

配信先：文部科学記者会、科学記者会、  
熊本県内報道機関、千葉県政記者クラブ



千葉大学  
CHIBA UNIVERSITY

令和6年5月16日

報道機関 各位

熊本大学

千葉大学

## 疾患を引き起こすゲノム状態を「地図化」 ーエピゲノムビッグデータの解析インフラを創出ー

### (ポイント)

- 世界最大規模のエピゲノム統合データベース ChIP-Atlas のメジャーアップデートを実施しました。
- 疾患関連ゲノム変異情報などの注釈づけデータを統合するとともに、エピゲノムデータを比較解析するためのオンラインツールを実装しました。
- エピゲノムビッグデータの解析基盤として、ChIP-Atlas は遺伝性疾患の解明、新規創薬標的の探索、再生医療などの分野への応用が期待されます。

### (概要説明)

熊本大学生命資源研究・支援センターの鄒 兆南助教、沖 真弥教授を中心とする研究グループは、千葉大学国際高等研究基幹・大学院医学研究院の大田達郎准教授との共同研究により、エピゲノム統合データベース ChIP-Atlas (<https://chip-atlas.org>) のメジャーアップデートを行い、ゲノムの三次元構造・疾患感受性ゲノム変異などの注釈づけ情報を統合し、遺伝子発現制御に関わるエピゲノム状態の変容を検出する比較解析ツールを実装しました。これにより、ChIP-Atlas は世界最大規模のデータ数を誇るエピゲノム情報インフラへと進化し、遺伝性疾患の発症メカニズムや薬物作用機序の解明、細胞分化転換の効率化などへの応用が期待されます。

本研究の成果は、グリニッジ標準時間 2024 年 5 月 16 日に英国オックスフォード大学出版局 (Oxford University Press) が発刊する学術誌『Nucleic Acids Research』(オンライン版)に掲載されました。

なお、本研究は国立研究開発法人科学技術振興機構 (JST) 「統合化推進プログラム」(JPMJND2202)、JST-ERATO 有田リピドームアトラスプロジェクト (JPMJER2101)、JST さきがけ (JPMJPR1942)、JST 次世代

研究者挑戦的研究プログラム（JPMJSP2110）、国立研究開発法人日本医療研究開発機構（AMED）革新的先端研究開発支援事業（PRIME）「加齢変容細胞のデコーディング技術の開発と応用」（23gm6710001h0002）、同機構生命科学・創薬研究支援基盤事業（BINDS）「空間オミクス解析の支援」（23ama121017j0002）、日本学術振興会科学研究費助成事業（22J15229, 23KF0048, 22H02819, 23H04954）、京都大学医学部教育研究支援基金、京都大学メディカルイノベーション大学院プログラムの研究支援を受け実施したものです。

## （説明）

### 〔背景〕

私たちの体の細胞がすべて同じゲノムを持つのに、なぜ細胞の種類ごとに異なる形をしていて、特異的なはたらきができるのでしょうか。その鍵を握っているのは、細胞タイプごとに特異な遺伝子セットを活性化する遺伝子発現制御機構です。ゲノム上に存在するエンハンサー<sup>※1</sup>領域の活性化、DNAのメチル化状態<sup>※2</sup>、そしてゲノムに結合する約1,000種類の転写因子<sup>※3</sup>が互いに協調して、DNAの塩基配列を変えずに、遺伝子の機能を決めています。このしくみのことを「エピゲノム」といいます。

十数年前に次世代シーケンサー<sup>※4</sup>が開発されて以来、このエピゲノムのしくみを理解するために、さまざまなシーケンス技術が登場しました。その中で、ゲノム上の転写因子結合場所を調べるChIP-seq<sup>※5</sup>、転写因子が結合しうる活性化エンハンサー領域を調べるATAC-seq<sup>※6</sup>、DNAのメチル化状態を調べるBisulfite-seq<sup>※7</sup>が広く利用されています。これまでに論文などで数十万件のエピゲノムに関する実験データが報告されてきましたが、これらのデータの意味を理解して活用するためには、高度な解析技術と豊富な計算資源が必要でした。

### 〔研究の内容と成果〕

熊本大学と千葉大学の共同研究グループは、国立遺伝学研究所のスーパーコンピュータシステムを利用し、ヒト、マウス、ラット、ショウジョウバエ、線虫、出芽酵母の6つのモデル生物の約40万件のChIP-seq、ATAC-seq、Bisulfite-seqビッグデータを収集・計算しました（図1）。

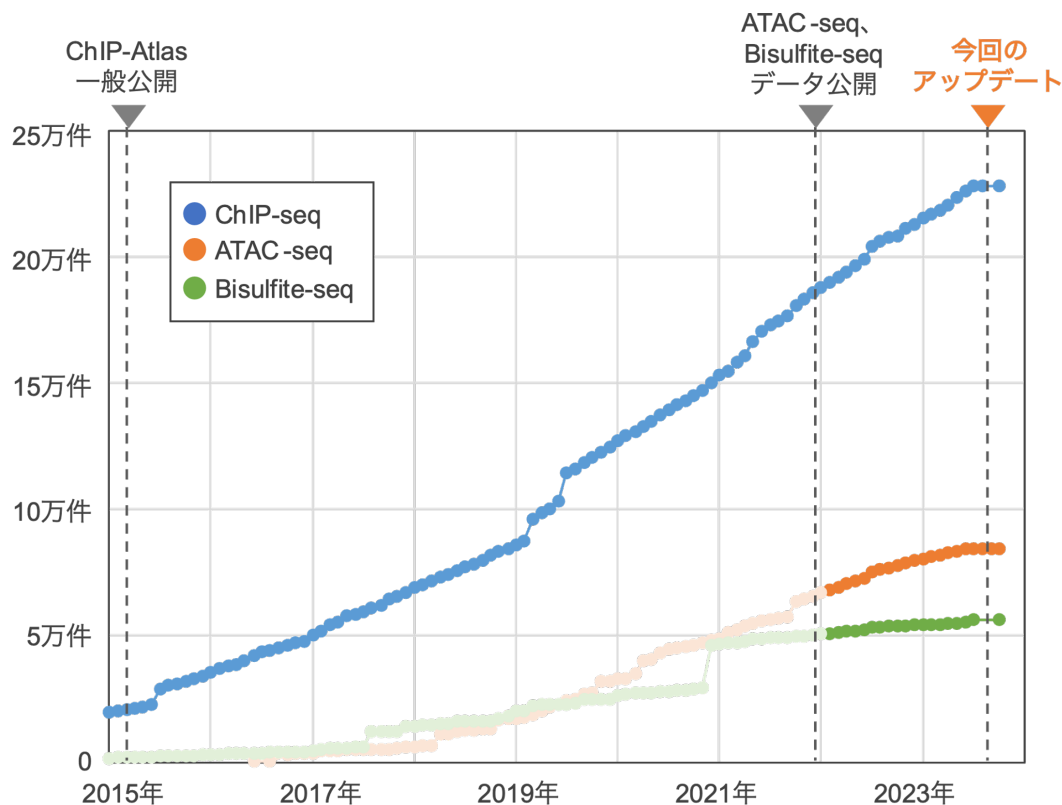


図1 ChIP-Atlas の公開データ

これは、これらのシーケンス技術の解析可能な公共データを ChIP-Atlas がほぼすべて網羅していることを意味します。ChIP-Atlas を利用することで、アトラス（地図帳）をめくるように自分の好きなゲノム領域にアクセスし、そこに結合する転写因子、活性化エンハンサー領域、DNAメチル化の情報をすべて閲覧できます。ChIP-Atlas のデータは、正常な細胞や組織だけでなく、遺伝子変異、薬物投与、加齢等による変容も含んでおり、これらはホームページからキーワード検索で簡単に入手できます。

これにより、ChIP-Atlas は世界最大規模のエピゲノム統合データベースという地位を確立しました。そこで共同研究グループは、後述の機能拡充を含め、このたび正式に論文として報告しました。

ChIP-Atlas による活性化エンハンサー領域と DNA メチル化状態の可視化により、従来多大なコストを必要とされたエンハンサーの特定を、ウェブ上の簡単な操作で予測できるようになりました。また、推定されたエンハンサー領域に結合する転写因子も視覚的に理解できるようになりました。しかしながら、これだけでは 1 つ 1 つのエンハンサーがどの遺伝子の発現制御に関わるかを特定することができません。これは、ゲノムは三次元構造を形成しているため、エンハンサーはその最も近傍の遺伝子を標的としているとは限らないからです。例えば、手足の元となる肢芽の形成に重要な *Shh* 遺伝子のスイッチとなるエンハンサーは、*Shh* 遺伝子から 80 万塩基以上離れたところに存在することが知られています。このような知見を

可視化するため、共同研究グループは ChIP-Atlas に「Annotation Track」として、さらにゲノム DNA の三次元構造を調べた実験データを統合しました。

ここからは利用事例を示しながら、Annotation Track について説明します。例えば、ATAC-seq と ChIP-seq データから、乳がんの発生に寄与するエストロゲン受容体をコードする *ESR1* 遺伝子の周辺に複数の活性化エンハンサー領域が存在することを理解できます（図 2）。

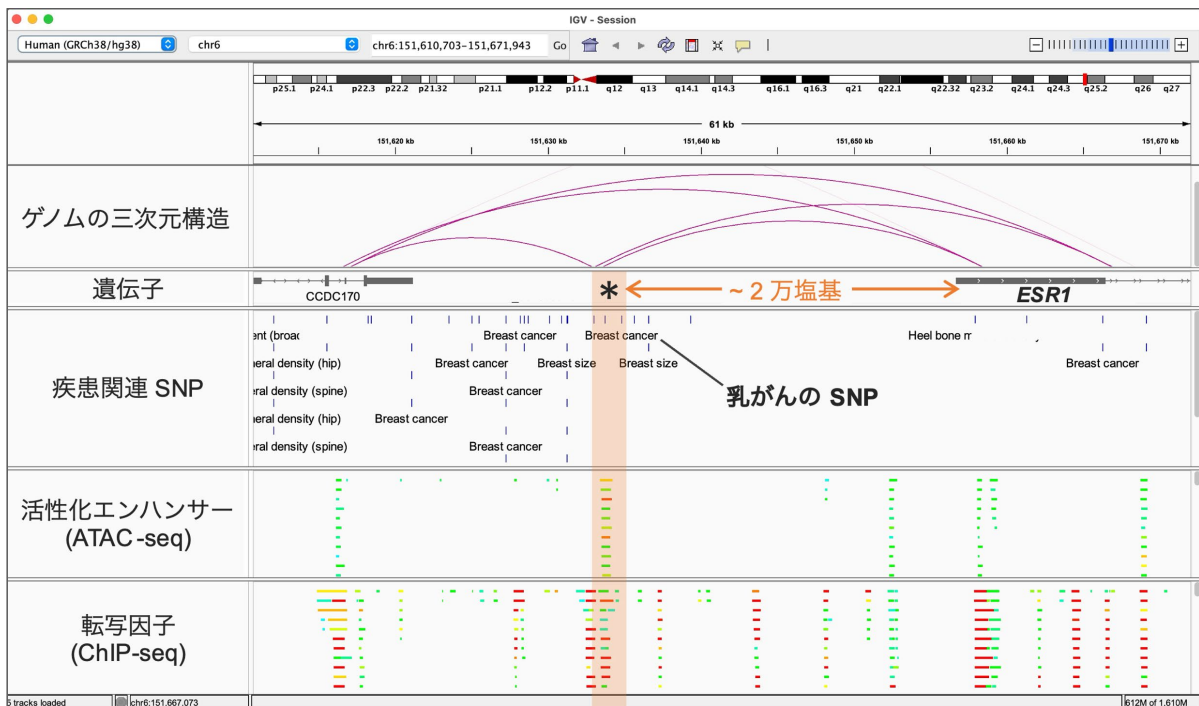


図 2 疾患関連 SNP 周辺エピゲノム状態の可視化

しかしながら、これだけでは *ESR1* の発現はどのエンハンサーによって調節されているかがわかりません。この度のゲノム三次元構造データの統合により、*ESR1* はその上流約 2 万塩基のゲノム領域（オレンジ）にコンタクトしていることがわかるようになりました。つまり、*ESR1* はおそらくこの領域に位置するエンハンサーによる制御を受けると推察できます。さらに今回のアップデートでは、疾患に関連するゲノム上の一塩基多型 (SNP) などの変異情報も Annotation Track に新たに追加されました。これにより、エンハンサー領域の位置情報に加え、そこに存在する疾患関連 SNP の情報を加味することで、疾患の背景にある遺伝子発現制御機構を一括で理解できるようになりました。*ESR1* の上流に位置する活性化エンハンサー領域（図 2、\* 印）には、乳がんの発症リスク上昇に関連する複数 SNP が報告されていることがわかります。このように、ウェブ上の簡単な操作のみで、*ESR1* 遺伝子のエンハンサー領域には乳がんの SNP があることが理解できます。このことからその SNP が *ESR1* のスイッチ役である転写因子群の結合異常をもたらす結果、エストロゲンの分泌異常が引き起こ

されるというメカニズムが推察できます。つまり、ChIP-Atlas の利活用により、ゲノム変異に起因する疾患の発症メカニズムに関するリーズナブルな仮説を形成することが可能です。

前述の通り、ChIP-Atlas によるデータ統合により、ゲノム上の転写因子結合、活性化エンハンサー領域、DNA メチル化などのエピゲノム状態を反映した実験データを視覚的に理解できるようになりました。また、このような可視化により、エピゲノム状態は細胞の分化段階、薬物を含む環境因子への曝露、疾患の発症などの影響を受け、ダイナミックに変化することがより鮮明にわかるようになりました。つまり、これらの生体内イベントのメカニズムに迫るためには、その分子基盤であるエピゲノム状態の変容を正確に評価することが重要ですが、これにもやはり情報学・統計学の知識と潤沢な計算資源が必要でした。そこで、ChIP-Atlas ではChIP-seq、ATAC-seq、Bisulfite-seq データの比較解析を支援するオンライン解析ツール「Diff Analysis」を実装しました。例えば多能性を有する ES 細胞と、筋肉の元となる筋芽細胞由来の ATAC-seq データを比較し、両方で活性化が異なるエンハンサー領域を簡単に抽出できます。図 3 で示すように、細胞の多能性を規定する *Oct4* 遺伝子の周辺は ES 細胞で特異的に活性化しており（オレンジ）、逆に筋肉のマーカー遺伝子である *Myod1* の近傍には筋芽細胞特異的な活性化エンハンサー領域（青）が存在することがわかります。このツールは、ChIP-Atlas にある数十万件の解析済み実験データによって成り立っています。これにより、ウェブ上で 2 群の実験データのアクセス番号を入力するだけで、わずか数分で計算が終了し比較結果を閲覧できるようになります。実験データの網羅性と操作の手軽さを兼ね備えたウェブサービスとしては世界初です。

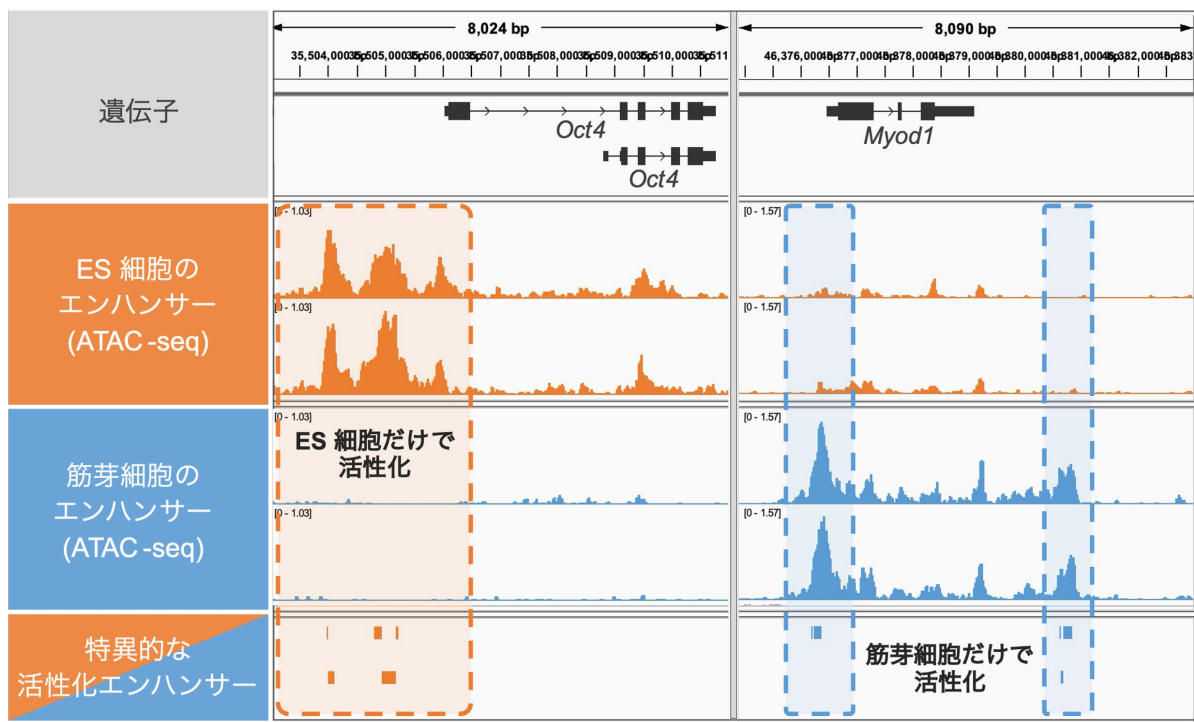


図3 エピゲノム状態の比較解析ツール

[展開]

ChIP-Atlas は 2015 年の一般公開以来、遺伝学、疾患メカニズム、創薬、発生生物学など幅広い研究分野で利用されており、これまでに 1,000 報に迫る論文に引用されています。この度のメジャーアップデートにより、疾患ゲノム情報をはじめとするゲノム・エピゲノム注釈づけデータが大幅に拡充され、ChIP-Atlas は遺伝子発現制御の異常に起因する遺伝性疾患の成り立ちを解明するデータ基盤へと進化しました。また、エピゲノムデータの比較解析ツールの実装は、薬物摂動によるエピゲノム状態の変容を突破口とした薬物の作用機序の解明、ひいては新規創薬標的の発見への応用が期待されます。さらに、細胞分化におけるエピゲノム変化に切り込むことで、例えば細胞の直接分化転換（いわゆるダイレクト・リプログラミング技術）<sup>※8</sup> を誘導できる司令塔的な転写因子もしくは低分子化合物をより簡便に見出すことが可能になります。将来的には組織再生医療への応用が期待されます。

[用語解説]

※1 エンハンサー

特定の遺伝子の発現を活性化する作用を持つ DNA 領域のことをいう。

※2 DNA のメチル化

シトシン塩基にメチル基が付加される可逆的な化学的修飾である。遺伝子のエンハンサー領域において、転写因子の結合を阻害し、遺伝子

の発現を抑制する役割を果たすとともに、遺伝子の安定性や構造の維持にも寄与することが知られている。

※3 転写因子

遺伝子の発現を制御する DNA 結合タンパク質のことをいう。エンハンサー領域などの特定の塩基配列を認識し、そこに直接結合することで、標的遺伝子の発現のオン・オフおよびその発現量の増減を制御する。

※4 次世代シーケンサー

断片化した 1 本鎖 DNA の相補鎖を合成しながら配列を決定できる Sequence-by-Synthesis 法に代表される手法を活用して DNA の塩基配列を読み出す装置である。従来型のサンガー法と比較して圧倒的に高速で大量のデータが得られる。

※5 ChIP-seq

断片化された DNA を転写因子などに対する抗体で免疫沈降し、得られた DNA 断片の配列解析により、目的の修飾ヒストンや転写因子の結合領域を特定するシーケンス技術である。

※6 ATAC-seq

Tn5 トランスポゼースの特性を利用し、ゲノム上の活性化エンハンサー領域を特定する技術である。Tn5 トランスポゼースは、DNA を切断し、その両端にシーケンスアダプターを付加する機能を持つ。

また、活性化されたエンハンサー領域にはトランスポゼースがアクセスしやすいことから、活性化エンハンサー領域の DNA のみがタグ化された断片として得られる。このタグ化された DNA 断片を増幅しシーケンスすることで、活性化エンハンサー領域の位置情報を得ることができる。

※7 Bisulfite-seq

ゲノム DNA を構成するシトシン塩基のメチル化状態を計測する技術である。細胞から抽出したゲノム DNA を重亜硫酸ナトリウム (bisulfite) で処理すると、ゲノム上のメチル化されていないシトシンのみがウリジンへと変換されるため、シーケンスによりシトシンのメチル化の有無を識別できる。

※8 直接分化転換 (ダイレクト・リプログラミング)

皮膚などの細胞にいくつかの転写因子を発現させ、肝臓、心筋、神経など別の分化状態に誘導する技術である。

(論文情報)

論文名 : ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data

著者 : Zhaonan Zou, Tazro Ohta, Shinya Oki

掲載誌 : Nucleic Acids Research

doi : 10.1093/nar/gkae358

**【お問い合わせ先】**

熊本大学 生命資源研究・支援センター  
機能ゲノミクス分野

担当 : 教授 沖 真弥

電話 : 096-373-6501

E-mail : okishinya@kumamoto-u.ac.jp