

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 6月 6日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2011～2012

課題番号：23710230

研究課題名（和文） 染色体高次構造の情報学的再構築法の検討

研究課題名（英文） The method for reconstructing chromosomal conformation through informatics

研究代表者

中戸 隆一郎 (NAKATO RYUICHIRO)

東京大学・分子細胞生物学研究所・助教

研究者番号：60583044

研究成果の概要（和文）：

タンパクのゲノム結合部位を網羅的に同定可能な「ChIP-seq 法」のためのデータ解析・可視化プログラム DROMPA を開発した。本手法は精度高く高速に結果を得ることができるだけでなく、結果を可視化して概観できる機能を持ち、ChIP-seq 解析に必要なコストを大幅に削減できる。本手法を用いた成果のひとつとして、姉妹染色体間接着因子であるコヒーシンの機能不全によって引き起こされる CdLS 遺伝病患者の細胞におけるコヒーシンのゲノム局在の変化を解析した。

研究成果の概要（英文）：

We have developed a cost-effective and time-efficient program DROMPA for ChIP-seq analysis. DROMPA can visualize a protein-binding map, which can reduce the total cost for an analysis. By using DROMPA, we comprehensively investigated the cohesin binding sites for human B-cells. We found that a part of cohesin sites colocalize with an enhancer marker with tissue-specific manner, and such sites were lost for CdLS, that is caused by the loss of cohesin function.

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	3,500,000	1,050,000	4,550,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・システムゲノム科学

キーワード：ゲノム情報学、ChIP-seq 法、データ可視化

1. 研究開始当初の背景

全ての生命現象の出発点は、ゲノム上で起こる DNA とタンパクの相互作用である。生物の全ての生命活動はゲノム上で転写・複製・分配・修復など様々な機能が密接に連携することで形作られており、これらの諸機能がどう連携・統合されているかを体系的に理解することは、生物学的・医学的観点から極めて重要である。

種々の DNA-タンパク相互作用によって制御されるゲノム機能を体系的に理解するためには、まず種々の DNA 結合タンパクがゲノムのどの部位に結合し、どのようにその相

互作用が規定されているかを知ることが必要である。「ChIP-seq 法」はクロマチン免疫沈降法 (ChIP) と次世代シーケンサ (高速に配列決定可能な装置) を組み合わせ、目的とするタンパクのゲノム結合部位情報 (各結合部位は「ピーク」として検出される) を網羅的かつ詳細に得られる手法である。近年の次世代シーケンサの改良に伴いシーケンシングのコストは大幅に削減され、ChIP-seq 法を用いた各種タンパクの結合部位の解析が近年急速に進められている。しかしながら、出力されるデータ量の増大に対して計算機の性能向上が追いついていないことや、解析

の最適なパラメータ条件（ピーク抽出の閾値など）はサンプル毎に異なるため、大規模な解析においては各サンプルの抽出条件を吟味するための多大な時間的コストを伴う、ことから、大量の ChIP-seq データを精度を保ちつつより高速に解析できる技術の需要が高まっている。

2. 研究の目的

本研究課題ではコヒーシオン（ゲノム複製時に姉妹染色体の接着と分配に中心的な役割を果たすタンパク）を足がかりとして主にヒトを対象に ChIP-seq 解析を行い、ゲノム上の DNA-タンパク相互作用に関する新規知見の獲得を目指す。近年の研究では、コヒーシオンは真核生物において転写制御にも関わっており、その機能不全はコルネリア・デ・ラング症候群 (CdLS, 先天性の奇形症候群) などの多臓器発生障害の原因となり得ることがわかっているが、その詳細は未だ明らかになっていない。本研究課題ではコヒーシオンによる転写制御のメカニズムを解明することを目標とする。

3. 研究の方法

ChIP-seq 法を用いてコヒーシオンのゲノム結合部位を網羅的に解析する。具体的には CdLS 患者の B 細胞を用いてそのコヒーシオンの局在を明らかにし、健常者との局在の差異を解析する。また、コヒーシオンの関連因子として知られる他のタンパクや、遺伝子マーカーであるヒストン修飾等の ChIP-seq データも合わせて解析を行い、コヒーシオン局在部位（特に CdLS で失われるピーク）との相関を調べることで、コヒーシオンによる転写制御機構を解明する。

また、本研究課題を遂行するにあたり必要となる高速かつ高精度な ChIP-seq 解析プログラムの開発を行う。ここでは、様々な生物種・タンパク結合・複製解析等に対応可能である柔軟性の高いプログラムの構築を目指す。さらに、コンピュータに不慣れた研究者にも結果を伝えるために、結果を視覚的に外観できるようにする可視化機能を実装する。

本研究課題は計算機を用いたインフォマティクスの解析であるが、次世代シーケンサのためのサンプル作製を含め、必要となる生物データ及び知見は本研究室の他の研究者から適宜提供を受けることができる。

4. 研究成果

2011 年度において、出芽酵母についてはゲノム上で有意に増幅している領域を抽出し、任意の解像度・スケールで描画するプログラムを完成させた。本プログラムは、ChIP-seq・ChIP-chip データに対して、遺伝子やセントロメア、複製開始点などの情報を

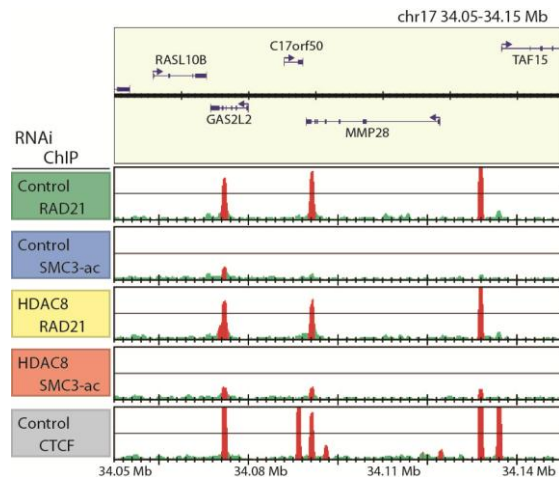


図 1: DROMPA を用いたピーク可視化例

[Deardorff *et al.* Nature. 2012]

を含め、複数のサンプルでも同時に可視化することができる。本プログラムを用いた共同研究による成果は論文で発表された [Tanaka *et al.* 2011, Masumoto *et al.* 2011, De Piccoli *et al.* 2012]。

2012 年度には、本プログラムの改良を更に進め、酵母・ヒトを含めたゲノム配列が公開されているあらゆる生物種に対して使用可能な ChIP-seq 解析プログラム DROMPA を完成させた [Nakato *et al.* 2013]。本プログラムの特徴としては以下が挙げられる。

(1) マップファイルを予め染色体毎の wig ファイル（ゲノムにマップされた断片配列の部分領域ごとのヒストグラム）に変換し、この wig ファイルを入力としてタンパク結合部位を推定する二段階手法により、計算に必要な時間・メモリを大幅に節約した。これにより、10 サンプルを超えるような大規模な解析を一般的なワークステーション上で無理なく実行できる。また、wig ファイルに変換する際、PCR 増幅に伴う配列の偏りや、全リード数の違いにより値を正規化するなど精度改善のための改良も施した。

(2) ピークリストだけでなく、wig データ（ゲノムの各領域にマップされたシーケンス配列数を定量化したもの）を遺伝子情報などと共に pdf 形式で可視化し出力する機能を実装した。これにより、シーケンス・マッピングされたデータがゲノム上でどのように分布しているかを視覚的に概観することができ、ChIP-seq の実験の成否や、タンパク結合抽出の閾値の妥当性を直観的に判断できる。既存の可視化ツールには、特別な可視化ソフトウェアを新たにインストールするか、Web ブラウザ上にアップロードすることで得られていたが、前者はコンピュータに不慣れた生物学研究者にとって敷居が高く、後者はサンプル数が増えた場合にアップロードの労力が増大してしまうという問題があった。

DROMPA は pdf 形式でデータを出力するため、特別なソフトウェアのインストールが不要で、全ての研究者にとって扱いやすい解析ソフトとなっている。

本プログラムを用いて 2011 年度から継続して ChIP-seq 解析を行い、CdLS 患者の B 細胞におけるコヒーシンの結合部位の同定と正常な細胞に対する変化を調べた。その結果、CdLS 患者の細胞では全てのコヒーシン結合部位において結合強度が 2 割程度減少していること、それにより得られるピーク数が 3 割程度減少することがわかった。更に、コヒーシン局在部位の大部分はインシュレータータンパクである CTCF と共局在することがわかっているが、一方で肝臓細胞において一部のコヒーシンが CTCF と独立に DNA に結合し、それらは肝臓特異的な転写因子と共局在することが近年報告されている [Schmidt *et al.* 2011]。我々の実験により、B 細胞においても同様の傾向が認められること、また、それらのコヒーシン結合部位は CdLS 患者の細胞で失われていることが確認された。失われた局在部位の大部分は p300 (エンハンサーマーカーとして知られる) と共局在していた。エンハンサーは細胞特異的な局在を示すことが報告されていることから [Ernst *et al.* 2011]、この結果は、一部のコヒーシンは CTCF と独立にエンハンサー領域に局在し、細胞特異的な遺伝子制御に関わっていることを示唆している。CdLS においては、この領域のコヒーシン結合不全により細胞特異的な形成異常の表現型を示すと考えられる。

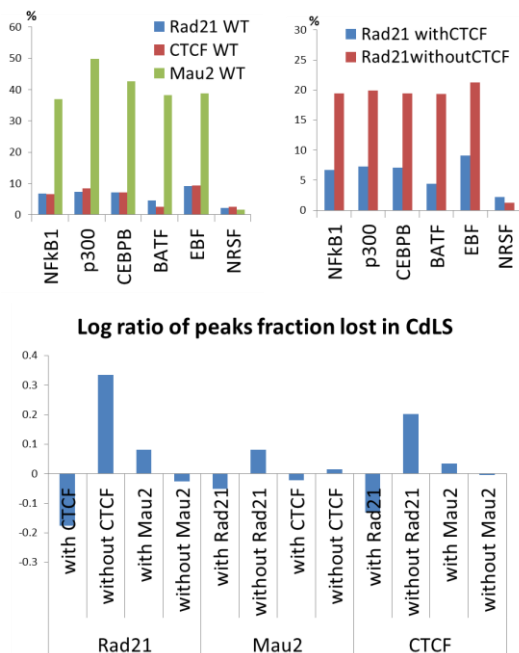


図 2: CdLS 患者のコヒーシン局在部位の消失と他のタンパク局在との相関解析

本研究課題で開発したプログラム DROMPA は、ChIP-seq 解析全体に必要な計算・時間的コストを削減し、大規模な ChIP-seq 解析を精度高くかつ効率的に行うことができるようになった。本システムを用いた解析の成果は、ヒト [Deardorff *et al.* 2012] (申請者は筆頭共著者)、マウス [Yamaji *et al.* 2013]、分裂酵母 [Tazumi *et al.* 2012] について発表された。また、DROMPA のウェブページを作成し、プログラムコードと実行・結果例を公開することで誰でも使用可能なシステムとしている。現在は、データベース上に存在する様々なタンパクの ChIP-seq データとコヒーシンとの相関を調べ、未だ明らかになっていないコヒーシンとの相互作用因子を探っている。また、DROMPA の拡張法として、最適なピーク抽出の閾値を現在は可視化されたピークから手動で設定しているが、サンプル毎に自動で決定できる機能を追加できれば、ChIP-seq 解析のコストを更に削減できると期待される。

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 8 件)

1. Nakato, R., Itoh T., Shirahige K., DROMPA: easy-to-handle peak calling and visualization software for the computational analysis and validation of ChIP-seq data, *Genes to Cells*, 査読有, 18, (2013).
2. Yamaji M., Ueda J., Hayashi K., Ohta H., Yabuta Y., Kurimoto K., Nakato R., Yamada Y., Shirahige K., Saitou M., PRDM14 Ensures Naive Pluripotency through Dual Regulation of Signaling and Epigenetic Pathways in Mouse Embryonic Stem Cells, *Cell Stem Cell*, 査読有, 12, 368-382 (2013).
3. Foltman M., Evrin C., De Piccoli G., Jones R.C., Edmondson R.D., Katou Y., Nakato R., Shirahige K., Labib K., Eukaryotic replisome components cooperate to process histones during chromosome replication, *Cell Rep.*, 査読有, 3, 892-904 (2013).
4. Tazumi A., Fukuura M., Nakato R., Kishimoto A., Takenaka T., Ogawa S., Song J.H., Takahashi T.S., Nakagawa T., Shirahige K., Masukata H., Telomere-binding protein Taz1 controls global replication timing through its localization near late replication origins in fission yeast, *Genes Dev.*, 査読有, 26, 2050-2062 (2012).

5. Deardorff M.A.*, Bando M.*, Nakato, R.* (他 39 名), HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle, Nature, 査読有, 489, 313-317 (2012). *筆頭共著者
6. De Piccoli, G., Katou, Y., Itoh, T., Nakato, R., Shirahige, K., and Labib, K., Replisome stability at defective DNA replication forks is independent of S-phase checkpoint kinases, Molecular Cell, 査読有, 45, 696-704 (2012).
7. Masumoto H., Nakato, R., Kanemaki M., Shirahige K., Hachinohe M., The inheritance of histone modifications depends upon the location in the chromosome in Saccharomyces cerevisiae, PLoS ONE, 査読有, 6, e28980 (2011).
8. Tanaka S., Nakato R., Katou Y., Shirahige K., Araki H., Origin Association of Sld3, Sld7, and Cdc45 Proteins Is a Key Step for Determination of Origin-Firing Timing, Current Biol., 査読有, 21, 2055-2063 (2011).

[図書] (計 1 件)

1. 中戸隆一郎, 伊藤武彦, 白髭克彦, 次世代シーケンサーによる染色体構造解析, 羊土社「実験医学」2012年4月号, P976-982.

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

DROMPA website:

http://www.iam.u-tokyo.ac.jp/chromosom_einformatics/rnakato/drompa/

6. 研究組織

(1) 研究代表者

中戸 隆一郎 (NAKATO RYUICHIRO)

東京大学・分子細胞生物学研究所・助教

研究者番号: 60583044