

統合的ゲノム解析のための Web データベース”CohesinDB”を開発

1. 発表者：

王 健康（東京大学大学院医学系研究科 分子細胞生物学専攻 細胞生物学・解剖学講座
医学博士課程 4年）
中戸 隆一郎（東京大学定量生命科学研究所附属高度細胞多様性研究センター
大規模生命情報解析研究分野 准教授）

2. 発表のポイント：

- ◆ 遺伝子発現やゲノム立体構造形成に重要な機能を持つコヒーシン複合体に関連する既存データを網羅的に収集したデータベース “CohesinDB” を開発しました。
- ◆ CohesinDB には、ヒトの 176 細胞種から成る転写因子結合、遺伝子発現変動、ゲノム立体構造など計 2,043 サンプルが含まれており、Web ブラウザシステムを通じて可視化、解析、データダウンロードなどが誰でも簡単に行うことができます。
- ◆ CohesinDB はコヒーシン関連解析だけでなく、転写制御や疾患関連解析などに関する全ての研究者にとって有用なリソースとなります。

3. 発表概要：

コヒーシンはゲノムの立体構造制御や遺伝子発現制御に重要な役割を果たすタンパク質複合体です。コヒーシンの変異は急性骨髄性白血病や複数の先天性疾患の要因となることから、これらの疾患の発症メカニズムを解明するうえで、コヒーシンの果たす機能を明らかにすることが求められています。しかしながら、コヒーシンの機能はエピゲノム・転写制御・ゲノム立体構造制御など多岐にわたるため、網羅的な全ゲノム解析は難易度が高く、これまで困難でした。

東京大学大学院医学系研究科の王健康大学院生、定量生命科学研究所附属高度細胞多様性研究センター 大規模生命情報解析研究分野の中戸隆一郎准教授は、そのような統合的大規模解析の解析コストを低減するため、コヒーシンに関するゲノム解析データを格納した大規模データベース”CohesinDB” (<https://cohesindb.iqb.u-tokyo.ac.jp/>) を開発しました。本データベースでは、過去に論文として報告されデータが公開されている全ゲノムデータを網羅的に収集し、独自のパイプラインを用いて統一的に再解析のうえ、得られたコヒーシンに関連するエピゲノム領域（プロモーター・エンハンサーなど）、遺伝子、クロマチンループ（注1）などを統一的にラベル付けし、データベースとして公開しています（図1）。本データベースは Web ブラウザシステムを通じて可視化、解析、データダウンロードなどが誰でも簡単に行うことができます。これにより、情報学の非専門家にとってこれまで難しかった大規模ゲノム解析が容易になり、コヒーシン研究の一層の推進が期待されます。

本研究は、AMED-PRIME「頑健なデータ駆動形エピゲノム解析を実現する情報解析システムの構築」（代表：中戸隆一郎、JP22gm6310012h0003）、からの支援を受けて実施されました。

本研究は、2022年9月27日付で Nucleic Acids Research 誌に掲載されました。

4. 発表内容：

コヒーシンはゲノムの立体構造制御や遺伝子発現制御に重要な役割を果たすタンパク質複合体です。コヒーシンの変異は急性骨髄性白血病や複数の先天性疾患の要因となることから、これらの疾患の発症メカニズムを解明するうえで、コヒーシンの果たす機能を明らかにすることが求められています。次世代シーケンサー（NGS）を用いた全ゲノム解析により、様々な生物種・細胞種を用いた実験が世界中で展開されており、コヒーシンの多彩な役割が明らかにされつつあります。その一方で、公開されたこれら大量の論文データを再解析し、自身の研究に活用することは高度な情報解析技術が要求されるため、多くの研究者にとってこれまで困難でした。

東京大学定量生命科学研究所 中戸隆一郎准教授らの研究グループは、そのような統合的解析を推進するため、コヒーシンに関する既存ゲノムデータを格納した大規模なデータベース”CohesinDB” (<https://cohesindb.iqb.u-tokyo.ac.jp/>) を開発しました。CohesinDB には、176 のヒト細胞種から生成された 2,043 のエピゲノム、遺伝子発現、ゲノム立体構造データが含まれています。これを我々独自の解析パイプラインを用いて統一的に再解析し、コヒーシンに関連するゲノム情報に詳細なラベルを与えた「コヒーシンオブジェクト」を生成しました（751,590 のコヒーシン結合ゲノム部位、957,868 のクロマチンループ、2,229,500 のシス制御モジュール（注2）を含む）。これらのデータは Web ブラウザを介して誰でもアクセス可能であり、Web ブラウザシステムを通じて可視化、機能解析、データダウンロードなどを誰でも簡単に行うことができます。これにより、情報解析の非専門家にとってこれまで難しかった大規模ゲノム解析が容易になり、コヒーシン研究の促進が期待されます。本データベースは転写制御や疾患関連解析などに関する全ての研究者にとって有用なリソースとなります。

5. 発表雑誌：

雑誌名：Nucleic Acids Research（発行日：9月27日）

論文タイトル：CohesinDB: A comprehensive database for decoding cohesin-related epigenomes, 3D genomes and transcriptomes in human cells

著者：Jiankang Wang and Ryuichiro Nakato*

DOI 番号：10.1093/nar/gkac795

URL：https://doi.org/10.1093/nar/gkac795

6. 問い合わせ先：

東京大学定量生命科学研究所 大規模生命情報解析研究分野

准教授 中戸 隆一郎（なかと りゅういちろう）

TEL: 03-5841-1471 FAX: 03-5841-7308

E-mail: rnakato@iqb.u-tokyo.ac.jp

7. 用語解説：

（注1）クロマチンループ：ゲノム上の離れた2つの領域が立体的に相互作用している部位を「ループ」と呼ぶ。

（注2）シス制御モジュール：複数の転写因子が結合し、遺伝子発現を制御するためのゲノム領域の総称。主として遺伝子のプロモーター領域やエンハンサー領域が該当する。

8. 添付資料 :

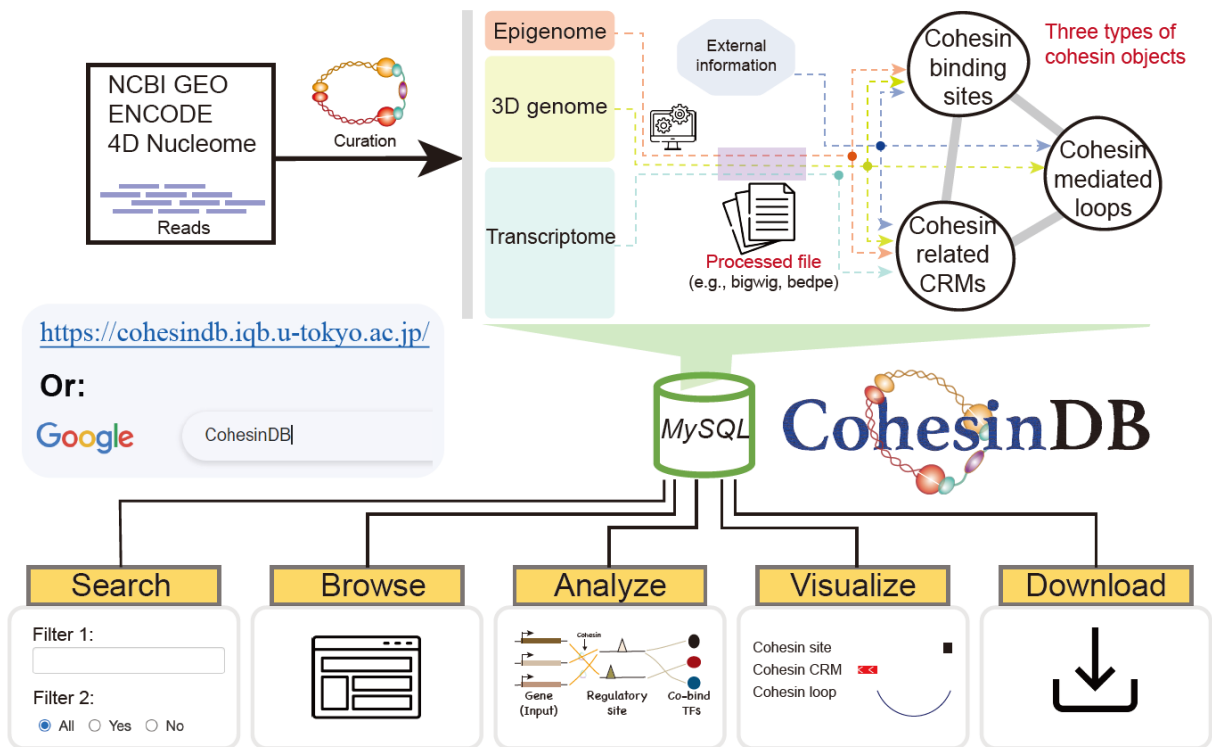


図1 : CohesinDBの概念図。CohesinDBは、エピゲノム、遺伝子発現、ゲノム立体構造などから成る計2,043のNGSデータを収集・処理した結果得られた3種類の「コヒーシンオブジェクト」をWebブラウザを通じて提供している。利用者はWebブラウザを通じてデータの検索、ダウンロード、可視化などが可能である。