

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



理化学研究所



大学共同利用機関法人 情報・システム研究機構

国立遺伝学研究所

令和 8 年 5 月 8 日

報道機関 各位

熊本大学
千葉大学
理化学研究所
国立遺伝学研究所

10年続くエピゲノム基盤の進化 － 遺伝子発現制御の「司令塔」を解き明かす－

(ポイント)

- エピゲノム統合データベース ChIP-Atlas が公開 10 周年を迎え、50 万件近くの実験データを統合し、世界最大級の解析基盤へ発展しました。
- 最新のアップデートでは、データの信頼性を可視化する新機能と、遺伝子発現制御を統合解析する新しい解析モジュールを実装しました。
- 長期運用されるデータ基盤として、疾患研究や創薬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、千葉大学の 大田 達郎准教授（国立遺伝学研究所 BSI（バイオデータ研究拠点）/DBCLS（ライフサイエンス統合データベース部門）/特命准教授）、理化学研究所生命医科学研究センターの 粕川 雄也チームディレクター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>) の公開 10 周年に際し、メジャーアップデートを実施しました。収録データの拡充の他に、個々の実験データの信頼性を可視化する新機能に加え、遺伝子発現を網羅的に測る RNA-seq データの上流解析をより簡便に行える新しいオンラインツールを実装しました。これにより、利用者はデータの確からしさを直感的に確認しながら、真核生物における遺伝子制御の仕組みをより柔軟に解析できます。

このような改良を通じて、ChIP-Atlas は利用者の多様な需要により即したエピゲノム解析インフラへと進化しました。

本研究の成果は、グリニッジ標準時間 2026 年 4 月 29 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）（NBDC「統合化推進プログラム」、ERATO「有田リピドームアトラスプロジェクト」）、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」、（AMED-CREST）「解釈可能なゲノム機能シミュレータの創出」、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」、日本学術振興会科学研究費助成事業（24K23417、25K21342、23K24081、23H04954、23KF0048）の支援を受け実施したものです。

（説明）

[背景]

私たちの体の細胞は全て同じ DNA を持っています。それにもかかわらず、神経細胞や筋肉細胞のように、形も働きも大きく異なります。この違いを生み出しているのは、DNA 配列そのものを変えずに、転写因子^{*1}の結合パターンやエンハンサー^{*2}の活性化、DNA のメチル化^{*3}などによる遺伝子の働き方を調節する仕組みで、エピゲノム^{*4}といいます。

DNA 配列を効率的に読み取る高速シーケンシング技術^{*5}が進歩した結果、エピゲノムに関する実験データは急速に蓄積されてきました。多くのデータは公共リソースとして公開されており、誰でもアクセスできるようになっていますが、これを有効に利活用するには、データ解析の専門知識や大規模な計算資源が必要であるため、多くの研究者にとって高いハードルとなっていました。

このような状況の中で、ChIP-Atlas は 2015 年に公開されました。論文などで報告されたほぼ全てのエピゲノム実験データを共通の解析手順で解析し、専門家によって整理されたサンプルメタ情報と共に、ウェブサービスとして誰でも利用しやすい形で提供してきました。その結果、ChIP-Atlas は遺伝子発現制御の理解を支える世界最大規模の研究基盤へと発展しました。公開から 10 年間で、国内外 1,500 報を超える論文に引用されています。

[研究の内容と成果]

熊本大学、千葉大学、理化学研究所、国立遺伝学研究所 BSI/DBCLS の共同研究グループは、公開されているエピゲノム実験データを継続的に収集・解析し、統合データベース ChIP-Atlas として整備してきました。現在は、ヒトやマウスなど 6 生物種を対象に、約 50 万件・数百 TB の ChIP-seq^{※6}、ATAC-seq^{※7}、Bisulfite-seq^{※8} データを網羅しています。つまり、これら主要なシーケンス技術に関する公共データのほぼ全てをカバーする規模に達しています（図 1）。

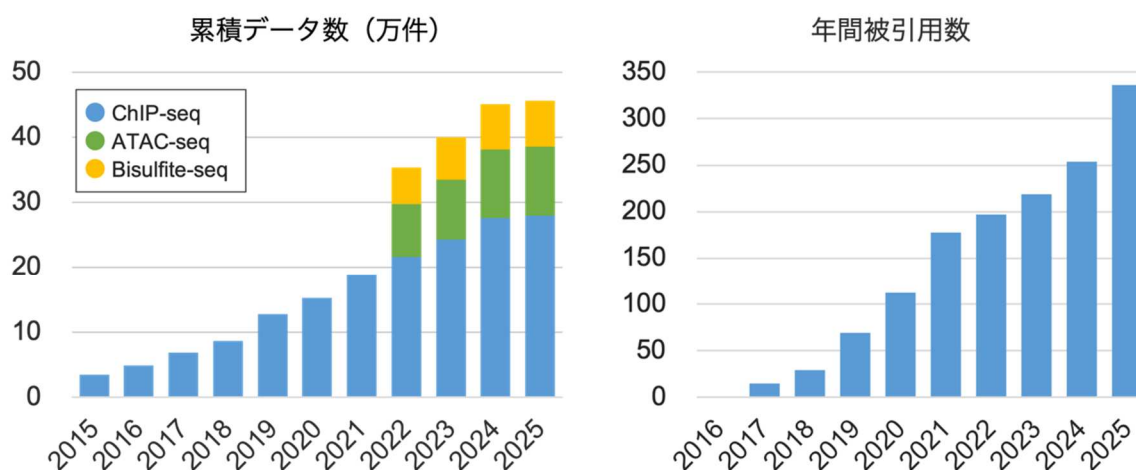


図1 ChIP-Atlas 10年のあゆみ

ChIP-Atlas を利用することで、研究者は高度な情報処理技術がなくても、ウェブ上の簡単な操作だけで、関心のある遺伝子やゲノム領域における転写因子の結合や DNA 状態を視覚的に確認できます。一方で、データが網羅的であるほど、品質が一様でないという課題も顕在化します。ChIP-Atlas では、解析過程で得られたデータ量^{※9}などの指標を公開しており、利用者がそれらを確認することで一定の品質評価は可能でした。しかし、その評価は体系化されておらず、最終的には利用者の経験や知識に依存していました。

この課題を解決するために、今回のアップデートでは、個々の実験データの信頼性を定量的に評価できる品質管理フレームワークを新たに導入しました。この機能では、まず各実験について、リード数やピーク数が同類の実験全体の中でどの位置にあるかを可視化します。これにより、そのデータが十分な情報量を持っているかどうかを把握できます。さらに、同じ条件で実施された他の実験と比較し、データの類似度も評価できるようにしました。これにより、そのデータが典型的な結果に近いのか、あるいは特徴的な挙動を示すのかを直感的に判断できます。

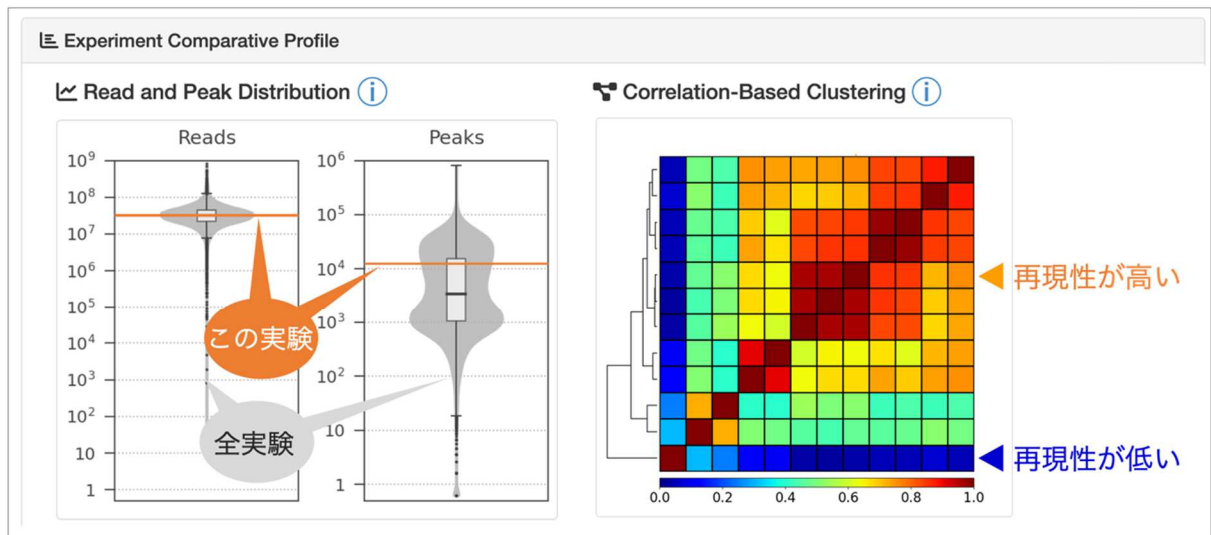


図2 実験データ品質管理フレームワーク

ここでは、肝臓由来の Hep G2 細胞で転写因子 HNF4A の結合を調べた ChIP-seq 実験 (SRX10829255) の例を用いて説明します (図 2)。実験の個別ページを開くと、下部に「Experiment Comparative Profile」と表示されたパネルがあり、これが本フレームワークによる評価結果です。左側の「Read and Peak Distribution」では、全ての ChIP-seq 実験におけるリード数とピーク数の分布が示されており、現在の実験の位置はオレンジの線で表示されます。この位置から、SRX10829255 は全体の中で平均以上の情報量を持つデータであることが分かります。しかし、情報量が多いことは必ずしも高品質を意味するわけではありません。そこで本機能では、「Correlation-Based Clustering」により、同じ Hep G2 細胞で HNF4A を対象とした他の実験と比較し、結果の類似度を評価します。図 2 の右側のパネルに示すように、対象データが主要な集団に含まれている場合、そのデータは典型的な結合パターンを反映している可能性が高く、再現性の高いデータと判断できます。一方で、同様の条件にもかかわらず他の実験と大きく異なる位置にある場合は、技術的なばらつきや特異的な要因の影響が考えられます。このように、本フレームワークにより、これまで専門知識や個別確認に依存していたデータの信頼性評価を、ウェブ上で体系的かつ直感的に行えるようになりました。

さらに ChIP-Atlas は、データを解析して公開するだけでなく、数十万件の実験データを統合し、横断的にデータマイニングを行える機能も備えています。その代表例が転写因子エンリッチメント解析です。この機能では、利用者が入力した遺伝子群に共通して結合する転写因子を網羅的に調べることによって、それらの遺伝子をまとめて制御する上流因子を特定することができます。例えば、薬剤投与の前後で発現が変動した遺伝子群を入力す

ることで、薬剤の作用機序の中心となる転写因子を推定することが可能です。このように、本機能は遺伝子発現制御メカニズムの理解に有力な手がかりを与えます。一方で、この解析では入力となる遺伝子群の定義が重要になります。実際に、発現変化の度合いに基づいて遺伝子を選別する際には、設定する基準によって対象が大きく変わることがあります。そのため、条件によっては十分な数の遺伝子が得られない場合や、変化が小さいものの重要な遺伝子が解析から外れてしまう可能性があります。

この課題に対応するため、今回のアップデートでは、RNA-seq^{*10} 実験で得られる遺伝子発現のカウントデータをそのまま入力として利用できる新たなエンリッチメント解析モジュールを導入しました。このモジュールでは、特定の遺伝子だけを選別するのではなく、全遺伝子の発現変化を連続値として統計的に扱います。そのうえで、蓄積された数十万件のエピゲノム実験データを参照し、各転写因子の標的遺伝子群が全体としてどの方向に変化しているかを評価します。これにより、発現変化の背後にある制御因子をより網羅的に推定することが可能となりました。

ここでは、この新たな解析モジュールを用いて、乳がん治療薬タモキシフェン（TAM）の作用機序を解析した例を紹介します（図3）。乳がん由来の MCF-7 細胞について、TAM 投与の有無で得られた RNA-seq カウントデータを用い、これをそのまま入力とします。解析結果はテーブル形式で出力され、各転写因子に対応する標的遺伝子群の変化の傾向が示されま

RNA-seq カウントテーブル

	投薬あり_1	投薬あり_2	投薬あり_3	投薬なし_1	投薬なし_2	投薬なし_3
<i>DDX11L1</i>	8	7	5	12	8	13
<i>WASH7P</i>	1,512	985	1,236	2,342	1,600	2,075
<i>FAM138A</i>	0	0	0	0	0	3
<i>OR4F5</i>	0	0	0	0	0	0
<i>DDX11L17</i>	4	5	3	11	8	11
<i>WASH9P</i>	2,011	1,327	1,758	3,031	2,174	2,735
<i>OR4F29</i>	1	1	0	0	1	2
<i>MIR12136</i>	4,156	2,678	3,469	6,927	5,240	6,825
<i>OR4F16</i>	15	10	20	21	20	28
<i>FAM87B</i>	19	15	14	39	29	37

解析結果

		転写因子		組織・細胞		統計量					
SRX16364104	TFs and others	ESR1	ESR1	Breast	MCF-7	7460	894/31605	-1.17/0.26	-243.4	-239.7	-33.37
SRX261274	TFs and others	ESR1	ESR1	Breast	MCF-7	6948	791/31605	-1.26/0.26	-242.6	-239.2	-33.31
SRX7030930	TFs and others	NELFA	NELFA	Breast	MCF-7	19348	10397/31605	-0.16/0.26	-240.7	-237.5	-33.18
SRX673753	TFs and others	ESR1	ESR1	Breast	MCF-7	8993	1140/31605	-0.99/0.26	-236.9	-233.9	-32.92
SRX730426	TFs and others	ESR1	ESR1	Breast	MCF-7	6929	794/31605	-1.24/0.26	-236.9	-233.9	-32.92
SRX3541102	TFs and others	ESR1	ESR1	Breast	MCF-7	10842	1227/31605	-0.94/0.26	-233.2	-230.2	-32.65
SRX1012747	TFs and others	ESR1	ESR1	Breast	MCF-7	8142	1189/31605	-0.95/0.26	-231.9	-229.0	-32.56
SRX5251415	TFs and others	ESR1	ESR1	Breast	MCF-7	6076	711/31605	-1.30/0.26	-229.1	-226.3	-32.37
SRX3285554	TFs and others	ESR1	ESR1	Breast	C49R	8716	1031/31605	-1.02/0.26	-222.2	-219.5	-31.88

図3 新しいエンリッチメント解析モジュール

す。その結果、複数の ChIP-seq 実験で同定されたエストロゲン受容体 ESR1 の標的遺伝子が、TAM 投与に応答して一貫して発現低下する傾向が検出されました。この結果は、TAM がエストロゲンシグナルを抑制するホルモン療法薬である、という既知の作用機序と一致しています。すなわち、本手法により、薬剤応答に伴う遺伝子発現変化を駆動する転写因子を推定できることが示されました。このように、ChIP-Atlas 上で、発現変化とその制御因子を一体として解釈できるようになり、薬剤作用の分子基盤の解明や創薬標的の同定などにつながることを期待されます。

[展開]

ChIP-Atlas が公開されてからの 10 年で、生体機能の分子基盤を担う遺伝子発現制御機構の理解を下支えし、遺伝学、疾患メカニズム解析、創薬、発生生物学など幅広い研究分野で利活用されてきました。今回のアップデートでは、個々の実験データの信頼性をシステムチックに評価できる品質管理機能と、遺伝子発現変化をもとに上流制御因子を特定する新たな統合的解析機能を実装しました。本機能拡張により、疾患発症の背景にある遺伝子発現制御の異常を体系的に捉え、その仕組みを解明する研究への応用が期待されます。また、薬剤に応答する遺伝子発現変化を手がかりとして、その作用機序の中核を担う転写因子を同定することで、新規創薬標的探索への応用が可能となり、創薬コストが高騰している現状を打破する突破口となることが期待されます。さらに、細胞分化や状態遷移における遺伝子発現制御を包括的に捉えることで、細胞運命を制御する転写因子やシグナル経路の同定が容易となります。これにより、細胞の性質を人為的に制御する技術の開発が進み、再生医療への応用も期待されます。

[用語解説]

※1 転写因子

遺伝子の発現を制御する DNA 結合タンパク質。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。

※2 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。

※3 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。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。

※4 エピゲノム

DNA の塩基配列を変えずに、遺伝子の働き方を調節する仕組み。細胞ごとの性質の違いに関わる。

※5 高速シーケンシング技術

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。従来型と比較して圧倒的に高速で大量のデータが得られる。

※6 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がどの遺伝子のどの領域に結合しているか特定するシーケンス技術。

※7 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。ゲノム上で「開いている（活性化している）領域」を特定する。

※8 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。

※9 データ量

ここでは主にリード数とピーク数のことをいう。リード数：シーケンサーで読み取った塩基断片の数；ピーク数：ゲノム上でリードが有意に集中している領域の数。

※10 RNA-seq

遺伝子発現量を網羅的に測定するシーケンス技術。逆転写酵素により細胞から抽出した RNA を cDNA に変換し、その断片を高速シーケンサーで読み取ることで、各遺伝子由来の配列の出現頻度を定量する。この解析により、各遺伝子について「どの程度発現しているか」を数値として表すことができる。最終的には、遺伝子ごとにサンプル間で比較可能なカウントテーブル（遺伝子 × サンプル行列）が得られる。

（論文情報）

論文名： ChIP-Atlas 2025 Update: 10-year anniversary of a data-mining platform for exploring epigenomic landscape

著者： Zhaonan Zou, Tazro Ohta, Takeya Kasukawa, Shinya Oki

掲載誌： *Nucleic Acids Research*

doi： 10.1093/nar/gkag378

URL : <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkag378/8664501>

【お問い合わせ先】

(研究に関するお問い合わせ)

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

(報道に関するお問い合わせ)

熊本大学総務部総務課広報戦略室

電話：096-342-3269

E-mail : sos-koho@jimu.kumamoto-u.ac.jp

千葉大学広報室

電話：043-290-2018

E-mail : koho-press@chiba-u.jp

理化学研究所 広報部 報道担当

Tel: 050-3495-0247

Email: ex-press@ml.riken.jp

情報・システム研究機構 国立遺伝学研究所
広報室

E-mail: prkoho@nig.ac.jp