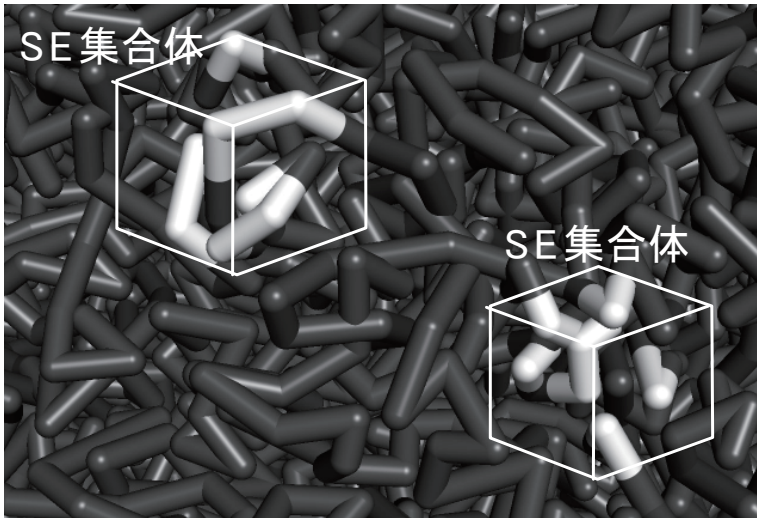
187　新規3Dゲノム構築法によるがん特異的転写集合体の同定	椙下　紘貴

【目的】近年、遺伝子発現制御において、クロマチンの空間的な制御が重要であることが次々と明らかになった。スーパーエンハンサーを含む「転写ファクトリー」は相分離し、核内ボディとして複数の集合体として核内に点在することが示された。しかし、それぞれの核内集合体にどのゲノム領域が集合しているのか、いかなる制御があるのか、については網羅的に調べる手法がなく、理解が進んでいなかった。そこで本研究は一細胞 Hi-C によって得られるコンタクト情報からゲノムの三次元空間マップを作成する独自の手法を用いて、そしてこれにエピゲノム情報を重ねることでスーパーエンハンサーをはじめとする個々の核内集合体の特徴の分類が可能であることを示す解析基盤の構築を目指した。

【方法】本研究では公共データベースで登録されている GM12878 細胞（Lymphoblastoid cell line：ヒトリンパ芽球様細胞株）の一細胞 Hi-C のコンタクトデータを三次元染色体構造の再構築に用いた。再構築した三次元染色体構造に対して ENCODE プロジェクトで収集された GM12878 細胞の各種ヒストン修飾（H3K27m3、H3K4me3、H3K36me3、H3K27ac、H3K9me3）の情報から分類した特徴情報を照らし合わせることでスーパーエンハンサーをはじめとする個々の核内集合体の実現性の実証を目指した。

【結果】GM12878 細胞の一細胞 Hi-C（No.06）のコンタクト情報から再構築した染色体構造は球体の形状を示し、三次元構造の構築を行うことができた。また、GM12878 細胞の各種ヒストン修飾を分類することによって、それぞれ転写活性化領域やヘテロクロマチンなどの詳細の分類ができた。この分類を再構築した三次元染色体構造に対して対応づけると、それぞれのクラスターは比較的集合している傾向にあり、またそれぞれのクラスターごとに密に集合している様な領域も見られた。これにより、三次元構築した染色体構造に対して、エピゲノム情報を対応づけることで、空間的な分類ができる可能性は見出した。今後、GM12878 細胞を含め、他のがん細胞種でも一細胞 Hi-C を実施し、自身で取得したデータを用いて、がん特異的な SE 集合体の探索を決定し、モチーフ解析を行うことで上流因子を同定していく予定である。

スーパーエンハンサー（S E）集合体や核内集合体の決定手法の開発



187. 新規 3D ゲノム構築法によるがん特異的転写集合体の同定

椙下　紘貴

東京大学 国際高等研究所 ニューロインテリジェンス国際研究機構

Key words：転写制御，スーパーエンハンサー，染色体高次構造，エピジェネティクス

緒　言

近年、遺伝子の発現を制御するメカニズムとして（転写因子だけでなく）クロマチン構造自体が大きな役割を果たしており、核内における空間配置が転写制御に重要であることが明らかになってきた［1］。例えば、エンハンサーが遺伝子プロモーターと近接するループを形成することが転写開始に重要である［2］。また近年、特に強く活性化しているエンハンサー（スーパーエンハンサー：SE）が核内で集合体を形成し、相分離していることが報告された［2］。SE は、細胞の運命決定に関与する重要な遺伝子群の制御に関わることが多い。例えば、がん細胞においては、そのがんの原因遺伝子の転写が SE に依存することが数多く報告されている。しかしながら、ひとつの SE 集合体の中にいかなるゲノム領域が参加しているのか（機能的に関連した遺伝子群が集合しているのか)、それぞれの SE 集合体に参加している遺伝子群は同時に共通のメカニズムで制御されているのか、は不明である。これは、従来の技術では、2 点以上のゲノム領域間の近接を網羅的に 1 細胞レベルで調べる手法が存在しなかったことが主な原因であった。そこで本研究では独自の手法を用いた一細胞 Hi-C によって得られるコンタクト情報からゲノムの三次元空間マップ［3〜6］を用いて染色体高次構造を再構築することで、各遺伝子が核内のどこに配置しているかを座標上で網羅的に特定できることに着目した。そしてこれにエピゲノム情報を重ねることで SE をはじめとする個々の核内集合体にどの遺伝子が参加しているかが初めて明らかになり、モチーフ解析などにより注目する核内集合体を制御する因子の予測ができることになる。エピゲノム情報を重ねることでSE をはじめとする個々の核内集合体の特徴（ユークロマチン、ヘテロクロマチンなど）の分類が可能であることを示せれば、網羅的に核内の転写制御で大きな役割を果たす転写ファクトリーを同定する解析手法として有用である。本研究ではこの実証性を示すことを目標とし、モデルケースとしてがん細胞である GM12878 細胞での転写活性化領域の同定を試みた。

方法および結果

まず、がん細胞において転写集合体やスーパーエンハンサーなどの核内ボディを分類できるかを明らかにするために、本研究では公共データベースで登録されている GM12878 細胞（Lymphoblastoid cell line：ヒトリンパ芽球様細胞株）　の一細胞 Hi-C のコンタクトデータを三次元染色体構造の再構築に用いた［7］。GM12878 細胞はさまざまな研究で用いられてきた一般的な細胞株であるため Hi-C データだけではなく遺伝子発現やエピゲノム情報（ChIP-Seq や ATAC-Seq など）のデータも多く収集されている。今回、大規模なエピゲノムデータ収集プロジェクトである ENCODE プロジェクトによって得られた GM12878 細胞のエピゲノム解析データが使用できることもあり、GM12878 細胞が核内凝集体の分類に適していると考えた。まず、GM12878 細胞の一細胞 Hi-C（No.06）のコンタクト情報を取得し［3，4］、の染色体三次元構造構築手法を用いることで、2 kb の解像度での染色体構築をおこなった（図 1）。本解析によって得られた座標情報をプロットすると球体の形状を示した染色体三次元構造の再構築が確認できた。

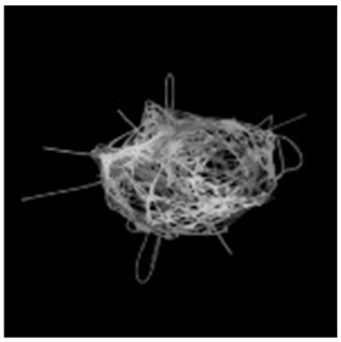
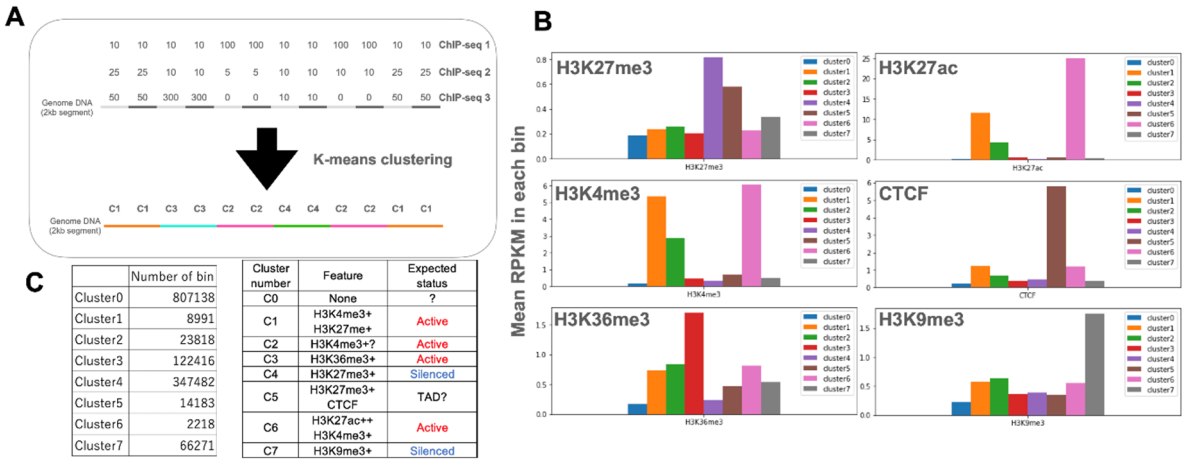


図 1. 独自手法によって一細胞 Hi-C データから再構築した GM12878 細胞の染色体構造
GM12878 細胞の細胞（No.06）を用いて染色体三次元構造を再構築し染色体の座標を 3 次元プロットした図。

再構成した GM12878 細胞の染色体三次元構造に対して、それぞれの核内空間がどのような特徴的な制御を有するかを検証するために、ENCODE プロジェクトによって収集されたヒストン修飾（H3K27m3、H3K4me3、H3K36me3、H3K27ac、H3K9me3）や染色体の高次構造形成に関わることが知られる CTCF の ChIP-Seq のデータを収集した。H3K27me3（ヒストン H3 の 27 番目リジンのトリメチル化）はポリコームによる抑制性ヒストン修飾、H3K4me3（ヒストン H3 の 4 番目リジンのトリメチル化）はユークロマチンにおけるプロモーターのヒストン修飾、H3K36me3（ヒストン H3 の 36 番目リジンのトリメチル化）、H3K27ac（ヒストン H3 の 27 番目リジンのアセチル化）は転写活性化、H3K9me3（ヒストン H3 の 9 番目リジンのトリメチル化）は抑制性のヘテロクロマチン領域の修飾として知られている。

GM12878 細胞の染色体三次元構造に対してこれらのヒストン修飾の情報を各点に対してカウントし、k-means クラスタリングによってそれぞれの空間座標がどのような特徴を有するかを分類した（図 2）。各染色体座標でクラスターを 8 つ分類すると H3K27me3 が集積したクラスター（C4、C5）、H3K4me3 が集積したクラスター（C1、C6）、H3K36me3 が集積したクラスター（C3）、H3K27ac が集積したクラスター（C1、C6）、CTCF が集積したクラスター（C5）、H3K9me3 が集積したクラスター（C）に分類できた。これらのクラスターの特徴から、各クラスターは図 2C の様な特徴を有すると考えられ、C1 または C6 が転写活性化に関わる座標であることが推定される。



C

	Number of bin
Cluster0	807138
Cluster1	8991
Cluster2	23818
Cluster3	122416
Cluster4	347482
Cluster5	14183
Cluster6	2218
Cluster7	66271

Cluster number	Feature	Expected status
C0	None	?
C1	H3K4me3+ H3K27me+	Active
C2	H3K4me3+?	Active
C3	H3K36me3+	Active
C4	H3K27me3+	Silenced
C5	H3K27me3+ CTCF	TAD?
C6	H3K27ac++ H3K4me3+	Active
C7	H3K9me3+	Silenced

図 2. GM12878 細胞における各染色体空間の特徴の分類
A） 染色体空間の特徴を分類するスキーム。
B） 分類後のクラスターごとのヒストン修飾の平均集積量（PKM 値）。
C） 各クラスターに分類された点（座標）の数（左）と、各クラスターで推定される転写状態の特徴（右）。

次に、ヒストン修飾や CTCF によって分類したクラスター情報を再構築した GM12878 細胞に対して対応づけを行なった（図 3)。それぞれのクラスターの空間位置関係を明確に分類することは困難であったが、それぞれのクラスターは比較的集合している傾向にあり、またそれぞれのクラスターごとに密に集合している様な領域も見られた。さらに、Cluster 6（転写活性化領域）と Cluster 7（ヘテロクロマチン領域）の情報を重ねて図示すると、Cluster 7 は集合して存在しているのに対して、Cluster 6 では Cluster 7 と近接する領域が少なく、比較的染色体が粗な領域に点在している傾向が見られユークロマチン空間としての特徴を示しているのかもしれない。

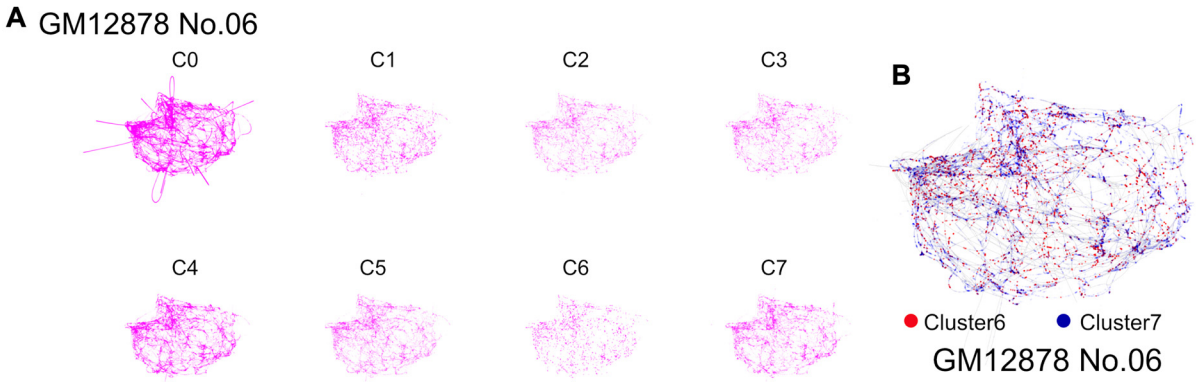


図 3. GM12878 細胞の 1 細胞（No.06）に分類したクラスター情報をカラーリングした図

A) GM12878（No.06）における各クラスターの空間位置。

B) GM12878（No.06）における Cluster 6（赤）と Cluster 7（青）の空間位置。

考　察

今回の研究では、一細胞 Hi-C から再構築された染色体三次元構造を用いて、染色体座が空間的にどの様な特徴を有するかを解析し、その実証性を検証した。GM12878 細胞の染色体三次元構造を再構築し、その構造に対してヒストン修飾などのエピゲノム情報で分類したそれぞれのクラスターが GM12878 細胞の染色体三次元構造内でどのように位置しているかを調べた。特定のクラスターは比較的集合して存在する傾向があり、またそれぞれのクラスターごとに密に集合している様な領域も見つかった。さらに、転写活性化領域とヘテロクロマチン領域を比較することで、それぞれの領域に特有の空間的特徴があることが示唆された。今回の解析で三次元構築した染色体構造に対して、エピゲノム情報を対応づけることで、空間的な分類ができる可能性は見出したが、しかし、今回用いた GM12878 細胞のデータのクオリティが十分でなく、細密な染色体構造の構築ができなかった。また 2 kb の解像度では距離行列の計算にかかる計算量が莫大であるため近接情報から核内凝集体を定義することが困難であった。現在、核内凝集体を定義するための解析手法の開発を行っている。これまでに ES 細胞や神経幹細胞において、自身で一細胞 Hi-C データの取得を行っており、現在、GM12878 細胞を含め、他のがん細胞種でも一細胞 Hi-C の実施を試みている。今後、自身で取得したデータを用いて、がん特異的な SE 集合体の探索を決定し、モチーフ解析を行うことで上流因子を同定していく予定である。この解析によって同定される「がん特異的 SE 集合体」の遺伝子群ならびに上流因子はがん化に貢献している可能性が高く、がん化の鍵因子を同定する全く新しい手法につながるポテンシャルがあり、それによって、将来、がん細胞におけるクロマチン制御異常の理解が進むことが期待されるとともに、「いかなる SE 集合体が形成されているか」に基づく新たな診断方法や治療薬の選択、予後の予測などへの応用も期待できると考えている。

共同研究者・謝辞

本研究の共同研究者は、東京大学大学院薬学系研究科分子生物学教室の後藤由季子教授、筑波大学システム情報系情報工学域の平田祥人准教授である。

文　献

1) Cook PR. The organization of replication and transcription. Science [Internet]. 1999 Jun 11;284(5421):1790–1795. Available from: http://dx.doi.org/10.1126/science.284.5421.1790 PMID: 10364545

2) Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, Li CH, Guo YE, Day DS, Schuijers J, Vasile E, Malik S, Hnisz D, Lee TI, Cisse II, Roeder RG, Sharp PA, Chakraborty AK, Young RA. Coactivator condensation at super-enhancers links phase separation and gene control. Science [Internet]. 2018 Jul 27;361(6400). Available from: http://dx.doi.org/10.1126/science.aar3958 PMCID: PMC6092193

3) Hirata Y, Oda A, Ohta K, Aihara K. Three-dimensional reconstruction of single-cell chromosome structure using recurrence plots. Sci Rep [Internet]. 2016 Oct 11;6:34982. Available from: http://dx.doi.org/10.1038/srep34982 PMCID: PMC5057099

4) Hirata Y, Kitanishi Y, Sugishita H, Gotoh Y. Fast reconstruction of an original continuous series from a recurrence plot. Chaos [Internet]. American Institute of Physics; 2021 Dec 1;31(12):121101. Available from: https://doi.org/10.1063/5.0073899

5) Hirata Y, Oda AH, Motono C, Shiro M, Ohta K. Imputation-free reconstructions of three-dimensional chromosome architectures in human diploid single-cells using allele-specified contacts. Sci Rep [Internet]. 2022 Jul 11;12(1):11757. Available from: http://dx.doi.org/10.1038/s41598-022-15038-4 PMCID: PMC9273635

6) Kitanishi Y, Sugishita H, Gotoh Y, Hirata Y. Changes of chromosomal architecture before establishment of chromosome territories revealed by recurrence plot reconstruction [Internet]. bioRxiv. 2022 [cited 2023 Apr 30]. p. 2021.05.20.444916. Available from: https://www.biorxiv.org/content/10.1101/2021.05.20.444916v2 doi.org/10.1101/2021.05.20.444916

7) Tan L, Xing D, Chang CH, Li H, Sunney Xie X. Three-dimensional genome structures of single diploid human cells. Science [Internet]. American Association for the Advancement of Science; 2018 Aug 31 [cited 2021 May 7];361(6405):924–928. Available from: https://science.sciencemag.org/content/361/6405/924. editor-summary PMID: 30166492