

## 論文の内容の要旨

論文題目 プロモーター領域中の塩基置換が転写活性化能へ与える影響の網羅的解析

氏名 榎原 雄太

### 背景

ヒトゲノムの解読の結果、ヒトゲノム中に存在する遺伝子の概要が明らかになった。しかし、それらがどのような時期にどのような量で転写されるのか、その発現制御についての全体像の解明は依然として大きな課題である。近年、転写制御機構の全体像の解明に向けた国際的な研究が進み、そのエピゲノム、トランスクリプトームについての情報集積が急速に進んでいる。例えば、ENCODEプロジェクトでは、100種類以上の細胞より取得したChIP-seq及びRNA-seqのデータを用い、ヒストン修飾や転写因子結合パターンから遺伝子発現量を予測するモデル構築も行われている(参考文献1)。しかし、遺伝子発現を説明する最も基盤的な要素としてのプロモーター領域のDNA配列として、その配列に内在的にコードされる転写活性化能すなわち1次元のDNA配列の転写活性化能については、依然として情報の集積が少ない。私は、本研究において、転写開始点近傍プロモーター領域の配列が転写制御にどの程度寄与しているか、体系的、網羅的に明らかにすべく方法論の開発を行った。転写開始点の同定には我々の研究室で開発、大規模解析を行ってきた完全長cDNAの5'端配列を指標とした。本研究では、その前半にこれらの転写開始点近傍プロモーター配列の体系的ルシフェラーゼアッセイについて、その手法の開発と実践を記す(参考文献2,3)。後半では、開発された手法と次世代シーケンス技術を融合させることで、そのさらなる効率化、大規模化の実現を試みた(論文準備中)。

### 材料と方法

#### 転写開始点近傍プロモーター配列の体系的ルシフェラーゼアッセイ

ヒト培養細胞であるHuman embryonic kidney(HEK) 293細胞由来の完全長cDNAの転写開始点を基点とし、上流1000bp、下流200bpをプロモーター領域とした。同定したプロモーター領域をPCRによりヒト正常ゲノムDNAより増幅、ホタルルシフェラーゼ遺伝子を有するpGL3 basicベクターへクローニングした。クローニングしたプロモーター配列はサンガーシーケンサーを用い確認し、クローンよりプラスミド精製、HEK293細胞へトランスフェクションした。この際、補正用にウミシイタケルシフェラーゼ遺伝子を有するphRL-TKベクターも同時にトランスフェクションし、活性値を測定した。各プロモーター配列の活性値はトリプリケートで測定し、ホタルルシフェラーゼの活性値をウミシイタケルシフェラーゼの活性値で除した値の平均値を、転写活性化能とした。レポーター遺伝子アッセイは $1 \times 10^3$  cellsを用いて96穴マイクロタイタープレートで行うことにより測定の効率化を図った。最終的に472のプロモーター領域の配列確認と転写活性化能の測定を実施した。

#### 次世代シーケンス技術を用いた変異プロモーター配列の大規模解析手法

野生型のプロモーター配列をError-Prone PCRを用い増幅することで、ランダムに変異の挿入さ

れた変異プロモーター配列を作成した。作成した変異プロモーター配列に、レポーター遺伝子アッセイに利用するGFP遺伝子及び作成された変異プロモーター配列を識別する12塩基のランダムタグ配列を付加することで、効率的な配列決定及びレポーター遺伝子アッセイを実現した。レポーター遺伝子アッセイは、 $1 \times 10^6$  cellsを10cm dishで24時間培養したHEK293細胞に、プールした変異プロモーター配列のプラスミドをトランスフェクション、48時間後にtotal RNAを回収した。回収したtotal RNAより、mRNAを抽出、cDNA合成、Illumina Hi-Seq2000の36-nt single-end reads法を利用しランダムタグ配列を計測した。

## 結果と考察

### 体系的レポーターアッセイ法によるプロモーター配列の解析

本研究では、プロモーター領域のDNA配列が転写活性化能へ与える影響を解き明かす第一歩として、特定された転写開始点を基点とした上流1000bp、下流200bpをプロモーター領域とし、ゲノム領域から増幅、472遺伝子に対して体系的ルシフェラーゼレポーター遺伝子解析を行った。また対照として、ランダムに選択した遺伝子間領域251箇所についても同様の解析を行った。得られたそれぞれの配列の特徴と転写活性化能の解析を行った。プロモーター領域の転写活性化能は二峰性の分布を示していた(図1)。転写活性化能が高値となるP1群(87%)及び低値となるP2群(13%)の2群の間には、既知のシスエレメントの存在割合など、配列上の相違点も存在していることが明らかとなった(表1)。また遺伝子間領域は配列上の特徴を大きく異ならせているが、その中にも十分な転写活性化能を有している配列が存在していた。

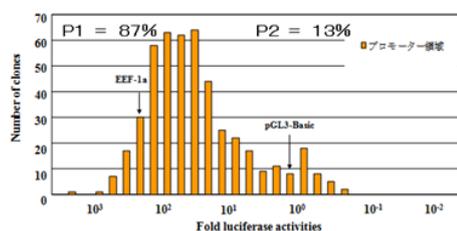


図1: 472遺伝子のプロモーター領域の転写活性化能の分布

表1: プロモーター領域の配列的特徴

	Promoter	P1	P2
CpG island	267 (57%)	259 (63%) <sup>+1</sup>	8 (13%) <sup>+1</sup>
TATA box: strict <sup>+1</sup>	34 (7%)	30 (7%)	4 (7%)
TATA box: less strict <sup>+2</sup>	103 (22%)	81 (20%) <sup>+2</sup>	22 (36%) <sup>+2</sup>
Average G+C content:	53%	54% <sup>+3</sup>	47% <sup>+3</sup>
TOTAL	472	411	61

<sup>+1</sup>p<1.0e-6 <sup>+2</sup>p<1.0e-2 <sup>+3</sup>p<6.0e-10

これらの結果から数百の規模でのプロモーター配列と転写活性化能の相関データを取得することが可能となったが、ゲノム全体におけるその関係を解き明かすためには、より多くのプロモーターについて、またひとつのプロモーター配列内でも1塩基レベルでの配列の差が転写活性化能へ与える影響を解析することが課題と考えられた。その手段として、塩基レベルでの欠損や塩基置換を入れた変異プロモーターを網羅的に作成、解析することが考えられたが、従来の変異プロモーターの取得、配列決定と転写活性化能の測定方法ではスループットに課題があり、より効率的な実験手法の確立が必要であった。

### 次世代シーケンス技術を用いたプロモーター配列の大規模解析手法の開発

そこで、遺伝子内のプロモーター領域の塩基レベルでの配列の差が、転写活性化能に及ぼす影響を解析するため、レポーターアッセイ系のさらなる効率化、大規模化を行った。遺伝子のプロモーター領域にランダムに変異の入った変異プロモーターを効率的に取得し、その配列決定及び転写活性化能の測定を行う、実験手法を確立した(図2-A)。

野生型のプロモーター配列をError-PCRを用い増幅することで、ランダムに変異の挿入された変異プロモーター配列を作成した(図2-A)。Error-Prone PCRは校正活性のないTaqポリメラーゼを用い、エラーの入りやすい条件で実施するPCR法である。また、非特異的なDNAの増幅を抑えるため、野生型のプロモーター領域をError-PCRで増幅した後、再度同様の条件でDNA断片内側に設計したプライマーを用い、Nested PCRを行った。

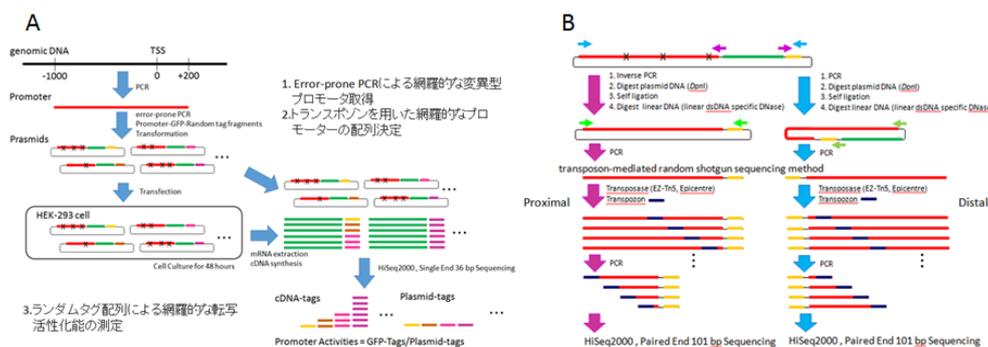


図2: 次世代シーケンス技術を用いたプロモーター配列の大規模解析手法の概要

次世代シーケンス技術を用いた、効率的な変異型プロモーターの作成、配列決定、転写活性化能の測定方法の概要。(A) 全体像 (B) トランスポゾンを用いたプロモーター領域の配列決定

作成した変異プロモーター配列に、レポーター遺伝子アッセイに利用するGFP遺伝子及び作成された変異プロモーター配列を識別する12塩基のランダムタグ配列を付加し、レポーター遺伝子アッセイベクターに挿入した。

レポーター遺伝子ベクターに挿入した多様な変異プロモーターを一度に配列決定するため、ランダムタグ配列及びトランスポゾンを利用し、効率的にdeletion mutantを作成、プロモーター領域の3'端側及び5'端側のそれぞれに偏ったIllumina HiSeq用のシーケンスライブラリーを準備した。得られたシーケンスライブラリーを用い、Illumina HiSeq2000の101bp paired end reads法にて配列決定を行った(図2-B)。配列を決定した結果、本手法では1度に約2万種類もの変異プロモーター配列を作成、配列決定が実施できた。また、変異プロモーター配列には、平均して約1~2%の変異が含まれていることが示された(表2)。

表2: 変異型プロモーター配列の特徴

	EEF1A1	GAPDH	DDX5	RPS12	PGK1	RBBP5	ITS	ARL6IP5
promoter_length	1103	1058	1111	1126	1079	1046	1088	1127
GC_content	0.59	0.65	0.63	0.55	0.57	0.46	0.38	0.44
# of mutant promoters	20575	24335	17139	22366	21902	23609	28833	20496
total mutation	396252	436298	354770	516061	471617	404006	534168	376678
mutation rate	1.7%	1.7%	1.9%	2.0%	1.9%	1.6%	1.7%	1.6%
A>T, T>A	79,802 (20.1%)	84,800 (19.4%)	51,389 (14.5%)	78,888 (15.3%)	97,723 (20.7%)	99,469 (24.6%)	127,501 (23.9%)	99,046 (26.3%)
A>G, T>C	64,674 (16.3%)	64,255 (14.7%)	64,270 (18.1%)	93,477 (18.1%)	74,374 (15.8%)	104,005 (25.7%)	147,400 (27.6%)	75,350 (20.9%)
A>C, T>G	65,085 (16.4%)	68,881 (15.8%)	45,846 (12.9%)	71,160 (13.8%)	74,004 (15.7%)	100,963 (25.0%)	109,917 (20.6%)	71,677 (19.0%)
G>A, C>T	80,943 (20.4%)	99,875 (22.9%)	99,419 (28.0%)	120,748 (23.4%)	119,436 (25.3%)	39,224 (9.7%)	55,736 (10.4%)	47,038 (12.5%)
G>T, C>A	54,799 (13.8%)	50,617 (11.6%)	52,129 (14.7%)	67,705 (13.1%)	53,631 (11.4%)	21,809 (5.2%)	23,930 (4.5%)	24,522 (6.5%)
G>C, C>G	21,004 (5.3%)	24,252 (5.6%)	20,520 (5.8%)	61,124 (11.8%)	25,444 (5.4%)	11,052 (2.7%)	16,125 (3.0%)	15,920 (4.2%)
deletion	29,189 (7.4%)	37,625 (8.6%)	20,510 (5.8%)	22,136 (4.3%)	26,245 (5.6%)	16,383 (4.1%)	34,247 (6.4%)	24,387 (6.5%)
insertion	756 (0.2%)	5993 (1.4%)	687 (0.2%)	823 (0.2%)	760 (0.2%)	11,821 (2.9%)	19,312 (3.6%)	18,738 (5.0%)

次に、次世代シーケンサーを用いてレポーター遺伝子ベクターに挿入された多様な変異プロモーター配列の転写活性化能を効率的に測定した。プールした変異プロモーター配列をバルクでHEK293細胞へトランスフェクションし、48時間後に変異プロモーター配列の下流に配置したGFP遺伝子のmRNAの発現量としてランダム配列のタグ数をIllumina Hi-Seq2000の36-nt

single-end reads法を用いて測定した (図2-A)。なお、転写活性化能の算出には、トランスフェクションしたプラスミド由来のランダムタグが基準となるため、下記の式で算出した。

$$\text{転写活性化能} = \frac{(\text{cDNA由来のランダム配列のタグ数} / \text{cDNA由来のすべてのランダム配列タグ数})}{(\text{プラスミド由来のランダム配列のタグ数} / \text{プラスミド由来のすべてのランダム配列タグ数})}$$

本手法で測定した変異プロモーター配列の転写活性化能を従来のルシフェラーゼアッセイを用いて確認した (図3)。その結果、相関係数は0.8と良好な値を示し、本手法での測定方法は、効率的、大規模でありながら、従来のルシフェラーゼアッセイと同様の転写活性化能測定が可能であると考えた。

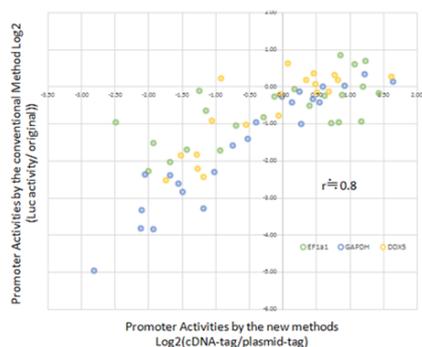


図3: 次世代シーケンズ技術を用いた手法とルシフェラーゼアッセイの比較  
測定した変異プロモーター配列の転写活性化能を、ルシフェラーゼアッセイで確認した。緑はEF1A1, 青はGAPDH, 黄はDDX5の変異型プロモーター

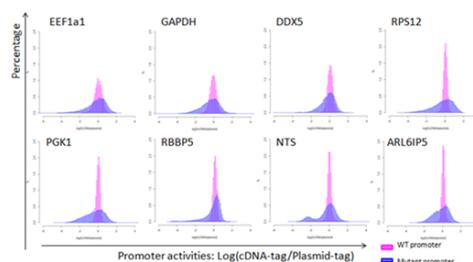


図4: プロモーター領域の転写活性化の分布

プロモーター配列の転写活性化能の分布。赤は野生型のプロモーター配列の転写活性化能、青は変異プロモーター配列の転写活性化能

図4に各遺伝子の野生型プロモーター配列及び変異プロモーター配列の転写活性化能の分布を示す。各遺伝子の変異プロモーターの転写活性化能について、野生型プロモーターに比して1/16~8倍の活性を示すものが得られた。EEF1A1遺伝子から得られた合計39万種類の配列-活性相関データに対し重回帰分析を用いてプロモーター領域内の各塩基の転写活性化能に対する影響を解析した。図5に示すように各塩基について、その塩基が置換した際に転写活性化能に及ぼす影響として、その塩基が固有に有する転写活性化能を算出することが可能であった。次世代シーケンサーを駆使した本研究によりDNA配列に内在的にコードされる塩基配列の持つ転写活性化能を解析することが可能となったと考えている。

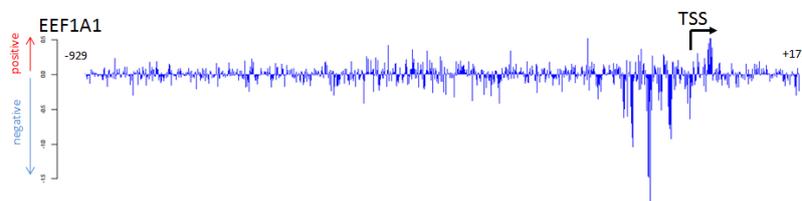


図5: EEf1A1遺伝子プロモーター配列の各塩基が有する転写活性化能

EEF1A1遺伝子のプロモーター配列の各塩基が、野生型より変異した時に、転写活性化能与える影響。変異がpositiveな影響は上に、negativeな影響は下にバーを示した

## 参考文献

1. Birney, E., et al., Nature, 447, 799-816, 2007
2. Sakakibara et al, DNA Res., 14, 71-77, 2007.
3. Irie T et al, Nucleic Acids Res., 2011 Jun 1;39(11):e75. Epub 2011 Apr 12.