

JST NBDC-DICP 2025年度採択課題（育成型）

4Dゲノム状態の理解と可視化 を支援するデータベースの構築

新海 創也

理化学研究所 生命機能科学研究センター 上級研究員

トーゴーの日シンポジウム2025「マルチモーダルデータ×AI」

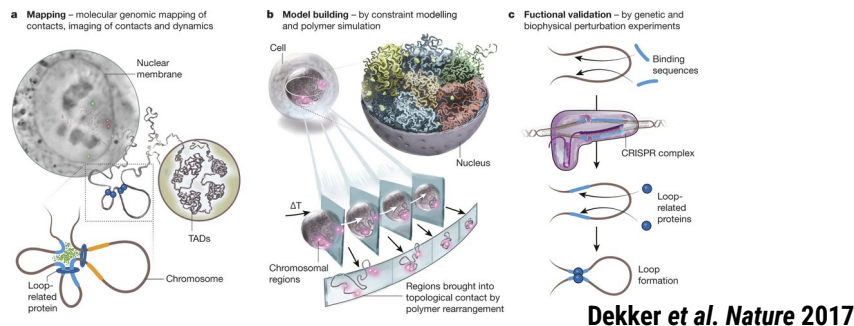
2025年10月20日

品川ザ・グランドホール

4Dヌクレオーム時代の3Dゲノムデータ基盤

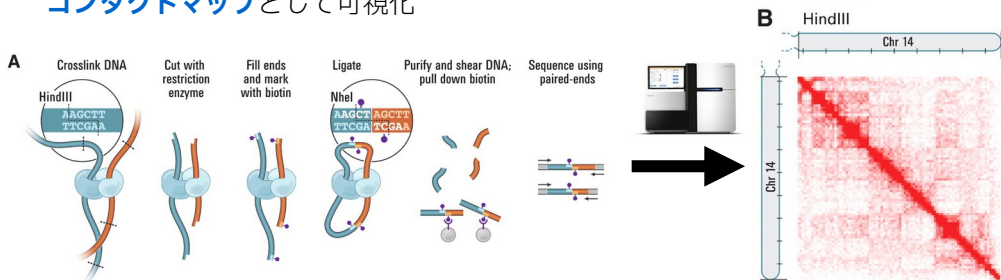
4Dヌクレオーム (4DN) 研究

時間軸 (1D) 上での3次元 (3D) 細胞核 (nucleus) 内のイベントを包括的に解析し、生命現象の理解を目指す学問領域



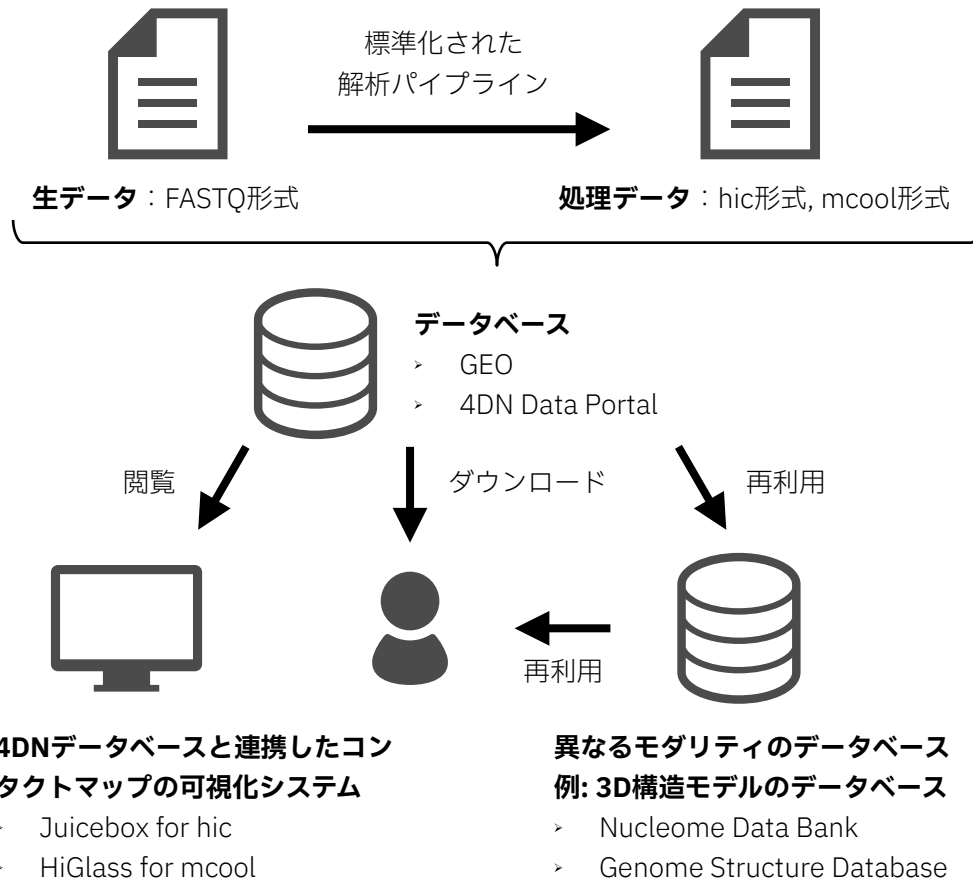
Hi-C技術の革新

次世代シーケンサーを用いてゲノムワイドに3次元ゲノム構造の特徴を検出し、**コンタクトマップ**として可視化



Lieberman-Aiden et al. Science 2009

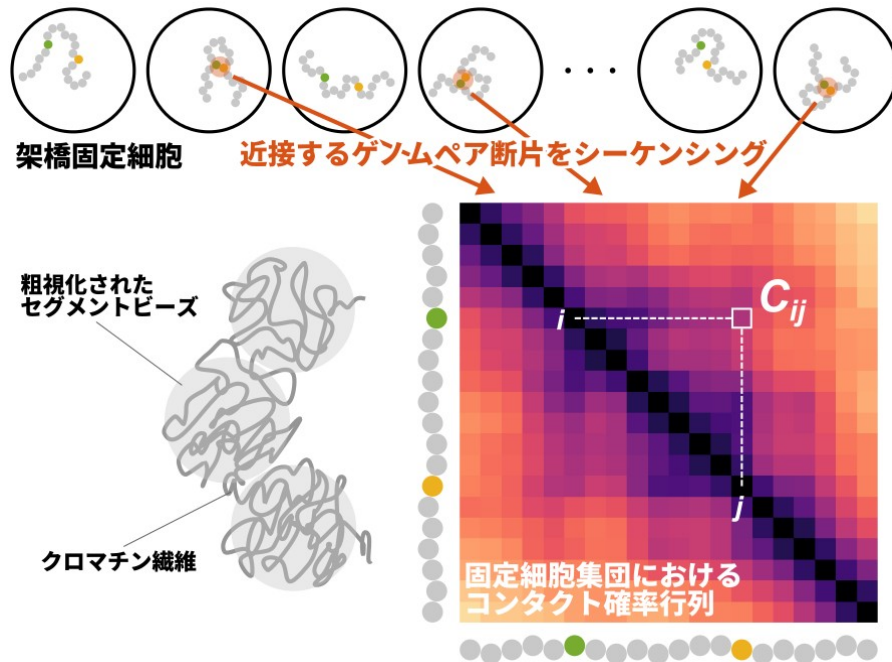
Hi-Cデータの蓄積・可視化・再利用システム



数理・物理モデリングの必要性：

Hi-Cデータは2次元、知りたいのは3D構造

- 最終的にHi-Cデータが描き出すのは**接触確率のマトリックス**
- 少なくとも知りたいことは、**その立体的な姿**



数理・物理モデリングと3D構造DBの群雄割拠

- 数理・物理モデリング手法は多数存在し、仮定や計算方法がそれぞれ異なる
- 再現性の評価指標が手法間で統一されておらず、比較が困難
- 3D構造DBも複数存在し、収録対象の偏りがある

Method "available online"	Representation	Scoring	
		U_{3c} $F_{ij} \rightarrow D_{ij}$ conversion	Functional form
ChromSDE* [37]	Points	$D_{ij} = \begin{cases} (\frac{1}{F_{ij}})^{\alpha} & \text{if } F_{ij} > 0 \\ \infty & \text{if } F_{ij} = 0 \end{cases}$ α is optimized	$\sum_{(i,j) D_{ij} < \infty} \frac{(r_{ij}^2 - D_{ij}^2)}{D_{ij}} - \lambda \sum_{(i,j)} r_{ij}^2$ where λ is set to 0.01
ShRec3D* [38]	Points	$D_{ij} = \begin{cases} (\frac{1}{F_{ij}})^{\alpha} & \text{if } F_{ij}' > 0 \\ \infty & \text{if } F_{ij}' = 0 \end{cases}$ F_{ij}' is the original F_{ij} corrected to satisfy all triangular inequalities with the shortest path reconstruction	N/A
TADbit* [43]	Spheres	$D_{ij} \propto \begin{cases} \alpha F_{ij} + \beta & \text{if } F_{ij} < \gamma' \text{ or } F_{ij} > \gamma \\ \frac{s_i + s_j}{2} & \text{if } i - j = 1 \end{cases}$ α and β are estimated from the max and the min F_{ij} , from the optimized max distance and from the resolution. $\gamma' < \gamma$ are optimized too. s_i is the radius of particle i	$\sum_{(i,j)} k_{ij} (r_{ij} - D_{ij})^2$ where $k_{ij} = 5$ if $ i - j = 1$ or proportional to F_{ij} otherwise
BACH* [45]	Points	$D_{ij} \propto \frac{B_i B_j}{r_{ij}^{\alpha}}$. The biases B_i and B_j and α are optimized	$b_{ij} D_{ij}^{1/\alpha} + c_{ij} \log(D_{ij})$ where b_{ij} and c_{ij} are optimized parameters
Giorgetti et al. [40]	Spheres	Particles interact with pair-wise well potentials of depths B_{ij} and contact radius a , which is larger than a hard-core radius and smaller than a maximum contact radius. The parameters are optimized over all the population of models	
Duan et al. [41]	Spheres	$F_{ i-j } = \frac{\sum_{k,q} F_{ik,jq}}{N- i-j }$ is the average of F_{ij} at genomic distance $ i - j $ expressed in kb. $D_{ij} = \overline{F}_{ i-j } \times 7.7 \times i - j $ assuming that α 1 kb maps onto 7.7 nm	$\sum_{(i,j)} (r_{ij} - D_{ij})^2$
MCMC5C* [49]	Points	$D_{ij} \propto \frac{1}{r_{ij}^{\alpha}}$ where α is optimized	$\sum_{(i,j)} (F_{ij} - r_{ij}^{-1/\alpha})^2$
PASTIS* [47]	Points	$D_{ij} \propto \frac{1}{r_{ij}^{\alpha}}$ where α is optimized	$b_{ij} D_{ij}^{1/\alpha} + c_{ij} \log(D_{ij})$ where b_{ij} and c_{ij} are optimized parameters

Serra et al. FEBS Lett 2015

Nucleome Data Bank

対象データ数：29件のHi-Cデータ

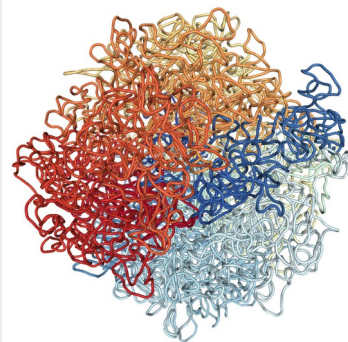
構造数：染色体ごとに数百個の3D構造

可視化：ブラウザ上でインタラクティブ閲覧可能

推定法：エピジェネティックマーカー + MD

一致度：Hi-Cとのピアソン相関 ≈ 0.95

課題：動的变化を考慮せず、静的構造のみ



Genome Structure Database

対象データ数：32件のHi-Cデータ

構造生成：染色体番号、数値アルゴリズム、解像度を指定

出力：各アルゴリズムごとに1構造のみ

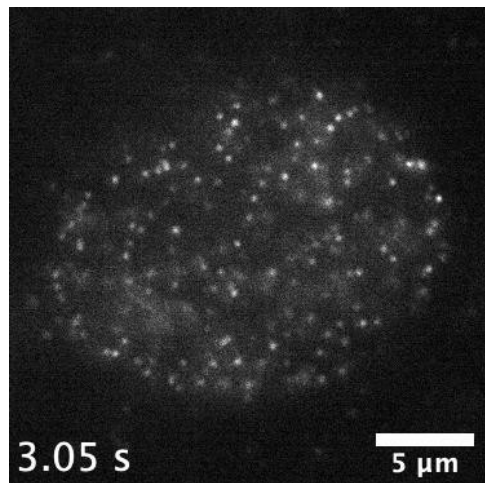
一致度：入力Hi-Cとの一致度定量評価なし

課題：アルゴリズム依存で不自然な構造が出る場合あり、動的变化を考慮しない

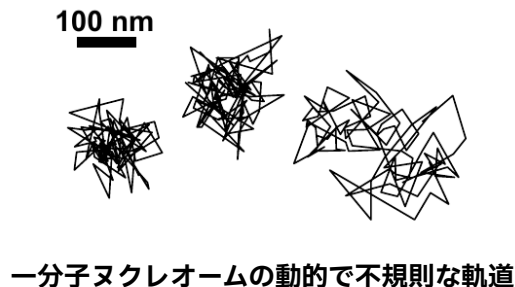


3D構造では不十分：時間軸を含めたゲノム理解へ

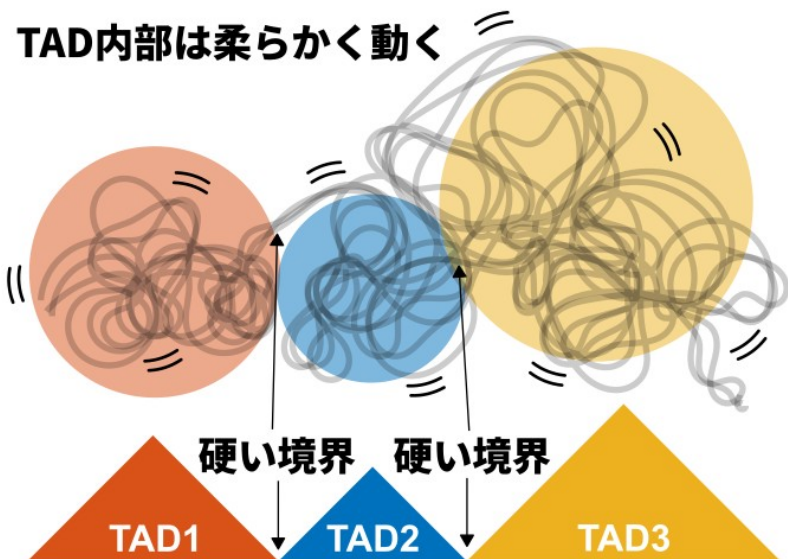
- ・ クロマチン構造は生細胞内で時間とともに揺らいでいる
- ・ TADは「ある／ない」ではなく、「どれくらいの時間、機能し続けるか」で語るべき構造
- ・ 境界の“硬さ”＝保持される時間の長さという定量的発想が重要
- ・ 真のゲノム構造理解は、時間軸を含む“4D構造”の視点から始まる



Shinkai, Maeshima et al. *PLOS Comput Biol* 2016
Shinkai, Maeshima et al. *Nucleus* 2017
Iida, Shinkai, Maeshima et al. *Sci Adv* 2022
Nozaki, Shinkai, Maeshima et al. *Sci Adv* 2023



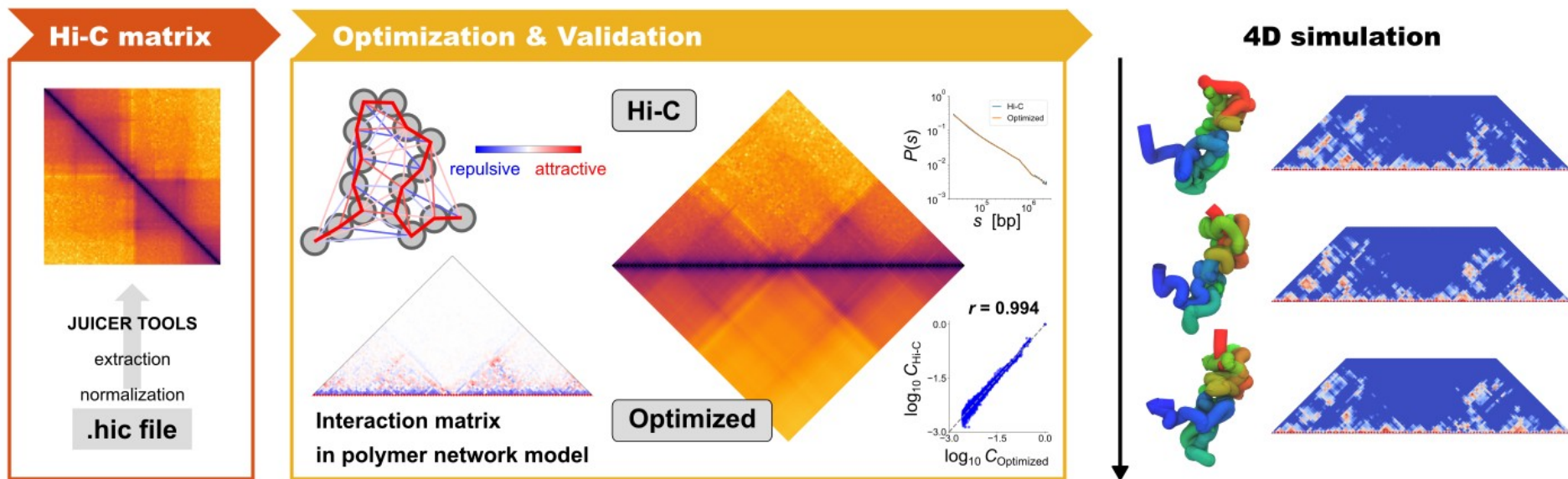
生細胞内での一分子ヌクレオソーム観測



Shinkai et al. *Biophys J* 2020
Shinkai et al. *Comput Struct Biotech J* 2020

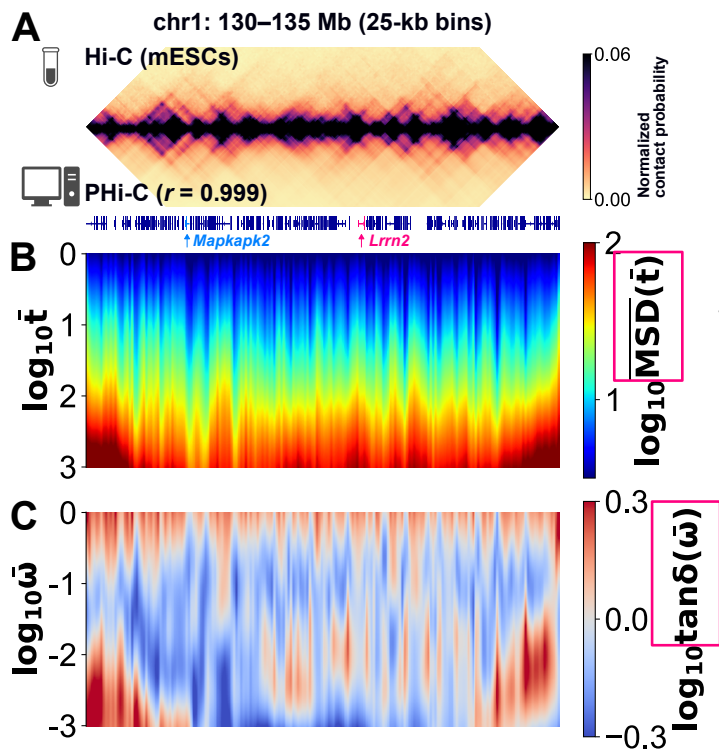
PHi-C (ファイシー) が拓く4Dゲノム解析の可能性

- Hi-Cデータをピアソン相関99%以上で高精度に再現
- 4Dシミュレーションにより、構造の時間的変動を表現



PHi-C（ファイシー）が拓く4Dゲノム解析の可能性

- Hi-Cデータをピアソン相関99%以上で高精度に再現
- 4Dシミュレーションにより、構造の時間的変動を表現
- ゲノム領域ごとの“動きやすさ”や“硬さ”を可視化する4D情報ヒートマップを出力



各ゲノム領域の“動きやすさ”の指標

各ゲノム領域の“液体・固体的”の指標

PHi-Cの現状と課題

現状

査読付き論文での実績

- ・ PHi-C論文掲載（NAR Genomics and Bioinformatics 2020年, 引用**48件**）
- ・ PHi-C2論文掲載（Bioinformatics 2022, 引用**13件**）

国際的な研究コミュニティでの評価

- ・ 代表ソフトに選出（米国4D Nucleome ハッカソン, Top5に採用）

ソフトウェアの利用実績

- ・ **17,000回以上**のダウンロード（PHi-C2, 2025/10/13現在）

downloads

17k

downloads/month

397

downloads/week

27

目標

- ・ PHi-CをHi-C解析の**国際標準へ発展**させる
- ・ 4Dゲノム状態の観点から**新たな生物学的意義を解明**する

課題

普及・理解の課題

- ・ 普及不足（宣伝不足で広まっていない）
- ・ 利用者の偏り（Pythonソフト → ドライ向け中心）
- ・ ウェット研究者に伝わらない（直感的な提示が必要）

技術的な課題

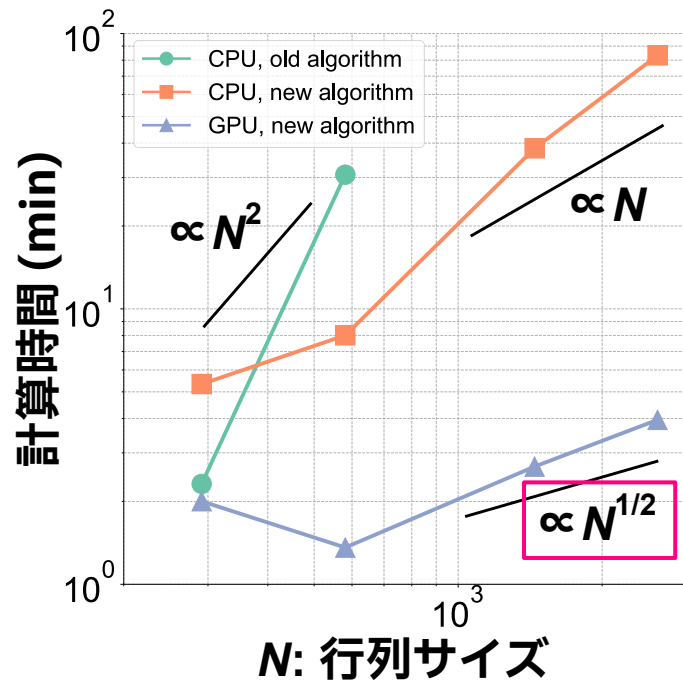
- ・ 計算コストが大きい（ヒト染色体200Mb → 50 kb解像度で4000×4000行列）
- ・ 可視化が分離（4D情報ヒートマップと動態シミュレーション）

技術課題の部分的解決

計算コストの課題

- 4000×4000行列（ヒト染色体を50kb解像度で解析）
- Phi-Cの最適化計算がボトルネック

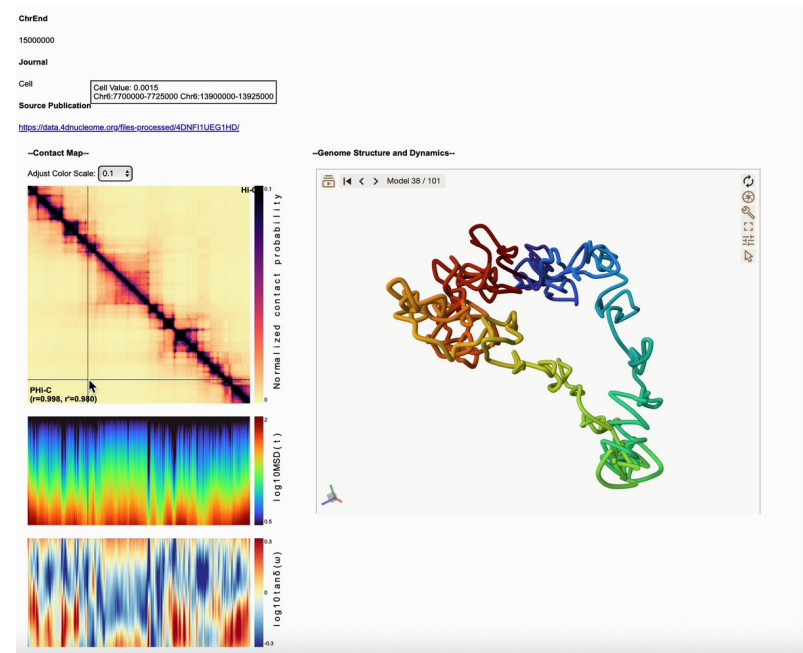
理論整備とGPU実装で高速化と並列化を実現



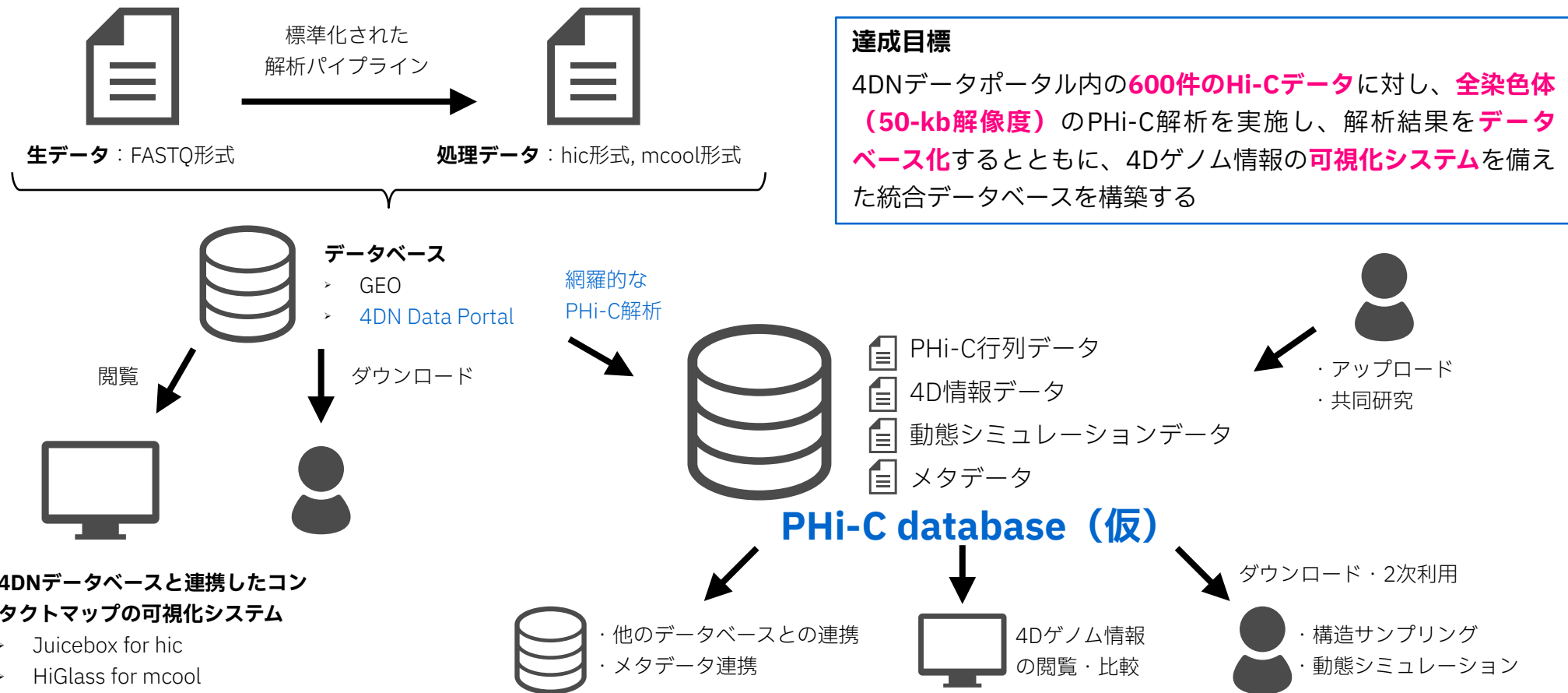
可視化の分離の課題

- 4D情報ヒートマップ → マニュアル操作
- 動態シミュレーション → VMDソフトで表示

同一ブラウザ上での統合可視化を実現



4Dゲノム情報の蓄積と活用に向けた PHi-Cデータベースの必要性



AI時代のゲノム構造研究フロンティアに立つPHi-C

nature

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

[nature](#) > [news](#) > article

NEWS | 25 June 2025

DeepMind's new AlphaGenome AI tackles the 'dark matter' in our DNA

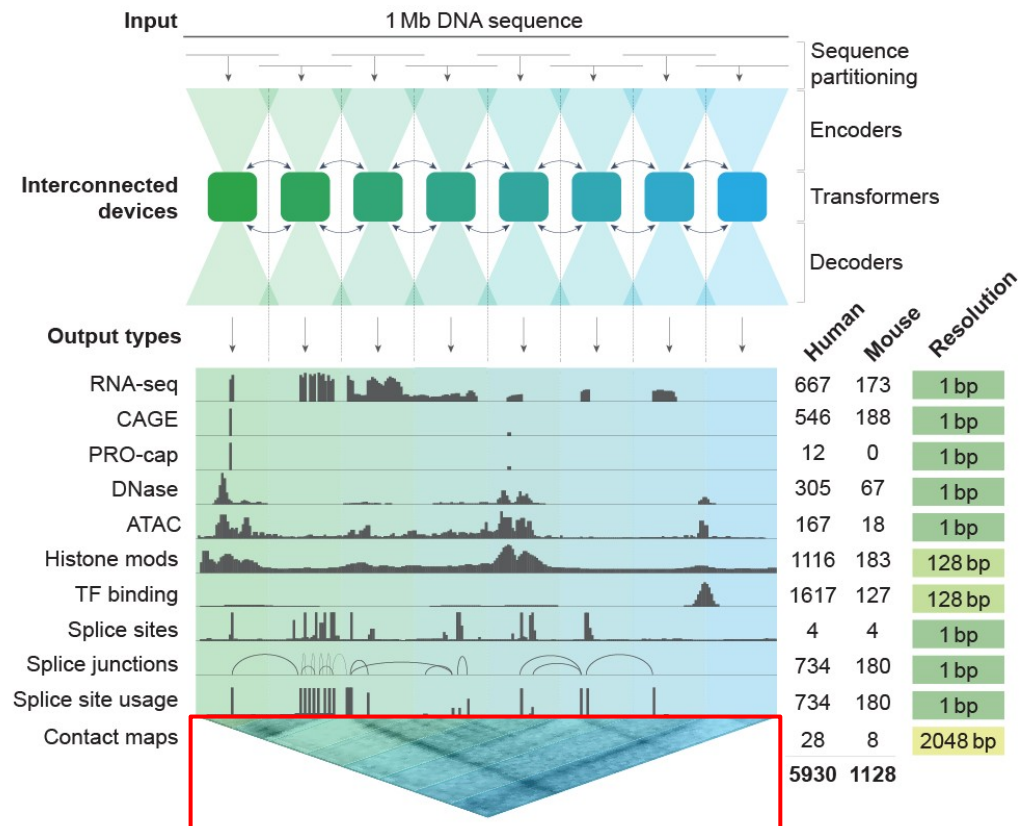
Tool aims to solve the mystery of non-coding sequences – but is still in its infancy.

By [Ewen Callaway](#)



AlphaGenome

from Google DeepMind



Acknowledgments



RIKEN BDR

ONAMI, Shuichi

ITOGA, Hiroya

KYODA, Koji

NAKAJIMA, Hiroki

HAGIWARA, Soya

ISHIDA, Teruo

Kyushu Univ

OCHIAI, Hiroshi



Genome
Modality

