

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 6 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26830137

研究課題名(和文)次世代シーケンサーを駆使したプロモーター配列の網羅的解析

研究課題名(英文)High-throughput analysis for mutant promoter using massively parallel sequencer

## 研究代表者

入江 拓磨(Irie, Takuma)

東京大学・新領域創成科学研究科・特任助教

研究者番号：50625944

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：一塩基単位の解像度でプロモーター領域の転写活性情報を抽出する手法の開発を行った。平均で1-2%ランダム突然変異を導入したヒトプロモーターに対し、次世代シーケンサーを用いた大規模なレポーターアッセイ系の構築を行い、プロモーター配列の解析と活性測定を行った。これらの情報を用い、プロモーター配列と活性に関する線形モデルを構築によって、事前の知識が無くても転写制御の情報を抽出を可能にした。また、これらの成果はプロモーター配列のデザインについての基礎情報として有用であると考えている。

研究成果の概要(英文)：We have developed a novel method to characterize potential promoter activities of proximal promoters of human genes at a single-base resolution. We introduced an average of 1-2 % mutations at random positions and subjected them to systematic reporter gene assays. Their sequences and encoding promoter activities were characterized by the next generation sequencing technology. We could extract information on transcriptional regulation at a single-base resolution from the multiple linear regression model. We consider that it is possible to design of the promoter sequence from the promoter prediction model.

研究分野：ゲノムシステム医療科学分野

キーワード：次世代シーケンサー 超並列レポーターアッセイ プロモーター 転写因子結合配列

## 1. 研究開始当初の背景

所属の研究室においては、これまでにヒトプロモーター領域約 500 種類を含む約 750 種類の DNA 断片のクローニングを行い、ヒト培養細胞の HEK293 細胞に一過的に導入し、ルシフェラーゼアッセイにより定量的なプロモーター活性データを収集した。(Sakakibara, Irie et al., DNA res., 2007). これら定量プロモーター活性データと DNA 配列情報を基に、転写因子結合配列を説明変数、プロモーター活性の“絶対量”を目的変数とした線形モデルとして、プロモーターの数理モデルの構築を行った。その結果、予測値と実測値の相関係数が約 0.6 となるようなモデルを構築することができた (Irie et al., Nucleic. Acids Res., 2011)。またプロモーター配列の機能エレメント抽出を目的として、ランダム点突然変異が入ったヒト EF1

1 プロモーター配列約 250 種類のルシフェラーゼアッセイを行った結果、変異型プロモーターは基の活性から約 0.01~6 倍の範囲で変化することを確認した。このことからランダム突然変異プロモーターからより高い活性、細胞内で望みの活性を持つプロモーター配列の創生が可能になると考えている。しかしながら、これまでのプロモーターの解析でプロモーター配列に対して1つずつルシフェラーゼアッセイを行っているため、スループットに関して課題があり、全プロモーター領域をカバーするような形で変異箇所を探索することが困難であった。そのため一塩基単位の変化とプロモーター活性の情報を体系的に収集するには別の実験的アプローチが必要であると考えられた。近年、次世代シーケンサーの登場により、これまでのレポーターアッセイのスループットをはるかに超える効率の実験系の構築が実現すると考えられた。

## 2. 研究の目的

プロモーターはその領域中に存在する転写因子結合配列と転写因子との相互作用によって遺伝子の転写量を制御している。したがってプロモーター配列から遺伝子の発現量を知ることが可能であれば遺伝子の機能解析などの基礎研究分野において有用な情報を与えることができると考えられる。また配列の改変することで転写量を自由にコントロールすることが可能であれば、組み換えタンパク産生など細胞内で転写量を厳密にコントロールしたい場合や合成生物学分野においてプロモーター活性のパラメーターを自由に変動できるツールとして応用研究分野においても有用な DNA 資源を提供する事が可能になると考えられる。本研究では、次世代シーケンサーを駆使した変異型プロモーター配列の網羅的な解析を通じて、一塩基単位の解像度でプロモーター配列の設計原理を明らかにする。さらに望みのプロモーター活性を持つプロモーター配列をデザ

インする実験的・情報学的方法論の開発に取り組んだ。

## 3. 研究の方法

### (1) エラープロオン PCR による変異導入

高効率にランダム変異を導入する目的とするエラープロオン PCR 法を行った。約 1kb のプロモーター-DNA に対して、dNTP の濃度のバランスを、Mg<sup>2+</sup>と Mn<sup>2+</sup>を加える、計 80 サイクルの PCR を行うことで変異プロモーターの作製を行った。

### (2) 次世代シーケンサーによる変異活性測定

変異プロモータープラスミドライブラリーの作製

次世代シーケンサーを使用したレポーターアッセイ用の変異プロモータープラスミドライブラリーを作製した。プロモーターのタグ付けおよび活性測定を行う目的として、レポーター遺伝子の 3' UTR 上にランダム 12 塩基のバーコード配列を付加した。変異プロモーター、GFP 遺伝子、バーコード DNA を連結しプラスミドに挿入した。数万種類のクローンが得られる条件でトランスフォーメーションを行い、培養後にプラスミド精製を行った。

### 変異プロモーターの活性測定

変異プロモーターの活性にはヒト培養細胞 (HEK293, A549, k562, hepG2, HeLaS3 など) を用いた。変異プロモータープラスミドライブラリーをトランスフェクションし、RNA 抽出、mRNA 精製後、レポーター遺伝子に特異的なプライマーを用い逆転写反応により cDNA を合成した。cDNA とプラスミドライブラリーをテンプレートとしバーコード配列について HiSeq2000 または 2500 を用いて single read シーケンシングをすることでタグカウント計測を行った。cDNA のタグカウントをプラスミドのタグカウントで除算した値をプロモーター活性とした。

### 変異プロモーターの配列決定

プロモーター配列決定についても HiSeq2000 または 2500 を用いた。プロモーター領域とバーコード配列を連結した DNA 断片を作製し、トランスポゾンランダムに挿入し、PCR で増幅したものを HiSeq2000 または 2500 で paired-end シーケンスにより配列決定した。バーコード配列ごとにデータを分割後、BWA, GATK を用いてマッピング、変異箇所を決定した。

### (3) 線形モデルによるプロモーター活性予測

変異プロモーターの配列情報を説明変数、プロモーター活性を目的変数とした線形モデルの構築を行うことで、変異情報からプロモーター活性の予測式を導出した。プロモ

ター上の各塩基における 変異の有無, 変異のパターンを説明変数とした時, 重回帰分析による解析を行った. 重回帰分析後, Lasso (Least absolute shrinkage and selection operator) によるモデル構築を行った. 計算には R を用いた.

#### (4) ルシフェラーゼアッセイによる変異プロモーター活性の検証

線形モデルの検証をルシフェラーゼアッセイによって行った. 線形モデルによりプロモーター活性が大きく変わると推定された箇所に対し, 点変異を導入したプロモーター配列の作製を行い, プロモーター活性の測定を行った.

### 4. 研究成果

#### (1) 変異プロモーター配列作製法の検討

プロモーター DNA に対してランダムに変異を導入する目的として, 高効率に変異を導入するエラープローン PCR の条件検討を行った. PCR の条件を, dNTP の濃度のバランスを  $dATP, dGTP \gg dTTP, dCTP$  とすること,  $Mg^{2+}$ ,  $Mn^{2+}$  を加えること, PCR サイクルを 80 サイクルまで増やすことなどの条件で高効率な変異の導入が可能であり, 約 2% 前後の変異率のプロモーター配列の作製が可能であった.

#### (2) 次世代シーケンサーを用いた変異プロモーターのレポーターアッセイ系の構築

変異プロモーター, GFP 遺伝子, 12nt のバーコード配列を連結した DNA 断片をプラスミド中に挿入し, 変異プロモーターのレポータープラスミドライブラリーを作製した. バーコード配列はプロモーターに対する標識として用いる. ヒト培養細胞 HEK293 細胞にトランスフェクションし, RNA 抽出, mRNA 精製後, cDNA を作製した. 次世代シーケンサーを用いてバーコード領域のタグカウントを測定した. このとき, プラスミドライブラリーのバーコードは入配列のタグカウントを同時に測定した. プラスミドタグカウントはライブラリー中のコピー数, PCR バイアスの補正に利用した. cDNA のタグカウントをプラスミドタグカウントで除算した値をプロモーター活性とした. 野生型のプロモーター活性と変異型のプロモーター活性の分布を調べた結果, 野生型と比較して変異型プロモーターにおいて活性が変動していることが確認され, 配列の変化が活性の変化を引き起こしているものと考えられた. また変異により活性が減少だけでなく増大する変異プロモーターもあることがわかった. EEF1a, GAPDH など転写活性が高く進化的にも保存されていると考えられる遺伝子のプロモーターにも見つかった.

プロモーターライブラリーから計 12 種類のプロモーター領域 (ヒト EEF1a1, GAPDH, DDX5, RPS12, PGK5, RBBP5, ARL6IP5, NTS, COMMD9, PSMD6, TPD52L1, CDK16) について数

万種類の変異プロモーターライブラリーを作製し, HEK293 細胞中の転写活性計測を行った. いずれのプロモーターについても転写活性の変動が観測された.

変異プロモーター配列の決定についても次世代シーケンサーを用いた決定法についても開発を行った. プロモーター領域とバーコード配列を同時に解析することでプロモーター領域の配列決定を行った. Illumina の HiSeq によるシーケンシングは数百 bp の短い DNA 断片の解析であるため, 断片化の必要があった. PCR, セルフライゲーションによりプロモーター領域とバーコード配列を連結した DNA 産物を作製後, トランスポゾン DNA をランダムに挿入し, PCR による増幅を行うことで, ランダムな長さのプロモーター領域 DNA とバーコード配列のペアを持つ DNA 断片の両側をシーケンスした. バーコードごとに配列結果を分割後マッピング, 変異箇所の検出を行った. その結果, 全体の変率が約 2% であり, 置換, 欠失, 挿入のパターンが観測された. 変異の検出率に関してはまだ課題は残るもの, ある程度実用可能であると考えている.

プロモーター活性の測定と配列の決定の結果, 各変異プロモーターライブラリーにおいて約 2 万種類の変異プロモーターのレポーターアッセイの実験系を構築することができた.

#### (3) プロモーター活性予測モデルの構築

バーコード配列によって, プロモーター活性情報と配列情報を関連付けることが可能になる. プロモーターの配列情報を基にプロモーター活性を予測する数理モデルの構築を試みた. まずプロモーター配列上の変異の有無を説明変数とし, プロモーター活性を目的変数として重回帰分析と LASSO によるモデル構築を行った. モデルの予測値と実験値は  $r=0.5$  となり, ある程度の予測精度が得られた. このモデルより得られた決定係数は変異によってプロモーター活性がどの程度変化するについての推定値となる. 係数値をプロモーター上に並べた結果, 高い係数を与えた塩基は転写開始点の上流 200bp 程度に集中しており, また高い係数を持つ塩基は数 bp 連続して出現する傾向にあり, 転写因子結合配列が存在する領域であると考えられた. このような中には TATA box などの既知の転写因子結合配列を含まれており, 転写因子結合配列を含むシスエレメントの抽出が可能になったと考えている. またこれら数理モデルにより得られた塩基とプロモーター活性の情報と進化上の保存度との比較を行った. その結果, 高いプロモーター活性を有する領域と保存度が比較的高い傾向にあることがわかった.

また変異のパターンを説明変数としたモデルの場合においても同様に線形モデルによる予測式の導出を試みた. 重回帰分析と

LASSO によるモデル構築を行った。モデルの予測値と実験値は  $r=0.6$  となった。この場合、塩基がどの塩基に置換・欠失とプロモーター活性との関係とした情報を得ることが可能であった。以上のことから、単純なモデルではあるものの中程度の精度をもつ予測式を導出できたと考えている。

(4) プロモーター配列の改変による活性の制御

以上より得られた数理モデルを基にして、プロモーター配列の改変により活性をコントロール可能かどうか検討を行った。

例として、EEF1a1 と GAPDH より高い活性可能を有する数 bp の領域を選び、点突然変異を全パターンでのプロモーターについてルシフェラーゼアッセイで活性測定を行った。その結果、数理モデルから予測と同等の変化が起こることが確認された。特に GAPDH の解析例では、領域全体では活性が落ちるものの、そのうちの変異 1 パターンは活性を増大させると予測された領域であったが、ルシフェラーゼアッセイでも確認することができた。活性を増大させた配列について TRANSFAC によって転写因子結合配列の探索を行った結果、このパターンでのみ NFY の結合サイトの出現が確認された。新規に転写因子結合配列が出現したものと考えられた。

同様のことを EEF1a1 プロモーターでも検討したところ、活性を増大させると予測された 3 塩基について、点変異プロモーターのルシフェラーゼアッセイによる測定を行った。その結果、予測どおりに活性が増大することが確認でき、TRANSFAC により解析した結果、変異により新規に CREB の結合サイトが出現し、検索結果のスコアと活性が相関することがわかった。またやや離れた箇所にプロモーター活性を増大させる塩基置換が存在し、CREB の領域とその他 2 箇所の塩基置換の組み合わせのパターンのプロモーター配列のルシフェラーゼアッセイによる測定を行った。その結果、複数の塩基置換によるプロモーター活性への影響の効果は相加的に働く傾向にあることが示された。すなわち、複数の塩基置換の組み合わせによってプロモーター活性を制御することが可能であることが示唆された。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計 5 件)

入江拓磨, 門城拓, 劉エイ, 関真秀, 菅野純夫, 矢田哲士, 鈴木穰, 次世代シークエンサーによる変異プロモーターの網羅的解析, 第 37 回日本分子生物学会年会, 2014 年 11 月 25 日 ~ 27 日, パシフィコ横浜 (神奈川県・横浜市)

門城拓, 入江拓磨, 菅野純夫, 鈴木穰, 次世

代シークエンサーを用いた肺腺がん細胞株におけるプロモーター活性の解析, 第 37 回日本分子生物学会年会, 2014 年 11 月 25 日 ~ 27 日, パシフィコ横浜 (神奈川県・横浜市)

入江拓磨・鈴木穰, 次世代シークエンサーを用いた変異ヒトプロモーターの網羅的解析, 2015 年 8 月, 日本進化学会 第 17 回大会, 2015 年 8 月 20 日 ~ 23 日, 中央大学後楽園キャンパス (東京都・文京区)

入江拓磨, 劉瑩, 門城拓, 関真秀, 榊原雄太, 菅野純夫, 矢田哲士, 鈴木穰, 次世代型シークエンサーを用いた変異プロモーターの網羅的解析, 第 38 回日本分子生物学会年会、第 88 回日本生化学会大会 合同大会, 2015 年 12 月 1 日 ~ 2015 年 12 月 4 日, 神戸ポートアイランド (兵庫県・神戸市)

劉瑩, 入江拓磨, 門城拓, 矢田哲士, 鈴木穰, 変異導入プロモーターの転写活性データを用いて転写制御コードを解読する, 第 38 回日本分子生物学会年会、第 88 回日本生化学会大会 合同大会, 2015 年 12 月 1 日 ~ 2015 年 12 月 4 日神戸ポートアイランド (兵庫県・神戸市)

[図書](計 1 件)

入江拓磨, 羊土社, 実験医学別冊 RNA-Seq 実験ハンドブック, 第 2 章 12 超並列レポーターアッセイによる転写制御解析, 2016 年, p.182 ~ 187

## 6. 研究組織

### (1) 研究代表者

入江 拓磨 (IRIE, Takuma)

東京大学大学院・新領域創成科学研究科・特任助教

研究者番号: 50625944