

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24年 6月 1日現在

機関番号：12608

研究種目：若手研究(A)

研究期間：2009～2011

課題番号：21687001

研究課題名（和文） 新型シーケンサーを用いたChIP-seq解析のための
情報解析基盤の研究開発研究課題名（英文） Research and development of biological information platform for
ChIP-seq analysis using next generation sequencer

研究代表者

伊藤 武彦 (ITO TAKEHIKO)

東京工業大学・大学院生命理工学研究科・教授

研究者番号：90501106

研究成果の概要（和文）：Solexa/SOLiDなどの次世代新型シーケンサーとクロマチン免疫沈降法(ChIP)を組み合わせた免疫沈降シーケンシング(ChIP-Seq)解析において、実験により得られたタグ配列情報と、対象とするゲノム情報とを組み合わせることにより、網羅的かつ定量的なタンパク質等結合情報を抽出するためのChIP-Seqに特化した情報解析プロトコルを導出することに成功した。

研究成果の概要（英文）：We developed a new algorithm for ChIP-seq analysis using next generation sequencer (i.e. Solexa / SOLiD) and targeted genome information, which can be picked up protein binding sites cyclopaedically and quantitatively from the entire genome.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	9,900,000	2,970,000	12,870,000
2010年度	5,600,000	1,680,000	7,280,000
2011年度	5,600,000	1,680,000	7,280,000
年度			
年度			
総計	21,100,000	6,330,000	27,430,000

研究分野：生物学

科研費の分科・細目：基礎生物学、遺伝・ゲノム動態

キーワード：バイオインフォマティクス・ChIP-seq解析

1. 研究開始当初の背景

(1) 染色体とは生命の本質であり、その維持、継承機構の破綻は癌化、老化をはじめ、種々の疾病の原因となる。複製、分配、転写、高次構造制御、組換え、修復といった染色体機能についての個別研究は我国でも盛んな研究分野であり、染色体機能に関わる重要なタンパク質が国内で多数同定されてきた。これらタンパク質が染色体上のどこにどの時期に局在するかを詳細に捉える技術として、ChIP-chip法（タンパク質の結合部位を同定する染色体免疫沈降法-Chromatin Im-muno-precipitation法とDNAチップを組み合わせた方法）が申請者を含むグルー

プにより開発され、染色体上のタンパク質の動態を高い解像度で捉えることが可能となった。申請者らは本方法を用い、複製タンパク質とチェックポイントタンパク質、分配タンパク質と転写装置、分配タンパク質とその制御因子が染色体上で働く動態を捉えることに酵母からヒトに至るまで幅広く成功している。

(2) 一方近年、Solexa/SOLiDに代表される次世代新型シーケンサーと呼ばれる25-35塩基ほどを数千万本高速に配列決定することが可能な機械が登場してきており、応用分野の一つとしてクロマチン免疫沈降法(ChIP)と組み合わせた免疫沈降シーケン

シング(ChIP-Seq)が挙げられている。しかし、今見られている ChIP-Seq による解析事例に関しては ChIP-chip 法の解析手法を流用しているに過ぎず、せつかくハイブリダイズという間接的な方法ではなく、配列情報として読み取れているという ChIP-Seq ならではのメリットを活かした研究事例は見当たらない。ChIP-chip ではクロスハイブリダイズの影響により解析ができていない繰り返し配列への結合情報や、セントロメアなどシークエンスされていなかったゲノム領域へのアクセスが可能となるの可能性を ChIP-seq 解析は秘めている。

(3) そこで、申請者らが、ChIP-chip 法に関して持っている大きなアドバンテージを持っている、ChIP-chip 法で得た情報解析プロトコルをベースとし、上述した ChIP-Seq ならではの特徴を活かした情報解析プロトコルを研究開発することはさらなるアドバンテージを生み、染色体動態研究分野において今後もリーダーシップを発揮していくことが可能になると期待される。

2. 研究の目的

本研究の目的は、Solexa/SOLiD などの次世代新型シークエンサーとクロマチン免疫沈降法(ChIP)を組み合わせた免疫沈降シークエンシング(ChIP-Seq)解析において、実験により得られたタグ配列情報と、対象とするゲノム情報とを組み合わせることにより、網羅的かつ定量的なタンパク質等結合情報を抽出するための ChIP-Seq に特化した情報解析プロトコルを導出することにある。本研究を通じ、ChIP-Seq の利点を広く認識されているコストによるメリットのみではなく、定量性の担保とゲノム中の繰り返し配列や非シークエンス領域までを対象とした網羅性の担保を可能にすると考えられる。

3. 研究の方法

(1) まず、ChIP-chip で一定の研究成果を挙げているコヒーシタンパク質を用い、ヒト HeLA 細胞における ChIP-Seq 解析を実施し、得られたデータをヒトゲノム上のユニーク領域にマッピングし、結合領域を同定することで ChIP-chip 成果の再現を図る。

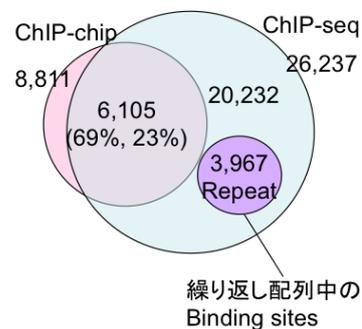
(2) 続いて、繰り返し配列を含む全ゲノム領域へのタンパク質結合情報を抽出するための情報解析プロトコルの研究開発を実施する。また、ゲノムへマッピングした結果張り付かなかったタグ情報をまとめて収集し、velvet など現在研究開発が進んでいるアセンブルプログラムによる解析を用いることで、未だゲノムシークエンスされていない領域への結合状況の解析も実施する。

(3) ChIP-seq 解析アルゴリズムの精度向上を目指し、mate-pair 情報を用いることでタ

ンパク質結合箇所の有意性判定を高めるためのアルゴリズム開発を実施する。これにより、繰り返し配列への結合状況を解析するためのプロトタイプ解析を実施する。

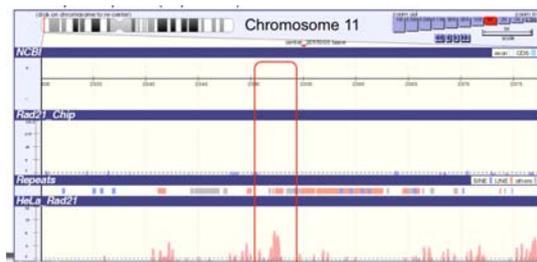
4. 研究成果

(1) ヒトのコヒーシデータ(SOLiD)を用い、ChIP-chip で得られた結果を再現可能な、ChIP-seq 解析プログラムの開発に成功した。ChIP-chip 解析では、ある Locus に対して設定された複数プローブから得られた結合情報を基に、Wilcoxon-rank-test を用いて有意に結合している領域を決定している。ChIP-seq の場合では、全ゲノム領域中への IP 分画由来および WCE 分画由来の結合状況の分布から有意に IP/WCE が高い領域を抽出することが理想である。しかしながら、IP/WCE の分布パターンはほぼ正規分布を取り、ゲノム全体の分布パターンから有意な領域を見いだすことは不可能である。このため、ChIP-seq 解析においても、ChIP-chip と同様にゲノムの局所局所に注目し、ある Window サイズを設定し、その Window 内に結合した IP 分画由来のタグ数と WCE 分画由来タグ数から濃縮度を算出し、注目している Window をさらに 50 分割し、個々のサブ window での IP, WCE 量を比べたデータを基に Wilcoxon-rank-test を用いることで優位性を算出する手法を開発した。本手法をヒトのコヒーシデータに適用することで、左図の通り、ほぼ



ChIP-chip で得られた結合サイトを包含する ChIP-seq 結果を得ることに成功した。

本手法の結果により、ChIP-seq では ChIP-chip で得られた結合サイト数よりも約3倍のサイト数が得られていることが確認できる。これは、ChIP-chip 解析ではアクセスできなかった繰り返し配列へのアクセスが可能になったこと(下図、赤枠を参照)



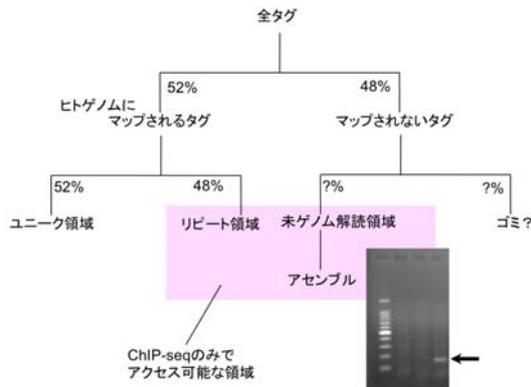
さらには、解析の分解能が高まったことで、

ChIP-chip 解析時には、1 結合サイトと判定されていた領域が、ChIP-seq 解析では複数の結合サイトと判定されるケースが認められることに起因する。(下図参照)



繰り返し配列へのアクセスが可能となった背景には、ChIP-chip ではクロスハイブリダイズの可能性を除去できないのに対し、ChIP-seq では塩基レベルでの精度を持ったデータが得られるため、一般的に繰り返し配列とされるゲノム領域でも区別がつくことが大きい。例えばヒトゲノムでは、50bp の配列長があれば、ゲノム全体の 91.1%、繰り返し配列中でも実に 85.3%を区別することが可能となる。

(2) 上述のアルゴリズムを用い、ChIP-seq データに対する結合サイト抽出プロトコルを確立し、種々の生物/タンパク質への適用を行った。これらの結果は論文①～⑤にまとめられている成果へと繋がっている。また、ヒトのコヒーシンに対する ChIP-seq 解析結果の概要は、以下の通りにまとめられる。



すなわち、マップされた全タグのうち 52%はユニークな領域に、残りの 48%は ChIP-seq のみでアクセス可能な繰り返し配列にマッピングされるのみならず、マップされない全体の 48%のタグ配列をアセンブルすることで、有意に濃縮された配列を認めることができた。この配列の存在は PCR で確認されている。このように、ゲノム未解読領域への結合状況も ChIP-chip では不可能であったが、次世代シーケンサから得られた配列をアセンブルすることでアクセスが可能となり、本手法の応用は、今後さらなる知見の獲得への応用が期待される。繰り返し配列領域への結合状況の解析では、

繰り返し配列のクラス毎に IP 分画由来の配列の割合、WCE 分画由来の配列の割合を算出し、その両者の比をとることで、どのクラスの繰り返し配列に濃縮がかかっているかを算出するといった、ChIP-chip では不可能な ChIP-seq ならではの解析手法も編み出せている。本手法も今後様々な応用例が期待できる。

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 5 件)

- ① De Piccoli G, Katou Y, Itoh T, et al. Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases. *Mol Cell*. 2012;45(5):696-704. 査読有
- ② A. Kegel, H. B. Lindroos, T. Kanno, K. Jeppsson, L. Strom, Y. Katou, T. Itoh, K. Shirahige et al. Chromosome length influences replication-induced topological stress *Nature*, 2011, 471 (7338):392-6. 査読有
- ③ B. Hu, T. Itoh, A. Mishra, Y. Katoh, K-L. Chan, W. Upcher, C. Godlee, M. B. Roig, K. Shirahige*, and K. Nasmyth* ATP Hydrolysis Is Required for Relocating Cohesin from Sites Occupied by Its Scc2/4 Loading Complex *Curr. Biol.* 21(1):12-24. (2011) 査読有
- ④ H. Maruyama, M. Shin, T. Oda, R. Matsumi, R.L. Ohniwa, T. Itoh, K. Shirahige, T. Imanaka, H. Atomi, S. H. Yoshimura, and K. Takeyasu Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Mol Biol Cell*. 2010. 査読有
- ⑤ A. Kurze, K. A. Michie, S.E. Dixon, A. Mishra, T. Itoh, S. Khalid, L. Strmecki, K. Shirahige, C. H. Hearing, J. Löwe, and K. Nasmyth A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion *EMBO J.* 2010, 査読有

〔学会発表〕(計 0 件)

〔その他〕

ホームページ等：特になし

6. 研究組織

(1) 研究代表者

伊藤 武彦 (ITO TAKEHIKO)

東京工業大学・大学院生命理工学研究科・
教授

研究者番号：90501106

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：