

博士論文

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

榊原 雄太

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

平成 28 年 9 月

東京大学大学院 新領域創成科学研究科
メディカル情報生命 専攻
ゲノムシステム医療科学分野

榊原 雄太

目次

1.序論	3
1.1プロモーター領域の配列情報にコードされる転写制御機構を解明することの重要性	
1.2細胞内に存在する分子メカニズムを理解に向けた情報蓄積	
1.3 完全長cDNAプロジェクトとヒト遺伝子プロモーター配列の情報蓄積	
1.4 ヒト遺伝子のプロモーター配列の差が転写活性化能へ与える影響の解析	
1.5 本研究の目的	
2.材料及び方法	7
2.1 培養細胞	
2.2 遺伝子のプロモーター領域の同定	
2.3 プロモーター配列間の差が転写活性化能に与える影響の解析	
2.4 プロモーター配列の塩基レベルの差が転写活性化能に与える影響の解析	
3.結果と考察	24
3.1 遺伝子のプロモーター配列間の差が転写活性化能へ与える影響の解析	
3.2 プロモーター配列の塩基レベルの差が転写活性化能へ与える影響の解析	
3.3 遺伝子のプロモーター配列の転写活性化能とRNA-seqの発現量の比較	
4.総括	95
5.1 プロモーター配列の配列差が転写活性化能に与える影響の解析	
5.2 今後の展望	
6.参考文献	97
7.論文目録	101
8.謝辞	102

1 序論

1.1 プロモーター領域の配列情報にコードされる転写制御機構を解明することの重要性

ヒトは約 37 兆といわれる細胞から構成されている。個々の細胞は、それぞれに固有の遺伝子発現プログラムを有し、その細胞機能を実現、維持している[1]。遺伝子の発現は、転写開始、転写伸長、転写終結、転写産物の輸送、スプライシングさらには分解といった多段階の制御を受ける。中でもその最初の段階である転写開始反応は非常に重要な制御段階であると考えられている。転写開始反応は、遺伝子の転写開始点近傍のプロモーター領域に基本転写因子が結合、RNA ポリメラーゼ II がリクルートされ、転写開始前複合体を形成[2,3]することで開始される。この効率、DNA 配列が規定する配列要素や、その配列因子に転写因子と呼ばれる一群のタンパク質が結合することで制御される。転写開始反応に関与する要素因子がいくつか示唆されている。すなわち転写開始反応では、プロモーターやエンハンサーと呼ばれる領域の配列因子（シス因子）[4]、プロモーター、エンハンサーへ結合する転写制御タンパク質因子（トランス因子）、さらにはクロマチン構造といった高次構造因子やメチル化を中心とした塩基修飾因子等が織りなす複雑な制御により、最終的に転写開始量が決定される。

私は本研究において、遺伝子の転写開始を制御する最も基本的な因子として、プロモーター、エンハンサーといった転写制御領域の DNA の配列情報に着目した。これらの領域の一次元の DNA 配列に内在的にコードされる転写活性化能が、遺伝子発現制御において最も基盤的な制御機構であり、その解析は転写制御の全体像の解明に向けて、その起点となると考えた。

1.2 細胞内に存在する分子メカニズムを理解に向けた情報蓄積

2001 年に、国際ヒトゲノム解読計画の成果として、ヒトゲノム配列のドラフト配列の解読が完了した[5,6]。さらに 2003 年には、ヒトゲノム配列の完全解読が発表された[7]。これにより、ヒトゲノム配列の配列情報を利用し、ヒトゲノム配列中に存在する個々の遺伝子の遺伝子機能を包括的に理解する試みが可能となった。ヒトゲノム配列解読の成果として、ヒトゲノム配列全体は 30 億塩基対で構成され、その配列中には約 22000 遺伝子が存在していることが明らかとなった。タンパク質コード領域については、アミノ酸配列が推定され、その機能推定を行ういわゆるアノテーションが活発に行われた。一方で、プロモーター領域など、その遺伝子発現制御に関わる機能性領域の解析も着手された。しかし、当時のヒトゲノムの配列情報には、遺伝子の正確なエクソン開始位置すなわち転写開始点情報、上流プロモーター領域情報、近傍のクロマチン構造、といった情報が欠如しており、ヒトゲノム配列だけでは遺伝子の配列及びそれに伴う遺伝子機能など、生物学的な情報を抽出することは困難であった。

米国の ENCODE (The Encyclopedia of DNA Elements) プロジェクトや日本におけるゲノムネットワークプロジェクトなど、ゲノム配列中の生物学的な情報の理解を目的とした大規模プロジェクトが実施された。ENCODE プロジェクトでは、ゲノム配列中の機能性領域のカタログ化を目的

に、100 細胞以上の網羅的な遺伝子発現パターン、転写因子の結合情報やクロマチン構造などのデータが蓄積された[3]。ゲノムネットワークプロジェクトでは、遺伝子発現調節、遺伝子発現、タンパク質間相互作用などの情報蓄積が実施された。また、これらの成果により、ゲノムレベルでの機能性領域の理解に向けた、情報の蓄積及び情報収集や情報活用のための技術開発が進んだ。

しかし、これらのプロジェクトでは、一次元の DNA 配列であるヒトゲノム配列に内在的にコードされる転写制御機構の情報収集は、その解析手法の大規模化が困難であったために、実施項目に含まれず、関連する情報蓄積は依然として限られたものであった。

1.3 完全長 cDNA プロジェクトとヒト遺伝子プロモーター配列の情報蓄積

私の所属する研究室においては、ENCODE プロジェクトやゲノムネットワークプロジェクトを補完する位置付けとして、完全長 cDNA プロジェクトを実施してきた。ヒト遺伝子配列およびその配列構造を理解するためには、ヒト遺伝子のエクソン構造から構成される mRNA の配列解析が必要であった。mRNA に相補的なヒト完全長 cDNA を利用することで、ヒト遺伝子配列の配列解析が各国におけるヒト完全長 cDNA プロジェクトの形で行われた。当研究室では、新エネルギー・産業技術総合開発機構(NEDO)におけるプロジェクトの 1 つとして、当研究室で開発されたオリゴキャッピング法[8]を用いたヒト完全長 cDNA プロジェクト (FLJ プロジェクト) を主導した[9]。

オリゴキャッピング法により得られた完全長 cDNA は mRNA の 5' 端を有していることから、mRNA の転写開始点の決定が可能であった。決定された転写開始点をゲノム配列へマッピングすることで、転写開始点近傍のゲノム配列をプロモーター領域として同定することが可能であった[10,11]。これらの情報は転写開始点データベース DBTSS (<http://dbtss.hgc.jp/>) [12] として公開されており、ヒトプロモーター領域の配列解析が可能であった。また、完全長 cDNA を利用した解析により、遺伝子をコードしない遺伝子間領域にも多数の転写産物が存在すること[13]や、複数の転写開始点を有する遺伝子[14]も見つかった。それまで考えられてきた転写制御機構の全体像よりもさらに複雑な制御が行われている可能性が示唆されていた。

1.4 ヒト遺伝子のプロモーター配列の差が転写活性化能へ与える影響の解析

私は、利用可能となったヒトゲノム配列に依拠して完全長 cDNA により転写開始点が規定された 5' 端近位プロモーター領域の配列に関する実験的解析を試みた。ある細胞内に存在する遺伝子群のプロモーター領域に対して実験的および計算機的に統一的な実験条件下で大規模にその転写活性化能の測定を行うことで、細胞内に存在する遺伝子群の転写制御機構を全体像として明らかにすべくその要素還元的アプローチの起点とすることができると考えた。

本研究では、その第一段階として、ある特定の細胞内に存在する遺伝子群のプロモーター領域が一次元の DNA 配列として内在的に有する転写活性化能が、その細胞内において全体としてどのように分布するのかを明らかにすることを目的として、細胞内の遺伝子のプロモーター配列について大規模に転写活性化能を測定できる体系的なルシフェラーゼアッセイ系の構築及び遺伝子のプロモーター配列の差が測定した転写活性化能へ与える影響の解析を実施した[15]。

本研究以前の遺伝子のプロモーター領域に対する解析は、解析ごとに異なる細胞、異なる実験条件下で実施された研究であり、特定の細胞における転写制御機構の全体像の解明を目指したものではありません [16]。マイクロアレイや SAGE 法といった、網羅的に遺伝子発現を解析する手法は存在した。しかし、これらの手法は細胞中の mRNA レベルを解析する手法である。これは、上述した遺伝子発現制御のうち、プロモーターやエンハンサー、クロマチン構造といった制御を受け転写開始された RNA が、スプライシングや分解といった多段階の制御を受けた最終結果を測定したものである [17]。そのために、これらの手法で得られた計測結果は、配列要素の活性を直接示すものではない。

1.5 本研究の目的

本論文の第一章では、体系的なルシフェラーゼアッセイによる転写活性化能の測定系の構築と実践について論じる。ルシフェラーゼアッセイによる転写活性化能は、目的とする DNA 配列の転写活性化能をルシフェラーゼの酵素反応による化学蛍光強度として検出することで実施した。レポーターアッセイはプラスミド DNA を外来的に導入し、レポーター遺伝子を指標に、そのプロモーター配列の転写活性化能の計測を行う。本計測系では、クロマチン構造や mRNA の分解など、それ以外の要素が排除されるために、DNA 配列が有する内在的に有する転写活性化能が直接的に測定可能であると考えた [15]。

しかし、この解析手法ではスループットに限界があり、より効率的な実験手法の確立が必要であることも同時に明らかとなった。そこで本論文の第二章では、次世代シーケンサーを用いたさらにハイスループットな測定系の構築を行った。近年、従来のサンガー法に加えていわゆる次世代シーケンス法が広く用いられるようになってきている。[18-20]。次世代シーケンサーは、数百 bp の DNA 配列を数億本同時に解析することが可能である [21,22,23]。次世代シーケンサーのスループットを利用して、トランスクリプトーム解析 (RNA-seq [24,25,26]、TSS-seq [27])、転写因子結合領域やクロマチン構造を決定する ChIP-seq [28,29] など、様々な応用が開発され実践されている。

次世代シーケンサーを利用してハイスループットにプロモーター領域やエンハンサー領域の転写活性化能を測定する方法がいくつか提案されている [30-34]。MPRA (Massively Parallel Reporter Assay) [35] は、目的の配列とランダムタグ配列を 142 mer のオリゴ DNA として合成し、クローニングする手法である。合成オリゴを用いてプラスミドに DNA ライブラリーを構築し、細胞にトランスフェクションして mRNA として転写されたランダムタグを、次世代シーケンサーでカウントする。STARR-Seq (Self-Transcribing Active Regulatory Region Sequencing) [36] は、断片化したゲノム配列をクローニング、プラスミドライブラリーを作成、細胞にトランスフェクション後に mRNA として転写された断片化されたゲノム配列をカウントする。これにより、転写活性を有するゲノム領域を同定することが可能であるという。これらの手法は、遺伝子のプロモーター領域に塩基レベルでの欠損や塩基置換を導入した活性測定系への拡張が容易ではない。塩基レベルでの転写活性化能の測定を安価、高効率に行うためにはさらなる技術的開発が必要であると考えられた。

本論文の第二章では、プロモーター配列内の 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ため、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

成し、その配列決定後に転写活性化能を測定する測定系の構築を行った。

2 材料及び方法

2.1 培養細胞

ヒト培養細胞である Human Embryonic Kidney (HEK) 293 細胞 (ATCC number: CRL-1573) を用いた。

HEK293 細胞は、10% Fetal Bovine Serum (FBS)、0.584 mg/L L-グルタミン酸 (GIBCO)、0.15% 炭酸水素ナトリウム、60 mg/L カナマイシン (GIBCO) を含んだ、Dulbecco's Modified Eagle's Medium (DMEM) を用い、37°C、5%CO₂ インキュベーター内で培養を行った。

2.2 遺伝子のプロモーター領域の同定

HEK293 細胞由来の mRNA より、オリゴキャッピング法を用いて、完全長 cDNA ライブラリーを作成した[8]。作成した完全長 cDNA ライブラリーより、ランダムに 12504 種類のクローンを選択し、5'端配列を取得した。取得した 5'端配列をヒトゲノム配列(hg16, UCSC genome browser, <https://genome.ucsc.edu/>)へ、配列解析ソフトウェアである BLAT[37]及び SIM4[38]を用いてマッピングし、転写開始点を同定した。同定した転写開始点を基点に、上流 1000bp 及び下流 200bp の領域をプロモーター領域として同定した。最終的に、2170 遺伝子のプロモーター領域を同定した (転写開始点から上流 1.0kb および下流 0.2kb の領域には、転写因子結合配列の 82%が含まれる) [8, 10, 39-42]。遺伝子の転写開始点を同定する際、ひとつの遺伝子に複数の転写開始点がマッピングされた[43]。このような場合、転写開始点は完全長 cDNA クローンによるサポートが最も多い転写開始点を採用した。

また、本研究においては、ヒトゲノム配列から抽出された 2170 遺伝子のプロモーター領域の配列情報よりランダムに選んだ、852 遺伝子のプロモーター領域の配列情報を利用した。

2.3 プロモーター配列間の差が転写活性化能に与える影響の解析

2.3.1 遺伝子のプロモーター領域へのプライマー作成

各遺伝子のプロモーター領域について、転写開始点を 0bp としたときの-1000bp から -900bp 及び 100bp から 200bp の配列中に、5'-プライマー及び 3'-プライマーをそれぞれ設計した。プライマーの設計にはプライマー設計ソフトウェアである PRIMER3 (<http://bioinfo.ut.ee/primer3/>) を用いた。T_m 55.0°Cの設定で利用した。(転写活性化能の測定を行なうことができた遺伝子のプロモーター領域に用いたプライマーの詳細について(表 3.1)に示した。)

また、クローニング時に利用した、組み換えを原理とした Gateway Cloning System (Invitrogen) に必要な配列を付加した。具体的には、5'-プライマーの 5'端に attB1 サイトの一部 (5'-AAAAAGCAGGT-3')、3'-プライマーの 5'端に attB2 サイトの一部 (5'-AGAAAGCTGGGT-3') を付与した。

2.3.2 遺伝子のプロモーター配列の増幅

human genomic DNA (Clontech) をテンプレートとし、2.3.1 で設計した 5'-プライマー及び 3'-プライマー、KOD-plus PCR kit (TOYOBO) を用いて PCR を行った。PCR 条件は、94°C1 分、58°C1 分、68°C2 分を 35 サイクルに設定した。

各遺伝子のプロモーター配列に Gateway Cloning System で用いる完全な attB1 サイト及び attB2 サイトを付与するため、前記の PCR 産物をテンプレート、完全な attB1 サイト (5'-GGGGACAAGTTTGTACAAAAAAGCAGGT-3') 及び完全な attB2 サイト (5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3') を 5'-プライマー及び 3'-プライマー、KOD-plus PCR kit を用いた PCR を行った。PCR 条件は、94°C15 秒、55°C30 秒、68°C2 分を 20 サイクルとした。

2 段階の PCR により増幅された各遺伝子のプロモーター配列は、ポリエチレングリコール沈殿処理により精製及び濃縮した。濃縮した各遺伝子のプロモーター配列溶液は、SPECTRA max Plus 384 (Molecular Devices) を用い吸光度を測定し、測定結果をもとに精製水で 45 ng/μL に濃度調整した。

2.3.3 遺伝子のプロモーター配列のクローニング

2.3.2 で増幅した各遺伝子のプロモーター配列を、Gateway Cloning System を利用し、ホタルルシフェラーゼ遺伝子を有する pGL3 Basic ベクター (Promega) へ挿入した。pGL3 Basic ベクターは、Gateway Cloning System にもちいる配列カセットを SmaI サイトに挿入した、pGL3 Basic CaA ベクターを利用した。

Gateway Cloning System の反応は、まず 2.3.2 で調整した遺伝子のプロモーター配列 6.63μL、150 ng/μL pDONR 1.33 μL、BP Buffer 2μL、BP clonase 4μL を加え、25°C で一晩反応した。一晩反応した溶液に 150 ng/μL pGL3 Basic CaA 2μL、0.75M NaCl 0.66μL、LR clonase 4μL を加え、一晩反応した。一晩反応した溶液に、proteinase K 2μL を加え、37°C 10 分反応した。

反応後の溶液 1.5μL を用い、ヒートショック法にて Competent High E. Coli DH5a 15μL を形質転換した。形質転換した Competent High E. Coli DH5a を LB 寒天培地 (5g/L Bacto Yeast Extract (DIFCO)、10g/L Bacto Tryptone (BD)、15g/L Bacto Agar (BD)、10g/L 塩化ナトリウム、50mg/L アンピシリン)、37°C にて一晩培養した。寒天培地中のアンピシリンにより選択形成されたコロニーを、185μL の LB 培地 (5 g/L Bacto Yeast Extract (DIFCO)、10 g/L Bacto Tryptone (BD)、10g/L 塩化ナトリウム、50mg/L アンピシリン) に移し、37°C で一晩培養した。培養後、45μL の 80% グリセロール水溶液を添加し、グリセロールストックとして保存した。

2.3.4 クローニングした配列の確認

グリセロールストックとして保存したクローン 1 μ L、5'-プライマー (5'-CTAGCAAAATAGGCTGTCCC -3') 及び 3'プライマー (5'- GACGATAGTCATGCCCCGCG -3') 3.2 pmol、Ex Taq (Takara) を用いた PCR にて、配列確認に供する配列を増幅した。PCR 条件は、95 $^{\circ}$ C 15 秒、55 $^{\circ}$ C 15 秒、72 $^{\circ}$ C 4 分を 30 サイクルとした。増幅した配列より、ExoSap-IT (USB Corporation) により、未反応の dNTPs やプライマーを除去した。ExoSap-IT 処理後の配列 2 μ L、5'-プライマー (5'- GCCAGAACATTTCTCTATCG -3') 又は 3'プライマー (5'-CTTTATGTTTTTGGCGTCTTCC-3')、BigDye Terminator v3.1 Cycle Sequencing Kit (ABI) を利用し、シーケンス用サンプルを作成した。PCR 条件は、95 $^{\circ}$ C 10 秒、50 $^{\circ}$ C 5 秒、60 $^{\circ}$ C 2 分 30 秒を 30 サイクルとした。

5'端及び 3'端から増幅したシーケンス用サンプルはエタノール沈殿処理を行い、20 μ L の精製水に溶解した。調整したシーケンス用サンプルは ABI PRISM 3730 DNA Analyzer (ABI) を用いて配列を決定した。

2.3.5 遺伝子のプロモーター配列ではないゲノム配列中の DNA 断片のクローニング

遺伝子のプロモーター配列のコントロールとして、ゲノム配列中のランダムな約 1.0kb の DNA 断片をクローニングした。

まず、DNA 断片の増幅には、human genomic DNA をテンプレートとし、attB1 サイト (5'-GGGGACAAGTTTGTACAAAAAAGCAGGT-3') 及び attB2 サイト (5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3') を 5'-プライマー及び 3'-プライマー、Ex Taq を用いた PCR を行った。PCR 条件は、プライマーがテンプレートに容易にアニーリングできる条件として、95 $^{\circ}$ C 1 分、40 $^{\circ}$ C 1 分、72 $^{\circ}$ C 1 分で 20 サイクルとした。

増幅した DNA 断片溶液を、1%アガロースゲルへ電気泳動した。電気泳動の結果より QIAquick Gel Extraction Kit (Qiagen)を用い、約 1kbp (750bp~1200bp) の DNA 断片を抽出した。抽出した DNA 断片は、SPECTRA max Plus 384 を用い吸光度を測定し、測定結果をもとに精製水で 45 ng/ μ L に濃度調整した。

調整した DNA 断片は、遺伝子のプロモーター配列と同様 (2.3.3、2.3.4) に Gateway Cloning System を利用して pGL Basic ベクターへ挿入、クローニングした。プラスミド DNA に挿入された DNA 断片の配列を決定し、既知の遺伝子のプロモーター配列でない約 1.0kbp の DNA 配列であることを確認後、ランダム領域とした。

2.3.6 転写活性化能の測定

2.3.6.1 プラスミド DNA の精製

グリセロールストックより目的の配列が挿入されたクローンを、2mL の LB 培地で 20 時間培養した。培養後、QIAwell 96 Ultra Plasmid Kit (Qiagen) を用いてプラスミド DNA を抽出した。抽出したプラスミド DNA はイソプロパノール沈殿処理によ

り濃縮し、SPECTRA max Plus 384 を用い吸光度を測定した。測定結果をもとに精製水で 25ng/ μ L に濃度調整した。

2.3.6.2 HEK293 細胞へのトランスフェクション

HEK293細胞 5×10^3 cells及びDMEM培地100 μ Lを96 WELL CULTURE CLUSTER (Coster) へ分注し、24時間培養した。24時間培養後のHEK293細胞に、2.3.6.1にて精製後のプラスミドDNA 50ngをトランスフェクションした。トランスフェクションにはFuGENE6 Transfection Reagent(Roche)0.3 μ L及びDMEM FCS(-)培地9.7 μ Lを用いた。補正用の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5ngも同時にトランスフェクションした。

2.3.6.3 レポーター遺伝子アッセイ

トランスフェクションから48時間後の細胞を、Phosphate Buffer Saline (PBS : 8g/L NaCl 0.2g/L KCl, 1.44g/L Na_2HPO_4 , 0.24g/L KH_2PO_4) で2回洗浄後、Passive Lysis Buffer (Promega) 15 μ Lを用いて溶解した。

溶解したサンプル3 μ Lに、Luciferase assay reagent II 15 μ Lを加え、2秒静置後、5秒間ホタルルシフェラーゼ活性の測定を行った。その後、Stop and Glo溶液 15 μ Lを加え、2秒静置後、5秒間ウミシイタケルシフェラーゼ活性の測定を行った。活性測定には、Dual-Luciferase Reporter Assay System (Promega) 及びCentro XS3 LB960 (BERTHOLD) を用いた。

2.3.6.4 転写活性化能の算出

測定されたホタルルシフェラーゼ活性をウミシイタケルシフェラーゼ活性で除した値として算出し、測定した3回の平均値を代表値とした。

2.3.7 TATA box の探索

ATA box の検索は、以前の論文に従い、緩やかな条件(less-strict TATA box)および厳しい条件(strict TATA box)、2種類の検索を行った。[10]

各領域における転写開始点近傍領域(転写開始点を0としたとき、-90 から+27)のセンス鎖に対して、マトリックス検索を用い、less-strict TATA box を検索した。マトリックス検索として、転写因子結合配列検索ソフトウェア MATCH を利用し、転写因子結合配列データベースである TRANSFAC7.2 に含まれる V\$TATA_C および V\$TATA_01 を検索した。[16]

strict TATA box として、同様の転写開始点近傍領域におけるセンス鎖に対して TATA[T or A][T or A]配列を検索した。

ランダム領域について TATA box を検索する際、ランダム領域の 3'端から 195 bp 目(クローニングされた各ヒト遺伝子のプロモーター領域における、3'端から転写開始点までの塩基数の平均値)を仮想的な転写開始点として定めた。そして、仮想的に定義された転写開始点を 0 とした際、転写開始点近傍となる領域のセンス鎖に対して TATA box の検索を行った。

2.3.8 GC 含有率

GC 含有率は、各領域のセンス鎖に対して次式($[C]$ の数 $+[G]$ の数)/(領域の塩基数)により求めた。

2.3.9 CpG island の検索

CpG island の検索は、以前の論文に従い、次のような条件で行った。[10]

評価対象となる領域のセンス鎖より、5'端から始点が 1bp ずつ異なる 100 bp の DNA 配列をそれぞれ抽出し、抽出された各 DNA 配列について次の 2 つの評価を行った。

1)100bp の DNA 配列における GC 含有率(方法 3-6-2 と同様)が 50%より多い。2)100bp の DNA 断片において次式($[CG]$ の数 \times 領域の塩基数)/($[C]$ の数 \times $[G]$ の数) > 0.6 を満たす。これら 2 つの条件を満たす DNA 配列の始点が 200bp 連続した場合に CpG island(+)とした。

2.4 プロモーター配列の塩基レベルの差が転写活性化能に与える影響の解析

2.4.1 変異プロモーター配列-GFP 遺伝子-ランダムタグ配列のクローニング

変異プロモーターは、遺伝子の野生型のプロモーター配列より高効率に塩基置換を導入する Error Prone PCR 法を用いて作成した。変異プロモーター配列、転写活性化能の測定時に利用する GFP 遺伝子、多様な変異プロモーター配列を識別するランダムタグ配列を連結した DNA 断片を作成し、pGL3 Basic ベクターへクローニングした。

2.4.1.1 変異プロモーター配列の増幅

各遺伝子の野生型プロモーター配列より、Error Prone PCR 条件[44]を用い、変異プロモーター配列を増幅した。Error Prone PCR 条件として、Taq ポリメラーゼ、不均衡な dNTP 濃度 (2mM dCTP 及び dTTP、10mM dGTP 及び dATP)、50mM MgCl₂ 及び 5mM MnCl₂ を用いた。PCR による増幅は 2 段階で行い、非特異な断片の増加を抑制した。25 ng/ μ L に調整された遺伝子の野生型プロモーター配列の挿入されたプラスミド DNA 溶液を用いて、1/4 希釈ずつ 7 段階の希釈を行い、8 種類の濃度のプラスミド DNA 溶液を調整した。調整した各濃度のプラスミド DNA 1 μ L をテンプレートとし、5'-プライマー (5'-CTAGCAAAATAGGCTGTCCC-3') 及び 3'-プライマ

ー (5'-CTTCCAGCGGATAGAATGGC-3') を用い、エラーの入りやすい PCR 条件により DNA 断片を増幅した。PCR は、95°C 15 秒、55°C 15 秒、72°C 2 分を 40 サイクルとした。各濃度のプラスミド DNA をテンプレートとして増幅した DNA 断片溶液を、1%アガロースゲルへ電気泳動した。PCR による増幅が確認できた最も薄いサンプルについて、QIAquick Gel Extraction Kit (Qiagen)を用い、30 μL の精製水で抽出した。

抽出された DNA 断片を 1/4 希釈ずつ 7 段階の希釈を行い、8 種類の濃度の DNA 断片溶液を調整した。調整した各濃度の DNA 断片溶液 1 μL をテンプレートとし、5'-プライマー (5'-GCCAGAACATTTCTCTATCG-3') 及び 3'-プライマー (5'-GACGATAGTCATGCCCCGCG-3') を用い、エラーの入りやすい PCR 条件により DNA 断片を増幅した。PCR は、95°C 15 秒、55°C 15 秒、72°C 2 分を 40 サイクルとした。各濃度の DNA 断片溶液をテンプレートとして増幅した DNA 断片溶液を、1%アガロースゲルへ電気泳動した。電気泳動の結果、PCR による増幅が確認でき且つラインの最も薄いサンプルについて、QIAquick Gel Extraction Kit (Qiagen)を用い、DNA 断片を抽出した。なお、30 μL の精製水で抽出した。

増幅した変異プロモーター配列の 5'端に In-Fusion クローニングによりベクターに組み込むための 15 塩基の配列を、3'端に GFP 遺伝子とオーバーラップする配列を付与するため、5'プライマー (5'-GATCTAAGTAAGCTTGGGGACAAGTTTGTACAAAAAAGCAGGCT-3') 及び 3'プライマー (5'-CTCACCATGGTGGCGGGGACCACTTTGTACAAGAAAGCTGGGT-3')、KOD-plus PCR kit を用いた PCR で増幅した。PCR 条件は、94°C15 秒、55°C30 秒、68°C2 分を 20 サイクルとした。増幅した変異プロモーター配列は QIAquick Gel Extraction Kit を用い、30 μL の精製水で抽出した。

2.4.1.2 GFP 遺伝子の増幅

pAcGFP1-1 ベクター(Clontech)の GFP 遺伝子領域を、5'プライマー (5'-CGCCACCATGGTGAGCAAGG-3') 及び 3'プライマー (5'-CGCTCACTTGTACAGCTCAT-3')、KOD-plus PCR kit を用いた PCR で増幅した。PCR 条件は、94°C15 秒、55°C30 秒、68°C2 分を 20 サイクルとした。不要なプラスミドを分解するため、2μL DpnI(New England Biolabs)を 1 時間処理し、反応後のサンプルを 1%アガロースゲルへ電気泳動、QIAquick Gel Extraction Kit を用い、目的の DNA 断片を抽出した。

2.4.1.3 ランダムタグ配列の作成

ランダムな 12 塩基配列であるランダムタグ配列を設計した。ランダムタグ配列の 5'端に GFP 遺伝子とオーバーラップする配列、3'端に In-Fusion クローニングにより

ベクターに組み込むための 15 塩基の配列を付与した (5'-CGCCCCGACTCTAGANNNNNNNNNNNNCTGCTGCGCTCACTTGTACAG-3')。ランダムタグ配列の作成は、eurofins のカスタムオリゴ DNA 合成サービスを利用した。

2.4.1.4 変異プロモーター配列-GFP 遺伝子-ランダムタグ配列の作成

2.4.1.1、2.4.1.2、2.4.1.3 で増幅した、変異プロモーター配列、GFP 遺伝子、ランダムタグ配列を Overlap extension PCR 法により結合した。

変異プロモーター配列 50 ng、GFP 遺伝子 50 ng、12 塩基のランダムタグ配列を含む DNA 断片を、KOD-plus PCR kit を用いた PCR により結合した。PCR は、94°C15 秒、55°C30 秒、68°C3 分を 10 サイクルとした。

反応後のサンプルをテンプレートとし、5'プライマー (5'-GATCTAAGTAAGCTTGGGGA-3') 及び 3'プライマー (5'-CGCCCCGACTCTAGA-3')、KOD-plus PCR kit を用いた PCR で増幅した。PCR 条件は、94°C15 秒、55°C30 秒、68°C2 分を 15 サイクルとした。

反応後のサンプルを 1%アガロースゲルへ電気泳動、QIAquick Gel Extraction Kit を用い、目的の DNA 断片を抽出した。

2.4.1.5 pGL3 Basic ベクターの増幅

In-Fusion クローニングを用い、pGL Basic ベクターへ変異プロモーター配列-GFP 遺伝子-ランダムタグ配列を挿入するため、pGL3 Basic ベクターのルシフェラーゼ遺伝子以外の領域を増幅した。

pGL3 Basic ベクターのルシフェラーゼ遺伝子以外の領域、5'プライマー (5'-TCTAGAGTCGGGGCGGCCGG-3') 及び 3'プライマー (5'-AAGCTTACTTAGATCGCAGA-3')、KOD-plus PCR kit を用いた PCR で増幅した。PCR 条件は、94°C15 秒、55°C30 秒、68°C2 分を 20 サイクルとした。

2µL DpnI(New England Biolabs)を 1 時間処理し、反応後のサンプルを 1%アガロースゲルへ電気泳動、QIAquick Gel Extraction Kit (Qiagen)を用い、目的の DNA 断片を抽出した。

2.4.1.6 変異プロモーター配列-GFP 遺伝子-ランダムタグ配列のクローニング

2.4.1.4 で作成した変異プロモーター配列-GFP 遺伝子-ランダムタグ配列を、2.4.1.5 で増幅した pGL3 Basic ベクターへ、In-Fusion HD Cloning Kit(Clontech)を用いて、クローニングした。

In-Fusion クローニングの反応条件は、In-Fusion HD 8 µL、pGL3 Basic ベクター6 µL (100 ng)、変異プロモーター配列-GFP 遺伝子-ランダムタグ配列 12µL (100 ng)、

精製水 14 μ L を 50 $^{\circ}$ C 15 分反応した。

反応後のサンプル 10 μ L 及び NEB 5-alpha Competent E. coli (New England Biolabs) 100 μ L を用い、ヒートショック法 (氷上 30 分、42 $^{\circ}$ C 45 秒、氷上 2 分、SOC 培地を 1000 μ L 加え、37 $^{\circ}$ C 30min) による形質転換を実施した。形質転換は 4 回実施した。形質転換した NEB 5-alpha Competent E. coli を LB 培地 (5g/L Bacto Yeast Extract (BD)、10 g/L Bacto Tryptone (BD)、10 g/L 塩化ナトリウム (WAKO) にて 500mL 調整し、50mg アンピシリン (WAKO) を添加)、37 $^{\circ}$ C にて一晩培養した。培養後、QIAfilter Plasmid Maxi Kit (Qiagen) を用いてプラスミド DNA を抽出した。抽出したプラスミド DNA はイソプロパノール沈殿処理による濃縮後、Nano Drop ND-1000 を用い吸光度を測定した。吸光度の測定結果をもとに精製水で 1 μ g/ μ L に濃度調整し、プラスミドライブラリーとして保存した。

形質転換した NEB 5-alpha Competent E. coli 20 μ L を LB 寒天培地 (5g/L Bacto Yeast Extract、10g/L Bacto Tryptone、15g/L Bacto Agar (BD)、10g/L 塩化ナトリウム、50mg/L アンピシリン)、37 $^{\circ}$ C にて一晩培養し、コロニー数をカウントすることでライブラリーの種類数を推定、ライブラリーサイズが 10,000 種類を超えていることを確認した。

2.4.2 クローニングした変異プロモーター配列-ランダムタグ配列の配列決定

変異プロモーター配列及びランダムタグ配列の全長は約 1.2kb となり、Hiseq2000 (Illumina) では一度に配列決定が行えない。このため、トランスポゾン配列をランダムに挿入し、トランスポゾン配列とランダムタグ配列を利用した 3' 端側及び 5' 端側双方のシーケンス用ライブラリーを作成の上、変異プロモーター配列の両端から全長配列の決定を行った。

2.4.2.1 変異プロモーター配列の 3' 端側シーケンス用ライブラリーの作成

2.4.2.1.1 変異プロモーター配列-ランダムタグ配列の増幅

2.4.1 で作成したプラスミド DNA より、5' 端より変異プロモーター配列-ランダムタグ配列となる DNA 断片を増幅した。

まず、2.4.1.6 で精製したプロモーター配列-GFP 遺伝子-ランダムタグ配列を挿入したプラスミド DNA をテンプレートとし、5' プライマー (5'-ACAAGTGAGCGCAGC-3') 及び 3' プライマー (5'-ACAAGAAAGCTGGGT-3')、KOD-plus PCR kit を用いた inverse PCR で、変異プロモーター配列-ランダムタグ配列の DNA 断片を増幅した。PCR 条件は、94 $^{\circ}$ C 15 秒、55 $^{\circ}$ C 15 秒、68 $^{\circ}$ C 6 分を 5 サイクルとした。PCR 後のサンプルより 2 μ L DpnI を 1 時間処理して、不要なプラスミドを分解した。

増幅した DNA 断片をセルフライゲーションにより環状化 DNA とした。1/10 希釈した PCR サンプル 2 μ L、Ligation High (TOYOBO) 5 μ L、T4 Polynucleotide

Kinase (TOYOBO) 1 μ L、精製水 7 μ L を用い、増幅した DNA 断片をセルフライゲーション反応を行った。反応は 16°C で 1 時間とした。セルフライゲーション反応後のサンプルは、Plasmid-Safe ATP-Dependent DNase を用い、セルフライゲーション未反応生成物である直鎖 DNA を分解した。反応は、25mM dATP 1 μ L、Plasmid-Safe ATP-Dependent DNase 1 μ L で、37°C 1 時間とした。反応後のサンプルより、AMPure XP Kit (BECKMAN COULTER) を用い、環状 DNA とした変異プロモーター配列-ランダムタグ配列を精製した。

環状化した変異プロモーター配列-ランダムタグ配列をテンプレートに、5'プライマー (5'-GCCAGAACATTTCTCTATCG-3') 及び 3'プライマー (5'-GTGGTTTGTCCAAACTCATC-3')、Q5 High-Fidelity 2x Master Mix (New England Biolabs) を用いた PCR で、変異プロモーター配列-ランダムタグ配列を増幅した。PCR 条件は、98°C 10 秒、55°C 30 秒、72°C 45 秒を 5 サイクルとした。増幅した DNA 断片は、QIAquick PCR Purification Kit (Qiagen) を用いて精製した。

2.4.2.1.2 トランスポゾンを用いたシーケンス用ライブラリーの作成

2.4.2.1.2 で増幅した変異プロモーター配列-ランダムタグ配列に、EZ-Tn5 <KAN-2> Insertion Kit (AR BROWN) を用いて、トランスポゾン配列をランダムに挿入した。反応は、EZ-Tn5 10x Reaction buffer 1 μ L、増幅した DNA 断片 150ng、Transposon 1 μ L、EZ-Tn5 Transposase 1 μ L を用い、37°C で 2 時間とした。トランスポゾン挿入後、AMPure XP Kit で精製した。

トランスポゾン挿入後の変異プロモーター配列-ランダムタグ配列をテンプレート、トランスポゾンの配列にシーケンス用アダプターを付加した 5'-プライマー (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTC CGATCTTGAGATGTGTATAAGAGACA -3') 及び 3'プライマー (5'-CAAGCAGAAGACGGCATAACGAGAT-illumina index sequence)-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT -(N or NN or NNN)-GCCGCCCCGACTCTA-3')、Q5 High-Fidelity 2x Master Mix を用いた PCR で、シーケンス用ライブラリーを作成した。PCR 条件は、98°C 10 秒、55°C 30 秒、72°C 45 秒を 15 サイクルとした。PCR 後のサンプルより、1 本鎖 DNA を分解するために、Exonuclease I 1 μ L を用い、37°C 1 時間反応した。反応後のサンプルより DNA 断片を AMPure XP Kit にて精製した後、2% アガロースゲルへ電気泳動し、350bp から 1000bp を QIAquick Gel Extraction Kit (Qiagen) を用い抽出した

2.4.2.2 変異プロモーター配列の 5' 端側シーケンス用ライブラリーの作成

2.4.2.2.1 ランダムタグ配列-変異プロモーター配列の増幅

2.4.1 で作成したプラスミド DNA より、5'端よりランダムタグ配列-変異プロモーター配列となる DNA 断片を増幅した。

2.4.1.6 で精製したプロモーター配列-GFP 遺伝子-ランダムタグ配列を挿入したプラスミド DNA をテンプレートとし、5'プライマー (5'-ACAAGAAAGCTGGGT-3') 及び3'プライマー (5'-CGCCCCGACTCTAGA-3')、KOD-plus PCR kit を用いた PCR で、変異プロモーター配列-GFP 遺伝子-ランダムタグ配列の DNA 断片を増幅した。PCR 条件は、94°C 15 秒、55°C 15 秒、68°C 6 分を 5 サイクルとした。2µL の DpnI を 1 時間処理して、未反応プラスミドを分解した。

増幅した DNA 断片をセルフライゲーションにより環状化 DNA とした。1/10 希釈した PCR サンプル 2 µL、Ligation High (TOYOBO) 5µL、T4 Polynucleotide Kinase (TOYOBO) 1 µL、精製水 7 µL を用い、増幅した DNA 断片をセルフライゲーション反応を行った。反応は 16°C で 1 時間とした。セルフライゲーション反応後のサンプルは、Plasmid-Safe ATP-Dependent DNase を用い、セルフライゲーションしなかった直鎖 DNA を分解した。反応は、25mM dATP 1uL、Plasmid-Safe ATP-Dependent DNase 1uL で、37°C 1 時間とした。反応後のサンプルより、AMPure XP Kit (BECKMAN COULTER) を用い、環状 DNA とした変異プロモーター配列-ランダムタグ配列を精製した。

環状 DNA とした変異プロモーター配列-ランダムタグ配列をテンプレート、5'プライマー (5'-ATGGATGAGCTGTACAAGTG-3') 及び 3'プライマー (5'-CGCCATTCAGCTCGATCAGG-3')、Q5 High-Fidelity 2x Master Mix (New England Biolabs) を用いた PCR で、ランダムタグ配列-変異プロモーター配列を増幅した。PCR 条件は、98°C 10 秒、55°C 30 秒、72°C 45 秒を 5 サイクルとした。増幅した DNA 断片は、QIAquick PCR Purification Kit (Qiagen) を用いて精製した。

2.4.2.2.2 トランスポゾンを用いたシーケンス用ライブラリーの作成

2.4.2.1.2 で増幅したランダムタグ配列-変異プロモーター配列に、EZ-Tn5 <KAN-2> Insertion Kit (AR BROWN) を用いて、トランスポゾン配列をランダムに挿入した。反応は、EZ-Tn5 10x Reaction buffer 1 µL、増幅した DNA 断片 150ng、Transposon 1 µL、EZ-Tn5 Transposase 1 µL を用い、37°C で 2 時間とした。トランスポゾン挿入後、AMPure XP Kit で精製した。

トランスポゾン挿入後の変異ランダムタグ配列-変異プロモーター配列をテンプレート、トランスポゾンの配列にシーケンス用アダプターを付加した 5'-
プ ラ イ マ ー (5'-
AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTTC

CGATCTTGAGATGTGTATAAGAGACA -3') 及び 3'プライマー (5'- C CAAGCAGAAGACGGCATAACGAGAT-(illumina index sequence)- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-(N or NN or NNN)- ACAAGTGAGCGCAGC-3')、Q5 High-Fidelity 2x Master Mix を用いた PCR で、シーケンス用ライブラリーを作成した。PCR 条件は、98°C 10 秒、55°C 30 秒、72°C 45 秒を 15 サイクルとした。PCR 後のサンプルより、1 本鎖 DNA を分解するために、Exonuclease I 1 µL を用い、37°C 1 時間反応した。反応後のサンプルより DNA 断片を AMPure XP Kit にて精製した後、2%アガロースゲルへ電気泳動し、350bp から 1000bp を QIAquick Gel Extraction Kit (Qiagen)を用い抽出した。

2.4.2.3 変異プロモーター配列-ランダムタグ配列のシーケンシング

2.4.2.1.2 及び 2.4.2.2.2 で抽出した DNA 断片をテンプレート、5'-プライマー (5'- AATGATACGGCGACCACCGAG -3') 及び 3'プライマー (5'- CAAGCAGAAGACGGCATAACG -3')、Q5 High-Fidelity 2x Master Mix を用いて、シーケンス用の DNA 断片を増幅した。PCR 条件は、98°C 5 秒、55°C 10 秒、72°C 45 秒を 10 サイクルとした。反応後のサンプルより DNA 断片を AMPure XP Kit にて精製した後、バイオアナライザ (Agilent) でシーケンス用 DNA 断片の品質を確認し、HiSeq2000 を用いた 101bp Paired-End Rapid Run でシーケンスを行った。なお、シーケンシングプライマーは、トランスポゾン配列を利用した Read1 プライマー (5'-ACACTCTTTCCTACACGACGCTCTTCCGATCTTGAGATGTGTATAAGAGACA-3') 及びランダムタグ配列側の Read2 プライマーを用いた。変異プロモーター配列の 5'端側及び 3'端側ライブラリーの Read2 プライマーは、それぞれ 5'端側ライブラリー Read2 プライマー (5'- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCGCCCCGACTCTAGA-3')、3'端側ライブラリー Read2 プライマー (5'- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACAAGTGAGCGCAGCAG-3') を用いた。Index プライマーは Illumina 純正のプライマーを用いた。

2.4.2.4 変異プロモーター配列-ランダムタグ配列の配列決定

2.4.2.3 で得られた変異プロモーター配列の 5'端側及び 3'端側ライブラリーのシーケンシング結果を利用し、次の手順で配列決定を行った。

シーケンスで得られたリードのうち、Read1 は 80 %以上の塩基でクオリティスコアが 20 以上のリード、Read2 はランダムタグ配列の全ての塩基でクオリティスコアが 20 以上のリードを用い、変異プロモーター配列とランダムタグ配列のリードペアとして解析に用いた。

ランダムタグ配列ごとに、リードペアを野生型プロモーター配列へマッピングし

た。マッピングにはアライメントツールである BWA (Burrows-Wheeler Aligner) [45] を用いた。マッピング結果より、予想されるストランドとは異なる方向へマッピングされたリード、プロモーター配列の半分以上の長さでマッピングされたリードは信頼性が低いとみなし除外した。

マッピングされたリードより、Picard Tools の MarkDuplicates を使い、重複配列を除外、SAM tools[46]を用い、プロモーター配列内の変異を特定した。

プロモーター配列内の 90%の塩基で、5 リード以上マッピングされた変異プロモーター配列を以後の解析に利用した。

2.4.3 転写活性化能の測定

2.4.3.1 HEK293 細胞へのトランスフェクション

HEK293細胞 1×10^6 cells及びDMEM培地10 mLを10 cm² Dish (Greiner bio-one) へ分注し、24時間培養した。24時間培養後のHEK293細胞に、2.4.1.6にて精製後のプラスミドDNA 6 μ gをトランスフェクションした。トランスフェクションには X-treme-Gene9 transfection reagent (Roche) 30 μ L及びOpti-MEM Reduced Serum Medium (Life Technology) 361.3 μ Lを用いた。この際、コントロール用として6種類のプラスミドDNA (異なる3種類のランダムタグ配列を付与したGAPDH遺伝子のプロモーター配列、異なる3種類のランダムタグ配列のみを挿入したプラスミドDNA) も同時にトランスフェクションした。

2.4.3.2 mRNA 抽出と cDNA 合成

トランスフェクションから48時間後のHEK293細胞を回収、RNeasy Midi Kit (Qiagen)を用いてtotal RNAを抽出した。その後、Magnosphere UltraPure mRNA Purification Kit (TaKaRa) を使い、poly-Aを有するmRNAを抽出した。抽出後のサンプルより、TURBO DNA-free Kit (Life Technologies) を使い、DNAを分解した。反応条件は、mRNA 500 ng、TURBO DNase 1 μ Lを用い、37°Cで25分とした。

次に、SuperScript III Reverse Transcriptase kit (Life Technologies) を使い、逆転写によりcDNAを合成した。反応には、mRNA 10 μ Lに逆転写用プライマー (5'-GCAGCTTATAATGGTTACAA-3') を用いた。反応後のサンプルに、Ribonuclease H (Life Technologies) 1 μ Lを37°C 20分処理することで、RNAを分解した。

反応後のサンプルを用い、リアルタイムPCRを利用することで、cDNAの合成を確認した。リアルタイムPCRには、5'-プライマー (5'-CCTGGGCAATAAGATGGAGT-3') 及び 3'-プライマー (5'-GAAGCGGCCGCGCCCGA-3')、Power SYBR Green PCR Master Mix (Applied Biosystems) を使い、7900HT Fast Real-Time PCR System (Applied Biosystems) で実施した。

2.4.3.3 ランダムタグ配列の検出

2.4.3.1 でトランスフェクションに用いたプラスミド DNA 及び 2.4.3.2 で合成した cDNA をテンプレート、シーケンス用アダプターを付加した 5'-プライマー (5'-CAAGCAGAAGACGGCATAACGAGAT-(illumina index sequence)-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTGGGCAATAAGATGGAGT-3') 及び 3' プライマー (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAGCGGCCGGCCGCCCGA-3')、Q5 High-Fidelity 2x Master Mix を用いた PCR で、シーケンス用ライブラリーを作成した。PCR 条件は、98°C 10 秒、55°C 30 秒、72°C 30 秒を 5 サイクルとした。反応後のサンプルより、DNA 断片を AMPure XP Kit にて精製した。

精製した DNA 断片をテンプレート、5'-プライマー (5'-AATGATACGGCGACCACCGAG -3') 及び 3' プライマー (5'-CAAGCAGAAGACGGCATAACG -3')、Q5 High-Fidelity 2x Master Mix を用いて、シーケンス用の DNA 断片を増幅した。PCR 条件は、98°C 10 秒、55°C 30 秒、72°C 30 秒を 10 サイクルとした。反応後のサンプルより DNA 断片を AMPure XP Kit にて精製した後、バイオアナライザ (Agilent) でシーケンス用 DNA 断片の品質を確認し、HiSeq2000 を用いた 36bp Single-End Run でシーケンスを行った。なお、シーケンシング プライマー は 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAGCGGCCGGCCGCCCGACTCTAGA-3'を用いた。Index プライマーは Illumina 純正のプライマーを用いた。

2.4.3.4 転写活性化能の算出

2.4.3.4 で得られたプラスミド DNA 及び cDNA のランダムタグ配列のシーケンシング結果を利用し、次の手順で転写活性化能を算出した。

シーケンス結果が得られたリードのうち、ランダムタグ配列 12 塩基すべてでクオリティスコアが 20 以上のリードを解析に用いた。

各リードよりランダムタグ配列を抽出し、各ランダムタグ配列をカウントした。プラスミド DNA は 100 以上、cDNA は 1 以上カウントされたランダムタグ配列を解析に用いた。コントロール用として加えた 6 種類のランダムタグ配列とシーケンス結果が 1 塩基違いのランダムタグ配列は、シーケンスエラーと判断し、カウントから除外した。

プラスミド DNA、cDNA のそれぞれに、各ランダムタグ配列数を総ランダムタグ配列数で除した値を算出、ランダムタグ配列ごとに cDNA の算出値をプラスミド DNA の算出値で除した値を転写活性化能として算出した。

2.4.3.5 野生型プロモーター配列の転写活性化能の測定

コントロールとして、野生型プロモーター配列についても、2.4.1、2.4.2、2.4.3.1-3と同様の方法で、転写活性化能の測定を行った。

25 ng/μL に調整された遺伝子の野生型プロモーター配列の挿入されたプラスミド DNA をテンプレートとし、5'-プライマー (5'-CTAGCAAAATAGGCTGTCCC-3') 及び 3'-プライマー (5'-CTTCCAGCGGATAGAATGGC-3') を用い、PCR により DNA 断片を増幅した。PCR 条件は、95°C 15 秒、55°C 15 秒、72°C 2 分を 40 サイクルとした。

増幅した DNA 断片は、2.4.1、2.4.2、2.4.3.1-3 と同様の方法で、野生型プロモーター配列-GFP 遺伝子-ランダムタグ配列を pGL3 Basic ベクターに挿入し、配列の確認、転写活性化能の測定を行った。

2.4.3.6 塩基レベルの変異が転写活性化能へ与える影響の解析

2.4.3.6.1 各塩基の変異有無が転写活性化能へ与える影響の解析

説明変数として転写活性化能、目的変数として塩基ごとの変異有無とした重回帰モデルを用い、変異の有無が転写活性化能へ与える影響を偏回帰係数として推定した。偏回帰係数が 0 か否かの検定として、推定された偏回帰係数及びその標準誤差を利用した t 検定を実施した。重回帰分析には統計解析ソフトウェア R[47]を用いた。

$$(1) \log_2 y = \beta_0 + \sum \beta_i x_i$$

それぞれの変数は下記を表す。

y: 転写活性化能

β_0 : 回帰切片

β_i : 転写開始点を基点とした i 番目の塩基の偏回帰係数

x_i : 転写開始点を基点とした i 番目の塩基の変異有無 (変異無=0/変異有=1)

2.4.3.6.2 各塩基の変異パターンが転写活性化能へ与える影響の解析

説明変数として転写活性化能、目的変数として塩基ごとの各変異パターン有無とした重回帰モデルを用い、各変異パターンが転写活性化能へ与える影響を偏回帰係数として推定した。偏回帰係数が 0 か否かの検定として、推定された偏回帰係数及びその標準誤差を利用した t 検定を実施した。なお、各変異パターンとは、欠損及び A、T、G、C それぞれへの変異を表す。重回帰分析には統計解析ソフトウェア R を用いた。

$$(2) \log_2 y = \beta_0 + \sum \beta_{ij} x_{ij}$$

それぞれの変数は下記を表す。

y: 転写活性化能

β_0 : 回帰切片

β_{ij} : 転写開始点を基点とした i 番目の塩基の各変異パターン j の偏回帰係数

x_{ij} : 転写開始点を基点とした i 番目の塩基の各変異パターン j 有無

(変異無=0/変異有=1)

2.4.3.6.3 10 分割交差検証

2.4.3.5.1 及び 2.4.3.5.2 で利用した重回帰モデルの予測性能を検証する目的で、10 分割交差検証を行った。重回帰分析に利用するデータを 10 分割し、9 割のデータを用いた重回帰モデルの構築、残り 1 割のデータを用いて重回帰モデルによる予測値と観測値の比較を行った。予測値と観測値の検証は、重回帰モデルの検証に用いるデータを変更しながら 10 回実施し、得られたすべての予測値と観測値のデータよりピアソンの相関係数を算出した。重回帰モデルの構築及び相関係数の算出には統計解析ソフトウェア R を用いた。

2.4.3.7 転写因子結合配列の探索

遺伝子のプロモーター配列中の転写因子結合配列の探索には、転写因子結合配列探索ソフトウェア MATCH を利用した[48]。転写因子のデータベースは TRANSFAC 2014.1 を用いて、False Positive を最小にするパラメーターセットで探索を行なった[49]。TATA ボックスの探索には、データベース上の転写開始点を基準(0)とした際に、-90~+27 の領域にデータベース中の VTATA_01 又は文字列 TATA[T/A][T/A]と一致するものを TATA ボックスとした。

2.4.3.8 プロモーター配列中 5 塩基ごとの GC 含有率

プロモーター配列中の 5 塩基あたりの GC 含有率は、各領域のセンス鎖の連続する 5 塩基を抽出し、次式([C]の数+[G]の数)/(領域の塩基数=5)で求めた。

2.4.3.9 配列保存度

プロモーター配列の各塩基の配列保存度は PhastCons[50]の conservation scores を利

用した。

2.4.4 転写活性化能の確認

2.4.4.1 本研究で測定した転写活性化能を従来のルシフェラーゼアッセイで確認

2.4.3.4 で算出した転写活性化能を、従来のルシフェラーゼアッセイを用いて算出した転写活性化能と比較し、確認を行った。

2.4.1.6 で保存したプラスミド DNA ライブラリーを用い、ヒートショック法にて NEB 5-alpha Competent E. coli を形質転換、LB 寒天培地 (5g/L Bacto Yeast Extract、10g/L Bacto Tryptone、15g/L Bacto Agar (BD)、10g/L 塩化ナトリウム、50mg/L アンピシリン)、37°Cにて一晚培養した。寒天培地中のアンピシリンにより選択形成されたコロニーを、ランダムに選択し、200 µL の LB 培地 (5 g/L Bacto Yeast Extract (DIFCO)、10 g/L Bacto Tryptone (BD)、10g/L 塩化ナトリウム、50mg/L アンピシリン) に移し、37°Cで一晚培養した。培養後、45µL の 80%グリセロール水溶液を添加し、グリセロールストックとして保存した。

グリセロールストックとして保存したクローン 1µL、5'-プライマー (5'-CTAGCAAAATAGGCTGTCCC -3') 及び 3' プライマー (5'-GACGATAGTCATGCCCCGCG -3') 3.2 pmol、Ex Taq を用いた PCR にて、配列確認に供する配列を増幅した。PCR 条件は、95°C 15 秒、55°C 15 秒、72°C 4 分を 30 サイクルとした。増幅した配列、ランダムタグ配列用プライマー (5'-GTGGTTTGTCCAAACTCATC-3')、5'-プライマー (5'-GCCAGAACATTTCTCTATCG -3') 又は 3' プライマー (5'-CTTTATGTTTTTGGCGTCTTCC-3')、BigDye Terminator v3.1 Cycle Sequencing Kit (ABI) を利用し、シーケンス用サンプルを作成した。PCR 条件は、95°C 10 秒、50°C 5 秒、60°C 2 分 30 秒を 30 サイクルとした。シーケンス用サンプルは、イソプロパノール沈殿を行い、風乾した後、20µL のホルムアミドに溶解した。調整したシーケンス用サンプルは ABI PRISM 3730 DNA Analyzer (ABI) を用いて配列を確認した。

ランダムタグ配列、変異プロモーター配列が確認されたクローンについて、グリセロールストックより、2mL の LB 培地で 16 時間培養した。培養後、QIAwell 96 Ultra Plasmid Kit (Qiagen) を用いてプラスミド DNA を抽出した。

抽出したプラスミド DNA をテンプレートとし、attB1 サイト (5'-GGGGACAAGTTTGTACAAAAAAGCAGGT-3') 及び attB2 サイト (5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3') を 5'-プライマー及び 3'-プライマー、KOD-plus PCR kit (TOYOBO) を用いた PCR にて、各遺伝子のプロモーター配列を増幅した。PCR 条件は、94°C1 分、58°C1 分、68°C2 分を 35 サイクルとした。増幅した DNA 断片溶液を、ポリエチレングリコール沈殿処理により精製及び濃縮した。濃縮した各遺伝子のプロモーター配列溶液は、SPECTRA max Plus 384 を用

い吸光度を測定し、測定結果をもとに精製水で 45 ng/μL に濃度調整した。

調整した DNA 断片は、遺伝子のプロモーター配列と同様 (2.3.3、2.3.4、2.3.6) に Gateway Cloning System を利用して pGL3 Basic ベクターへ挿入、クローニングし、転写活性化能の測定を行った。

2.4.4.2 各塩基の変異が転写活性化能へ与える影響の確認

各塩基に変異を挿入した点変異プロモーター配列を作成し、ルシフェラーゼアッセイによる転写活性化能の測定を実施した。点変異プロモーター配列は、inverse PCR による点変異導入法を用い、作成した。まず、野生型プロモーター配列を pGL3 Basic ベクターに挿入したプラスミド DNA をテンプレートとし、変異導入用プライマー及び各プロモーターの共通プライマー、KOD-plus PCR kit を用い、DNA 断片を増幅した。PCR 条件は、94°C 15 秒、55°C 15 秒、68°C 6.5 分を 5 サイクルとした。なお、利用したプライマーは表 3.22 に示した。PCR 後のサンプルより不要なプラスミド DNA を分解するため、2μL DpnI を 1 時間処理した。

増幅した DNA 断片をセルフライゲーションによる環状 DNA とした。まず、1/10 希釈した PCR サンプル 2 μL、Ligation High (TOYOBO) 5μL、T4 Polynucleotide Kinase (TOYOBO) 1 μL、精製水 7 μL を用い、増幅した DNA 断片をセルフライゲーション反応を行った。反応は 16°C で 1 時間とした。セルフライゲーション反応後のサンプルは、Plasmid-Safe ATP-Dependent DNase を用い、セルフライゲーションしなかった直鎖 DNA を分解した。反応は、25mM dATP 1μL、Plasmid-Safe ATP-Dependent DNase 1μL で、37°C 1 時間とした。反応後のサンプルより、AMPure XP Kit (BECKMAN COULTER) を用い、環状 DNA とした点変異プロモーター配列を精製した。

精製した点変異プロモーター配列は、遺伝子のプロモーター配列と同様 (2.3.3、2.3.4、2.3.6) に、クローニングし、転写活性化能の測定を行った。

3 結果及び考察

3.1 遺伝子のプロモーター配列間の差が転写活性化能へ与える影響の解析

3.1.1 遺伝子のプロモーター配列の転写活性化能の測定

今回解析対象に選定したヒト胎児腎臓由来細胞株 HEK293 細胞においては、完全長 cDNA データの解析結果から、2170 遺伝子についてその発現と cDNA の構造が記載されていた。これらの遺伝子よりランダムに選んだ 852 遺伝子を対象としてヒトゲノム配列からプロモーター領域の配列情報を取得した。それぞれについて PCR プライマーを作成し、健常人ゲノムより該当領域を増幅し、レポーターアッセイプラスミドへとクローニングし、転写活性化能の測定を行った。その結果、472 遺伝子のプロモーター領域について転写活性化能を測定することができた(図 3.1)。472 遺伝子のプロモーター領域のクローニングおよび HEK293 細胞を用いたルシフェラーゼレポータージーンアッセイは統一的な実験条件下で行なった。各遺伝子のプロモーター領域に対する転写活性化能の測定は 3 回ずつ行った。全転写活性化能測定データのうち 86% で、測定データに対する標準誤差率が 25% 以下であった(図 3.1)。転写活性化能の測定された 472 遺伝子のプロモーター領域に対する詳細を(表 3.1)として示す。

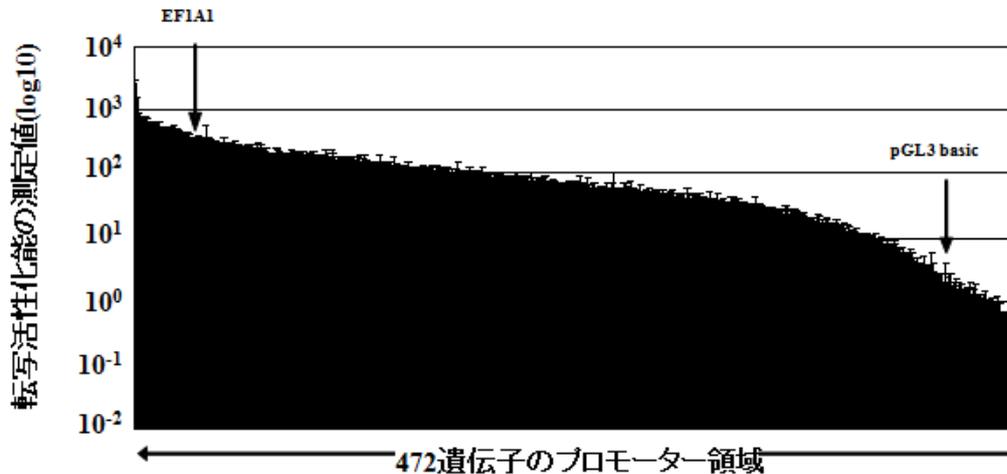


図 3.1 : 遺伝子のプロモーター配列の転写活性化能

転写活性化能の測定を行なうことができた 472 遺伝子のプロモーター領域が有する転写活性化能の測定値を、転写活性化能の強い順に左から並べた。コントロールとして、強い転写活性化能を有することが知られる EEF1A1 および空ベクターである pGL3 basic vector についてそれぞれ矢印で示す。

表 3.1 : 転写活性化能を測定した 472 遺伝子のプロモーター領域の詳細

PPR ID	NM ID	Gene Name	primer (fw)	primer (rev)	Amplified position (-xx to +yy; TSS=0)	転写活性化能	標準偏差
PPR_0001	NM_001673	ASNS	TGCTCTCCAAGCATTGACTG	ATCTCCACCCTTGGCTTC	-931,+185	2750.57	277.72
PPR_0002	NM_001016	RPS12	ACAACGGCTCGTTAAAAACC	GGGTTACGGGTGAAGTTGAA	-936,+189	1540.54	104.38
PPR_0003	NM_001967	EIF4A2	GCAGTTTACGCAGCAAGGAT	CCGTGCCATCCACTATCATC	-980,+148	800.72	75.11
PPR_0004	NM_003016	SRSF2	ATGGTGAGTGGAGAAGGCAG	TCATAGCTCTGAGTGGCGG	-928,+179	796.37	10.31
PPR_0005	NM_006466	POLR3F	GGGGTGAACCTAATGCCTTG	CTGCACCTTCACCTTCACCT	-965,+134	768.34	40.62
PPR_0006	NM_016299	HSPA14	TTTGCAAAGCAATTTTCACC	CCAGGTGAACCTCCGATGG	-976,+118	735.84	32.89
PPR_0007	NM_006276	SRSF7	AGCAAGGCTGCTATTTCCAC	CTCGAGTGCCTACTACACC	-905,+188	690.52	25.96
PPR_0008	NM_005057	RBBP5	TGGGCAACATGAAAATGCT	CTCAGCTCCGGCAACAAC	-952,+93	640.83	27.94
PPR_0009	NM_001030	RPS27	GTTTGGTGAGCCGAGATG	CCCTCTACGATCCGAAATCA	-932,+126	631.19	41.83
PPR_0010	NM_000973	RPL8	GCTCGACAGTCTCACTCAG	GCCCTCAGGGAAGAGGATG	-954,+86	594.75	69.94
PPR_0011	NM_003753	EIF3D	GGCGAGAGGACCACCTTGA	CCGCAGGAACCTATGAAACA	-953,+150	566.61	43.95
PPR_0012	NM_004396	DDX5	AAGACCCTATCCAGGAAACC	TAGAAAAGCGTCCGACAAGT	-976,+135	565.23	24.56
PPR_0013	NM_007208	ALAS1	TCCAAATTTCTACAACAGTAATCTGA	CGCTCTGCTCCGAGGAGT	-937,+119	553.97	11.88
PPR_0014	NM_001254	CDC6	AGAATTGGGTTCTGGCCTCT	CCTCCTCGAGCAATCCTCTT	-974,+132	553.58	112.97
PPR_0015	NM_016368	ISYNA1	GTCCACTTCCAACCTCCCTCA	GTCAGCGGGGACTCTAAGC	-899,+164	541.69	10.31
PPR_0016	NM_001895	CSNK2A1	TAAAGGAGGAGGCAAGGAT	CCTAGCTGCTGGGTGGA	-940,+111	529.44	27.25
PPR_0017	NM_007043	KRR1	TCGTGGGTTCTCTGTCCAT	AACATCACCTTGGTCTCCG	-969,+134	526.55	9.55
PPR_0018	NM_002799	PSMB7	GGTGGTTTATGCAGGCATTT	CTCTTCTCCCTCCCTGAACC	-957,+113	513.62	41.46
PPR_0019	NM_002106	H2AFZ	TTGCCGAACATTTGTCATGT	GCCCTGAGGACTCGAGGAGT	-990,+124	506.57	32.95
PPR_0020	NM_006425	SLU7	CACTTTCACGCTAGTCTCTGACA	CCTCAACACCTGAGGATCA	-924,+178	498.34	40.54
PPR_0021	NM_002816	PSMD12	GACGGAGTCTTGCTCTGTCCG	TCCTCACCCTTGGCTAGCTTC	-999,+164	487.21	64.43
PPR_0022	NM_000688	ALAS1	AAGGCCCTTTCTCAAGAA	GCCCTGAGGACTCGAGGAGT	-958,+120	477.39	85.34
PPR_0023	NM_005324	H3F3B	GGCAAAGTAGGATTAGCA	TACCCAACGCCGAAGTTTA	-966,+119	466.74	54.84
PPR_0024	NM_024662	NAT10	TTCAGTCTCTTGTCCATCC	ATCCAGCCAGTGGGAAAG	-984,+119	437.54	67.55
PPR_0025	NM_001316	CSE1L	CTGGGCACCCAGTTCAAT	CTCACCTCGTGGGATACAG	-937,+111	421.31	53.19
PPR_0026	NM_006275	SRSF6	TCTCTTTCTGCGCTTAAA	TGTAGCTCAGGCGCTATG	-933,+186	400.84	27.15
PPR_0027	NM_006362	NXF1	GGAGACACCTGCCAGAAAT	TTACCGCTGACGACTTCCC	-979,+130	392.96	28.50
PPR_0028	NM_005973	PRCC	TAAATGGGCTGGACAAGGAT	GGCAGATGAGGCTACTTGA	-993,+197	390.05	15.31
PPR_0029	NM_003404	YWHAB	CTGGGCGACAGAAAGAGACT	CGAGATGACCTTCTCCTCCT	-984,+139	381.84	62.64
PPR_0030	NM_016407	RTFDC1	GTCTCACCATGTTGTCCAGG	TTTCTGGGGTCTGCAAAATC	-954,+188	363.96	45.53
PPR_0031	NM_001417	EIF4B	GCCACTCTGTCTCCACAAA	GCAGCCCTCAATTCAGAAAG	-944,+162	352.50	11.05
PPR_0032	NM_002815	PSMD11	CCTCCTTTTATGGGACAGCA	AGTGGAGGATGTCGATGGAG	-935,+124	350.14	32.76
PPR_0033	NM_007358	MTF2	GGAGTGGAAAGGTTATGACCA	AGGGTGGGGGTAGAATGAG	-951,+133	343.39	33.55
PPR_0034	NM_003435	ZNF134	TTCAGACACTGTCGGTGAATG	CAGAGCCAGACGCGACTTC	-970,+73	342.00	7.74
PPR_0035	NM_005002	NDUF9	TGGCTCTTTGTAGAACAATGC	AACCTTAAATCCCCTGCTCG	-938,+125	341.57	60.95
PPR_0036	NM_021058	HIST1H2BJ	ATCTGGACCACCTTCCCTCT	TGGACCTGCTTCAGAACCTT	-925,+157	341.14	29.37
PPR_0037	NM_003145	SSR2	TGCCCTAACTTCAAGTTTCATCC	CTCTCCGCTCTCCCGTTC	-957,+93	334.54	47.62
PPR_0038	NM_001402	EEF1A1	AAAAGTACCATGAGGCACAC	GGGATCAAGAATCAGCTACT	-945,+142	330.83	7.20
PPR_0039	NM_006306	SMC1A	TGGTTTACCCTGCCATTACA	GGCCAGGTTTATCTCACCAG	-982,+190	328.67	237.11
PPR_0040	NM_032299	CUN1D5	AAATTGCTTTCTGGTGTCTCA	CAGCCAGTCCAGAGACAG	-939,+176	320.89	28.92
PPR_0041	NM_000985	RPL17	ACCCGGAGGAGAAGACAAAG	GCCTCCAGCGGAGGATTAGT	-939,+189	315.99	45.78
PPR_0042	NM_005850	SF3B4	TTTTCTATGGGAATGAGGCG	GGAAGGGAGGAGTCCAGTCT	-956,+177	315.01	33.90
PPR_0043	NM_000969	NPL5	CTTCTGATGCTTTTGCCTC	GGGAACCTCCATGCATCTAA	-974,+124	314.69	18.29
PPR_0044	NM_006294	UQCRLB	TGAGCCTCAGCTTTCTCCTC	TGCAGCAAAAATAAACGGTG	-960,+129	310.66	11.01
PPR_0045	NM_016422	RNF141	GGATTGCTGTGAAGATTAACG	TTTCTCTCCAGACCCAGA	-911,+150	306.40	25.35
PPR_0046	NM_016485	VTA1	GCTGGAGATTTTGGCTCTTG	CTGAGCCGTCCCTCAGATGAT	-943,+125	305.67	10.96
PPR_0047	NM_024814	CBLL1	TTTTAACCTCTCCTTCCGGC	TAATCCCTTTTGCCTGTGT	-943,+121	283.55	33.02
PPR_0048	NM_006782	ZFP1	CCCACAGGCAAACTGTTTT	CTGCACTGGTCCGGAAGG	-961,+150	282.66	17.35
PPR_0049	NM_002592	PCNA	GTTGCAATGAGCTGAGATCG	GCCGGAAAGGAGGAAAGTCTA	-915,+140	278.94	89.03
PPR_0050	NM_006328	RBM14	CCTCAGCCTCCCAAGTAGC	GACGTTGCCACGAATATCT	-924,+116	277.83	7.12
PPR_0051	NM_006773	DDX18	GCTGAAGGGTCTTGGGTAT	AACGCATCTCACCTGAAAC	-927,+152	275.68	35.57
PPR_0052	NM_004537	NPL1L1	TGAGAATGAAAAGGCAGTC	CCGTTACCGGCACTAGTAT	-922,+153	273.63	45.26
PPR_0053	NM_006392	NOP56	AAAATCCCCTTCGAATGACTC	AACGCAACCTCAGCGTCT	-978,+181	270.58	32.86
PPR_0054	NM_014412	CACYBP	TCCATTACCTGGGCAAAATC	CAGCCTGGTTACTTTCAAAACG	-922,+126	270.48	10.16
PPR_0055	NM_003457	ZNF207	TCCCTTACGGAGATACCC	TGAGCCCTCAACTGTGTGTT	-892,+171	269.68	50.29
PPR_0056	NM_006830	UQCRL1	TCCATTCTTATCAGGAGGC	GCACAATAACGGAACGCGAG	-964,+165	261.10	23.89
PPR_0057	NM_015950	MRPL2	TCTTCAAAGACGCCCAATC	CTTCCCGGGTAAGATGGATT	-966,+164	255.31	28.95
PPR_0058	NM_005085	NUP214	GCCCGGCTTTTATTTACTT	CATCCATCTCGCTCCCATC	-964,+136	250.70	9.15
PPR_0059	NM_018252	TMEM206	TGAGTGGAGGGTTACCAAG	TCCCTTCCGCTCTCACAAC	-950,+183	249.38	25.29
PPR_0060	NM_024619	ZNF419	AAGCTCCAAGAAGGAGGGAG	GGAGACAATGACCGAGGAAA	-969,+195	248.62	21.14
PPR_0061	NM_030969	TMEM14B	ATCTGGAAGTAGGTGCTGC	CGGCCGTGAAACTCCATAAT	-917,+149	243.91	32.70
PPR_0062	NM_004741	NOLC1	CACTATGCCTGGCCAATTTT	CGAACTTATTGGCCACCTCT	-958,+143	242.26	42.81
PPR_0063	NM_001697	ATP5O	CAGGGATTCTTTGCTTTTGC	TGGTCTCGAGGTTTGGTA	-955,+141	240.25	12.64
PPR_0064	NM_004343	CALR	AGCCGTTTAAATTGAAAAGC	TACCTGCTCACAGAAAGCTGC	-940,+165	239.35	39.38
PPR_0065	NM_002635	SLC25A3	GGCCATGTGTACTGTACCA	CTTTTCTTCCCGTCTCCAC	-939,+180	238.97	15.76
PPR_0066	NM_006409	ARPC1A	CAGCCTAAGCAACACAGCA	GACAAGAGAGGCGACAGG	-957,+136	238.81	6.46
PPR_0067	NM_000269	NME1	GGTCTTTTGGTGTCTGTAA	AACTCCCTCACCTAACCCACT	-982,+137	235.91	71.42
PPR_0068	NM_007158	CSDE1	TCCAGTGGTTACACATACAACA	TCCCTGGGTTTCTGCTAC	-937,+122	234.77	12.79
PPR_0069	NM_003091	SNRPB	GGCAAAACCCGCTCTACT	CAGCTGTGCCCTCCTTAC	-929,+112	233.98	45.25
PPR_0070	NM_024513	FYCO1	CCCCAAACTGACTACTCCA	TAAGGGCAACAGGTGACAAG	-983,+198	230.32	26.92
PPR_0071	NM_016504	MRPL27	CAAAAAGAATTCCTGCAGCC	CGTCCCTCAAACCTGGAGTA	-940,+196	229.08	35.78
PPR_0072	NM_000967	RPL3	AATTCGAGGAAACACGAG	GCCTAACGGGTCCGCTAT	-923,+137	223.09	37.80
PPR_0073	NM_014633	CTR9	CATGTTGAAATGAGATTTTGGGA	GTCAAAAGGCGCAAGTGT	-937,+130	213.90	5.29
PPR_0074	NM_001418	EIF4G2	TAGTTCAACTTCCCGGCTTT	CCTCAAACCTCAGCTCAGAGGA	-920,+146	213.68	5.08
PPR_0075	NM_006096	NDRG1	GCTCCTCGGACTTCTCAT	TACCTAACGCGAGGAGAAA	-920,+121	211.75	6.81

表 3.1 (続き) : 転写活性化能を測定した 472 遺伝子のプロモーター領域の詳細

PPR ID	NM ID	Gene Name	primer (fw)	primer (rev)	Amplified position (-xx to +yy; TSS=0)	転写活性化能	標準偏差
PPR_0076	NM_007192	SUPT16H	CTGGCTAAAGGCAAGGGAAT	TACAGGATCTTCTCACCCG	-997,+178	208.98	9.73
PPR_0077	NM_001320	C5NK2B	AAAGCTCGGGTTTCAGTTTT	CAGGACGACTTCAGACTCA	-978,+136	208.16	42.08
PPR_0078	NM_004539	NARS	ACAACGGACTCGGGGAAG	CGTTGCATCAGAGAGCGTAG	-932,+126	207.03	26.85
PPR_0079	NM_002482	NASP	TTGAACTCCTGACCTCGTGA	ACTCCTGAGCCATCGTTC	-921,+119	205.39	22.88
PPR_0080	NM_006170	NOP2	GAGGAGGTAAGTGGCATGGA	GGGTTAATGTCGAGAGTGC	-963,+90	205.29	10.00
PPR_0081	NM_018077	RBM28	GTGAAACCCCTGTCTCGCAA	GGCCACAAATAGGTCAAG	-922,+123	204.44	12.49
PPR_0082	NM_016070	MRPS23	GCTGCATAGGGTTCAAAAGA	CACCTGCATAGCCGCTC	-939,+190	203.49	11.46
PPR_0083	NM_005911	MAT2A	GCACATTCAAGGGAACGAAT	TGTCGGTTTACGTTGGTGC	-986,+91	202.29	2.75
PPR_0084	NM_004960	FUS	GCCTGTAAATCAGGCACAAT	GGTCCCACTGAAACGAAAA	-977,+152	201.59	11.76
PPR_0085	NM_005570	CFI1	TATGCTCTGCAAGGTACCC	TTCCGGAAACGGAAAGGAG	-988,+139	199.58	40.75
PPR_0086	NM_020239	CDC42SE1	TCAATTAAGGCGTGTCCCA	CTCACCTGCTTCTTCCACC	-892,+175	199.09	11.54
PPR_0087	NM_017816	LYAR	CCCAACACTGCTCACTTTCA	CGTCCCTCCCTACTCACTC	-976,+184	195.96	24.86
PPR_0088	NM_007355	HSP90AB1	CCCGCAGGTAAGTCCACT	CAGCACTAAGGGAAAGGAGC	-918,+141	195.85	4.55
PPR_0089	NM_007202	AKAP10	TGTCTCTTCTGGGCTGTTT	CTCCCTCATTAGCAACC	-926,+162	194.80	21.53
PPR_0090	NM_003286	TOP1	ATGCAAGTAGTGGCTGTT	GTAAGCGGGGCTGACTGT	-952,+107	194.73	12.63
PPR_0091	NM_006739	MCM5	AGATTGAGCCACTGCATCTC	TCCAAGTACCCCGGAATC	-953,+131	188.80	41.12
PPR_0092	NM_004661	CDC23	AGTGCAATGGCAGATCTC	CCGCAATAAGCAGCAGTTG	-961,+131	187.21	41.68
PPR_0093	NM_005687	FARSB	CTTTGTTTCTGGTGGCTGT	CTCCGAGAAGAGGCGTAGG	-935,+121	186.68	7.74
PPR_0094	NM_001791	CDC42	GAGCTCGAGATTGCAGTGAG	GAAGCTCAGGTTGGAGCTC	-924,+119	184.17	33.54
PPR_0095	NM_000983	RPL22	GCTTGAGGCTGAGTTGAAA	AGCGAGCCTGGTAGATG	-952,+122	183.61	14.94
PPR_0096	NM_006112	PIPE	ATAAAGCCTGCGTTTGGCTG	TCCTCTCTGGTTCAGAGATGC	-950,+141	183.26	2.76
PPR_0097	NM_017850	C1orf109	GCCTGATTGTTGGATTGG	CGCCGCTCTTCTCTGACAT	-922,+150	182.44	41.53
PPR_0098	NM_015946	PELO	CTGCCCTCCACTCCATCTAA	TTTCTAAGGCACTTGCTTG	-996,+156	179.24	1.78
PPR_0099	NM_000291	PGK1	TTGTGCATTTCGTCAGTTC	TACAGCTGGGAGAGAGGCTC	-979,+99	179.00	15.90
PPR_0100	NM_020414	DDX24	GGATGCTGTAGCCGGTTTTA	TGTTACTACCGTGGAGAGA	-970,+127	175.72	18.82
PPR_0101	NM_002129	HMGB2	ACACTCGGGCTTGGAAAT	GAACCCGAGGATTGGAG	-949,+143	174.71	18.27
PPR_0102	NM_021974	POLR2F	AGTCCCTGGCTGTAGGAA	GCTCCCGCACTCACTTGT	-958,+125	174.62	23.54
PPR_0103	NM_017858	TIPIN	CCGGCCCAAGTAACATA	CTGGACAATTAGCTTTCC	-985,+178	173.96	19.85
PPR_0104	NM_006397	RNASEH2A	CCTGCAATCCCAACTTTG	GACAGCGCCTGTATTGTCT	-935,+148	172.29	68.91
PPR_0105	NM_004597	SNRPD2	CCGCTCCAAATGGTGTAGA	CAACCTTCCCACACTCAAT	-918,+136	172.22	44.04
PPR_0106	NM_003366	UQCRC2	ACGAAGTCAATCCAGGATCT	AGCCGCTCTGGTTAGTAGC	-931,+169	169.56	21.30
PPR_0107	NM_003707	RUVBL1	CCACCTGGGTACCTGTCT	CTTTCACGTGGCTGTGGG	-946,+141	168.42	65.45
PPR_0108	NM_014142	NUDT5	CCATTCTAAGCCAGTTACACATTTT	TCCTCTCTCGACACTTCGGTA	-953,+121	164.64	47.32
PPR_0109	NM_004663	RAB11A	CGCTTTTCTGGAATGACAAA	GGGAGACAGCGTGGATCTG	-956,+130	164.15	21.00
PPR_0110	NM_002793	PSMB1	AGGTCATCTCTGTGTTTCCA	GCCGAATACATGGCTGTAGA	-975,+79	164.07	22.19
PPR_0111	NM_000687	AHCY	TCCACAGAGGCTCCTGAAACC	CACCGACTTTGTAGGGCAGT	-989,+97	163.88	21.37
PPR_0112	NM_015414	RPL36	CTGCCACCCGATTACAGTTAT	ATCCCTCAACCTGCACAGAG	-929,+134	162.82	15.32
PPR_0113	NM_032177	PHAX	GCCCTGGCATGCTTTTACTA	CCTTTCACACTCACTGGCAA	-969,+125	162.38	10.00
PPR_0114	NM_017953	ZNHIT6	AAACCCATCATGATCCCGAC	GCCCTGCTAAGTCCCTTAC	-970,+191	161.97	10.20
PPR_0115	NM_003756	EIF3H	TTTCATATACACTTGAATGAGTGA	CTTGCTTACCGCTGAATCT	-981,+131	161.45	13.40
PPR_0116	NM_012433	SF3B1	TCATCGTGAAGTCAAAGACC	CAGCTCCTACGAAAGAAACCA	-917,+171	157.55	9.05
PPR_0117	NM_017841	SDHAF2	TTGAAACCACTGCCAGTTTG	AAGCAAAAGGGTAAATCCCGC	-983,+156	157.01	12.71
PPR_0118	NM_003542	HIST1H4C	TCAAACCAATTCAGACCCA	GCGAATAGCCGGTTTTGTAA	-938,+124	153.33	24.47
PPR_0119	NM_017769	GZE3	TGCTGTATGGTTGGTTTG	CACGGTCTGCTTCTCTCTC	-922,+163	149.81	14.72
PPR_0120	NM_018471	ZC3H15	CATATAGGCTGAAGCCGGAC	GGGTGGTACTCACTTCGAT	-921,+187	147.60	17.81
PPR_0121	NM_004757	AIMP1	AGGGCCAGAAAACCAACCA	GAGAAAGAACCCCTACGA	-944,+120	147.21	35.81
PPR_0122	NM_004236	COPS2	AAGACACAGGATCTTTGGAA	ACTGAGAGAGGCTGTAACCG	-951,+194	146.84	50.15
PPR_0123	NM_014610	GANAB	GATGCAGCTGGTGCAGAG	GGAAAGAGTGGAAATGGACT	-980,+142	145.43	9.02
PPR_0124	NM_004147	DRG1	CATGGATACATTCGCTTTGAAAA	TGCTTACTCTCCGGGATCT	-957,+124	145.35	25.39
PPR_0125	NM_000947	PRIM2	TCATTCAGGACATTTGTAA	GTCTGTCTTCTGCTACTGC	-979,+186	144.96	48.31
PPR_0126	NM_006292	TSG101	GCCAACTTTGAGGGGAATCA	CCTCACCTTGGACACCATT	-976,+129	144.95	15.38
PPR_0127	NM_015176	FBXO28	AAAGTGATTCTCCCGTTCA	GAGGCTGTCCGGGTAG	-929,+117	143.15	4.87
PPR_0128	NM_018040	GPATC2	ACCATGTTGCTCGGTGGT	CACACCTTCCCTGACCTC	-933,+199	141.87	12.71
PPR_0129	NM_003757	EIF3I	CCTCTCCCTCACCTCTGCTA	CTGCAGTAGGATCGGCTTCT	-999,+164	141.24	6.56
PPR_0130	NM_005869	CWC27	AGGAACTGAGGCCAGAAAG	AGTCCAGTGGTGCAGGAAAG	-943,+116	139.97	11.48
PPR_0131	NM_016399	TRIAP1	TCACACCTGTAATCCAGCA	AGCTGTCCCTTGTAGAAAT	-962,+120	139.55	4.63
PPR_0132	NM_016067	MRPS18C	CGGGTCTTCAATAACAGCA	GTGAGTCCCGGATGTGTAA	-958,+116	139.14	5.83
PPR_0133	NM_007363	NONO	GCGAGTTAGGCATTACTGAA	CTGATTGTTCCAAGAAATCC	-920,+146	136.78	57.34
PPR_0134	NM_003191	TARS	AGCAGATTGCCGTGATCATA	GACGAAGCATGAGAGAACC	-938,+120	136.37	17.07
PPR_0135	NM_001219	CALU	TGCCACAGGTTTACATTTCA	TGGGCATCACCGTACTTACC	-982,+129	135.93	11.19
PPR_0136	NM_002515	ORC5	GATCCTTTTGACCATCACAAATC	AAAGCCACAGTTTTCCAAG	-947,+148	135.85	3.92
PPR_0137	NM_003564	TAGLN2	CAGCAGGTAAGGGTAGAGCA	CTTTCTACACAGCAGCACA	-945,+142	135.11	11.00
PPR_0138	NM_005319	HIST1H1C	CCTACAATTTTTCATGGGCTGA	CTTTTGGCCGCTTCTT	-950,+121	134.68	23.74
PPR_0139	NM_002716	PPP2R1B	TCCAAGCTGCTTAAATTTCA	TACCTGCAGTCTTCAATTGC	-948,+131	134.15	14.96
PPR_0140	NM_006447	USP16	ACCTCTGCCTCCAGACTTCA	ATAACTGCATCCCTCATGG	-989,+90	132.70	44.92
PPR_0141	NM_012087	GTF3C5	CTGGCCAACATGGTGAAC	CATGTCTTCGGGAAACGAC	-920,+128	131.62	3.47
PPR_0142	NM_006899	IDH3B	GCAGGAGAATGGGTTGAAC	GCTTTGAGACATCCAGACCC	-974,+168	130.40	12.57
PPR_0143	NM_017670	OTUB1	CTGCCCCCACTAACATCTTG	GGACATCCCTCTGGATCT	-940,+120	129.91	4.53
PPR_0144	NM_032353	VPS25	AGGTGAGAGATGGTGGTTG	TAAGGACTGAGGTTGGTTG	-979,+197	127.51	4.53
PPR_0145	NM_022770	GINS3	CCCAAGATGTCAAAGCATCA	CCGACTCCACTCGGAAATAA	-959,+165	127.16	8.38
PPR_0146	NM_006351	TIMM44	AAGCTGCCTCCCTTTGAAAT	CCATGTTGGAGAACTCGTGTG	-992,+66	126.12	10.40
PPR_0147	NM_018195	C11orf57	TTGTGGATCAGCACCACTTT	CGGTAACCTTCCCTCAAAA	-965,+118	125.37	9.06
PPR_0148	NM_005008	SNU13	TGAGAGGGGGTTTCTGCC	AGTGAACGGAACCTTGAGTC	-917,+161	124.35	28.82
PPR_0149	NM_018116	MSTO1	GTCGCTGGCTAACTTTTTG	GGGACTGTAGATCCCGAAGG	-951,+148	123.40	8.00
PPR_0150	NM_004461	FARSA	CATGGCCAAAACCCATACTT	CTGAAGGCTTTCACGGC	-989,+152	123.25	6.52

表 3.1 (続き) : 転写活性化能を測定した 472 遺伝子のプロモーター領域の詳細

PPR ID	NM ID	Gene Name	primer (fw)	primer (rev)	Amplified position (-xx to +yy; TSS=0)	転写活性化能	標準偏差
PPR_0151	NM_002804	PSMC3	GGGCTCACCTCAAAGGATAA	GCCTTCTCTTGACCAGTGGGA	-940,+121	121.20	9.16
PPR_0152	NM_001668	ARNT	TGTCCCTGTCAGTACCCAAA	CTTAGTTTGTGACGCCCTTTCG	-953,+178	120.62	8.98
PPR_0153	NM_031206	LAS1L	CAGCTGTACAGATGGGAAAACCTGA	ACTTTCCGTACCACGGCACTC	-934,+130	120.43	7.76
PPR_0154	NM_000521	HEXB	TGGTCTGCCTAGCGTTTTCT	CAGTGTGCGCAACAGCAG	-943,+182	119.98	3.04
PPR_0155	NM_004218	RAB11B	CCTATCACTACCTGACAATACAGCA	CCGCACCTTTGAATAGGTAGT	-985,+141	119.58	1.98
PPR_0156	NM_005175	ATP5G1	TTTCCCCGAGACGGGAAT	CCAAGTCACCTGCACCTCCTA	-918,+127	118.14	19.13
PPR_0157	NM_001628	AKR1B1	GCAGTTGGGGTGAGGTAGAG	CGTTGTTGAGCAGCAACGC	-975,+129	117.16	6.68
PPR_0158	NM_001688	ATP5F1	CAACACTCTCAGGCTCCTCC	TGTCACCCCTAGATTGCTT	-988,+169	116.57	9.78
PPR_0159	NM_006948	HSPA13	CAGTGACCAAGATGCGAGAA	ACAGCCGAAAACCGTCTCTA	-951,+195	115.66	9.66
PPR_0160	NM_001253	CDCS5L	GCCAGAGATGTTCTGGAAGG	CTCGGTATTCTCCATACGC	-925,+163	115.10	14.31
PPR_0161	NM_019071	ING3	TTTCCCAGAAAATCGAAAA	GCCCTTAGAGCTGAGGGAAC	-988,+131	115.02	16.78
PPR_0162	NM_006842	SF3B2	GGGAAAAGGGCCTGAGTTTA	CTGTGTTCCCTACCCCTGGAT	-928,+162	114.19	9.13
PPR_0163	NM_001007	RPS4X	GGCAGTTTCCGAAAGCAGAT	CCAGTAGGAATCCGCAAAACC	-921,+144	113.40	16.95
PPR_0164	NM_006345	SLC30A9	CCCCGAGATAGTTATTTTGCAG	CAGGGAGGACCAGCTACATC	-981,+172	111.21	23.01
PPR_0165	NM_013293	TRA2A	TCGACATTACACAGGGCAAA	GATGGCTAATTAACCCGCT	-991,+124	108.98	17.56
PPR_0166	NM_013234	E1F3K	TAAACCTCTGCTGGGTGCTT	GCTCAGACCTGCTGCATACCC	-905,+175	107.92	9.16
PPR_0167	NM_003262	SEC62	AGCTTGGGCAATATAGAGCG	CTTTGCTCGCCCTTTTAG	-967,+142	107.89	13.09
PPR_0168	NM_004550	NDUFS2	CCAGGCTCAGAGTTGGAGAG	TGAGAAACCCTAGCCTTGGGA	-983,+162	107.88	11.63
PPR_0169	NM_001814	CTSC	CAGGAGGCACCTGGAATA	CCAGTAGGAGTGGAGTTGGCA	-886,+197	107.00	16.34
PPR_0170	NM_017828	COMMD4	GGATGTGCACAGGTTATTTGCG	TCCCTACCAGACTCACAGCC	-919,+188	105.21	11.43
PPR_0171	NM_002337	LRPAP1	AGATGCCCGAGGCTCTATTTCA	CAGCTGGTCAACTTCTCCCA	-920,+197	105.09	0.41
PPR_0172	NM_025114	CEP290	GATGAGCAAACTTGACAGA	CTCTCACTACCGCAACACC	-954,+160	104.53	11.96
PPR_0173	NM_003518	HIST1H2BG	GGTGTCAACCTCGGTTTTC	GGGGTGAACCTGTTTTAGCA	-960,+198	104.50	3.30
PPR_0174	NM_014062	NOB1	CACAAAACCAAAACATCCCC	CAGCCAGATCTCTCTCGCT	-993,+114	103.47	15.87
PPR_0175	NM_020365	E1F2B3	CACACTACCATGCCAGCTA	CCGCTCTCTACCCCTTTCG	-953,+177	103.11	42.29
PPR_0176	NM_021141	XRCC5	CAGCTGGGGTTCATATGGTA	CCAAGTCCATGGCTTTCTTT	-926,+151	103.07	6.71
PPR_0177	NM_020191	MSPS22	CACATCCTGGATAGCTGGAAA	CTCGAAAGAGCAAGGTAGCG	-978,+147	101.56	9.18
PPR_0178	NM_014285	EXOSC2	GTGCATGGGGAGTTACCATC	GCATGAATCTGCTGCTCGTA	-919,+138	101.21	10.47
PPR_0179	NM_000581	GPX1	TATCCAGGTTCCCTCCCAGA	AGAAGGCATACACCAGCTGG	-956,+103	99.71	13.99
PPR_0180	NM_031465	RHNO1	CTTTCCAGGCTTCCCTCTCT	CGTCCCCAAACTCTCCCT	-943,+165	99.17	8.53
PPR_0181	NM_004853	STX8	GCTGGTAAAACCATCTCAC	GGACTCTCTCCTGCTTAGC	-964,+196	98.11	28.22
PPR_0182	NM_001641	APEX1	ATTGAGGCAACTGGAATCT	GAAGCCCTCTTATGAGCAA	-992,+122	97.22	13.98
PPR_0183	NM_006949	STXBP2	CGGGTCTGATTCACTCCCTA	GAAACCCAGGATCCAGAG	-950,+167	97.13	8.96
PPR_0184	NM_001724	BPGM	GCTTGCCTGTCACAAACCTA	AAAGCCACTCAGCTGCAAAAG	-930,+152	96.76	9.90
PPR_0185	NM_004792	PPIG	GGACCCTTGCCCTTAAAAT	CACCTACCCCGCAACAGT	-970,+117	96.39	13.03
PPR_0186	NM_002634	PHB	CAGTATTCTGGGAGGGTGA	AAGTGACATCATGGAGCAG	-974,+114	96.11	8.94
PPR_0187	NM_005445	SMC3	CCTGGTTAGCAGAGGCATTT	CGAAGCCCTTACCTGCTTTA	-997,+109	93.49	19.27
PPR_0188	NM_018120	ARMC1	TAACCGAGAACAAAGGGAACG	CCACACCTACGGATGAGGTAG	-996,+167	93.48	31.35
PPR_0189	NM_015640	SERBP1	TCAATATTTCCCTCCGCTCA	CCGAAGCCTTCCCTGTAAGTG	-939,+121	93.42	4.42
PPR_0190	NM_003288	TPD52L2	GCCCTGAACAGGAAGTGAGA	GGGAGACGGAGAACAACG	-989,+159	92.37	4.34
PPR_0191	NM_020806	GPHN	AACGTAGAATAGCCACAGGTA	CAAGTCCGGCAAGAAGGAG	-943,+136	91.93	1.91
PPR_0192	NM_015235	CSTF2T	GGGGAAACACACACTGAAT	CAGAACAACCTCCGAGAAA	-991,+165	91.79	7.37
PPR_0193	NM_001363	DKC1	CACCTTTACAGGCTCACCTCA	AGGGAACACCCGACAGCT	-926,+159	89.43	18.36
PPR_0194	NM_018126	TMEM33	CCTCTGATTGGTTGCTTTTTG	CACCGCCGAGATCAAAGC	-945,+106	88.12	12.92
PPR_0195	NM_001239	CCNH	ACCTGTGTGGGCTTTTCAAG	GACCAGGTCAGAGGGTCT	-988,+107	87.98	2.04
PPR_0196	NM_015599	PGM3	CCCGCACTAGATAGGGGAAT	AGAATCTGACCTGCCCCTG	-924,+157	87.51	10.79
PPR_0197	NM_030759	NRF2F2	GTGCGGTAGCTCATGCCTAT	TTCTGTGGGACACTCACCG	-992,+183	87.30	4.62
PPR_0198	NM_001878	CRABP2	GTGACACCTGAGACCCTGCT	CGATCGGATGATTTCCAGT	-935,+176	87.13	11.71
PPR_0199	NM_004819	SYMFK	GGTGAGAACTCTGGAGCCCTT	GAGCCTCTCCTTTCTGCTCT	-922,+158	86.70	18.15
PPR_0200	NM_004330	BNIP2	AAAATAAACCCGACCGTTCCC	CAGTCTCGAAAAGGAGG	-958,+120	86.44	7.48
PPR_0201	NM_016004	IFT52	AATTGTGAAAGCAACACGA	CCCTGGATGTTCTATGACCG	-976,+197	86.19	3.30
PPR_0202	NM_032335	PHF6	GCCTAGGCATTATTCTGTGTA	ATCCCTCTGCTCTCCCGC	-970,+192	85.26	5.53
PPR_0203	NM_016391	NOP16	GAGGGAGTAAGTCCCACGGT	AGAGCCTTTCGGTTGACACT	-984,+125	84.49	3.60
PPR_0204	NM_000249	MLH1	GGCTTTAAAGTCCCTGGCTC	CCCTCCGTACCAGTTCTCAA	-958,+159	83.98	5.82
PPR_0205	NM_003157	NEK4	GCACACCTGTAATAACGGCA	GGCAACAAGAAGCTCGGTT	-974,+127	83.82	7.60
PPR_0206	NM_000414	HSD17B4	TGTGTTCCACCACTCTCTG	AACCTTCGGATGCTCACCT	-926,+133	80.91	20.07
PPR_0207	NM_004134	HSPA9	GTTACGCGCTTCTCCTG	ATCATGGCGGATAAATGGAG	-987,+72	80.43	19.73
PPR_0208	NM_024813	RPAP2	TGGAAAGTTTTTGTGTTGTTTT	CTGAAGCAGCCTTTCAGGAG	-927,+187	79.53	7.01
PPR_0209	NM_014676	PUM1	TTGCGATTGTCTGTTTTCAA	ATGAAGGACAATCTGCTCG	-938,+153	79.18	15.05
PPR_0210	NM_025103	IFT74	GCAAGCAATAACTCATGTCCC	ATTTGTTTTGTTGGAGCGG	-977,+196	79.09	10.65
PPR_0211	NM_014325	CORO1C	AGAGCTGAGGCTCAAATTC	CACATGCTCAGAGGACAGGT	-937,+160	78.20	3.08
PPR_0212	NM_006743	RBM3	GGGTTCAAGCGATTTCCGAG	TTTCCGACGACATATGCAAC	-922,+129	77.80	6.48
PPR_0213	NM_032120	RBM48	ACTACAGCTGTGTTCCGCGT	AGGAGAGCGAATGAAACCAA	-996,+164	77.75	15.71
PPR_0214	NM_002219	STT3A	GATGATGGAATTGCTGCTCA	CTGCCGCTGATTTTCAAAT	-923,+127	77.53	23.29
PPR_0215	NM_000574	CD55	AGCTCCCACGCTGATTCTAA	ACAACAGCACCAAGCAGCA	-978,+153	77.18	5.99
PPR_0216	NM_004328	BCS1L	ACGCTGAGCCGTCAGATTA	GAACCCCTCCTGAACCTC	-922,+134	76.25	8.76
PPR_0217	NM_021004	DHRS4	CACCTGGCAGAAGTGCAGAC	TACGAGGCCACCTTATTTG	-972,+121	75.68	11.91
PPR_0218	NM_004440	EPHA7	CTTGCGAAGAGAAGGTCTTG	TTATTGTGCTCCTTGCACTG	-983,+125	75.26	2.35
PPR_0219	NM_005789	PSME3	TTGCCTTATTGAGATGTCAG	TGAGAAATCTCGGCTGACCT	-950,+135	74.08	7.61
PPR_0220	NM_018491	CBWD1	CGCGTGTAGCCACACAATAG	GACTTTTCTCCTCCTCGCT	-955,+192	73.94	8.63
PPR_0221	NM_005190	CCNC	CAGACCAGCCTAGCGAAATC	CACAGCTTCCCTGATAAAA	-938,+167	73.27	11.64
PPR_0222	NM_004551	NDUFS3	GAGAGCAAAAGTGGTCCAGA	CTGAGGAGCTCTGCACTG	-937,+167	73.21	18.18
PPR_0223	NM_006601	PTGES3	GAGTAGGTTCCAGGTTTCTG	GTCGGGAGAAAGGAAAAGT	-932,+134	71.86	4.35
PPR_0224	NM_000895	LTA4H	CCCCCAGCATCTTCTGATT	ATCACACAGCACAGCGACCT	-980,+69	71.86	3.69
PPR_0225	NM_019067	GNL3L	GCTCCTTGAGATGGCTTCAG	GGAAGGGATCCTGCTGTTTA	-922,+131	71.74	3.02

表 3.1 (続き) : 転写活性化能を測定した 472 遺伝子のプロモーター領域の詳細

PPR ID	NM ID	Gene Name	primer (fw)	primer (rev)	Amplified position (-xx to +yy; TSS=0)	転写活性化能	標準偏差
PPR.0226	NM.006429	CCT7	TCTGGCCTTAGCTACTTC	GTCTACAGAGCAAGCCAG	-952,+166	71.47	8.74
PPR.0227	NM.006467	POLR3G	CCAGGTCACGGTCTCAAAGT	CCAAGTCTTCGATTGCTTCC	-956,+156	71.44	3.65
PPR.0228	NM.022163	MRPL46	AGGCAGGGAAGGTTATTGGT	GCTTGAGACTACAGGTGGC	-960,+195	71.14	2.77
PPR.0229	NM.004541	NDUFA1	GCCACCCATGCTTGAATCT	CTGGAATCAACAAGCACACG	-927,+119	70.18	4.02
PPR.0230	NM.005762	TRIM28	CCGACTCGCCTAAGAAAGTG	AGTGCTCCAGCAGCTCCA	-952,+196	69.38	5.48
PPR.0231	NM.003090	SNRPA1	ATGCAAACTGCATCAGAAC	CAAGAGCGGAAAAAAGG	-999,+184	68.34	5.04
PPR.0232	NM.032178	SLC7A6OS	TGCCCTCCAAGAGCTTAGACT	GCCACCAAGTGAAGACATT	-995,+194	67.81	5.76
PPR.0233	NM.005870	SAP18	TTGCCAATCTAGTGAAGTGGG	ATCGGTTTCTGCGCTCCTT	-927,+129	67.53	4.59
PPR.0234	NM.005470	ABII	CCCAGTTACTCTGGAGGCTG	CCCAGTCAAGTCTGGTAA	-912,+191	66.80	3.57
PPR.0235	NM.012341	SNRBP4	CCTCTGCTGCCTGTTAGCTT	CTGAGGAGACGCGAAAGTT	-954,+140	66.75	10.31
PPR.0236	NM.031480	RIOK1	CGCAAAGGCCTTGTTCCTTA	CCGGTAGTCCATGACTGGAG	-939,+196	66.34	5.46
PPR.0237	NM.002358	MAD2L1	TGGTAATTCGTTATGCAGCAA	ATGGCCAGGACACACAAA	-933,+118	64.97	1.88
PPR.0238	NM.004893	H2AFY	GGGTGGGGTCTCTAATTACC	CTTTTCTCTCCGGCTCCT	-924,+133	64.25	9.79
PPR.0239	NM.004275	MED20	TGGATACATGCTAGGAGCACTT	CCCTGCAGTCTACCAAGTC	-955,+133	63.58	12.78
PPR.0240	NM.032390	NIFK	TGCCCTCAGGCTGAAGTTC	GCCTGCTCACCTGGGTTAT	-972,+145	63.17	2.30
PPR.0241	NM.016185	HNI	ATGCAAGACATGAGAAGGGC	GCCGATTCCTGCTGTTGGG	-935,+137	62.81	27.85
PPR.0242	NM.014186	COMMD9	TTCACAAATGGGCATTCTTTC	CCAATTCCTGCTCCGTCTAA	-950,+138	62.67	11.72
PPR.0243	NM.018403	DCP1A	AAGAGCGAGACTCCGTTTCA	GGGAGAGAGGTACAGAGC	-995,+136	62.15	7.95
PPR.0244	NM.003093	SNRBP	GGTGGATCACGACTCACTA	CCCAGTACTCTCCCATCCC	-935,+153	61.67	4.21
PPR.0245	NM.004371	COPA	AAACAATGTGAATTGGGGGA	TCTCCACTACCTTTGACCCG	-915,+153	61.58	21.67
PPR.0246	NM.025132	WDR19	TGCCCTGAATCTTTCTTTG	TCAGAAGGAGAGGGACAGA	-988,+132	60.65	6.14
PPR.0247	NM.002811	PSMD7	TGTTTTTCCACCACCACAAA	ACCACCACCTTCTGGACTG	-932,+132	60.26	7.86
PPR.0248	NM.002109	HARS	GCTGTATGCCCTTCCAAAAA	CTAAATCTCACAGCTCGGC	-955,+164	59.58	14.30
PPR.0249	NM.002388	MCM3	AAGAAAAAGCCAACCCAGCAG	CGTCCAGGAAAGTCCAGGTA	-971,+138	59.15	14.67
PPR.0250	NM.005610	RBP4	CAGGTGGCTTTTTCTTCCG	ACCCTACCTTCCCTTTCG	-971,+109	59.12	2.16
PPR.0251	NM.006594	AP4B1	CAGCTCTCGGAGACCGACT	TAAGTCCGAGGCTCCTTCTC	-874,+174	58.79	3.32
PPR.0252	NM.014828	TOX4	AGCAGTTATGCTTCTGGTGG	CTCTCTCTCCTCTCCCGTCA	-987,+171	58.40	4.79
PPR.0253	NM.019040	ELP4	CCACTGTCCCTGCTCAGTTT	CTCTAGGACCCCTCCTCTGG	-968,+133	58.39	3.22
PPR.0254	NM.024908	WDR76	GGACAATCTCAGGCAAGGA	CCGCATAACCCCTCACAAAAA	-933,+157	58.27	3.28
PPR.0255	NM.012333	MYCBP	TGGGAAAAACACATGTTCCAGA	GACCAGCGGCTTGGCTACC	-925,+136	58.04	10.44
PPR.0256	NM.005906	MAK	TGGTTTTCCACGCGTGTATT	AGATTGGCCCTGGTAGCAC	-980,+187	57.72	2.13
PPR.0257	NM.006083	IK	GCCCAGCTCTATGTGCAAAA	GGATGGCCTGAGCCTTACTA	-922,+131	57.44	3.64
PPR.0258	NM.016311	ATPIF1	TACTGCATTGCACCACATTG	GGCTTGGATTGGTCTCTCAC	-994,+113	56.78	14.19
PPR.0259	NM.032796	SYAP1	TCCCTGACCTTGTGATCCTCC	CAGACTCAGCCACCCGCTCTC	-953,+196	56.35	47.31
PPR.0260	NM.003341	UBE2E1	GCTAAAGAAAGAGAAAAATGGCTTA	TCCCGTACCTCTCTTTGTG	-957,+144	56.25	4.82
PPR.0261	NM.032364	DNAJC14	GGGAGCAAGGAAATAGGAG	GCAGGGAGAGACGTAGGAG	-997,+157	55.24	5.10
PPR.0262	NM.003932	ST13	AGCAAGCTCTCACGCCTAAC	ACACATTTTTCACAAAAGGCC	-966,+144	55.17	4.91
PPR.0263	NM.031280	MRPS15	GTTAGGCTCCTACCCCTCCCA	GTCGGAATCAAACTCAGCGT	-978,+95	54.88	4.02
PPR.0264	NM.018492	PBK	GGGGGGGAGAGGTA AAAA	GGAAAGGTGGGAGGTGAAA	-908,+132	54.87	15.25
PPR.0265	NM.003086	SNAPC4	ATTTGTGGCGCGGTATTTAG	CCTCAGTTTCCCGCAGTTT	-951,+179	54.26	3.57
PPR.0266	NM.005370	RAB8A	GATAGGACCAGGATCCTCA	GCCCTTATTGGCTCCAG	-925,+168	54.25	16.64
PPR.0267	NM.006158	NEFL	TGGTGGTAGCAAGCAGGAAT	GTAGGGCTCGTAGCTGAAGG	-942,+128	53.67	8.70
PPR.0268	NM.030579	CYB5B	TTCTACCCCAATTGCTGTC	GACATCGTAGACTCGCCCAT	-974,+194	52.84	22.13
PPR.0269	NM.000386	BLMH	CACACACACTCAACCTCATCG	GCGGGTAAACGGTAGCTT	-976,+155	52.65	4.26
PPR.0270	NM.004786	TXNL1	CTCTTCCCGACAAAATCCAA	AGTCTTCTCCCGGACTCTC	-930,+153	51.44	10.17
PPR.0271	NM.000988	RPL27	TTCTGGTGAGTGATGGGAAA	GAGTGACCTCACTTCTGTGCC	-925,+128	51.36	4.29
PPR.0272	NM.012124	CHORDC1	GGTAGGAGGACGTGAGCAGA	TCCTCTTACCGTCGGAAATG	-947,+181	51.35	1.24
PPR.0273	NM.000373	UMPS	TCCTGTCTCCGGCTTGTTC	GATGTAGATGGGGAGGAAAAG	-915,+144	50.73	15.91
PPR.0274	NM.003134	SRP14	GTTGGCTAGTGGCAGAGAGG	GGCAGAGGTCTGAGTAACG	-935,+121	50.51	3.18
PPR.0275	NM.006384	CIB1	ACTTCTCTCCAGCAGTT	CTGCTCCCGAAGAGAGGC	-987,+159	49.32	4.04
PPR.0276	NM.004299	ABCB7	GTTGGAGAATGCAGCAGTGA	GACTAAAGGCCGGATCAGAA	-920,+133	48.71	7.02
PPR.0277	NM.001240	CCNT1	AAGTTGGATCACTGGCGAAC	TTCTTTATCTGGGCTCCACGC	-937,+162	47.60	1.24
PPR.0278	NM.006812	OS9	TGGTGAGATCCCCAGAGAG	ACTGAGCTCCTCCAGGTTCA	-944,+145	47.59	9.12
PPR.0279	NM.014167	CCDC59	TAGAGCCAAAGCTAGCCTGC	CGCCATGTCTTCTGTCTCAC	-961,+136	47.57	4.31
PPR.0280	NM.002389	CD46	TTACCGGTCATTCCGGTTTA	GCGGAAGACGCTGTTATTTTC	-943,+129	47.03	5.56
PPR.0281	NM.002915	RFC3	CCTTACCTGCAGGAGTAATTCA	CGCCTGCTCCTTGTGATAGT	-896,+169	46.82	9.40
PPR.0282	NM.004649	C21orf33	GGGATCTTCACTGACCCGT	ACAGGACGCTGAATCGAGAT	-937,+141	46.78	6.02
PPR.0283	NM.004099	STOM	TTGCAGGCTGAACCCCTTA	GCTGTCAAGCCCTCGGAGC	-968,+184	46.53	8.83
PPR.0284	NM.018248	NEIL3	TGAAAAATGAAAACAGAAATGGCT	GCGGGAATCTTCTCTCCA	-905,+164	45.61	5.81
PPR.0285	NM.014771	RNF40	TACGCGAGCACAGTAACCC	CACAGAGCAGAGATGTGGGA	-938,+141	45.55	4.14
PPR.0286	NM.018428	UTP6	CAATATTGCCTCCAAAGCTGT	GCCATGAGGTCCCAAGGCTCA	-959,+84	45.39	10.03
PPR.0287	NM.015904	EIF5B	GTGTCCGACCTTGTCTTTC	TTTCTTCCCATTTGCTTGTG	-914,+180	45.22	0.62
PPR.0288	NM.006009	TUBA1A	TTCCAGACAAAAGTCCAATTCA	TTTTCCAAGTAGAGCCTGGG	-943,+165	44.35	2.22
PPR.0289	NM.021203	SRPRB	TCTTCATGTCTCTACTCAGCCA	CAAGAACCAGCCACCACTACT	-976,+181	44.01	3.84
PPR.0290	NM.007062	PWP1	ACAATGCCACTGGGAAAGAC	CGGTTTCAATGGTCTCAAGTC	-955,+111	43.34	8.42
PPR.0291	NM.012322	LSM5	GGGTTCAAGCAGTTCCTCTG	ATGTCAGCTCCCTGTTCTCTG	-932,+111	43.26	13.12
PPR.0292	NM.016940	RWD2B	TGCATATAGTTGGATTGGGAAG	CCCCTCACTGCTGTAACC	-935,+115	43.04	0.69
PPR.0293	NM.004563	PKC2	GTGGGCAGATGAGTTTCTCTG	AGGCCAAGGGCGGTACAAT	-965,+117	42.97	7.20
PPR.0294	NM.001230	CASP10	TTGTTGGGATGGTGAATGA	GGGAAATTCCTGAAAACAA	-943,+139	42.86	4.95
PPR.0295	NM.000454	SOD1	CAGGTACTTGGGAGGCTGAG	GCCTTCGTCGCGCAATACTC	-957,+85	42.57	4.37
PPR.0296	NM.005500	SAE1	CTCCCGGATTCAGCAATTA	GGATCTGCCGGTCACTAGT	-969,+126	42.30	5.34
PPR.0297	NM.015966	ERGIC3	TGTTAACTTGGATAGCTGCTT	GGGTAGCATCGAAGCTGCT	-959,+86	42.11	6.88
PPR.0298	NM.001786	CDK1	GGATATGATGGATGGCAATG	CCCTAGACACAGCCCTGACC	-920,+122	41.81	17.48
PPR.0299	NM.000449	RFX5	GCTGCGCAAAAACAAAATACA	TTTTTCTGCTCTCCTCCCAA	-959,+157	41.26	16.62
PPR.0300	NM.014860	SUPT7L	CTCCAAAAGCGAACACATGC	GACCTGAACCGAGACAAGGA	-920,+146	40.94	7.55

表 3.1 (続き) : 転写活性化能を測定した 472 遺伝子のプロモーター領域の詳細

PPR ID	NM ID	Gene Name	primer (fw)	primer (rev)	Amplified position (-xx to +yy; TSS=0)	転写活性化能	標準偏差
PPR_0301	NM_004208	AIFM1	CCATTGTGAGGGGTAAGTCG	ACCTCCTCCTTCCCTTTTCT	-929,+147	40.90	4.08
PPR_0302	NM_022916	VPS33A	AGGCACCCAGTTAAGCCTCT	CACCTTGTCCAGGAAGCTCGC	-958,+131	40.77	6.48
PPR_0303	NM_007048	BTN3A1	TGGCAGCAGGAGAAACAAAT	GCATTCCTCCTGTAATAATCA	-916,+142	40.24	3.78
PPR_0304	NM_002013	FKBP3	GGAAAGTGAACGTGGTCTGGT	GCCTCTGGTACCGAATCTGA	-903,+172	40.15	5.39
PPR_0305	NM_017807	OSGEP	TCAAGGGAATGACAAACACC	CTGGCAATGTCAGGAGCTG	-932,+118	39.29	8.94
PPR_0306	NM_020357	PCNP	GGTAGCTCTGTAGGCAGTAAGCA	ACCCTGGGAGCTCAAAGAAA	-922,+168	38.90	1.00
PPR_0307	NM_017707	ASAP3	CCACTGTCAATCCACTCCCT	CCCTCTCTGTGATAATCCCT	-952,+141	38.53	0.83
PPR_0308	NM_003295	TPT1	GCTACCATGCAGGACCTGTT	GACTGAAGGGAGACGACGAC	-979,+101	37.94	6.60
PPR_0309	NM_004181	UCHL1	CCCACCTTTGTTCTGCAGCTT	CTCACCTCGGGGTTGATCT	-917,+136	37.84	2.40
PPR_0310	NM_014748	SNX17	TCCTAACTTCCACAAGCTACCTG	GGGAAAGGAAAAGCTACATGT	-935,+166	37.81	3.15
PPR_0311	NM_014303	PES1	AAAAGCTCGAATATGTTGAATGT	ATCCCGTCTTTCCAGTCTCC	-950,+199	37.28	15.28
PPR_0312	NM_015169	RRS1	CCTTAAACGGGCTCAACAA	CAACTTCTCTGCCTGCTCT	-944,+148	36.03	5.66
PPR_0313	NM_004814	SNRNP40	GGAGCAGTAGGAGGCATTTCC	CGCCTGGCTTCTCTCACTACC	-927,+180	35.95	3.23
PPR_0314	NM_017686	GDAP2	AGGCTAAATCGTTTGCCAGT	GGATCAGGCTCCTGGTAGAA	-920,+137	35.90	1.92
PPR_0315	NM_002796	PSMB4	GGTCTTGAATTCCTGGGTTCC	ATGAAGGAATCGGAGAGTGG	-922,+122	34.98	12.59
PPR_0316	NM_018605	RBM26	AGGGCTCTACCAGCAAGTCA	ATCTCTGGGCTCCGCCTATCT	-940,+172	34.79	3.36
PPR_0317	NM_016221	DCTN4	TTTCTTCATGGTGCAGGACA	TCGCTATAGGGGCACTCTGT	-941,+180	34.30	1.28
PPR_0318	NM_004280	EEF1E1	TTCTGGGATAGGGAAGGTGT	CCTCTCCTTCTCACCTCTCG	-975,+131	34.07	5.47
PPR_0319	NM_001168	BIRC5	AGTGAGTGGATGTGATGCC	AAGGGCCAGTTCCTGAATGT	-923,+127	33.99	1.69
PPR_0320	NM_015907	LAP3	GCACCTGTGGAACAGATGAA	AGCGGACGTGCGTCTAGG	-920,+150	33.90	1.03
PPR_0321	NM_003825	SNAP23	TGGGGTTTTCTCTATGTTGGC	GACCCCAACCACTTTTCAAC	-935,+177	33.82	2.99
PPR_0322	NM_005872	BCAS2	GCAGCAGCTTGAGTGGTTC	CTAGCAAGACACCCCGCACT	-960,+145	33.36	9.72
PPR_0323	NM_004184	WARS	CAGCTCATCTCCAATGGAAC	CACTCCCTCCCTTGTTCCT	-948,+119	33.17	7.50
PPR_0324	NM_001064	TKT	GTTCAAGACCTGCCTGGGTA	CAGGCTTGTGGTAGCTCTCC	-968,+90	33.06	9.71
PPR_0325	NM_003174	SVIL	GAGCTCGCCACTGGCTCT	GGAGGCACTTGGCTCACTA	-963,+191	32.53	5.42
PPR_0326	NM_000016	ACADM	TCCCTTTTCCGAATTTTCT	ACAATACCATGTTCCAGCC	-998,+145	32.52	7.04
PPR_0327	NM_002394	SLC3A2	CGCTGGCTTTGCATAAAGAT	GTCGTGACCGTCAGTCCAG	-927,+119	31.77	2.44
PPR_0328	NM_017897	OXSM	GCCTGGGATAGCATGTGAAA	AGCATGGCAGAGAAAAGA	-994,+195	31.44	5.01
PPR_0329	NM_016147	PPME1	CTTGCCCAACATCACACAAC	GCACGCAGTAGTCAAACAG	-933,+122	31.19	4.65
PPR_0330	NM_003678	THOC5	TCAGTACTGGTGGTGCAGC	CGGTAGTGAGGGAGGAGA	-945,+187	31.11	3.27
PPR_0331	NM_016437	TUBG2	GCCAGACCCAAAAGACAAA	CTCGCTTGTCTACTCTGGTT	-955,+150	31.06	1.86
PPR_0332	NM_017812	CHCHD3	CACTGACTCACTCAGGGCAA	ATGTTCTCATTCTCGTCCGC	-949,+195	30.96	2.44
PPR_0333	NM_019023	PRMT7	CGCTCCTTACATCCCTTATC	CGGAGGATGCCAAGTAGGC	-978,+200	30.94	2.54
PPR_0334	NM_003522	HIST1H2BF	GTCAAGGCTGTCTGGAACCTC	CACGGATAGCTCTCCTTGG	-961,+177	30.74	0.49
PPR_0335	NM_014706	SART3	GGCTCGCTGAGTCTGAATCT	GTCTAGCCGCCTTAACCTC	-944,+116	30.70	12.56
PPR_0336	NM_017425	SPA17	TGTCACGCTAGGGTATTTCC	TGTGTCTGGGATGCTGAGAG	-936,+123	29.92	1.53
PPR_0337	NM_032365	ATAD2	TTATTATGTCAAGCAGCGC	TGTGCTCACAGACTGAGGAG	-961,+177	29.49	3.50
PPR_0338	NM_006924	SRSF1	CGGTATTTTCAAACCTGGTG	CCACGAATCACACCACCTC	-990,+101	28.64	4.59
PPR_0339	NM_006824	EBNA1BP2	TGGTGTGCTCAGGAAAAGCT	CTGCTTTCAGCCCTTACT	-938,+118	28.07	3.56
PPR_0340	NM_018157	RIC8B	CGGGTTCTCTTTTATTTCCA	TCGGTGTAAATCCCTCAGGAC	-999,+143	27.52	1.16
PPR_0341	NM_014175	MRPL15	GCCCTCCAAAATTTGTCTCA	CCATTTACCGGTTTCTTGGGA	-987,+146	27.42	1.77
PPR_0342	NM_003336	UBE2A	TTATTGCTTTTGCAGGTCCA	ATACACTGGGGTCTCGGGTT	-924,+133	26.75	3.67
PPR_0343	NM_031292	PUS7L	CGCCACTATGCCAGCTAAAT	TGTCCAAAGGCCACGACACTA	-964,+196	26.26	4.50
PPR_0344	NM_001610	ACP2	CCCCAGGACCTATGATCTGA	CAGGGAAGTCTTTGGTGAGC	-920,+176	26.21	2.15
PPR_0345	NM_004135	IDH3G	CTGGAGTTCAGACCTCAGCC	ACTCCAGACTGCTTCGGGT	-961,+197	26.18	1.69
PPR_0346	NM_015434	INTS7	TAACAGGGGTCTCTCTGGC	GTTGGCATCCAGTTCCTGTT	-988,+172	25.45	0.05
PPR_0347	NM_012429	SEC14L2	TCCTGGACCTAGCTCTGCAT	CAGGGCTACAGCTCACCTTG	-981,+158	25.32	1.05
PPR_0348	NM_015472	WWTR1	CGCACCTCTCTACTCCGAG	AGGCTTGGCTGACAAATCC	-929,+144	25.29	1.95
PPR_0349	NM_000981	RPL19	TTGCGAGATAATGGGAGGAG	CAGTACTCACGGGAGTTGG	-929,+121	25.08	1.59
PPR_0350	NM_005452	WRD46	GGATGAGAGGAAAGGCAATTT	CAGGGAGGCTCTACCTTTC	-931,+151	24.31	2.45
PPR_0351	NM_022374	ATL2	AGCTGTTTGTAGACCCGGAG	GTACCTAGGGAGGTCGTGGA	-931,+141	24.28	7.20
PPR_0352	NM_006623	PHGDH	AGGAAGTGGAGGAAAGGAA	AAAAGCACTTGTGAGGTTG	-973,+172	24.20	7.53
PPR_0353	NM_003537	HIST1H3B	GATTATGCCGCCATAAAGA	CGGTAAACGGTGAGGCTTTT	-961,+167	24.17	2.86
PPR_0354	NM_016297	PCYOX1	TCCTGCTTTAGGCAAAATGG	GATTTTATCTGGCGGAGCAC	-988,+151	24.02	2.79
PPR_0355	NM_004450	ERH	ATGTGTTTACGAGGCACCTT	GGGAGGGGAAAACGTATGG	-928,+154	23.96	2.50
PPR_0356	NM_002805	PSMC5	CCAGCTCAGAACCCAGAAAG	TCTCCACTCAGATCACGCTC	-898,+177	23.57	2.73
PPR_0357	NM_005760	CEBPZ	ACCACCATCAGCCAAGTAG	GAGCCGTAACACTTCTCTCCA	-973,+163	22.66	4.34
PPR_0358	NM_006280	SSR4	GTGGGAGAAGTCAGCATCGT	GTGGGAGGCTCTCACTTT	-939,+122	21.41	4.12
PPR_0359	NM_031488	L3MBTL2	TTTGAAAAGCTCCACAACAG	GTTTATCCATCCTCCACCC	-991,+190	21.33	2.08
PPR_0360	NM_015710	GLTSCR2	GAAGTAGCCCGCTCAAGATG	AGCCCGCTTCTTATTTCTT	-987,+154	20.89	1.35
PPR_0361	NM_001188	BAK1	AGACATGGTCTCACTTTGTTGC	GGCATCTGGATGTAGCCTTT	-927,+122	20.61	3.05
PPR_0362	NM_001412	EIF1AX	TGGCCATATGTGGCTACTAAC	CGGGCTCTGACTTCTTCT	-921,+178	20.11	1.72
PPR_0363	NM_014750	DLGAP5	GAGTGGTGTGATGTGGTG	GTGTTCTCCGTCAGCTCTC	-900,+161	19.27	1.67
PPR_0364	NM_006407	ARL6IP5	GCCACTTAGGGGCTTAGAGA	CGGAGTGGGGCATATAAAC	-990,+136	18.40	1.97
PPR_0365	NM_004050	BCL2L2	CTCTTCTGGCCCTTTGGTTG	CCTGCCGACACCATCTCTAT	-975,+151	17.46	1.49
PPR_0366	NM_007067	KAT7	TCCTTCTAGGCCCTCTCAT	CTCCCGTTTCTCACCTTCT	-899,+195	17.34	2.37
PPR_0367	NM_024945	RMI1	GCTCACACAAAGAGAGCGG	AAACACGGGCATTAAGCTGG	-928,+166	17.30	1.06
PPR_0368	NM_004939	DDX1	CGGCTTCTTCTGTGTGTGTG	ACCCCACAGATTTCCACAA	-968,+96	17.30	0.38
PPR_0369	NM_004982	KCNJ8	CAGACCATCTAGACAGAGG	GAAAAGTCATCCACTCACTG	-978,+188	17.12	6.00
PPR_0370	NM_005104	BRD2	GATACGGTGTGCCTTTGG	CACCGTCTTGAGATTTTCC	-926,+155	17.11	4.72
PPR_0371	NM_015679	TRUB2	TGGTGCATGACAGATTCTC	CAGAAAGCGGAGATGGAATTC	-966,+199	16.99	1.44
PPR_0372	NM_018010	IFT57	CAGGCATCCACCCTTACTCT	GACCTAGGCCACCCCACTTCT	-993,+148	16.19	1.17
PPR_0373	NM_002946	RPA2	CCAGGAACAAGGAGGCATAA	CACGGATGCTAGCTTGT	-933,+124	16.06	4.99
PPR_0374	NM_002697	POU2F1	GGGTGAAGGCAGAAAGGA	AGCCGGGTTGAGTATGAAT	-923,+119	15.80	2.99
PPR_0375	NM_005005	NDUFB9	ACTACGGCCATGAACACTC	GTCCCATCCCTTACTCTGT	-963,+156	15.18	1.16

表 3.1 (続き) : 転写活性化能を測定した 472 遺伝子のプロモーター領域の詳細

PPR ID	NM ID	Gene Name	primer (fw)	primer (rev)	Amplified position (-xx to +yy; TSS=0)	転写活性化能	標準偏差
PPR_0376	NM_003135	SRP19	CCAAGCAACTTCTTTTCAGGC	CTCAGAACCACCTGTCCTG	-954,+134	14.97	1.60
PPR_0377	NM_015957	APIP	GGCCTCAATGAAAACACCTC	GCCTTCTCTCTGGGAATAC	-927,+160	14.92	2.01
PPR_0378	NM_015510	DHRS7B	GTGGGCTCTGGAGAACAGAA	CTCGACCACCTACCCGCTTC	-945,+148	14.84	1.49
PPR_0379	NM_004493	HSD17B10	TTATGGCCTGAAGACCTTGG	ATGACCCTTTCCTCGAAAC	-943,+150	14.75	3.83
PPR_0380	NM_002413	MGST2	CACCATGCCCTGCTAATTTT	GCTGACCCTTCTTATCGGGTA	-938,+140	13.95	1.66
PPR_0381	NM_033084	FANCD2	TCCACCCTAGGAAGGAAAT	CAGAAAGGTACCTGTGGGGA	-940,+199	13.87	1.07
PPR_0382	NM_012465	TLL2	CAGGGCCATTTTACTTGCAT	GGACTCGGTGGTTACACAGG	-997,+131	13.53	0.35
PPR_0383	NM_002787	PSMA2	AAGCTTTTCTTTCAAGGCC	CCAACCTCATTCCCTACTGG	-934,+180	13.45	2.66
PPR_0384	NM_003958	RNF8	GGCGAGACCCTGTCTGTAAT	TCTCCTGTGACGAAGAAGCC	-971,+179	12.42	1.01
PPR_0385	NM_014267	C11orf58	CCAAAATGCTGGGATTACA	GCATACCTGGCCGGAATAAAA	-935,+176	12.40	3.40
PPR_0386	NM_017979	UNC45A	CCTCACTTGTGCCAGACCAC	CTCATGGATGGGAGAAAGG	-992,+108	12.06	2.85
PPR_0387	NM_024102	WDR77	CAGGGGAAGGTCACAAAGGA	CTCCCACTGCCGTTCCAT	-942,+140	11.65	1.08
PPR_0388	NM_003142	SSB	TGGGGATTATGGGATTACA	AGCTTCTCATTAGCCACCGA	-982,+143	11.56	1.62
PPR_0389	NM_024094	DSCC1	TTTAGCTCTCAGGGACAGCC	GCATTGAGCTTGGCGATCT	-931,+139	11.53	1.07
PPR_0390	NM_005174	ATP5C1	TAAGTGGCCCTGCTTGACTT	AAAGCTTCCATCCCTTTCG	-922,+141	11.35	2.11
PPR_0391	NM_014764	DAZAP2	GGCTCAAGTGATCCTCTTC	AGCCTCAGAAAGTGCACAGG	-922,+136	11.03	0.98
PPR_0392	NM_006460	HEXIM1	TCCTGAAACAGAATGGTACGG	TTGCTAGTTTGGGCTGGTG	-934,+147	10.33	0.25
PPR_0393	NM_002913	RFC1	CCAGATTTACAGGTTGGGATG	GGCCTTGGGAGAAAGACTT	-967,+138	10.10	0.83
PPR_0394	NM_016059	PP1L1	GCTCCTCTCTCAAGGAGCTG	CGTTCAAGCCCTTCGATACC	-942,+135	10.07	0.12
PPR_0395	NM_022839	MRPS11	CCTCACACAGATAGCCCAT	AAGGGACACGAGTAGGAGCA	-952,+158	10.07	0.97
PPR_0396	NM_001757	CBR1	ACAATTCGCTGATGCCTGTT	GTCACCAAGCCGTACATGGAT	-966,+161	9.86	0.92
PPR_0397	NM_012450	SLC13A4	GACTCTGCTCAGCATCTCCC	GGCTCCTGTTGGTAAAGCT	-951,+196	9.85	0.53
PPR_0398	NM_003594	TTF2	CCTTCTGTGGTTCCACGTT	GCTCTTCTTTTATTCGGGC	-929,+190	9.69	1.16
PPR_0399	NM_015497	TMEM87A	GGAGTCCCTCCAGATTCTC	ATACTCACCGCGGTATCGG	-919,+189	9.59	0.40
PPR_0400	NM_007054	KIF3A	AATGGAGGGTTTTGAGCAGA	ACTGCTCTCACCGGATCTTT	-960,+129	9.51	2.29
PPR_0401	NM_006209	ENPP2	GCCCTTCCCTTTTGTGTTAG	GTAGAGAGAGGCGCATACCG	-996,+151	8.92	0.71
PPR_0402	NM_012138	AATF	CTAGGCTGATTACCGTTGGG	AACTGTTCCAGTTGCAGCG	-969,+144	8.70	0.27
PPR_0403	NM_000466	PEX1	CGCAATAAGCCATAGGACGA	TGGTGAAGGCCACAGCTCAC	-951,+123	8.63	1.77
PPR_0404	NM_003287	TPD52L1	TCGTGGGAGGAAGTCAAAC	ACCACTCACCTTGTGCCTG	-967,+176	8.26	0.80
PPR_0405	NM_018206	VPS35	TGCTGTTTTCTTGCCTTGG	CTAGATGGACGGACCTGAGC	-921,+185	7.76	0.38
PPR_0406	NM_002914	RFC2	ATGACAAATTTGGGCTATGG	AAACGGCCCTTCTTTTAC	-999,+177	7.56	1.36
PPR_0407	NM_031299	CDCA3	AATACTGACAGCAGCGCCTC	TGCGGAGTTAATGACTGCTG	-935,+153	7.48	1.32
PPR_0408	NM_014666	CLINT1	AGATCCTCCTGCCTCAACCT	GCTGGTTTCAAGTGCAGACA	-920,+178	7.16	0.47
PPR_0409	NM_002916	RFC4	CGGTGGACATTTCTGAAGCA	GTTCTCCAGCAGGTTACCA	-935,+188	7.10	0.65
PPR_0410	NM_005032	PLS3	TCCCACAAGTCAGTCTCTTGA	GCTGAGCTTAAACGAGATGC	-925,+186	6.82	1.88
PPR_0411	NM_004199	P4HA2	TTTGTGTTTCAACCTCCTTGG	CAAGACTCGTGACCACTGGA	-923,+122	6.74	0.99
PPR_0412	NM_021800	DNAJC12	TGGAGTCTTGCTCTGTAGCG	AATCAGTCTCTTCTCCCTCG	-958,+85	6.54	0.55
PPR_0413	NM_015361	R3HDM1	ACTTCCAGCCAGTCACACCT	TTACCCGGTTACTGAGGAG	-961,+168	5.87	0.43
PPR_0414	NM_006826	ZBTB6	GGTGCAGGAACACAGATGC	CCCTCGGAACTCAGTGTGAT	-929,+147	5.25	0.44
PPR_0415	NM_003754	EIF3F	AAAAATGGAGTAAACAACCCCA	GTCTGAGGATGAGGCTGGAG	-949,+168	5.11	1.89
PPR_0416	NM_014673	EMC2	CCAATAAATTTCCGGCTGGAT	GACACCAGAGAACCAGGCTC	-974,+153	4.94	0.34
PPR_0417	NM_018290	PGM2	CCCCTGACCAAACTACTGG	GTGCCCCCTCACCTTGTCC	-922,+140	4.62	1.26
PPR_0418	NM_014814	PSMD6	TGCAATCATGTTTGTGAATGA	CATCTCCAGTCTGCGAGGTA	-965,+153	4.49	1.02
PPR_0419	NM_018061	PRPF38B	TTCCATAAATTTGCATCACAA	GATCTGTAGGGGCTGTGTA	-975,+201	4.37	2.10
PPR_0420	NM_016047	SF3B6	GGGACGGAGATTCTGGGTAT	GCGGGCTGATGAAGTTACC	-976,+107	4.29	1.31
PPR_0421	NM_006109	PRMT5	CTCAGGACTCCTTGCCAGTC	GCCGGGATTCTTGACTACTA	-992,+55	4.18	0.57
PPR_0422	NM_006090	CEPT1	GGTGGCTCTCCCTGTAAT	TCAACCCCTTCCCAAAAAG	-923,+123	3.70	0.81
PPR_0423	NM_020165	RAD18	TTGCAATTGGTAAAAGCTGG	AGGCATCCTCCTCAAAGGG	-932,+187	3.61	0.43
PPR_0424	NM_004990	MARS	TGGAGAGAAAGGGTATGGA	CTCCACCTGGCAAAGACATC	-921,+122	3.56	0.16
PPR_0425	NM_004637	RAB7A	AGCGGCAAAACCAAGTATGA	TCCACAGCAGAGAGGCTTAG	-944,+122	3.28	0.31
PPR_0426	NM_006431	CCT2	CGCCCGGCTAATTTTTGTAT	GCAAGAGGCGAGGGAGCT	-968,+117	3.24	1.91
PPR_0427	NM_012255	XRN2	GGGACCCAGCAAAAACAAA	GCCTCTCACCTTCTCTTCC	-911,+161	3.15	0.19
PPR_0428	NM_007317	KIF22	CCCAAGTTAAGAAATGGAGGT	ATCACTTGCTCCCGCTGTT	-923,+120	3.15	0.30
PPR_0429	NM_017749	AMBRA1	AAGGGTGTTCACAGCAGTGT	CCCTCATTACCCTTCCAAGC	-950,+121	3.03	0.58
PPR_0430	NM_012245	SNW1	CTGGACAGTTTATCCCTGAAGT	GTACGGGGAGGTTCTCTTC	-980,+119	2.86	2.77
PPR_0431	NM_018685	ANLN	CAGGAGGATTCTCTTTTCC	ACGCATCATCAGCACTTGAG	-931,+134	2.83	1.13
PPR_0432	NM_033018	CDK16	GAGCCTGTTCTATGCCAAGG	CCCCTACTCGGTCTCACTTG	-968,+168	2.83	0.11
PPR_0433	NM_006183	NTS	AACAGGAAAGAAAGCCGAT	TGATTTTCTATTCTGCCAT	-955,+132	2.46	0.32
PPR_0434	NM_005758	HNRNPA3	GCTTATCTGCCCCAAAT	CTGGTCTCTTGGATCATGG	-983,+118	2.08	0.53
PPR_0435	NM_002118	HLA-DMB	AAAGCAAGCTCCTCTTGTG	GGGTAGCAGCCACAGAAGTG	-928,+151	2.02	0.71
PPR_0436	NM_015425	POLR1A	CCTGAGGAGTGGGAGGAG	CACCTGTAAAGCTGCCAACC	-920,+135	1.98	0.61
PPR_0437	NM_004643	PABPN1	CCAGCCCTGCACACTTCTA	GGTGTACCTGCACCAGTCT	-962,+122	1.85	2.07
PPR_0438	NM_012474	UCK2	GTGCTTACCTCGGAGCAG	GGAAGGAGGTTCTCAGTGCAG	-949,+150	1.77	0.57
PPR_0439	NM_000183	HADHB	TTGAAAAAGTTGCAAGTGCAGTA	CCATTGAGTATCTGTGCGAGA	-948,+150	1.71	1.08
PPR_0440	NM_024755	SLTM	GGCTACATATTTGAGGGGAGG	TGACAGAGGCTTGTGCTCTG	-995,+163	1.68	0.91
PPR_0441	NM_004882	CIR1	CCGGCCAAAGTTTATCTTTA	TGGAGGCGAGTGAAGTCT	-827,+125	1.56	0.39
PPR_0442	NM_004969	IDE	GCACGCTGTTGCCTTTT	TTTGCTCTTGA AAAACAGATCC	-922,+144	1.51	0.23
PPR_0443	NM_023074	ZNF649	CGCAGTGTATTGTGCCCTG	CTCAGAAGCTCCCTACCCCT	-988,+139	1.51	0.67
PPR_0444	NM_006904	PRKDC	AAGGCTAGTTTTGGCTTTTC	CCCAAGTACCCACCTCTGTG	-936,+140	1.41	0.54
PPR_0445	NM_001380	DOCK1	AGATGCCGTCGTGATAAG	CTGAAAGCCCTGAACAAGC	-944,+120	1.39	0.09
PPR_0446	NM_002695	POLR2E	ACCTGGCCAGCTAATTTTT	ACGAGAGTACGAGAGACGC	-974,+199	1.39	0.05
PPR_0447	NM_005493	RANBP9	ATGCAGTCTCTTTGGGATA	CCAGATTTCAACTTTCCAA	-999,+126	1.31	0.16
PPR_0448	NM_018149	SMG8	CTCTTTTCCGCCCAAGAT	TTCAAAAATGGAAAGGCTGTG	-955,+135	1.26	0.40
PPR_0449	NM_002802	PSMC1	TTCCATTCTCAACCAGTCTCT	CAGCATCTGGTCCCTTTGTT	-923,+157	1.25	0.73
PPR_0450	NM_012343	NNT	TTGGATACATGATATGGGGTAGG	CAGTGGTGTGAGGAGGAGCAG	-993,+126	1.24	0.38

表 3.1 (続き) : 転写活性化能を測定した 472 遺伝子のプロモーター領域の詳細

PPR ID	NM ID	Gene Name	primer (fw)	primer (rev)	Amplified position (-xx to +yy; TSS=0)	転写活性化能	標準偏差
PPR_0451	NM_014003	DHX38	CTTAAGGGGAGATGGGGAAG	TGCAACAAGAAAAAGGCTGA	-931,+123	1.23	0.31
PPR_0452	NM_001752	CAT	CATGGGGGAGAACATGCAGAC	ACCAATCAGCCAAATAAACCC	-920,+139	1.22	0.57
PPR_0453	NM_005087	FXR1	GTGGGCAAAAGTGAACAGACA	CCTGGATGCCAAACAAGAAC	-948,+126	1.17	0.10
PPR_0454	NM_005891	ACAT2	GCTGGTGTCAAACCTGCTGAC	TGATGGGTCTTTTGTCTCTC	-962,+121	1.14	0.19
PPR_0455	NM_001879	MASP1	TGTTGTTGTGGAGGGGGTAA	ATTTGACACTGGTCCCAAAA	-921,+126	1.11	0.15
PPR_0456	NM_006819	STIP1	CCAAACTTTCCCGCACTTT	ATACCGCTGCTTGATTGGTA	-921,+119	1.09	0.20
PPR_0457	NM_014742	TM9SF4	GGAGCCCTGAGGTTGTTTTT	GCTCCAAAAATCCAGGAACA	-982,+143	1.06	0.40
PPR_0458	NM_015387	MOB4	CCTTCCCTCAGCATTCTCAA	CCTAGTTGCCGACAAGAAA	-987,+118	0.89	0.16
PPR_0459	NM_001398	ECH1	TTGAGATCTTGCTCTGTGCG	CCCATAAGGCAAGAGGTGAC	-935,+130	0.86	0.19
PPR_0460	NM_002810	PSMD4	AATACAAAATTAGCCGGGCA	GTCCAGCGGGCTCCTTAC	-956,+200	0.86	0.07
PPR_0461	NM_000972	RPL7A	ATGAGCATCTGGGCTCTTTC	CAAACAGGGGATTCACCACT	-920,+119	0.83	0.30
PPR_0462	NM_004526	MCM2	CTGGGCGTGACTCATTCTTC	CACACACACCCCTGAGTC	-950,+129	0.80	0.04
PPR_0463	NM_006659	TUBGCP2	CCCACCTTGGGTGTCTGTGA	TGTAGACCTCAGCCCATCT	-919,+124	0.77	0.19
PPR_0464	NM_002567	PEBP1	TGTGTGGGAAAACGGACAA	GCTCCCTGGGCTCTGATT	-957,+151	0.77	0.40
PPR_0465	NM_024658	IPO4	CAACTTGGCAAAGACCTGT	ACAGGGAAGGAGAGTCCCAG	-978,+126	0.68	0.26
PPR_0466	NM_000108	DLD	GCCCTTAGGAAATGGTAGCC	GAAAGAACTGTCAGCTAATGTGA	-951,+135	0.63	0.05
PPR_0467	NM_030877	CTNBNL1	GAGTGTGGGAGAGTCTGAGATG	CCCATTGCCGGAAGACTTT	-931,+197	0.63	0.03
PPR_0468	NM_001863	COX6B1	CCCTGGATGATGGAGAAAAA	TTGTTACTGTTGGGGTGAA	-964,+160	0.61	0.06
PPR_0469	NM_024725	CCDC82	GATCTCCCTGCATCTTTCCA	TTCTCGCTTTTGTGCAGCTA	-991,+133	0.57	0.06
PPR_0470	NM_012086	GTF3C3	TGGGCAACAATGTAGTAACC	CCTAGCATCGCCTCACTTTC	-953,+191	0.48	0.04
PPR_0471	NM_006153	NCK1	GAAACCCCGTCACTGCTAAA	ACAAGCCGCTCTCACCTTC	-924,+133	0.33	0.05
PPR_0472	NM_007344	TTF1	CCACAAGGTGGGCTAATAA	GAGACCTCAGGCCAAAAAGG	-918,+163	0.30	0.07

(図 3.3) に測定された 472 遺伝子のプロモーター領域の転写活性化能のヒストグラムを示す。図はランダム領域 (後述) が有する転写活性化能の測定値の平均を基準(10^0)とした。(図 3.3)に示すように、各遺伝子のプロモーター領域が有する転写活性化能の分布は転写活性化能 $10^{0.8}$ を境界とする二峰性の分布様式を示すように考えられた。転写活性化能の値が $10^{0.8}$ 以上の強い転写活性化能を有するプロモーター領域群を P1(遺伝子のプロモーター領域群全体に対し 87%)、転写活性化能の値が $10^{0.8}$ 未満の弱い転写活性化能を有するプロモーター領域群を P2(遺伝子のプロモーター領域群全体に対し 13%)として、以下の解析を行った(表 3.4)。

プロモーター領域群 P1 および P2 に対し、転写活性化能の分布様式における差異を統計的に調べるため、Kolmogorov-Smirnov 検定および chi-square 検定を行なった (表 3.3)。ひとつの正規分布からなる一峰性の分布であることが否定された。統計的に二峰性の分布であることが示された (表 3.3)。

プロモーター領域以外のヒトゲノム配列中からランダムに抽出した、遺伝子のプロモーター領域と同等の長さの DNA 断片に対しても遺伝子のプロモーター領域と同様の方法を用いて、転写活性化能の測定を行った (ランダム領域、と称する)。クローニングされたランダム領域の配列を決定し、ゲノム配列との配列相同性を比較した。既存のプロモーター領域と重複するプラスミド DNA を除去し、ヒトゲノム配列中からランダムに抽出され、遺伝子のプロモーター領域同様の約 1.0kb の DNA 断片であるものを選択した。最終的に 251 種類のランダム領域が有する転写活性化能を測定することができた(図 3.2)。測定に用いたランダム領域の詳細を(表 3.2)に示す。

表 3.3 : 遺伝子のプロモーター領域群 P1 および P2 が有する転写活性化能の分布の差異

(P1+P2) プロモーター領域群 P1、P2 を含めた転写活性化能の分布がひとつの正規分布に所属するか Kolmogorov-Smirnov 検定および chi-square 検定による p 値を示した。(P1) (P2)プロモーター領域群 P1 および P2 が有する転写活性化能の分布がそれぞれひとつの正規分布に所属するか Kolmogorov-Smirnov 検定および chi-square 検定による p 値を示した

	K. S. test	chi-sq test
P1+P2	5.6E-05	<1.0e-200
P1	0.85	0.55
P2	0.54	0.92

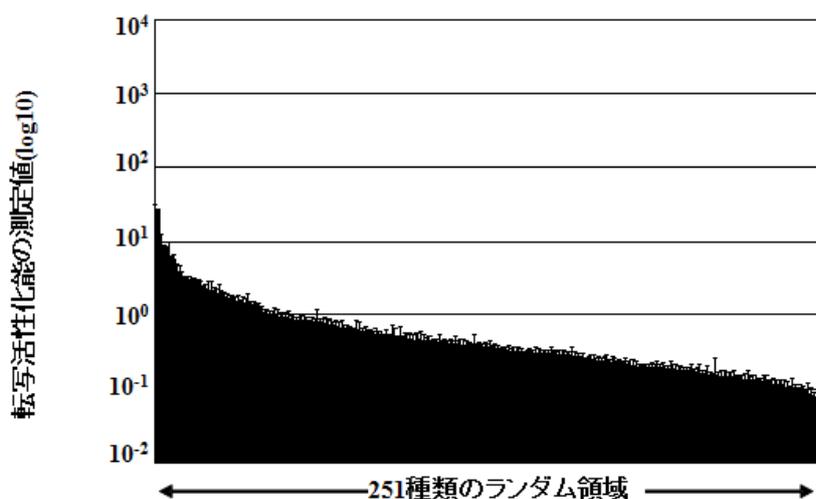


図 3.2 : ランダム配列の転写活性化能

転写活性化能の測定を行なうことができた 251 種類のランダム領域が有する転写活性化能の測定値を、転写活性化能の強い順に左から並べた。

表 3.2 : ランダム配列の転写活性化能

Genomic fragment ID	Gene Location	Amplified length	転写活性化能	標準偏差
Random_0001	chr15,35590094,35591033,+	940	27.383	4.293
Random_0002	chr8,4078000,4077123,-	878	25.689	1.140
Random_0003	chr8,9422941,9424092,+	1152	11.198	1.098
Random_0004	chr12,78333346,78332275,-	1072	8.761	0.206
Random_0005	chr3,50980609,50979539,-	1071	8.115	0.382
Random_0006	chr3,138579032,138579998,+	967	7.477	2.363
Random_0007	chr17,36401439,36400570,-	870	5.950	0.325
Random_0008	chr6,34055810,34056719,+	910	5.676	0.731
Random_0009	chr10,31713744,31712678,-	1067	4.042	0.718
Random_0010	chr9,69636078,69637141,+	1064	3.983	0.549
Random_0011	chr18,7525872,7524792,-	1081	3.343	0.500
Random_0012	chr1,17777280,17778327,+	1048	3.228	0.145
Random_0013	chr10,109842571,109843520,+	950	3.131	0.261
Random_0014	chr8,30426631,30427734,+	1104	3.099	0.246
Random_0015	chr7,78744649,78743599,-	1051	2.991	0.066
Random_0016	chr2,18015555,18014700,-	856	2.914	0.238
Random_0017	chr14,89108498,89109455,+	958	2.736	0.332
Random_0018	chr17,13415383,13416421,+	1039	2.683	0.279
Random_0019	chr4,93883061,93882047,-	1015	2.416	0.118
Random_0020	chr18,7524793,7525873,+	1081	2.256	0.328
Random_0021	chr4,150560397,150561283,+	887	2.253	0.682
Random_0022	chr8,9422944,9424086,+	1143	2.160	0.775
Random_0023	chr1,64743892,64742850,-	1043	2.118	0.308
Random_0024	chr12,109063555,109062644,-	912	1.989	0.308
Random_0025	chr13,105386498,105385484,-	1015	1.980	0.615
Random_0026	chr3,99562555,99561526,-	1030	1.907	0.211
Random_0027	chr12,39036693,39037632,+	940	1.850	0.191
Random_0028	chr1,67029525,67030514,+	990	1.687	0.264
Random_0029	chr11,64202879,64203817,+	939	1.683	0.145
Random_0030	chr21,28986440,28987384,+	945	1.660	0.214
Random_0031	chr14,16268811,16267879,-	933	1.596	0.193
Random_0032	chr5,171465160,171466190,+	1031	1.587	0.259
Random_0033	chr12,2508525,2509408,+	884	1.542	0.076
Random_0034	chr1,64742851,64743893,+	1043	1.480	0.240
Random_0035	chr9,115881822,115880862,-	961	1.480	0.234
Random_0036	chr17,40653879,40654839,+	961	1.474	0.465
Random_0037	chr1,211893852,211894882,+	1031	1.439	0.081
Random_0038	chr6,45949582,45950494,+	913	1.424	0.107
Random_0039	chr3,184093542,184094513,+	972	1.370	0.099
Random_0040	chr9,107653477,107652312,-	1166	1.297	0.134
Random_0041	chr12,39036694,39037633,+	940	1.187	0.140
Random_0042	chr6,168991558,168992717,+	1160	1.093	0.075
Random_0043	chr17,27291470,27292342,+	873	1.056	0.156
Random_0044	chr7,147232508,147231463,-	1046	1.048	0.093
Random_0045	chr8,97583412,97584608,+	1197	1.047	0.074
Random_0046	chr17,57675488,57676558,+	1071	1.010	0.261
Random_0047	chr1,31123646,31122469,-	1178	0.981	0.183
Random_0048	chr18,5774633,5775787,+	1155	0.969	0.080
Random_0049	chrX,31888240,31887377,-	864	0.959	0.066
Random_0050	chr1,25097460,25096441,-	1020	0.946	0.113
Random_0051	chr3,84426552,84427550,+	999	0.932	0.174
Random_0052	chr12,26530719,26529886,-	834	0.908	0.055
Random_0053	chr7,2883037,2881994,-	1044	0.860	0.154
Random_0054	chr14,56467624,56468526,+	903	0.855	0.063
Random_0055	chr1,67029526,67030515,+	990	0.855	0.067
Random_0056	chr13,61915987,61917004,+	1018	0.843	0.133
Random_0057	chr3,194851983,194853120,+	1138	0.841	0.186
Random_0058	chr17,68269645,68270599,+	955	0.831	0.086
Random_0059	chr5,78253375,78254366,+	992	0.825	0.022
Random_0060	chr9,98901353,98900417,-	937	0.825	0.115
Random_0061	chr20,58893940,58892840,-	1101	0.822	0.077
Random_0062	chr3,188609139,188610122,+	984	0.819	0.381
Random_0063	chr6,166771950,166771034,-	917	0.805	0.068
Random_0064	chr6,138748164,138747238,-	927	0.780	0.099
Random_0065	chr3,84426553,84427551,+	999	0.777	0.156
Random_0066	chr4,78546652,78545607,-	1046	0.768	0.070
Random_0067	chr10,3262701,3263683,+	983	0.753	0.143
Random_0068	chr14,49006331,49005317,-	1015	0.735	0.042
Random_0069	chr3,194851982,194853119,+	1138	0.711	0.128
Random_0070	chr12,125350037,125350848,+	812	0.685	0.126
Random_0071	chr10,117838397,117839597,+	1201	0.685	0.143
Random_0072	chr6,45949583,45950495,+	913	0.680	0.056
Random_0073	chr2,141769918,141770944,+	1027	0.661	0.059
Random_0074	chr5,145442232,145441048,-	1185	0.661	0.031
Random_0075	chr9,117525322,117526216,+	895	0.632	0.025

表 3.2 (続き): ランダム配列の転写活性化能

Genomic fragment ID	Gene Location	Amplified length	転写活性化能	標準偏差
Random_0076	chr21,29687700,29688550,+	851	0.625	0.042
Random_0077	chr3,58008950,58009870,+	921	0.623	0.215
Random_0078	chr5,171465234,171466151,+	918	0.613	0.164
Random_0079	chr11,122348497,122347316,-	1182	0.605	0.052
Random_0080	chr13,36112022,36110962,-	1061	0.600	0.088
Random_0081	chr2,206996906,206997889,+	984	0.592	0.049
Random_0082	chr5,45823982,45824957,+	976	0.585	0.049
Random_0083	chr1,36284480,36285464,+	985	0.568	0.040
Random_0084	chr1,23552615,23553715,+	1101	0.553	0.112
Random_0085	chr6,7185303,7184323,-	981	0.548	0.047
Random_0086	chr9,108709197,108710066,+	870	0.544	0.063
Random_0087	chr8,75114785,75113796,-	990	0.531	0.030
Random_0088	chr22,44770179,44771306,+	1128	0.528	0.110
Random_0089	chr12,93356395,93357244,+	850	0.516	0.045
Random_0090	chr3,27922735,27923574,+	840	0.516	0.198
Random_0091	chr9,115880863,115881823,+	961	0.507	0.169
Random_0092	chr6,47657861,47658877,+	1017	0.507	0.022
Random_0093	chr20,57656371,57657303,+	933	0.499	0.188
Random_0094	chr7,139473298,139472344,-	955	0.491	0.035
Random_0095	chr11,13677781,13678908,+	1128	0.487	0.083
Random_0096	chr6,47657862,47658878,+	1017	0.479	0.076
Random_0097	chr1,4243062,4244011,+	950	0.477	0.099
Random_0098	chr3,12444331,12445416,+	1086	0.474	0.075
Random_0099	chr4,2378533,2379627,+	1095	0.469	0.106
Random_0100	chr17,68269658,68270593,+	936	0.464	0.122
Random_0101	chr22,37126691,37125788,-	904	0.459	0.151
Random_0102	chr14,68808129,68809119,+	991	0.458	0.099
Random_0103	chr6,33671148,33672260,+	1113	0.458	0.041
Random_0104	chr12,130550130,130551050,+	921	0.444	0.088
Random_0105	chr10,26957677,26956756,-	922	0.441	0.013
Random_0106	chr9,87357107,87358231,+	1125	0.439	0.024
Random_0107	chr10,31713679,31712745,-	935	0.436	0.035
Random_0108	chr3,110212732,110213735,+	1004	0.433	0.116
Random_0109	chr15,54382959,54382033,-	927	0.420	0.022
Random_0110	chr15,32030703,32029838,-	866	0.419	0.045
Random_0111	chr3,100083431,100084542,+	1112	0.415	0.040
Random_0112	chr9,114603182,114602236,-	947	0.414	0.051
Random_0113	chr1,153266139,153267294,+	1156	0.412	0.033
Random_0114	chr2,66135844,66136742,+	899	0.409	0.104
Random_0115	chr1,13628440,13627389,-	1052	0.405	0.048
Random_0116	chr10,20765002,20766008,+	1007	0.403	0.066
Random_0117	chr14,16267863,16268812,+	950	0.401	0.076
Random_0118	chr11,36023409,36024376,+	968	0.396	0.056
Random_0119	chr11,106098182,106097050,-	1133	0.394	0.042
Random_0120	chr9,5538611,5539699,+	1089	0.390	0.013
Random_0121	chr18,5774635,5775788,+	1154	0.387	0.161
Random_0122	chr2,50675556,50674390,-	1167	0.384	0.017
Random_0123	chr6,155592633,155593624,+	992	0.382	0.049
Random_0124	chr22,20880391,20881284,+	894	0.380	0.021
Random_0125	chr9,123160657,123161565,+	909	0.378	0.051
Random_0126	chr12,55425523,55424541,-	983	0.370	0.028
Random_0127	chr4,72337174,72336255,-	920	0.367	0.089
Random_0128	chr11,106098182,106097074,-	1109	0.365	0.020
Random_0129	chr4,4844389,48443206,-	1184	0.363	0.035
Random_0130	chr9,114672083,114673059,+	977	0.363	0.028
Random_0131	chr9,123160658,123161566,+	909	0.346	0.011
Random_0132	chr8,81741625,81740527,-	1099	0.343	0.024
Random_0133	chr4,93365632,93364606,-	1027	0.338	0.025
Random_0134	chr12,15639018,15639970,+	953	0.335	0.052
Random_0135	chr12,22189035,22189927,+	893	0.331	0.026
Random_0136	chr13,61917002,61915976,-	1027	0.328	0.036
Random_0137	chr3,126115008,126116017,+	1010	0.326	0.051
Random_0138	chr1,58828326,58827362,-	965	0.321	0.048
Random_0139	chr4,111521043,111520111,-	933	0.320	0.032
Random_0140	chr16,27992657,27993671,+	1015	0.318	0.002
Random_0141	chr8,108936804,108935957,-	848	0.314	0.024
Random_0142	chr3,45338387,45337383,-	1005	0.311	0.044
Random_0143	chr16,29636005,29634831,-	1175	0.311	0.030
Random_0144	chr16,88053439,88052469,-	971	0.310	0.051
Random_0145	chrX,29436543,29437652,+	1110	0.308	0.045
Random_0146	chr12,129530303,129531305,+	1003	0.306	0.033
Random_0147	chrX,65926946,65927824,+	879	0.306	0.026
Random_0148	chr7,143834673,143835481,+	809	0.306	0.032
Random_0149	chr3,12444331,12445416,+	1086	0.305	0.038

表 3.2 (続き): ランダム配列の転写活性化能

Genomic fragment ID	Gene Location	Amplified length	転写活性化能	標準偏差
Random_0150	chr18,2230034,2231080,+	1047	0.304	0.057
Random_0151	chr15,18616521,18617275,+	755	0.300	0.040
Random_0152	chr1,152674897,152675878,+	982	0.291	0.029
Random_0153	chr12,109062645,109063556,+	912	0.291	0.039
Random_0154	chr13,59835160,59834144,-	1017	0.290	0.050
Random_0155	chr15,50888152,50889065,+	914	0.285	0.046
Random_0156	chr8,70215386,70214368,-	1019	0.284	0.033
Random_0157	chr12,11238804,11239629,+	826	0.282	0.038
Random_0158	chr18,39930407,39931236,+	830	0.280	0.086
Random_0159	chr1,200460869,200461837,+	969	0.279	0.038
Random_0160	chr20,49359353,49360328,+	976	0.278	0.020
Random_0161	chr14,68809124,68808120,-	1005	0.270	0.004
Random_0162	chr9,114603112,114602294,-	819	0.267	0.037
Random_0163	chr7,118017457,118018508,+	1052	0.250	0.008
Random_0164	chr8,72163693,72164655,+	963	0.249	0.028
Random_0165	chr14,90549392,90550489,+	1098	0.249	0.033
Random_0166	chr3,141000512,140999588,-	925	0.247	0.035
Random_0167	chr2,49485074,49486218,+	1145	0.245	0.019
Random_0168	chr9,121462084,121463158,+	1075	0.243	0.034
Random_0169	chr2,200936234,200937265,+	1032	0.243	0.019
Random_0170	chr2,173110787,173111922,+	1136	0.238	0.034
Random_0171	chr3,49978242,49979159,+	918	0.238	0.045
Random_0172	chr9,29642003,29642915,+	913	0.235	0.016
Random_0173	chr21,28986441,28987385,+	945	0.235	0.034
Random_0174	chr1,157976003,157977122,+	1120	0.234	0.032
Random_0175	chr2,207414071,207415111,+	1041	0.231	0.010
Random_0176	chr4,2379626,2378531,-	1096	0.227	0.026
Random_0177	chr19,29923445,29924295,+	851	0.226	0.014
Random_0178	chr7,102976086,102977086,+	1001	0.222	0.044
Random_0179	chr11,83963240,83964312,+	1073	0.221	0.021
Random_0180	chr21,32636026,32634971,-	1056	0.214	0.034
Random_0181	chr21,30679357,30680262,+	906	0.214	0.040
Random_0182	chr20,19165117,19166165,+	1049	0.211	0.024
Random_0183	chr1,118094193,118093109,-	1085	0.203	0.016
Random_0184	chr3,157187460,157188324,+	865	0.202	0.010
Random_0185	chr8,72164654,72163692,-	963	0.200	0.004
Random_0186	chr18,39930406,39931235,+	830	0.199	0.018
Random_0187	chr1,33062335,33063236,+	902	0.199	0.013
Random_0188	chr3,191927960,191928999,+	1040	0.199	0.029
Random_0189	chr6,34525191,34526273,+	1083	0.198	0.014
Random_0190	chr13,18704170,18705029,+	860	0.198	0.036
Random_0191	chr6,130816506,130817438,+	933	0.196	0.014
Random_0192	chr15,87401162,87400122,-	1041	0.195	0.028
Random_0193	chr13,36110970,36112020,+	1051	0.192	0.009
Random_0194	chr4,136431757,136432569,+	813	0.191	0.016
Random_0195	chr3,111285533,111284469,-	1065	0.191	0.040
Random_0196	chr15,87400123,87401163,+	1041	0.186	0.027
Random_0197	chr20,51548661,51547794,-	868	0.185	0.025
Random_0198	chr1,25096442,25097461,+	1020	0.184	0.017
Random_0199	chr3,55530689,55531669,+	981	0.184	0.025
Random_0200	chr1,77973824,77974798,+	975	0.183	0.008
Random_0201	chr6,155592632,155593623,+	992	0.179	0.014
Random_0202	chr6,159758884,159759936,+	1053	0.178	0.015
Random_0203	chr21,40008229,40009152,+	924	0.177	0.032
Random_0204	chr16,68475827,68474966,-	862	0.173	0.013
Random_0205	chr1,212049559,212050576,+	1018	0.172	0.020
Random_0206	chr2,5426753,5427908,+	1156	0.166	0.043
Random_0207	chr1,172359536,172360620,+	1085	0.164	0.014
Random_0208	chr17,49214221,49215244,+	1024	0.161	0.015
Random_0209	chr13,105386565,105385435,-	1131	0.158	0.041
Random_0210	chr20,33072349,33073270,+	922	0.155	0.015
Random_0211	chr7,86967814,86966905,-	910	0.153	0.008
Random_0212	chr3,111284477,111285525,+	1049	0.152	0.114
Random_0213	chr5,45671439,45670251,-	1189	0.149	0.015
Random_0214	chr13,103480279,103481359,+	1081	0.149	0.026
Random_0215	chr2,207414072,207415112,+	1041	0.148	0.023
Random_0216	chr16,51367095,51368078,+	984	0.148	0.017
Random_0217	chr13,46529117,46530042,+	926	0.148	0.023
Random_0218	chr15,54565558,54566729,+	1172	0.147	0.024
Random_0219	chr2,141769908,141770943,+	1036	0.146	0.031
Random_0220	chr6,129489324,129490175,+	852	0.145	0.002
Random_0221	chr20,11928613,11929766,+	1154	0.145	0.009
Random_0222	chr13,37885372,37886519,+	1148	0.137	0.011
Random_0223	chr6,67472457,67473454,+	998	0.136	0.023
Random_0224	chr18,29144186,29143002,-	1185	0.134	0.041
Random_0225	chr5,147332192,147333352,+	1161	0.133	0.007

表 3.2 (続き) : ランダム配列の転写活性化能

Genomic fragment ID	Gene Location	Amplified length	転写活性化能	標準偏差
Random_0226	chrY,9560073,9561189,+	1117	0.132	0.021
Random_0227	chrX,65926945,65927824,+	880	0.131	0.021
Random_0228	chr16,6522055,6521060,-	996	0.131	0.007
Random_0229	chr9,89301760,89302767,+	1008	0.131	0.009
Random_0230	chr17,22976294,22977162,+	869	0.130	0.019
Random_0231	chr17,36948122,36949226,+	1105	0.128	0.020
Random_0232	chr10,32319629,32318589,-	1041	0.127	0.008
Random_0233	chr3,45337394,45338388,+	995	0.117	0.018
Random_0234	chr6,3331380,3332517,+	1138	0.117	0.023
Random_0235	chr18,24972366,24973291,+	926	0.116	0.014
Random_0236	chr16,29634832,29636006,+	1175	0.116	0.019
Random_0237	chr3,123462548,123463551,+	1004	0.114	0.020
Random_0238	chr7,34063962,34064997,+	1036	0.109	0.023
Random_0239	chr6,144340326,144341259,+	934	0.107	0.021
Random_0240	chr13,101659429,101660482,+	1054	0.106	0.004
Random_0241	chr20,59920738,59921836,+	1099	0.106	0.033
Random_0242	chr6,5731876,5732697,+	822	0.104	0.014
Random_0243	chr3,28955643,28954674,-	970	0.099	0.005
Random_0244	chr14,89108497,89109455,+	959	0.098	0.017
Random_0245	chr21,32634979,32636025,+	1047	0.094	0.014
Random_0246	chr14,22915173,22916154,+	982	0.088	0.012
Random_0247	chr4,37610540,37611563,+	1024	0.087	0.031
Random_0248	chr7,118570410,118571419,+	1010	0.082	0.023
Random_0249	chr8,104432116,104433308,+	1193	0.082	0.010
Random_0250	chr11,13678929,13677773,-	1157	0.069	0.008
Random_0251	chr21,33700085,33701164,+	1080	0.012	0.001

ランダム領域の転写活性化能(図 3.2)を、その平均値を基準(10^0)としたヒストグラムを(図 3.3)として示した。(図 3.3)から、ランダム領域群が有する転写活性化能の分布のうち、転写活性化能 10^0 以上のランダム領域群が有する転写活性化能の分布が、プロモーター領域群 P2 の転写活性化能の分布と重なっていることが明らかになった。

ランダム領域について、プロモーター領域群 P2 と転写活性化能の分布が重なるもの(転写活性化能 10^0 以上のもの)を R1(ランダム領域群全体おける 18%)、残りを R2(ランダム領域群全体おける 82%)とした(表 3.4)。

表 3.4 : 各領域群の転写活性化能とクローン数

転写活性化能	Promoter	P1	P2	random	R1	R2
$>10^3$	2	2	0	0	0	0
10^3-10^2	175	175	0	0	0	0
10^2-10^1	217	217	0	3	3	0
10^1-10^0	63	17	46	43	43	0
10^0-10^{-1}	15	0	15	196	0	196
$<10^{-1}$	0	0	0	9	0	9
Total	472	411	61	251	46	205

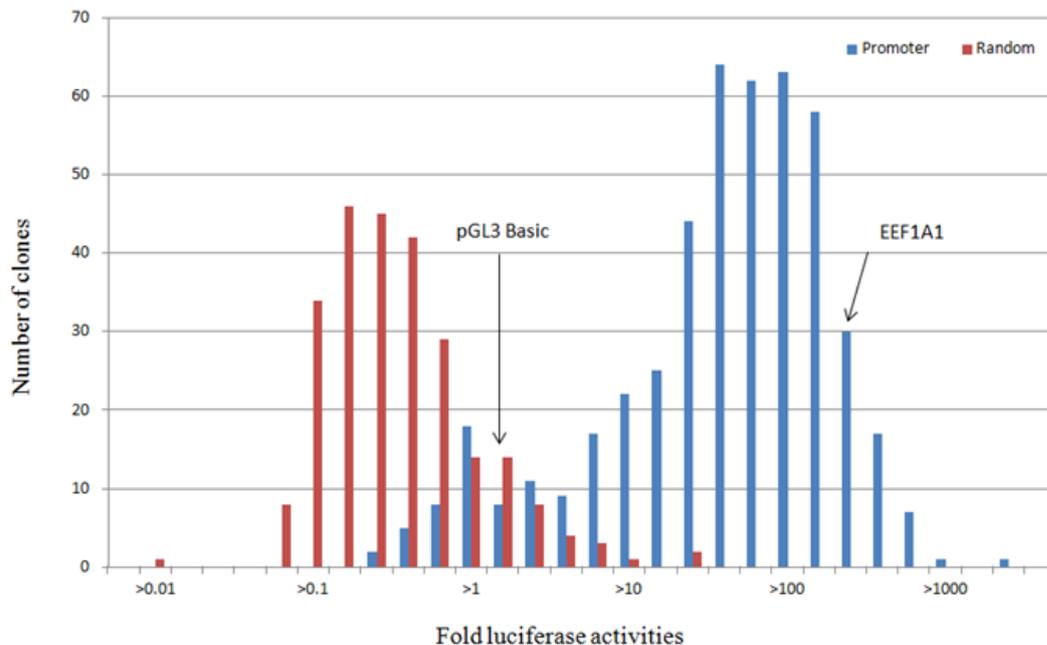


図 3.3 : 遺伝子のプロモーター配列とランダム配列の転写活性化能の分布

転写活性化能を測定できた、472 遺伝子のプロモーター領域 (青)、251 種類のランダム領域(赤)が有する転写活性化能をヒストグラムとして示した。このとき、ランダム領域の転写活性化能の平均値を 1 とした。また、コントロールとして、強い転写活性化能を有することが知られる EEF1A1 および空ベクターである pGL3 basic vector についてそれぞれ矢印で示す。

3.1.2 遺伝子のプロモーター配列の転写活性化能の確認

弱い転写活性化能を有したプロモーター配列群 P2 が、誤って抽出されたプロモーター配列であるため、強い転写活性化能を示していない可能性について検討した。既報の論文で解析されたプロモーター領域および完全長 cDNA が少なくとも 3 クローン以上対応するプロモーター領域 445 種について、転写活性化能の分布および配列解析を行った(図 3.4、表 3.5、3.6)。その結果は全体のプロモーター領域を用いて行ったものと同様であった(図 3. 表 3.5、3.6)。

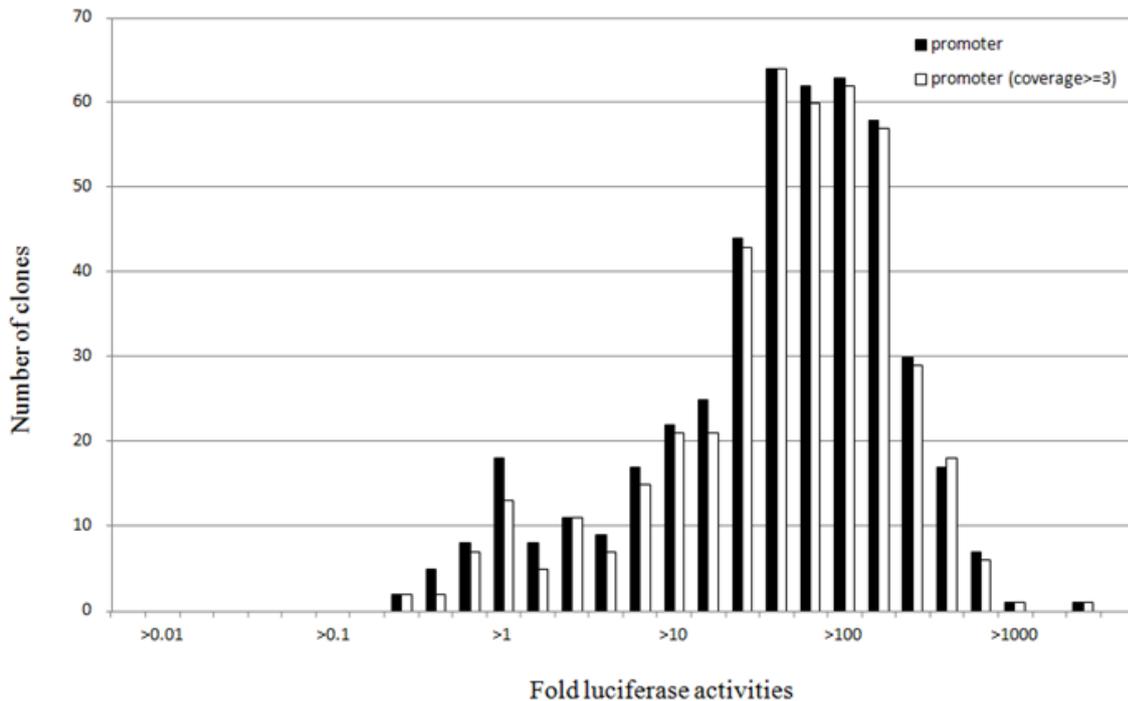


図 3.4: ヒト完全長 cDNA1 クローンおよび 3 クローンにサポートされた遺伝子のプロモーター領域における転写活性化能の分布

少なくとも 1 クローンのヒト完全長 cDNA クローンにサポートされる 472 遺伝子のプロモーター領域 (黒)、少なくとも 3 クローンのヒト完全長 cDNA クローンにサポートされる 445 遺伝子のプロモーター領域 (白) が有する転写活性化能をヒストグラムとして示した。このとき、ランダム領域の転写活性化能の平均値を 1 とした。

表 3.5: 少なくとも 3 クローン以上のヒト完全長 cDNA にサポートされる遺伝子のプロモーター

領域群 P1 および P2 が有する転写活性化能の分布における統計的な差異

(P1+P2) プロモーター領域群 P1、P2 を含めた転写活性化能の分布がひとつの正規分布に所属するか Kolmogorov-Smirnov 検定および chi-square 検定による p 値を示した。(P1) (P2) プロモーター領域群 P1 および P2 が有する転写活性化能の分布がそれぞれひとつの正規分布に所属するか Kolmogorov-Smirnov 検定および chi-square 検定による p 値を示した

	K. S. test	chi-sq test
P1+P2	2.2E-04	4.4E-16
P1	0.88	0.76
P2	0.37	0.42

*Promoters with >3 supporting oligo-cap cDNAs were used

表 3.6 : 少なくとも 3 クローン以上のヒト完全長 cDNA にサポートされる遺伝子のプロモーター領域群が有する配列上の特徴

	P	P1	P2
CpG island	262 (59%)	254 (64%) ^{*2}	8 (17%) ^{*2}
TATA box: strict	29 (7%)	28 (7%)	1 (2%)
TATA box: less strict	94 (21%)	79 (20%) ^{*3}	15 (32%) ^{*3}
Average G+C content	53%	55% ^{*1}	48% ¹
TOTAL	445	398	47

*Promoters with >3 supporting oligo-cap cDNAs were used

^{*1} p<2.0e-7 ^{*2} p<7.0e-5 ^{*3} p<8.0e-2

3.1.3 TATA box の探索、GC 含有率、CpG island の探索

遺伝子のプロモーター領域群全体とプロモーター領域群 P1、P2 のそれぞれに対し、遺伝子のプロモーター領域に存在するエレメントである CpG island および TATA box を有するプロモーター領域の存在率と、GC 含有率を計算した。結果を表 3.7 に示す。TATA box の検索については strict TATA box(TATA[T/A] [T/A]配列の完全マッチを求めた厳しい検索条件)と less-strict TATAbox(マトリックス検索を用いた緩やかな検索条件)の 2 種類の方法で行った。

遺伝子のプロモーター領域群全体に対する配列解析の結果では、解析対象となった遺伝子のプロモーター領域群の GC 含有率は 53%であった。ヒトゲノム配列の GC 含有率 (41%)より高い値を示していた。これは先行研究のプロモーター領域に対する解析結果と一致していた[10]。プロモーター領域群 P1 および P2 について比較したところ、GC 含有率は P1 が 54%、P2 が 47%と P1 のほうが高かった(p<6.0e-10)。CpG island を有するプロモーター領域の存在率は、P1 が 63%、P2 が 13%と CpG island を有するプロモーター領域も P1 が多かった (p<1.0e-6)。TATA box を有するプロモーター領域の存在率は、strict TATA box は P1 および P2 共に 7%と同等であったが、less-strict TATA box を有するプロモーター領域は P1 が 20%、P2 が 36%と P2 に多かった(p<1.0e-2)。二峰性の転写活性化能の分布様式を形成するプロモーター領域群 P1 および P2 の間には、配列上の特徴にも差異が存在していることが明らかとなった。

ランダム領域群全体、ランダム領域群 R1 と R2 のそれぞれに対し、同様に、CpG island および TATA box を有するランダム領域の存在率、GC 含有率について配列解析を行った(表 3.7)。ランダム領域群の GC 含有率は 45%とヒトゲノム配列全体における GC 含有率 (41%)と類似していた。CpG island を有しているランダム領域群は存在しなかった。ランダム領域群 R1 と R2 を比較したところ、strict TATA box を有するランダム領域の存在率は R1 が 24%、R2 が 16%、less-strict TATA box を有するランダム領域の存在率は R1 が 70%、

R2 が 51% と、TATA box を有するランダム領域が R1 に多かった。GC 含有率および CpG island を有するランダム領域の存在率については、R1 と R2 の間に差異はみられなかった (%)。

表 3.7 : 遺伝子のプロモーター領域及びランダム領域の配列上の特徴

	promoter	P1	P2	Random	R1	R2
CpG island	267 (57%)	259 (63%) ^{*1}	8 (13%) ^{*1}	0 (0%)	0 (0%)	0 (0%)
TATA box: strict	34 (7%)	30 (7%)	4 (7%)	47 (19%)	11 (24%)	32 (16%)
TATA box: less strict	103 (22%)	81 (20%) ^{*2}	22 (36%) ^{*2}	140 (56%)	32 (70%)	104 (51%)
Average G+C content:	53%	54% ^{*3}	47% ^{*3}	45%	44%	45%
クローン数	472	411	61	251	46	205

^{*1}p < 1.0e-6 ^{*2}p < 1.0e-2 ^{*3}p < 6.0e-10

3.1.4 転写因子結合配列の探索

遺伝子のプロモーター領域群全体とプロモーター領域群 P1、P2 及びランダム領域群全体とランダム領域群 R1、R2 のそれぞれに、転写因子結合配列候補となる既知モチーフの検索を行った。モチーフの検索には、転写因子結合配列ソフトウェア MATCH[48]を利用した。転写因子のデータベースとして TRANSFAC 2008.3 の vertebrate_non_redundant_minFP.prf を用いた。220 種類の position weight matrix (PWM) で構成され、False Positive を最小にするパラメーターセットである[49]。

遺伝子のプロモーター領域では平均 27.1 個 (P1: 27.4 個、P2: 24.9 個)、ランダム領域全体では 22.0 個 (R1: 21.1 個、R2: 22.2 個) の既知モチーフが検出された。転写因子結合配列が P1 に濃縮されている可能性が示唆された。

遺伝子のプロモーター領域群 P1、P2 間又はランダム領域群 R1、R2 間で存在割合に差が検出されたモチーフの一覧を示した (表 3.8)。モチーフの存在割合の差は Fisher 正確確率検定の p 値 0.05 未満を指標とした。P1、P2 の比較で、P1 に濃縮されているモチーフが 10 個、P2 に濃縮されているものが 6 個検出された。R1、R2 の比較で、R1 に濃縮されているモチーフが 2 個検出された。R2 に濃縮されたものはなかった。

P1、P2、R1 のそれぞれに濃縮された各モチーフについて、モチーフがマッピングされた配列の GC 含有量を算出した (表 3.8)。各群に濃縮されたモチーフがマッピングされた配列の GC 含有量は、平均して P1: 68.7%、P2: 36.5%、R1: 41.8% であった。転写活性化能の強いプロモーター領域群 P1 には GC 含量の高い転写因子結合配列が濃縮されている可能性が考えられた。

表 3.8 : 遺伝子のプロモーター領域及びランダム領域と既知モチーフ

遺伝子のプロモーター領域及びランダム領域に対し既知モチーフを検索し、P1、P2 間又は R1、R2 間で既知モチーフの存在割合に差が検出されたモチーフを一覧として示した。既知モチーフの存在割合の差は Fisher 正確確率検定の p 値=0.05 を閾値として判断した。既知モチーフが検索された配列の平均 GC 含量 (%) も算出した。

モチーフ名	Promoter(n=472)		P1 (n=411)		P2 (n=61)		P1-P2 p-value	Random (n=251)		R1 (n=46)		R2 (n=205)		R1-R2 p-value	平均GC 含量(%)
	度数	%	度数	%	度数	%		度数	%	度数	%	度数	%		
V\$ZF5_B	277	58.7	263	64.0	14	23.0	1.E-09	33	13.1	6	13.0	27	13.2	1.E+00	75.6
V\$PAX5_02	195	41.3	183	44.5	12	19.7	2.E-04	20	8.0	2	4.3	18	8.8	5.E-01	64.1
V\$CETS1P54_03	208	44.1	194	47.2	14	23.0	3.E-04	25	10.0	4	8.7	21	10.2	1.E+00	61.7
V\$KID3_01	426	90.3	377	91.7	49	80.3	8.E-03	183	72.9	32	69.6	151	73.7	6.E-01	80.0
V\$RFX1_02	66	14.0	64	15.6	2	3.3	9.E-03	18	7.2	3	6.5	15	7.3	1.E+00	55.7
V\$VMYB_02	171	36.2	158	38.4	13	21.3	1.E-02	26	10.4	6	13.0	20	9.8	6.E-01	56.0
V\$AP2_Q6	83	17.6	79	19.2	4	6.6	1.E-02	6	2.4	1	2.2	5	2.4	1.E+00	87.2
V\$E2F_03	34	7.2	34	8.3	0	0.0	1.E-02	1	0.4	0	0.0	1	0.5	1.E+00	64.3
V\$CACD_01	94	19.9	89	21.7	5	8.2	2.E-02	37	14.7	8	17.4	29	14.1	6.E-01	82.8
V\$RFX_Q6	114	24.2	106	25.8	8	13.1	4.E-02	30	12.0	5	10.9	25	12.2	1.E+00	59.7
V\$HNF4ALPHA_Q	12	2.5	6	1.5	6	9.8	2.E-03	6	2.4	2	4.3	4	2.0	3.E-01	42.1
V\$FOXJ2_02	69	14.6	53	12.9	16	26.2	1.E-02	43	17.1	8	17.4	35	17.1	1.E+00	18.1
V\$POU3F2_01	76	16.1	59	14.4	17	27.9	1.E-02	76	30.3	16	34.8	60	29.3	5.E-01	21.4
V\$TTF1_Q6	55	11.7	42	10.2	13	21.3	2.E-02	35	13.9	3	6.5	32	15.6	2.E-01	55.9
V\$TST1_01	158	33.5	129	31.4	29	47.5	2.E-02	58	23.1	12	26.1	46	22.4	6.E-01	29.4
V\$SMAD3_Q6	16	3.4	11	2.7	5	8.2	4.E-02	20	8.0	1	2.2	19	9.3	1.E-01	52.3
V\$YY1_Q6	88	18.6	78	19.0	10	16.4	7.E-01	14	5.6	6	13.0	8	3.9	3.E-02	48.2
V\$MRF2_01	15	3.2	14	3.4	1	1.6	7.E-01	12	4.8	5	10.9	7	3.4	5.E-02	35.5

*p-value: Fisher 正確確率検定

3.2 プロモーター配列の塩基レベルの差が転写活性化能へ与える影響の解析

3.1 の体系的なルシフェラーゼアッセイによる測定系により、数百の規模で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。しかし、構築した測定系で得られたデータでは、プロモーター配列中約 1000 塩基について、どの塩基がその全体の差異にどのように影響しているのかを解析するには不十分であった。より高精細に DNA 配列としてのプロモーター配列が転写活性化能へ与える影響を解析するためには、プロモーター配列内のそれぞれの塩基について 1 塩基単位の DNA 配列の差異が全体として転写活性化能へ与える影響を測定する実験系を構築する必要がある

3.2.1 効率的な変異プロモーター配列の取得、配列決定、転写活性化能の測定系の構築

本研究では、エラーの入りやすい PCR 条件 (error prone PCR) を用いた効率的な変異プロモーター配列の取得、次世代シーケンサーを応用した変異プロモーターの配列決定及び転写活性化能の測定の一連の実験系を構築した。(図 3.5) 以下の項において、構築した各工程の詳細について示す。一連の実験系の構築及びその後の解析には、8 遺伝子のプロモーター配列を用いた (表 3.9)。これらの遺伝子は、前章の体系的なルシフェラーゼアッセイにおいて転写活性化能が高値を示す遺伝子より選択した。TATA box の有無、GC 含有量、CpG island 有無の配列上の特徴についても、それぞれの群に典型的なものを選択した。

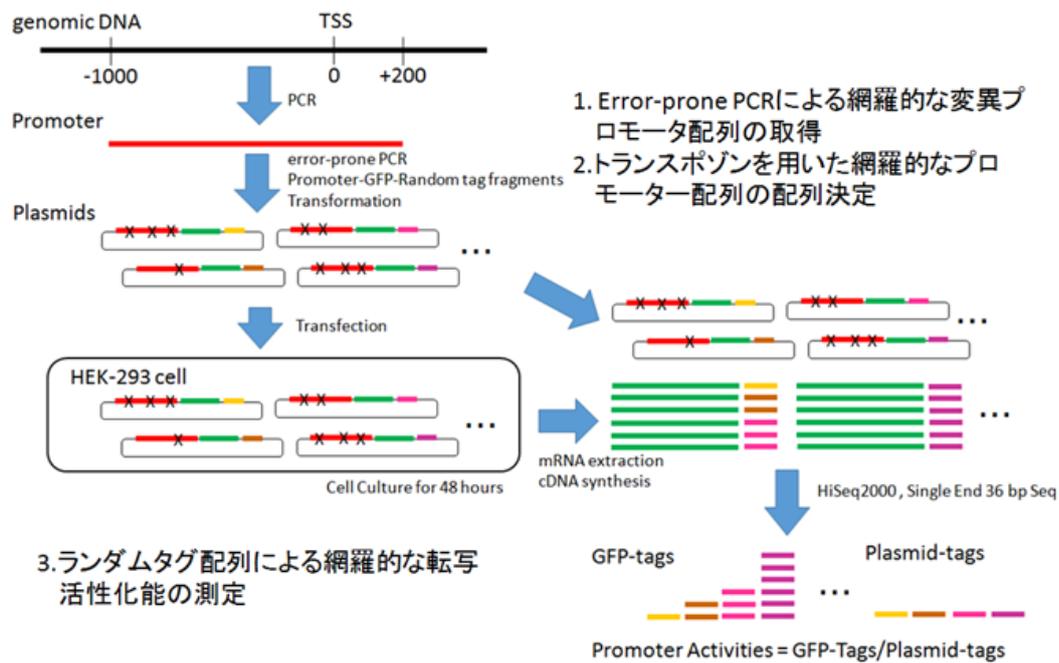


図 3.5 : 多様な変異プロモーターの取得、配列決定、転写活性化能の測定の全体像

本研究で構築した実験系の全体像を示した。error prone PCR による効率的な変異プロモーター配列の取得、次世代シーケンサーを利用した変異プロモーターの配列決定及び転写活性化能の測定の一連の実験系の概要。

表 3.9 : 本研究に用いた 8 遺伝子のプロモーター配列

NM ID	GeneName	ChrNo	Direction	Start	End	Promoter Length	log10(Luc)	TATA box	CpG island	GC content
NM_006407	ARL6IP5	chr3	+	69133098	69134224	1127	1.26	-	+	0.44
NM_004396	DDX5	chr17	-	62503385	62502275	1111	2.75	+	+	0.63
NM_001402	EEF1A1	chr6	-	74231684	74230582	1103	2.52	+	+	0.59
NM_002046	GAPDH	chr12	+	6642712	6643769	1058	1.83	+	+	0.65
NM_006183	NTS	chr12	+	86267117	86268204	1088	0.39	+	-	0.38
NM_000291	PGK1	chrX	+	77358751	77359829	1079	2.25	-	+	0.57
NM_005057	RBBP5	chr1	-	205092071	205091026	1046	2.81	-	-	0.46
NM_001016	RPS12	chr6	+	133134776	133135901	1126	3.19	-	+	0.55

3.2.1.1 変異プロモーター配列の取得

本研究にて構築した、error prone PCR およびランダムタグ配列を指標として次世代シーケンシング技術を駆使する、ハイスループットな変異プロモーター配列解析方法の概略を図 3.6 に示す。

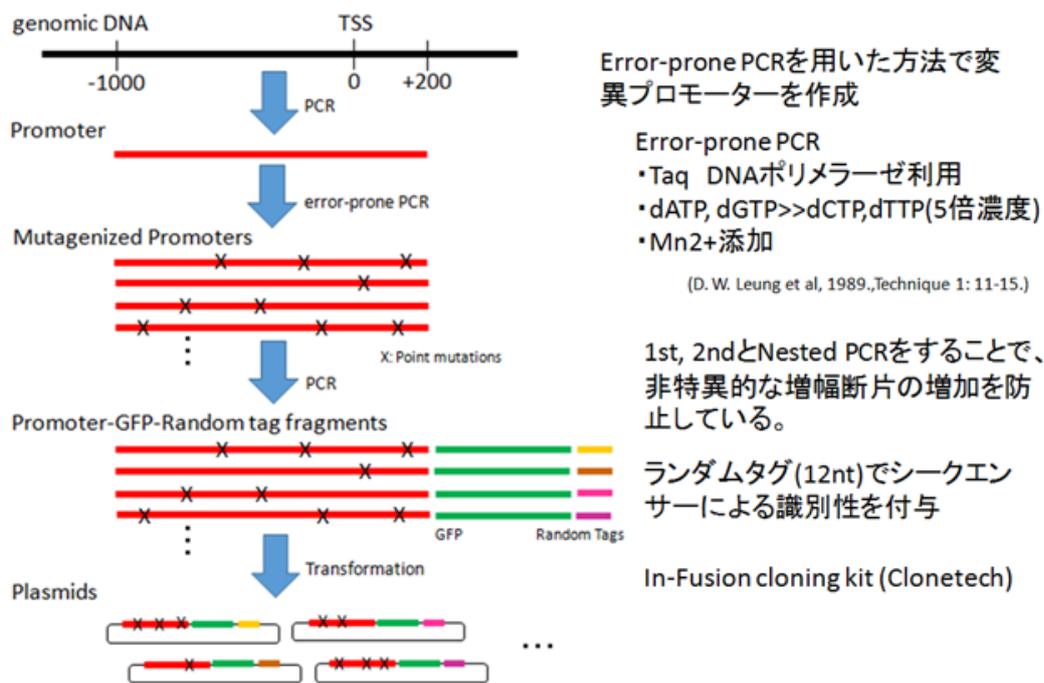


図 3.6 : 多様な変異プロモーター配列の効率的な取得方法

野生型プロモーター配列に、error prone PCR を用いてランダムに変異を導入した。12 塩基のランダムタグ配列を連結することで個々の断片に識別性を与えた。変異プロモーターは配列には DNA 融合に用いる 15 塩基の DNA 配列を挿入し、In-Fusion クローニングにて、プラスミド DNA へ挿入した。

変異プロモーター配列の取得のために、野生型プロモーター配列をテンプレートとし、error prone PCR を行った。PCR は、error prone PCR 条件として、Taq ポリメラーゼを利用し、不均衡な dNTP 濃度 (2mM dCTP 及び dTTP, 10mM dGTP 及び dATP)、50mM MgCl₂ 及び 5mM MnCl₂ を用いて、95°C 15 秒、55°C 15 秒、72°C 2 分で 40 サイクルの反応を行った。非特異的な DNA 断片の増幅を防ぐため、PCR による増幅は 2 段階で実施した。利用するテンプレート量を最小限にする目的で、テンプレートとする DNA 断片を 1/4 希釈ずつ 7 段階の希釈を行い、8 種類の濃度で PCR を実施した。PCR 後、電気泳動を行い、最もバンドの薄い増幅産物を選択した。

作成した変異プロモーターに識別性を付与するため、12 塩基のランダムタグ配列を作成した。変異プロモーターと Overlap extension PCR 法にて連結して pGL3 Basic ベクターに挿入した。転写活性化能の測定時には GFP 遺伝子を用いることとした。変異プロモーター配列とランダムタグ配列の間に GFP 遺伝子を挿入した形で、pGL3 Basic ベクターへと導入した。作成した変異プロモーター配列-GFP 遺伝子-ランダムタグ配列の DNA 断片の 5' 端及び 3' 端に In-Fusion クローニングに用いる 15 塩基の配列を付与し、In-Fusion クローニングシステムを用いて、変異プロモーターを一括してクローニングできるようにした。作成したプラスミド DNA ライブラリーで、NEB 5-alpha Competent E. coli を形質転換した。LB 培地で培養する際、形質転換後のサンプル 4400µL から 20µL を LB 寒天培地で一晚培養しコロニー数が 45 を超えるこ

とを確認した。これによりプラスミド DNA ライブラリーの多様性が 10000 種類を超えることを保証した

3.2.1.2 変異プロモーター配列の配列決定

図 3.7 にプロモータークローンの塩基配列決定のスキームを示す。トランスポゾン配列とランダムタグ配列を利用した 3'端側及び 5'端側双方のシーケンス用ライブラリーを用い、多様な変異プロモーター配列について、一括で配列決定を行った。変異プロモーター配列にランダムにトランスポゾン配列を挿入し、PCR でトランスポゾン配列とランダムタグ配列を保持した変異プロモーター配列の 3'端側、5'端側の多様な deletion mutant を作成した。作成した deletion mutant をシーケンスライブラリーとして HiSeq2000 にてシーケンシング、野生型プロモーター配列へマッピングすることで、変異塩基の同定を行った。

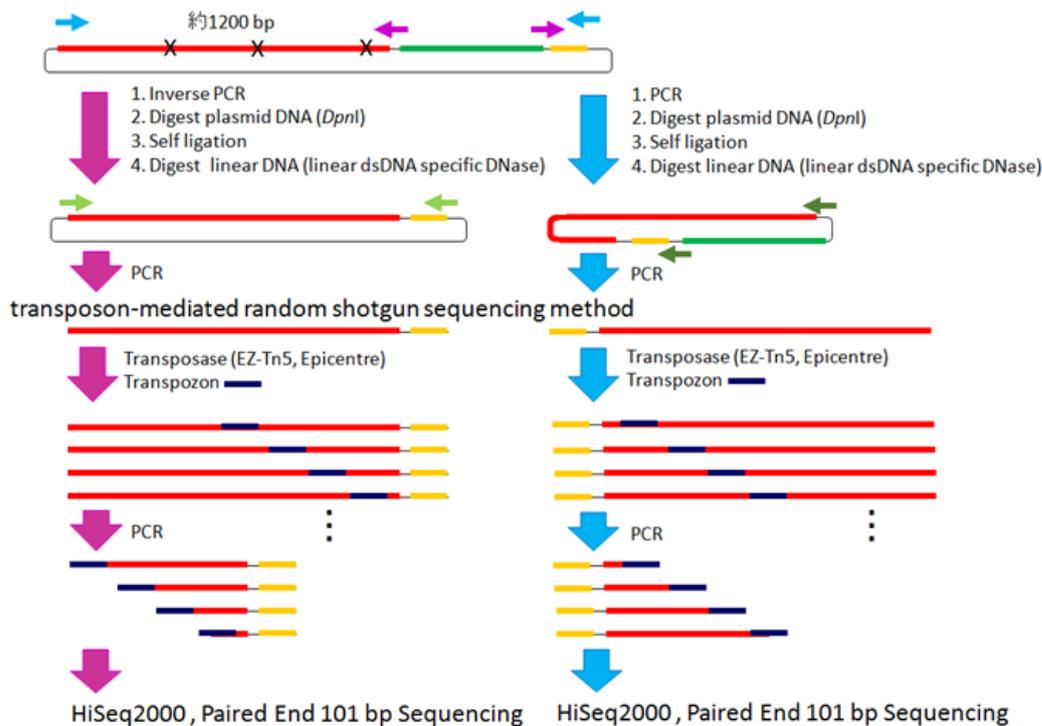


図 3.7 : 多様な変異プロモーター配列の一括した配列決定

本研究で構築した変異プロモーター配列決定方法の全体像を示す。変異プロモーター配列にランダムにトランスポゾン配列を挿入し、PCR にて効率的にトランスポゾン配列とランダムタグ配列を保持した変異プロモーター配列 3'端側、5'端側の多様な deletion mutant を作成した。作成した多様な deletion mutant をシーケンシングし、野生型プロモーター配列へマッピングすることで、野生型プロモーターからの変異の検出を行った。

変異プロモーター配列の 3'端側及び 5'端側シーケンスライブラリーのシーケンス結果を表 3.10 に示した。解析に用いる上で十分量かつ高品質のリードが得られていることを確認した。

表 3.10 変異プロモーター配列の3'端側及び5'端側ライブラリーのシーケンスサマリー

遺伝子	ライブラリー	Yield (Mbases)	% PF	# Reads	% of raw clusters per lane	% of >= Q30 Bases (PF)	Mean Quality Score (PF) Max =40
EEF1A1	3'端側	14409	93.28	152,933,498	36.06	91.34	34.84
EEF1A1	5'端側	7471	93.00	79,534,614	18.75	85.49	33.21
GAPDH	3'端側	14285	90.82	155,731,370	31.29	84.50	33.14
GAPDH	5'端側	7829	90.78	85,388,600	17.16	85.71	33.36
RPS12	3'端側	9,109	93.74	96,214,584	21.95	87.22	34.23
RPS12	5'端側	8,796	93.20	93,438,094	21.32	90.10	35.01
PGK1	3'端側	11,080	93.19	117,722,910	26.05	87.39	34.26
PGK1	5'端側	8,918	93.18	94,757,298	20.97	89.66	34.95
DDX5	3'端側	10,894	85.00	126,892,972	27.92	83.72	33.52
DDX5	5'端側	6,624	92.31	71,046,288	15.63	79.10	32.56
ARL6IP5	3'端側	12,265	89.52	137,005,190	31.29	88.74	34.76
ARL6IP5	5'端側	7,395	94.51	78,241,882	17.87	90.14	35.20
NTS	3'端側	8,277	93.57	88,460,050	22.35	92.17	35.67
NTS	5'端側	7,562	96.11	78,680,214	19.88	92.33	35.93
RBBP5	3'端側	11,635	89.39	130,160,264	28.24	87.23	34.35
RBBP5	5'端側	20,845	96.20	216,680,594	57.42	81.87	33.50

本研究で確立した測定系では、1度の実験で平均 22538 種類の変異プロモーターについて配列決定及び転写活性化能データの取得が可能であった。これは 1000 変数の重回帰分析で偏相関係数 0.05 の結果を、有意水準 5% で検出する際、検出力が 99% 以上となるデータ量に相当した。配列決定及び転写活性化能を測定できた変異プロモーターは平均して 1.6 ~ 2.0% の変異を有していた (表 3.11)。プロモーター配列内の各塩基レベルでの変異は平均 387 のデータが得られた (表 3.12)。配列が決定され転写活性化能が測定 (詳細は以下に示す) された変異プロモーター配列について、遺伝子ごとにそのプロファイルを表 3.11 に示す。プロモーター配列の各塩基に挿入された変異の平均頻度を表 3.12 に示した。

表 3.11 : 各遺伝子の変異プロモーター配列のプロファイル

	ARL6IP5	DDX5	EEF1A1	GAPDH	NTS	PGK1	RBBP5	RPS12
promoter_length	1127	1111	1103	1058	1088	1079	1046	1126
GC_content	0.44	0.63	0.59	0.65	0.38	0.57	0.46	0.55
# of mutatnt promoters	20486	17139	20575	24335	28833	21902	23809	22366
total_mutation	376678	354770	396252	436298	534168	471617	404006	516061
mutation rate	1.60%	1.90%	1.70%	1.70%	1.70%	1.90%	1.60%	2.00%
A>T, T>A	99,046(26.3%)	51,389 (14.5%)	79,802 (20.1%)	84,800 (19.4%)	127,501(23.9%)	97,723 (20.7%)	99,469(24.6%)	78,888 (15.3%)
A>G, T>C	75,350(20.9%)	64,270 (18.1%)	64,674 (16.3%)	64,255 (14.7%)	147,400(27.6%)	74,374 (15.8%)	104,005(25.7%)	93,477 (18.1%)
A>C, T>G	71,677(19.0%)	45,846 (12.9%)	65,085 (16.4%)	68,881 (15.8%)	109,917(20.6%)	74,004 (15.7%)	100,963(25.0%)	71,160 (13.8%)
G>A, C>T	47,038(12.5%)	99,419 (28.0%)	80,943 (20.4%)	99,875 (22.9%)	55,736(10.4%)	119,436 (25.3%)	39,224(9.7%)	120,748 (23.4%)
G>T, C>A	24,522(6.5%)	52,129 (14.7%)	54,799 (13.8%)	50,617 (11.6%)	23,930(4.5%)	53,631 (11.4%)	21,809(5.2%)	67,705 (13.1%)
G>C, C>G	15,920(4.2%)	20,520 (5.8%)	21,004 (5.3%)	24,252 (5.6%)	16,125(3.0%)	25444 (5.4%)	11,052(2.7%)	61,124 (11.8%)
deletion	24,387(6.5%)	20,510 (5.8%)	29,189 (7.4%)	37,625 (8.6%)	34,247(6.4%)	26245 (5.6%)	16,383(4.1%)	22,136 (4.3%)
insertion	18,738(5.0%)	687 (0.2%)	756 (0.2%)	5993 (1.4%)	19,312(3.6%)	760 (0.2%)	11,821(2.9%)	823 (0.2%)

表 3.12 : プロモーター配列の各塩基に挿入された変異の平均頻度

	ARL6IP5	DDX5	EEF1A1	GAPDH	NTS	PGK1	RBBP5	RPS12
promoter_length	1127	1111	1103	1058	1088	1079	1046	1126
GC_content	0.44	0.63	0.59	0.65	0.38	0.57	0.46	0.55
# of mutatnt promoters	20486	17139	20575	24335	28833	21902	23809	22366
mutation	334.2(1.632 %)	314.8(1.863 %)	351.6(1.746 %)	387.1(1.695 %)	474(1.703 %)	418.5(1.996 %)	358.5(1.622 %)	457.9(2.049 %)
A>T, T>A	87.9(0.429 %)	45.6(0.223 %)	70.8(0.346 %)	75.2(0.367 %)	113.1(0.552 %)	86.7(0.423 %)	88.3(0.431 %)	70.0(0.342 %)
A>G, T>C	66.9(0.326 %)	57(0.278 %)	57.4(0.28 %)	57(0.278 %)	130.8(0.638 %)	66.0(0.322 %)	92.3(0.450 %)	82.9(0.405 %)
A>C, T>G	63.6(0.310 %)	40.7(0.199 %)	57.8(0.282 %)	61.1(0.298 %)	97.5(0.476 %)	65.7(0.321 %)	89.6(0.437 %)	63.1(0.308 %)
G>A, C>T	41.7(0.204 %)	88.2(0.431 %)	71.8(0.351 %)	88.6(0.433 %)	49.5(0.241 %)	106.0(0.517 %)	34.8(0.170 %)	107.1(0.523 %)
G>T, C>A	21.8(0.106 %)	46.3(0.226 %)	48.6(0.237 %)	44.9(0.219 %)	21.2(0.104 %)	47.6(0.232 %)	19.4(0.094 %)	60.1(0.293 %)
G>C, C>G	14.1(0.069 %)	18.2(0.089 %)	18.6(0.091 %)	21.5(0.105 %)	14.3(0.070 %)	22.6(0.110 %)	9.8(0.048 %)	54.2(0.265 %)
deletion	21.6(0.106 %)	18.2(0.089 %)	25.9(0.126 %)	33.4(0.163 %)	30.4(0.148 %)	23.3(0.114 %)	14.5(0.071 %)	19.6(0.096 %)
insertion	16.6(0.081 %)	0.6(0.003 %)	0.7(0.003 %)	5.3(0.026 %)	17.1(0.084 %)	0.7(0.003 %)	10.5(0.051 %)	0.7(0.004 %)

1塩基あたりの変異データが10以上又は50以上データの観測された塩基を表3.13に示す。各遺伝子のプロモーター領域において、平均99.4%の塩基で変異プロモーターが10以上、平均91.8%の塩基で変異プロモーターが50以上対応していた。各塩基について、10以上の変異プロモーターが対応したものの割合は、A:86.2%、T:85.1%、G:72.6%、C:69.4%、deletion:24.7%であった(表3.13)。プロモーター領域の約70%以上の領域で、野生型の塩基から他の種類の塩基へ置換したプロモーター活性のデータが取得された。deletionについては、24.7%と限られた領域での情報集積であった。

表 3.13-1: 変異データが10以上集積された塩基数と割合

各遺伝子のプロモーター配列中に野生型の配列から、変異(mutation)又はそれぞれの変異パターン(toA:Aに変異、toT:Tに変異、toG:Gに変異、toC:Cに変異、del:delに変異)に10以上データが集積された塩基数とその割合を示した。

GeneName	length	Mutaton(>=10)		ToA(>=10)		toT(>=10)		toG(>=10)		toC(>=10)		del(>=10)	
		度数	%	度数	%	度数	%	度数	%	度数	%	度数	%
ARL6IP5	1127	1115	98.9	630	79.0	671	80.9	652	72.2	547	64.2	242	21.5
DDX5	1111	1101	99.1	768	86.0	776	84.3	504	69.0	543	68.7	230	20.7
EEF1A1	1103	1098	99.5	758	87.8	766	85.8	556	72.9	524	66.3	288	26.1
GAPDH	1058	1057	99.9	790	90.4	763	87.6	550	73.6	512	75.1	320	30.2
NTS	1088	1088	100.0	632	88.5	692	88.3	670	78.1	658	72.5	297	27.3
PGK1	1079	1076	99.7	780	90.2	711	85.6	551	72.3	571	73.3	267	24.7
RBBP5	1046	1027	98.2	612	78.6	607	80.8	547	68.0	556	69.2	242	23.1
RPS12	1126	1123	99.7	751	88.9	794	87.8	589	75.0	557	66.0	268	23.8

表 3.13-2: 変異データが 50 以上集積された塩基数と割合

各遺伝子のプロモーター配列中に野生型の配列から、変異 (mutation) 又はそれぞれの変異パターン (toA : A に変異, toT : T に変異, toG : G に変異, toC : C に変異, del : del に変異) に 50 以上データが集積された塩基数とその割合を示した。

GeneName	length	Mutaton(>=50)		ToA(>=50)		toT(>=50)		toG(>=50)		toC(>=50)		del(>=50)	
		度数	%	度数	%	度数	%	度数	%	度数	%	度数	%
ARL6IP5	1127	984	87.3	366	45.9	416	50.2	364	40.3	343	40.3	85	7.5
DDX5	1111	997	89.7	484	54.2	477	51.8	283	38.8	266	33.7	67	6.0
EEF1A1	1103	1009	91.5	515	59.7	499	55.9	315	41.3	279	35.3	97	8.8
GAPDH	1058	998	94.3	540	61.8	559	64.2	291	39.0	277	40.6	107	10.1
NTS	1088	1053	96.8	455	63.7	478	61.0	453	52.8	397	43.7	108	9.9
PGK1	1079	1003	93.0	522	60.3	470	56.6	309	40.6	355	45.6	82	7.6
RBBP5	1046	911	87.1	397	51.0	382	50.9	304	37.8	300	37.4	64	6.1
RPS12	1126	1063	94.4	502	59.4	502	55.5	359	45.7	298	35.3	77	6.8

表 3.14 に、導入された変異が 50 又は 10 未満であった塩基の特徴を示す。変異の導入が 50 未満であった塩基は 90.5 % が GC であった。これは taq polymerase を用いた error prone PCR が GC に対する変異導入が低効率であるとする先行研究の結果[51] と一致していた。本研究では、プロモーター配列中の平均 99.4 % の塩基で変異が 10 以上、平均 91.8 % の塩基で変異が 50 以上観測され、十分量のデータが蓄積されたものと考えているが、より効率的にデータを集積するためには塩基によって変異を導入する割合の偏りの少ない手法を用いる必要があると考えられた[52]。

表 3.14: 変異データの集積が 50 又は 10 未満であった塩基の特徴

GeneName	塩基	<50		<10		GeneName	塩基	<50		<10			
		度数	%	度数	%			度数	%	度数	%		
ARL6IP5	A	1	20	14.0	1	8.3	NTS	A	1	2	5.7	0	0.0
ARL6IP5	T	2	13	9.1	0	0.0	NTS	T	2	1	2.9	0	0.0
ARL6IP5	G	3	60	42.0	8	66.7	NTS	G	3	17	48.6	0	0.0
ARL6IP5	C	4	50	35.0	3	25.0	NTS	C	4	15	42.9	0	0.0
DDX5	A	1	5	4.4	1	10.0	PGK1	A	1	4	5.3	0	0.0
DDX5	T	2	0	0.0	0	0.0	PGK1	T	2	3	3.9	0	0.0
DDX5	G	3	59	51.8	5	50.0	PGK1	G	3	38	50.0	0	100.0
DDX5	C	4	50	43.9	4	40.0	PGK1	C	4	31	40.8	0	0.0
EEF1A1	A	1	0	0.0	0	0.0	RBBP5	A	1	6	4.4	2	10.5
EEF1A1	T	2	2	2.1	0	0.0	RBBP5	T	2	22	16.3	4	21.1
EEF1A1	G	3	54	57.4	2	40.0	RBBP5	G	3	52	38.5	5	26.3
EEF1A1	C	4	38	40.4	3	60.0	RBBP5	C	4	55	40.7	8	42.1
GAPDH	A	1	0	0.0	0	0.0	RPS12	A	1	2	3.2	2	66.7
GAPDH	T	2	0	0.0	0	0.0	RPS12	T	2	3	4.8	0	0.0
GAPDH	G	3	30	50.0	0	0.0	RPS12	G	3	42	66.7	1	33.3
GAPDH	C	4	30	50.0	1	100.0	RPS12	C	4	16	25.4	0	0.0

3.2.1.3 変異プロモーター配列の転写活性化能の測定

次世代シーケンサーを駆使したランダムタグ配列カウントによる高効率レポーターアッセイ系のスキームを図 3.8 に示す。変異プロモーターについて、HEK293 細胞へプラスミド DNA ライブラリーをトランスフェクションした。48 時間後、細胞中の mRNA を回収、GFP 遺伝子-ランダムタグ配列の cDNA を合成した。次世代シーケンサーを用いてランダムタグ部分の配列を決定した。基準として用いたラン

スフェクション前のプラスミド DNA ライブラリーについても同様に、ランダムタグ配列を決定した。

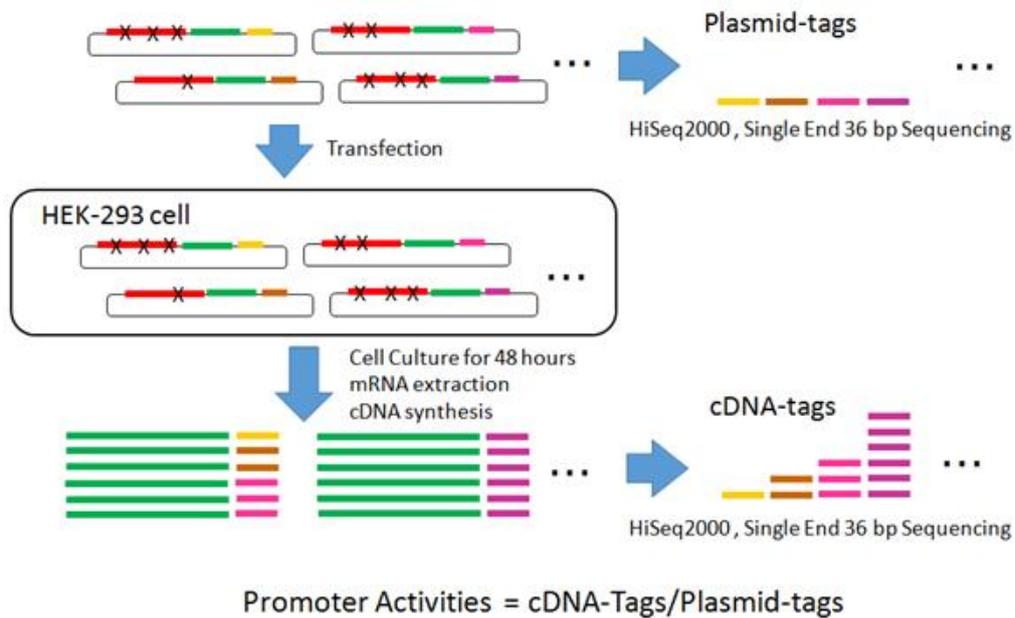


図 3.8 : 変異プロモーター配列の転写活性化能の測定

変異プロモーター配列を一括して HEK293 細胞へトランスフェクションし、48 時間後の GFP 遺伝子-ランダムタグ配列を mRNA として回収。逆転写による cDNA 合成後、次世代シーケンサーにてランダムタグの検出を行った。各変異プロモーター配列の基準となるトランスフェクション前のプラスミド DNA のランダムタグ配列も同様に検出した。

表 3.15 に、GFP 遺伝子-ランダムタグ配列の mRNA 由来の cDNA およびトランスフェクション前のプラスミド DNA のランダムタグ配列の決定について、そのシーケンス結果を示す。十分量かつ高品質なリードが得られていることを確認した。

表 3.15 : ランダムタグ配列のシーケンスサマリー

遺伝子	ライブラリー	Yield (Mbases)	% PF	# Reads	% of raw clusters per lane	% of >= Q30 Bases (PF)	Mean Quality Score (PF) Max =40
EEF1A1	cDNA	278	81.57	9,472,022	12.49	82.98	34.19
EEF1A1	plasmid DNA	125	81.11	4,283,935	5.65	82.84	34.15
GAPDH	cDNA	309	81.66	10,495,358	13.84	83.07	34.21
GAPDH	plasmid DNA	227	81.59	7,711,689	10.17	83.12	34.23
RPS12	cDNA	1,812	96.38	52,214,000	32.32	72.64	31.11
RPS12	plasmid DNA	790	97.3	22,562,937	13.97	72.71	31.13
PGK1	cDNA	1,113	98.61	31,366,805	23.16	81.1	33.22
PGK1	plasmid DNA	1,225	98.49	34,548,459	25.51	81.22	33.23
DDX5	cDNA	603	97.5	17,188,125	13.88	92.65	37.01
DDX5	plasmid DNA	517	97.48	14,721,002	11.88	92.71	37.02
ARL6IP5	cDNA	1,030	88.01	32,497,328	23.23	61.56	29.22
ARL6IP5	plasmid DNA	1,164	87.80	36,840,409	26.34	61.48	29.20
NTS	cDNA	1,254	98.51	35,350,906	26.77	97.38	38.13
NTS	plasmid DNA	1,185	98.47	33,416,511	25.30	97.34	38.12
RBBP5	cDNA	1,007	94.19	29,692,552	20.32	58.95	28.84
RBBP5	plasmid DNA	1,053	94.09	31,078,148	21.27	59.06	28.87

図 3.9 に、トランスフェクション前のプラスミド DNA のランダムタグ配列数の分布を示す。本研究で作成した各遺伝子の変異プロモーター配列ライブラリーは、10 ~30 ppm をピークとした分布を示した。

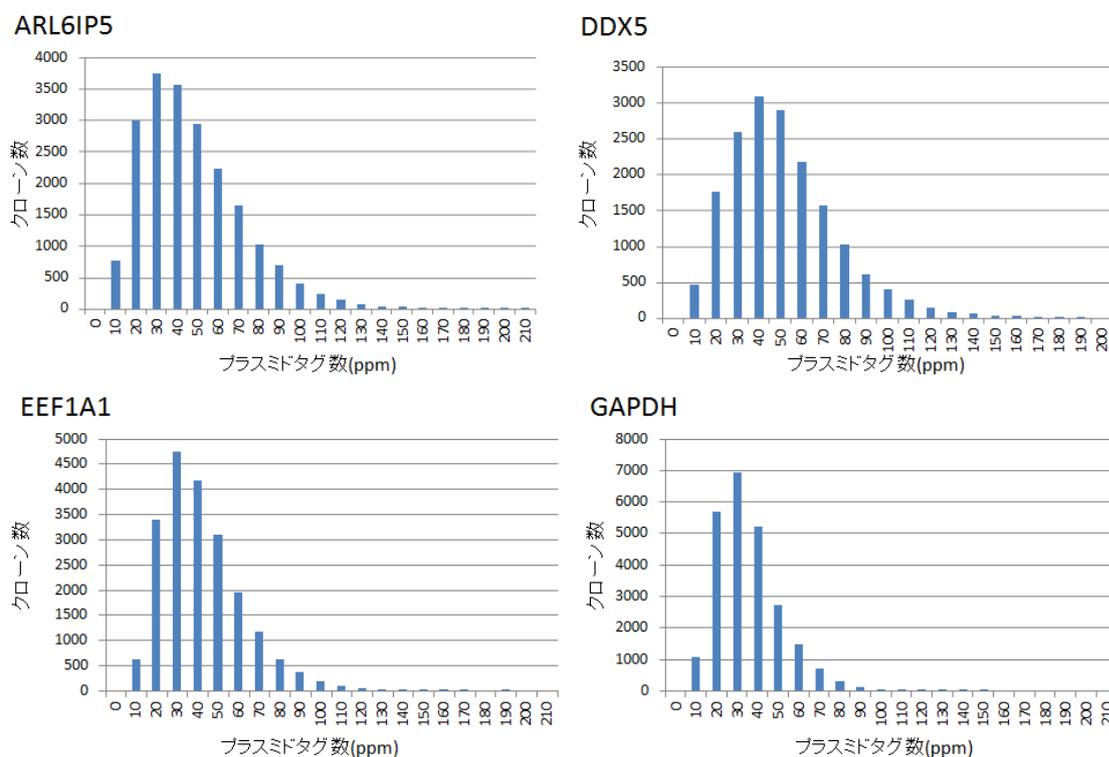


図 3.9 : プラスミド DNA のランダムタグ配列の分布

トランスフェクション前のプラスミド DNA のランダムタグ配列数の分布について、X 軸にプラスミド DNA のランダムタグ配列数、Y 軸にクローン数とした、ヒストグラムを示した。プラスミド DNA のランダムタグ配列数は ppm (parts per million) 値として表記した。

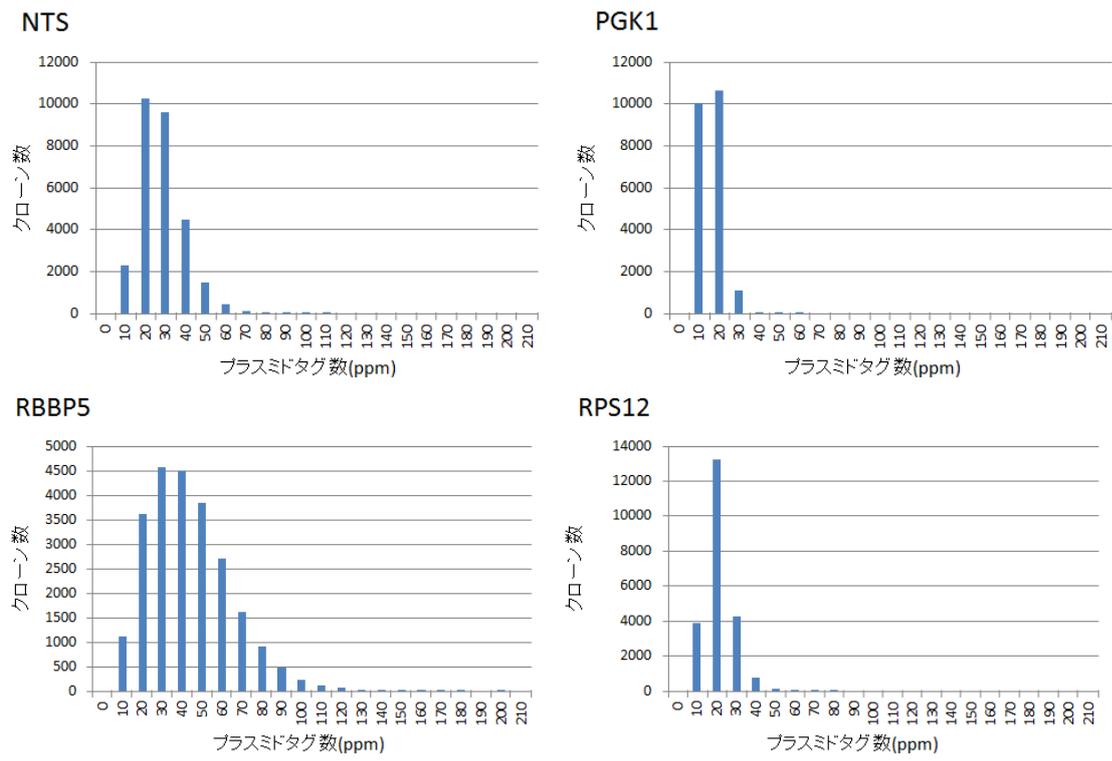


図 3.9 (続き) : プラスミド DNA のランダムタグ配列の分布

トランスフェクション前のプラスミド DNA のランダムタグ配列数の分布について、X 軸にプラスミド DNA のランダムタグ配列数、Y 軸にクローン数とした、ヒストグラムを示した。プラスミド DNA のランダムタグ配列数は ppm (parts per million) 値として表記した。

表 3.16 及び図 3.10 に、測定された変異プロモーター配列の分布を示す。鋳型となった野生型プロモーター配列の転写活性化能も図中に示した。変異プロモーター配列の転写活性化能は、野生型プロモーター配列の転写活性化能に比して、2.07~4.16 倍に変動していた。その平均値は野生型のものとはほぼ一致していた。プロモーター配列に変異を導入することで、転写活性化能に影響があり、その影響は転写活性化能が強くなる方向及び弱くなる方向の両方に及ぶことが明らかになった。野生型プロモーター配列の転写活性化能の平均値を基準としたときに、基準以下の活性を示す変異型プロモーター配列は平均 56.1%であった。

表 3.16 : 転写活性化能の要約

各遺伝子の変異プロモーター配列 (Mutant) 及び野生型プロモーター配列 (Wild Type) の転写活性化能について要約統計量を算出した。N : クローン数、Mean : 平均値、Std : 標準偏差、Median : 中央値、Min : 最小値、Max : 最大値

Gene	Promoter	N	Mean	Std	Median	Min	Max
ARL6IP5	Mutant	20486	-0.163	0.829	-0.063	-3.592	4.034
ARL6IP5	Wild Type	34157	0.076	0.262	0.089	-3.497	3.380
DDX5	Mutant	17139	-0.238	0.781	-0.127	-6.030	2.523
DDX5	Wild Type	5328	0.042	0.292	0.049	-1.999	1.298
EEF1A1	Mutant	20575	-0.239	1.086	-0.044	-7.069	2.466
EEF1A1	Wild Type	22765	0.110	0.525	0.120	-4.419	2.520
GAPDH	Mutant	24332	-0.432	0.999	-0.316	-6.309	4.554
GAPDH	Wild Type	38233	-0.086	0.419	-0.073	-4.433	5.266
NTS	Mutant	28833	-0.342	1.062	-0.090	-4.405	4.262
NTS	Wild Type	26620	-0.015	0.316	0.002	-3.337	2.011
PGK1	Mutant	21902	-0.421	1.050	-0.282	-4.892	4.417
PGK1	Wild Type	5129	0.000	0.289	0.018	-3.446	1.948
RBBP5	Mutant	23809	-0.137	0.937	0.127	-5.056	2.467
RBBP5	Wild Type	42635	0.097	0.225	0.112	-3.198	1.115
RPS12	Mutant	22366	-0.286	1.095	-0.118	-5.138	2.085
RPS12	Wild Type	18835	0.025	0.284	0.042	-4.533	0.999

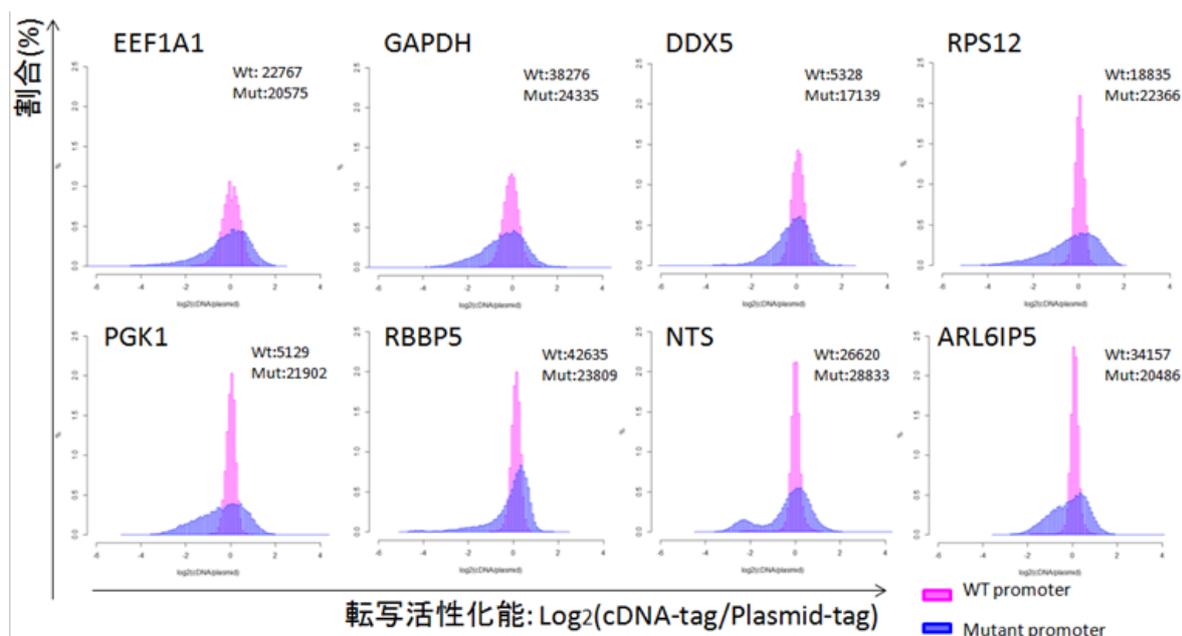


図 3.10 : 変異プロモーター配列の転写活性化能の分布

各遺伝子の変異プロモーター配列及び野生型プロモーター配列について、転写活性化能の強さ (X 軸) と存在割合 (Y 軸) を示した。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。

3.2.2 構築した測定系の確認

3.2.2.1 再現性の確認

本研究で構築した変異プロモーター配列の転写活性化能の測定方法について、その再現性を検証した。EEF1A1 遺伝子の変異プロモーターライブラリーについて、

その作成から、転写活性化能の測定まで、独立した2回の実験(Library1及びLibrary2)を行った。表3.17及び表3.18に、変異プロモーターの配列決定および転写活性化能測定時の概略を示す。

表 3.17 : 独立した2回の実験の変異プロモーター配列の配列決定時のシーケンスサマリー

遺伝子	library ID	ライブラリー	Yield (Mbases)	% PF	# Reads	% of raw clusters per lane	% of >= Q30 Bases (PF)	Mean Quality Score (PF) Max =40
EEF1A1	library1	3'端側	14409	93.28	152,933,498	36.06	91.34	34.84
EEF1A1	library1	5'端側	7471	93.00	79,534,614	18.75	85.49	33.21
EEF1A1	library2	3'端側	10000	91.03	108,767,056	22.76	89.53	34.40
EEF1A1	library2	5'端側	8330	90.97	90,670,790	18.98	84.26	32.92

表 3.18: 独立した2回の実験の変異プロモーター配列の転写活性化能測定時のシーケンスサマリー

遺伝子	library ID	ライブラリー	Yield (Mbases)	% PF	# Reads	% of raw clusters per lane	% of >= Q30 Bases (PF)	Mean Quality Score (PF) Max =40
EEF1A1	library1	cDNA	278	81.57	9,472,022	12.49	82.98	34.19
EEF1A1	library1	plasmid DNA	125	81.11	4,283,935	5.65	82.84	34.15
EEF1A1	library2	cDNA	606	97.39	17,963,560	13.96	92.64	37.00
EEF1A1	library2	plasmid DNA	285	97.64	8,116,061	6.55	92.80	37.05

表3.19に、配列決定および転写活性化能の測定を行った変異プロモーター配列に関する情報を示す。独立して作成したLibrary1及びLibrary2から得られた変異プロモーター配列数はそれぞれ20575、13121と約1.6倍の差異があった。これはプラスミドDNAライブラリーを大腸菌へ形質転換する際の効率が影響したものと考えられた。いずれのライブラリーにおいても10000クローンの多様性を有していた。

表3.19及び図3.11に、得られた変異プロモーター配列内に挿入された変異数および変異パターンを示す。Library1とLibrary2について、取得した変異プロモーター配列中に導入された変異およびその変異パターンは、同様のものではあった。塩基ごとに生じる変異パターンの割合はLibrary間で高い相関($r=0.816$, $p<0.0001$)を示した(図3.11)。

表 3.19 : 2 回の独立した実験間で得られた変異プロモーター配列数と挿入された変異パターン

2 回独立して作成した EEF1A1 遺伝子の変異プロモーター配列ライブラリーにおいて、各塩基に挿入された変異の有無及び各変異パターン（野生型プロモーター配列から A、T、G、C、deletion への変異）の数及び割合の平均値を示した。

	Library1	Library2
promoter length	1103	
GC content	0.59	
# of mutant promoter	20575	13121
#mutation	358.4(1.74 %)	241.5(1.84 %)
#toA	110.2(0.57 %)	73(0.60 %)
#toT	87.6(0.52 %)	59.7(0.56 %)
#toG	66.4(0.46 %)	45.3(0.49 %)
#toC	67.9(0.45 %)	45.4(0.48 %)
#del	26.3(0.12 %)	18.1(0.13 %)

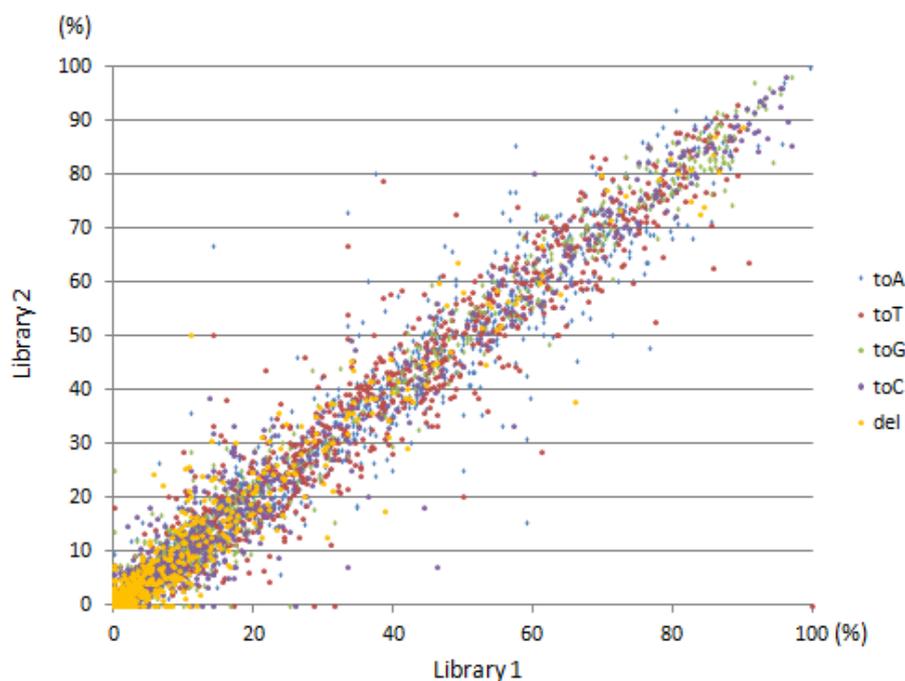


図 3.11 : 各塩基の変異パターン構成割合の実験間比較

2 回独立して作成した EEF1A1 遺伝子の変異プロモーター配列について、各塩基に含まれる野生型プロモーター配列からの変異パターンの構成割合を比較した。青 : A への変異、赤 : T への変異、緑 : G への変異、紫 : C への変異、橙 : 欠損をそれぞれ示した。相関係数は Pearson の積率相関係数を用いて算出した。

2 回の独立した EEF1A1 遺伝子の変異プロモーター配列について、転写活性化能の測定結果を比較した（表 3.20、図 3.12）。独立した 2 回の実験で測定した転写活性化能の測定結果は相互に類似しており、相関係数は 0.95 ($P < 0.001$) であった。

これらの結果から、本研究にて確立した多様な変異プロモーター配列の取得、配列決定、転写活性化能の測定を実施する一連の実験系が、再現性のある測定系とし

て構築されたものと考えられた。

表 3.20 : 2 回独立した実験間で得られた転写活性化能の要約統計量の比較

2 回の独立した EEF1A1 遺伝子の変異プロモーター配列ライブラリーである Library1 及び Library2 の転写活性化能について要約統計量を算出した。N : クローン数、Mean : 平均値、Std : 標準偏差、Median : 中央値、Min : 最小値、Max : 最大値

Gene	Library	Promoter	N	Mean	Std	Median	Min	Max
EEF1A1	Library1	Mutant	20575	-0.239	1.086	-0.044	-7.069	2.466
EEF1A1	Library2	Mutant	13121	-0.296	1.050	-0.091	-6.304	2.782

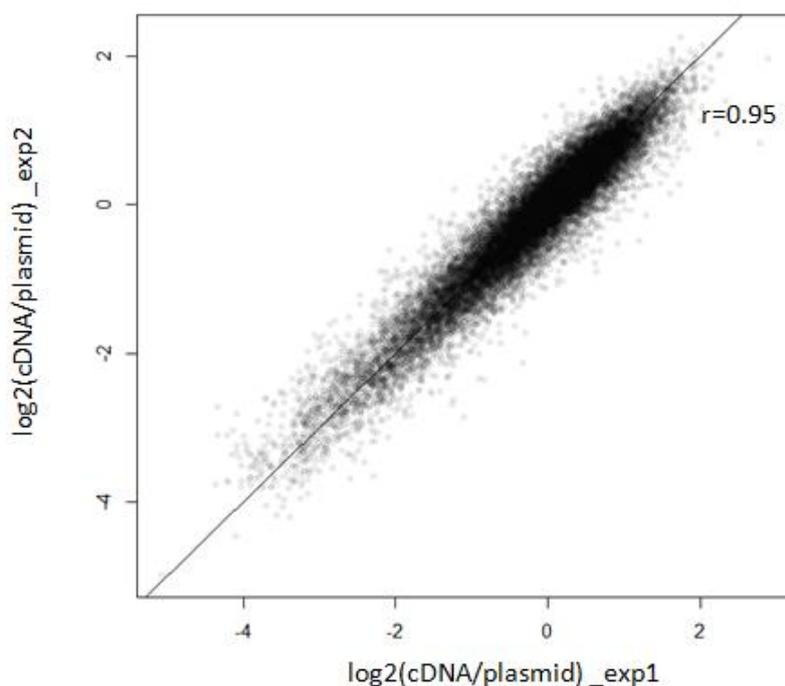


図 3.12 : 独立に 2 回測定した EEF1A1 遺伝子の変異プロモーターの転写活性化能

独立して 2 回測定した EEF1A1 遺伝子の変異プロモーター配列ライブラリーより得られた転写活性化能データを、X 軸に測定 1、Y 軸に測定 2 としてプロットした。相関係数は Pearson の積率相関係数を用いて算出した。

3.2.2.2 従来のルシフェラーゼアッセイとの比較

本研究の実験系が従来の実験系と同様の精度が得られるかを検証した。本研究手法を用いて測定した転写活性化能データと従来の測定方法であるルシフェラーゼアッセイにより測定した転写活性化能データを比較した (図 3.13)。比較には、本研究にて転写活性化能を測定した遺伝子の変異プロモーター配列から、3 遺伝子 (EEF1A1、GAPDH、DDX5)、66 種類の変異プロモーター配列を選択した。これらについてルシフェラーゼアッセイにより転写活性化能を個別に測定した。本研究手法で測定した転写活性化能とルシフェラーゼアッセイで測定

した転写活性化能の相関係数は 0.816 (P<0.001) と高い相関を示した。このことから、本研究にて確立した転写活性化能の測定系は、従来のルシフェラーゼアッセイによる転写活性化能データと同様の結果が得られる手法であると考えられた。

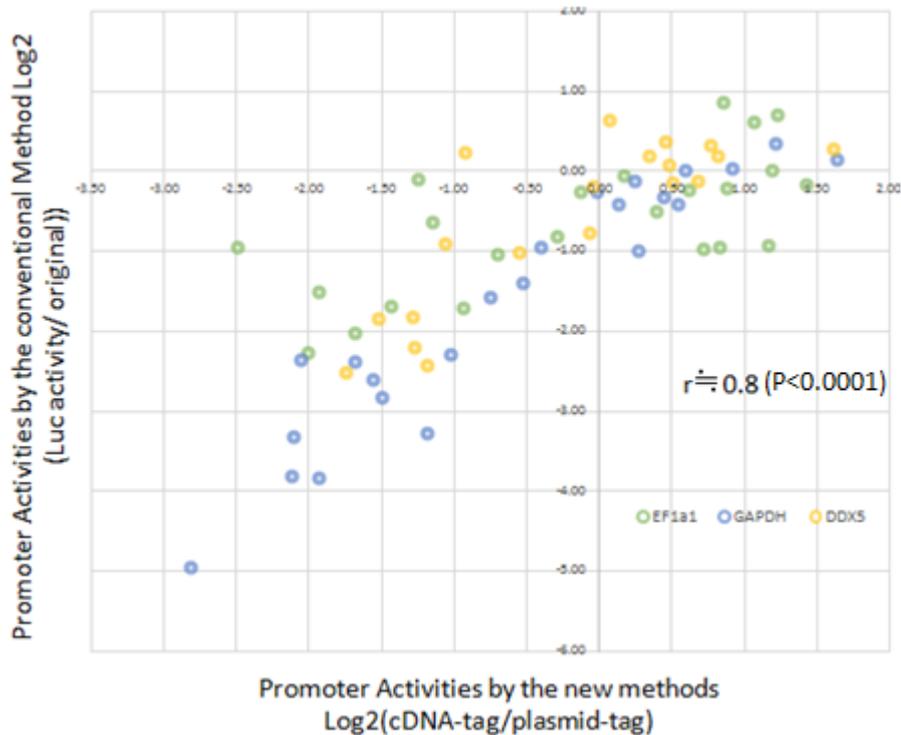


図 3.13 : 次世代シーケンス技術を用いた手法とルシフェラーゼアッセイの比較

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 DOX5 の変異型プロモーター。相関係数は Pearson の積率相関係数を用いて算出した。

3.2.3 プロモーター配列の塩基レベルの差が転写活性化能へ与える影響の解析

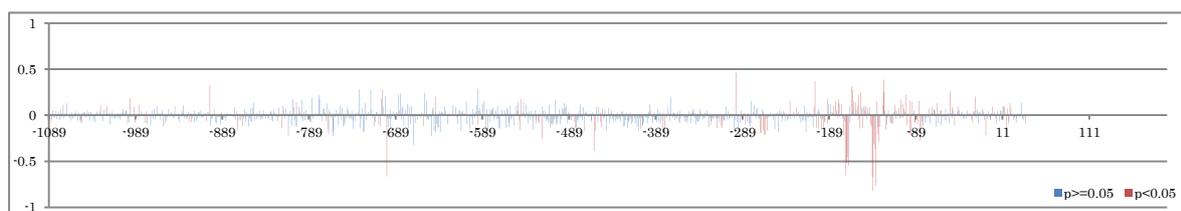
測定した変異プロモーター配列と転写活性化能のデータより、各塩基に導入された変異又は変異パターンが転写活性化能へ与える影響を、重回帰モデルの偏回帰係数として推定した。重回帰分析は、目的変数は変異プロモーターの転写活性化能、説明変数は各塩基の変異有無又は各変異パターンの有無とした。変異パターンとは、野生型塩基からそれぞれ A、T、G、C、deletion 変異を指す。1 塩基レベルの差が転写活性化能に与える影響を解析する上で、プロモーター配列中すべての塩基に対応する偏回帰係数を算出することが可能であるため、重回帰モデルを採用した。

3.2.3.1 重回帰分析を利用した各塩基の変異が転写活性化能へ与える影響の推定

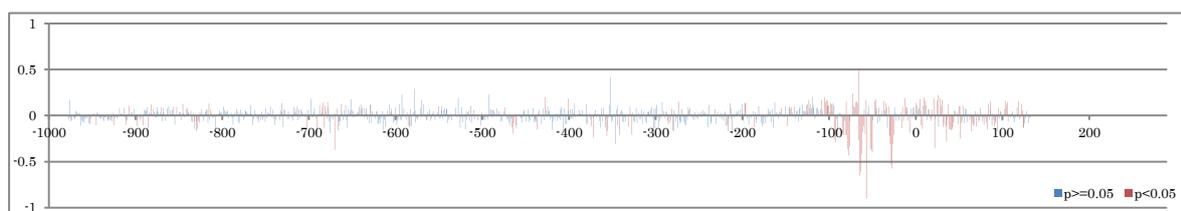
3.2.3.1.1 各塩基の変異が転写活性化能へ与える影響の推定

各塩基に変異が導入された際に、その変異が転写活性化能へどの程度の影響を与えるのかを解析した。本研究にて測定した転写活性化能を目的変数に、各塩基の変異有無を説明変数とした重回帰分析にて、各塩基の転写活性化能へ与える影響を偏回帰係数として推定した (図 3.14)。

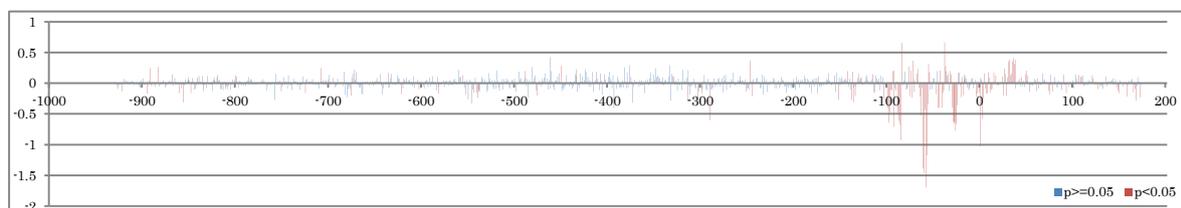
ARL6IP5



DDX5



EEF1A1



GAPDH

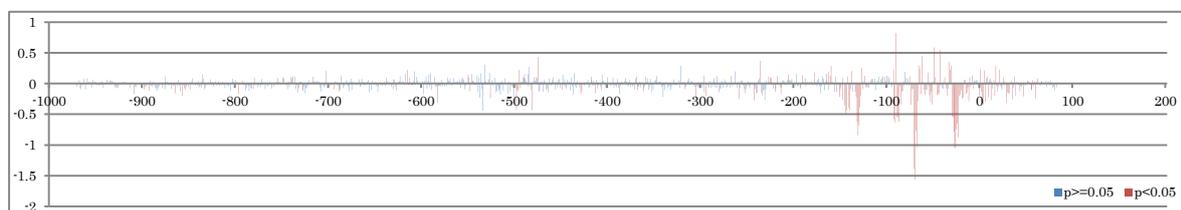
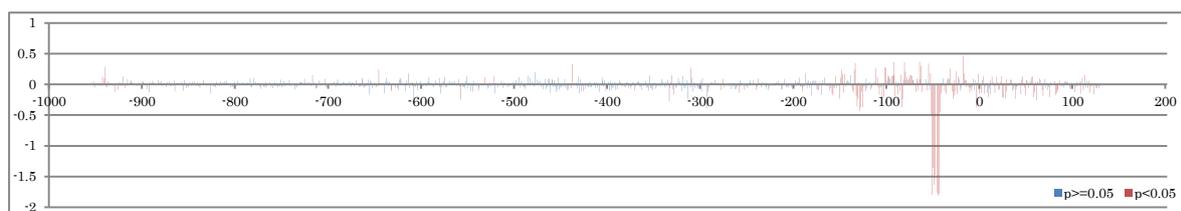


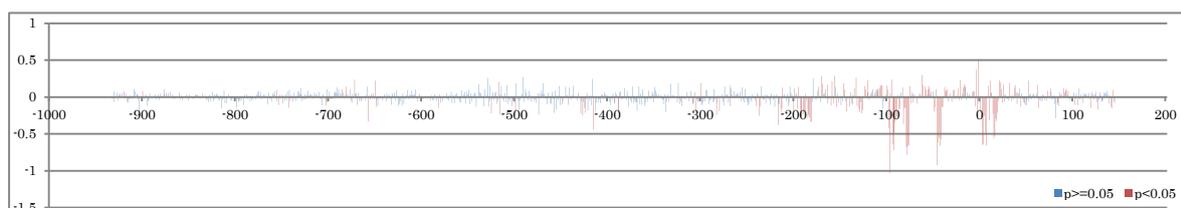
図 3.14 : プロモーター配列の変異が転写活性化能へ与える影響

遺伝子ごとに転写開始点を基準(0)としたときの、プロモーター配列中の位値 (X 軸) とその位値の塩基に変異が導入されることで転写活性化能へ与える影響 (Y 軸) を棒グラフとして示した。転写活性化能へ与える影響は正であれば positive な影響、負であれば negative な影響。棒グラフの色は推定された転写活性化能が 0 か否かの検定結果として赤 : $p < 0.05$ 、青 : $p \geq 0.05$ を示した。

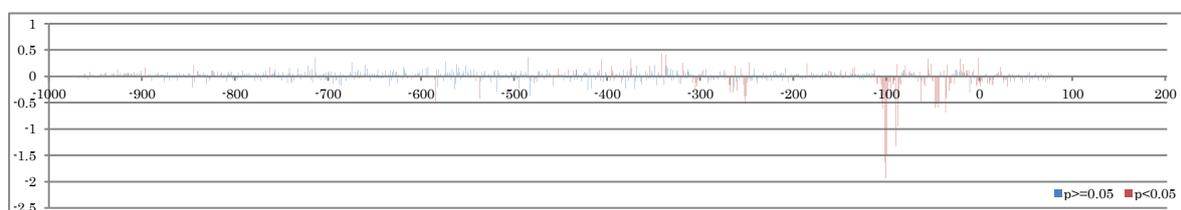
NTS



PGK1



RBBP5



RPS12

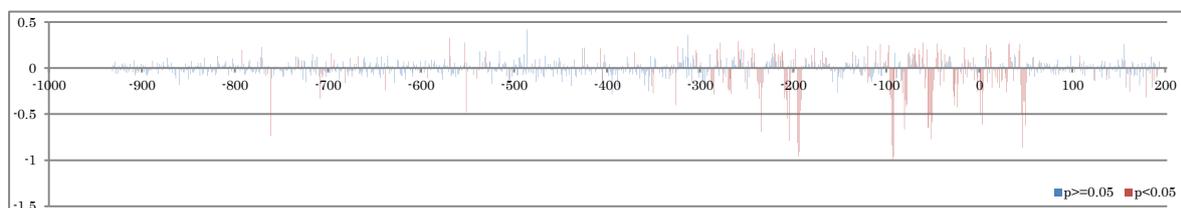
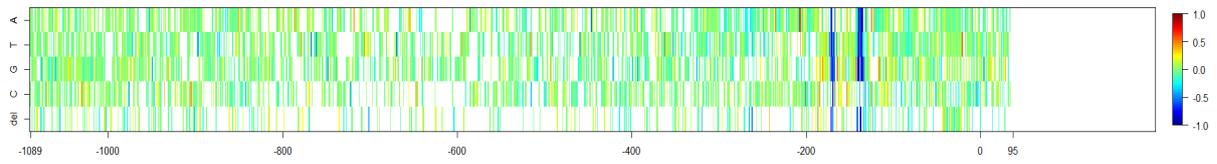


図 3.14 (続き) : プロモーター配列の変異が転写活性化能へ与える影響

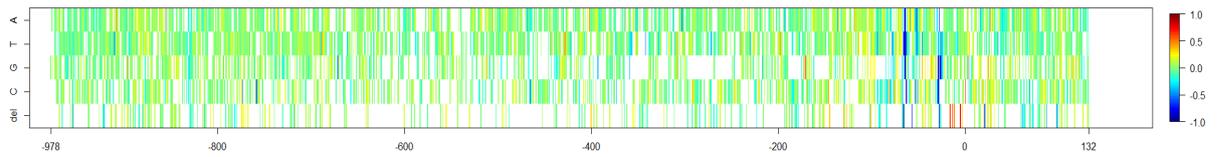
遺伝子ごとに転写開始点を基準(0)としたときの、プロモーター配列中の位値 (X 軸) とその位値の塩基に変異が導入されることで転写活性化能へ与える影響 (Y 軸) を棒グラフとして示した。転写活性化能へ与える影響は正であれば positive な影響、負であれば negative な影響。棒グラフの色は推定された転写活性化能が 0 か否かの検定結果として赤 : $p < 0.05$ 、青 : $p \geq 0.05$ を示した。

各塩基に変異を導入した際、野生型の塩基からそれぞれ A、T、G、C、deletion の変異パターンごとに、転写活性化能への程度の影響を与えるのかを解析した。本研究にて測定した転写活性化能を目的変数に、各塩基の変異パターンを説明変数とした重回帰分析にて、各塩基の変異パターンが転写活性化能へ与える影響をその偏回帰係数として推定した (図 3.15)。

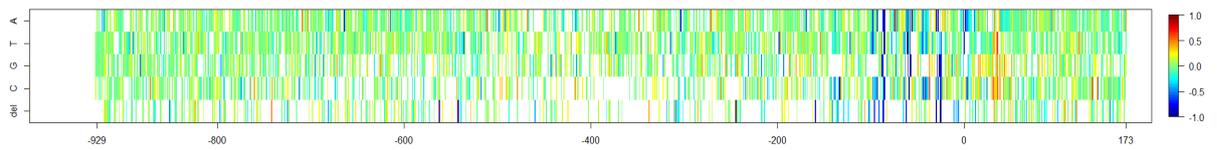
ARL6IP5



DDX5



EEF1A1



GAPDH

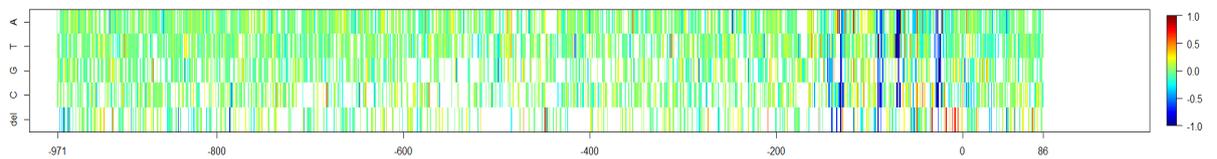
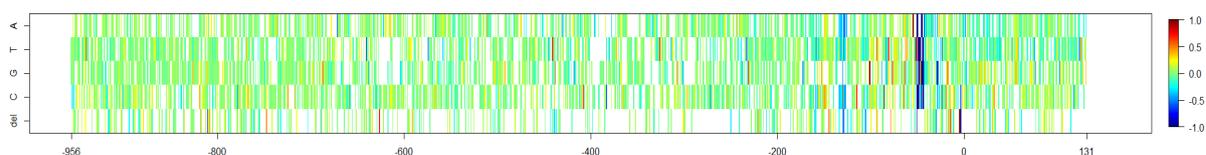


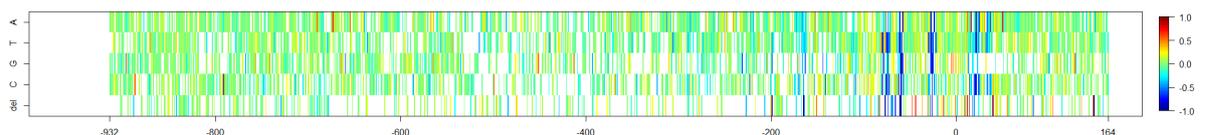
図 3.15 : プロモーター配列の変異が変異パターンごとに転写活性化能へ与える影響

遺伝子ごとに転写開始点を基準(0)としたときの、プロモーター配列中の位値 (X 軸) とその位値の塩基に変異が導入されることで転写活性化能へ与える影響 (Y 軸) をヒートマップとして示した。転写活性化能へ与える影響は正であれば positive な影響 (赤)、負であれば negative な影響 (青)

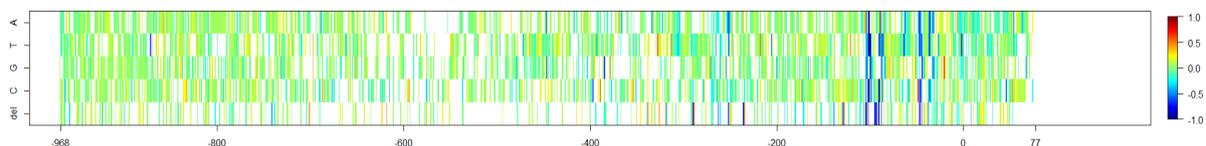
NTS



PGK1



RBBP5



RPS12

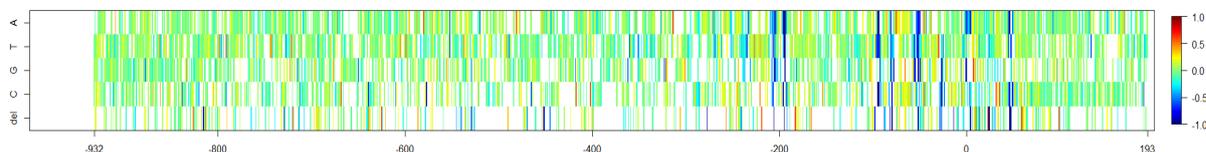


図 3.15 (続き) : プロモーター配列の変異が変異パターンごとに転写活性化能へ与える影響

遺伝子ごとに転写開始点を基準(0)としたときの、プロモーター配列中の位値 (X 軸) とその位値の塩基に変異が導入されることで転写活性化能へ与える影響 (Y 軸) をヒートマップとして示した。転写活性化能へ与える影響は正であれば positive な影響 (赤)、負であれば negative な影響 (青)

3.2.3.1.2 推定された塩基レベルの転写活性化能へ与える影響の再現性

本研究での各塩基の変異が転写活性化能へ与える影響の推定値に再現性があるか確かめた。2 回独立して作成した EEF1A1 遺伝子の変異プロモーター配列ライブラリーの転写活性化能データを用い、重回帰分析にて変異の導入が転写活性化能へ与える影響を解析した。(図 3.16、図 3.17)。

2 回の独立した実験データから算出された各塩基への変異が転写活性化能へ与える影響の推定値について、相関係数 0.693 ($p < 0.0001$) の相関が認められた。特に、偏回帰係数が $p < 0.05$ の塩基では、相関係数は 0.972 ($p < 0.0001$) で、高い相関を示した (図 3.18)。各塩基の変異パターンごとに転写活性化能へ与える影

響の推定値については、2回の推定値間で相関係数として 0.543 ($p < 0.0001$)、 $p < 0.05$ で推定された値のみを用いると相関係数として 0.914 ($p < 0.0001$) であった (図 3.19)。

本研究で構築した実験系及び重回帰分析を利用した、プロモーター配列の 1塩基レベルの配列差が転写活性化能へ与える影響の解析結果は、再現性が得られたものと考えた。

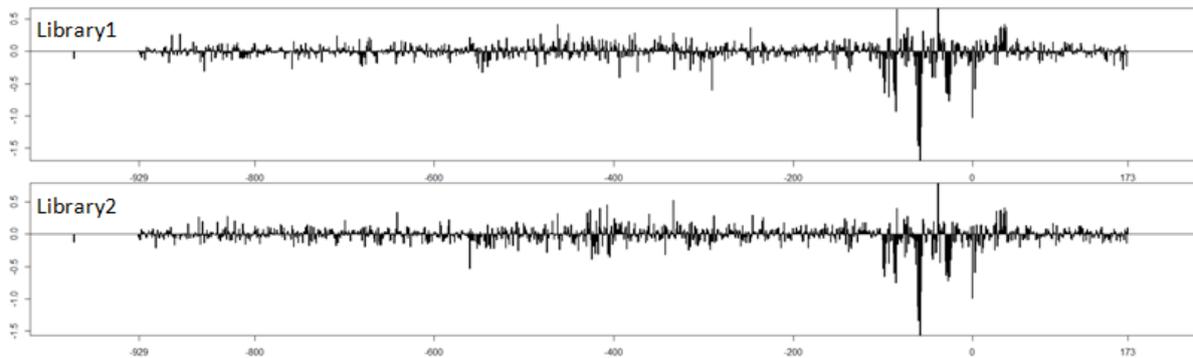


図 3.16 : 2 回の独立した実験データから得られた変異有無が転写活性化能へ与える影響

独立して 2 回作成した EEF1A1 遺伝子の変異プロモーター配列ライブラリーより得られた転写活性化能データを用い、各塩基への変異が転写活性化能へ与える影響を重回帰分析の偏回帰係数として推定した。X 軸: 転写開始点を基準としたときのプロモーター配列内の位値、Y 軸: 各塩基に変異が入った際の転写活性化能への影響。

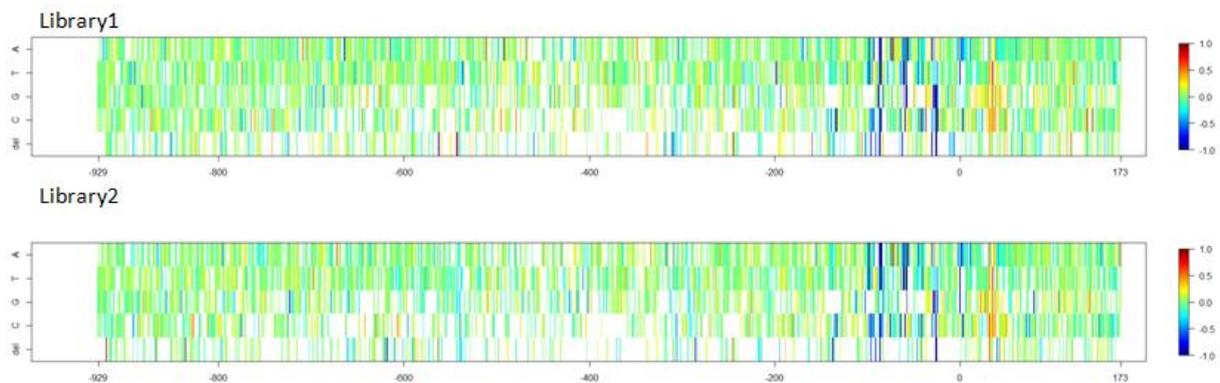


図 3.17 : 2 回の独立した実験データから得られた変異パターンごとの転写活性化能へ与える影響

独立して 2 回作成した EEF1A1 遺伝子の変異プロモーター配列ライブラリーより得られた転写活性化能データを用い、各塩基に変異が入った際、変異パターンごとの転写活性化能への影響を重回帰分析の偏回帰係数として推定した。X 軸: 転写開始点を基準としたときのプロモーター配列内の位値、Y 軸: 各塩基に変異が入った際の転写活性化能への影響。

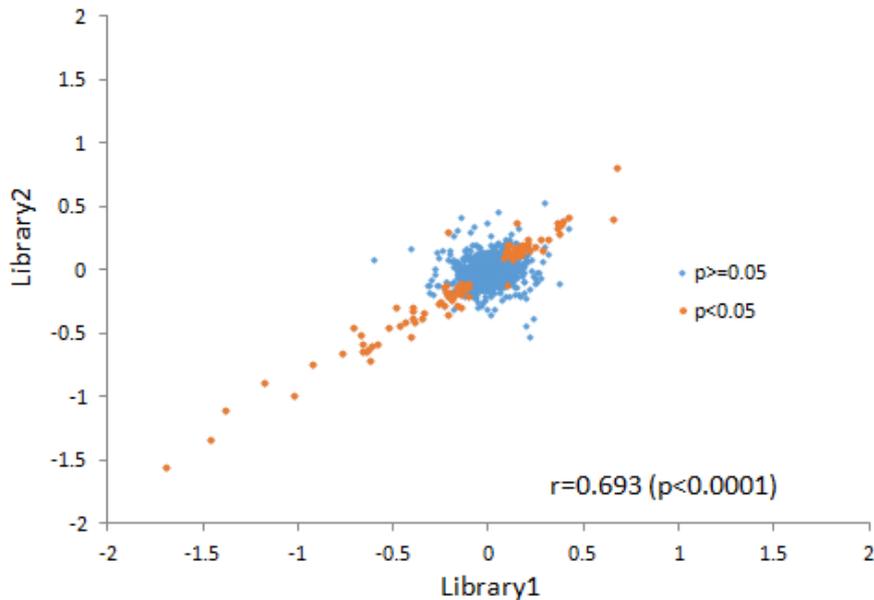


図 3.18 : 2 回独立したデータから推定した変異有無の転写活性化能への影響の比較

独立して 2 回作成した EEF1A1 遺伝子の変異プロモーター配列ライブラリーより得られた転写活性化能データを用い、各塩基に変異が入った際の転写活性化能への影響を重回帰分析の偏回帰係数として推定し、X 軸に Library1、Y 軸に Library2 をプロットした。橙: 推定値が $p < 0.05$ 、青: 推定値が $p \geq 0.05$ 。相関係数は Pearson の積率相関係数を用いて算出した。

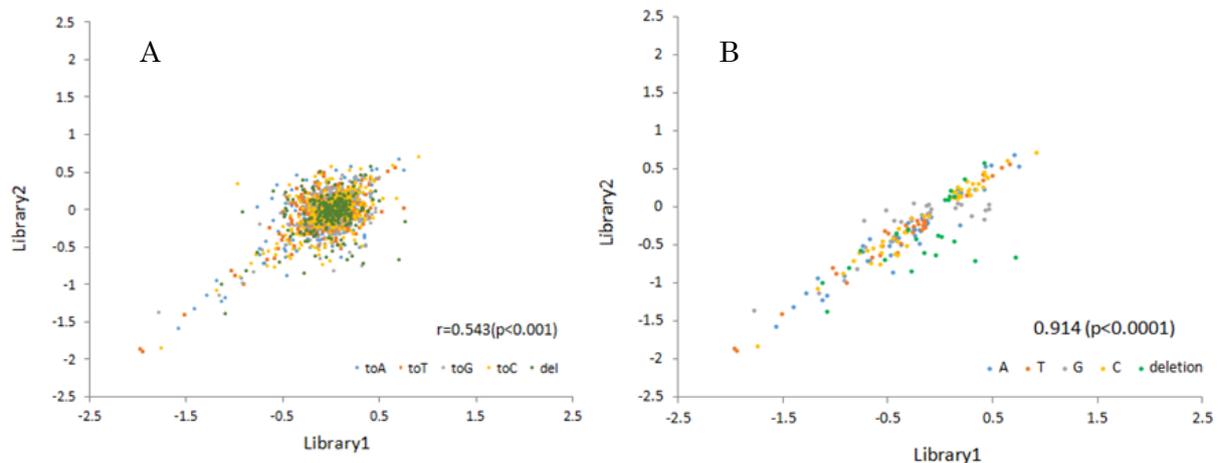


図 3.19 : 2 回独立したデータから推定した変異パターン of 転写活性化能への影響の比較

独立して 2 回作成した EEF1A1 遺伝子の変異プロモーター配列ライブラリーより得られた転写活性化能データを用い、各塩基に導入される変異パターンによる転写活性化能への影響を重回帰分析の偏回帰係数として推定し、X 軸に Library1、Y 軸に Library2 をプロットした。A. 推定された全ての偏回帰係数のプロット、B. $p < 0.05$ で推定された偏回帰係数のプロット。青: A への変異、赤: T への変異、紫: G への変異、橙: C への変異、緑: deletion。相関係数は Pearson の積率相関係数を用いて算出した。

3.2.3.1.3 塩基レベルの配列差が転写活性化能へ与える影響の特徴

本研究で推定された各塩基の変異が転写活性化能へ与える影響について、その特徴を示す。

表 3.21 に、プロモーター配列中の変異が転写活性化能へ影響を与える影響を推定できた塩基数とその割合を示す。塩基に変異が入ることで転写活性化能へ与える影響を推定できた塩基は平均して全体の 99.5 % であった。そのうち、変異が導入されることで転写活性化能に正 (positive) 又は負 (negative) に影響すると推定された塩基はそれぞれ平均して 49.0 %、51.0 % と、ほぼ同程度であった。偏回帰係数が 0 か否かの検定結果である p 値 0.05 及び 0.01 を閾値とし、変異が導入されることで転写活性化能に正負の影響があると推定された塩基を抽出した。抽出された塩基はそれぞれ平均して、プロモーター配列中の 17.5 % (p<0.05)、10.8 % (p<0.01) であった。抽出された塩基のうち正、負のそれぞれに影響を与える塩基数は、平均して全体の 7.6 %、10.0 % (p<0.05)、4.3 %、6.6 % (p<0.01) で、負の影響が推定された塩基が多かった。

表 3.21: 変異が転写活性化能へ与える影響を推定できた塩基数とその割合

各塩基の変異が転写活性化能へ与える影響を推定できた塩基数とその割合を示す。影響の方向性別の塩基数と割合を示す。推定値である偏回帰係数が 0 か否かの検定結果を基に p<0.05 及び p<0.01 を閾値とした際の、それぞれの塩基数及び割合も示す。

Gene Name	length	推定可能であった塩基数				p<0.05		p<0.01		
		度数	%	効果の方向	度数	%	度数	%	度数	%
ARL6IP5	1127	1117	99.1	positive	504	45.1	70	6.3	44	3.9
				negative	613	54.9	95	8.5	52	4.7
DDX5	1111	1106	99.5	positive	572	51.7	104	9.4	57	5.2
				negative	534	48.3	100	9.0	61	5.5
EEF1A1	1103	1101	99.8	positive	551	50.0	79	7.2	50	4.5
				negative	550	50.0	99	9.0	68	6.2
GAPDH	1058	1057	99.9	positive	453	42.9	71	6.7	36	3.4
				negative	604	57.1	140	13.2	94	8.9
NTS	1088	1088	100.0	positive	535	49.2	95	8.7	64	5.9
				negative	553	50.8	118	10.8	67	6.2
PGK1	1079	1078	99.9	positive	516	47.9	87	8.1	55	5.1
				negative	562	52.1	120	11.1	85	7.9
RBBP5	1046	1026	98.1	positive	591	57.6	72	7.0	31	3.0
				negative	435	42.4	82	8.0	57	5.6
RPS12	1126	1122	99.6	positive	535	47.7	82	7.3	38	3.4
				negative	587	52.3	118	10.5	85	7.6

遺伝子ごとの変異の転写活性化能へ与える影響の要約統計量を表 3.22 に示した。P<0.05 で推定された変異が転写活性化能へ与える影響は、正 (Positive) の影響として 0.042~0.823、負 (Negative) の影響として -0.049~-1.941 の範

圏であった。

表 3.22 : 変異の転写活性化能へ与える影響の要約統計量

推定された変異の転写活性化能へ与える影響について要約統計量を算出した。N : 塩基数、Mean : 平均値、Std : 標準偏差、Median : 中央値、Min : 最小値、Max : 最大値。効果の方向性は、偏回帰係数が 0 か否かの検定結果が $p < 0.05$ の推定値うち、正のものを Positive、負のものを Negative とした。

Gene	効果の方向	N	Mean	Std	Median	Min	Max
ARL6IP5	Positive	70	0.141	0.089	0.110	0.042	0.464
ARL6IP5	Negative	95	-0.181	0.162	-0.129	-0.821	-0.054
ARL6IP5	Others	952	-0.006	0.067	-0.006	-0.326	0.290
DDX5	Positive	104	0.124	0.055	0.116	0.051	0.507
DDX5	Negative	100	-0.190	0.141	-0.144	-0.896	-0.057
DDX5	Others	902	0.002	0.062	0.004	-0.305	0.418
EEF1A1	Positive	79	0.188	0.116	0.155	0.050	0.669
EEF1A1	Negative	99	-0.321	0.305	-0.203	-1.699	-0.087
EEF1A1	Others	923	0.006	0.078	0.002	-0.407	0.422
GAPDH	Positive	71	0.175	0.137	0.123	0.045	0.823
GAPDH	Negative	140	-0.264	0.262	-0.168	-1.562	-0.052
GAPDH	Others	846	-0.006	0.067	-0.008	-0.441	0.308
NTS	Positive	95	0.144	0.089	0.123	0.045	0.463
NTS	Negative	118	-0.239	0.371	-0.126	-1.816	-0.049
NTS	Others	875	-0.001	0.051	0.001	-0.385	0.201
PGK1	Positive	87	0.147	0.073	0.130	0.059	0.506
PGK1	Negative	120	-0.254	0.201	-0.194	-1.036	-0.054
PGK1	Others	871	-0.001	0.064	-0.001	-0.289	0.271
RBBP5	Positive	72	0.150	0.090	0.117	0.051	0.443
RBBP5	Negative	82	-0.303	0.349	-0.160	-1.944	-0.054
RBBP5	Others	872	0.011	0.072	0.013	-0.369	0.369
RPS12	Positive	82	0.174	0.061	0.170	0.067	0.330
RPS12	Negative	118	-0.301	0.227	-0.223	-0.990	-0.064
RPS12	Others	922	0.000	0.070	-0.002	-0.286	0.418

3.2.3.1.4 塩基レベルの配列差が転写活性化能へ与える影響と配列長

各遺伝子のプロモーター配列中の位置と塩基レベルの配列差が転写活性化能へ与える影響から、転写活性化能へ影響の大きい塩基が転写開始点近傍でクラスターとして存在していることが示された (図 3.14)。これは、転写因子結合配列を検出していることが考えられるため、詳細に解析をした。

変異の導入が転写活性化能へ影響があると推定される領域の配列長とその存在割合を明らかにするため、偏回帰係数の p 値が 0.05 未満を示した塩基が、連続する塩基長を集計した (表 3.23)。その結果、転写活性化能へ影響があると推定された領域の平均 2.7 % が 6 塩基以上のクラスターを形成していることが明らかになった。転写因子結合部位の共通配列の長さは典型的には 6~10 塩基[4]、転写因子結合配列検索ソフトウェアで用いられるモチーフの長さとして 5~30

塩基程度[65]である。一方、平均して 97.3 %で 5 塩基以下の配列長であった。配列長が短い領域は、クラスター間に偶然 $p < 0.05$ で影響が推定されなかったことで本来は長い配列長になる領域が、過小評価されている可能性が考えられた。この問題を調べるため、効果の推定された 2 塩基の間に 5 塩基まで間隔があっても連続とみなす緩やかな条件で集計した。この集計においても 5 塩基未満の配列長となる領域は 94.0 %存在した。

表 3.23: 変異の導入が転写活性化能へ影響があると推定された領域の連続する配列長

塩基数	8遺伝子		ARL6IP5		DDX5		EEF1A1		GAPDH	
	度数	%	度数	%	度数	%	度数	%	度数	%
1	898	77.6	116	80.0	125	79.1	104	80.0	123	79.9
2	131	11.3	20	13.8	15	9.5	11	8.5	15	9.7
3	54	4.7	6	4.1	10	6.3	5	3.9	7	4.6
4	25	2.2	1	0.7	3	1.9	1	0.8	2	1.3
5	19	1.6	1	0.7	0	0.0	4	3.1	1	0.7
6	8	0.7	0	0.0	2	1.3	1	0.8	1	0.7
7	8	0.7	0	0.0	2	1.3	1	0.8	2	1.3
8	5	0.4	0	0.0	0	0.0	2	1.5	1	0.7
9	2	0.2	0	0.0	0	0.0	1	0.8	1	0.7
10	2	0.2	1	0.7	0	0.0	0	0.0	0	0.0
11	3	0.3	0	0.0	1	0.6	0	0.0	0	0.0
12	1	0.1	0	0.0	0	0.0	0	0.0	0	0.0
13	1	0.1	0	0.0	0	0.0	0	0.0	0	0.0
16	1	0.1	0	0.0	0	0.0	0	0.0	1	0.7

塩基数	NTS		PGK1		RBBP5		RPS12	
	度数	%	度数	%	度数	%	度数	%
1	122	76.3	103	71.0	96	78.7	109	75.7
2	15	9.4	25	17.2	13	10.7	17	11.8
3	9	5.6	4	2.8	7	5.7	6	4.2
4	7	4.4	6	4.1	1	0.8	4	2.8
5	3	1.9	2	1.4	3	2.5	5	3.5
6	1	0.6	1	0.7	1	0.8	1	0.7
7	2	1.3	1	0.7	0	0.0	0	0.0
8	1	0.6	0	0.0	0	0.0	1	0.7
9	0	0.0	0	0.0	0	0.0	0	0.0
10	0	0.0	1	0.7	0	0.0	0	0.0
11	0	0.0	0	0.0	1	0.8	1	0.7
12	0	0.0	1	0.7	0	0.0	0	0.0
13	0	0.0	1	0.7	0	0.0	0	0.0
16	0	0.0	0	0.0	0	0.0	0	0.0

変異が転写活性化能に影響のある塩基が連続する配列長と、そのプロモーター配列中の位置を解析した。転写活性化能に影響する塩基の連続する配列長ごとに、転写開始点を 0 とした際の位置を解析した (図 3.20)。各連続塩基の位置はその中央値とした。転写活性化能へ影響のある配列の長さが 6 塩基以上連続する領域は、転写開始点の近傍領域 (-143~+43) に集中していた。5 塩基以下

の配列長となる領域についても、配列長が長くなるほど、転写開始点近傍に位置していることが示された。一方、1塩基又は2塩基の配列長で転写活性化能への影響が推定された塩基は、プロモーター配列中の広範囲に位置していた。

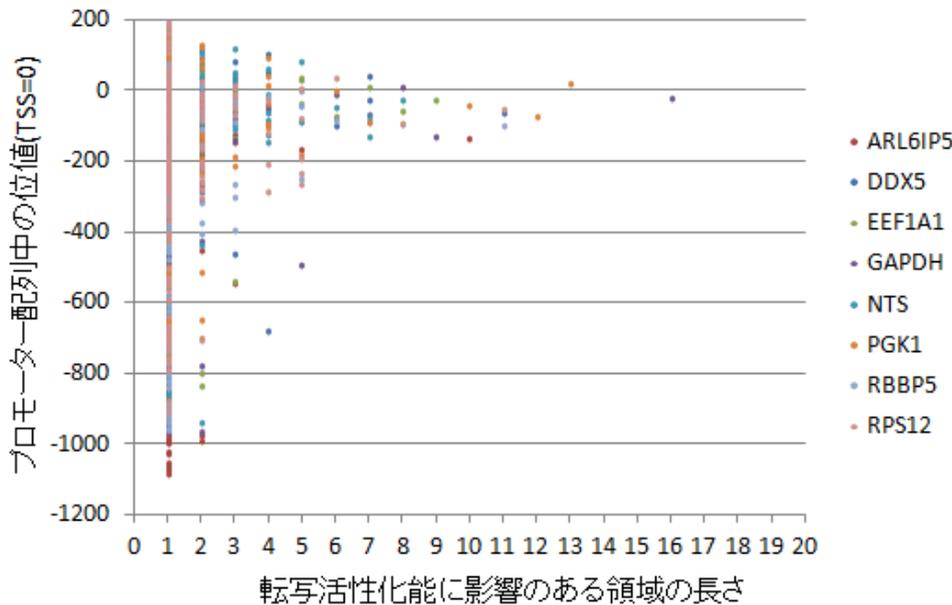


図 3.20: 転写活性化能に影響のある領域の配列長とプロモーター配列中の位値

プロモーター配列中の塩基置換について、転写活性化能への影響が $p < 0.05$ で推定された塩基が連続する領域の配列長 (X 軸) ごとに、転写開始点を基準(0)としたプロモーター配列中の位値 (Y 軸) をプロットした。各連続塩基の位値はその中心とした。各色は、各遺伝子のプロモーター配列を示す (赤:ARL6IP5、紺: DDX5、緑: EEF1A1、紫: GAPDH、青:NTS、橙: PGK1、水色: RBBP5、ピンク: RPS12)。

各塩基の変異が転写活性化能に影響のある塩基の連続する配列長と、変異が導入されることで転写活性化能へ与える影響の方向 (正負) を比較した (図 3.21、表 3.24)。連続する塩基の転写活性化能へ与える影響はその平均値を用いた。連続する転写活性化能へ影響のある塩基が 5 塩基以下の領域では、転写活性化能へ影響する方向性として、正 47.0%、負 53.0% とほぼ同程度であったのに対し、6 塩基以上連続する領域では正 4.4%、負 95.7% と負の影響のある領域が濃縮されていることが明らかとなった ($p < 0.0001$) (表 3.24)。

各遺伝子のプロモーター配列に対し転写因子結合配列検索ソフトウェア MATCH にて転写因子結合配列を探索し、連続する転写活性化能へ影響のある塩基が 6 塩基以上の領域と比較した (表 3.25)。TATA ボックスを有するプロモーター領域では、転写開始点上流 30 塩基程度の 6 塩基以上連続する領域と TATA ボックスの検索結果が一致した。その領域への変異は転写活性化能への負の影響が推定された (詳細は以下に示す)。このように連続する転写活性化能へ影響のある塩基が 6 塩基以上の領域は、転写因子結合配列を検出したものと考えら

れた。遺伝子発現に影響のある cis-eQTL が転写開始点近傍にエンリッチされるという推定結果が報告されている[63]。プロモーター領域の広い範囲に分布した変異が転写活性化能へ影響を与える短い領域が、転写活性化能への影響を介した潜在的な cis-eQTL を検出したものだとすると興味深い。

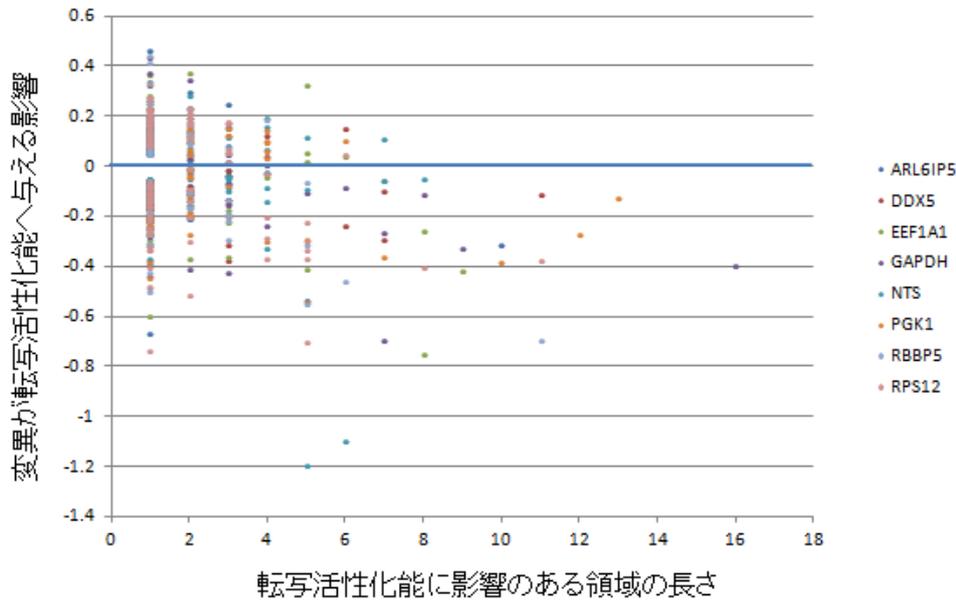


図 3.21: 各塩基の変異が転写活性化能に影響のある領域の配列長と影響の方向と強さ

プロモーター配列中の塩基置換について、転写活性化能への影響が $p < 0.05$ で推定された塩基が連続する領域の配列長 (X 軸) ごとに、変異が転写活性化能へ与える影響の推定値 (Y 軸) にプロットした。連続する塩基の転写活性化能へ与える影響は平均値を用いた。各色は、各遺伝子のプロモーター配列を示す (赤:ARL6IP5、紺: DDX5、緑: EEF1A1、紫: GAPDH、青:NTS、橙: PGK1、水色: RBBP5、ピンク: RPS12)。

表 3. 24 : 転写活性化能に影響のある領域の配列長と影響の方向性

プロモーター配列中の塩基置換について、転写活性化能への影響が $p < 0.05$ で推定された塩基が連続する領域の配列長が 6 塩基以上又は 6 塩基未満で、転写活性化能への正又は負の影響が推定された数を集計した。連続する塩基の転写活性化能へ与える影響は平均値を用いた。また、検定は chi-square test を実施した。

	positive		negative	
	度数	%	度数	%
6塩基未満	411	47.0	464	53.0
6塩基以上	1	4.4	22	95.7

$P < 0.0001$ (chi-square test)

表 3.25 : 転写活性化能への影響が 6 塩基以上連続する領域と転写因子結合配列

各遺伝子のプロモーター配列に対し検索された転写因子結合配列候補のうち、変異が転写活性化能へ影響のある塩基が 6 塩基以上連続する領域と重複するものを一覧に示す。

Gene	6塩基以上連続して影響のある配列			転写因子結合配列検索結果			
	Start	End	配列	偏回帰係数平均値	モチーフ名	配列	strand
ARL6IP5	-143	-134	CGCCCGCCCA	-0.312	V\$GKLF_Q4	CCGCCca	+
ARL6IP5	-143	-134	CGCCCGCCCA	-0.312	V\$ZFP161_04	ccCGCCcactc	+
ARL6IP5	-143	-134	CGCCCGCCCA	-0.312	V\$SNKX25_Q6	ccCACTCccaa	+
DDX5	-106	-101	CTATGC	0.149	V\$ZFP161_04	ggcctatG	-
DDX5	-106	-101	CTATGC	0.149	V\$AP2ALPHA_Q	tgccgag	-
DDX5	-82	-77	AGGTCA	-0.235	V\$ERALPHA_Q	taGGTCA	-
DDX5	-82	-77	AGGTCA	-0.235	V\$RHOX11_01	aggtcatA	-
DDX5	-82	-77	AGGTCA	-0.235	V\$CREBP1_01	ttaGGTCA	-
DDX5	-82	-77	AGGTCA	-0.235	V\$REVERBALP	cgattaGG	-
DDX5	-82	-77	AGGTCA	-0.235	V\$CRX_Q4_01	GATTAg	-
DDX5	-71	-61	CGGCTCCCAGC	-0.111	V\$IK_Q5	ggCTCCCa	-
DDX5	-71	-61	CGGCTCCCAGC	-0.111	V\$PAX_Q6	gcggtCC	-
DDX5	-71	-61	CGGCTCCCAGC	-0.111	V\$RHOX11_01	gcggtcC	-
DDX5	-32	-26	ACTATAA	-0.293	V\$RHOX11_01	gactataA	-
DDX5	-32	-26	ACTATAA	-0.293	V\$CDX2_Q5_0	tATAAA	-
DDX5	-32	-26	ACTATAA	-0.293	V\$HOXC13_01	agacTATA	-
DDX5	-32	-26	ACTATAA	-0.293	V\$HMX1_02	agactaTA	-
DDX5	-32	-26	ACTATAA	-0.293	V\$SMAD4_Q6_	cCAGAcT	-
DDX5	37	43	TCGGCTG	-0.097	V\$BEN_01	ttcGGCTG	-
DDX5	37	43	TCGGCTG	-0.097	V\$HDX_01	tcgcTGG	-
DDX5	37	43	TCGGCTG	-0.097	V\$CPHX_01	gcTGGTg	-
EEF1A1	-99	-92	CGCGGGGT	-0.257	V\$HMX1_02	cggggtAAACT	-
EEF1A1	-99	-92	CGCGGGGT	-0.257	V\$E2F_Q6_01	aggTGCGCg	-
EEF1A1	-76	-71	TGTCGT	0.044	V\$CPHX_01	agTGATGtcgt	-
EEF1A1	-61	-54	CCGCCTTT	-0.750	V\$MAF_Q6_01	cTGGCTccgcc	-
EEF1A1	-61	-54	CCGCCTTT	-0.750	V\$HOXD12_01	cctttTCCCG	-
EEF1A1	-61	-54	CCGCCTTT	-0.750	V\$HOXC13_01	ctttTCCCGa	-
EEF1A1	-61	-54	CCGCCTTT	-0.750	V\$NFAT1_Q4	tTTTCC	-
EEF1A1	-32	-24	CGTATATAA	-0.418	V\$HMX1_02	cgataTAAGT	-
EEF1A1	-32	-24	CGTATATAA	-0.418	V\$HOXC13_01	accgTATAa	-
EEF1A1	-32	-24	CGTATATAA	-0.418	V\$HOXB13_01	aaccgTATAa	-
EEF1A1	-32	-24	CGTATATAA	-0.418	V\$RUSH1A_02	atATAAGtc	-
EEF1A1	-32	-24	CGTATATAA	-0.418	V\$SNKX25_Q6	atatAAGTca	-
EEF1A1	-32	-24	CGTATATAA	-0.418	V\$GRE_C	gAGAACcgat	-
EEF1A1	-32	-24	CGTATATAA	-0.418	V\$TATA_01	gagaaccgTAT	-
EEF1A1	7	13	GCAACGG	-0.056	V\$RFX_Q6	tcGCAACgg	-
EEF1A1	7	13	GCAACGG	-0.056	V\$MYB_05	ttcgcAACGGg	-
EEF1A1	7	13	GCAACGG	-0.056	V\$CEBPA_Q6	ttTTTCGcaac	-
EEF1A1	7	13	GCAACGG	-0.056	V\$NF1_Q6	acgggtttgcc	-
GAPDH	-134	-126	TTTACGGGC	-0.324	V\$CHCH_01	CGGGCg	+
GAPDH	-134	-126	TTTACGGGC	-0.324	V\$RHOX11_01	ctagcGGTTt	+
GAPDH	-134	-126	TTTACGGGC	-0.324	V\$HOMEZ_01	ctactaGCGGT	+
GAPDH	-93	-87	GGA CTGG	-0.267	V\$CTF_01	CTGGCt	+
GAPDH	-93	-87	GGA CTGG	-0.267	V\$PAX_Q6	CTGGGactggc	+
GAPDH	-93	-87	GGA CTGG	-0.267	V\$IK_Q5	gctGGGACTg	+
GAPDH	-93	-87	GGA CTGG	-0.267	V\$RHOX11_01	cttggGCTGGg	+
GAPDH	-93	-87	GGA CTGG	-0.267	V\$P53_04	ggaCTGGCtga	+
GAPDH	-72	-66	AGGCGGG	-0.692	V\$CHCH_01	CGGGGt	+
GAPDH	-72	-66	AGGCGGG	-0.692	V\$AP2ALPHA_03	ctGGCGGgagg	+
GAPDH	-72	-66	AGGCGGG	-0.692	V\$IK_Q5	ggcGGGAGgc	+
GAPDH	-72	-66	AGGCGGG	-0.692	V\$CHCH_01	CGGGAg	+
GAPDH	-33	-18	TTTCTATAAATTGAGC	-0.397	V\$CDPCR1_01	tATAAAAttga	+
GAPDH	-33	-18	TTTCTATAAATTGAGC	-0.397	V\$HOXB13_01	gtttctATAAA	+
GAPDH	-33	-18	TTTCTATAAATTGAGC	-0.397	V\$HMX1_02	ggtttcTATAA	+
GAPDH	-33	-18	TTTCTATAAATTGAGC	-0.397	V\$TATA_01	ctATAAAAttga	+
GAPDH	-33	-18	TTTCTATAAATTGAGC	-0.397	V\$HMX1_02	tctataAATTG	+
GAPDH	-33	-18	TTTCTATAAATTGAGC	-0.397	V\$MEF2_03	ccggtttctat	+
GAPDH	-33	-18	TTTCTATAAATTGAGC	-0.397	V\$SIX1_01	ccccggTTTC	+
GAPDH	-16	-11	CGCAGC	-0.084	V\$RNF96_01	gCCCGCagcc	+
GAPDH	-16	-11	CGCAGC	-0.084	V\$ZIC1_05	gccccGAGCct	+
GAPDH	-16	-11	CGCAGC	-0.084	V\$AP2ALPHA_03	gcAGCCTcccg	+
GAPDH	2	9	TCTCTGCT	-0.112	V\$ZFP161_04	ttCGCTCtctg	+
GAPDH	2	9	TCTCTGCT	-0.112	V\$HOMEZ_01	cccgtTCGCT	+

表 3.25 (続き) : 転写活性化能への影響が 6 塩基以上連続する領域と転写因子結合配列

Gene	6塩基以上連続して影響のある配列			転写因子結合配列検索結果			
	Start	End	配列	偏回帰係数平均値	モチーフ名	配列	strand
NTS	-137	-131	AAAGGGG	-0.058	V\$IK_Q5	aagGGGAG	+
NTS	-137	-131	AAAGGGG	-0.058	V\$FPM315_01	aggGGAGG	+
NTS	-137	-131	AAAGGGG	-0.058	V\$NFAT1_Q4	GGAAAa	+
NTS	-137	-131	AAAGGGG	-0.058	V\$HELIOSA_0	gggAGGAG	+
NTS	-78	-72	CTGCAAA	0.113	V\$LEF1_Q5_0	gCAAAGa	+
NTS	-78	-72	CTGCAAA	0.113	V\$KAISO_01	atCCTGCa	+
NTS	-78	-72	CTGCAAA	0.113	V\$CEBPA_Q6	acatccTG	+
NTS	-53	-48	AATGAC	-1.095	V\$CPHX_01	tcaatgaC	+
NTS	-53	-48	AATGAC	-1.095	V\$HDX_01	tcaatgaC	+
NTS	-53	-48	AATGAC	-1.095	V\$SIX1_01	tcaatgaC	+
NTS	-53	-48	AATGAC	-1.095	V\$CPHX_01	atgacatC	+
NTS	-53	-48	AATGAC	-1.095	V\$HNF6_Q4	aaTCAATg	+
NTS	-53	-48	AATGAC	-1.095	V\$CDPCR1_01	aATCAATg	+
NTS	-53	-48	AATGAC	-1.095	V\$CREBP1_01	TGACAtca	+
NTS	-53	-48	AATGAC	-1.095	V\$HOXB13_01	caatcaAT	+
NTS	-30	-23	ATATATAG	-0.047	V\$TATA_01	atATATAg	+
NTS	-30	-23	ATATATAG	-0.047	V\$CDPCR1_01	tATATAta	+
NTS	-30	-23	ATATATAG	-0.047	V\$CEBPA_Q6	atatagGG	+
NTS	-30	-23	ATATATAG	-0.047	V\$TATA_01	atATATAt	+
NTS	-30	-23	ATATATAG	-0.047	V\$TATA_01	ttATATAt	+
NTS	-30	-23	ATATATAG	-0.047	V\$IRX2_01	cttatATA	+
PGK1	-95	-89	GGACAGC	-0.359	V\$BEN_01	CAGCGcca	+
PGK1	-83	-72	GGAGCAATGGCA	-0.271	V\$DBP_Q6	AGCAAAtg	+
PGK1	-83	-72	GGAGCAATGGCA	-0.271	V\$CTF_01	ATGGCa	+
PGK1	-83	-72	GGAGCAATGGCA	-0.271	V\$CDPCR1_01	gAGCAAtggc	+
PGK1	-83	-72	GGAGCAATGGCA	-0.271	V\$ZIC1_05	aatggCAGCG	+
PGK1	-83	-72	GGAGCAATGGCA	-0.271	V\$IK_Q5	ccaGGGAGca	+
PGK1	-83	-72	GGAGCAATGGCA	-0.271	V\$HELIOSA_0	ccaGGGAGca	+
PGK1	-49	-40	TGGCCAATAG	-0.381	V\$CTF_01	GTGGCc	+
PGK1	-49	-40	TGGCCAATAG	-0.381	V\$NFY_01	tgtggCCAAT	+
PGK1	-7	-2	GCCGGG	0.102	V\$CHCH_01	CGGGAA	+
PGK1	-7	-2	GCCGGG	0.102	V\$IK_Q5	gccGGGAagg	+
PGK1	-7	-2	GCCGGG	0.102	V\$BEN_01	CAGCGgcc	+
PGK1	14	26	AGGCGGGGTGTGG	-0.127	V\$CHCH_01	CGGGGt	+
PGK1	14	26	AGGCGGGGTGTGG	-0.127	V\$TBX5_01	cggGGTGTgg	+
PGK1	14	26	AGGCGGGGTGTGG	-0.127	V\$RHGX11_01	ggcggGGTGT	+
PGK1	14	26	AGGCGGGGTGTGG	-0.127	V\$MAZ_Q6_01	ggaGGCGGgg	+
PGK1	14	26	AGGCGGGGTGTGG	-0.127	V\$IK_Q5	tgcGGGAGgc	+
PGK1	14	26	AGGCGGGGTGTGG	-0.127	V\$CHCH_01	CGGGAg	+
RBBP5	-106	-96	GTCAGGCATAT	-0.697	V\$HOXB13_01	aggcaTATAt	-
RBBP5	-106	-96	GTCAGGCATAT	-0.697	V\$ERALPHA_Q6_	ttGGTCA	-
RPS12	-99	-92	GCCCGGCA	-0.404	V\$ZFP161_04	gcCCGGCactg	+
RPS12	-99	-92	GCCCGGCA	-0.404	V\$PAX_Q6	CCGGCactgcg	+
RPS12	-99	-92	GCCCGGCA	-0.404	V\$P53_Q4	aggCCCGGcac	+
RPS12	-99	-92	GCCCGGCA	-0.404	V\$ZIC1_05	cggcaCTGCGg	+
RPS12	-99	-92	GCCCGGCA	-0.404	V\$AP2ALPHA_	aaCTCCcagg	+
RPS12	-59	-49	GCACATGCGCA	-0.378	V\$IRX2_01	agtgcACATGc	+
RPS12	-59	-49	GCACATGCGCA	-0.378	V\$ZFP161_04	caTGCGCacag	+
RPS12	-59	-49	GCACATGCGCA	-0.378	V\$ZFP161_04	tgCGCACaggg	+
RPS12	29	34	GAGGCG	0.050	V\$ZFP161_04	gtCGCGCggag	+
RPS12	29	34	GAGGCG	0.050	V\$ZFP161_04	cgCGCGGaggc	+
RPS12	29	34	GAGGCG	0.050	V\$AP2ALPHA_	ggAGCGGagg	+
RPS12	29	34	GAGGCG	0.050	V\$E2F_Q6_01	gtcgCGCGGag	+

3.2.3.1.5 TATA ボックスへの変異と転写活性化能への影響

プロモーター領域に TATA ボックスが存在した 4 遺伝子 (DDX5、EEF1A1、GAPDH、NTS) について、本研究にて推定したプロモーター配列の各塩基の変異が転写活性化能へ与える影響を精査した。TATA ボックスは転写開始点を基準

とした-90+27 の領域に転写因子結合配列検索ソフトウェア MATCH を用いた V\$TATA_01 の検索及び配列 TATA[T/A][T/A]を検索した結果を TATA ボックスとした。DDX、EEF1A1、GAPDH、NTS はそれぞれ-30~-25、-38~-28、-30~-19、-30~-23 に TATA ボックスを有した。

図 3.23 に、各遺伝子のプロモーター配列中の TATA ボックスについて、各塩基に変異が入った際の転写活性化能への影響を示す。遺伝子 DDX、EEF1A1、GAPDH のプロモーター配列は、TATA ボックスに該当する領域に変異が入ることで、転写活性化能には大きな負の影響があると推定された。このことは、従来の研究結果の通り、TATA ボックスに変異が入ることで、基本転写因子である TATA ボックス結合タンパク質 (TBP) の結合が阻害され、転写活性化能に負の影響があることを検出できたものと考えられた。また、NTS 遺伝子のプロモーター配列中の TATA ボックスにおいても、-27、-25 の塩基を除いて、変異が入ることで転写活性化能に負の影響があることが検出された。

このように、本研究で確立した測定系及び測定データの解析結果より、既知の転写因子結合配列が検索可能であることが示された。

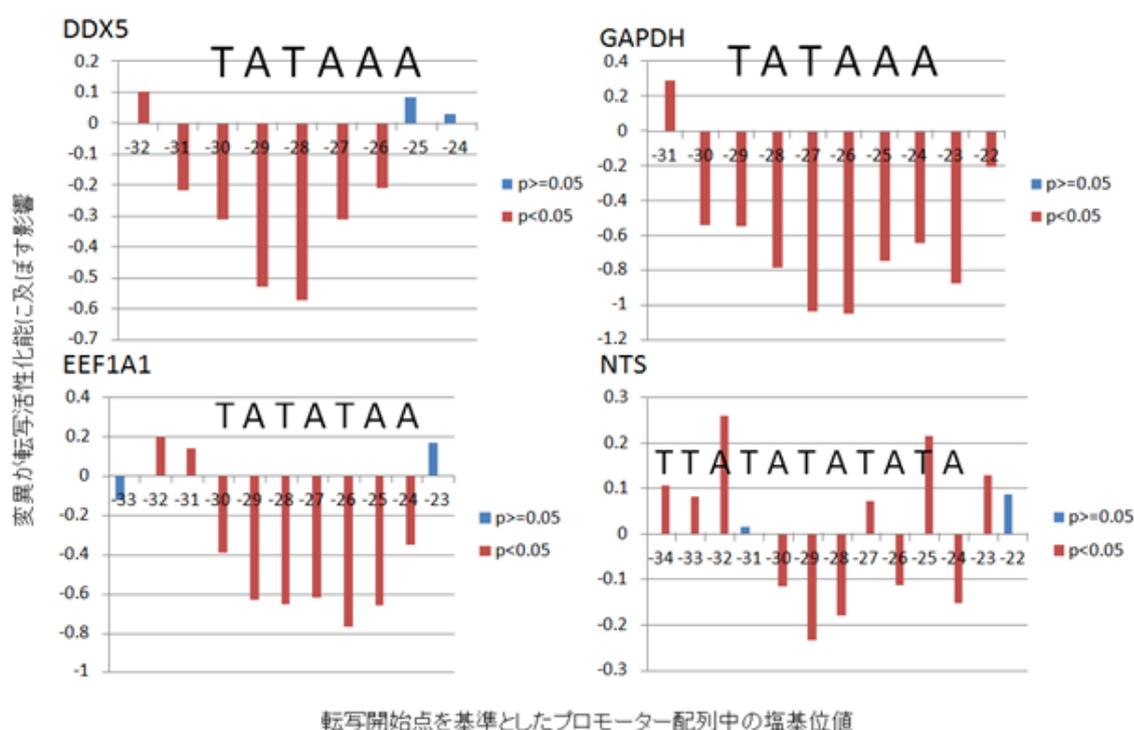


図 3.23: 既知の転写因子結合配列である TATA ボックスと推定された転写活性化能への影響
遺伝子 DDX5、EEF1A1、GAPDH、NTS のプロモーター配列について、TATA ボックスの各塩基に導入される変異が転写活性化能へ与える影響を示した。X 軸は転写開始点を基準としたときのプロモーター配列中の位値、Y 軸は変異が転写活性化能に与える影響を示している。赤の棒は推定された偏回帰係数が $p < 0.05$ 、青は $p \geq 0.05$ を示した。

NTS 遺伝子について、-27、-25 の塩基に変異が入ることで、転写活性化能へ

正の影響があると推定された。この領域について、より詳しく解析を行った。当該領域について各塩基に導入される変異パターンごとに転写活性化能へ与える影響を解析した (図 3.24)。

-27 の塩基について、野生型の塩基配列 T から A に置換された場合に、転写活性化能に対して、0.231 の正の影響があることが $p < 0.05$ で推定された。-25 の塩基について、野生型の塩基配列 T から A に置換されることで、転写活性化能に 0.105 の正の影響があることが $p < 0.05$ で推定された。-27 の塩基 T が A に変異すると、31~-27 の領域の野生型プロモーター配列 TATATA が TATAAA へ、-25 の塩基 T が A に変異すると-29~-24 の領域の野生型プロモーター配列 TATATA が TATAAA へ変化する。

これは、TATA ボックスが 5 塩基目の影響で構造が変化し、5 塩基目が T の場合に TBP と TATA ボックスの結合及び乖離の回転が速まることで、転写が抑制されるとの報告と一致する[64]。本研究により新たなシス因子の生成を予測できたとすれば興味深い。

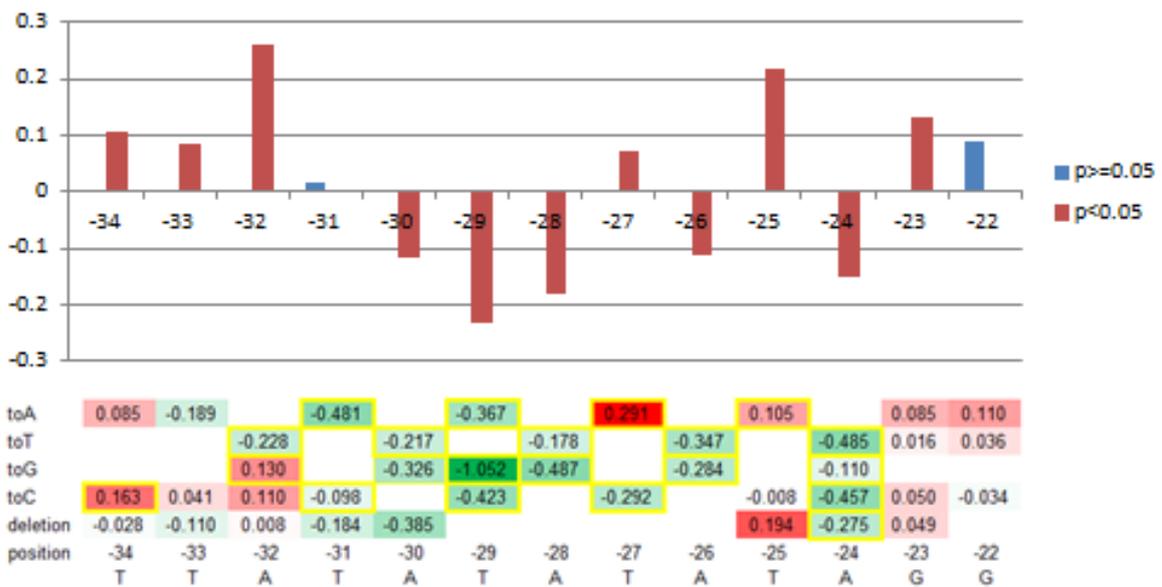


図 3.24: NTS 遺伝子の TATA ボックスへの変異が転写活性化能へ与える影響

NTS 遺伝子のプロモーター配列について、TATA ボックスの各塩基に導入される変異が転写活性化能へ与える影響 (上段)、その変異パターンごとの転写活性化能に与える影響 (下段) を示した。上段: X 軸は転写開始点を基準としたときのプロモーター配列中の位値、Y 軸は変異が転写活性化能に与える影響を示している。赤の棒は推定された偏回帰係数が $p < 0.05$ 、青は $p \geq 0.05$ を示した。下段: 横軸は転写開始点を基準としたときのプロモーター配列中の位値、縦軸は野生型塩基から各塩基又は欠損の変異パターンが生じた時に転写活性化能へ与える影響を示した。正の影響は赤、負の影響は緑。黄色枠で囲まれた値は、 $p < 0.05$ で偏回帰係数が推定されたことを示した。

3.2.3.1.6 各塩基の変異が転写活性化能へ与える影響の実験的確認

重回帰分析により推定した各塩基の変異が転写活性化能へ与える影響を実験的に検証した。野生型プロモーター配列に点変異を導入した点変異プロモーター配列を作成し、ルシフェラーゼアッセイによる転写活性化能の測定を行った。

以下に示すように、各塩基の変異が転写活性化能へ 8 塩基連続して正に影響していると推定された *EEF1A1* 遺伝子のプロモーター配列の+31~+38 領域について測定を行った。変異が転写活性化能へ 7 塩基連続で影響を与えると推定され、その配列内で転写活性化能への影響が正負に変化している領域として、*GAPDH* 遺伝子のプロモーター配列の-93~-87 領域についても同様に測定を行った。実験的に確認を行った領域の、各塩基の変異が転写活性化能へ与える影響の推定値を図 3.25 に示す。

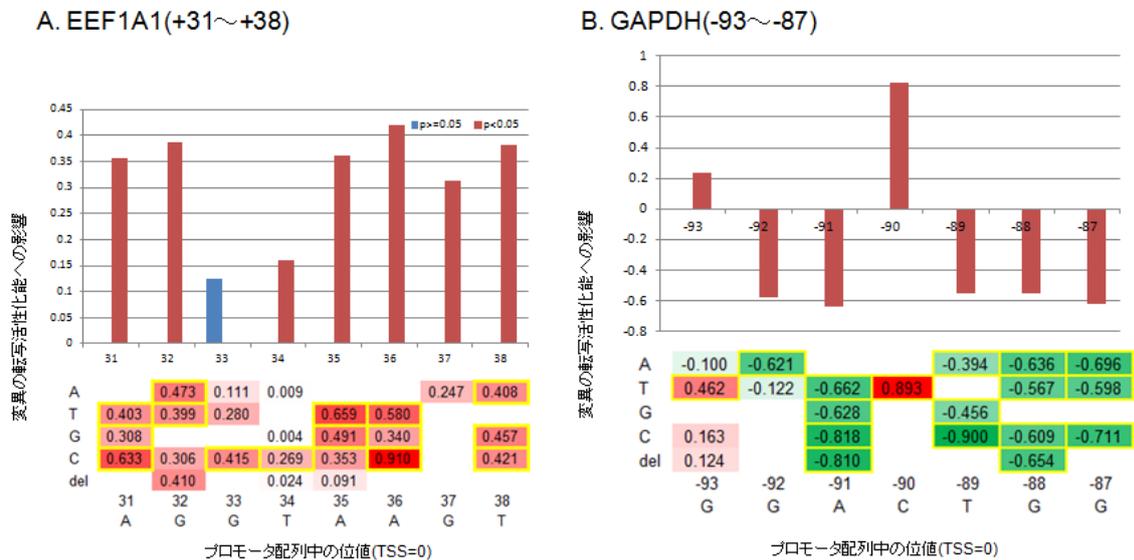


図 3.25 : EEF1A1+31~+38 領域、GAPDH-93-87 領域の変異が転写活性化能への影響

(A) EEF1A1 遺伝子の+31~+38 領域に変異及びその変異パターンが転写活性化能へ与える影響 (B) GAPDH 遺伝子の-93~-87 領域に変異及びその変異パターンが転写活性化能へ与える影響。

上段 : X 軸は転写開始点を基準としたときのプロモーター配列中の位値、Y 軸は変異が転写活性化能へ与える影響を示している。赤の棒は推定された偏回帰係数が $p < 0.05$ 、青は $p \geq 0.05$ を示した。下段 : 横軸は転写開始点を基準としたときのプロモーター配列中の位値、縦軸は野生型塩基から各塩基又は欠損の変異パターンが生じた時に転写活性化能へ与える影響を示した。正の影響は赤、負の影響は緑。黄色枠で囲まれた値は、 $p < 0.05$ で偏回帰係数が推定されたことを示した。

各領域について、inverse PCR による点変異導入法を用い、点変異プロモーター配列をクローニング、ルシフェラーゼアッセイによる転写活性化能の測定を行った。点変異プロモーター配列クローンを作成する際に用いたプライマーを表 3.26 に示す。

表 3. 26 : 点変異プロモーター配列のクローニングに用いたプライマー

GAPDH

Gene Name	location from TSS	mutation	primer sequenses
GAPDH	-93 - -87	7bp deletion	CTGAGCCTGGCGGGAGGCGG
GAPDH	-93	1 G deletion	GACTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-91	3 A deletion	GGCTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-90	4 C deletion	GGATGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-89	5 T deletion	GGACGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-88	6 G deletion	GGACTGCTGAGCCTGGCGGGAGGCGG
GAPDH	-93	1 G -> A	AGACTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-93	1 G -> T	TGACTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-93	1 G -> C	CGACTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-92	2 G -> A	GAAGTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-92	2 G -> T	GACTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-92	2 G -> C	GCACTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-91	3 A -> T	GGTCTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-91	3 A -> G	GGGCTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-91	3 A -> C	GGCCTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-90	4 C -> A	GGAATGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-90	4 C -> T	GGATTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-90	4 C -> G	GGAGTGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-89	5 T -> A	GGACAGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-89	5 T -> G	GGACGGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-89	5 T -> C	GGACCGGCTGAGCCTGGCGGGAGGCGG
GAPDH	-88	6 G -> A	GGACTAGCTGAGCCTGGCGGGAGGCGG
GAPDH	-88	6 G -> T	GGACTTGCTGAGCCTGGCGGGAGGCGG
GAPDH	-88	6 G -> C	GGACTCGCTGAGCCTGGCGGGAGGCGG
GAPDH	-87	7 G -> A	GGACTGACTGAGCCTGGCGGGAGGCGG
GAPDH	-87	7 G -> T	GGACTGTCTGAGCCTGGCGGGAGGCGG
GAPDH	-87	7 G -> C	GGACTGCCTGAGCCTGGCGGGAGGCGG
GAPDH	-	共通プライマー	CAGCCCAAGGTCTTGAGGCC

EEF1A1

Gene Name	location from TSS	mutation	primer sequenses
EEF1A1	31-38	8bp deletion	GCCGTGTGTGGTTCCCGCGG
EEF1A1	31	1A deletion	GGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	32	2G deletion	AGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	34	4T deletion	AGGAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	35	5A deletion	AGGTAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	37	7G deletion	AGGTAATGCCGTGTGTGGTTCCCGCGG
EEF1A1	38	8T deletion	AGGTAAGGCCGTGTGTGGTTCCCGCGG
EEF1A1	31	1 A -> T	TGGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	31	1 A -> G	GGGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	31	1 A -> C	CGGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	32	2 G -> A	AAGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	32	2 G -> T	ATGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	32	2 G -> C	ACGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	33	3 G -> A	AGATAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	33	3 G -> T	AGTTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	33	3 G -> C	AGCTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	34	4 T -> A	AGGAAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	34	4 T -> G	AGGGAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	34	4 T -> C	AGGCAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	35	5 A -> T	AGGTTAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	35	5 A -> G	AGGTGAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	35	5 A -> C	AGGTCAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	36	6 A -> T	AGGTATGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	36	6 A -> G	AGGTAGGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	36	6 A -> C	AGGTACGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	37	7 G -> A	AGGTAATGCCGTGTGTGGTTCCCGCGG
EEF1A1	37	7 G -> T	AGGTAATTGCCGTGTGTGGTTCCCGCGG
EEF1A1	37	7 G -> C	AGGTAAGTGCCGTGTGTGGTTCCCGCGG
EEF1A1	38	8 T -> A	AGGTAAGAGCCGTGTGTGGTTCCCGCGG
EEF1A1	38	8 T -> G	AGGTAAGGGCCGTGTGTGGTTCCCGCGG
EEF1A1	38	8 T -> C	AGGTAAGCGCCGTGTGTGGTTCCCGCGG
EEF1A1	-	共通プライマー	GTGTTCTGGCGCAAACCG

点変異プロモーター配列について、ルシフェラーゼアッセイを用いて転写活性化能を測定した結果を図 3.26 に示す。

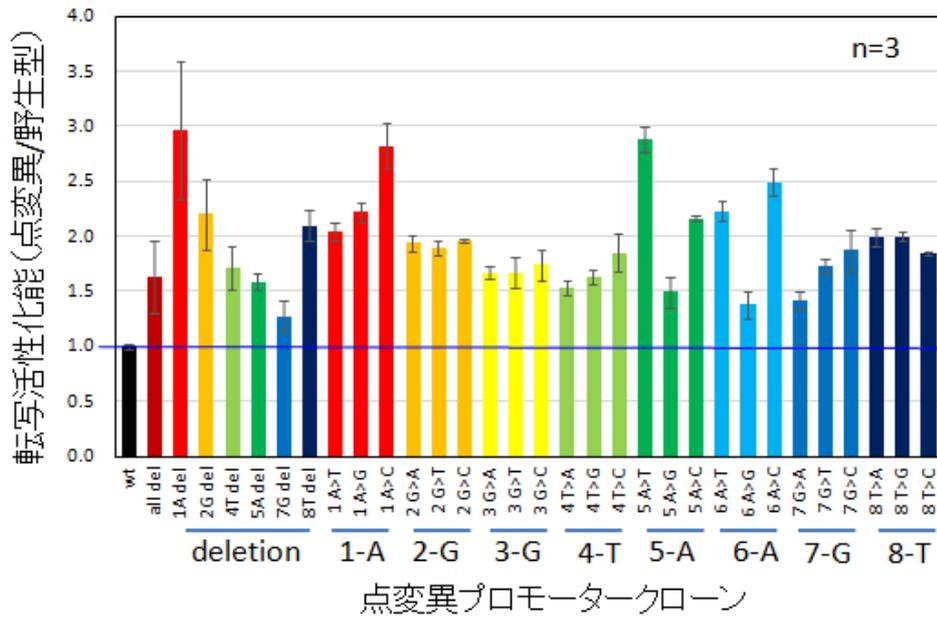
EEF1A1 (+31~+38) では、全ての点変異プロモーター配列で野生型プロモーター配列に比べて転写活性化能が増加していた。これは、重回帰分析で推定された各塩基の変異が転写活性化能へ与える影響と一致した。EEF1A1 遺伝子のプロモーター配列に対し、転写因子結合配列検索ソフトウェア MATCH にて転写因子結合配列を検索すると、転写を抑制する転写因子である HMX1 が検索された。HMX1 の結合配列に変異が入ることで、転写活性化能に正の影響があった可能性が考えられた。

GAPDH(-93~-87)では、塩基に変異が入ることで転写活性化能に正の影響が得られると推定された-93 及び-90 の塩基で、点変異プロモーター配列の転写活性化能が野生型プロモーター配列に比べて増加していた。変異パターンの転写活性化能へ与える影響の推定値から、-93 及び-90 ではそれぞれ G から T 及び C から T への変異することで、転写活性化能への正に影響すると推定されていた。点変異プロモーター配列の転写活性化能の測定結果は、-93 及び-90 ともに T への点変異プロモーター配列のみ転写活性化能が野生型プロモーター配列より上昇する結果で、重回帰分析による解析結果と一致していた。(図 3.26)

推定された各塩基の変異パターンが転写活性化能へ与える影響とそれぞれの点変異プロモーター配列の転写活性化能を比較した (図 3.27)。その結果、EEF1A1 (+31~+38) 及び GAPDH(-93~-87)では、相関係数としてそれぞれ 0.601 ($p<0.001$)、0.857 ($p<0.001$) と高い相関を示した。

これらの結果から、本研究で測定した多様な変異プロモーター配列の配列情報及び転写活性化能のデータから、重回帰分析にて推定した各塩基の変異が転写活性化能へ与える影響が、1 塩基レベルの情報粒度にて、実験的にも確認できたと考えられた。

EEF1A1 (+31~+38)



GAPDH(-93~87)

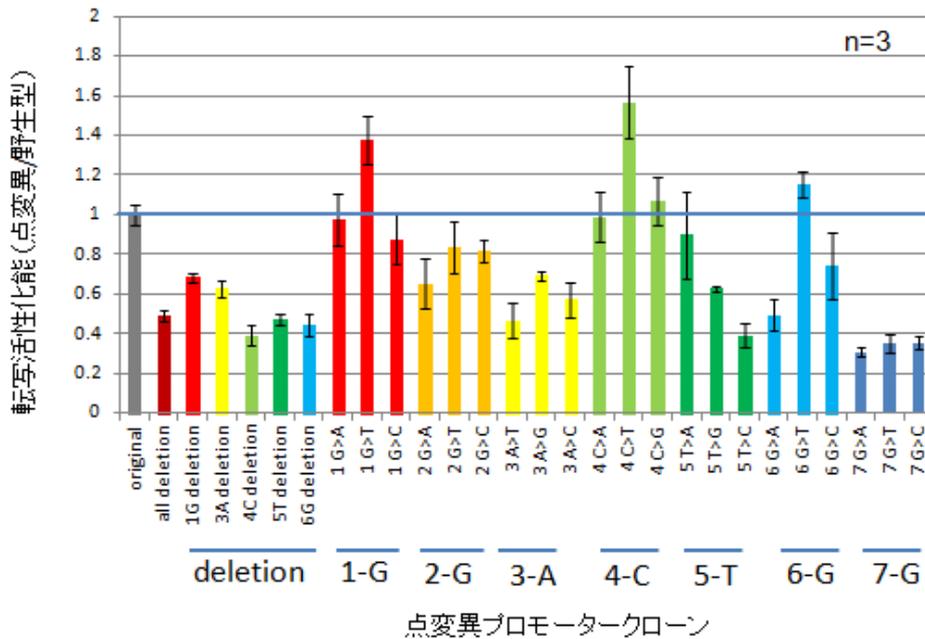


図 3.26 ; 点変異プロモーター配列の転写活性化能

野生型プロモーター配列及び EEF1A1+31~+38 領域、GAPDH-93-87 領域の各塩基に塩基置換又は欠損を挿入した点変異プロモーター配列のルシフェラーゼアッセイで測定した転写活性化能を示した。転写活性化能は n=3 で測定し、野生型プロモーター配列の平均値を 1 とした。棒グラフは平均値±標準偏差で示した。

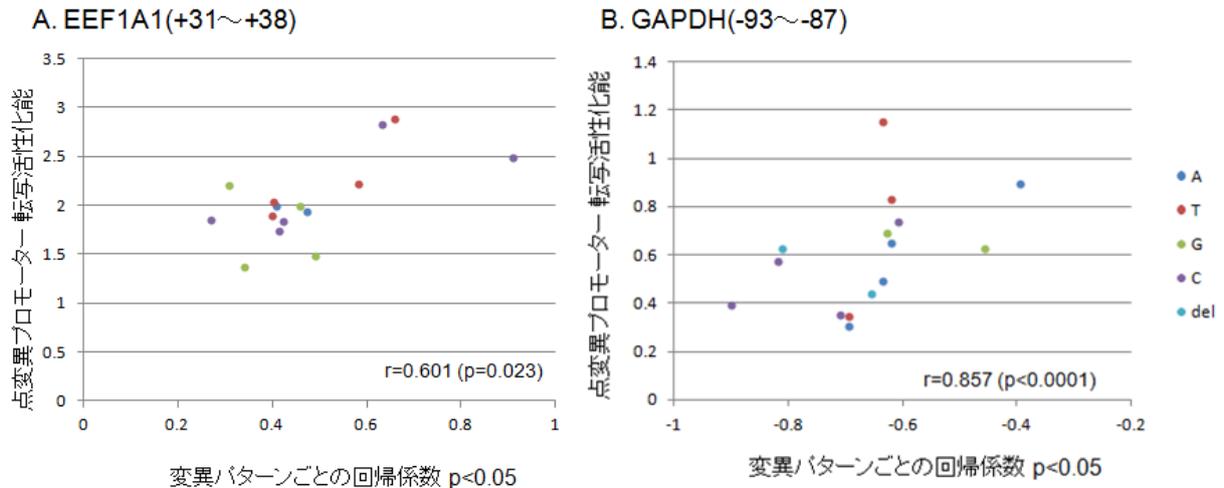


図 3.27：推定された転写活性化能への影響と点変異クローンを用いた変異導入後の転写活性化能重回帰分析の偏回帰係数として推定された各塩基への変異パターンが転写活性化能へ与える影響（X 軸）と実験的に点変異を挿入した点変プロモーター配列の転写活性化能の平均値（Y 軸）を散布図で示した。（A）EEF1A1+31～+38 領域、（B）GAPDH-93-87 領域。各プロットの色は、青：A への変異、赤：T への変異、緑：G への変異、紫：C への変異、水色：deletion を示した。相関係数は Pearson の積率相関係数を用いて算出した。

3.2.3.1.7 5 塩基ごと GC 含量と変異が転写活性化能へ与える影響

プロモーター配列中の各塩基周辺の GC 含量の違いによって、各塩基の変異が転写活性化能に与える影響が異なるのかを検証した。プロモーター配列中の連続する 5 塩基ごとの GC 含量と変異が転写活性化能に与える影響について解析した（表 3.27、図 3.28、図 3.29）。変異が転写活性化能に与える影響について、 $p < 0.05$ で負の影響があるもののうち、偏回帰係数の中央値 (-0.160) より小さい値を Negative1、それ以外を Negative2 とした。 $p < 0.05$ で正の影響があるもののうち、偏回帰係数の中央値 (0.128) より大きい値を Positive1、それ以外を Positive2 とした。

5 塩基ごとの GC 含量と変異が転写活性化能に与える影響を表 3.27 に示す。遺伝子のプロモーター配列ごとに 5 塩基ごとの GC 含量と変異が転写活性化能に与える影響について図 3.28、図 3.29 に示す。これらの解析結果から、5 塩基あたりの GC 含量 20-100 %については、GC 含量が高くなると変異が入ることによって影響の大きい塩基が濃縮される傾向があった（表 3.27、図 3.28）。GC 含量の高い領域は、CG メチル化の影響にもかかわらず進化的に保存されてきた領域であって、転写活性化能においても進化的に保存された領域であることを反映している可能性が示唆された。これは CpG island を有するプロモーター領域で転写因子の結合割合[3]や推定された cis-eQTL の存在割合[64]が上昇するという先行研究の内容とも一致する。

GC 含量 0 %については、変異の転写活性化能への負の影響が大きい塩基が濃

縮されていた。TATA ボックスを有する遺伝子においては、GC 含量が低い TATA ボックスに変異が入れば転写活性化能に大きな負の影響があることを反映しているのかもしれない (図 3.28)。各塩基の変異が転写活性化能へ与える正の影響は、負の影響とは異なり、GC 含量が中程度の領域(5 塩基あたりの GC 含量 40 %、60%) に濃縮されていた (表 3.27)。

表 3. 27 : 5 塩基ごとの GC 含量と変異が転写活性化能に与える影響

5 塩基あたりの GC 含量ごとに、変異が与える転写活性化能への影響を示した。変異が与える転写活性化能への影響は、次のカテゴリで示した。p<0.05 で負の影響があるもののうち、中央値より小さい値を Negative1、それ以外を Negative2。p<0.05 で正の影響があるもののうち、中央値より大きい値を Positive1、それ以外を Positive2。

	5塩基あたりの GC含量(%)	Negative1		Negative2		Positive2		Positive1		Others		合計
		度数	%	度数	%	度数	%	度数	%	度数	%	
8遺伝子	0	19	5.9	15	4.7	7	2.2	4	1.3	275	85.9	320
	20	39	3.1	40	3.2	30	2.4	29	2.3	1120	89.0	1258
	40	82	3.6	82	3.6	79	3.5	87	3.9	1930	85.4	2260
	60	124	4.9	138	5.4	100	3.9	123	4.8	2063	81.0	2548
	80	124	6.8	116	6.3	89	4.9	72	3.9	1433	78.1	1834
	100	51	9.9	42	8.1	23	4.4	17	3.3	385	74.3	518
ARL6IP5	0	1	2.1	4	8.5	1	2.1	1	2.1	40	85.1	47
	20	4	1.7	6	2.5	4	1.7	7	2.9	221	91.3	242
	40	7	1.7	18	4.4	15	3.7	10	2.4	360	87.8	410
	60	10	3.2	19	6.2	15	4.9	12	3.9	253	81.9	309
	80	4	4.2	17	17.9	4	4.2	1	1.1	69	72.6	95
	100	4	16.7	1	4.2	0	0.0	0	0.0	19	79.2	24
DDX5	0	3	42.9	0	0.0	0	0.0	0	0.0	4	57.1	7
	20	4	5.3	3	4.0	3	4.0	4	5.3	61	81.3	75
	40	10	4.6	4	1.8	12	5.5	9	4.1	183	83.9	218
	60	8	2.2	20	5.6	19	5.3	18	5.0	295	81.9	360
	80	14	4.2	20	6.0	24	7.2	8	2.4	266	80.1	332
	100	4	3.4	10	8.4	5	4.2	2	1.7	98	82.4	119
EEF1A1	0	4	33.3	0	0.0	0	0.0	1	8.3	7	58.3	12
	20	7	6.4	3	2.8	2	1.8	4	3.7	93	85.3	109
	40	14	5.8	4	1.6	6	2.5	18	7.4	201	82.7	243
	60	17	4.7	12	3.3	10	2.7	16	4.4	310	84.9	365
	80	17	5.8	10	3.4	8	2.7	9	3.1	249	85.0	293
	100	7	8.6	4	4.9	2	2.5	3	3.7	65	80.2	81
GAPDH	0	5	35.7	0	0.0	0	0.0	0	0.0	9	64.3	14
	20	7	13.7	3	5.9	0	0.0	1	2.0	40	78.4	51
	40	8	5.0	14	8.7	6	3.7	6	3.7	127	78.9	161
	60	18	4.8	20	5.4	13	3.5	12	3.2	310	83.1	373
	80	24	6.7	19	5.3	14	3.9	10	2.8	289	81.2	356
	100	12	11.7	10	9.7	6	5.8	3	2.9	72	69.9	103
NTS	0	2	1.9	6	5.8	5	4.9	1	1.0	89	86.4	103
	20	4	1.3	16	5.3	10	3.3	5	1.6	269	88.5	304
	40	14	3.7	18	4.7	19	5.0	19	5.0	311	81.6	381
	60	14	6.3	26	11.6	8	3.6	12	5.4	164	73.2	224
	80	8	11.6	10	14.5	8	11.6	5	7.2	38	55.1	69
	100	0	0.0	0	0.0	1	14.3	2	28.6	4	57.1	7
PGK1	0	0	0.0	0	0.0	0	0.0	0	0.0	40	100.0	40
	20	3	2.5	3	2.5	0	0.0	2	1.7	111	93.3	119
	40	8	3.4	6	2.6	6	2.6	4	1.7	209	89.7	233
	60	23	6.9	20	6.0	14	4.2	20	6.0	254	76.7	331
	80	18	6.6	20	7.3	17	6.2	15	5.5	203	74.4	273
	100	14	16.9	5	6.0	5	6.0	4	4.8	55	66.3	83
RBBP5	0	1	1.5	4	6.0	1	1.5	1	1.5	60	89.6	67
	20	2	1.0	4	1.9	9	4.3	5	2.4	190	90.5	210
	40	5	1.5	12	3.6	14	4.2	7	2.1	294	88.6	332
	60	16	6.0	7	2.6	13	4.9	12	4.5	218	82.0	266
	80	12	9.2	7	5.4	4	3.1	4	3.1	103	79.2	130
	100	4	9.8	8	19.5	1	2.4	1	2.4	27	65.9	41
RPS12	0	3	10.0	1	3.3	0	0.0	0	0.0	26	86.7	30
	20	8	5.4	2	1.4	2	1.4	1	0.7	135	91.2	148
	40	16	5.7	6	2.1	1	0.4	14	5.0	245	86.9	282
	60	18	5.6	14	4.4	8	2.5	21	6.6	259	80.9	320
	80	27	9.4	13	4.5	10	3.5	20	7.0	216	75.5	286
	100	6	10.0	4	6.7	3	5.0	2	3.3	45	75.0	60

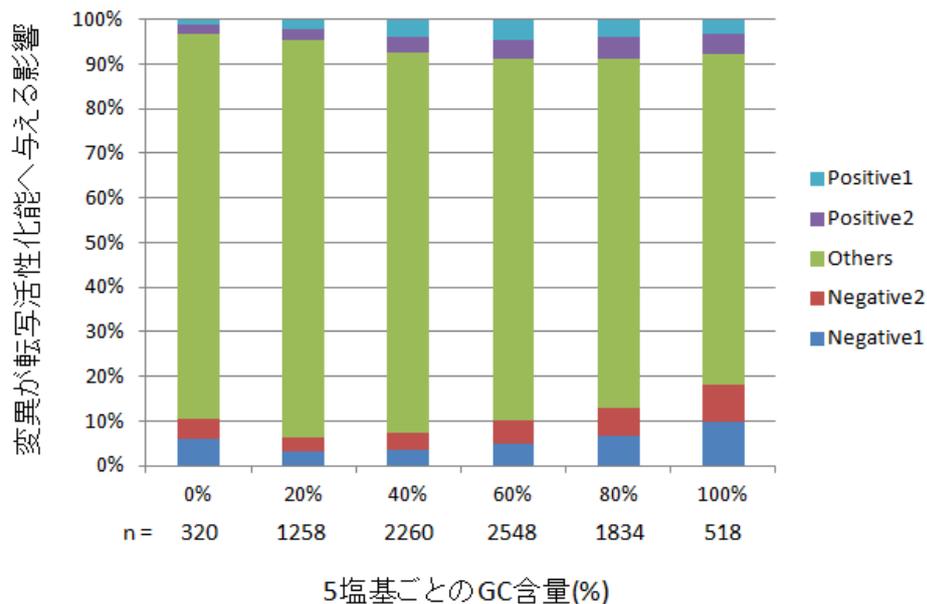


図 3.28: 5 塩基ごとの GC 含量と変異が転写活性化能に与える影響

プロモーター配列中の 5 塩基ごとの GC 含量 (X 軸) と各塩基への変異が転写活性化能に与える影響の割合 (Y 軸) を示す。変異が与える転写活性化能への影響は、次のカテゴリにて示した。p<0.05 で負の影響があるもののうち、中央値より小さい値を Negative1、それ以外を Negative2。p<0.05 で正の影響があるもののうち、中央値より大きい値を Positive1、それ以外を Positive2 とした。

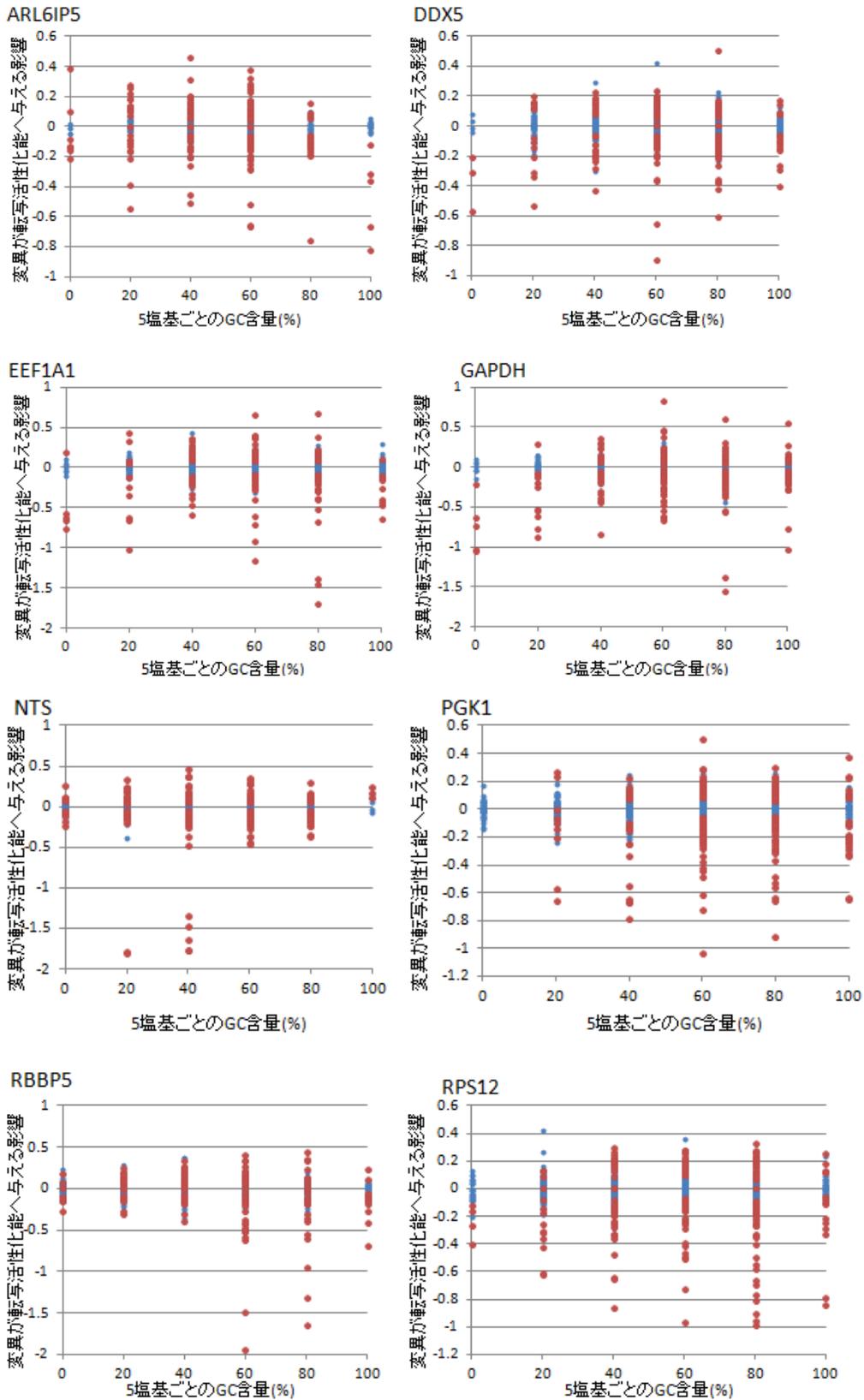


図 3.29: 5 塩基ごとの GC 含量と変異が転写活性化能に与える影響

プロモーター配列中の 5 塩基ごとの GC 含量 (X 軸) と各塩基への変異が転写活性化能へ与える影響 (Y 軸) に示した。赤 : $p < 0.05$ で推定された転写活性化能への影響、青 : $p \geq 0.05$ で推定された転写活性化能への影響

3.2.3.1.8 配列保存度と変異が転写活性化能へ与える影響

各塩基の配列保存度の違いにより、各塩基の変異が転写活性化能に与える影響が異なるのか調べるため、配列保存度の指標として PhastCons Score[50]を用い、塩基レベルの PhastCons Score と変異が転写活性化能に与える影響について解析した。変異が転写活性化能に与える影響について、 $p<0.05$ で負の影響があるもののうち、偏回帰係数の中央値 (-0.160) より小さい値を Negative1、それ以外を Negative2 とした。 $p<0.05$ で正の影響があるもののうち、偏回帰係数の中央値 (0.128) より大きい値を Positive1、それ以外を Positive2 とした。表 3.28 及び図 3.30、3.31 に、配列保存度ごとの変異が転写活性化能に与える影響を示す。

配列保存度が低い塩基と高い塩基で、変異が入った場合の転写活性化能への影響強度の分布は異なっていた ($p<0.0001$)。配列保存度が低い塩基では、変異が転写活性化能へ影響を与える塩基は 13.1 % (Negative1 2.5 %、Negative2 4.1 %、Positive1 3.1 %、Positive2 3.4 %) であったのに対し、配列保存度の高い塩基では 41.4 % (Negative1 18.5 %、Negative2 9.5 %、Positive1 7.6 %、Positive2 5.9 %) と濃縮されていた (表 3.28、図 3.30)。特に、負の強い影響を示す Negative1 の存在割合が 18.5 % であった。すべての遺伝子のプロモーター配列において、配列保存度の低い塩基と比較して、配列保存度の高い塩基で Negative1、Positive1 の存在割合が増加していた (表 3.28)。

表 3.28 : 配列保存度と変異が転写活性化能へ与える影響

各塩基の配列保存度を示す PhastCons Score ごとの、変異が与える転写活性化能への影響を示した。変異が与える転写活性化能への影響は、次のカテゴリにて示した。 $p<0.05$ で負の影響があるもののうち、中央値より小さい値を Negative1、それ以外を Negative2。 $p<0.05$ で正の影響があるもののうち、中央値より大きい値を Positive1、それ以外を Positive2。

	PhastCons	Negative1		Negative2		Positive2		Positive1		Others		合計
		度数	%	度数	%	度数	%	度数	%	度数	%	
8遺伝子	>0.5	252	18.5	130	9.5	80	5.9	103	7.6	799	58.6	1364
	<=0.5	187	2.5	303	4.1	248	3.4	229	3.1	6407	86.9	7374
ARL6IP5	>0.5	13	8.1	17	10.6	10	6.2	8	5.0	113	70.2	161
	<=0.5	17	1.8	48	5.0	29	3.0	23	2.4	849	87.9	966
DDX5	>0.5	30	9.7	19	6.1	17	5.5	22	7.1	223	71.7	311
	<=0.5	13	1.6	38	4.8	46	5.8	19	2.4	684	85.5	800
EEF1A1	>0.5	45	23.2	12	6.2	5	2.6	23	11.9	109	56.2	194
	<=0.5	21	2.3	21	2.3	23	2.5	28	3.1	816	89.8	909
GAPDH	>0.5	35	23.2	12	8.0	8	5.3	9	6.0	87	57.6	151
	<=0.5	39	4.3	54	6.0	31	3.4	23	2.5	760	83.8	907
NTS	>0.5	26	10.4	40	16.0	24	9.6	19	7.6	141	56.4	250
	<=0.5	16	1.9	36	4.3	27	3.2	25	3.0	734	87.6	838
PGK1	>0.5	22	40.7	6	11.1	4	7.4	4	7.4	18	33.3	54
	<=0.5	44	4.3	48	4.7	38	3.7	41	4.0	854	83.3	1025
RBBP5	>0.5	27	23.9	12	10.6	9	8.0	9	8.0	56	49.6	113
	<=0.5	13	1.4	30	3.2	33	3.5	21	2.3	836	89.6	933
RPS12	>0.5	54	41.5	12	9.2	3	2.3	9	6.9	52	40.0	130
	<=0.5	24	2.4	28	2.8	21	2.1	49	4.9	874	87.8	996

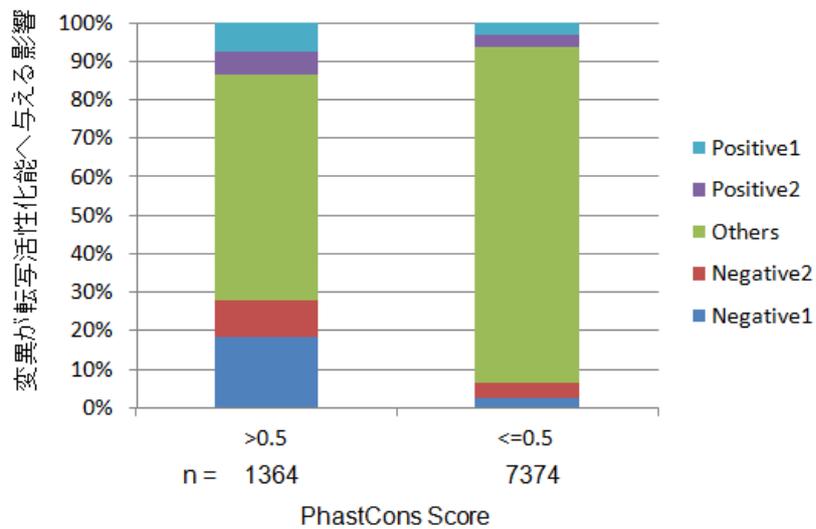
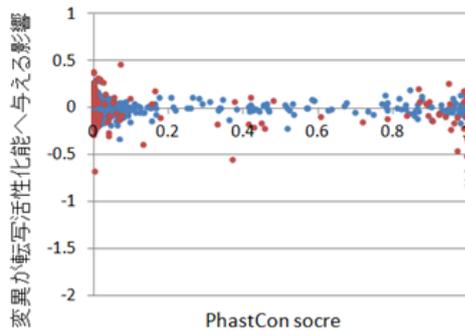


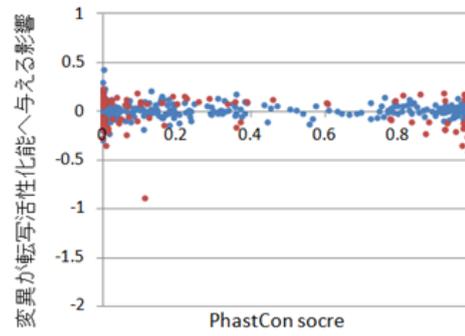
図 3.30: 配列保存度と変異が転写活性化能へ与える影響

各塩基の配列保存度と変異の転写活性化能へ与える影響を棒グラフで示した。配列保存度は、PhastCons Score>0.5、PhastCons Score <=0.5 を用いた。変異が与える転写活性化能への影響は、次のカテゴリにて示した。p<0.05 で負の影響があるもののうち、中央値より小さい値を Negative1、それ以外を Negative2。p<0.05 で正の影響があるもののうち、中央値より大きい値を Positive1、それ以外を Positive2 とした。

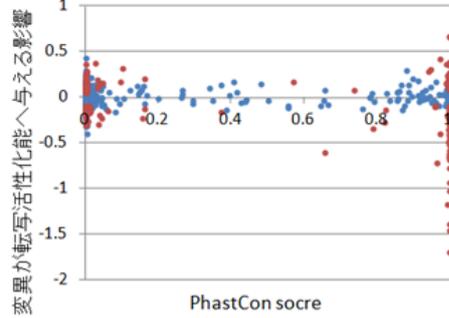
ARL6IP5



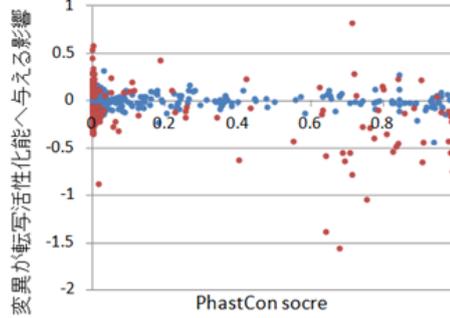
DDX5



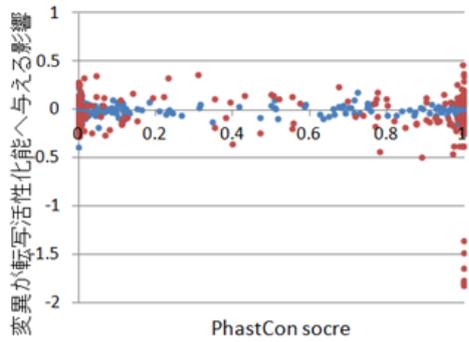
EEF1A1



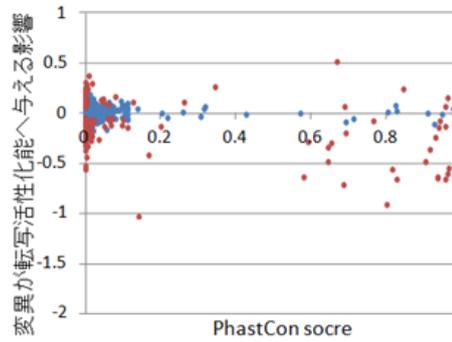
GAPDH



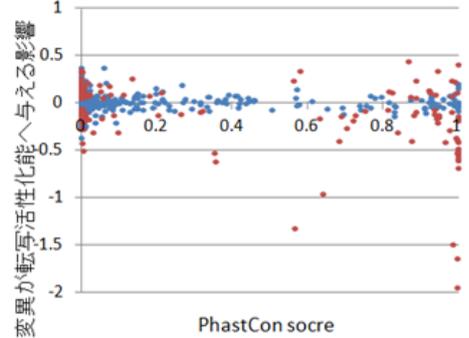
NTS



PGK1



RBBP5



RPS12

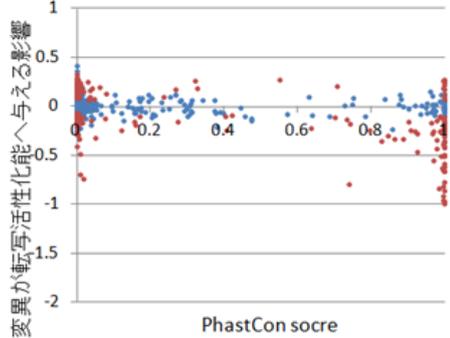


図 3.31: 配列保存度と変異が転写活性能へ与える影響

遺伝子のプロモーター配列ごとに、塩基ごとの配列保存度を示す PhastCons Score (X 軸) と変異が与える転写活性能への影響 (Y 軸) を散布図として示した。赤 : $p < 0.05$ で推定された転写活性能への影響、青 : $p \geq 0.05$ で推定された転写活性能への影響。

3.2.3.1.7 では、5 塩基あたりの GC 含量が高い又は低いグループで Negative1 が濃縮されていた。これは、GC 含量が高い又は低いグループでは、配列保存度が高く、進化的に転写活性化能にも重要な影響を有しているためと考えられた。そこで、配列保存度が高く、GC 含量が高い又は低いグループに、変異が転写活性化能へ与える負の影響の大きい塩基が濃縮されているか調べた (図 3.32)。

その結果、期待していた通り、配列保存度の高い塩基では、配列保存度の低い塩基と比較して、GC 含量の高い又は低い塩基に、変異の転写活性化能への負の影響が大きい塩基が濃縮されていた。配列保存度の低い塩基では、GC 含量が高い塩基に、変異の転写活性化能への影響が大きい塩基が濃縮されていた。

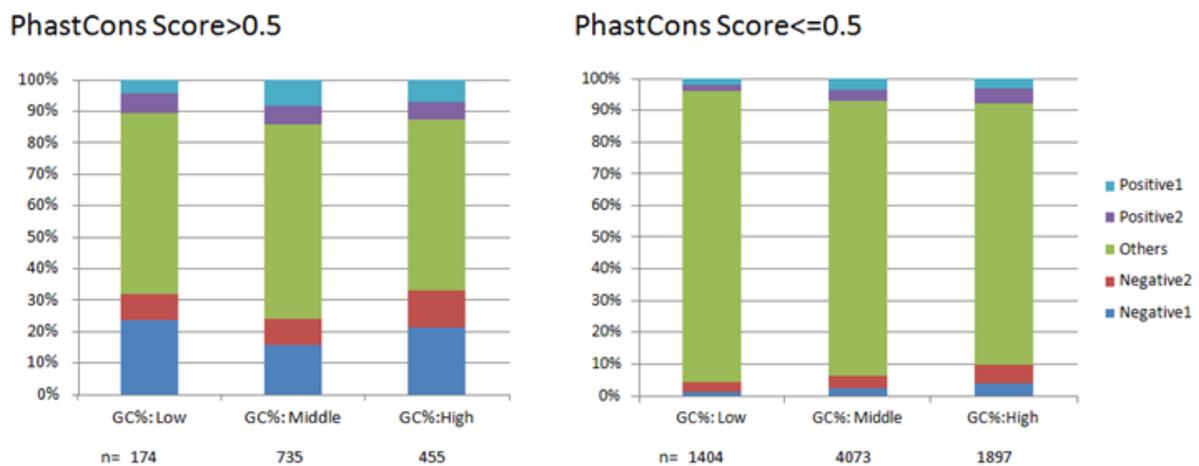


図 3.32: 配列保存度別の 5 塩基あたりの GC 含量と変異の転写活性化能へ与える影響

各塩基の配列保存度別に、5 塩基あたりの GC 含量と変異の転写活性化能へ与える影響を棒グラフで示した。配列保存度は、PhastCons Score > 0.5、PhastCons Score ≤ 0.5 を用いた。5 塩基あたりの GC 含量は、GC%: Low : GC 含量として 0 % 又は 20 %、GC%: Middle : GC 含量として 40 % 又は 60 %、GC%: High : GC 含量として 80 % 又は 100 % とした。変異が与える転写活性化能への影響は、次のカテゴリにて示した。p < 0.05 で負の影響があるもののうち、中央値より小さい値を Negative1、それ以外を Negative2。p < 0.05 で正の影響があるもののうち、中央値より大きい値を Positive1、それ以外を Positive2 とした。

プロモーター配列中の塩基の位置ごとに配列保存度と変異の転写活性化能へ与える影響を評価した (図 3.33)。配列保存度の高い塩基のうち、各塩基の変異が転写活性化能に与える影響の大きい塩基は、プロモーター配列中の転写開始点を基点(0)としたとき、-200 塩基程度から下流に主に分布していた。

これは、転写開始点近傍に存在するコアプロモーター領域及び転写開始点上流 200 塩基程度までに存在する近位プロモーター領域に TATA ボックスや GC ボックスなどの進化的に保存されている配列が存在し、転写制御に重要な役割を担っているとこのこれまでの通念と一致する。また、eQTL が転写開始点近傍

のコアプロモーター領域にエンリッチされていることが推定され、その領域の配列保存度が高いことを示した報告とも一致した[63]。

一方で、-200 よりも上流では、配列保存度は高い塩基は存在しているが、そのうち各塩基の変異が転写活性化能へ与える影響が大きい塩基は少なかった(図 3.33)。配列保存度から、これらの塩基も重要な生物学的機能を担っている可能性が考えられるが、それは一次配列がコードする転写活性化能の観点からのものではないと思われた。

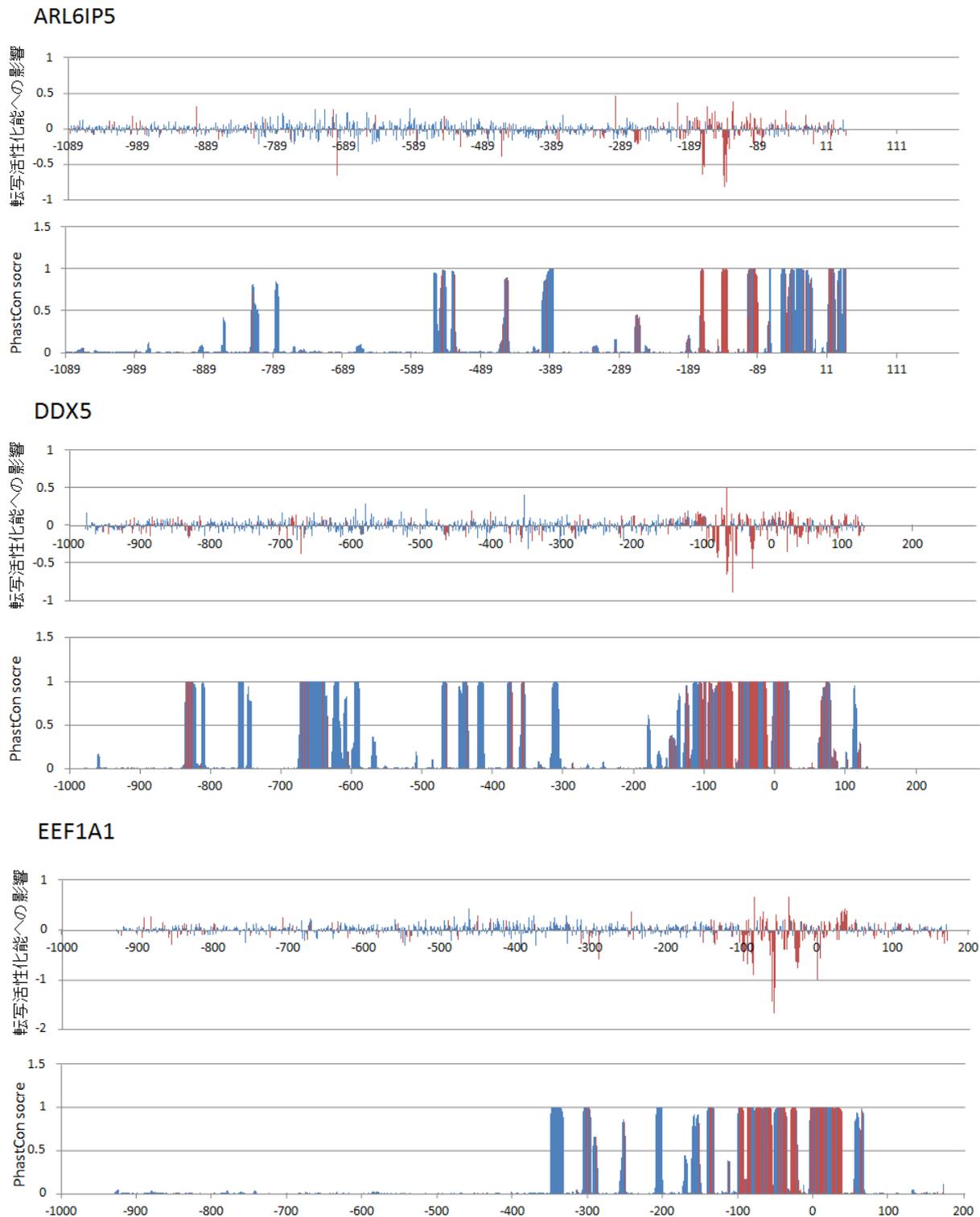
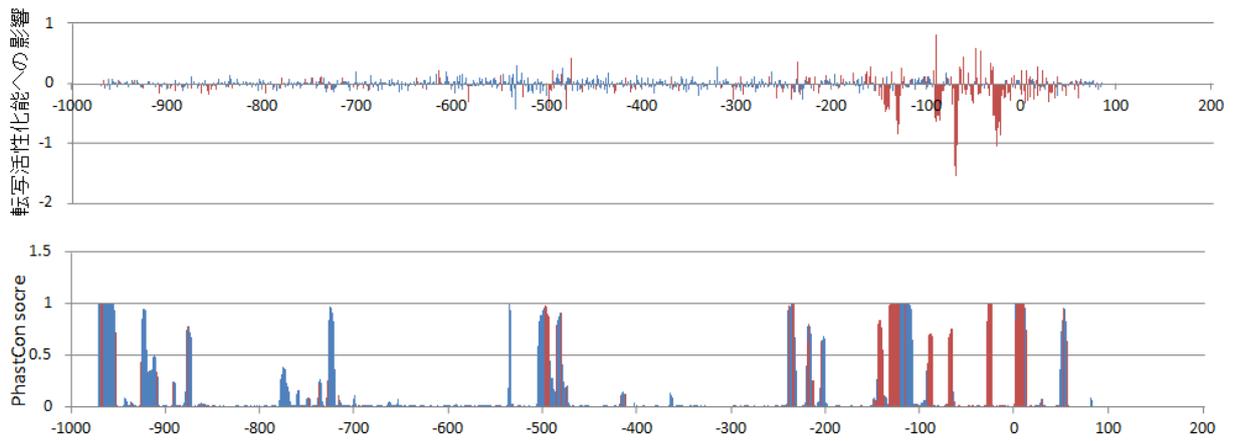


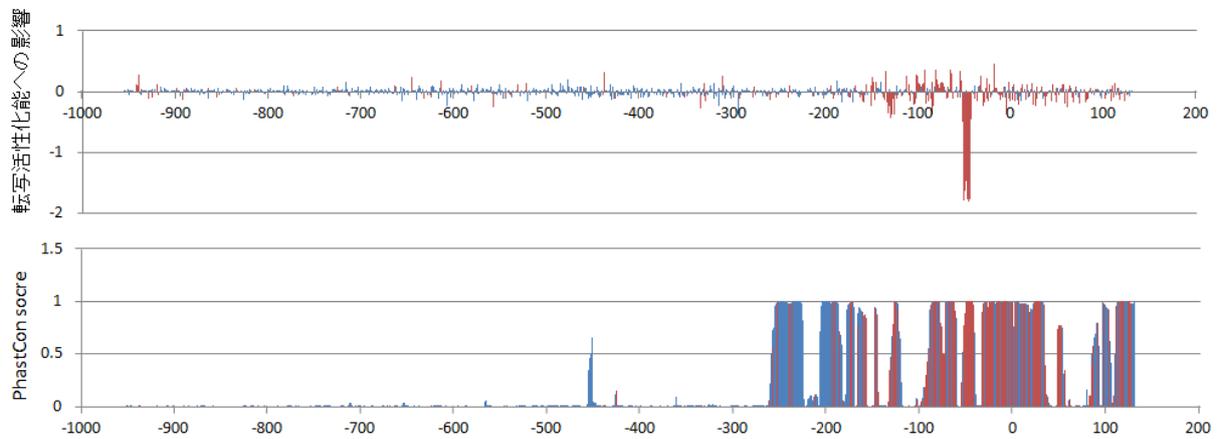
図 3.33: プロモーター配列内の各塩基における配列保存度と変異の転写活性化能への影響

プロモーター配列内の塩基の位値 (X 軸) ごとに、変異が入った場合の転写活性化能への影響 (Y 軸上段) と配列保存度 (Y 軸下段) に棒グラフとして示した。配列保存度は PhastCons Score を用いた。赤 : $p < 0.05$ で推定された変異の転写活性化能への影響、青 : $p \geq 0.05$ で推定された変異の転写活性化能への影響

GAPDH



NTS



PGK1

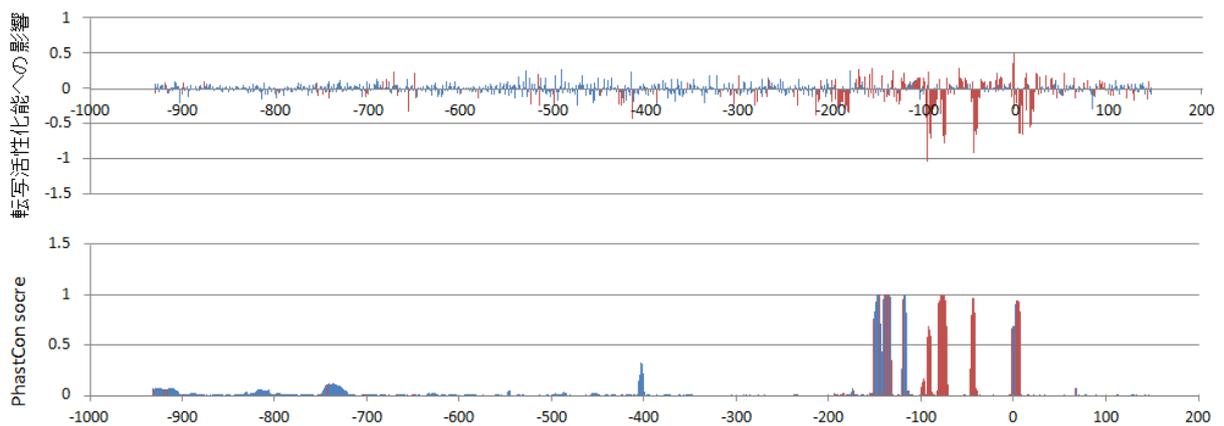
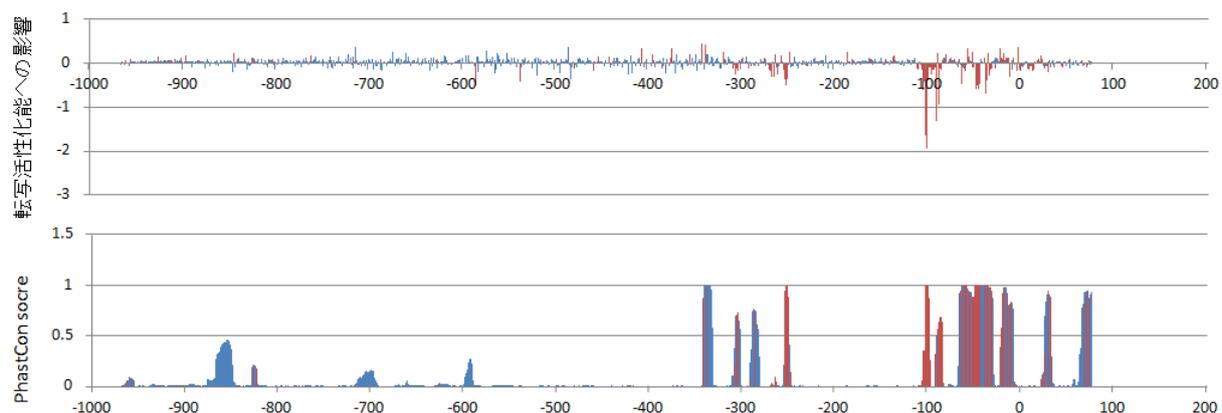


図 3.33 (続き) : プロモーター配列内の各塩基における配列保存度と変異の転写活性化能への影響

プロモーター配列内の塩基の位置 (X 軸) ごとに、変異が入った場合の転写活性化能への影響 (Y 軸上段) と配列保存度 (Y 軸下段) に棒グラフとして示した。配列保存度は PhastCons Score を用いた。赤 : $p < 0.05$ で推定された変異の転写活性化能への影響、青 : $p \geq 0.05$ で推定された変異の転写活性化能への影響

RBBP5



RPS12

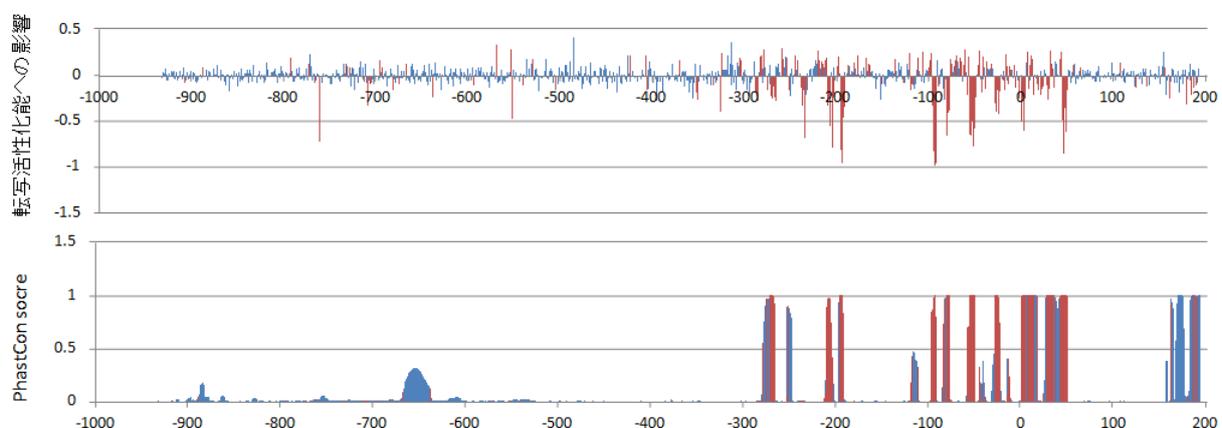


図 3.33 (続き) : プロモーター配列内の各塩基における配列保存度と変異の転写活性化能への影響

プロモーター配列内の塩基の位値 (X 軸) ごとに、変異が入った場合の転写活性化能への影響 (Y 軸上段) と配列保存度 (Y 軸下段) に棒グラフとして示した。配列保存度は PhastCons Score を用いた。赤 : $p < 0.05$ で推定された変異の転写活性化能への影響、青 : $p \geq 0.05$ で推定された変異の転写活性化能への影響

3.2.3.1.9 dbSNP との比較

dbSNP[53]に登録されている遺伝子のプロモーター配列内の既知の SNP について、転写活性化能に影響が強い塩基に生じている可能性を検討した (表 3.29、図 3.34)。dbSNP に登録された SNP であるか否かで、変異が転写活性化能へ与える影響強度の分布に有意な差は検出されなかった ($p < 0.293$)。ただし、dbSNP に登録されていない塩基では、変異が転写活性化能へ影響を与える塩基は 17.4 % (Negative1 5.0 %、Negative2 4.9 %、Positive1 3.8 %、Positive2 3.7 %) であったのに対し、SNP として登録のある塩基では 22.4 % (Negative1 7.5 %、Negative2 6.1 %、Positive1 3.7 %、Positive2 5.1 %) と濃縮される傾向があった (表 3.29、図 3.34)。dbSNP に SNP として登録されている塩基は、転写活性化能へ影響を与えている可能性が考えられた。

dbSNP でプロモーター配列中に位置していた SNP のうち、転写活性化能へ強い負の影響のある Negative1、転写活性化能に強い正の影響がある Positive1 に位置するものも見出された (表 3.30)。rs1991401 は DDX5 の-28 の逆鎖に A/G と登録された SNP である。この塩基に変異が入ることで偏回帰係数として-0.583 (p=1.25E-39)、T より C の変異パターンで-0.467 (p=5.96E-17) の転写活性化能への影響があった。rs1991401 は The Genotype-Tissue Expression (GTEx) [62]データを用いた Single-Tissue eQTL 解析より、各組織中で DDX5 遺伝子の発現に負の影響を与える cis-eQTL と考えられた (表 3.31)。組織中においても、rs1991401 が転写活性化能へ影響を与え、遺伝子発現の変化に寄与する可能性が示唆された。rs45461995 は GAPDH の-130 に C/T と登録された SNP である。この塩基に変異が入ることで転写活性化能への影響は偏回帰係数として-0.678(p=1.62E-40)、C より T の変異パターンの転写活性化能への影響は-0.512(p=7.90E-19)であった。rs45461995 は、rs1991401 と同様の解析では、cis-eQTL として特定されなかった。

表 3.29 : dbSNP に登録のある SNP と変異が転写活性化能に与える影響

dbSNP に SNP の登録がある塩基か否か別に、変異が与える転写活性化能への影響を分割表として示した。変異が与える転写活性化能への影響のカテゴリは、 $p < 0.05$ で負の影響があるもののうち、中央値より小さい値を Negative1、それ以外を Negative2。 $p < 0.05$ で正の影響があるもののうち、中央値より大きい値を Positive1、それ以外を Positive2 とした。

		Nrgative1		Negative2		Positive2		Positive1		Others		合計
		度数	%	度数	%	度数	%	度数	%	度数	%	
8遺伝子	SNP有	16	7.5	13	6.1	11	5.1	8	3.7	166	77.6	214
	SNP無	423	5.0	420	4.9	317	3.7	324	3.8	7040	82.6	8524
ARL6IP5	SNP有	1	5.0	1	5.0	0	0.0	2	10.0	16	80.0	20
	SNP無	29	2.6	64	5.8	39	3.5	29	2.6	946	85.5	1107
DDX5	SNP有	2	6.3	2	6.3	1	3.1	1	3.1	26	81.3	32
	SNP無	41	3.8	55	5.1	62	5.8	40	3.7	881	81.7	1079
EEF1A1	SNP有	2	8.3	0	0.0	0	0.0	1	4.2	21	87.5	24
	SNP無	64	5.9	33	3.1	28	2.6	50	4.6	904	83.8	1079
GAPDH	SNP有	2	7.1	1	3.6	2	7.1	1	3.6	22	78.6	28
	SNP無	72	7.0	65	6.3	37	3.6	31	3.0	825	80.1	1030
NTS	SNP有	0	0.0	5	26.3	0	0.0	0	0.0	14	73.7	19
	SNP無	42	3.9	71	6.6	51	4.8	44	4.1	861	80.5	1069
PGK1	SNP有	2	9.1	1	4.6	1	4.6	1	4.6	17	77.3	22
	SNP無	64	6.1	53	5.0	41	3.9	44	4.2	855	80.9	1057
RBBP5	SNP有	0	0.0	2	9.5	2	9.5	1	4.8	16	76.2	21
	SNP無	40	3.9	40	3.9	40	3.9	29	2.8	876	85.5	1025
RPS12	SNP有	7	14.6	1	2.1	5	10.4	1	2.1	34	70.8	48
	SNP無	71	6.6	39	3.6	19	1.8	57	5.3	892	82.8	1078

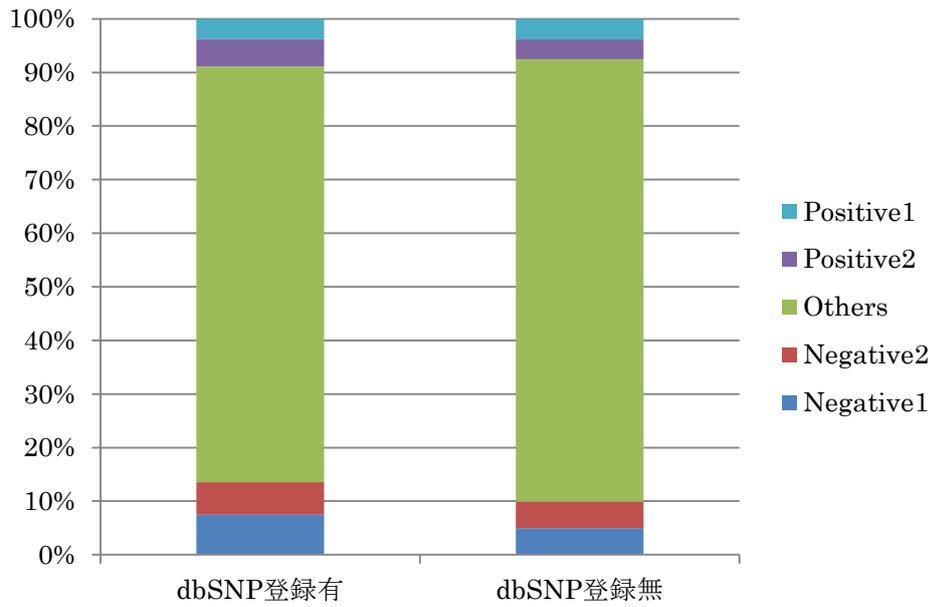


図 3.34 各塩基の dbSNP 登録有無と変異が転写活性化能へ与える影響

dbSNP に SNP の登録がある塩基か否か別に、変異が与える転写活性化能への影響を棒グラフとして示した。変異が与える転写活性化能への影響のカテゴリは、 $p < 0.05$ で負の影響があるもののうち、中央値より小さい値を Negative1、それ以外を Negative2。 $p < 0.05$ で正の影響があるもののうち、中央値より大きい値を Positive1、それ以外を Positive2 とした。

表 3.30 : $p < 0.05$ で変異が転写活性化能に与える影響が推定された塩基と dbSNP に登録のある SNP

$p < 0.05$ で変異が転写活性化能に与える影響が推定された塩基のうち、dbSNP に SNP として登録のある塩基を一覧として示す。

Gene	position	塩基	変異の転写活性化能への影響	カテゴリー	p値	SNP Strand	RS ID
ARL6IP5	-202	A	-0.152	Negative2	1.349E-04	+	rs375096389
ARL6IP5	-168	A	-0.453	Negative1	1.800E-26	+	rs185429527
ARL6IP5	-153	G	0.249	Positive1	4.064E-04	+	rs79392449
ARL6IP5	-65	T	0.129	Positive1	6.200E-06	+	rs146642345
DDX5	-215	A	0.134	Positive1	2.328E-03	+	rs78841668
DDX5	-141	C	0.096	Positive2	2.601E-02	+	rs9905596
DDX5	-77	A	-0.336	Negative1	1.010E-31	+	rs374811154
DDX5	-28	T	-0.573	Negative1	1.270E-39	+	rs1991401
DDX5	65	C	-0.084	Negative2	1.779E-02	+	rs377373286
DDX5	124	G	-0.126	Negative2	6.501E-04	+	rs201924128
EEF1A1	-476	A	-0.201	Negative1	1.048E-02	+	rs376145722
EEF1A1	26	A	0.240	Positive1	1.600E-05	+	rs188135961
EEF1A1	93	C	-0.161	Negative1	8.888E-03	+	rs148339214
GAPDH	-613	T	0.092	Positive2	2.318E-02	+	rs45612631
GAPDH	-130	C	-0.678	Negative1	1.620E-40	+	rs45461995
GAPDH	-64	T	0.241	Positive1	3.780E-06	+	rs111720901
GAPDH	-44	C	-0.162	Negative1	1.130E-07	+	rs45561238
GAPDH	30	C	-0.143	Negative2	5.560E-08	+	rs367946267
GAPDH	37	T	0.114	Positive2	1.886E-03	+	rs3199979
NTS	-117	G	-0.082	Negative2	2.366E-02	+	rs367713513
NTS	-21	G	-0.099	Negative2	2.216E-02	+	rs117953598
NTS	67	G	-0.090	Negative2	3.233E-02	+	rs142420800
NTS	82	G	-0.144	Negative2	4.840E-11	+	rs369679972
NTS	83	G	-0.119	Negative2	2.651E-02	+	rs373597970
PGK1	-931	T	-0.068	Negative2	1.508E-02	+	rs181500091
PGK1	-305	C	-0.221	Negative1	2.502E-04	+	rs372898993
PGK1	3	G	-0.651	Negative1	3.680E-155	+	rs377286545
PGK1	21	G	0.223	Positive1	2.660E-09	+	rs182927805
PGK1	95	C	0.081	Positive2	4.449E-02	+	rs192936390
RBBP5	-627	A	0.094	Positive2	4.607E-02	-	rs6698188
RBBP5	-161	T	0.094	Positive2	1.025E-02	+	rs189633547
RBBP5	-2	C	0.357	Positive1	3.050E-23	+	rs12033568
RBBP5	29	G	-0.067	Negative2	4.598E-02	+	rs11240362
RBBP5	44	C	-0.075	Negative2	2.332E-03	+	rs375728850
RPS12	-675	C	-0.093	Negative2	1.328E-02	+	rs112447165
RPS12	-279	G	0.277	Positive1	1.642E-02	+	rs188213983
RPS12	-204	C	-0.244	Negative1	1.180E-05	+	rs58748092
RPS12	-62	A	0.116	Positive2	3.545E-02	+	rs57172226
RPS12	-12	T	0.067	Positive2	2.420E-02	+	rs41286194
RPS12	-3	G	-0.257	Negative1	9.890E-06	+	rs113122367
RPS12	0	C	-0.500	Negative1	1.100E-09	+	rs145134922
RPS12	9	T	0.109	Positive2	5.491E-03	+	rs112860722
RPS12	13	G	-0.213	Negative1	8.900E-12	+	rs11543926
RPS12	48	A	-0.362	Negative1	3.890E-42	+	rs146445896
RPS12	57	A	0.111	Positive2	7.789E-03	+	rs149114330
RPS12	156	T	0.088	Positive2	3.630E-10	+	rs12202302
RPS12	161	A	-0.258	Negative1	1.318E-04	+	rs112352320
RPS12	179	C	-0.317	Negative1	4.082E-02	+	rs188597053

表 3.31: GTEx データを用いた Single-Tissue eQTL 解析結果

GTEx データの Single-Tissue eQTL 解析で cis-eQTL であった SNP のうち、 $p < 0.05$ で変異が転写活性化能に与える影響が推定された塩基と一致した SNP を一覧として示す。

Gene Symbol	SNP Id	Tissue	P-Value	Effect Size
DDX5	rs1991401	Whole Blood	2.30E-13	-0.23
DDX5	rs1991401	Artery - Tibial	2.00E-11	-0.26
DDX5	rs1991401	Testis	3.80E-07	-0.22
DDX5	rs1991401	Skin - Sun Exposed (Lower leg)	6.50E-07	-0.20
DDX5	rs1991401	Lung	9.50E-07	-0.19
DDX5	rs1991401	Adipose - Subcutaneous	6.10E-06	-0.19

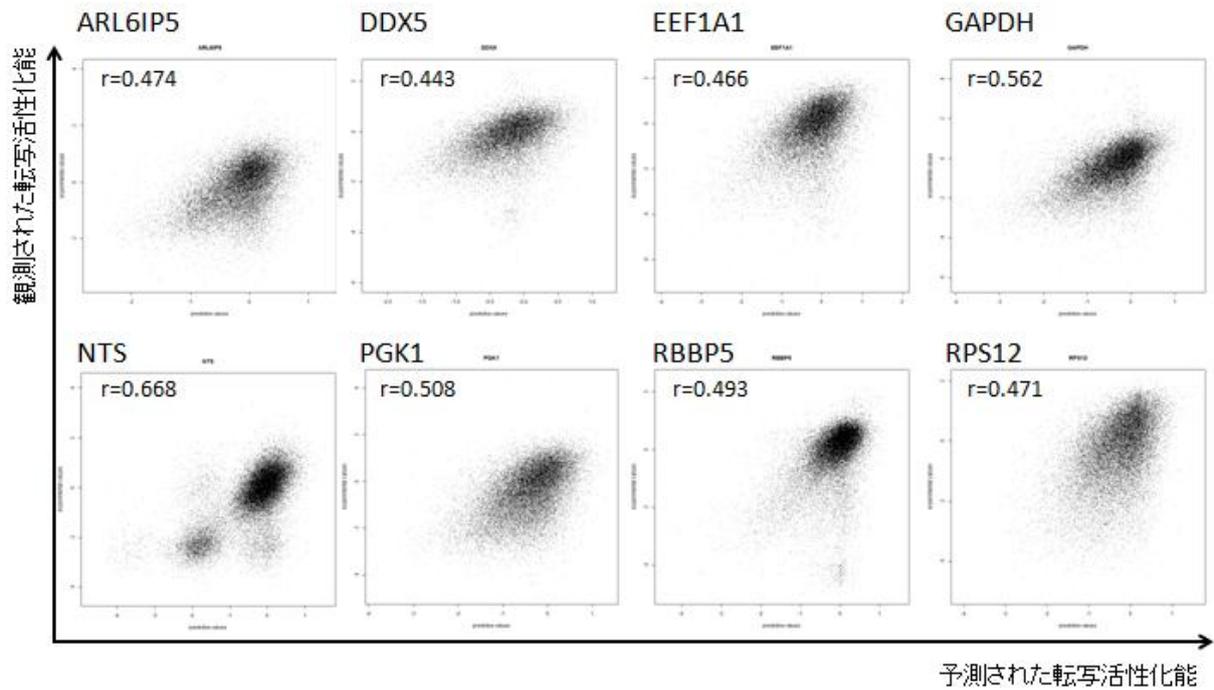
3.2.3.1.10 構築された重回帰モデルの転写活性化能の予測精度の評価

本研究で構築されるモデルがどの程度の精度で変異プロモーターの一次配列からその転写活性化能を予測できるかを検討した。10 分割交差検証を実施し、得られた転写活性化能の予測値と測定された実測値を比較した。その相関図を(図 3.35) に示す。

遺伝子ごとに構築された重回帰モデルより予測された転写活性化能と実際に測定された転写活性化能の相関係数は、説明変数を各塩基の変異有無とした場合に 0.443~0.668、説明変数を各塩基の変異パターンとした場合に、0.455~0.693 であった。

本研究で用いた重回帰モデルによる予測精度は、同種先行研究に比して同程度であった[54]。先行研究では、説明変数に塩基配列情報以外の情報を加味することや、変数選択を実施することで、予測精度の向上を図っている。本研究で得られた各塩基が転写活性化能へ与える影響についても基盤情報としてモデルに加味することで、転写活性化能の予測精度を向上させた予測モデルの構築が可能になるものと期待される。

説明変数：各塩基の変異の有無



説明変数：各塩基の変異パターン

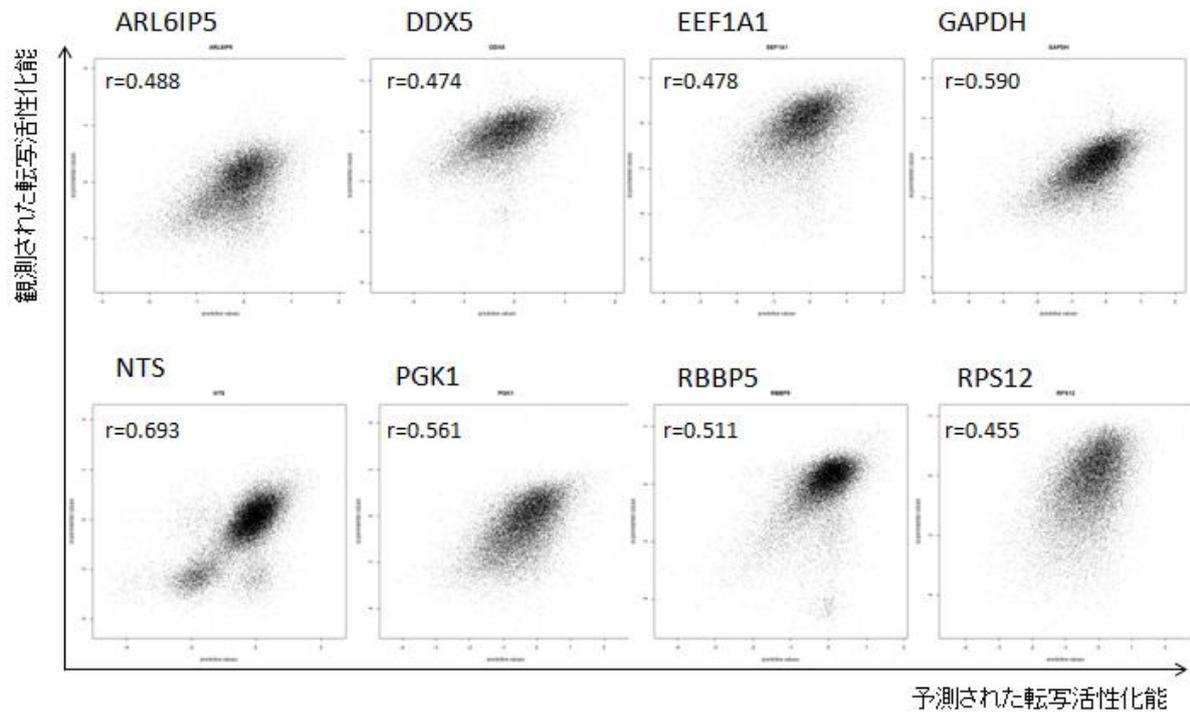


図 3.35：10 分割交差検証で得られた予測値と観測値の比較

遺伝子ごとに目的変数を測定された転写活性化能、説明変数を各塩基の変異又は各塩基の変異パターンとして構築した重回帰モデルについて 10 分割交差検証を実施し、得られた転写活性化能の予測値と測定された実測値を散布図として示した。

3.3 遺伝子のプロモーター配列の転写活性化能と RNA-seq の発現量の比較

本研究で体系的なルシフェラーゼアッセイにて測定した遺伝子のプロモーター配列の転写活性化能と、RNA-seq により測定された HEK293 細胞内での遺伝子の mRNA の発現レベルを比較した (図 3.36)。転写活性化能と mRNA の発現レベルは Spearman の相関係数 0.163 と弱い相関にとどまった。転写活性化能の強い P1 は 0.184、転写活性化能の弱い P2 は 0.004 であった。

一次配列の転写活性化能が低い、mRNA の発現レベルが高い遺伝子は、本研究で用いた約 1000bp のプロモーター領域以外に存在するエンハンサーや、プロモーター配列以外の転写制御機構の影響により、mRNA の発現レベルを高めている可能性が考えられた。転写活性化能は高いが、mRNA の発現レベルが低い遺伝子は、遺伝子のプロモーター領域のメチル化やクロマチン構造によりプロモーター配列自体は転写活性化能を有するものの実際の転写は開始されていない可能性があると考えられた。あるいは転写は開始されているが、mRNA の分解などの転写開始後の転写制御機構により、mRNA レベルが低下しているのかもしれない。特に転写活性化能の弱い遺伝子のプロモーター配列は、mRNA の発現レベルが転写活性化能以外の制御機構に強く影響される可能性が考えられた。

本研究で得られた遺伝子のプロモーター配列の有する転写活性化能、遺伝子のプロモーター配列内の塩基レベルの転写活性化能への影響を起点に、ChIP-Seq[29]によるヒストン修飾の状態、NET-Seq[56]による RNA の伸長の状態、BRIC-Seq[57]による RNA の分解の状態等を加味することで、遺伝子の転写制御機構の全体像が詳解できるのかもしれない。

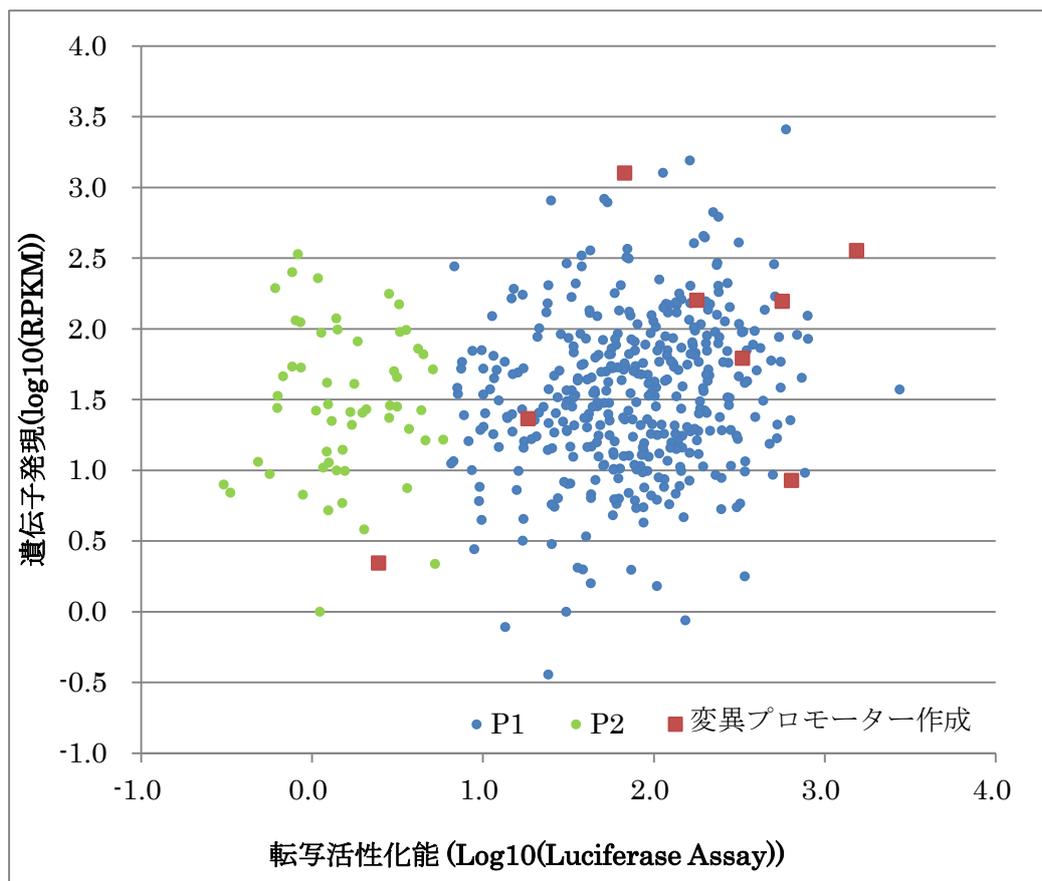


図 3.36: 遺伝子のプロモーター配列の転写活性化能と mRNA の発現レベルの比較

本研究で体系的なルシフェラーゼアッセイで測定した遺伝子のプロモーター配列の転写活性化能 (X 軸) と、RNA-seq で測定した遺伝子の mRNA の発現レベル (Y 軸) を散布図に示した。青: 転写活性化能高い P1、緑: 転写活性化能の低い P2、赤: 本研究で変異プロモーター配列の作成及び転写活性化能の測定に用いた遺伝子。

4 総括

本研究では、遺伝子のプロモーター配列に一次配列として内在的にコードされている転写活性化能の解析を試みた。その基盤となる測定技術を確立し、データ収集およびデータ解析の方法論の開発を行った。

前半では、HEK293 細胞において、体系的なルシフェラーゼレポータージーンアッセイを用いた実験系を構築した。ヒト遺伝子の転写に関わる遺伝子のプロモーター領域群の配列、転写活性化能における全体像を初めて示すことができたものと考えている。遺伝子のプロモーター領域以外の領域であるランダム領域についても、遺伝子のプロモーター領域と同様の解析を行った。遺伝子のプロモーター領域以外の領域についても、その潜在的な転写活性化能について全体像を記載できたと考えている[15]。

後半では、前半で課題と考えられた、プロモーター配列中の 1 塩基レベルの配列差が転写活性化能へ与える影響を解析することを目的に、大規模測定が可能な実験系の構築を行った。Error prone PCR とランダムタグ配列を用いた多様な遺伝子の変異プロモーター配列の取得、次世代シーケンサーを駆使して効率的な変異プロモーターの配列決定及び転写活性化能の測定を行うことのできる実験系を構築することができた。さらに、一連の実験系を用いて取得した 8 遺伝子の変異プロモーター配列と転写活性化能データを解析し、遺伝子のプロモーター配列中の 1 塩基レベルの配列差が転写活性化能に与える影響について評価した。

興味深いことに、予想に反し、体系的なルシフェラーゼアッセイにより得られた遺伝子のプロモーター配列の転写活性化能と mRNA レベルの発現は Spearman の相関係数として 0.163 という弱い相関であった (図 3.36)。これは、転写制御機構全体を解明するためには、本研究で得られた一次元の DNA 配列が有する転写活性化能のデータを基盤に、ChIP-Seq[29]によるヒストン修飾の状態、NET-Seq[56]による RNA の伸長の状態、BRIC-Seq[57]による RNA の分解の状態、RNA-Seq[25]による mRNA レベルの発現などの各転写制御段階が、最終的な mRNA 発現量の決定に大きく寄与している可能性を示していた。

本研究では HEK293 細胞を利用し、転写活性化能のデータを蓄積した。本研究は細胞内での遺伝子発現制御、特にその近位プロモーター塩基配列要素を精査しようという基盤的な基礎研究である。しかし、疾患解析研究へと本研究を発展させることも可能かもしれない。すなわち、利用する細胞として、疾患由来及び正常細胞由来の初代培養細胞を用い、プロモーター配列の 1 塩基レベルの配列差が転写活性化能へ与える影響を推定、比較解析を行う。これにより、疾患の発症メカニズムに関わるプロモーター配列上の原因配列が特定できれば興味深い。ゲノム編集により、疾患病態への介入も可能となるかもしれない[61]。

また、本研究では、転写活性化能の予測モデルを構築した。本研究で構築された重回帰モデルは、予測値と観測値の予測精度が相関係数として 0.443~0.693 であった (図 3.35)。先行研究においてもその予測精度は依然として十分でなく[54, 55]、さらなる改善が求められる。モデルの予測精度の向上のためには、モデルに含める変数の選択、転写因子結合配列マトリクスや GC 含量などの情報、非線形モデルの適応などが考えられる。予測精度の高い転写活

性化能予測モデルが構築できれば、多数の実験を実施せずとも、1塩基ごとの配列差が転写活性化能へ与える影響を知ることができるかもしれない。さらに、予測精度の高い予測モデルが構築できれば、目的とする転写活性化能を実現するための、人工的なプロモーター配列の設計などにも利用可能かもしれない。

本研究で構築した、プロモーター領域やプロモーター領域中の1塩基レベルの配列差が転写活性化能に与える影響を解析する一連の実験系及び解析方法が、幅広い分野で利用され、該当分野の研究の発展に繋がることを期待する。

5 参考文献

1. Levine M, Tjian R., Transcription regulation and animal diversity., *Nature*. 2003 Jul 10;424(6945):147-51.
2. International Human Genome Sequencing Consortium., Finishing the euchromatic sequence of the human genome., *Nature*. 2004 Oct 21;431(7011):931-45.
3. Birney E, et al, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project., *Nature*. 2007 Jun 14;447(7146):799-816.
4. Wittkopp, P. J., et al., Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 2012. 13 (1), p.59-69.
5. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
6. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
7. Lincoln D. Stein. End of the beginning. *Nature* **431**, 915-916(2004)
8. Suzuki, Y. & Sugano, S. Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. *Methods Mol Biol* 221, 73-91 (2003).
9. Imanishi T, et al, Integrative annotation of 21,037 human genes validated by full-length cDNA clones., *PLoS Biol*. 2004 Jun;2(6):e162. Epub 2004 Apr 20.
10. Suzuki, Y. et al. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res* **11**, 677-84 (2001).
11. Suzuki Y, et al, Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites., *EMBO Rep*. 2001 May;2(5):388-93.
12. Yamashita R, et al, DBTSS provides a tissue specific dynamic view of Transcription Start Sites., *Nucleic Acids Res*. 2010 Jan;38(Database issue):D98-104
13. Carninci P, et al, The transcriptional landscape of the mammalian genome., *Science*. 2005 Sep 2;309(5740):1559-63.
14. Kimura K, Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes., *Genome Res*. 2006 Jan;16(1):55-65.
15. Sakakibara, Y., et al., Intrinsic promoter activities of primary DNA sequences in the human genome. *DNA Research*, 2007. 14 (2), p.71-7
16. Matys, V. et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31, 374-8 (2003).
17. Khodursky AB, et al, Life after transcription--revisiting the fate of messenger RNA., *Trends Genet*. 2003 Mar;19(3):113-5.
18. Sanger F, et al, Nucleotide sequence of bacteriophage phi X174 DNA., *Nature*. 1977 Feb 24;265(5596):687-95.
19. Mardis, E. R., The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 2008. 24 (3), p.133-41.

20. Bentley, D. R., et al., Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 2008. 456 (7218), p.53-9.
21. Margulies M, Genome sequencing in microfabricated high-density picolitre reactors., *Nature*. 2005 Sep 15;437(7057):376-80.
22. Bentley DR, et al, Accurate whole human genome sequencing using reversible terminator chemistry., *Nature*. 2008 Nov 6;456(7218):53-9.
23. Metzker ML, et al., Sequencing technologies - the next generation., *Nat Rev Genet*. 2010 Jan;11(1):31-46.
24. Wilhelm BT, et al., RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing., *Methods*. 2009 Jul;48(3):249-57.
25. Wang Z, et al, RNA-Seq: a revolutionary tool for transcriptomics., *Nat Rev Genet*. 2009 Jan;10(1):57-63.
26. Sultan M, et al., A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome., *Science*. 2008 Aug 15;321(5891):956-60.
27. Tsuchihara K, et al, Massive transcriptional start site analysis of human genes in hypoxia cells., *Nucleic Acids Res*. 2009 Apr;37(7):2249-63.
28. Barski A, et al., High-resolution profiling of histone methylations in the human genome., *Cell*. 2007 May 18;129(4):823-37.
29. Park PJ, et al., ChIP-seq: advantages and challenges of a maturing technology., *Nat Rev Genet*. 2009 Oct;10(10):669-80.
30. Patwardhan, R. P., et al., Massively parallel functional dissection of mammalian enhancers in vivo. *Nature Biotechnology*, 2012. 30 (3), p.265-70.
31. Sharon, E., et al., Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotechnology*, 2012. 30 (6), p.521-30.
32. Kwasnieski, J. C., et al., Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proceedings of the National Academy of Sciences*, 2012. 109 (47), p.19498-503.
33. Dickel, D. E., et al., Function-based identification of mammalian enhancers using site-specific integration. *Nature Methods*, 2014. 11 (5), p.566-71.
34. Akhtar, W., et al., Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*, 2013. 154 (4), p.914-27.
35. Melnikov, A., et al., Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnology*, 2012. 30 (3), p.271-7.
36. Arnold, C. D., et al., Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science*, 2013. 339 (6123), p.1074-7.
37. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-64 (2002).
38. Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M. & Miller, W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* **8**, 967-74 (1998).
39. Ota, T. et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat*

- Genet* **36**, 40-5 (2004).
40. Carninci, P. & Hayashizaki, Y. High-efficiency full-length cDNA cloning. *Methods Enzymol* **303**, 19-44 (1999).
 41. Okazaki, Y. et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563-73 (2002).
 42. Suzuki, Y., Yamashita, R., Sugano, S. & Nakai, K. DBTSS, DataBase of Transcriptional Start Sites: progress report 2004. *Nucleic Acids Res* **32 Database issue**, D78-81 (2004).
 43. Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M. & Miller, W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* **8**, 967-74 (1998).
 44. Leung DW, Chen E, Goeddel DV (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* 1(1):11-15.
 45. Li, H., et al., Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009. 25 (14), p.1754-60.
 46. Li, H., et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 2009. 25 (16), p.2078-9.
 47. Ross Ihaka and Robert Gentleman (1996) R: A Language for Data Analysis and Graphics. *Journal of computational and graphical statistics* 5: 299-314.
 48. Kel, A. E., et al., MATCHM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research*, 2003. 31 (13), p.3576-9.
 49. Wingender, E., The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Briefings in Bioinformatics*, 2008. 9 (4), p.326-32.
 50. Siepel A, et al., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes., *Genome Res*. 2005 Aug;15(8):1034-50.
 51. Cadwell, R.C. and Joyce, G.F. (1992) *PCR MethodsAppl*.
 52. McCullum EO, et al., Random mutagenesis by error-prone PCR., *Methods Mol Biol*. 2010;634:103-9.
 53. Sherry ST, et al., dbSNP: the NCBI database of genetic variation., *Nucleic Acids Res*. 2001 Jan 1;29(1):308-11.
 54. Landolin, J. M., et al., Sequence features that drive human promoter function and tissue specificity. *Genome Research*, 2010. 20 (7), p.890-8.
 55. Irie, T., et al., Predicting promoter activities of primary human DNA sequences. *Nucleic Acids Research*, 2011. 39 (11), p.e75.
 56. Churchman, L. S., et al., Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 2011. 469 (7330), p.368-73.
 57. Imamachi, N., et al, BRIC-seq: A genome-wide approach for determining RNA stability in mammalian cells. *Methods*, 2014. 67 (1), p.55-63.
 58. Versteeg, R. et al. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res* **13**, 1998-2004 (2003).

59. Rockman, M. V. & Wray, G. A. Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol* **19**, 1991-2004 (2002).
60. Wakabayashi-Ito N, et al., Characterization of the regulatory elements in the promoter of the human elongation factor-1 alpha gene., *J Biol Chem.* 1994 Nov 25;269(47):29831-7.
61. Nelson CE, et al., In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy., *Science.* 2016 Jan 22;351(6271):403-7
62. GTEx Consortium., Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans., *Science.* 2015 May 8;348(6235):648-60.
63. Veyrieras JB, et al., High-resolution mapping of expression-QTLs yields insight into human gene regulation., *PLoS Genet.* 2008 Oct;4(10):e1000214.
64. Ravarani CN, et al., Affinity and competition for TBP are molecular determinants of gene expression noise., *Nat Commun.* 2016 Feb 2;7:10417.
65. Stewart AJ, et al., Why transcription factor binding sites are ten nucleotides long., *Genetics.* 2012 Nov;192(3):973-85.

6 論文目録

原著論文

Intrinsic Promoter Activities of Primary DNA Sequences in the Human Genome

Yuta Sakakibara, Takuma Irie, Yutaka Suzuki, Riu Yamashita, Hiroyuki Wakaguri, Akinori Kanai, Joe Chiba, Toshihisa Takagi, Junko Mizushima-Sugano, Shin-ichi Hashimoto, Kenta Nakai, Sumio Sugano

DNA Res (2007) 14 (2): 71-77.

7 謝辞

本研究の遂行に当たり、修士課程より長年ご指導頂きました、東京大学大学院新領域創成科学研究科 メディカル情報生命専攻 ゲノムシステム医療科学分野 菅野純夫教授、生命システム観測分野 鈴木穰教授に御礼申し上げます。常日頃からの温かいご指導のおかげで、研究を遂行することができました。心より感謝申し上げます。

お忙しい中、本論文の審査をお引き受けくださり、貴重なご助言をいただきました。東京大学大学院新領域創成科学研究科 中井謙太 教授、松田浩一 教授、佐藤均 准教授、九州工業大学 大学院情報工学研究院 矢田哲士 教授に心より御礼申し上げます。

本研究において、共同研究者として多大なるご協力をいただきました。菅野研究室 入江琢磨 特任助教、鈴木研究室 劉 瑩 様、菅野研究室卒業生 門城 拓 様、心より感謝申し上げます。

鈴木研究室 今村聖実 様をはじめ、実験補助員のみなさま、株式会社ダイナコム 若栗 浩幸 博士、堀内 映実 様には、次世代シーケンサーのデータ取得及び情報解析を主に大変なご協力をいただきました。深く感謝申し上げます

博士論文の審査過程において、様々なご助言をいただき、大変お世話になりました。菅野研究室 特任研究員 関本 登 様、心より御礼申し上げます。

研究生生活を送る上で、お世話になりました、東京大学大学院新領域創成科学研究科 メディカル情報生命専攻 ゲノムシステム医療科学分野並びに生命システム観測分野に所属するみなさまに感謝いたします。

最後に、大学院進学に快く賛成し、支えてくれた家族に感謝いたします。

平成 28 年 9 月