分裂期クロマチンの制御機構を理解するプロテオミクス解析

太田信哉*

*E-mail: shinya.ohta@kochi-u.ac.jp

高知大学医学部生化学講座: 783-8505 高知県南国市岡豊町小蓮

(受付 2017 年 10 月 2 日, 改訂 2017 年 10 月 30 日, 受理 2017 年 11 月 8 日)

これまで分裂期染色体上に,約4,000種類のタンパク質を同定した.そこに含まれる非特異的結合タンパク質を除いた "真の"染色体タンパク質の抽出や,隠されたタンパク質間の相互作用を見いだすために,機械学習を基盤としたコンピュータシミュレーションが応用できる.一方で,分裂期の制御を理解する上で,タンパク質リン酸化の包括的な理解は欠かせない.これまでに,染色体上の分裂期特異的リン酸化と脱リン酸化を決定しており,これらは,今後の染色体動態の 制御機構の解明に向けて大きな足がかりとなる.

1 序 論

分裂期の染色体は、1882年に Walther Flemming によっ て観察されて以来、その特徴的な構造から多くの生物学 研究者の興味の対象となってきた. 1977 年に Uli Laemmli らは、Chromosome Scaffold とそれによる染色体の分裂 期特異的な規則的な折りたたみのモデルを提唱し¹⁾, そ の後, DNA トポイソメラーゼ IIa (TopoIIa) と Structure maintenance of chromosome 2 (SMC2) という 2 つの Chromosome Scaffold 因子が決定されている。特に SMC2 It SMC4, Chromosome-associated protein D2 (CAP-D2), CAP-H, CAP-Gとともにコンデンシン複合体を形成し, 分裂期特異的染色体凝縮に重要であることが示されてい る^{2),3)}.後に、この複合体はコンデンシンIとなり、SMC2、 SMC4, CAP-D3, CAP-H2, CAP-G2からなる複合体, コ ンデンシン II も近年見いだされている⁴⁾. 最近では, こ れらのコンデンシンと Kinesin family member 4A (Kif4A) の働きが、反対の作用を持つ TopoIIa とバランスよく組み 合わされることで、分裂期染色体の構造が形成されている ことが示されている⁵⁾.しかし、この分裂期の染色体凝縮 は未だに未知部分が多く、PMM 標識と呼ばれるヒストン H3の修飾(H3pT3K4me2R8me2)が染色体構造を形成し ている原動力であるという報告もなされている⁶⁾.

一方で、染色体が正確に娘細胞に分配される際には、染 色体狭窄部のセントロメアが重要な領域となる.近年、こ の領域のヒストンは一部のH3がCentromere protein A (CENP-A) に置き換わっているだけではなく⁷⁾、CENP-A を含むヌクレオソームのヒストン H4 の 20 番目のリジン がメチル化されていることが報告されている⁸⁾. さらには, 両端のペリセントロメアの領域に CENP-B が特異的に結 合し⁹⁾, ヒストンが高度に修飾され, ヘテロクロマチンを 形成していることもわかっている¹⁰⁾. このセントロメア 上には分裂期には数十種類のタンパク質からなる巨大複合 体, キネトコアが形成される¹¹⁾. キネトコアは微小管と の接合や, 分裂期チェックポイントのモニタリングなどの 機能を担っている.

このように, 正確な分裂期の染色体構造の構築機構を解 明することは非常に重要であるが、未だに解明されていな い点が多い.我々は、これまでに、こうした重要な機構を 作り上げている分裂期染色体のタンパク質組成を決定する ことを目的として研究を進めてきた. もちろん同様に分裂 期の染色体タンパク質組成の決定を目的とした研究はこれ までもなされてきたが^{12),13)},染色体という巨大なタンパ ク質複合体の組成決定には、非特異的結合タンパク質を考 慮しなくてはならず、タンパク質組成の完全決定には至っ ていなかった.我々は、この問題を解決するために、機 械学習を応用した Multi-classifier combinational proteomics (MCCP) という解析方法を作り上げ, 意味のある分裂期 染色体のタンパク質組成決定に至った.そして,近年では それをさらに発展させ、分裂期染色体上の相互作用をシ ミュレーションする手法, nano Random Forest (nanoRF) を開発した. また, これらとは別に我々は分裂期染色体上 のタンパク質リン酸化サイトを網羅的に決定している.

2 分裂期染色体をターゲットとしたプロテオミクス解析

様々な機能のタンパク質が結合している巨大複合体であ る染色体の組成の決定を試みる研究は、染色体の制御メカ ニズムを明らかにする上で必須となる.近年いくつかのグ ループにより分裂期染色体プロテオームの決定が試みら れたが、明確な知見が得られてこなかった^{12),13)}.それは、 染色体は非常に大きく複雑な複合体であるため単離が困難 であり、加えて核膜消失後に大量の細胞質タンパク質が非 特異的に結合してしまうことが原因であった.我々はニワ トリ DT40 細胞より分裂期染色体の単離を試み,質量分析 でその組成を確認したところ,4,027 種類の染色体タンパ ク質を同定した.この中には,122 のこれまでに見いださ れているほぼ全ての分裂期セントロメア局在タンパク質が 含まれており (Fig. 1A),4,027 種類の染色体タンパク質に, 新規の機能的な染色体タンパク質が含まれている可能性が 高く存在する.

一方で,他グループの先行研究と同様に,非特異的染 色体結合タンパク質と考えられるものも含まれていた



Fig. 1 Proteomic analysis of mitotic chromosome proteins with Random forest analysis predicting novel proteins of interest

(A) The 27 classes of proteins found in chromosomes. (B) The rank of all proteins observed for proteomic abundance, enrichment, retention, SMC2 dependency is plotted as vertical color coded lines (rug plots). Proteins are ranked from the highest (left) to lowest values. Centromere proteins are indicated by longer bars. (C) Separation of the two training sets by RF analysis. The line provides optimal separation of the two training sets. RF analysis of B includes the bioinformatic analysis. (D) Position of novel cloned proteins and remaining uncloned proteins in the same 2d analysis with classification of newly cloned proteins from regions to the left and right of the dividing line, respectively. This figure was modified from Ohta *et al.*, 2010 (15).

(Fig. 1A). 質量分析スペクトラムから定量すると,それら が全体に占める質量の割合は16%に過ぎず,ごく少量で はあった.しかし,その種類は1,700種類ほどもあり,有 益なタンパク質の発見を試みるためにはこの数字は少ない とは言えない.そこで,得られたプロテオームから意味の あるタンパク質とそれ以外を分類する手法を兼ね備えた新 しい解析法を取り入れたプロテオミクスが必要となった.

3 分裂期染色体上のタンパク質コピー数の決定

最初に、質量分析により決定した 4,027 種類のタンパク 質全てにおいて、染色体に結合しているタンパク質のコ ピー数を決定することを試みた.その結果、質量分析によ るペプチドのスペクトラルカウント数とタンパク質コピー 数の相関を基にした下記の式と定数 a を決定することで、 全ての染色体タンパク質のコピー数を推測することが可能 となった.定数 a の決定のため、数種類の染色体タンパク 質を大腸菌の大量発現系より精製、濃度を決定した.それ を標準タンパク質として、単離した炭素 13 でラベルした 染色体に混合し、質量分析でその比を測定し、aの決定に 至った.

estimated copy number

$= a^{*}(10^{[b^{*}(\text{spectral count})/(\text{protein-specific normalization})]} - 1)$

本定量法の正確性を示すために, Stable Isotope Labeling with Amino acids in Cell culture (SILAC) を用いてラベル した、二種類の染色体を単離し、全タンパク質の相対差を 比較した. 同時に上記の式を用いて、この二種類の染色体 に含まれる全タンパク質の定量を行ない、その結果より 個々のタンパク質の相対差を比較した. その結果, 2つの 定量法による相対差は、80%のタンパク質で2倍以内に、 97%で4倍以内となった.この事実は、本項で開発した方 法が, SILAC を用いた方法と同程度に信頼できる定量性 を持つこと、なおかつこれまで困難であった網羅的絶対量 の推定を可能にするものであることを示している. 解析結 果の一例として、ニワトリ DT40 細胞においては、1つの セントロメアには CENP-A を含むヌクレオソームが約12 分子存在することを示した. これは,報告されているヒト における CENP-A の数の 1/15 程度である. このことから. ヒトとニワトリでは、セントロメアの形成過程が異なるこ とが推測される. また, CCAN, Ndc80, Mis12 複合体は1 つのセントロメアにそれぞれ 51 ± 27,47 ± 5,40 ± 22 と、ほぼ50分子の複合体が存在することを明らかにした. この事実とこれまでの電子顕微鏡による先行研究などから, ニワトリ細胞のセントロメアには、6つの微小管が結合し それぞれに各8分子の複合体が存在することが示唆された. この数は酵母における微小管あたりの Ndc80 の数と完全 に一致した.

しかし、コピー数を基盤とした切り口のみでは、実際の 染色体たんぱく質を決定するための強力なパラメータとは ならず (Fig. 1B),異なるパラメータの導入が必要である ことが示唆された.

4 多次元プロテオミクスによる組成の決定

分裂期においては、「目的とする染色体タンパク質は染 色体上に多く存在し、細胞質内には少ない」と仮定し、そ のようなタンパク質を抽出する定量プロテオミクスを展開 することを試みた.実際には、SILACを基盤に、安定同 位体でラベルした細胞質と非ラベルの染色体を単離し、1 対1に混合し質量分析を行い、全てのタンパク質の染色体 特異性 "Enrichment"を決定した.得られた結果は、染色 体タンパク質を抽出する傾向にはあるが、完全に分裂期 染色体タンパク質とそれ以外を分けるには至らなかった (Fig. 1B).

そこで、単離した染色体と細胞質を用いて、全ての タンパク質の in vitro における置換速度 "Retention" を, SILAC を応用することで決定した(Fig. 1B). また, 分裂 期の染色体凝縮に必須と考えられる複合体、コンデンシン の存在下と非存在下の染色体における各タンパク質の量 の比較 "SMC2 dependency" (Fig. 1B), そして新規に見い だしたアウターキネトコア(動原体外層部)局在タンパ ク質 Ska3 の存在下と非存在下の染色体における各タンパ ク質の量の比較"Ska3 dependency"を決定した¹⁴⁾. しか し、どのパラメータ "Classifier" も完全に分裂期染色体タ ンパク質とそれ以外を分けるには至らない事実から、全て の classifier を組み合わせる事でそれが改善できないかを 検討した.実際に全てをクラスター解析で組み合わせると, 複合体ごとに一つのブランチに収まる傾向があり、プロテ オミクスの結果を多次元で解析することが、実際の染色体 タンパク質を決定するために有効であることが示唆された. しかし、このような解析法でもやはり非特異的染色体結合 タンパク質が、染色体タンパク質を多く含むブランチに、 無視できない程度含まれており、異なる解析法を開発する 必要があった.

5 機械学習を応用した分裂期染色体プロテオームの理解

前述の全ての定量プロテオミクスを、機械学習によって 染色体タンパク質と非特異的染色体結合タンパク質に分 類することを試みた.実際にはランダムフォレストを用 い、同定したタンパク質のうち、分裂期染色体に結合する ことがよく知られている、ヒストン、セントロメアタンパ ク質、テロメアタンパク質そしてコヒーシンなどの染色体 タンパク質を"True"、膜タンパク質やミトコンドリアタ ンパク質などを"False"として"Learning set"を構築し、 これまでのプロテオミクスの結果をもとにランダムな決定 木, "Decision tree", を数百通り作成した. 一つ一つの決 定木に従い, 個々のタンパク質が染色体タンパク質と認識 されるかどうかシミュレーションしたところ, その確率を 計算することが可能であると分かり, 高い確率で"真の" 分裂期染色体タンパク質の抽出に成功した. 既知の染色体 タンパク質と非染色体タンパク質で, この手法の正確性を 検証したところ, 89%のタンパク質を正確に分類できた (Fig. 1C).

実際にその有効性を示すためには、すべてのタンパク質 の細胞内局在を細胞生物学的手法で確認しなくてはならな い.したがって、未知タンパク質を分類し、そのうち染色 体タンパク質と分類された 34 種類のタンパク質について、 そのヒト類似タンパク質を GFP と融合した形でヒト細胞 内において発現させ、蛍光顕微鏡を用いてその局在を確認 した.結果、その中から 30 種類の新規染色体局在タンパ ク質を見いだすことができた(Fig. 1C).同様に、非染色 体タンパク質と分類された 16 種類には 2 種類の染色体タ ンパク質しか見いだせなかった(Fig. 1C).以上の結果は、 93%の新規染色体タンパク質と 83%の非染色体タンパク 質を正確に分類したことを示している.この MCCP は分 裂期染色体のみならず、精製が困難で明確な答えのでない あらゆる細胞内巨大複合体プロテオーム決定に応用できる 手法 "Fuzzy proteomics"である^{15),16}.

6 機械学習による分裂期染色体上のタンパク質相互作用 解析

我々は、プロテオミクス解析のモデル細胞としてニワト リ DT40 細胞を用いている. この細胞の最大の利点は、遺 伝学的ノックアウト細胞を作製でき、これまでに構築した ノックアウト細胞のライブラリーとプロテオミクス解析と を融合することが可能な点にある.従って,染色体構造 の構築とその維持に重要な役割を担っている(あるいは その可能性がある)遺伝子の欠失変異体から分裂期の染 色体を単離し、その欠失が染色体のタンパク質組成に及 ぼす影響を、前述の SILAC を用いることで定量的かつ網 羅的に調べることが可能である.実際に,染色体構造に 関わる SMC タンパク質を含む複合体(コンデンシン I や コンデンシン II. あるいはその両方, そして, コヒーシン やSMC5/6複合体)をそれぞれ欠失させた変異体について, 分裂期染色体上の全タンパク質の組成変化を測定し、それ ぞれの複合体がどのように分裂期染色体構造に寄与してい るかを明らかにした¹⁷⁾. さらには, ここで得られた5種 類の変異体における分裂期染色体タンパク質の組成変化を 基にして, 分裂期染色体上のタンパク質間の相互作用や 制御構造を,網羅的に予測することを試みた(Fig. 2A, B). そのために MCCP を発展させ、定量プロテオミクスデー タからコンピュータシミュレーションで標的タンパク質の

相互作用タンパク質を予測する解析アルゴリズム nanoRF を開発した¹⁸⁾. これにより,標的としたタンパク質複合 体の新規構成因子の予測を可能としただけでなく,タンパ ク質複合体間の制御構造を理解することも可能であること を示した(Fig. 2A). 実際に,GFP 融合タンパク質の形で の局在観察や siRNA を用いたノックダウン表現型の観察 により,これらの機能未知タンパク質がセントロメアや染 色体腕部などに局在し,分裂期の制御に関与することを確 認している¹⁷⁾(Fig. 2C).

7 分裂期染色体のリン酸化プロテオーム解析

現在知られている制御機構だけでは染色体動態の全体像 を説明することは困難であり、未知の制御機構がまだ数多 く残されていると考えられる.そこで、その機構に近づく ことを目的として染色体上の未知の分裂期特異的リン酸化 を同定することを試みた.

チタニアを利用した Hydroxy acid-modified metal oxide chromatography (HAMMOC) 法を用いて¹⁹⁾,精製された 分裂期染色体からリン酸化ペプチドのみを濃縮した. 続け て,分裂期染色体のリン酸化部位の決定を行い,その結果 4,274 のリン酸化部位を決定した (Fig. 3A, B).

さらに、同定したリン酸化が、どの程度分裂期に特異的 であるかを調べるために、分裂期の細胞から単離した複合 体と, 間期から単離したそれとを SILAC を用いて比較し た. その結果 2,761 種類のリン酸化ペプチドについて分裂 期と間期の間で起こる増減を調べることができた. さらに, 分裂期と間期の個々のタンパク質の量比も決定し、この結 果も合わせて考察することで、351の分裂期特異的リン酸 化サイトと167の分裂期特異的脱リン酸化サイトを決定す るに至った. その詳細を見てみると, CENP-T や INCENP といったセントロメアタンパク質や Ki-67 やコンデンシン, TopoIIa などが分裂期特異的に高度にリン酸化されている ことがわかった(Fig. 3C). 逆に, DNA 複製や組換え機構 に関わるタンパク質群は分裂期において特異的に脱リン酸 化されていることを示した²⁰⁾ (Fig. 3C). これらの分裂期 特異的なリン酸や脱リン酸化は、染色体動態の制御機構を 知る上で大きな足がかりとなる.

8 終わりに

質量分析装置に関する技術向上,あるいは新しい解析方 法の開発は,非常に展開が早く,SILACを含めた定量質 量分析は今日では頻繁に目にする手法となっている.しか し,未だそこから得られる膨大なデータをシームレスに生 物の基本的システムの理解へと結びつけられる手法は限ら れている.本稿でふれた機械学習を取り入れた解析に加え, 主成分分析や k-Neighbor 法など,数学,統計分野を取り 入れた新手法を確立し駆使することで,質量分析を基盤と



Fig. 2 Nano Random Forest to figure out the novel chromosomal regulation

(A) Two-dimensional scatter graph plots of RF scores for (a) CPC versus condensin II and (b) CPC versus condensin I. Key proteins are color coded by category: condensin I, pink; condensin II, purple; CPC, blue; kinetochore, orange; and CCAN, green. (B) Magnification of the part of the heat map depicting the combined results of 5,058 proteins from nanoRF targeting condensin, condensin I, condensin II, cohesin, SMC5/6, CPC, histones, CCAN, Nup-Ran complexes, kinetochore, chromosomal scaffolds, and ribosomal proteins. (C) Relative mitotic index of HeLa cells after 28 individual siRNA treatments against human orthologues. * in (B) shows the siRNA target in. This figure was modified from Ohta *et al.*, 2016 (17).

したプロテオミクス解析は、染色体構造だけでなく、多く の未解明のメカニズムの一端を解き明かす歩みに貢献でき ると考えている.

著者に開示すべき利益相反状態は無い.

文 献

1) Adolph KW, Cheng SM, Laemmli UK. Role of nonhistone proteins in metaphase chromosome structure. Cell. 1977;12:805– 816.

- Hirano T, Kobayashi R, Hirano M. Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. Cell. 1997;89:511–521.
- Hudson DF, Vagnarelli P, Gassmann R, Earnshaw WC. Condensin is required for nonhistone protein assembly and structural integrity of vertebrate mitotic chromosomes. Dev Cell. 2003;5:323–336.
- 4) Ono T, Losada A, Hirano M, Myers MP, Neuwald AF, Hirano T. Differential contributions of condensin I and condensin II



Fig. 3 Quantitative analysis of mitosis-specific phosphorylation

(A) Experimental design for quantification of mitosis-specific phosphorylation. (B) Immunoblot analysis of the abundance of chromosomal markers and actin in each fraction. Proteins were resolved by SDS-PAGE and stained with Coomassie Brilliant Blue. Xs: isolated mitotic chromosomes; mitotic WCE: whole-cell extract of mitotic cells; nuclear: nuclear fraction of unsynchronized cells; WCE: whole-cell extract of unsynchronized cells. (C) Heat map showing the population category of the 28 protein categories; clustering tendency is shown at the bottom. This figure was modified from Ohta *et al.*, 2016 (20).

to mitotic chromosome architecture in vertebrate cells. Cell. 2003;115:109–121.

- Samejima K, Samejima I, Vagnarelli P, *et al.* Mitotic chromosomes are compacted laterally by KIF4 and condensin and axially by topoisomerase IIα. J Cell Biol. 2012;199:755–770.
- 6) Markaki Y, Christogianni A, Politou AS, Georgatos SD. Phosphorylation of histone H3 at Thr3 is part of a combinatorial pattern that marks and configures mitotic chromatin. J Cell Sci. 2009;122:2809–2819.
- Tachiwana H, Kagawa W, Shiga T, *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. Nature. 2011;476:232–235.
- Hori T, Shang W-H, Toyoda A, *et al.* Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential for kinetochore assembly. Dev Cell. 2014;29:740–749.
- 9) Tanaka Y, Nureki O, Kurumizaka H, et al. Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA.

EMBO J. 2001;20:6612-6618.

- 10) Nakagawa H, Lee J-K, Hurwitz J, et al. Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. Genes Dev. 2002;16:1766–1778.
- Musacchio A, Salmon ED. The spindle-assembly checkpoint in space and time. Nat Rev Mol Cell Biol. 2007;8:379–393.
- 12) Morrison C, Henzing AJ, Jensen ON, *et al.* Proteomic analysis of human metaphase chromosomes reveals topoisomerase II alpha as an Aurora B substrate. Nucleic Acids Res. 2002;30:5318–5327.
- Uchiyama S, Kobayashi S, Takata H, *et al.* Proteome analysis of human metaphase chromosomes. J Biol Chem. 2005;280:16994–17004.
- 14) Ohta S, Bukowski-Wills J-C, Wood L, et al. Proteomics of isolated mitotic chromosomes identifies the kinetochore protein Ska3/Rama1. Cold Spring Harb Symp Quant Biol. 2010;75:433–438.

- 15) Ohta S, Bukowski-Wills J-C, Sanchez-Pulido L, et al. The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. Cell. 2010;142:810–821.
- 16) Kustatscher G, Hégarat N, Wills KLH, *et al.* Proteomics of a fuzzy organelle: interphase chromatin. EMBO J. 2014;33:648– 664.
- 17) Ohta S, Montaňo-Gutierrez LF, Alves F de L, *et al.* Proteomics analysis with a nano Random Forest approach reveals novel functional interactions regulated by SMC complexes on mitotic chromosomes. Mol Cell Proteomics. 2016;15:2802–2818.
- 18) Montano-Gutierrez LF, Ohta S, Kustatscher G, Earnshaw

WC, Rappsilber J. Nano Random Forests to mine protein complexes and their relationships in quantitative proteomics data. Mol Biol Cell. 2017;28:673–680.

- 19) Sugiyama N, Masuda T, Shinoda K, Nakamura A, Tomita M, Ishihama Y. Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. Mol Cell Proteomics. 2007;6:1103–1109.
- 20) Ohta S, Kimura M, Takagi S, Toramoto I, Ishihama Y. Identification of Mitosis-Specific Phosphorylation in Mitotic Chromosome-Associated Proteins. J Proteome Res. 2016;15:3331–3341.

Proteomics Analysis Understanding the Regulation System of the Mitotic Chromosome

Shinya Ohta*

*E-mail: shinya.ohta@kochi-u.ac.jp

Department of Biochemistry, Medical School, Kochi University, Kohasu, Oko-cho, Nankoku, Kochi 783-8505, Japan

(Received: October 2, 2017; Revised: October 30, 2017; Accepted: November 8, 2017)

Packaging of the DNA into condensed chromosomes during mitosis is essential for the faithful segregation of the genome into daughter nuclei. Although intensely studied for over 100 years, mitotic chromosome structure and composition are still yet to be fully elucidated. We developed the multi-classifier combinatorial proteomics (MCCP) and nano Random Forest to reveal the protein composition in mitotic chromosomes purified from chicken DT40 cells. One of the greatest advances from the MCCP approach to analysis of the mitotic chromosome proteome is the ability to combine with genetic knock-out and the stable labeling with amino acids in cell culture. These approaches open the door to the identification of all functional components of mitotic chromosomes despite the adventitious binding of cellular background proteins during mitosis. Furthermore, they can be extended by adding additional classifiers to delineate protein complexes and define functional dependencies between them in the context of intact mitotic chromosomes. This will serve both as a starting point for systematic determination of the full range of functions involved in mitotic chromosome segregation, and as a basis for the development of detailed structural and functional interaction maps of key chromosomal sub-domains. Indeed, we have discovered and characterized several number of novel centromere or chromosomal proteins. Currently we succeeded the identification of 4,274 total phosphorylation sites and 350 mitosis-specific phosphorylation sites in mitotic chromosome-associated proteins with using TiO₂-based hydroxy acid-modified metal oxide chromatography. In future, phosphoproteome analysis combined with MCCP and nanoRF approaches based on machine learning would be useful for the analysis of much more complex hierarchical regulation network on mitotic chromosomes.

Keywords: chromatin; mitosis; fuzzy proteomics; machine learning; phosphoproteome; SILAC.