

ゲノム立体構造のさまざまな特徴量を抽出する新規手法を開発 ～ゲノムにひそむ重要な機能領域の同定～

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 博士後期課程3年）
中戸 隆一郎（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 講師）

2. 発表のポイント：

- ◆ゲノム立体構造を網羅的に観測する Hi-C 法によって得られたデータから様々な次元特徴量を抽出可能な新規手法“HiC1Dmetrics”を開発しました。
- ◆HiC1Dmetrics では Hi-C 解析において用いられる既存指標の計算だけでなく、これまで定量的な計測が難しかった特殊な立体構造を抽出できる新規指標を提案しました。
- ◆本アプローチは、複数の Hi-C サンプルの比較、可視化、エピゲノムデータとの統合において優れたパフォーマンスを示しました。

3. 発表概要：

Hi-C 法はゲノム立体構造情報を全ゲノム的に得ることができる強力な手法です。一方でこの方法は計算量が多大である、二次元ヒートマップの可視化による視覚的な比較に頼らざるを得ない、エピゲノムデータとの統合が難しい等の問題点がありました。

東京大学大学院医学系研究科の王健康大学院生、東京大学定量生命科学研究所の中戸隆一郎講師らは、Hi-C データから多種多様な次元特徴情報を効率的に抽出可能な新規手法“HiC1Dmetrics”を開発しました (<https://h1d.readthedocs.io/en/latest/index.html>)。本手法では Hi-C 解析において用いられる種々の既存指標を統一的に計算できるほか、これまで同定が難しかった特殊な立体構造（クロマチンハブ（注1）等）を定量的に抽出できる新規指標を提案しました。これにより、多サンプル間の立体構造の定量的比較が容易になるだけでなく、計算に多大な時間を要する高解像度データに対しても高速・効率的に解析を行うことが可能になりました。本アプローチを用いた複数 Hi-C サンプルの網羅的比較により、ゲノム構造制御機構研究の推進が期待されます。

4. 発表内容：

Hi-C 法はゲノム立体構造情報を全ゲノム的に得ることができる強力な手法です。一般的な Hi-C 解析では、得られた全ゲノム的な相互作用情報を二次元行列形式で保存します。この二次元行列を可視化し（二次元ヒートマップ）、クロマチンループ、トポロジカルドメイン（注2）、コンパートメント（注3）など特定のゲノム構造情報を抽出し、それらをサンプル間で比較することによってゲノム構造の変動を調査するという方法が用いられてきました。一方でこの方法は計算量が多大である（二次元行列の解像度の二乗に比例）、サンプル間の立体構造の違いを二次元ヒートマップの可視化による視覚的な（恣意的になりがちである）同定に頼らざるを得ない、ChIP-seq 法などで得られたエピゲノムデータとの比較・統合が難しい等の問題点がありました。また、近年初めて同定された特殊なゲノム構造（Stripe TAD（注4）、クロマチンハブ等）を全ゲノム的に抽出することはこれまで困難でした。

今回、研究グループは、Hi-C 法によって得られた二次元行列データを入力に、多種多様な一次元特徴情報を効率的に抽出可能な新規手法“HiC1Dmetrics”を開発しました

(<https://h1d.readthedocs.io/en/latest/index.html>)。HiC1Dmetrics では Hi-C 解析において用いられる種々の既存指標 (第一主成分、Insulation score 等) を統一的に計算できるほか、Stripe TADs、クロマチンハブなどを定量的に網羅抽出できる新規指標 (intra-TAD score, adjusted interaction frequency) を提案しました。更にこれらの新規指標を ChIP-seq 法で得られたヒストン修飾データと統合し「クロマチン状態解析 (注5)」を実施することにより、クロマチンループを伴うエンハンサー領域と伴わないエンハンサー領域が存在すること、それらのゲノム分布が細胞種特異的であることなど、複数の知見が明らかになりました。

更に directional relative frequency という新規指標を提案し、トポロジカルドメイン間の相互作用変化と遺伝子発現変動の関連性を一部明らかにしました。本手法を用いることで、これまで難しかった多サンプル間の定量的な立体構造比較が容易になっただけでなく、ChIP-seq データとの統合による、より詳細なエピゲノム解析も可能であることが実験で明らかになりました。本手法は二次元行列データを直接用いる従来法に比べて高速・メモリ効率的であり、計算に多大な時間を要する高解像度データに対しても高速・効率的に解析を行うことが可能になります。本手法は Python3 パッケージとしてコマンドライン上でインストール可能である他、Web アプリケーションも提供されています (<http://hic1d.herokuapp.com/>)。本手法を用いてより多数・高解像度の複数 Hi-C サンプルを網羅的比較することが可能になり、ゲノム構造制御機構研究の推進が期待されます。

5. 発表雑誌：

雑誌名：Briefings in Bioinformatics

論文タイトル：HiC1Dmetrics: framework to extract various one-dimensional features from chromosome structure data

著者：Jiankang Wang and Ryuichiro Nakato*

DOI 番号：10.1093/bib/bbab509

URL：https://doi.org/10.1093/bib/bbab509

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

講師 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) クロマチンハブ：

他の多数のゲノム領域と相互作用を持つゲノム領域。

(注2) トポロジカルドメイン (TAD)：

ゲノムの立体相互作用のまとまりとして定義される 1Mbp 程度のゲノム領域。

(注3) コンパートメント：

Hi-C データに基づく主成分分析により、ゲノム全体をアクティブ領域・非アクティブ領域に分けたもの。それぞれ A コンパートメント・B コンパートメントと呼ばれる。

(注 4) Stripe TADs :

TAD 構造のうち、片側端に顕著な相互作用が認められるもの (図参照)。

(注 5) クロマチン状態解析 :

複数の 1 次元データ (通常はヒストン修飾データ) のパターンをもとに各ゲノム領域を特定のクロマチン状態 (プロモーター、エンハンサーなど) に分類する手法。

8. 添付資料 :

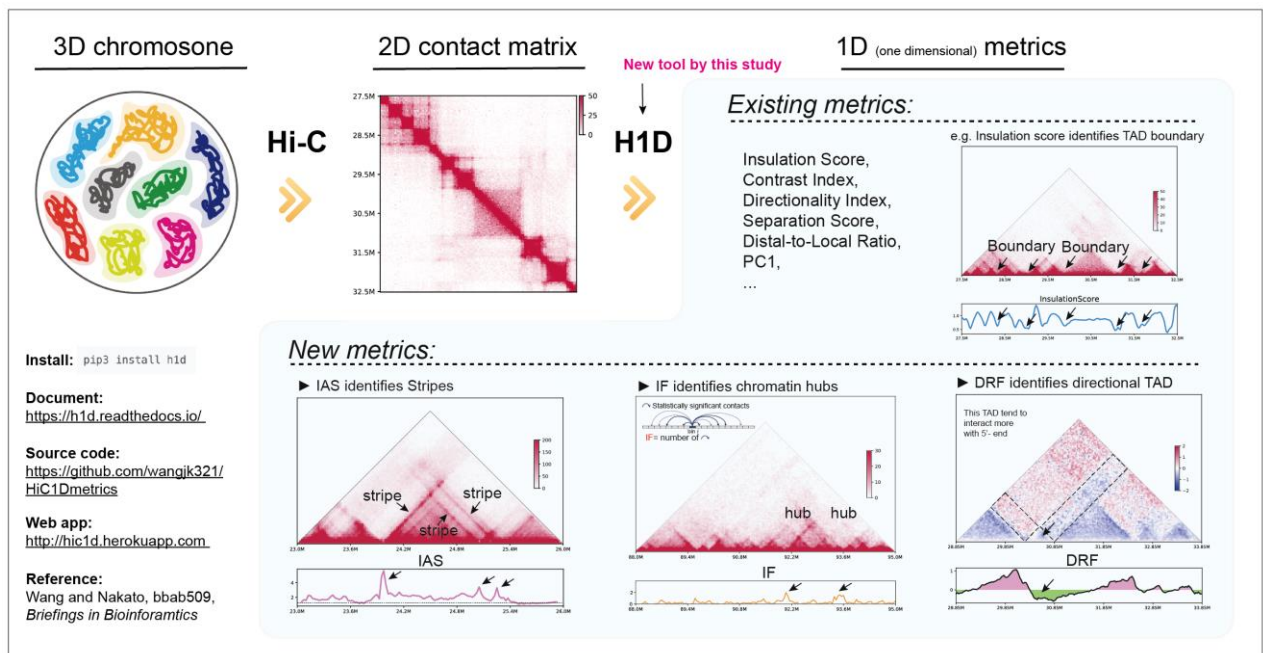


図 : HiC1DmetricsはHi-Cデータを入力に、多種多様な1次元特徴量 (新規手法含む) を計算・出力します。これにより、様々なゲノム情報を定量的に比較することが可能です。HiC1DmetricsはコマンドラインもしくはWebインターフェースから利用可能です。