

## 科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成25年5月1日現在

機関番号：17102

研究種目：挑戦的萌芽研究

研究期間：2011～2012

課題番号：23651195

研究課題名（和文） 高精度遺伝子座同定法の開発

研究課題名（英文） The accurate determination of gene loci by genome-wide approach.

研究代表者

大川 恭行 (OHKAWA YASUYUKI)

九州大学・医学研究院・准教授

研究者番号：80448430

研究成果の概要（和文）：

細胞機能は厳密にコントロールされた mRNA 量のバランスによって成立している。mRNA を転写する RNA polymerase II (RNAPII) は、触媒サブユニット RPB1 のリン酸化により分類される。ゲノムワイドな RNAPII のリン酸化レベルと mRNA 発現量の関係は未だに不明である。我々はリン酸化 RNAPII の分布と mRNA の発現量の関係について解析した。mRNA の発現量が Ser2 と Ser5 の段階的なリン酸化と強く相関していることを明らかにした。結果、複雑な mRNA 発現制御機構は RNAPII のリン酸化の状態によって3つのカテゴリーに分類可能であることが示された。

研究成果の概要（英文）：

Cellular function is regulated by the balance of stringently regulated amounts of mRNA. Previous reports revealed that RNA polymerase II (RNAPII), which transcribes mRNA, can be classified into the pausing state and the active transcription state according to the phosphorylation state of RPB1, the catalytic subunit of RNAPII. We examined the correlation between the phosphorylation of RNAPII and mRNA expression levels using a combined analysis by CHIPseq and RNAseq. We could visualize the recruitment of various phosphorylated RNAPIIs. Our analysis indicated that the complexity of quantitative regulation of mRNA levels could be classified into three categories according to the phosphorylation state of RNAPII.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
交付決定額	3,000,000	900,000	3,900,000

研究分野：エピジェネティクス

科研費の分科・細目：ゲノム科学・ゲノム医科学

キーワード：クロマチン、転写、RNAポリメラーゼII、発現プロファイリング、リン酸化

## 1. 研究開始当初の背景

細胞の機能はゲノム上の遺伝子の正確かつ制御された転写によって成り立っている。タンパク質をコードしている遺伝子から転写された mRNA の量は多様であり、その転写の制御はクロマチン構造上の様々な核内因子によって成されている。その中の一つの重要な制御メカニズムは、RNA polymerase

II (RNAPII) の活性制御である。RNAPII は蛋白質を転写するコード遺伝子のすべてと多くの非コード遺伝子を転写しているが、その RNAPII の活性は、RNAPII の巨大な触媒サブユニットの1つである RPB1 のリン酸化の状態と相関している。RPB1 は他のタンパク質にはみられない C 末端ドメイン (CTD) を持っている。それは種間で保存性の高い7つのアミ

ノ酸のリピート配列(N - Tyr1 - Ser2 - Pro3 - Thr4 - Ser5 - Pro6 - Ser7 - C)から成っており、哺乳類ではそのリピートは 52 回繰り返している。このリピート構造を構成しているアミノ酸は、リン酸化修飾や単糖修飾といったアミノ酸修飾の潜在的な標的となっている。転写制御において、ゲノム DNA に結合していないフリーの低リン酸化 RNAPII は、最初に遺伝子プロモーター領域にリクルートされる。RNAPII がプロモーター領域から離れて転写を開始するためには、CTD Ser5 をリン酸化させる役割を持つ general transcription factors の一つである TFIIH が必要である。プロモーターから離れた RNAPII は、転写開始点(TSS)の下流まで移動する可能性があるが、それだけでは NELF や DSIF などの pausing factor により、mRNA の生産的な伸長反応は妨げられてしまう。この現象は Promoter proximal pausing として知られている。mRNA の生産的な伸長反応は CTD Ser2 残基のリン酸化と共に引き起こされる。RNAPII の Promoter proximal Pausing は遺伝子発現レベルの調節に寄与している可能性がある。Pausing のため全長の mRNA は全く検出されない可能性も考えられる一方で、RNAPII のエントリーとエスケープが様々な割合で起こることによって、高い発現量も含めて様々なレベルの mRNA が発現される可能性もある。最近の研究ではヒト線維芽細胞や ES 細胞において、不活化遺伝子のプロモーター上に RNAPII が結合している場合があることが示されている。さらにマウス ES 細胞では、bivalent gene において、Ser5 がリン酸化され Ser2 がリン酸化されていない RNAPII が TSS 周囲に集積している。これらの遺伝子は、分化マーカーとして知られているが、未分化な状態にも関わらず低いレベルで検出され得る。近年、次世代型の高速シーケンサー技術とそれを使った cDNA 解析法が革命的なツールとして使われる様になってきたが、これらのシーケンスデータが、「active transcription」の状態の遺伝子から発現したものを示しているのか、「Pausing」の状態の遺伝子から発現したものを示しているのか、in vivo でゲノムワイドな RNAPII のリン酸化状態は検証されていない。「pausing」の状態にある RNAPII が存在する遺伝子のうちいくつかは、分化において重要な役割を担っている。それ故、RNASeq と RNAPII の関係を理解することは非常に重要である。RNASeq で同定されたすべての遺伝子上にある RNAPII のリン酸化状態を評価するために、我々は Ser2 及び Ser5 がリン酸化されていないフリーの RNAPII を除外し、RNAPII の Ser2 と Ser5 残基がリン酸化されることによって活発に転写されている遺伝子と、Ser5 残基のみがリン酸化された RNAPII が存在する「Pausing

state」にある遺伝子を区別する必要がある。RNAPII のリン酸化状態と mRNA の発現レベルとの関連を評価することにより、活発に転写されている遺伝子と Pausing 状態にある遺伝子の同定が可能になる。

## 2. 研究の目的

Northern blotting 法や RT-qPCR 法、SAGE 法や発現アレイ法など様々な技術が遺伝子の発現レベルを定量し解析するために開発されてきた。近年開発された次世代シーケンサーは一度の実験で何千万ものタグを読み込むことにより、これまでの技術と比較して非常にバイアスの少ない mRNA 発現解析を可能とした (RNASeq)。RNASeq は既知の転写産物の定量のみに留まらず、新規の転写産物の同定、組織特異的な alternative splicing を調べることが可能である。この技術は、従来の発現アレイと比較して 1000 倍以上の感度で発現遺伝子の定量化が可能であり、現在行うことのできる mRNA 発現量定量ツールの中では最も優れている様にみえる。しかしながら、RNASeq は独自の問題点も有している。その一つとしてあげられるのは、参照配列を必要とする点である。次世代シーケンサーは、以前の技術と異なり、25-200bp の短いフラグメントを一度の実験で何千万個もシーケンスする。このフラグメントは「リード」とも言われ、既知のトランスクリプトームにマッピングすることにより遺伝子の発現が定量される。しかしながら、よく研究されたマウスやヒトのトランスクリプトームですら未だに不完全であり、RNASeq の解析は参照配列によって制約をうけ、新規転写産物を同定するためには別の計算が必要であった。TopHat は既知のトランスクリプトームの情報によらず、新規のスプライシングサイトを含めて、新しい転写産物を評価する道を開いた。更に Cufflinks は、リードをゲノム参照配列にマップすることで、スプライシングを考慮しつつ 1kbp あたりのすべての転写産物の定量を可能にする。これらのマッピング技術による定量法の弱点は、ゲノム参照配列にマップする際に比較的短いタグをゲノムにマッピングするという点である。シーケンスをマッピングできるかどうかは mRNA の構造に依存しており、mRNA に共通の構造であるホモログが存在する時には、統計的な解析結果にバイアスが生じる可能性がある。それゆえ、これらのバイアスを克服するためには、ゲノムへ 1 対 1 対応でマップされるユニークな情報だけでなく、ゲノム上の 2 か所以上の場所にマップされる「multiple hit」の情報も利用する必要がある。TopHat ではパラメータの 1 つにこのマッピング数をコントロールできる「Max multihits」があり、マッピング効率の最適化を行うことが可能であ

る。しかしながらこのパラメータが mRNA の同定に与える影響の詳細は、いまだ検証されていない。それゆえ、我々は ChIPSeq のデータと RNASeq のデータを組み合わせることによって、リン酸化 RNAPII の分布と mRNA の発現量の関係について解析した。

### 3. 研究の方法

我々は次世代シーケンサーを用いて CTD のリン酸化による RNAPII の制御機構を解析することにより、どの様にして mRNA 発現量がコントロールされているのかを明らかにした。我々はすべての遺伝子について、RNAPII のリン酸化状態を分類した。さらに、これらのデータと最適化された「Max multihit」パラメータを用いた RNASeq から得られたゲノムワイドな遺伝子発現データを組み合わせることにより、mRNA の多様な発現と RNAPII のリン酸化の状態の関連性を明らかにした。

### 4. 研究成果

「pausing」状態と定義される遺伝子のどのくらいが mRNA を発現し、どの程度の量を発現しているかもまだ明らかになっていない。そこで RNASeq で検出された遺伝子について、どのくらいの「active transcribing gene」と「pausing gene」が存在するのかを調べた。「Max multihits」パラメータを増加させた時、RNASeq と ChIPSeq で共に見つかる共通の遺伝子の数よりも、RNASeq でのみ見つかる遺伝子の数の方が増える傾向がある。Multihit を過剰にすると発現していない遺伝子までカウントしてしまうリスクを増やしてしまうため、我々は「max multihit = 10」という最適化されたパラメータを用いて FPKM>0 となる遺伝子を検出した。ChIPSeq において陽性の peak を抽出する際には、Peakcaller の一つである PeakSeq を用いて、P value<0.05, FDR<0.05 の基準に当てはまるものを抽出した。遺伝子については、RefSeq で定義されるヒト遺伝子(計 23,821 遺伝子)について評価を行った。結果、62.7%にあたる 14,954 遺伝子が RNASeq にて FPKM>0 または ChIPSeq にて Ser2 または Ser5 のリン酸化が陽性である RNAPII が存在すると判定された。Ser2P が陽性と判定された 7,918 遺伝子のうち、87%にもあたる 6,860 遺伝子が Ser5P も陽性と判定された。一方で、Ser5P が陽性と判断された 11,590 遺伝子のうち、6,860 遺伝子(59%)が Ser2P も陽性と判断された。この結果は、Ser5, Ser2 が順序だってリン酸化される必要があることを示唆している。しかしながら、13%にあたる 1,058 遺伝子は Ser2P 単独陽性と判定されている。これらの遺伝子について UCSC(University of California, Santa Cruz) のゲノムブラウザを用いて ChIPSeq の結果を可視化すると、Ser2P 単独

陽性遺伝子は比較的遺伝子が密集している領域に存在している。さらに、ChIP-qPCR のデータも合わせて検証すると、例えば Ser2P 単独陽性と判定された SOX15 については、本来 Ser2P は TES に近づくにつれ徐々にシグナルが増えてくるのだが、Ser2P は TES よりも TSS 付近で同定されていることが分かる。このことから、Ser2P 単独陽性遺伝子は、周囲の遺伝子の影響を受けることによって生じた偽陽性かこの領域に存在する未知の転写産物を示している可能性が考えられる。CTD がリン酸化されていない RNAPII はまずプロモーター領域にリクルートされ、Ser5 がリン酸化されるとプロモーターから解放される。次に Ser2 がリン酸化されることによって「active transcription」が開始されるが、転写が終了したとしても RNAPII は停止反応が起こるまで走り続ける。このことが、ChIPSeq の解像度の悪化を惹起し、遺伝子が密集している領域での偽陽性の一因になっている可能性がある。この弱点を克服するために、我々は遺伝子にまたがっている peak のみを陽性とする基準を設けている。このことは「Promoter proximal pausing」の状態にある RNAPII が検出される効率に悪影響を及ぼす可能性があるが、この条件下でも TSS 周囲に pausing している Ser5 リン酸化 RNAPII は十分検出されている。面白いことに、RNASeq は「active transcription」遺伝子(Ser5P+, Ser2P+)のみならず、「Promoter Proximal Pausing」遺伝子(Ser5P+, Ser2P-)についても比較的高い遺伝子発現量を検出しており、両者は FPKM>0 遺伝子の大部分を占めていた。これらの結果は Ser5 と Ser2 のリン酸化が 2 段階で遺伝子の発現量と関連していることを示している。そのことはさらに、RNASeq は、その高すぎる感度のために、背景にあるリン酸化 RNAPII によるエピジェネティックな転写調節機構を無視してしまう可能性があることを示唆している。マウスの ES 細胞では分化マーカーのいくつかは Ser5P 単独陽性遺伝子であり、低いレベルでしか mRNA を発現していないことが示されている。しかしながら、RNASeq の高い感度の結果、これらの遺伝子は RNASeq で同定される可能性があるため、幹細胞における分化マーカーの発現の解釈には気を付ける必要がある。RNASeq で FPKM>0 と判定された 13,462 遺伝子のうち、ChIPSeq にて Ser2P もしくは Ser5P が陽性と判定された遺伝子は 11,156 遺伝子(83%)存在した。残りの 2,306 遺伝子(17%)は RNASeq にて FPKM>0 であったが、ChIPSeq では Ser2P も Ser5P も同定されなかった。ChIPSeq にて Ser2P もしくは Ser5P が陽性と判定された 12,648 遺伝子のうち 88% あたる 11,156 遺伝子は RNASeq にて FPKM>0 と判定された。さらに「pausing」遺伝子や

「active transcription」遺伝子についてその機能を調べるために、我々はGene Ontologyの中に有意な関連 term がないかFisher 検定を用いて抽出を試みた。「active transcription」遺伝子については数百の GO term が有意と判断され、「Pausing」遺伝子についてはミトコンドリア遺伝子を含めた少数の GO term が有意と判断された。いずれの解析でもハウスキーピング遺伝子以外に特異的な遺伝子の有意な集積は認められなかった。

これまでに示した通り、mRNA の発現量は RNAPII の Ser2 と Ser5 の段階的なリン酸化によってグループ分けされ、リン酸化のレベルに応じて発現量は増加する傾向にあった。このことは遺伝子発現量を3つのカテゴリー、つまり「High pausing, High elongation」「High pausing, Low elongation」「Low pausing, Low elongation」、に分けることができることを示唆していた。このことから、ロジスティック回帰分析を用いると、遺伝子は FPKM の値に応じてカテゴリー化することができる。解析は RNASeq で FPKM>0 と判定された 13,462 遺伝子を対象とした。これまで述べた通り、Ser2P 単独陽性と判断された 588 遺伝子については、その Ser2P は周囲の遺伝子の影響を受けたことによる偽陽性の可能性があったため、この解析では Ser2P, Ser5P double negative に加えてカウントした。さらに、FPKM の値に応じて、RNAPII の Ser2/Ser5 のリン酸化の状態がどのくらいの割合で認められるか、確率プロットを作成した。RNASeq によって得られる FPKM の値が上昇するにつれて、RNAPII がリン酸化される割合(ChIPSeq で P-value<0.05, FDR<0.05 の基準で判定されるリン酸化の割合)は上昇した。最終的に FPKM が 0 付近では約半数程度の遺伝子が ChIPSeq ではリン酸化 RNAPII の存在が証明できない「Low pausing, Low elongation」群に属していることが分かる。一方でリン酸化 RNAPII が証明された遺伝子のうち、FPKM が 1 以下の遺伝子に関しては半分以上が Ser5P single positive の遺伝子であった。Ser5P single positive 遺伝子は FPKM の値が約 6 に達するまで上昇し続けた。このことから、FPKM の値が比較的高くても意外に「pausing」遺伝子と判定された遺伝子が多いといえる。Ser2P/Ser5P double positive RNAPII に関連する遺伝子の割合は FPKM の値が増えるにつれて上昇し、最終的には 90%以上となった。我々の解析は様々な遺伝子発現量 FPKM から Ser2/Ser5 のリン酸化の割合を推測することができる。それぞれのカテゴリーの境界はロジスティック回帰分析の確率プロット上では水平方向に傾いており、遺伝子発現量の決定には RNAPII の Ser2/Ser5 リン酸化以外の別の因子の存在が疑われるが、

遺伝子発現量と RNAPII のリン酸化は概ね相関していると言える。これはエピジェネティックな因子を使って様々な遺伝子発現量を説明した初めてのモデルである。将来的には転写因子やヒストン修飾など別のエピジェネティックな因子を組み合わせると同様のモデルを用いることにより、遺伝子発現量を説明することができる。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 9 件)

- ① Maehara K, Odawara J, Harada A, Yoshimi T, Nagao K, Obuse C, Akashi K, Tachibana T, Sakata T, Ohkawa Y. A co-localization model of paired ChIP-seq data using a large ENCODE data set enables comparison of multiple samples. *Nucleic Acids Res.* 査読有、Vol. 41, No. 1, 2013, pp. 54-62. DOI: 10.1093/nar/gks1010.
- ② Saiwai H, Kumamaru H, Ohkawa Y, Kubota K, Kobayakawa K, Yamada H, Yokomizo T, Iwamoto Y, Okada S. Ly6C(+) Ly6G(-) Myeloid-derived suppressor cells play a critical role in the resolution of acute inflammation and the subsequent tissue repair process after spinal cord injury. *J Neurochem.* 査読有、Vol. 125, No. 1, 2013, pp. 74-88. DOI: 10.1111/jnc.12135.
- ③ Harada A, Okada S, Konno D, Odawara J, Yoshimi T, Yoshimura S, Kumamaru H, Saiwai H, Tsubota T, Kurumizaka H, Akashi K, Tachibana T, Imbalzano AN, Ohkawa Y. Chd2 interacts with H3.3 to determine myogenic cell fate. *EMBO J.* 査読有、Vol. 31, No. 13, 2012, pp. 2994-3007. DOI: 10.1038/emboj.2012.136.
- ④ Kumamaru H, Ohkawa Y, Saiwai H, Yamada H, Kubota K, Kobayakawa K, Akashi K, Okano H, Iwamoto Y, Okada S. Direct isolation and RNA-seq reveal environment-dependent properties of engrafted neural stem/progenitor cells. *Nat Commun.* 査読有、Vol. 3, 2012, 1140. DOI: 10.1038/ncomms2132.
- ⑤ Fujii Y, Shiota M, Ohkawa Y, Baba A, Wanibuchi H, Kinashi T, Kurosaki T, Baba Y. Surf4 modulates STIM1-dependent calcium entry. *Biochem Biophys Res Commun.* 査読有、Vol. 422, No. 4, 2012, pp. 615-620.

- DOI: 10.1016/j.bbrc.2012.05.037.
- ⑥ Kumamaru H, Saiwai H, Ohkawa Y, Yamada H, Iwamoto Y, Okada S. Age-related differences in cellular and molecular profiles of inflammatory responses after spinal cord injury. *J Cell Physiol*. 査読有、Vol.227、No.4、2012、pp.1335-1346.  
DOI: 10.1002/jcp.22845.
- ⑦ Xiao H, Leblanc SE, Wu Q, Konda S, Salma N, Marfella CG, Ohkawa Y, Imbalzano AN. Chromatin accessibility and transcription factor binding at the PPAR $\gamma$ 2 promoter during adipogenesis is protein kinase A-dependent. *J Cell Physiol*. 査読有、Vol.226、NO.1、2011、pp.86-93.  
DOI: 10.1002/jcp.22308.
- ⑧ Okada S, Saiwai H, Kumamaru H, Kubota K, Harada A, Yamaguchi M, Iwamoto Y, Ohkawa Y. Flow cytometric sorting of neuronal and glial nuclei from central nervous system tissue. *J Cell Physiol*. 査読有、Vol.226、No.2、2011、pp.552-558.  
DOI: 10.1002/jcp.22365.
- ⑨ Odawara J, Harada A, Yoshimi T, Maehara K, Tachibana T, Okada S, Akashi K, Ohkawa Y. The classification of mRNA expression levels by the phosphorylation state of RNAPII CTD based on a combined genome-wide approach. *BMC Genomics*. 査読有、Vo.12、2011、516.  
DOI: 10.1186/1471-2164-12-516.

[学会発表] (計26件)

- ① 大川 恭行、ヒストンバリエントによる骨格筋分化制御機構、第1回ヒストンバリエント研究会、2013.3.8、九州大学総合研究棟、福岡
- ② 大川 恭行、Cd2 incorporates H3.3 to mark myogenic genes.、新学術領域「ゲノム支援」国際シンポジウム”Expanding Frontiers of Genome Science”、2013.1.10、東京大学 伊藤国際学術研究センター、東京(招待講演)
- ③ 前原 一満, 小田原 淳, 原田 哲仁, 吉見 智彦, 長尾 恒治, 小布施 力史, 立花 太郎, 坂田 年男, 大川 恭行、共局在回帰モデルと ENCODE データセットを介した ChIP-seq データの多サンプル比較解析、第35回日本分子生物学会年会、2012.12.14、福岡国際会議場・マリンメッセ福岡、福岡

- ④ 林 正康, 小田原 淳, 原田 哲仁, 前原 一満, 立花 太郎, 赤司 浩一, 大川 恭行、CHD5(Chromodomain helicase DNA binding protein 5)はマウスの胚性幹細胞において H3.1 と結合して遺伝子発現を抑制する、第35回日本分子生物学会年会、2012.12.13、福岡国際会議場・マリンメッセ福岡、福岡
- ⑤ 大川 恭行、高次クロマチン構造制御による骨格筋分化制御、第35回日本分子生物学会年会、2012.12.12、福岡国際会議場・マリンメッセ福岡、福岡
- ⑥ Akihito Harada, Anthony N. Imbalzano, Yasuyuki Ohkawa、Chd-2 dependent deposition of H3.3 is required for myogenic cell fate.、KEystone SYMPOSIA、2012.1.19、Keystone Resort、コロラド、USA
- ⑦ Jun Odawara, Akihito Harada, Yasuyuki Ohkawa、Control of gene expression through the phosphorylation of Ser2 and Ser5 of RNAP II: A combined analysis for ChIPseq and RNAseq.、KEystone SYMPOSIA、2012.1.19、Keystone Resort、コロラド、USA
- ⑧ Yasuyuki Ohkawa, Jun Odawara, Akihito Harada、The deposition of histone variants is the basis of chromatin structural changes in cell fate decision.、第34回日本分子生物学会年会、2011.12.16、パシフィコ横浜、神奈川県
- ⑨ Akihito Harada, Jun Odawara, Kazumitsu Maehara, Tomohiko Yoshimi, Saori Yoshimura, Taro Tachibana, Hitoshi Kurumizaka, Hiroshi Kimura, Yasuyuki Ohkawa、Chd-2 dependent deposition of H3.3 is crucial for Brg1 recruitment in myogenesis.、第34回日本分子生物学会年会、2011.12.15、パシフィコ横浜、神奈川県

[図書] (計4件)

- ① 大川 恭行, 小田原 淳, 原田 哲仁、羊土社、実験医学 Vol.31 No.4「クローズアップ実験法: ChIPseq のためのクロマチン免疫沈降法」、2013、6(pp.577-582)
- ② 原田 哲仁, 大川 恭行、秀潤社、細胞工学 Vol.31 No.8「骨格筋分化における高次クロマチン構造解析」、2012、5(pp.877-881)
- ③ Ohkawa Y, Mallappa C, Vallaster CS, Imbalzano AN.、Springer 社、Methods Mol Biol. ”Isolation of nuclei from skeletal muscle satellite cells and myofibers for use in chromatin immunoprecipitation assays.”、2012、

14(pp. 517-530)

- ④ Ohkawa Y, Mallappa C, Vallaster CS, Imbalzano AN.、Springer 社、Methods Mol Biol. ” An improved restriction enzyme accessibility assay for analyzing changes in chromatin structure in samples of limited cell number. ”、2012、12(pp. 531-542)

[その他]

ホームページ等

<http://chromatin.med.kyushu-u.ac.jp/>  
九州大学 医学研究院 先端医療医学部門  
エピジェネティクス分野 / 大川研究室

## 6. 研究組織

### (1) 研究代表者

大川 恭行 (OHKAWA YASUYUKI)  
九州大学・医学研究院・准教授  
研究者番号：80448430