigate chromosomal conformation in yeast (Dekker et al, 2002), and has been used to study interactions of genetic elements that are separated in long distance and/or in different chromosomes. In 3C, DNA-protein (chromatin) structures are formaldehyde cross-linked in vivo, and chromatins are fragmented by restriction enzyme digestion. DNA fragments tethered by DNA binding proteins are then joined together by ligation, and the junctions of two suspected known elements are detected by PCR. The detection of chromatin interactions mediated by specific DNA binding protein or transcription factors may be further enhanced by chromatin immunoprecipitation (3C-ChIP), in which the chimerical DNA fragments cross-linked with protein resulted in the 3C procedure are enriched by antibody pull-down.

[007] Though each individual technique and some combinations have been demonstrated to be useful in identifying some specific intra- and inter-chromosomal interactions, these approaches relied on existing knowledge or conjecture as to what possible distal chromatin interactions may be present and primers designed to detect such junctions by PCR one region at a time. Therefore, the current technologies for study of chromatin interactions are extremely limited for identification of novel chromatin interactions and large scale at whole genome level.

[008] Despite considerable interest in the way that chromosomes are spatially organized within the nucleus and how that may regulate transcription of distal genes in concert, only scattered and indirect information is currently available. The lack of information in this aspect is largely due to the lack of robust technologies that may effectively address three-dimensional questions of chromosomal interactions.

[009] There is a need in the art for more efficient methods and robust technologies that may effectively address three-dimensional questions of chromosomal interactions that may overcome the disadvantages and limitations of the existing art.

Summary of the Invention

The present invention solves the problems mentioned above by providing a new method of detecting, identifying and/or preparing at least one nucleic acid sequence or fragment from a nucleic acid complex, in particular from a nucleic acid-protein complex. In particular, the method according to the invention provides a method of detecting, identifying and/or preparing at least two nucleic acid sequences or fragments from a nucleic acid-protein complex. The present invention also relates to oligonucleotides prepared with a method according to any embodiment of the invention. There is also provided a method to identify chromatin interaction events mediated by specific DNA binding proteins across long distances and between different chromosomes.

In another aspect, the present invention solves the problem mentioned above by providing an isolated oligonucleotide comprising at least one first tag and at least

one second tag, wherein the first and second tags are tags of a nucleic acid-protein complex. In particular, there is provided an isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex. The first and second polynucleotides may be part of the same nucleic acid region or from different nucleic acid regions of the nucleic acid-protein complex.

The isolated oligonucleotide may further comprise at least one restriction enzyme recognition site at least one linker. In particular, the at least one restriction enzyme recognition site may be included in a linker. The linker may be inserted between the tags or the linker may flank at least one tag (that is, positioned upstream and/or downstream to at least one tag). The at least one restriction enzyme recognition site may be asymmetric. The at least one restriction enzyme recognition site may be, for example, for a type IIs restriction enzyme or for a homing restriction enzyme.

[0010] The at least one first tag may comprise a 5' terminus and a 3' terminus from the first polynucleotide and the at least one second tag comprises a 5' terminus and a 3' terminus from the second polynucleotide. The isolated oligonucleotide may further comprise at least one linker. The linker may be inserted between the tags or the linker may positioned upstream and/or downstream to at least one tag. The linker comprises at least one restriction enzyme recognition site. The at least one restriction enzyme recognition site may be asymmetric. The at least one restriction enzyme recognition site may be, for example, for a type IIs restriction enzyme or for a homing restriction enzyme.

The linker may comprise a first restriction recognition site recognized by a restriction enzyme capable of cleaving the first polynucleotide to obtain the first tag, and a second restriction recognition site recognized by a restriction enzyme capable of cleaving the second polynucleotide to obtain the second tag. In particular, the linker may comprise a first restriction recognition site recognized by a restriction enzyme capable of cleaving the first polynucleotide to obtain a 3' terminus of the first polynucleotide, and a second restriction recognition site recognized by a second restriction enzyme capable of cleaving the second

polynucleotide to obtain a 5' terminus of the second polynucleotide. The first and second restriction recognition sites may be recognized by the same or different restriction enzyme.

[0011] According to another aspect, the first polynucleotide may be further cleaved by a third restriction enzyme recognizing a third recognition site to obtain a 5' terminus of the first polynucleotide, and the second polynucleotide is cleaved by a fourth restriction enzyme recognizing a fourth recognition site to obtain a 3' terminus of the second polynucleotide. According to this embodiment, at least one first tag is obtained from ligating the 5' and 3' termini of the first polynucleotide and at least one second tag is obtained from ligating the 5' and 3' termini of the second polynucleotide. The third and fourth recognition sites may be recognized by the same or different restriction enzymes. The further recognition sites may be included in adaptors ligated to the 5' end of the first polynucleotide and to the 3' end of the second polynucleotide, respectively. Alternatively, the third and fourth restriction sites may be present in a vector into which the structure first polynucleotide-linker-second polynucleotide is inserted. In this case, the third restriction site flanks the 5' end of the first polynucleotide and the fourth restriction site flanks the 3' end of the second polynucleotide.

[0012] The nucleic acid-protein complex of the isolated oligonucleotide may be part of a chromatin structure. The nucleic acid fragment to which a protein of interest binds may be any nucleic acid fragment comprising a region to which a protein of interest binds, for example, histone binding site. The polynucleotides may be DNA or RNA.

In another aspect, the present invention provides a concatemer of oligonucleotides comprising at least two isolated oligonucleotides, each isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first and second tags are tags of a nucleic acid-protein complex. IN particular, there is provided a concatemer of oligonucleotides comprising at least two isolated oligonucleotides, each isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is a tag of a first polynucleotide and the second tag is a tag of a second polynucleotide, the first and second polynucleotides are from a nucleic acid-protein complex. The concatemer may further comprise at least one linker. The linker may comprise at

least one restriction enzyme recognition site. Further, each isolated oligonucleotide of the concatemer may comprise at least one restriction enzyme recognition site. The at least one restriction enzyme recognition site may be comprised in at least one linker and/or in at least one adapter.

The linker may be inserted between the tags or the linker may be positioned upstream and/or downstream to at least one tag. The linker may comprise at least one restriction enzyme recognition site. The at least one restriction enzyme recognition site may be for a type II restriction enzyme or for a homing restriction enzyme. The first tag of each oligonucleotide of the concatemer may comprise a 5' terminus and a 3' terminus from the first polynucleotide, and the second tag may comprise a 5' terminus and a 3' terminus from the second polynucleotide.

[0013] The oligonucleotide or the concatemer according to the invention may be inserted into a vector and/or a cell. The cell may be a bacterial cell.

[0014] The polynucleotide nucleic acid-protein complex may be part of a chromatin structure. The polynucleotides may be located on the same chromosome or may be located on different chromosomes.

[0015] In another aspect, the present invention provides a library of oligonucleotides or concatemer of oligonucleotides comprising at least one oligonucleotide, the oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex.

[0016] The at least one oligonucleotide of the library may comprise at least one linker. The linker may be inserted between the tags or the linker may be positioned upstream and/or downstream to at least one tag. The first tag may comprise a 5' terminus and a 3' terminus from the first polynucleotide and the second tag further comprises a 5' terminus and a 3' terminus from the second polynucleotide.

In another aspect, the present invention provides a method of preparing at least one isolated oligonucleotide comprising:

- (a) providing a nucleic acid-protein complex;
- (b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first and second tags are obtained from a nucleic acid-protein complex; and

(c) isolating the oligonucleotide.

In particular, the present invention provides a method of preparing at least one isolated oligonucleotide comprising:

- (a) providing a nucleic acid-protein complex;
- (b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex; and
- (c) isolating the oligonucleotide.

[0017] In one embodiment, step (b) of this aspect of the invention comprises:

- (i) inserting at least one linker comprising at least one restriction enzyme recognition site, and
- (ii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing the at least one recognition site in the linker to form an oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide and the linker between the tags.

[0018] In another embodiment, step (b) of this aspect of the invention comprises:

- (i) inserting at least one linker comprising at least one restriction enzyme recognition site between the first polynucleotide and the second polynucleotide of the complex;
- (ii) adding at least one restriction enzyme recognition site to the 5' terminus of the first polynucleotide and to the 3' terminus of the second polynucleotide;
- (iii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing at least one recognition site to obtain cleaved fragments; and
- (iv) ligating cleaved fragments to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide, the tags comprising 5' and 3' termini of each polynucleotide, and the linker inserted between the tags and the at least one restriction enzyme recognition site in step (ii) is part of an adaptor or part of a vector.

[0019] The polynucleotides may be obtained from the nucleic acid-protein complex by incorporating a photoactivatable moiety into the nucleic acid (for example, DNA) and/or the protein of interest and isolation of nucleic acid/protein complex by

antibody-mediated precipitation or by affinity-mediated technique. Examples of such affinity-based techniques include streptavidin/ biotin, Glutathione-S-transferase/ glutathione matrix, maltose-binding protein/ amylose matrix interactions.

[0020] In another aspect, the present invention provides a method of detecting and/or identifying at least two polynucleotides from a nucleic acid-protein complex, the method comprising:

- (a) providing a nucleic acid-protein complex;
- (b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex;
- (c) sequencing the oligonucleotide; and
- (d) mapping the at least two polynucleotides based on the nucleotide sequences of the first and second tags,

thereby detecting and/or identifying the at least two polynucleotides.

[0021] The oligonucleotide obtained in step (b) may be amplified before being sequenced in step (c). The amplification may be by polymerase chain reaction. The amplified oligonucleotides may be subjected to at least one purification step after amplification but before being sequenced in step (c). The at least one purification step may be gel electrophoresis.

[0022] The oligonucleotide according to the invention may be concatenated with at least one further oligonucleotide obtained by steps (a) to (b) before being sequenced.

[0023] The sequencing may be carried out by the Sanger method or by multiplex sequencing such as pyrosequencing.

[0024] The detecting and/or identifying may be for transfection or translocation of the polynucleotides. Accordingly, the method may be used to detect and/or identifying polynucleotides and/or genes in proximity with each other in nucleic acid-protein complexes such as those in chromatin. In addition, the nucleic acid fragment to which a protein of interest binds may be any nucleic acid fragment comprising a region to which a protein of interest binds, for example, histone binding site. The polynucleotides may be DNA or RNA.

[0025] The oligonucleotide may be transfected into at least one cell. The transfection may be by electroporation. The cell may be a bacterial cell. The polynucleotides may be DNA or RNA.

[0026] In another aspect, the present invention provides a vector comprising the oligonucleotide, the concatemer of oligonucleotides or the library of oligonucleotides or concatemers.

Brief Description of the Drawings

[0027] FIG.1 illustrates the overview of one method of the present invention for Chromatin Interaction Analysis by paired end (di)tag(s) (CIA-PET) sequencing.

[0028] FIG.2 illustrates the overview of another method of the present invention for CIA-diPET method.

[0029] FIG.3 illustrates mapping of CIA-PET. CIA-PETs representing real interacting regions are expected to span two different genomic regions, and multiple CIA-PETs are expected to be clustered.

[0030] FIG. 4 illustrates the relevant restriction enzyme recognition sequences for CIA-PET. Each base indicated with N or X may be any nucleotide (either A, C, G or T). The regions comprising N and X represent portions obtained from different polynucleotides. The sequences which appear in the figure from top to bottom are as follows: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:13, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:14, SEQ ID NO:24, SEQ ID NO:25 (SEQ ID Nos:24 and 25 repeated thrice).

[0031] FIG.5 illustrates mapping CIA-diPET. CIA-diPETs representing the interacting regions are expected to span two different genomic regions, and multiple CIA-diPETs are expected to be clustered. The additional information provided by the two PETs of the CIA-diPET will be useful in mapping the CIA-diPETs to the genome.

[0032] FIG. 6 illustrates the relevant restriction enzyme recognition sequences for CIA-diPET. The bases N and X represent any nucleotide (either A, C, G or T) and they are obtained from different polynucleotides. The numerals 1 and 2 represent nucleotides from two regions or termini obtained from one polynucleotide while the numerals 3 and 4 represent nucleotides from two regions or termini obtained from another polynucleotide. The numerals (1, 2, 3, and/or 4) may represent any nucleotide (either A, C, G or T). The sequences which appear in the figure from top to bottom are as follows: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35 (SEQ ID Nos:24 and 25 repeated thrice).

[0033] Fig.7 illustrates the vector pGIS8 containing multiple unique cloning sites, which is an example of a vector that may be used under the present invention.

[0034] FIG. 8 illustrates how oligonucleotides that do not have a linker inserted will not be cut by the restriction enzymes and hence are too long and will be removed by electrophoresis. The sequences which appear in the figure from top to bottom are as follows: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44 and SEQ ID NO:45.

[0035] FIG 9 illustrates how if only a part of a linker is inserted into an oligonucleotide, it will result in that oligonucleotide that is incorrectly cut and hence will be removed by electrophoresis as it will be too long. The sequences which appear in the figure from top to bottom are as follows: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

[0036] FIG. 10 illustrates how an incorrectly inserted linker will give rise to oligonucleotides that are too long and variable in length and hence can be removed by electrophoresis. The sequences which appear in the figure from top to

bottom are as follows: SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62 and SEQ ID NO:63.

[0037] FIG. 11 shows the nucleic acid sequence of the vector pGIS8 used in the examples with the various restriction enzyme recognition sites indicated. The sequences appear as SEQ ID NO:18 and SEQ ID NO:19 in the sequence listing.

Detailed Description of the Invention

[0038] Definitions.

[0039] Restriction enzyme - A restriction enzyme (or restriction endonuclease) is an enzyme that cuts double-stranded DNA. The enzyme makes two incisions, one through each of the phosphate backbones of the double helix without damaging the bases. The chemical bonds that the enzymes cleave may be reformed by other enzymes known as ligases so that restriction fragments obtained from different chromosomes or genes may be spliced together, provided their ends are complementary. Type II enzymes recognized specific nucleic sequences and cut DNA at defined positions close to or within their recognition sequences sites. They produce discrete restriction fragments and distinct gel banding patterns. Type IIs enzymes cleave outside of their recognition sequence to one side. Mmel as well as most of the type IIs restriction enzymes produce variable end lengths (Dunn et al, 2002 showed that Mmel can cut 18/20 or 19/21 bases away in a rough proportion of 1:1. Therefore, the sequences given in all figures each represent one common variant of the use of Mmel.

[0040] The PET intermediates also have variable lengths (that is, M and G adapter DNA sequences after cloning into pGIS8 plasmids and M PET intermediate), because the polynucleotide can be of differing lengths. Type III enzymes are also large combination restriction-and-modification enzymes. They cleave outside of their recognition sequences and require two such sequences in opposite orientations within the same DNA molecule to accomplish cleavage. Homing endonucleases are rare double-stranded DNases that have large,

asymmetric recognition sites (12–40 base pairs) and coding sequences that are usually embedded in either introns (DNA) or inteins (proteins). Restriction enzymes may make cuts that leave either blunt end or sticky ends with overhangs. A sticky-end fragment can be ligated not only to the fragment from which it was originally cleaved, but also to any other fragment with a compatible cohesive or sticky end. As such, ends produced by different enzymes may also be compatible. A sticky end can thus be also referred as an end capable of being ligated. Many type II restriction enzymes cut palindromic DNA sequences. If a restriction enzyme has a non-degenerate palindromic cleavage site, all ends that it produces are compatible. A “palindromic” sequence is one where the sequence on one strand reads the same in the opposite direction on the complementary strand. As such, it is possible for nucleic acid strands treated to obtain palindromic cohesive ends to have its ends mate and the nucleic acid strands to self-circularize. The meaning of “palindromic” in this context is different from its linguistic usage. For example, the sequence GTAATG is not a palindromic DNA sequence, while the sequence GTATAC is. Examples of restriction enzymes leaving cohesive or sticky ends include BamH1, EcoR1 and HindIII. Examples of restriction enzymes leaving blunt, non-cohesive or non-sticky ends include BseR1 and AluI.

[0041] Nucleotide – a phosphoric ester of nucleoside; the basic structural unit of nucleic acids (DNA or RNA). Nucleotides form base pairs - one of the pairs of chemical bases joined by hydrogen bonds that connect the complementary strands of a DNA molecule or of an RNA molecule that has two strands; the base pairs are adenine with thymine and guanine with cytosine in DNA and adenine with uracil and guanine with cytosine in RNA. Short strands of nucleotides are referred to as oligonucleotides; longer strands are referred to as polynucleotides. Nucleotides may be joined with or concatenated with other nucleotides. The term nucleotide may be used interchangeably with the term nucleic acid. A stretch of nucleic acids possess a 5' end and a 3' end. The end regions of a stretch of nucleic acid may be referred to as the 5' terminus and the 3' terminus respectively. With 5' or 3' terminus of a polynucleotide, it is understood any region, fragment or whole piece of a polynucleotide that comprises the actual 5' or 3' terminus of the polynucleotide are included.

[0042] Concatemer – It is composed by at least two nucleotide monomers sequences linked end to end, optionally separated by a linker or spacer. For the purpose of the present invention, a concatemer comprises at least two oligonucleotides prepared according to the method of the invention.

[0043] Clone, cloning – To transfer nucleotides, such as a gene from one organism to another and/or to replicated the nucleotide by genetic engineering techniques.

[0044] Library - a collection of cloned nucleic acid sequences, oligonucleotides or polynucleotides, usually comprised in one or more plasmids.

[0045] Vector - A bacteriophage, plasmid, or other agent that transfers genetic material from one cell to another.

[0046] Obtain, derive – to use molecular biology and genetic engineering and manipulation techniques on biological material such as nucleic acids and proteins to confer upon the material certain desired characteristics. The terms obtain and derive may be used interchangeably under the present invention.

[0047] Amplification – increasing the copy number of nucleic acids. One method commonly used is that of polymerase chain reaction (PCR). Other amplification methods known to a skilled person may also be used.

[0048] Transfection or transformation - any method for introducing foreign molecules into a cell. Lipofection, calcium phosphate precipitation, retroviral deliver, electroporation and biolistic transformation are just a few of the teachings that may be used.

[0049] Chromatin - A complex of nucleic acids and proteins, primarily histones, in the cell nucleus that stains readily with basic dyes and condenses to form chromosomes during cell division. Chromatin is an example of a nucleic acid-protein complex. Regions of chromosomes may interact with other regions either on the same or different chromosome. The interaction event may thus be an inter- or intra- chromosomal event and may involve rearrangement of the genetic material at the regions involved.

[0050] Transfusion – rearrangement of genetic information at the RNA processing level to form a new chimerical transcript.

[0051] Translocation - rearrangement of genetic information at the genomic DNA level.

[0052] Nucleic acid-protein complex – an interaction of genetic material and a protein such as that found in chromatin or when a transcription factor binds to stretch of nucleic acid. A DNA-protein-DNA (DPD) complex is a more specific structure wherein a protein binds between two stretches of nucleic acid (DNA) of interest. A stretch of nucleic acid such as DNA may be manipulated to become a tag or an identifiable sequence of DNA.

[0053] Tag, tag-linker structure - A tag or signature is an identifiable sequence of nucleic acids refers to either the 5'- or 3'-most terminal nucleic acid sequence (terminus; usually 18-20 bp) derived from any contiguous DNA region, or a tag may comprise both the 5' and 3' most terminal nucleic acid sequences or termini of any contiguous DNA region. A linker is an artificial sequence of nucleic acids. The tag-linker-tag structure is thus an arrangement of nucleic acids wherein a linker is inserted between two tags. Another possible arrangement is a linker-tag-tag-linker structure where a linker flanks a tag (that is, it is positioned upstream and/or downstream to at least one of the tag). The terms tag and signature may be used interchangeably under the present invention.

[0054] Ditag – A short (usually 12-60 bp) nucleic acid fragment derived terminal tags or signatures of polynucleotides. A ditag may be prepared according to US 20050255501 and/or US 20050059022, the contents of which are herein incorporated by reference.

Sequencing - The methods used to determine the order of constituents in a biopolymer, in this case, a nucleic acid. Sequencing techniques used include Sanger method and modified variations thereof, as well as pyrosequencing or the "454 method" of sequencing.

[0055] In the following description, details and specific quantities and parameters are provided to describe the embodiments of the present invention. It shall be apparent to one skilled in the art, however that the invention may be practiced without such details. Some of the details may not be described at length so as not to obscure the invention.

[0056] For the performance of the methods of the present invention for a particular embodiment, any description disclosed for the purpose of carrying out other embodiments of this invention may also be used and are herein incorporated by reference. In particular, technique(s), reagents, experimental conditions,

restrictions sites, enzymes, vectors, primers, and the like. In particular, it will be evident to any skilled person how to adapt techniques and material disclosed for the other embodiments to the present embodiment of the invention.

[0057] A person skilled in the art will appreciate that techniques not specifically taught herein may be found in standard molecular biology reference books such as *Molecular Cloning: A Laboratory Manual* by Sambrook and Russell, Third Edition, 2001, published by Cold Spring Harbor Laboratory Press.

[0058] Description

[0059] The present invention relates to a new method of detecting, identifying and/or preparing at least one nucleic acid sequence or fragment from a nucleic acid complex, in particular from a nucleic acid-protein complex. In particular, the method according to the invention provides a method of detecting, identifying and/or preparing at least two nucleic acid sequences or fragments from a nucleic acid-protein complex. The invention also provides for oligonucleotides and/or concatamers of oligonucleotides.

[0060] According to one aspect, the present invention provides a method for chromatin interaction analysis (CIA). The CIA is designed to capture novel information about distal control regions and inter-chromosomal interactions de novo. This method is designed to identify chromatin interaction events mediated by specific DNA binding proteins, such as histones, across long distances and between different chromosomes.

[0061] Two embodiments of the CIA method are provided under the present invention to detect such DNA junctions, namely the CIA-PET method that extracts a single tag signature (about 20bp) from each of the two ligated DNA fragments to form a "tag1-linker-tag2" (also referred to as "first tag-linker-second tag) paired end ditag (PET) structure (Figure 1), and the CIA-diPET method that obtains two paired end ditags (PET) to represent the two related DNA fragments in a structure of "PET1-linker-PET2", so called diPET (Figure 2). The tags of CIA-PET and CIA-diPET may be directly sequenced a multiplex sequencing technique such as the "454" pyrosequencing method, or concatenated for cloning and sequencing using conventional sequencing method.

Accordingly, the present invention provides a method of preparing at least one isolated oligonucleotide comprising:

- (a) providing a nucleic acid-protein complex;
- (b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first and second tags are obtained from a nucleic acid-protein complex; and
- (c) isolating the oligonucleotide.

The first tag may be obtained from a first polynucleotide and the second tag is obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex.

The step (b) may comprise:

- (i) inserting at least one linker comprising at least one restriction enzyme recognition site, and
- (ii) cleaving the first and second polynucleotides with a least one restriction enzyme recognizing the at least one recognition site in the linker to form an oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide and the linker.

In particular, step (b) may comprise:

- (i) inserting at least one linker comprising at least one restriction enzyme recognition site between the first polynucleotide and the second polynucleotide of the complex;
- (ii) adding at least one restriction enzyme recognition site each to the 5' terminus of the first polynucleotide and the 3' terminus of the second polynucleotide;
- (iii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing at least one recognition site to obtain cleaved fragments; and
- (iv) ligating cleaved fragments to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide, the tags comprising 5' and 3' termini of each polynucleotide, and the linker inserted between the tags.

[0062] The present invention also provides a method of detecting and/or identifying at least two polynucleotides from a nucleic acid-protein complex, the method comprising:

- (a) providing a nucleic acid-protein complex;
- (b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex;
- (c) sequencing the oligonucleotide; and
- (d) mapping the at least two polynucleotides based on the sequence, thereby detecting and/or identifying the at least two polynucleotides.

[0063] As the linker sequence carries two Type II restriction enzyme recognition sites at each end (FIGS 2 and 4), therefore, after a Type II restriction digestion, one tag sequence signature each (about 20bp) may be excised from the two linked DNA fragments, to form a tag-linker-tag structure, in which one tag represents one DNA region of a chromosome while the other tag represents a locus from a far away region on the same or in different chromosome. This paired end ditag structure is referred to as "CIA-PET", which may be efficiently sequenced through concatenation into longer stretch of DNA or directly analyzed by sequencing.

[0064] Alternatively, linker sequences may flank the two tags to yield a linker-tag-tag-linker structure.

[0065] In this method, the nucleic acid-protein complexes, such as native DNA-protein-DNA (DPD) complexes, are cross-linked by a suitable fixative such as formaldehyde, glutaldehyde or methanol. The cross-linked DPD complex may then be fragmented by sonication, hydroshearing (Hydroshear, Gene Machines), repeated drawing through a hypodermic syringe needle or by restriction enzyme digestion. The DNA fragments originated from different chromosomes or at long distance are tethered by DNA binding proteins in the DPD complexes. Unlike the 3C technique that requires existing knowledge or conjecture as to what these DNA fragments are so as to generate PCR primers, the ends of DNA fragments with distal relationship bounded in a DPD complex are joined by specific linker through ligation.

[0066] The linker (about 20bp) contains two Type II restriction enzyme recognition sites to join the ends of different DNA fragments tethered by proteins in each DPD complex. The DNA junction of the two related DNA fragments may then be tagged by the paired end ditagging (PET) strategy (US 20050255501). Type II restriction enzyme recognition sites are preferred. Besides Type II restriction enzymes, any other suitable restriction enzyme, including Type III or homing restriction enzymes may be used.

[0067] Thus the present invention provides in one aspect, a method to identify chromatin interaction events mediated by specific DNA binding proteins, such as histones, across long distances and between different chromosomes. In another aspect, the present invention provides an isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex. The tags correspond to regions of chromatin in nucleic acid-protein complexes. These tags may then be sequenced to analyse, identify, and/or detect chromatin interaction events (FIGS 3 and 5).

[0068] The isolated oligonucleotide may further comprises at least one linker. The linker may be inserted between the tags or a linker may be positioned upstream and/or downstream the at least one tag. The linker may comprises at least one restriction enzyme recognition site; the at least one restriction enzyme recognition site may be asymmetric, the at least one restriction enzyme recognition site may be for a type II restriction enzyme or for a homing restriction enzyme.

[0069] Alternatively, the at least one first tag may comprise a 5' terminus and a 3' terminus from the first polynucleotide and the at least one second tag comprises a 5' terminus and a 3' terminus from the second polynucleotide. The isolated oligonucleotide may further comprise at least one linker. The linker may be inserted between the tags or a linker may be positioned upstream and/or downstream the at least one tag. The linker comprises at least one restriction enzyme recognition site; the at least one restriction enzyme recognition site may be asymmetric, the at least one restriction enzyme recognition site may be for a type II restriction enzyme or for a homing restriction enzyme.

[0070] The linker may comprise a first restriction recognition site recognized by a restriction enzyme capable of cleaving the first polynucleotide to obtain the first tag, and a second restriction recognition site recognized by a restriction enzyme capable of cleaving the second polynucleotide to obtain the second tag (FIGS. 1 and 4).

[0071] The linker may comprise a first restriction recognition site recognized by a restriction enzyme capable of cleaving the first polynucleotide to obtain a 3' terminus of the first polynucleotide, and a second restriction recognition site recognized by a second restriction enzyme capable of cleaving the second polynucleotide to obtain a 5' terminus of the second polynucleotide. Alternatively, the linker may comprise a first restriction recognition site recognized by a restriction enzyme capable of cleaving the first polynucleotide to obtain a 5' terminus of the first polynucleotide, and a second restriction recognition site recognized by a second restriction enzyme capable of cleaving the second polynucleotide to obtain a 3' terminus of the second polynucleotide.

[0072] The first polynucleotide may be further cleaved by a third restriction enzyme recognizing a third recognition site to obtain a 3' terminus of first polynucleotide and the second polynucleotide may be further cleaved by a fourth restriction enzyme recognizing a fourth recognition site to obtain a 5' terminus of the second polynucleotide; the at least one first tag obtained from ligating the 5' and 3' termini of the first polynucleotide and the at least one second tag obtained from ligating the 5' and 3' termini of the second polynucleotide (FIGS. 2 and 6). The first and third recognition sites may possess the same sequence and the second and fourth recognition sites may possess the same sequence.

[0073] The nucleic acid-protein complex of the isolated oligonucleotide may be part of a chromatin structure. The nucleic acid fragment to which a protein of interest binds may be any nucleic acid fragment comprising a region to which a protein of interest binds, for example, histone binding site. The polynucleotides may be DNA or RNA.

[0074] The resulting CIA-PET sequences may be mapped to reference genome sequences to localize the chromatin interaction junction points, which may be either long range intra-chromosome (on the same chromosome) or inter-

chromosome (on different chromosomes). This information may be used for study of chromosomal 3-dimensional organization structures in nuclei, coordinated gene regulation in long distance and across different chromosomes mediated by transcription factors, and other epigenetic related questions (FIGS. 3 and 5).

[0075] With complete genome sequences available, it is possible to develop whole genome approaches to identify all potential chromatin interactions. Whole genome tiling array is an attractive approach for genome interrogation, in which 20-60mer oligonucleotides are tiled to cover the entire genome in microarrays, and DNA probes with biological contents are hybridized to the array, and the profiles of hybridization signal intensity of all the array elements is the readout for interrogation of genetic elements. Tiling arrays have been demonstrated useful for identifying exons, and for localizing transcription factor binding sites when coupled with ChIP (ChIP-chip). Though the array-based approach is efficient due to its highly multiplex and parallel nature, it is inconceivable that the hybridization-based detection will be capable of detecting non-linear relationship of two DNA fragments in chromatin interactions.

Preparation of the oligonucleotide according to one embodiment

[0076] The nucleic acid-protein complex may be obtained through Chromatin immunoprecipitation (ChIP).

Chromatin immunoprecipitation (ChIP)

[0077] ChIP has been used to enrich and thereby allow the identification of genomic regions associated with specific proteins such as histones and other proteins binding to nucleic acids in nucleic-acid protein complexes (reviewed in Taverner et al., Genome Biol, 2004. 5(3): p. 210). The aim is to cross-link proteins with DNA at their sites of interaction. This is accomplished quickly and efficiently by adding a suitable fixative such as formaldehyde, glutaldehyde or methanol directly to living cells in culture.

[0078] Crude extracts of these fixed cells are then prepared, and the chromatin sheared by sonication, hydroshearing, repeated drawing through a hypodermic syringe needle or by restriction enzyme digestion to an average size of usually about 1 kb, then used in immunoprecipitation reactions with antibodies raised

against the DNA-associated protein of interest (e.g. transcription factors or histones). DNA fragments enriched in each immunoprecipitation are then de-linked and purified to allow their identification by a variety of methods. The advantage of using ChIP is that this approach is able to “freeze” the in vivo gene regulatory network by rapid cross-linking of chromatin and other non-histone proteins, thereby in theory representing a “true” picture of the regulatory system at any point in time, free of potential artifacts imposed by heterologous expression, for instance.

[0079] Recently, ChIP has been combined with whole-genome (Lieb et al., *Nat Genet*, 2001. 28(4): p. 327-34), whole-chromosomal (Euskirchen et al., *Mol Cell Biol*, 2004. 24(9): p. 3804-14) and CpG island (Weinmann et al., *Genes Dev*, 2002. 16(2): p. 235-44) or microarrays in a “ChIP-chip” or “ChIP-on-chip” approach that promises to enable the genome-level localization of protein-binding sites such as transcription factor binding sites (TFBS) (reviewed in Buck and Lieb, 2004). While the usefulness of this approach has been demonstrated for small genomes such as yeast (Lieb et al., *Nat Genet*, 2001. 28(4): p. 327-34), the cost and complexity of producing whole-genome microarrays for more complex organisms still remains a limiting factor.

[0080] CpG island microarrays contain human genomic fragments of high CpG content, and because CpG islands often correspond to promoter regions (Antequera and Bird, *Proc Natl Acad Sci USA*, 1993. 90(24): p. 11995-9), such microarrays represent a possible compromise. However, the location of putative protein-binding sites still has to be indirectly inferred by examining genomic DNA upstream and downstream (usually 1-2 kb, as this is the approximate size of sonicated ChIP fragments) of the CpG-rich probe spotted on the array.

[0081] As an alternative, cloning and sequencing of the ChIP-enriched DNA fragments has previously been attempted but with limited success. The problem is that the targets of ChIP enrichment are obtained against a high background of the entire genome. Even a 100-fold enrichment of specific targets would still represent only a small fraction of clones in a ChIP library, making standard DNA sequencing a very costly solution. Therefore, sequencing ChIP clones under these circumstances is not a good approach for identifying the enriched targets. Serial Analysis of Gene Expression (SAGE) and Massively Parallel Signature

Sequencing (MPSS) (Brenner et al., Nat Biotechnol, 2000. 18(6): p. 630-4) have also been suggested as useful quantitative tools for detecting ChIP enrichment, the underlying principle being that the tags generated from ChIP-enriched DNA fragments would be present in larger numbers compared to the non-specific background.

[0082] These tags could then be mapped to the genome sequence for identification of the general region of interest (i.e. assumed to be 1-2kb, representing the sonicated fragments). Although the 20bp SAGE and MPSS tags should be specific enough in most instances to define the specific genome location, one still has to examine all sequences approximately 1-2kb upstream and downstream of the tag when mapping to the genome. This is the same problem faced by the CpG island microarray approach. Furthermore, complete coverage using these methods depends on the availability of prerequisite restriction enzyme recognition sites (mapping-enzyme sites); if a recognition site is absent from a certain genomic location, that particular tag will be missing from the corresponding ChIP fragment, and hence that location will be a "blind spot" within the genome.

[0083] From the issues described above, it is clear that what is required to facilitate genome level transcriptional regulatory analysis is a method to accurately and rapidly pinpoint the nucleic acid sequences flanking protein binding regions, as an alternative to whole genome arrays. In this regard, the novel approach provided by the present invention, possesses several advantages: (i) the tag sequences generated by one method of the present invention provide higher specificity for mapping, because each tag would already be known to have been derived from a contiguous DNA segment encompassed by the 5' and 3' signatures. This information facilitates precise localization of the genomic region of interest, and obviates the need to repeatedly examine every sequence an arbitrary 1-2kb upstream and downstream of a standard SAGE or MPSS tag; (ii) the method of the present invention is thus independent of any requirement for the presence of mapping-enzyme sites; (iii) the concatenation of tags prior to sequencing means that several tags may be identified within one sequencing read; (iv) the region that is common to (i.e. overlapped by) all mapped tags in that

cluster therefore defines the DNA regions involved in the nucleic acid-protein complex in question.

Tags and ditags

[0084] For the purpose of the present application, a tag is a nucleotide sequence or signature obtained from a nucleic acid molecule and represents the polynucleotide from which the tag was obtained or derived from. The polynucleotide which is intended to shrink or represent may be RNA, mRNA, genomic DNA, full-length cDNA, or cDNA.

[0085] Under the present invention, two tags that are present in an oligonucleotide of the present invention may also each be called a ditag. Like tags, a ditag is shorter than the original nucleic acid molecule from which it originates or which it represents. Preferably, the ditag must be much shorter than the original nucleic acid molecule. As consequence of the "shrinking", the ditag may essentially comprises either or both the 5' end region (also indicated as 5' tag) and 3' end region (also indicated as 3' tag) of the original nucleic acid molecule. Hence, the portion of the original nucleic acid molecule that is originally between or inside the 5' tag and 3' tag is not included in the ditag. The ditag according to the invention retains the most informative features of the original nucleic acid molecule, namely the start and the end signatures of the nucleic acid.

[0086] The 5' tag and 3' tag forming the ditag may have the same or different size. Preferably, they have the same number of nucleotides. The ditag may be of any size, but needs to be meaningful and advantageous over the size of the parental sequence from which it is derived. The preferred size of a tag or ditag is determined by genome complexity. For a bacterial genome a tag from about 8 bp to about 16 bp may be sufficient whereas for a complex genome like the human genome, a 16-20 bp tag (or in other words, a 32-40bp ditag) may be considered. In general, the size of the ditag is from about 12-60 bp.

[0087] For the purpose of the present application, the terms 5'-terminus, 5'-end and 5'-tag are equivalent to each other and may be used interchangeably. In the same way, the terms 3'-terminus, 3'-end and 3'-tag are equivalent to each other and may be used interchangeably. In an original nucleic acid molecule or portion inside a nucleic acid molecule that one intends to reduce or represent, each of the

5'-end and 3'-end represents a region or portion closest to the extremity and most far from the middle region of the molecule.

[0088] According to one aspect of the present invention, the 5'-tag and 3'-tag comprised in the ditag are the regions of the molecule cleaved by a restriction enzyme closest to the 5'-end and 3'-end, respectively, of the nucleic acid molecule or portion thereof which is intended to be reduced or represented. Accordingly, the size of the ditag may be determined by the restriction enzyme or enzymes used.

[0089] Accordingly, the invention provides an isolated oligonucleotide comprising an isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex. The oligonucleotide may further comprise at least one linker inserted between the tags. The oligonucleotide may also further comprise at least one linker inserted to flank a tag (that is, the at least one linker may be positioned upstream and/or downstream the at least one tag).

Linker

[0090] In particular, each linker may comprise at least: one first restriction site and at least a second adjacent restriction site. Therefore, the number of restriction sites present in each linker may be one or more, preferably two. The restriction site may be an asymmetric restriction site. Examples of asymmetric restriction sites are homing endonuclease asymmetric recognition sites, and some type II (or class II) recognition sites. Type II's restriction enzymes that cut to one side of their recognition sites are preferable.

[0091] However, any recognition site known in the art may be used. Restriction enzyme recognizing at least one recognition site within the nucleic acid molecule and which may be used will be evident to those skilled in the art (see for example, Current Protocols in Molecular Biology, Vol. 2, 1995, Ed. Ausubel, et al., Greene Publish. Assoc. & Wiley Interscience, Unit 3.1.15; New England Biolabs Catalog, 2005). A list of possible restriction sites and corresponding restriction enzymes recognizing is reported below.

[0092] As an example, a restriction enzyme recognizing a restriction site may be used for the purpose of the preparation of the ditag according to the invention. In particular a type II enzyme, for example MmeI. When MmeI is used, this enzyme recognizes a sequence inside each of the two adaptors that flank the nucleic acid molecule that one intends to reduce, but cuts inside the nucleic acid molecule forming a tag comprising 17-21 nucleotides (see MmeI cuts indicated in FIGS. 1 and 4). Two such tags may be additionally processed by blunting and ligation to form a ditag comprising 34-38 nucleotides. The ditag is hence obtained by splicing together or ligating the 5' terminus and the 3' terminus of the same nucleic acid molecule.

[0093] As an example, asymmetric sites may be introduced. Asymmetric site sequences useful for the purpose of the present invention are: i) two homing endonuclease asymmetric recognition site sequences or ii) restriction endonuclease asymmetric cleavage sites sequences recognizable by type II restriction enzymes.

[0094] Homing endonucleases are sold and described by New England Biolabs, Inc.; a description of the asymmetric site sequences is also available in the New England Biolabs Catalog. These homing endonuclease asymmetric recognition site sequences are from 18 to 39 bp. However, in the present invention the recognition site sequences are not limited to those sequences nor to these sizes. Preferably, the restriction homing endonucleases capable of cutting the asymmetric site sequences are selected from the group consisting of: I-CeuI, PI-SceI, PI-PspI and I-SceI. The list mentioned above however is not exhaustive. Other homing endonucleases known in the art and those which may be later discovered are included in the scope of the present invention.

[0095] Examples of type II restriction enzymes include: AarI, AclII, Alol, BaeI, Bbr7I, BbvI, BbvII, BccI, Bce83I, BceAI, Bcefl, Bcgl, BciVI, Bfil, BinI, BpII, BsaXI, BscAI, BseMII, BseRI, BsgI, BsmI, BsmAI, BsmFI, Bsp24I, BspCNI, BspMI, Bsrl, BsrDI, BstF5I, BtgZI, BtsI, CjeI, CjePI, EciI, Eco31I, Eco57I, Eco57MI, Esp3I, Fall, Faul, FokI, GsuI, HaeIV, HgaI, Hin4I, HphI, HpyAV, Ksp632I, MboII, MlyI, MmeI, MnlI, PfiI, Ppil, PsrI, RleAI, SapI, SfaNI, SspD5I, Sth132I, StsI, TaqII, TspDTI, TspGWI, TspRI and Tth111II (the list in the web site of Rebase Enzymes®: <http://rebase.neb.com/cgi-bin/outsidelist>; see also Szybalski, W., 1985, Gene,

40:169). The list mentioned above however is not exhaustive. Other type II enzymes known in the art and those which may be later discovered are included in the scope of the present invention.

[0096] Examples of recognition sites and cleavage sites of several class II restriction enzymes are (into parenthesis are the recognition site and the cleavage site): BbvI (GCAGC 8/12), HgaI (GACGC 5/10), BsmFI (GGGAC 10/14) SfaNI (GCATC 5/9), and Bsp I (ACCTGC 4/8).

[0097] Artificial restriction endonucleases may also be used. These endonucleases may be prepared by protein engineering. For example, the endonuclease FokI has been engineered by insertions so that it cleaves one nucleotide further away from its recognition site on both strands of the DNA substrates. See Li and Chandrasegaran, Proc. Nat. Acad. Sciences USA 90:2764-8, 1993. Such techniques may be applied to prepare restriction endonucleases with desirable recognition sequences and desirable distances from recognition site to cleavage site.

[0098] Under the present invention, the isolated oligonucleotide of the present invention may be joined or concatenated with other isolated oligonucleotide to form a concatemer of oligonucleotides. Any number of oligonucleotides may be joined together for the purposes of sequencing or for cloning into a suitable plasmid or vector.

[0099] Accordingly, in another aspect, the present invention is a concatemer of oligonucleotides comprising at least two isolated oligonucleotides, each isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex.

[00100] The isolated oligonucleotides of the concatemer of oligonucleotides may further comprise at least one linker. The linker may be inserted between the tags. Alternatively, a linker may flank a tag. The linker may comprise at least one restriction enzyme recognition site; the at least one restriction enzyme recognition site may be for a type II restriction enzyme.

[00101] In one embodiment, the first tag of the concatemer of oligonucleotides may comprise a 3' terminus from the first polynucleotide and the second tag further

comprises a 5' terminus from the second polynucleotide. The isolated oligonucleotides of the concatenated polynucleotide may each further comprise at least one linker inserted between the tags. In another embodiment, the first tag of the concatemer of oligonucleotides may comprise a 5' terminus from the first polynucleotide and the second tag further comprises a 3' terminus from the second polynucleotide.

[00102] In another embodiment, the first tag of the concatemer of oligonucleotides may comprise a 5' terminus and a 3' terminus from the first polynucleotide and the second tag further comprises a 5' terminus and a 3' terminus from the second polynucleotide. The isolated oligonucleotides of the concatemer of oligonucleotides may each further comprise at least one linker inserted between the tags.

[00103] The linkers of either of these embodiments may comprise at least one restriction enzyme recognition site; the at least one restriction enzyme recognition site may be for a type IIs restriction enzyme.

[00104] The concatemer of oligonucleotides may be inserted into a vector or a cell; the cell may be a bacterial cell.

[00105] The concatemer of oligonucleotides may be from a chromatin structure. The polynucleotides are located on the same chromosome or the polynucleotides are located on different chromosomes.

[00106] While these are preferred concatemer(s), it will be apparent that the number of oligonucleotides of the present invention that may be concatenated depends on the length of the oligonucleotides and may be readily determined by those of skill in the art without undue experimentation. After formation of concatemers, multiple tags may be cloned into a vector for sequence analysis, or ditags or concatemers may be directly sequenced without cloning by methods known to those of skill in the art. Hence, the concatenation of ditags allows an efficient analysis of the nucleic acid molecules, like full-length cDNAs, in a serial manner by sequencing multiple ditags within a single vector or clone.

[00107] While term vector or recombinant vector it is intended, a plasmid, virus or other vehicle known in the art that has been manipulated by insertion or incorporation of the ditag genetic sequences may also be used. Such vectors

contain a promoter sequence that facilitates the efficient transcription. The vector typically contains an origin of replication, a promoter, as well as specific genes that allow phenotypic selection of the transformed cells. Vectors suitable for use in the present invention include for example, pBlueScript (Stratagene, La Jolla, CA); pBC, pZErO-1 (Invitrogen, Carlsbad, CA) and pGEM3z (Promega, Madison, WI) or modified vectors thereof as well as other similar vectors known to those of skill in the art. As a particular realization, the pGEM3z vector has been modified, and will be referred to as pGIS8 (FIGS. 7 and 11). pGEM vectors have also been disclosed in US 4,766,072, herein incorporated by reference.

[00108] For the production of the parental polynucleotide or nucleic acid molecule from which the tags or ditags were derived, as full-length libraries, suitable vectors may be used. Accordingly, suitable vectors, which are within the scope of the present invention, are those wherein the backbone of the vector does not comprise the same restriction site comprised in the adaptors flanking the polynucleotide or the ditag, after insertion of the polynucleotide. Preferably, the invention provides a vector wherein the vector backbone (other than within the stuffer region containing multiple cloning sites that is removed during insertion of the polynucleotide) does not comprise the first restriction site and the second or further restriction sites that are comprised into the adaptors. In particular, the vector does not comprise the at least II restriction site (for example type IIs restriction site) and the at least second or further restriction site comprised in the adaptors. More preferably, the vector backbone (other than within the stuffer region that is removed during insertion of the polynucleotide) does not comprise Mmel and BamHI.

[00109] An example of such a vector not comprising Mmel in any region outside of the stuffer is the vector pGIS8 shown in FIGS. 7 and 11. In pGIS8 the Mmel recognition sites were deleted by mutagenesis.

[00110] Accordingly, the present invention provides a library of oligonucleotides comprising at least one oligonucleotide, the oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex.

There is also provided a library of concatemer(s) of oligonucleotides according to the invention.

[00111] The at least one oligonucleotide of the library of oligonucleotides may comprise at least one linker is inserted between the tags. Alternatively, a linker may instead flank a tag. The first tag may comprise a 5' terminus and a 3' terminus from the first polynucleotide and the second tag further comprises a 5' terminus and a 3' terminus from the second polynucleotide.

[00112] The polynucleotide nucleic acid-protein complex may be part of a chromatin structure. The polynucleotides may be located on the same chromosome or the polynucleotides may be located on different chromosomes.

[00113] According to one aspect, the oligonucleotide is amplified. For example, by using PCR or any other known amplification methods. The PCR primers and the probes sequences may be prepared based on the information of the sequence of the ditags. Accordingly, suitable PCR primers corresponding to specific regions inside the vector are used. Such regions flank the oligonucleotide comprising the ditag and adaptors. PCR may be performed directly on the ligation (self-circularization) reaction to obtain short (for example 200bp) PCR products.

[00114] These PCR products that contain the required ditags may then be cut with an enzyme recognizing the at least second restriction site (inside the adaptors) to generate the required short cohesive ditags. As restriction enzyme recognizing the second or further restriction site, BamHI may for example be used, and cohesive ditags of 50bp are generated. The advantage of this amplification step is that of generating ditags circumvents the need to produce a ditag library amplification, which may be avoided by not transforming the self-circularized tagged plasmids. The amplified oligonucleotide may then subsequently be excised from the vector (in this example, by digestion with BamHI) and concatenated in long stretches of DNA or RNA for subsequent cloning and sequencing analysis (FIGS 1 and 2).

[00115] As a particular aspect, the invention discloses a cDNA library wherein the oligonucleotide(s) comprises at least one ditag, and wherein the ditag comprises about 34-38 nucleotides and is obtained by splicing nucleotides from the 5' terminus and nucleotides from the 3' terminus of a full-length cDNA or fragment thereof.

[00116] The ditag library according to the invention is representative of the library comprising the original nucleic acid molecules. For example, when the library comprising the nucleic acid molecules is a full-length polynucleotide library, the ditag library is representative of the full-length ditag library. Each ditag clone comprises sufficient information characterizing the specific full-length clone. More important, the ditag of the invention comprises the 5'-end and 3'-end of an original full-length polynucleotide derived from the nucleic acid-protein complex. Hence, the ditag is representative of the structure of the polynucleotide.

[00117] Accordingly, it is sufficient to sequence and analyze the ditag clones of the ditag library. In case a ditag of interest is found, the corresponding full-length polynucleotide may be selected and prepared from the full-length polynucleotide library, for example by PCR or directly from target RNA samples by reverse transcription-polymerase chain reaction (RT-PCR).

[00118] Sequencing the oligonucleotides of the invention may be performed by the "454" sequencing (pyrosequencing) method (Margulies et al, 2005), or concatemerized into longer DNA stretches for cloning to make CIA-PET library and followed by sequencing using the Sanger capillary method.

[00119] Accordingly, the present invention provides methods to prepare the oligonucleotide of the invention, to detecting and/or identifying at least two polynucleotides from a nucleic acid-protein complex, and to prepare vectors comprising the oligonucleotides and concatenated oligonucleotides of the present invention.

Thus, in another aspect, the present invention provides a method of preparing at least one isolated oligonucleotide comprising:

- (a) providing a nucleic acid-protein complex;
- (b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first and second tags are obtained from a nucleic acid-protein complex; and
- (c) isolating the oligonucleotide.

[00120] In particular, there is provided a method of preparing at least one isolated oligonucleotide comprising:

- (a) providing a nucleic acid-protein complex;

(b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex; and

(c) isolating the oligonucleotide.

[00121] In one embodiment, step (b) of this aspect of the invention comprises:

(i) inserting at least one linker comprising at least one restriction enzyme recognition site, and

(ii) cleaving the first and second polynucleotides with a least one restriction enzyme recognizing the at least one recognition site in the linker to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide and the linker between the tags.

[00122] In another embodiment, step (b) of this aspect of the invention comprises:

(i) inserting at least one linker comprising at least one restriction enzyme recognition site between the first polynucleotide and the second polynucleotide of the complex;

(ii) adding at least one restriction enzyme recognition site each to the 5' terminus of the first polynucleotide and the 3' terminus of the second polynucleotide;

(iii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing at least one recognition site to obtain cleaved fragments; and

(iv) ligating cleaved fragments to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide, the tags comprising 5' and 3' termini of each polynucleotide, and the linker inserted between the tags and the at least one restriction enzyme recognition site in step (ii) is part of a vector.

[00123] The nucleic acid-protein complex may be obtained by chromatin immunoprecipitation.

[00124] In another aspect, the present invention provides a method of detecting and/or identifying at least two polynucleotides from a nucleic acid-protein complex, the method comprising:

(a) providing a nucleic acid-protein complex;

- (b) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex;
- (c) sequencing the oligonucleotide; and
- (d) mapping the at least two polynucleotides based on the nucleotide sequences of the first and second tags, thereby detecting and/or identifying the at least two polynucleotides.

[00125] In one embodiment, step (b) of this aspect of the invention comprises:

- (i) inserting at least one linker comprising at least one restriction recognition site, and
- (ii) cleaving the first and second polynucleotides with a least one restriction enzyme recognizing the at least one recognition site in the linker to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide and the linker between the tags.

[00126] In another embodiment, step (b) of this aspect of the invention comprises:

- (i) inserting at least one linker comprising at least one restriction enzyme recognition site between the first polynucleotide and the second polynucleotide of the complex;
- (ii) adding at least one restriction enzyme recognition site each to the 5' terminus of the first polynucleotide and the 3' terminus of the second polynucleotide;
- (iii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing at least one recognition site to obtain cleaved fragments; and
- (iv) ligating cleaved fragments to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide, the tags comprising 5' and 3' termini of each polynucleotide, and the linker inserted between the tags; the at least one restriction enzyme recognition site in step (ii) may be part of an adaptor or a vector.

[00127] The nucleic acid-protein complex under the present invention may be obtained by chromatin immunoprecipitation. The nucleic acid-protein complex may be obtained by incorporating a photoactivatable moiety into the DNA and/or the

protein of interest and isolation of DNA/protein complex by antibody-mediated precipitation or by affinity-mediated technique. Examples of such affinity-based techniques include streptavidin/ biotin, glutathione-S-transferase/glutathione matrix, and maltose-binding protein/amylose matrix interactions.

[00128] The oligonucleotide obtained in step (b) may be amplified before being sequenced in step (c); the amplification may be by polymerase chain reaction. The amplified oligonucleotides may be subjected to at least one purification step after amplification but before being sequenced in step (c). The at least one purification step may be gel electrophoresis.

[00129] The oligonucleotide according to the invention may be concatenated with at least one other oligonucleotide obtained by steps (a) to (b) before being sequenced to make a concatemer of oligonucleotides.

[00130] The sequencing may be by the Sanger method or by multiplex sequencing. The multiplex sequencing may be pyrosequencing. Any suitable method of sequencing such as those described by Bonetta (2006) may be employed. The nucleic acid fragment to which a protein of interest binds may be any nucleic acid fragment comprising a region to which a protein of interest binds, for example, a histone binding site. The polynucleotides may be DNA or RNA.

[00131] The polynucleotides may be located on the same chromosome or the polynucleotides may be located on different chromosomes.

[00132] The detecting and/or identifying may be for transfection or translocation of the polynucleotides.

[00133] The oligonucleotide may be transfected into a cell. The transfection may be by electroporation. The cell may be a bacterial cell.

[00134] The embodiments of this aspect of the invention are described in more detail below as the CIA-PET and CIA-diPET techniques below.

[00135] In another aspect, the present invention is a vector comprising the oligonucleotide, the concatemer of oligonucleotides or the library of oligonucleotides of the present invention.

CIA-PET

[00136] In the CIA-PET method (FIGS. 1 and 4), DNA fragments tethered by protein in DPD complexes will be joined by linker sequences through ligation. The

linker sequence contains two Mmel sites. After reversal of the cross-linking, the ligated DNA will be digested by Mmel to release the paired end ditag (CIA-PET). Each of the CIA-PETs contains a linker with two flanked tags (about 20bp each) (Step 2 labeled in FIG. 1). The two tags contained in a CIA-PET thus represent two distal genetic regions that are far away from each other in linear genome sequence, but interacting together and mediated by a specific protein such as a histone protein.

[00137] After the CIA-PETs are gel-purified, sequence specific adaptors are added to each side of the CIA-PETs, which are then amplified by PCR (Step 3 in FIG. 1). The amplified CIA-PETs may be directly sequenced by a multiplex sequencing method such as the "454" pyrosequencing method or any other suitable sequencing methods, or concatemerized into longer DNA stretches for cloning to make CIA-PET library and followed by sequencing using the Sanger capillary method. The CIA-PET sequences will be mapped to the reference genome sequences. True chromatin interaction sites may be identified based on frequent occurrence of CIA-PET clusters in specific loci, which will be distinguished from background noise randomly scattered as singletons (FIG. 3).

CIA-diPET

[00138] In the CIA-diPET method (FIGS. 2 and 6), paired end ditag sequences representing each of the two related DNA fragments are extracted and sequenced compared to only one tag per sequence. As such, the resulting CIA-diPET will contain longer tag sequences than the tags of the CIA-PET method. It will be appreciated by a person skilled in the art that the CIA detection method of the present invention is thus more specific in mapping chromatin interactions than other methods of the prior art. The DNA fragments in DPD complexes are joined by a linker sequence that contains a Mmel site and a Gsu1 site. After reverse cross-linking, the ligated DNA will be randomly broken by sonication.

[00139] The DNA will then be size fractionated and cloned into a pGIS8 vector (Step 2 of FIG. 2) that contains an immediate Mmel site and a Gsu site at its cloning site (FIG. 7). After transformation and propagation in bacteria cells, the library clones are subjected by sequential manipulations of digestion, self-ligation, and transformation to create the single diPET library. The plasmid DNA of the

single diPET library will be digested by BamH1 to release the diPET structure (Step 4 of FIG. 2), which may be directly sequenced using the MS-PET sequencing method, or further concatenated for cloning and sequencing. The CIA-diPET sequences will be mapped to the reference genome sequences. True chromatin interaction sites may be identified based on frequent occurrence of CIA-diPET clusters in specific loci, which will be distinguished from background noise randomly scattered as singletons (FIG. 5).

[00140] Having now generally described the invention, the same will be more readily understood through reference to the following examples that are provided by way of illustration and are not intended to be limiting of the present invention.

[00141] For quality control and to ensure correct insertion of linkers or adaptors, gel electrophoresis is performed to remove unwanted fragments (FIGS. 8 to 10) during the process.

Examples

Methodology of CIA

[00142] The Nanog transcription factor in mouse ES cell (E14) and ChIP-PET data for Nanog and other key ES cell transcription factors such as Oct4 and Sox2 were used as the biological system to practice the present invention. The ChIP-PET data provides a linear map of where these transcription factors bind to, and will be used as reference to validate the chromatin interaction data. The interactions between chromatins were detected by two methods, CIA-PET and CIA-diPET.

[00143] The sequences used in the examples are as below, and in the sequence listing. Where two strands are shown, the top strand is the sense strand and the bottom strand is the antisense strand. All sequences are shown in the 5' to 3' direction. Where a nucleotide is denoted by N or n (or X or a numeral in the Figures), it means that any nucleotide (A, C, G or T) may be represented in that position.

M&M linker

5' GTCGGAGGCCAAGGCGGCCGTACGTCCAACT 3' (31 nt) (SEQ ID NO:1)

5' GTTGGACGTACGGCCGCCTTGGCCTCCGAC 3' (31 nt) (SEQ ID NO:2)

5' ends are phosphorylated

M&G linker

5' GTCGGAGGCCAAGGCGGCCGTACGCTGGAGT 3' (31 nt) (SEQ ID NO:3)

5' CTCCAGCGTACGGCCGCTTGGCCTCCGACT 3' (31 nt) (SEQ ID NO:4)

5' ends are phosphorylated

P1 concatenating adapter (PMR 011)

5' GGATCCCTTAATCGCCTTGCAGCACATC 3' (28 nt) (SEQ ID NO:5)

5' GATGTGCTGCAAGGCGATTAAGGGATCCNN 3' (30 nt) (SEQ ID NO:6)

5' end of top adapter is phosphorylated

5' end of bottom adapter is not phosphorylated

P2 concatenating adapter (PMR 012)

5' GGATCCCCTGTGTGAAATTGTTATCCGCT 3' (29 nt) (SEQ ID NO:7)

5' AGCGGATAACAATTTACACAGGGGATCCNN 3' (31 nt) (SEQ ID NO:8)

5' end of top adapter is phosphorylated

5' end of bottom adapter is not phosphorylated

D1 diPETing adapter (PMR 011)

5' GGATCCCTTAATCGCCTTGCAGCACATC 3' (28 nt) (SEQ ID NO:9)

5' GATGTGCTGCAAGGCGATTAAGGGATCCNN 3' (30 nt) (SEQ ID NO:10)

5' end of top adapter is phosphorylated

5' end of bottom adapter is not phosphorylated

D2 diPETing adapter (PMR 012)

5' GGATCCAATGCTCCTCCCTGTGTGAAATTGTTATCCGCT 3' (39 nt) (SEQ ID NO:11)

5' AGCGGATAACAATTTACACAGGGAGGAGCATTGGATCCNN 3' (41 nt) (SEQ ID NO:12)

5' end of top adapter is phosphorylated

5' end of bottom adapter is not phosphorylated

PMR011 primer

5' GATGTGCTGCAAGGCGATTAAG 3' (22 nt) (SEQ ID NO:13)

5' ends are phosphorylated

PMR012 primer

5' AGCGGATAACAATTTACACAGG 3' (23 nt) (SEQ ID NO:14)

5' ends are phosphorylated

RecA selection oligo

5' AGTCGGAGGCCAAGGCGGCCGTACGCTGGAGT 3' (SEQ ID NO:15)

(Biotinylated)

pGIS8 vector

This vector was derived from a pGEM vector (Promega) and contains the multiple unique cloning sites for the following restriction enzymes in this order:

BamH1→Mme1 ; Cleavable region ; Gsu1→BamH1→BseR1

The following oligonucleotides are used:

5' AATTGGATCCGACTCGAGGATGAATTCTCCAGGATCCCTCCTC 3' (43 nt)
(SEQ ID NO:16)

5' TCGAGAGGAGGGATCCTGGAGAATTCATCCTCGAGTCGGATCC 3' (43 nt)
(SEQ ID NO:17)

5' ends are phosphorylated.

The rest of the vector do not contain any BseR1, BamH1, Mme1 or Gsu1 sites.

The pGIS8 plasmid is one example of a vector that may be used (FIG. 7). The sequence of the sense strand is given in SEQ ID NO:18 and the sequence of the antisense strand is SEQ ID NO:19. A multiple cloning site containing restriction enzyme recognition sites (represented by SEQ ID NOS: 16 and 17) is inserted into the vector. The sequence listing for pGIS8 showing the various restriction enzyme recognition sites are shown in FIG. 11. Any other vector fulfilling the requirements may be prepared and used by any skilled person in the art.

M&M adapter PET

5'NNNNNNNNNNNNNNNNNNNAGTCGGAGGCCAAGGCGGCCGTACGTCCAACTN
NNNNNNNNNNNNNNNNNNNN 3' (69 nt) (SEQ ID NO:20)

5'NNNNNNNNNNNNNNNNNNNAGTTGGACGTACGGCCGCCTTGGCCTCCGACT
NNNNNNNNNNNNNNNNNNNN 3' (69 nt) (SEQ ID NO:21)

M&M adapter PET with adapter sequences

5'GATGTGCTGCAAGGCGATTAAGGGATCCNNNNNNNNNNNNNNNNNNNNNNNNNAGT
CGGAGGCCAAGGCGGCCGTACGTCCAACTNNNNNNNNNNNNNNNNNNNNNNNGGA
TCCCCTGTGTGAAATTGTTATCCGCT 3'(128 nt) (SEQ ID NO:22)

5' AGCGGATAAC AATTTACACAC AGGGGATCCNNNNNNNNNNNNNNNNNNNNNNNA
GTTGGACGTACGGCCGCCTTGGCCTCCGACTNNNNNNNNNN NNNNNNNNNNN
GGATCCCTTA ATCGCCTTGCAGCACATC 3' (128 nt) (SEQ ID NO:23)

M&M final PET (after cleavage of adapter sequences)

5'GATCCNNNNNNNNNNNNNNNNNNNNNNNNNAGTCGGAGGCCAAGGCGGCCGTACG
TCCAACTNNNNNNNNNNNNNNNNNNNNNNNNNG (77 nt) (SEQ ID NO:24)

5' GATCCNNNNNNNNNNNNNNNNNNNNNNNNNAGTTGGACGTACGGCCGCCTTGG
CCTCCGACTNNNNNNNNNNNNNNNNNNNNNNNG 3' (77 nt) (SEQ ID NO:25)

M&G adapter DNA sequences after sonication

5'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGTCGGAGGCCAAGGCGGCCGTACGC
TGGAGTNN 3' (SEQ ID NO:26)

5' NNNACTCCAGCG TACGGCCG
CCTTGGCCTCCGACTNN 3' (SEQ ID NO:27)
(variable nt)

M&G adapter DNA sequences after cloning into pGIS8 plasmids

(note: rest of the plasmid is not shown)

5'GGATCCGACNN
NNNNNNNNNNNAGTCGGAGGCCAAGGCGGCCGTACGCTGGAGTNNNNNNNNNN
NNCTCCAGGATCC 3' (140
nt) (SEQ ID NO:28)

5' GGATCCTGGAGNN
NNNNNNNNNNNNACTCCAGCGTACGGCCGCCTTGGCCTCCGACTNNNN
NN NGTCG
GATCC 3' (140 nt) (SEQ ID NO:29)

M PET intermediate

5' GGATCCGACNNAGT
CGGAGGCCAAGGCGGCCGTACGCTGGAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNCTCCAGGATCC 3' (132 nt) (SEQ ID NO:30)

5' GGATCCTGGAGNN
NNACTCCAGCGTACGGCCGCCTTGGCCTCCGACTNNNN
NN NGTCGATCC 3' (132 nt) (SEQ ID NO:31)

M and G diPET

5' ATCCGACNNAGT
CGGAGGCCAAGGCGGCCGTACGCTGGAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNCTCCAGGATCC 3' (116 nt) (SEQ ID NO:32)

5' GGATCCTGGAGNNACTCAGCGTA
CGGCCGCCTTGGCCTCCGACTNN
NN NNNNNNNNGTCGATC 3' (116 nt) (SEQ ID NO:33)

5' GATCCGACNNAGT
CGGAGGCCAAGGCGGCCGTACGCTGGAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNN NNNNNCTCCAG 3' (111 nt) (SEQ ID NO:34)

M&M adapter PET before Mmel digestion

[illegible][illegible]

Oligonucleotide of FIG. 8 inserted into plasmid (rest of plasmid not shown)

[illegible][illegible]

[illegible][illegible]

5' GATCCGANN
NN
NNNNNNNNNNNNNCTCCAG 3' (112 nt) (SEQ ID NO:44)

5' GATCCTGGAGNN
NN
NNNNNNNNNNNNNNNNNNNTCG 3' (112 nt) (SEQ ID NO:45)

**5' NNNNNNNNNNNNNNNNNNNNNNTCCGACNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNN 3' (59 nt) (SEQ ID NO:46)**

5' NNGTCGANNNNNNNNNNNN
NNNNNNNNNNNNNNNN 3' (59 nt) (SEQ ID NO:47)

[illegible]

5' GGATCCTGGAGNN
NNNGTCGANNNNNNNNNNN
NNNGTCGGATC
C 3' (140 nt) (SEQ ID NO:49)

[illegible]

5' GATCCGACNNNTCCGA
 CNN
 NNNNNNNNNNNNNNNNNNNNCTCCAG 3' (120 nt) (SEQ ID NO:52)

5' GATCCTGGAGNN
 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTCGGANNNNNNNNNNNNNNNNNN
 NNNNNNNNNNNNNNNNNNNNNNGTCG 3' (120 nt) (SEQ ID NO:53)

[illegible]

5' GGACTGGAGNN
NNNNNNNAGTCGGAGGCCAAGGCGGCCGTACGCTGGAGTNNNNNNNNNN
NNNGTCG
GATCC 3' (148 nt) (SEQ ID NO:56)

5' GGATCCGACNN
NNNNNNNNNNNNNNNNNNNACTCCAGCGTACGGCCGCCTTGCCCTCCGACT

Intermediate oligonucleotide 1 of FIG. 10

[illegible]

**5' GGATCCGACNN
NNNNNNCTCCAGTCC 3' (62 nt) (SEQ ID NO:61)**

5' GATCCTGGAGNN
NN
NNNNNNNNNNNNNNNNNNNTCG 3' (114 nt) (SEQ ID NO:63)

[00144] The CIA-PET method (FIGS. 1 and 4) comprises five sections: (1) Generate ChIP DNA-protein-DNA complex, (2) Prepare CIA-PET, (3) Amplify CIA-PET, (4) Sequence CIA-PET, and (5) Map the CIA-PET sequences. (FIG. 3)

(b) About 1 to 2×10^8 cells are collected and crosslinked with formaldehyde (final concentration of 1%; Sigma) for 10 minutes at room temperature.

(c) Cell lysis and chromatin preparation:

- c1. Cells are lysed in lysis buffer (50 mM HEPES, 1 mM EDTA, 0.15 M NaCl, 1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate, all from Ambion).
- c2. The chromatin is solubilized by sonication using a Branson 450 ultrasonic cell disruptor (20% duty power output, 30 second, 5 to 8 times;).
- c3. The chromatin is diluted 10 times to lower the SDS to 0.1%.
- c4. The extract is then cleared by centrifugation at 14,000 rpm for 10 minutes at 4°C.
- c5. This extract is stored at -80°C until use.

(d) Immuno-precipitation

- d1. Two microgrammes of of monoclonal antibody (F7, Santa Cruz) are bound to protein G sepharose (Pharmacia).
- d2. The antibody coated beads are incubated with the chromatin extract at 4°C for 16 hours.
- d3. The beads are then washed with the following buffers (reagents from Sigma Chemical Company):
 - Wash buffer 1 (50 mM HEPES, 1 mM EDTA, 0.15 M NaCl, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate) 2 times
 - Wash buffer 2 (50 mM HEPES, 1 mM EDTA, 0.5 M NaCl, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate) 1 time
 - Wash buffer 3 (20 mM Tris.HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 0.5% NP40, 0.5% sodium deoxycholate) 1 time
 - Wash buffer 4 (20 mM Tris.HCl pH 8.0, 1 mM EDTA) 1 time
- d4. The protein-DNA complexes are then eluted from the beads with elution buffer (50 mM Tris.HCl pH 8.0, 1 mM EDTA, 1 % SDS) for 20 min at 65°C.
- d5. The eluant is then dialyzed in PBS (Ambion) to remove SDS for 3 hours at 4°C.

[00146] (2) Prepare CIA-PET

(a) End repair

Perform end-repair with the Epicentre End-It kit and chemicals:

Chromatin (up to 5 ug)

2.5 ul 10x

End Repair Buffer	5 ul
2.5 mM dNTP Mix	5 ul
10 mM ATP	5 ul
End-Repair Enzyme Mix	1 ul
Nuclease free water	to 50 ul

Vortex briefly to mix and then incubate at room temperature for 45 minutes, stop reaction by heating at 70°C for 10 minutes. Adjust the concentration to 65ng/ul by adding 27ul of nuclease free water.

(b) A-tailing

DNA	20 ul
10 mM dATP (Roche)	0.5 ul
10x ExTaq buffer (Takara)	2.5 ul
ExTaq polymerase (Takara)	0.5 ul
Nuclease free water	to 25 ul

Incubate in a PCR machine at 72°C for 30 min, then 4°C. It is best to take out the tubes and perform the ligation immediately once incubation at 72°C ends.

(c) Ligate DNA with M&M linker (SEQ ID NOS: 1 and 2)

Chromatin DNA (200ug)	3.1 ul
M&M adapter – T tailed (38ng)	3.1 ul
5X ligase buffer with PEG (Invitrogen)	6 ul
T4 ligase (5U/ul) (Invitrogen)	1 ul
Nuclease free water	12.3ul

Ligate overnight at 16°C to obtain oligonucleotides (SEQ ID NOS: 20 and 21) inserted with M&M linker.

(d) Reverse cross-links with Proteinase K

Divide DNA sample into 20 ul aliquots, and reverse cross-links by overnight incubation at 65°C in the presence of 15 ul of 20 mg/ml proteinase K (Ambion). The next day, add 1 ul of 10 mg/ml RNase A (Qiagen) to degrade RNA is degraded 45 min at 37°C, followed by phenol extraction and ethanol

precipitation of DNA. Resuspend in 20 ul of Elution Buffer and store at -20°C. The concentration obtained is usually 500 ng/ul.

Quantitate DNA and perform quality control by running 0.5 ul of the DNA obtained on a 1% gel with Takara Wide-Range Ladders and Invitrogen Low-Mass ladders and 0.5 ul of material from (a). The material from (d) should be less smeared on the gel and should not show any bright band at around the 50-100 bp mark.

(e) Mme1 cut

10 ug DNA	20 ul
10x NEBuffer 4 (New England Biolabs)	20 ul
10x SAM (New England Biolabs)	20 ul
Mme1 (2U/ul) (New England Biolabs)	20 ul
Nuclease-free water	120 ul

Split into two tubes and incubate overnight at 37°C. SAM should be prepared fresh.

Phenol chloroform and ethanol precipitation with glycoblu.

Quantitate DNA and perform quality control by running 3 ul on a 2% gel or PAGE gel together with Takara Wide-Range Ladders and Invitrogen Low-Mass ladders and with 0.5 ul of material from step (d).

(f) Gel purification

Load DNA obtained onto a 2% agarose gel in a Scie-Plas medium-sized unit (60 ul per well) together with appropriate ladders, eg 20 ul of Takara Wide Range ladder. Run at 80V, approximately 1.5 hours, and visualize at 365 nm UV. Oligonucleotides with incorrectly inserted linkers will be removed by the gel electrophoresis (for examples of incorrectly inserted linkers, see FIGS. 8 to 10).

Excise ditag band and electroelute using disposable Fermentas ElutaTubes according to manufacturer's recommendations. Electroelution is performed for 1-1.5 hours at 90 V and the harvested ditags are ethanol precipitated thus:

For every 200 ul eluate, add

3 M NaOAc pH 5.2 (Amresco)	20 ul
1 M MgCl ₂ (Ambion)	4.5 ul
Glycoblu (Ambion)	2 ul

100% ethanol

800 ul

Resuspend precipitated ditags in 12 ul Elution Buffer (Qiagen) and run 2-5 ul on a 4-20% PAGE minigel together with Low Mass Ladders (Invitrogen) for purity check and visual quantification. Correctly obtained ditags have the sequences given in SEQ ID NOS: 22 and 23).

[00147](3) Amplify PET

(a) Adapter ligation (Concatenating adaptors SEQ ID NOS: 5 to 8)

DNA (100 ng)	6 ul
Adapter (10 ug)	6 ul
10x ligase buffer (with spermidine)	1.5 ul
T4 DNA ligase (5U/ul, Invitrogen)	1 ul
Nuclease Free Water	0.5 ul

The total volume is 15 ul. Incubate at 16°C for 16 hours.

10x ligation buffer with Spermidine is made up of:

60mM Tris-HCl pH7.5 (Ambion)
 60mM MgCl₂ (Ambion)
 50mM NaCl (Ambion)
 1mg/ml BSA (New England Biolabs)
 70mM Beta-mercaptoethanol (Sigma)
 1mM ATP (Invitrogen)
 20mM DTT (Invitrogen)
 10mM spermidine (Sigma)

(b) PCR amplification

Amplify with primers PMRs 11 and 12 (SEQ ID NOS: 13 and 14 respectively) and the Hotstartaq kit from Qiagen:

DNA	1 ul
10x PCR buffer (Qiagen)	10 ul
dNTP mix (10 mM of each) (Invitrogen)	2 ul
PMR11	1 ul (0.2 uM)

PMR12	1 ul (0.2 uM)
HotStarTaq DNA Polymerase (Qiagen)	0.5 ul
Nuclease Free Water	to 100 ul

Mix well. Incubate in PCR machine:

1. 15 minutes, 95°C
2. 0.5 min 94°C
3. 0.5 min 55°C
4. 1 min 72°C
5. Repeat from step 2 25 times.
6. 10 min 72°C.

Purify using PCR purification kit (Qiagen).

[00148](4) Sequence the CIA-PETs

The CIA-PETs may be directly sequenced according to the protocols for the 454 multiplex sequencing machine (454 life sciences). The technique is taught in Margulies et al (2005) and US Application No. 20030068629. These references are hereby incorporated in their entirety by reference.

[00149](5) Map the CIA-PETs

Mapping may be performed using the Compressed Suffix Array. Multiple joins across two different DNA fragments ($n > 3$) shall be taken to represent real distal control regions (FIG. 3).

Example 3 - Method of CIA-diPET

[00150] The CIA-diPET method (FIGS. 2 and 6).comprises the following sections: (1) Generate ChIP DNA-protein-DNA complex, (2) Prepare CIA library (3) Prepare CIA-PET library (4) Sequence and (5) Map the CIA-PET sequences (FIG. 5).

[00151](1) Generate ChIP DNA-protein-DNA complex (as above)

(a) Mouse embryonic stem (ES) cells are cultured under feeder-free conditions in the presence of leukemia inhibitory factor (Chemicon).

(b) About 1 to 2 x 10⁸ cells are collected and crosslinked with formaldehyde (final concentration of 1%; Sigma) for 10 minutes at room temperature.

(c) Cell lysis and chromatin preparation:

- c1. Cells are lysed in lysis buffer (50 mM HEPES, 1 mM EDTA, 0.15 M NaCl, 1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate, all from Ambion).
- c2. The chromatin is solubilized by sonication using a Branson 450 ultrasonic cell disruptor (20% duty power output, 30 second, 5 to 8 times;).
- c3. The chromatin is diluted 10 times to lower the SDS to 0.1%.
- c4. The extract is then cleared by centrifugation at 14,000 rpm for 10 minutes at 4°C.
- c5. This extract is stored at -80°C until use.

(d) Immuno-precipitation

- d1. Two micrograms of monoclonal antibody (F7, Santa Cruz) are bound to protein G sepharose (Pharmacia).
- d2. The antibody coated beads are incubated with the chromatin extract at 4°C for 16 hours.
- d3. The beads are then washed with the following buffers (reagents from Sigma Chemical Company):
 - Wash buffer 1 (50 mM HEPES, 1 mM EDTA, 0.15 M NaCl, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate) 2 times
 - Wash buffer 2 (50 mM HEPES, 1 mM EDTA, 0.5 M NaCl, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate) 1 time
 - Wash buffer 3 (20 mM Tris.HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 0.5% NP40, 0.5% sodium deoxycholate) 1 time
 - Wash buffer 4 (20 mM Tris.HCl pH 8.0, 1 mM EDTA) 1 time
- d4. The protein-DNA complexes are then eluted from the beads with elution buffer (50 mM Tris.HCl pH 8.0, 1 mM EDTA, 1 % SDS) for 20 min at 65°C.
- d5. The eluant is then dialyzed in PBS (Ambion) to remove SDS for 3 hours at 4°C.

[00152](2) Prepare CIA-diPET library

[00153] (a) End repair (as for preparation of CIA-PET), perform end-repair with the Epicentre End-It kit:

Chromatin (up to 5 ug)	2.5 ul
10x End Repair Buffer (Epicentre)	5 ul
2.5 mM dNTP Mix (Epicentre)	5 ul
10 mM ATP (Epicentre)	5 ul
End-Repair Enzyme Mix (Epicentre)	1 ul
Nuclease Free Water	31.5 ul

Briefly vortex to mix, then incubate at room temperature for 45 minutes, stop by heating at 70°C for 10 minutes. Adjust the concentration to 65 ng/ul by adding 27 ul of Nuclease Free Water.

(b) A-tailing, as above for CIA-PET

DNA	20 ul
10 mM dATP (Roche)	0.5 ul
10x ExTaq buffer (Takara)	2.5 ul
ExTaq polymerase (Takara)	0.5 ul
Nuclease Free Water	1.5 ul

Incubate in a PCR machine at 72°C for 30 minutes, then 4°C forever. It is best to take out the tubes and perform the ligation immediately once the incubation at 72°C ends.

(c) Ligate DNA with M&G linker (SEQ ID NOS: 3 and 4)

Chromatin DNA (200 ng)	3.1 ul
M&G adapter- T tailed (38 ng)	7.6 ul
5x ligase buffer with PEG (Invitrogen)	6 ul
T4 ligase (5U/ul) (Invitrogen)	1 ul
Final volume	12.3 ul

Ligate overnight at 16°C to obtain oligonucleotides (SEQ ID NOS: 26 and 27) inserted with M&G linker.

(d) Reverse cross-links with proteinase K, as above for CIA-PET.

Divide DNA sample into 20 ul aliquots, and reverse cross-links by overnight incubation at 65°Celsius in the presence of 15 ul of 20 mg/ml proteinase K (Ambion). The next day, add 1 ul of 10 mg/ml RNase A (Qiagen) to degrade RNA is degraded 45 min at 37°Celsius, followed by phenol extraction and ethanol precipitation of DNA. Resuspend in 20 ul of Elution Buffer.and store at -20°C. The concentration is usually 500 ng/ul.

Quantitate and perform quality control check by running 0.5 ul on a 1% gel together with Takara Wide-Range Ladders and Invitrogen Low-Mass ladders and 0.5 ul of material from (a). The material from (d) should be less smeared out on the gel as well as not show any bright band at around the 50-100 bp mark.

(e) Digest with Nla III

Store Nla III at -80°C (half-life is 6 months at -80°C). Place on ice just before use.

DNA (approximately 1 ug)	2 ul
10x NEBuffer 4 (New England Biolabs)	5 ul
NlaIII (New England Biolabs)	1 ul
100x BSA (New England Biolabs)	0.5 ul
Nuclease-Free Water	41.5 ul

Prepare five tubes of the above reaction. Incubate for 1 hour at 37°C. Phenol chloroform ethanol precipitate with glycoblue, resuspend in 10 ul of Elution Buffer (Qiagen).

(f) Polish ends with the Epicentre End-It kit:

DNA (up to 5 ug)	34 ul
10x End Repair Buffer (Epicentre)	5 ul
2.5 mM dNTP Mix (Epicentre)	5 ul
10 mM ATP (Epicentre)	5 ul
End-Repair Enzyme Mix (Epicentre)	1 ul

Incubate at room temperature for 45 minutes; stop by heating at 70°C for 10 minutes.

(i) Clone into a pGIS8 vector (SEQ ID NOS: 18 and 19; FIGS. 7 and 11) with Mme1 and Gsu1 flanking sites

Set up on ice, using 1.7 ml microfuge tubes (use multiple tubes):

40 ng/ul precut pGIS	1 ul
----------------------	------

DNA	6 ul
-----	------

5x ligase buffer with PEG (Invitrogen)	2 ul
--	------

T4 DNA ligase (5U/ul) (Invitrogen)	1 ul
------------------------------------	------

Incubate overnight (12-16 hours) at 16°C to obtain oligonucleotides represented by SEQ ID NOS: 28 and 29).

Also set up a vector self-ligation control.

Transform 1µl of the ligation reaction per 50µl of electrocompetent TOP10 cells (Invitrogen) by electroporation. Recover each aliquot in 1ml LB media at 37°C for 1hr, then plate out in a series of several dilutions (on LB agar + ampicillin) for quality control and titering.

Then scale up this process by plating the remaining culture on large agar plates (Q-trays), and performing maxipreps using the Qiagen HiSpeed Plasmid Maxi kit.

[000153] (3) CIA-diPET library

(a) Mme1 cut

10 ug DNA	100 ul
-----------	--------

10x NEBuffer 4 (New England Biolabs)	20 ul
--------------------------------------	-------

10x SAM (New England Biolabs)	20 ul
-------------------------------	-------

Mme1 (2U/ul) (New England Biolabs)	12 ul
------------------------------------	-------

Nuclease-free water	48 ul
---------------------	-------

Incubate overnight at 37°C to obtain oligonucleotides represented by SEQ ID NOS: 30 and 31. SAM should be prepared fresh.

Phenol chloroform and ethanol precipitation with glycobblue, resuspend in 12 ul of Elution Buffer.

Quantitate DNA and perform quality control check by running 1 ul on a 2% gel or PAGE gel together with Takara Wide-Range Ladders and Invitrogen Low-Mass ladders and with 1 ul of material from step (e).

(b) Circularization

Set up a 96 well plate with each well containing the following (MJ Research):

100 ng DNA	50 ul
------------	-------

Ligation Solution 1 (Takara Ligation Kit ver 2)	50 ul
---	-------

Seal plate tightly to prevent evaporation.

Incubate at 16°C, overnight.

Perform PCR purification (Qiagen) with three columns, resuspending in 40 ul of Elution Buffer each, giving rise to a total of approximately 120 ul.

(c) Gsu1 cut

Set up 9 tubes of:

Circularized DNA	12 ul
------------------	-------

10x buffer TANGO (Fermentas)	8.6 ul
------------------------------	--------

10x SAM (New England Biolabs)	8.6 ul
-------------------------------	--------

Gsu1 (5U/ul) (Fermentas)	1 ul
--------------------------	------

Nuclease Free Water	55.8 ul
---------------------	---------

Digest at 30°C for at least 2 h, but do not cut overnight. Oligonucleotides represented by SEQ ID NOS: 32 and 33 are obtained,

(d) Circularization

Set up 9 tubes containing the following:

Approx 100 ng DNA	50 ul
-------------------	-------

Ligation Solution 1 (Takara Ligation Kit ver 2)	50 ul
---	-------

Incubate at 16°C, overnight.

Phenol chloroform ethanol precipitate, resuspend in 12 ul of Elution Buffer.

(e) Amplification with Rolling Circle Amplification (Templiphi kit, Amersham)

Thaw solutions on ice. Prepare 3-4 tubes of the following:

2.5 ng of DNA	1 ul
---------------	------

Templiphi kit denature buffer (Amersham)	10 ul
--	-------

Heat at 95°C for 3 minutes, then briefly cool on ice.

To the reaction buffer, add:

Templiphi kit premix (Amersham)	10 ul
---------------------------------	-------

Mix well by tapping or gentle vortexing.

Incubate at 30°C, no shaking, 16 – 18 hours.

Examine the material, for example, by micropipetting 1 ul. It should be viscous.

Quantitate double-stranded DNA by picogreen fluorimetry (Quant-iT DNA assay kit, Molecular Probes)

(f) BamH1 cut

100 ug DNA	1 ul
10x unique BamHI buffer (New England Biolabs)	10 ul
100x Bovine Serum Albumin (New England Biolabs)	1 ul
BamHI (20 U/ul; 2-fold excess) (New England Biolabs)	10 ul
Nuclease Free Water	78 ul

Prepare more tubes as needed to digest the DNA; incubate at 37°C overnight to obtain oligonucleotides represented by SEQ ID NOS: 34 and 35.

(g) Gel purification

Load onto a 2% agarose gel in a Scie-Plas medium-sized unit (60 ul per well) together with appropriate ladders, eg 20 ul of Takara Wide Range ladder. Run at 80V, approximately 1.5 hours, and visualize at 365 nm UV. Oligonucleotides with incorrectly inserted linkers will be removed by the gel electrophoresis (for examples of incorrectly inserted linkers, see FIGS. 8 to 10).

Excise ditag band and electroelute using disposable Fermentas ElutaTubes according to manufacturer's recommendations. Electroelution is performed for 1-1.5 hours at 90 V and the harvested ditags are ethanol precipitated thus:

For every 200 ul eluate, add

3 M NaOAc pH 5.2	20 ul
1 M MgCl ₂	4.5 ul
Glycoblu	2 ul
100% ethanol	800 ul

Resuspend precipitated diPETs in 12 ul Elution Buffer and run 2-5 ul on a 4-20% PAGE minigel together with Low Mass Ladders (Invitrogen) for purity check and visual quantification.

[00154] (4) Sequence the CIA-diPET**(a) Concatenation of gel-purified BamHI-cohesive diPETs:**

CIA-diPETs 200-1000ng	6 ul
10x ligase buffer (with spermidine)	1ul
T4 DNA ligase (5U/ul) (Invitrogen)	1ul
Nuclease Free Water	2 ul

Incubate at 16°C for 2 hrs to overnight. Then heat-inactivate at 65°C for 10 min.

(b) Partial BamHI redigestion of CIA-diPET concatemers

Purify the concatemer DNA using the Qiagen PCR purification QuickSpin kit. Then quantitate the DNA by using 1 ul for Nanodrop (Nanodrop technologies), and do a short BamHI re-digest:

Concatemers	20ul
10x BamHI buffer (New England Biolabs)	3ul
BamHI (diluted to 1U/ul) (New England Biolabs)	0.2 ul
100x BSA (New England Biolabs)	0.5ul
Nuclease Free Water	6.3 ul

Incubate at 37°C for 30 min, not any longer.

Quickly add 6ul of loading dye, heat at 65°C for 15 min and chill on ice before loading on PAGE gel.

(c) PAGE purification of concatenated CIA-diPETs

Load the entire sample preferably into 1 well of a 4-20% gradient PAGE minigel, flanked by Takara Wide-Range and Invitrogen Low Mass ladders to allow sizing. Electrophorese at 200V for about 1hr. Stain for 15-30min in SYBR Green I, and visualize on the Dark Reader transilluminator (Clare Chemical) for gel excision.

(d). Excision of concatemers

Excise the concatenated DNA in 3 separate fractions, low (400-1000bp); medium (1000-2000bp) and high (>2000bp). Place the gel slice of each excised size-fraction into a 0.6ml microfuge tube that has been pierced at the bottom with a 21G needle. This pierced tube is placed inside a 1.7ml microfuge tube, and

centrifuged at 16110 g, 4°C for 5min. The gel pieces are thus conveniently shredded and collected in the bottom of each 1.7ml tube. Add 200µl of LoTE:NH₄OAc (167:33) (LoTE according to the recipe below, NH₄OAc from Ambion) to each tube and elute by heating at 65°C for 2hrs.

LoTE buffer:

3 mM Tris-HCl pH 7.5 (Ambion)

0.2 mM EDTA (Ambion)

Separate the supernatant (containing the eluted concatenated DNA) away from the gel pieces with the aid of microspin filter units as before, by spinning at 16110 g, 10min, 4°C. Perform phenol/chloroform extraction on each eluted size-fraction, then ethanol precipitate:

Eluted DNA fraction	200ul
3M Sodium Acetate pH 5.2	20ul
GlycoBlue	2.2ul
100% Ethanol	800ul

Keep at -80°C, 30min, then spin 16110 g, 4°C, 30min; wash 1x with 75% Ethanol. Resuspend the pellet in 6ul of LoTE buffer.

(e) Ligation to pZErO-1 vector

Prior to use, the pZErO-1 cloning vector is prepared by digesting 2ug of pZErO-1 plasmid DNA (Invitrogen) with 10 units of BamHI (New England Biolabs) for 2 hours at 37°C. The digested plasmid DNA is phenol chloroform extracted and ethanol precipitated, then resuspended in 60 ul of LoTE at a concentration of 33ng/µl. The plasmid may be validated by setting up a vector self-ligation as a control (there should be few colonies), as well as by running on an agarose gel.

Set up the ligation as follows:

Concatemer DNA fraction	6 ul
BamHI/pZErO-1	1 ul
5x ligase buffer (with PEG) (Invitrogen)	2ul
T4 DNA ligase (5U/µL) (Invitrogen)	1ul

Incubate at 16°C overnight. Do not heat inactivate.

Also set up self-ligation of the vector in parallel as a control. When preparing the vector for self-ligation, replace the concatemer DNA fraction with nuclease free water.

Purify each ligation reaction to remove salts before electroporation: adjust volume to 200ul with nuclease free water, perform phenol/chloroform extraction (pH 7.9) and ethanol precipitate (with GlycoBlue) at -80°C for 30min. Spin, wash the pellet at least twice with 70% ethanol, and resuspend in 12ul LoTE.

Add 1ul of the purified ligation reaction into 25ul electrocompetent cells (e.g. E. coli from Lucigen; Top10 from Invitrogen), in pre-chilled 1.7ml microfuge tubes. Do NOT pipette up and down to mix; instead, gently stir with the pipette tip. Stand on ice for 5min, then transfer to pre-chilled Biorad electroporation cuvettes (0.1cm gap).

Stand on ice for another 5min. Electroporate using Biorad Micropulser unit, single pulse, program EC1. The time constant is usually between 4.5 to 5 ms.

Add 1ml room temperature plain LB media within 10sec of pulsing; transfer to 15ml Falcon tube, and recover at 37°C, 1hr, 200rpm shaking.

Plate between 20-50ul (out of 1ml) on a small agar plate containing Low Salt LB agar plus Zeocin (Immedia Zeocin Agar, Invitrogen) and incubate overnight at 37°C.

(f) Library quality control

Count the numbers of colonies and determine library efficiency after eliminating the self-ligation background. Pick 24-48 colonies for PCR screening using primers PMR011 and PMR012. Include a control PCR on the pZErO-1 vector itself.

Based on the PCR results, pick 1 to 4x 96-well plates of colonies for overnight culture (in Low Salt LB + Zeocin, Immedia, Invitrogen), plasmid purification and sequencing to determine the average number of ditags per insert.

At this stage, the library can be stored in the form of purified ligation mix at -20°C, until one wishes to perform large scale transformations, plasmid extractions and sequencing.

(g) Sequencing library plating and colony picking

The transformed TOP10 (Invitrogen) bacteria cells were plated out on 22x22cm agar plates (Q-trays, Genetix) with colony density of less than 2,000 per plate to facilitate robotic picking. Individual colonies were picked and cultured in 384-well plates with LB plus Zeocin (see above) at 37°C overnight. Multiple copies of 384-well plates are replicated and stored in -80°C in the presence of 15% glycerol (Sigma).

(h) Template preparation

Plasmid DNA from the pZERO-1- derived clones is prepared using the Sprintprep solid-phase kit (Agencourt).

(i) DNA sequencing

Plasmids are sequenced using the sequencing primers PMR011 and PMR012 (SEQ ID NOS: 13 and 14 respectively) to sequence in both directions.

[00155] (5) Map the CIA-diPETs

Mapping may be performed using the Compressed Suffix Array. Multiple joins across two different DNA fragments ($n > 3$) shall be taken to represent real distal control regions (FIG. 5). More than 3 PETs have to show rearrangement of the same stretch of DNA – that is, tag 1 and tag 2 separated by a distance of more than 10 kB or located on different chromosomes, before any chimera consisting of tags that map to different locations of the genome, is taken to be representative of genomic rearrangements.

Example 4 - Preparation of M and G vector pGIS8 for use

(a) Obtain pGIS8 vector.

(b) Amplify by Rolling Circle Amplification as before (Templiphi kit, Amersham)

Thaw solutions on ice. Prepare 3-4 tubes of the following:

1 ng of DNA (from Maxiprep)	1 ul
Templiphi kit denature buffer	10 ul

Heat at 95°C for 3 minutes, then briefly cool on ice.

To the reaction buffer, add:

Templiphi kit premix	10 ul
----------------------	-------

Mix well by tapping or gentle vortexing.

Incubate at 30°C, no shaking, 16 – 18 hours.

Examine the material. It should be viscous. Quantitate DNA by picogreen fluorimetry (Quant-It kit, Molecular Probes).

(c) Perform restriction enzyme digest

DNA	25 ul
XhoI (20U/ul) (New England Biolabs)	1 ul
EcoRI (20U/ul) (New England Biolabs)	1 ul
10x Buffer EcoRI (New England Biolabs)	5 ul
100x BSA (New England Biolabs)	1 ul
Nuclease Free Water	17 ul

Incubate at 37°C for 16 hours.

Purify with PCR purification kit (Qiagen).

Test 5 ul on a 1% agarose gel.

(d) End-it blunt

Prepare two tubes of:

DNA	22.5 ul
10x End Repair Buffer (Epicentre)	5 ul
2.5 mM dNTP Mix (Epicentre)	5 ul
10 mM ATP (Epicentre)	5 ul
End-Repair Enzyme Mix	1 ul
Nuclease Free Water	11.5 ul

Incubate at room temperature for 45 minutes, stop by heating at 70°C for 10 minutes.

[00156] It will be appreciated that various modifications and improvements may be made by a person skilled in the art without departing from the spirit and scope of the present invention.

[00157] For example, the CIA method of the present invention may be used in other kinds of cells such as yeast cells instead of mammalian cells. Also, instead of using ChIP, the method may be performed by cross-linking with a suitable fixative without need for immunoprecipitation. In this variation, the protocol is:

1. Harvest cells.
2. Cross-link with formaldehyde for 10 min at 36°C.
3. Cell lysis with bead-beater followed by centrifugation to obtain a supernatant.
4. Shear DNA by sonication, hydroshearing, repeated drawing through a hypodermic syringe needle or restriction enzyme digestion.
5. Remove unwanted proteins with SDS and Triton-X treatment.

Thereafter, the DNA may be further processed (end blunting, ligation) as described above.

[00158] Further, in another variation, rolling circle amplification may be used instead of PCR. In such a variation, the protocol following removal of unwanted proteins is:

1. End-blunting of DNA.
2. A-tailing of DNA.
3. Ligation.
4. Rolling circle amplification; may be performed with a suitable commercial kit such as the Templiphi kit from Amersham Biosciences.
5. Quantification of DNA; may be performed with the Invitrogen/Molecular Probes' PicoGreen fluorimetry kit.
6. Digestion with Mmel restriction enzyme to obtain the isolated oligonucleotides.

[00159] Further, in another variation, the A-tailing step may be omitted and a suitable blunt-ended adaptor may be used.

References

- Antequera and Bird (1993), Proc Natl Acad Sci U S A. 90(24): 11995-9
- Ausubel (1995) Current Protocols in Molecular Biology, Vol. 2, 1995, Ed. Ausubel, et al., Greene Publish. Assoc. & Wiley Interscience, Unit 3.1.15
- Bonetta (2006) Nature Methods 3(2): 141-147.
- Brenner et al (2000)., Nat Biotechnol, 2000. 18(6): p. 630-4
- Buck and Lieb (2004) Genomics 83(3):349-60
- Dekker J, Rippe K, Dekker M and Kleckner N (2002) Science 295 (5558):1306-11.
- Dunn et al (2002) Genome Research 12(11):1756-1765
- Euskirchen et al., Mol Cell Biol, 2004. 24(9): p. 3804-14
- Li and Chandrasegaran (1993) Proc. Nat. Acad. Sciences USA 90:2764-8
- Lieb et al., Nat Genet, 2001. 28(4): p. 327-34
- Margulies et al, 2005 Nature 437, 376-380 (15 Sep 2005)
- New England Biolabs Catalog 2005. New England Biolabs (Ipswich, MA)
- Szybalski, W., 1985, Gene, 40:169
- Taverner et al., Genome Biol, 2004. 5(3):210.
- US 20050255501
- US 20050059022
- US 20030068629
- US 4,766,072
- Weinmann et al., Genes Dev, 2002. 16(2):235-44)

SEQUENCE LISTING

<110> Agency for science, technology and research

<120> Nucleic Acid Interaction Analysis

<130> FP3534

<160> 63

<170> PatentIn version 3.4

<210> 1

<211> 31

<212> DNA

<213> Artificial

<220>

<223> M&M linker sense strand; 5' end is phosphorylated

<400> 1

gtcggaggcc aaggcggccg tacgtccaac t 31

<210> 2

<211> 31

<212> DNA

<213> Artificial

<220>

<223> M&M linker antisense strand; 5' end is phosphorylated

<400> 2

gttggacgta cggccgcctt ggcctccgac t 31

<210> 3

<211> 31

<212> DNA

<213> Artificial

<220>

<223> M&G linker sense strand; 5' end is phosphorylated

<400> 3

gtcggaggcc aaggcggccg tacgctggag t 31

<210> 4

<211> 31

<212> DNA

<213> Artificial

<220>

<223> M&G linker antisense strand; 5' end is phosphorylated

<400> 4

ctccagcgta cggccgcctt ggcctccgac t 31

<210> 5
<211> 28
<212> DNA
<213> Artificial

<220>
<223> P1 concatenating adaptor (PMR 011) sense strand; 5' end is phosphorylated

<400> 5
ggatccctta atcgcccttgc agcacatc 28

<210> 6
<211> 30
<212> DNA
<213> Artificial

<220>
<223> P1 concatenating adaptor (PMR 011) antisense strand; 5' end is not phosphorylated

<220>
<221> misc_feature
<222> (29)..(30)
<223> n is a, c, g, or t

<400> 6
gatgtgctgc aaggcgatta agggatccnn 30

<210> 7
<211> 29
<212> DNA
<213> Artificial

<220>
<223> P2 concatenating adaptor (PMR 012) sense strand; 5' end is phosphorylated.

<400> 7
ggatcccctg tgtgaaattg ttatccgct 29

<210> 8
<211> 31
<212> DNA
<213> Artificial

<220>
<223> P2 concatenating adaptor (PMR 012) antisense strand; 5' end not phosphorylated.

<220>

<221> misc_feature
<222> (30)..(31)
<223> n is a, c, g, or t

<400> 8
agcggataac aatttcacac aggggatccn n 31

<210> 9
<211> 28
<212> DNA
<213> Artificial

<220>
<223> D1 diPETing adaptor (PMR011) sense strand; 5' end phosphorylated

<400> 9
ggatccctta atcgcccttgc agcacatc 28

<210> 10
<211> 30
<212> DNA
<213> Artificial

<220>
<223> D1 diPETing adaptor (PMR011) antisense strand; 5' end not
phosphorylated.

<220>
<221> misc_feature
<222> (29)..(30)
<223> n is a, c, g, or t

<400> 10
gatgtgctgc aaggcgatta agggatccnn 30

<210> 11
<211> 39
<212> DNA
<213> Artificial

<220>
<223> D2 diPETing adaptor (PMR012) sense strand; 5' end phosphorylated.

<400> 11
ggatccaatg ctccctccctg tgtgaaattg ttatccgct 39

<210> 12
<211> 41
<212> DNA
<213> Artificial

<220>
<223> D2 diPETing adaptor (PMR012) antisense strand; 5' end not

phosphorylated.

<220>
<221> misc_feature
<222> (40)..(41)
<223> n is a, c, g, or t

<400> 12
agcggataac aatttcacac agggaggagc attggatccn n 41

<210> 13
<211> 22
<212> DNA
<213> Artificial

<220>
<223> PMR011; 5' end phosphorylated.

<400> 13
gatgtgctgc aaggcgatta ag 22

<210> 14
<211> 23
<212> DNA
<213> Artificial

<220>
<223> PMR012; 5' end is phosphorylated

<400> 14
agcggataac aatttcacac agg 23

<210> 15
<211> 32
<212> DNA
<213> Artificial

<220>
<223> RecA selection oligonucleotide; biotinylated.

<400> 15
agtcggaggc caaggcggcc gtacgctgga gt 32

<210> 16
<211> 43
<212> DNA
<213> Artificial

<220>
<223> Sense strand of multiple cloning site stuffer for pGIS8 vector.

<400> 16
aattggatcc gactcgagga tgaattctcc aggatccctc etc 43

<210> 17
 <211> 43
 <212> DNA
 <213> Artificial

<220>

<223> Antisense strand of multiple cloning site stuffer for pGIS8 vector.

<400> 17

tcgagaggag ggatcctgga gaattcatcc tcgagtcgga tcc 43

<210> 18
 <211> 3541
 <212> DNA
 <213> Artificial

<220>

<223> Sense strand of pGIS8 vector.

<400> 18

gggcgaattc gatatcgcg cgcgcctgg ataaagtcag cagcttcac gccagcttca 60
 cacaaaaagt gactgacggt agcggcgcg cggcgcagga aggtcagggc gatctgtggg 120
 tgaaacgtcc aaacttattc aactggcata tgacacaacc tgatgaaagc attctggttt 180
 ctgacggtaa aacactgtgg ttctataacc cgttcgttga gcaagctacg gcaacctggc 240
 tgaaagatgc caccggtaat acgccgttta tgctgattgc ccgcaaccag tccagcgact 300
 ggcagcagta caatatcaaa cagaatggcg atgactttgt cctgacgccg aaagccagca 360
 atggcaatct gaagcagttc accattaacg tgggacgtga tggcacaatc catcagttta 420
 gcgcggtgga gcaggacgat cagcgcagca gttatcaact gaaatcccag caaaatgggg 480
 ctgtggatgc agcgaattt accttcaccc cgcgcgaagg cgtcacggta gatgatcaac 540
 gtaagtagag gcacctgagt gagcaatctg tcgctcgatt ttcggataa tacttttcaa 600
 cctctggccg cgcgtatgcg gccagaaaat ttagcacagt atatcgcca gcaacatttg 660
 ctggctgcg ggaagccgtt gccgcgcgct atcgaagccg ggcatattaca ttctatgatc 720
 ctctgggggc cgcgggtac cggcaaaaca actctcgctg aagtgattgc ccgctatgcg 780
 aacgctgatg tggaacgtat ttctgccgta agtcgaattg gatccgactc gaggatgaat 840
 tctccaggat ccctcctctg agtattctat agtgtcacct aaatagcttg gcgtaatcat 900
 ggtcatagct gtttcctgtg tgaaattgtt atccgctcac aattccacac aacatacgag 960
 ccggaagcat aaagtgtaaa gcctgggggtg cctaattgagt gagctaactc acattaattg 1020

cgttgcgctc	actgcccgt	ttccagtcgg	gaaacctgtc	gtgccagctg	cattaatgaa	1080
tcggccaaacg	cgcggggaga	ggcggtttgc	gtattgggcg	ctcttccgct	tcctcgctca	1140
ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcgg	atcagctcac	tcaaaggcgg	1200
taatacggtt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	gcaaaaaggcc	1260
agcaaaaaggc	caggaaccgt	aaaaaggccg	cgttgctggc	gtttttcgat	aggctccgcc	1320
cccctgacga	gcatcacaaa	aatcgacgct	caagtcagag	gtggcgaaac	ccgacaggac	1380
tataaaagata	ccaggcgttt	ccccctggaa	gctcccctcg	gcgctctcct	gtaccgaccc	1440
tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	ctttctcata	1500
gctcacgctg	taggtatctc	agttcggtgt	aggctggtcg	ctccaagctg	ggctgtgtgc	1560
acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	cttgagacca	1620
acccggtaag	acacgactta	tgcgcaactg	cagcagccac	tggtaacagg	attagcagag	1680
cgaggatatgt	aggcgggtgt	acagagttct	tgaagtgggtg	gcctaactac	ggctacacta	1740
gaaggacagt	atttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	aaaagagttg	1800
gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggttttttt	gtttgcaagc	1860
agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	tctacgggggt	1920
ctgacgctca	gtggaacgaa	aactcacgtt	aagggaattt	ggcatgaga	ttatcaaaaa	1980
ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	taaagtatat	2040
atgagtaaac	ttggtctgac	agttaccaat	gcttaatcag	tgaggcacct	atctcagcga	2100
tctgtctatt	tcgttcatcc	atagttgcct	gactccccgt	cgtgtagata	actacgatac	2160
gggagggctt	accatctggc	cccagtgtg	caatgatacc	gcgagaccca	cgctcaccgg	2220
ctccggattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	agtggctctg	2280
caactttatc	cgctccatc	cagtctatta	attgttgccg	ggaagctaga	gtaagtagtt	2340
cgccagttaa	tagtttgccg	aacgttggtg	gcattgctac	aggcatcgtg	gtgtcacgct	2400
cgtcgtttgg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	gttacatgat	2460
cccccatggt	gtgcaaaaaa	gcggttagct	ccttcgggtc	tccgatcggt	gtcagaagta	2520
agttggccgc	agtgttatca	ctcatgggtta	tggcagcact	gcataattct	cttactgtca	2580
tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	ttctgagaat	2640
agtgtatgcg	gcgaccgagt	tgctcttgcc	cggcgtcaat	acgggataat	accgcgccac	2700
atagcagaac	tttaaaagtg	ctcatcattg	gaaaacgttc	ttcggggcga	aaactctcaa	2760

```

ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc aactgatctt 2820
cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg 2880
caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc ctttttcaat 2940
attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt 3000
agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc cctgacgtct 3060
aagaaaccat tattatcatg acattaacct ataaaaatag gcgtatcacg aggccctttc 3120
gtctcgcgcg tttcgggtgat gacggtgaaa acctctgaca catgcagctc ccggagacgg 3180
tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg 3240
gtgttgggcg gtgtcggggc tggcttaact atgcggcatc agagcagatt gtactgagag 3300
tgcaccatat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac cgcacagggc 3360
gccattcgcc attcaggctg cgcaactgtt ggggaagggcg atcggtgcgg gcctcttcgc 3420
tattacgcca gctggcgaaa ggggggatgtg ctgcaaggcg attaagttgg gtaacgccag 3480
ggttttccca gtcacgacgt tgtaaaacga cggccagtga attgtaatac gactcactat 3540
a 3541

```

```

<210> 19
<211> 3541
<212> DNA
<213> Artificial

```

```

<220>
<223> Antisense strand of pGIS8 vector.

```

```

<400> 19
tatagtgagt cgtattacaa ttacttgcc gtcgttttac aacgtcgtga ctgggaaaac 60
cctggcggtta cccaacttaa tcgccttgca gcacatcccc ctttcgccag ctggcgtaat 120
agcgaagagg cccgcaccga tcgcccttcc caacagttgc gcagcctgaa tggcgaatgg 180
cgctgatgc ggtatTTTTt ccttacgcat ctgtgcggta tttcacaccg catatgggtg 240
actctcagta caatctgctc tgatgccgca tagttaagcc agccccgaca cccgccaaca 300
cccgtgacg cgccctgacg ggcttgctct ctcccgcat ccgcttacag acaagctgtg 360
accgtctccg ggagctgcat gtgtcagagg ttttcaccgt catcaccgaa acgcgcgaga 420
cgaaagggcc tcgtgatacg cctatTTTTt taggttaatg tcatgataat aatggtttct 480
tagacgtcag gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc 540
taaatacatt caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa 600

```

tattgaaaaa ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt	660
gcggcatttt gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct	720
gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc	780
cttgagagtt ttgcctccga agaacgtttt ccaatgatga gcacttttaa agttctgcta	840
tgtggcgcggt tattatcccc tattgacgcc gggcaagagc aactcggtcg ccgcatacac	900
tattctcaga atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc	960
atgacagtaa gagaattatg cagtgcgtcc ataaccatga gtgataaacac tgcggccaac	1020
ttactttctga caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg	1080
gatcatgtaa ctgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac	1140
gagcgtgaca ccacgatgcc tgtagcaatg ccaacaacgt tgcgcaaact attaaactggc	1200
gaactactta ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt	1260
gcaggaccac ttctgcgtc ggcccttcct gctggctggg ttattgctga taaatccgga	1320
gccggtgagc gtgggtctcg cggatcatt gcagcactgg ggccagatgg taagccctcc	1380
cgtatcgtag ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag	1440
atcgctgaga taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca	1500
tatatacttt agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc	1560
ctttttgata atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca	1620
gaccccgtag aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc	1680
tgcttgcaaa caaaaaaacc accgctacca gcggtgggtt gtttgccgga tcaagagcta	1740
ccaactcttt ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt	1800
ctagtgtagc cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc	1860
gctctgctaa tctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg	1920
ttggtctcaa gacgatagtt accggataag gcgcagcggg cgggctgaac ggggggttcg	1980
tgcacacagc ccagcttggg gcgaacgacc tacaccgaac tgagatacct acagcgtgag	2040
ctatgagaaa gcgccacgt tcccgaaggg agaaaggcgg acaggatatcc ggtaagcggc	2100
agggtcggta caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat	2160
agtccgtcg ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg	2220
gggcggagcc tatcgaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc	2280
tggccttttg ctacatggt ctttctgctg ttatcccttg attctgtgga taaccgtatt	2340

```

accgcctttg agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca 2400
gtgagcgagg aagcggaaga ggcaccaata cgcaaaccgc ctctccccgc gcgttgggccg 2460
attcattaat gcagctggca cgacagggtt cccgactgga aagcgggcag tgagcgcaac 2520
gcaattaatg tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg 2580
gctcgtatgt tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac 2640
catgattacg ccaagctatt taggtgacac tatagaatac tcagaggagg gatcctggag 2700
aattcatcct cgagtcggat ccaattcgac ttacggcaga aatacgttcc acatcagcgt 2760
tcgcatagcg ggcaatcact tcagcgagag ttgttttgcc ggtacccggc ggcccccaga 2820
ggatcataga atgtaaatgc ccggcttcga tagcgcgcgg caacggcttc ccgcagcca 2880
gcaaagtgtg ctggccgata tactgtgcta aattttctgg ccgcatacgc gcggccagag 2940
gttgaaaagt attatccgaa aaatcgagcg acagattgct cactcaggtg cctctactta 3000
cgttgatcat ctaccgtgac gccttgccgc ggggtgaagg taaatttcgc tgcattccaca 3060
gccccatttt gctgggattt cagttgataa ctgctgcgct gatcgtcctg ctccaccgcg 3120
ctaaactgat ggattgtgcc atcacgtccc acgttaatgg tgaactgctt cagattgcca 3180
ttgctggctt tcggcgtcag gacaaaagtca tcgccattct gtttgatatt gtactgctgc 3240
cagtcgctgg actgggttgc ggcaatcagc ataaacggcg tattaccggg ggcatctttc 3300
agccagggtt ccgtagcttg ctcaacgaac gggttataga accacagtgt ttaccgtca 3360
gaaaccagaa tgctttcatc aggttgtgtc atatgccagt tgaataagtt tggacgtttc 3420
accacagat cgccctgacc ttctgcacc gccgcgccgc taccgtcagt cactttttgt 3480
gtgaagctgg cgtggaagct gctgacttta tccaggcgcg gccgcgatat cgaattcgcc 3540
c 3541

```

<210> 20

<211> 69

<212> DNA

<213> Artificial

<220>

<223> M&M adaptor PET sense strand.

<220>

<221> misc_feature

<222> (1)..(17)

<223> n is a, c, g, or t

<220>

10

<221> misc_feature
<222> (50)..(69)
<223> n is a, c, g, or t

<400> 20
nnnnnnnnnnnn nnnnnnnnagt cggaggccaa ggcggccgta cgtccaactn nnnnnnnnnnn 60
nnnnnnnnnnnn 69

<210> 21
<211> 69
<212> DNA
<213> Artificial

<220>
<223> M&M adaptor PET antisense strand

<220>
<221> misc_feature
<222> (1)..(18)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (51)..(69)
<223> n is a, c, g, or t

<400> 21
nnnnnnnnnnnn nnnnnnnnnag ttggacgtac ggcgccttg gcctccgact nnnnnnnnnnn 60
nnnnnnnnnnnn 69

<210> 22
<211> 128
<212> DNA
<213> Artificial

<220>
<223> M&M adaptor PET sense strand with adaptor sequences

<220>
<221> misc_feature
<222> (29)..(47)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (80)..(99)
<223> n is a, c, g, or t

<400> 22
gatgtgctgc aaggcgatta agggatccnn nnnnnnnnnnn nnnnnnnnagt cggaggccaa 60
ggcggccgta cgtccaactn nnnnnnnnnnn nnnnnnnnnng gatccctgt gtgaaattgt 120

tatccgct 128

<210> 23
 <211> 128
 <212> DNA
 <213> Artificial

<220>
 <223> M&M adaptor PET antisense strand with adaptor sequences

<220>
 <221> misc_feature
 <222> (30)..(49)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (82)..(100)
 <223> n is a, c, g, or t

<400> 23
 agcggataac aatttcacac aggggatccn nnnnnnnnnn nnnnnnnnna gttggacgta 60
 cggccgcctt ggctccgac tnnnnnnnnn nnnnnnnnnn ggatccctta atcgcccttg 120
 agcacatc 128

<210> 24
 <211> 77
 <212> DNA
 <213> Artificial

<220>
 <223> M&M final PET sense strand after cleavage of adaptor sequences

<220>
 <221> misc_feature
 <222> (6)..(24)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (57)..(76)
 <223> n is a, c, g, or t

<400> 24
 gatccnnnnn nnnnnnnnnn nnnnagtcgg aggccaaggc ggccgtacgt ccaactnnnn 60
 nnnnnnnnnn nnnnnng 77

<210> 25
 <211> 77

<220>
<223> M&G adaptor antisense strand after sonication

<220>
<221> misc_feature
<222> (1)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (63)..(85)
<223> n is a, c, g, or t

<400> 27
nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn actccagcgt acggccgcct tggcctccga 60
ctnnnnnnnnn nnnnnnnnnnn nnnnnn 85

<210> 28
<211> 140
<212> DNA
<213> Artificial

<220>
<223> M&G adaptor sense strand after cloning into pGIS8 vector, rest of
vector not shown.

<220>
<221> misc_feature
<222> (10)..(54)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (87)..(129)
<223> n is a, c, g, or t

<400> 28
ggatccgacn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnagtcgg 60
aggccaaggc ggccgtacgc tggagtnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 120
nnnnnnnnnnc tccaggatcc 140

<210> 29
<211> 140
<212> DNA
<213> Artificial

<220>
<223> M&G adaptor antisense strand after cloning into pGIS8 vector,
rest of vector not shown.

<220>
<221> misc_feature
<222> (12)..(54)
<223> n is a, c, g, or t

14

<220>
<221> misc_feature
<222> (87)..(131)
<223> n is a, c, g, or t

<400> 29
ggatcctgga gnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnactcca 60
gcgtacggcc gccttggcct ccgactnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 120
nnnnnnnnnn ngtcggatcc 140

<210> 30
<211> 132
<212> DNA
<213> Artificial

<220>
<223> M PET intermediate sense strand

<220>
<221> misc_feature
<222> (10)..(46)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (79)..(121)
<223> n is a, c, g, or t

<400> 30
ggatccgacn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnagtc ggaggccaag 60
gcggccgtac gctggagtnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 120
nctccaggat cc 132

<210> 31
<211> 132
<212> DNA
<213> Artificial

<220>
<223> M PET intermediate antisense strand

<220>
<221> misc_feature
<222> (12)..(64)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (97)..(123)

15

<223> n is a, c, g, or t

<400> 31

ggatcctgga gnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 60

nnnnactcca gcgtacggcc gccttggcct ccgactnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 120

nnngtcggat cc 132

<210> 32

<211> 116

<212> DNA

<213> Artificial

<220>

<223> M&G diPET sense strand

<220>

<221> misc_feature

<222> (9)..(45)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (78)..(105)

<223> n is a, c, g, or t

<400> 32

gatccgacnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnagtcg gaggccaagg 60

cggccgtacg ctggagtnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnctcca ggatcc 116

<210> 33

<211> 116

<212> DNA

<213> Artificial

<220>

<223> M&G diPET antisense strand

<220>

<221> misc_feature

<222> (12)..(37)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (70)..(108)

<223> n is a, c, g, or t

<400> 33

ggatcctgga gnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnact ccagcgtacg gccgccttgg 60

cctccgaactn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnngt cggatc 116

<210> 34
<211> 111
<212> DNA
<213> Artificial

<220>
<223> Released diPET sense strand with BamH1 sticky ends

<220>
<221> misc_feature
<222> (9)..(45)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (78)..(105)
<223> n is a, c, g, or t

<400> 34
gatccgacnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnagtcg gaggcccaagg 60
cggccgtacg ctggagtnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnctcca g 111

<210> 35
<211> 111
<212> DNA
<213> Artificial

<220>
<223> Released diPET antisense strand with BamH1 sticky ends

<220>
<221> misc_feature
<222> (11)..(39)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (72)..(107)
<223> n is a, c, g, or t

<400> 35
gatcctggag nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnna ctccagcgta cggccgcctt 60
ggcctccgac tnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnngtc g 111

<210> 36
<211> 91
<212> DNA
<213> Artificial

<220>

<223> M&M PET sense strand before Mmel digestion

<220>

<221> misc_feature

<222> (1)..(28)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (61)..(91)

<223> n is a, c, g, or t

<400> 36

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnag tcggaggcca aggcggccgt acgtccaact 60

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn n 91

<210> 37

<211> 91

<212> DNA

<213> Artificial

<220>

<223> M&M PET antisense sequence before Mmel digestion

<220>

<221> misc_feature

<222> (1)..(31)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (64)..(91)

<223> n is a, c, g, or t

<400> 37

nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nagttggacg tacggccgcc ttggcctccg 60

actnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn n 91

<210> 38

<211> 53

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 8 sense strand before insertion of M&G linker.

<220>

<221> misc_feature

<222> (1)..(53)

<223> n is a, c, g, or t

<400> 38
 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnn 53

<210> 39
 <211> 53
 <212> DNA
 <213> Artificial

<220>
 <223> Oligonucleotide of FIG. 8 antisense strand before insertion of
 M&G linker

<220>
 <221> misc_feature
 <222> (1)..(30)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (32)..(53)
 <223> n is a, c, g, or t

<400> 39
 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn tnnnnnnnnnnn nnnnnnnnnnnn nnn 53

<210> 40
 <211> 140
 <212> DNA
 <213> Artificial

<220>
 <223> Oligonucleotide of FIG.8 sense strand inserted into plasmid (rest
 of plasmid not shown)

<220>
 <221> misc_feature
 <222> (10)..(129)
 <223> n is a, c, g, or t

<400> 40
 ggatccgacn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 60
 nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 120
 nnnnnnnnnnc tccaggatcc 140

<210> 41
 <211> 140
 <212> DNA
 <213> Artificial

<220>

<223> Oligonucleotide of FIG.8 antisense strand inserted into plasmid
(rest of plasmid not shown)

<220>

<221> misc_feature

<222> (12)..(131)

<223> n is a, c, g, or t

<400> 41

ggatcctgga gnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 60

nnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 120

nnnnnnnnnnn ngtcggatcc 140

<210> 42

<211> 140

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG.8 sense strand inserted into plasmid (rest
of plasmid not shown) after Mmel cut

<220>

<221> misc_feature

<222> (10)..(129)

<223> n is a, c, g, or t

<400> 42

ggatccgacn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 60

nnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 120

nnnnnnnnnnnc tccaggatcc 140

<210> 43

<211> 140

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG.8 antisense strand inserted into plasmid
(rest of plasmid not shown) after Mmel cut

<220>

<221> misc_feature

<222> (12)..(131)

<223> n is a, c, g, or t

<400> 43

ggatcctgga gnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 60

nnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 120

nnnnnnnnnnn ngtcggatcc 140

<210> 44

<211> 112

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 8 sense strand without M&G linker
inserted removed by electrophesis

<220>

<221> misc_feature

<222> (8)..(106)

<223> n is a, c, g, or t

<400> 44

gatccgannn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 60

nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnctcc ag 112

<210> 45

<211> 112

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 8 antisense strand without M&G linker
inserted removed by electrophesis

<220>

<221> misc_feature

<222> (11)..(109)

<223> n is a, c, g, or t

<400> 45

gacctggag nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn 60

nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnt cg 112

<210> 46

<211> 59

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 9 sense strand before insertion of M&G
linker

<220>

<221> misc_feature
<222> (1)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (30)..(59)
<223> n is a, c, g, or t

<400> 46
nnnnnnnnnn nnnnnnnnnn nnntccgacn nnnnnnnnnn nnnnnnnnnn nnnnnnnnn 59

<210> 47
<211> 59
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide of FIG. 9 sense strand before insertion of M&G linker

<220>
<221> misc_feature
<222> (1)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (37)..(59)
<223> n is a, c, g, or t

<400> 47
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gtcggannnn nnnnnnnnnn nnnnnnnnn 59

<210> 48
<211> 140
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide of FIG. 9 sense strand with part of M&G linker inserted (rest of plasmid not shown)

<220>
<221> misc_feature
<222> (10)..(58)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (65)..(129)
<223> n is a, c, g, or t

<400> 48

22

ggatccgacn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnntc 60
cgacnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnc tccaggatcc 140

<210> 49
<211> 140
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide of FIG. 9 antisense strand with part of M&G linker inserted (rest of plasmid not shown)

<220>
<221> misc_feature
<222> (12)..(76)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (83)..(131)
<223> n is a, c, g, or t

<400> 49
ggatcctgga gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nnnnnngtcg gannnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn ngtcggatcc 140

<210> 50
<211> 126
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide of FIG. 9 sense strand with partial M&G linker inserted after restriction enzyme digestion (rest of plasmid not shown)

<220>
<221> misc_feature
<222> (10)..(44)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (51)..(115)
<223> n is a, c, g, or t

<400> 50
ggatccgacn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnntccgac nnnnnnnnnn 60

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnctcca 120

ggatcc 126

<210> 51

<211> 126

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 9 antisense strand with partial M&G linker inserted after restriction enzyme digestion (rest of plasmid not shown)

<220>

<221> misc_feature

<222> (12)..(76)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (83)..(117)

<223> n is a, c, g, or t

<400> 51

ggatcctgga gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

nnnnnnnnnn nnnnnngtcg gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngtc 120

ggatcc 126

<210> 52

<211> 120

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 9 sense strand excised from plasmid and removed by electrophoresis as it is too long

<220>

<221> misc_feature

<222> (9)..(43)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (50)..(114)

<223> n is a, c, g, or t

<400> 52

gatccgacnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nntccgacn nnnnnnnnnn 60

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnctccag 120

<210> 53

<211> 120

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 9 antisense strand excised from plasmid
and removed by electrophoresis as it is too long

<220>

<221> misc_feature

<222> (11)..(75)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (82)..(116)

<223> n is a, c, g, or t

<400> 53

gatcctggag nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

nnnnnnnnnn nnnnngtcgg annnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngtcg 120

<210> 54

<211> 85

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide of FIG. 10 sense strand with M&G linker inserted
in the opposite orientation

<220>

<221> misc_feature

<222> (1)..(23)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (56)..(85)

<223> n is a, c, g, or t

<400> 54

nnnnnnnnnn nnnnnnnnnn nnnagtcgga ggccaaggcg gccgtacgct ggagtnnnnn 60

nnnnnnnnnn nnnnnnnnnn nnnnn 85

<210> 55

<211> 85

<212> DNA

```

<213> Artificial

<220>
<223> Oligonucleotide of FIG. 10 antisense strand with M&G linker
      inserted in the opposite orientation

<220>
<221> misc_feature
<222> (1)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (63)..(85)
<223> n is a, c, g, or t

<400> 55
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn actccagcgt acggccgcct tggcctccga      60

ctnnnnnnnn nnnnnnnnnn nnnnnn                                     85

<210> 56
<211> 148
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide of FIG. 10 sense strand inserted into plasmid
      (rest of plasmid not shown)

<220>
<221> misc_feature
<222> (10)..(54)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (87)..(139)
<223> n is a, c, g, or t

<400> 56
ggactggagn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnagtcgg      60

aggccaaggc ggccgtacgc tggagtnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn      120

nnnnnnnnnn nnnnnnnnnng tcggatcc                                     148

<210> 57
<211> 148
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide of FIG. 10 antisense strand inserted into plasmid

```

(rest of plasmid not shown)

```

<220>
<221> misc_feature
<222> (10)..(62)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (95)..(139)
<223> n is a, c, g, or t

<400> 57
ggatccgacn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn      60

nnactccagc gtacggcgcg cttggcctcc gactnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn      120

nnnnnnnnnnn nnnnnnnnnnc tccagtcc                                     148

<210> 58
<211> 62
<212> DNA
<213> Artificial

<220>
<223> Intermediate oligonucleotide 1 sense strand of FIG. 10

<220>
<221> misc_feature
<222> (10)..(53)
<223> n is a, c, g, or t

<400> 58
ggactggagn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnngtcggat      60

cc                                                                    62

<210> 59
<211> 62
<212> DNA
<213> Artificial

<220>
<223> Intermediate oligonucleotide 1 antisense strand of FIG. 10

<220>
<221> misc_feature
<222> (10)..(53)
<223> n is a, c, g, or t

<400> 59
ggatccgacn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnctccagt      60

```

27

cc 62

<210> 60

<211> 62

<212> DNA

<213> Artificial

<220>

<223> Intermediate oligonucleotide 2 sense strand of FIG. 10

<220>

<221> misc_feature

<222> (10)..(53)

<223> n is a, c, g, or t

<400> 60

ggactggagn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnngtcggat 60

cc 62

<210> 61

<211> 62

<212> DNA

<213> Artificial

<220>

<223> Intermediate oligonucleotide 2 antisense strand of FIG. 10

<220>

<221> misc_feature

<222> (10)..(53)

<223> n is a, c, g, or t

<400> 61

ggatccgacn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnctccagt 60

cc 62

<210> 62

<211> 114

<212> DNA

<213> Artificial

<220>

<223> Oligonucleotide sense strand generated by incorrect insertion of
M&G linker removed by electrophoresis

<220>

<221> misc_feature

<222> (8)..(108)

<223> n is a, c, g, or t

28

<400> 62
gatccgannn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnct ccag 114

<210> 63
<211> 114
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide antisense strand generated by incorrect insertion
of M&G linker removed by electrophoresis

<220>
<221> misc_feature
<222> (11)..(111)
<223> n is a, c, g, or t

<400> 63
gatcctggag nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ntcg 114

28

1. An isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first and second tags are tags of a nucleic acid-protein complex.
2. The isolated oligonucleotide according to claim 1, wherein the first tag is of a first polynucleotide and the second tag is of a second polynucleotide, and wherein the first and second polynucleotides are polynucleotides of a nucleic acid-protein complex.
3. The isolated oligonucleotide according to claim 1 or claim 2, wherein the oligonucleotide further comprises at least one restriction enzyme recognition site.
4. The isolated oligonucleotide according to any one of claims 1 to 3, wherein the oligonucleotide further comprises at least one linker.
5. The isolated oligonucleotide according to claim 4, wherein the linker comprises at least one restriction enzyme recognition site.
6. The isolated oligonucleotide according to claim 4 or claim 5, wherein the at least one linker is inserted between the tags or positioned upstream to the first tag or downstream to the second tag.
7. The isolated oligonucleotide according to any one of claims 3, 5 or 6, wherein the at least one restriction enzyme recognition site is for a type II_s restriction enzyme.
8. The isolated oligonucleotide according to claims 3, 5 or 6, wherein the at least one restriction enzyme recognition site is for a homing restriction enzyme.

9. The isolated oligonucleotide according to any one of claims 1 to 8, wherein the at least one first tag comprises a 5' terminus and a 3' terminus of a first polynucleotide and the at least one second tag comprises a 5' terminus and a 3' terminus of a second polynucleotide, and wherein the first and second polynucleotides are polynucleotides of a nucleic acid-protein complex.
10. The isolated oligonucleotide according to claim 5 or claim 6, wherein the linker comprises a first restriction recognition site recognized by a restriction enzyme capable of cleaving a first polynucleotide to obtain the first tag, and a second restriction recognition site recognized by a restriction enzyme capable of cleaving a second polynucleotide to obtain the second tag, the first and second polynucleotides being polynucleotides of a nucleic acid-protein complex.
11. The isolated oligonucleotide according to claim 5 or claim 6, wherein the linker comprises a first restriction recognition site recognized by a first restriction enzyme capable of cleaving a first polynucleotide to obtain a 3' terminus of the first polynucleotide, and a second restriction recognition site recognized by a second restriction enzyme capable of cleaving a second polynucleotide to obtain a 5' terminus of the second polynucleotide, the first and second polynucleotides being polynucleotides of a nucleic acid-protein complex.
12. The isolated oligonucleotide according to claim 5 or claim 6, wherein the linker comprises a first restriction recognition site recognized by a first restriction enzyme capable of cleaving a first polynucleotide to obtain a 5' terminus of the first polynucleotide, and a second restriction recognition site recognized by a second restriction enzyme capable of cleaving a second polynucleotide to obtain a 3' terminus of the second polynucleotide, the first and second polynucleotides being polynucleotides of a nucleic acid-protein complex.

13. The isolated oligonucleotide according to any one of claims 5, 6 or 11, wherein the linker comprises a first restriction recognition site recognized by a first restriction enzyme capable of cleaving a first polynucleotide to obtain a 3' terminus of the first polynucleotide, and a second restriction recognition site recognized by a second restriction enzyme capable of cleaving a second polynucleotide to obtain a 5' terminus of the second polynucleotide, the first polynucleotide further comprises a third recognition site recognised by a third restriction enzyme capable of cleaving the first polynucleotide to obtain a 5' terminus of first polynucleotide, and the second polynucleotide comprises a fourth recognition site recognised by a fourth restriction enzyme capable of cleaving the second polynucleotide to obtain a 3' terminus of the second polynucleotide; the first and second polynucleotides being polynucleotides of a nucleic acid-protein complex; the at least one first tag obtained from ligating the 5' and 3' termini of the first polynucleotide and the at least one second tag obtained from ligating the 5' and 3' termini of the second polynucleotide.
14. The isolated oligonucleotide according to any one of claims 1 to 13, wherein the nucleic acid-protein complex is part of a chromatin structure.
15. The isolated oligonucleotide according to any one of claims 1 to 14, wherein the polynucleotides are DNA or RNA.
16. A vector comprising the oligonucleotide of any one of claims 1 to 15.
17. A concatemer of oligonucleotides comprising at least two isolated oligonucleotides, each isolated oligonucleotide comprising at least one first tag and at least one second tag, wherein the first and second tags are tags of a nucleic acid-protein complex.
18. The concatemer of oligonucleotides according to claim 17, wherein the first tag is a tag of a first polynucleotide and the second tag is a tag of a second polynucleotide, the first and second polynucleotides being from a nucleic

acid-protein complex, and wherein each isolated oligonucleotide comprises at least one restriction enzyme recognition site.

19. The concatemer of oligonucleotides according to claim 18, wherein the at least one recognition site is included in at least one linker.
20. The concatemer of oligonucleotides according to claim 19, wherein the at least one linker is inserted between the tags or positioned upstream to the first tag and/or downstream to the second tag.
21. The concatemer of oligonucleotides according to any one of claims 18 to 20, wherein the at least one restriction enzyme recognition site is for a type II restriction enzyme.
22. The concatemer of oligonucleotides according to any one of claims 17 to 21, wherein the first tag comprises a 5' terminus and a 3' terminus from the first polynucleotide and the second tag further comprises a 5' terminus and a 3' terminus from the second polynucleotide.
23. An oligonucleotide library comprising at least one oligonucleotide according to any one of claims 1 to 15.
24. A method of preparing at least one isolated oligonucleotide comprising:
 - (d) providing a nucleic acid-protein complex;
 - (e) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first and second tags are obtained from a nucleic acid-protein complex; and
 - (f) isolating the oligonucleotide.
25. The method according to claim 24, wherein the first tag is obtained from a first polynucleotide and the second tag is obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex.

26. The method according to claim 24 or 25, wherein step (b) comprises:
- (ii) inserting at least one linker comprising at least one restriction enzyme recognition site, and
 - (ii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing the at least one recognition site in the linker to form an oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide and the linker.
27. The method according to claim 24 or 25, wherein step (b) comprises:
- (i) inserting at least one linker comprising at least one restriction enzyme recognition site between the first polynucleotide and the second polynucleotide of the complex;
 - (ii) adding at least one restriction enzyme recognition site each to the 5' terminus of the first polynucleotide and the 3' terminus of the second polynucleotide;
 - (iii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing at least one recognition site to obtain cleaved fragments; and
 - (iv) ligating cleaved fragments to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide, the tags comprising 5' and 3' termini of each polynucleotide, and the linker inserted between the tags.
28. The method according to claim 27, wherein the at least one restriction enzyme recognition site in step (ii) is part of an adaptor or of a vector.
29. The method according to any one of claims 24 to 28, wherein the nucleic acid-protein complex is obtained by chromatin immunoprecipitation.
30. A method of detecting and/or identifying at least two polynucleotides from a nucleic acid-protein complex, the method comprising:

- (g) providing a nucleic acid-protein complex;
- (h) preparing an oligonucleotide comprising at least one first tag and at least one second tag, wherein the first tag is obtained from a first polynucleotide and the second tag obtained from a second polynucleotide, the first and second polynucleotides obtained from a nucleic acid-protein complex;
- (i) sequencing the oligonucleotide; and
- (j) mapping the at least two polynucleotides based on the nucleotide sequences of the first and second tags, thereby detecting and/or identifying the at least two polynucleotides.

31. The method according to claim 30, wherein step (b) comprises:

- (i) inserting at least one linker comprising at least one restriction recognition site, and
- (ii) cleaving the first and second polynucleotides with a least one restriction enzyme recognizing the at least one recognition site in the linker to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag obtained from the second polynucleotide and the linker between the tags.

32. The method according to claim 30, wherein step (b) comprises:

- (i) inserting at least one linker comprising at least one restriction enzyme recognition site between the first polynucleotide and the second polynucleotide of the complex;
- (ii) adding at least one restriction enzyme recognition site each to the 5' terminus of the first polynucleotide and the 3' terminus of the second polynucleotide;
- (iii) cleaving the first and second polynucleotides with at least one restriction enzyme recognizing at least one recognition site to obtain cleaved fragments; and
- (iv) ligating cleaved fragments to form the oligonucleotide comprising a first tag obtained from the first polynucleotide, a second tag

obtained from the second polynucleotide, the tags comprising 5' and 3' termini of each polynucleotide, and the linker inserted between the tags.

33. The method according to claim 32, wherein the at least one restriction enzyme recognition site in step (ii) is part of an adaptor or of a vector.
34. The method according to any one of claims 30 to 33, wherein the nucleic acid-protein complex is obtained by chromatin immunoprecipitation.
35. The method according to any one of claims 30 to 34, wherein the oligonucleotide is concatenated with at least one further oligonucleotide obtained by steps (a) to (b) before being sequenced.
36. The method according to any one of claims 30 to 35, wherein the polynucleotides are located on the same chromosome or the polynucleotides are located on different chromosomes.

1. Abstract

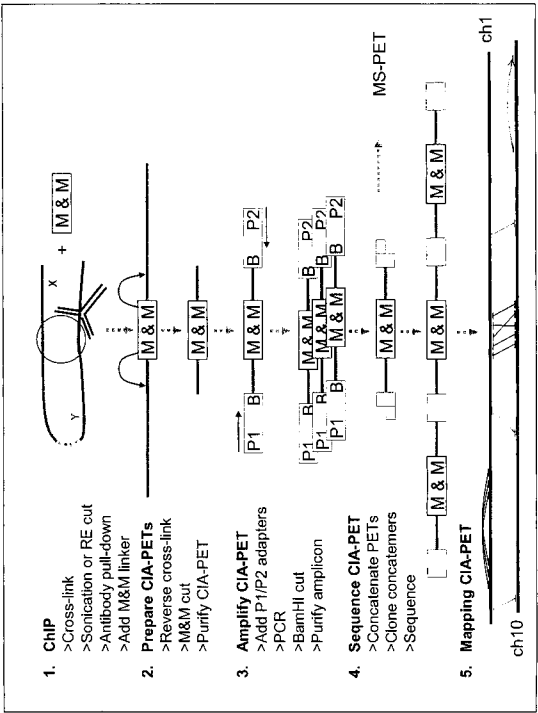
The present invention provides an isolated oligonucleotide and a method using the isolated oligonucleotide to detect and/or identify at least two polynucleotides from a nucleic acid-protein complex. The oligonucleotide comprises at least one first tag and at least one second tag, wherein the first and second tags are obtained from a nucleic acid-protein complex.

2. Representative Drawing

FIG. 2

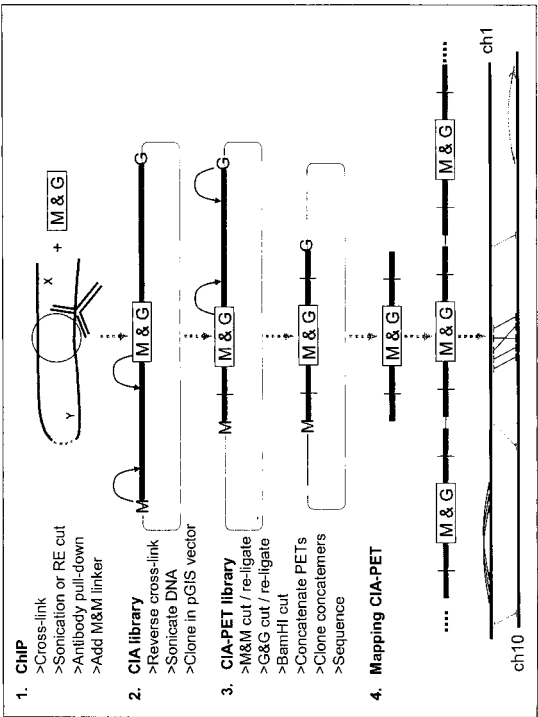
1/14

FIGURE 1



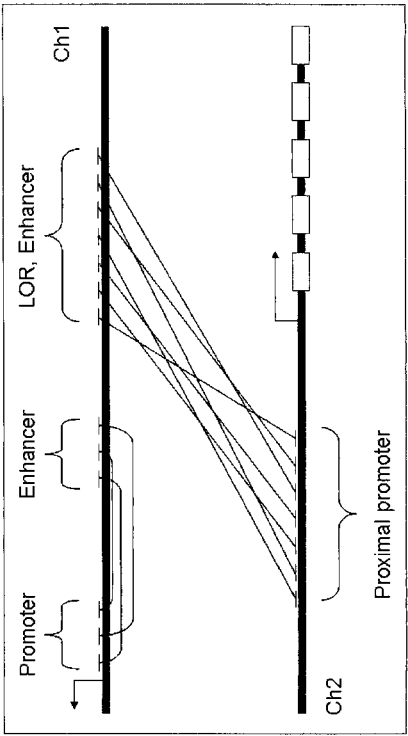
2/14

FIGURE 2



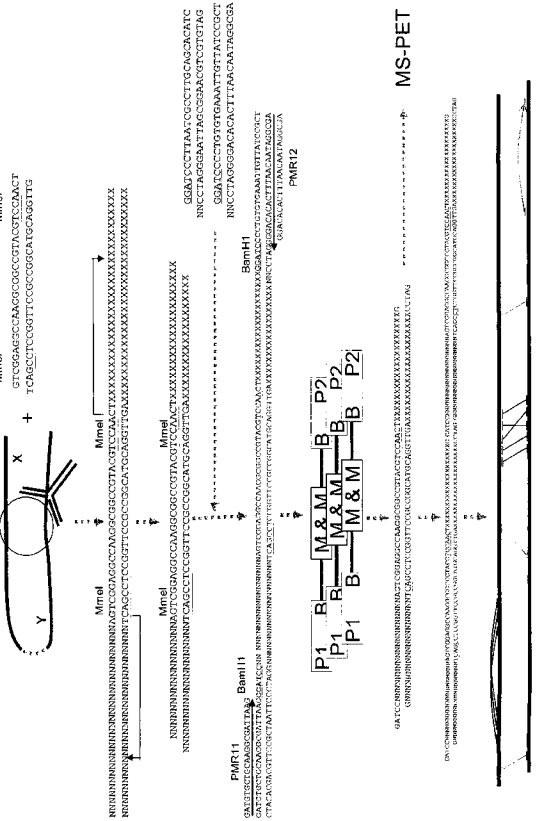
3/14

FIGURE 3



4/14

FIGURE 4



13/14

FIGURE 11 continued

2101 TCTCTCTT TCTTCTTCT ACATCTCTT GACTCTCTT GCTTGTATG ACTAGTATG GCGAGCTCT GCGAGTCTG CAGATATCC
 AGACATATA AGCATATGAG TATCAAGGTA CTAGAGGGA CAGATCTAT TATCTATG CCTCTCGAA TGTATAGCG GGTCTACAC GTTACTATG
 2201 GCGAGGGA GCTTACCTG CTTGGGATT ATGCGATA ACTAGGAGG CTGAGAGGG GAGTGTCTG GACTTATC GCGTCTATC
 GCTCTGGT GCGATGGCG GAGCTTAA TATCTGTAT TGTGTCTG GCTCTTCC GCTGTCTCT TGCAGAGC GTTAAATAG GCGAGGTAG
 2301 GAGTCTTA TGTCTTCC GAGATGTA GTATGAGT CAGATTTA TGTCTTCC AGCTTCTG GATCTTAC AGGATCTG GGTCTCTG GTCTAGCT
 GTAGATAT TACAGGCG CTTGATCT CATATGTA GCGTATAT ATAGAGCG TGCAGAC GTTATGATG TCTGTAGC CAGATGGA
 2401 GCGTCTG TGTAGGCG TGTAGCTCT GTTCTAGCG TGTAGGCG GTTATGAT CCGATCTT GCGAAGAA GCGATCTCT CTTCTCTG
 GCGAGGCG ATAGGAGT AGTGTAGCG CAGAGTTC TATCTCTT GAGTATCT GCGGTATG CAGCTTCT GCGAATGA GAGAGCAG
 2501 TGTCTCTG GGTAGGCG GGTAGCTCT GTTCTAGCG TGTAGGCG GTTATGAT CCGATCTT GCGAAGAA GCGATCTCT CTTCTCTG
 AGCTAGGA GATCTCTAT TGCAGCG TACATAT GATATCAT ACTGTGTA GTATATGA GATGAGCT AGCTAGGA TTTAGAAA
 2601 TGTCTCTG GGTAGGCG GGTAGCTCT GTTCTAGCG TGTAGGCG GTTATGAT CCGATCTT GCGAAGAA GCGATCTCT CTTCTCTG
 AGCTAGGA GATCTCTAT TGCAGCG TACATAT GATATCAT ACTGTGTA GTATATGA GATGAGCT AGCTAGGA TTTAGAAA
 2701 ATAGGAGC TTTATAGTG CTATCATG GAAAGTTC TGTAGGCG AACTTCTA GATCTTAC GGTGTAGA TGTATGTA TGTATCTG
 TATCTCTG AGTATCTG GATATCTT TACTTAC AGCTCTCT GTTGTAGC AGAGAGCG CAGATCTG CAGATCTG AGTATCTG
 2801 TGTCTCTG TGTAGGCG GGTAGCTCT GTTCTAGCG TGTAGGCG GTTATGAT CCGATCTT GCGAAGAA GCGATCTCT CTTCTCTG
 AGCTAGGA GATCTCTAT TGCAGCG TACATAT GATATCAT ACTGTGTA GTATATGA GATGAGCT AGCTAGGA TTTAGAAA
 2901 TGTCTCTG TGTAGGCG GGTAGCTCT GTTCTAGCG TGTAGGCG GTTATGAT CCGATCTT GCGAAGAA GCGATCTCT CTTCTCTG
 AGCTAGGA GATCTCTAT TGCAGCG TACATAT GATATCAT ACTGTGTA GTATATGA GATGAGCT AGCTAGGA TTTAGAAA

211

3001 AGAATATA AGAATGCG GTTCTCTG CATTCTCG AAGTCTCA CTTAGCTT AGAATAT TATATATG AATATCTT ATAAATAG
 TCTTCTAT TCTTCTCT CAGCTCTT GTAGAGGCG TTTCTATG GATCTAGA TCTTCTGA ATATATG TATATGTA TATCTTAT
 3101 GGTATGCG AGCTCTCT GTTCTCTG TTTCTATG GATCTAGA ACTTCTCA CTTCTCTG CAGCTCTT TGTATGTA TATCTTAT
 CAGATCTG TGTAGGCG CAGAGCG AGCTCTT CTTCTATG TGTATGTA TATCTTAT TGTATGTA TATCTTAT
 3201 GATCTCTG GGTATGCG CATTCTCG CTTCTCTG CATTCTCG TGTATGTA TATCTTAT TGTATGTA TATCTTAT
 CAGATCTG TGTAGGCG CAGAGCG AGCTCTT CTTCTATG TGTATGTA TATCTTAT TGTATGTA TATCTTAT

14/14

FIGURE 11 continued

3301 TGTATGCG GGTATGCG CATTCTCG CTTCTCTG CATTCTCG TGTATGTA TATCTTAT TGTATGTA TATCTTAT
 CAGATCTG TGTAGGCG CAGAGCG AGCTCTT CTTCTATG TGTATGTA TATCTTAT TGTATGTA TATCTTAT
 3401 ATCTCTCG GATCTCTG TATATGTA GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG
 TATCTCTG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG
 3501 TGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG
 TATCTCTG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG GGTATGCG

专利名称(译)	核酸相互作用分析		
公开(公告)号	JP2007289152A	公开(公告)日	2007-11-08
申请号	JP2007063117	申请日	2007-03-13
[标]申请(专利权)人(译)	新加坡科技研究局		
申请(专利权)人(译)	四局科学技术研究		
[标]发明人	イー・ジュン・ルアン メリッサ・ジェイン・フルウッド チアリン・ウェイ ハック・ファイ・ワン		
发明人	イー・ジュン・ルアン メリッサ・ジェイン・フルウッド チアリン・ウェイ ハック・ファイ・ワン		
IPC分类号	C12N15/09 C40B40/06 C12Q1/68 G01N33/53 G01N33/566 C12Q1/02		
CPC分类号	C12N15/1065 C12Q1/6804 C12Q2525/197 C12Q2525/131 C12Q2521/313 C12Q2531/113 C12Q2525/191 C12Q2523/101		
FI分类号	C12N15/00.ZNA.A C40B40/06 C12Q1/68.Z G01N33/53.M G01N33/566 C12Q1/02 C12N15/00.A C12N15/00.AZN.A C12N15/10.P C12N15/10.Z C12Q1/6804.C C12Q1/6804.Z		
F-TERM分类号	4B024/AA11 4B024/AA20 4B024/CA01 4B024/CA10 4B024/CA11 4B024/CA20 4B024/HA11 4B024/HA20 4B063/QA01 4B063/QA13 4B063/QQ02 4B063/QQ08 4B063/QQ42 4B063/QQ52 4B063/QQ91 4B063/QR08 4B063/QR14 4B063/QR20 4B063/QR32 4B063/QR35 4B063/QR42 4B063/QR62 4B063/QR77 4B063/QR80 4B063/QS12 4B063/QS16 4B063/QS25 4B063/QS28 4B063/QS31		
优先权	11/373519 2006-03-13 US		
外部链接	Espacenet		

摘要(译)

分离的寡核苷酸和使用该分离的寡核苷酸从核酸-蛋白质复合物中检测和/或鉴定至少两个多核苷酸的方法。 解决方案：与DNA /蛋白质/ DNA复合体结合的DNA片段通过接头序列连接。 接头序列含有限制酶切割位点。 通过交联固定蛋白质复合物后，用限制酶切割接头。 然后使用释放的末端作为标签分析碱基序列。 [选择图]图2

FIGURE 2

