

2025 年 1 月 14 日

ヒトゲノムにおけるシス制御エレメントの大規模機能解析 ーゲノムに秘められた機能解読を目指してー

概要

ヒトゲノム中には、「シス制御エレメント」と呼ばれる DNA 配列が無数に存在し、近傍に存在する遺伝子の転写を制御しています。シス制御エレメント¹の変異は、ヒトの疾患や進化の要因と考えられており、ENCODE² プロジェクトなどの国際コンソーシアムプロジェクトにより大規模に解析が進められています。しかしながら、無数に存在するシス制御エレメントの活性、あるいはその変異がおよぼす影響を網羅的に解析することは非常に困難でした。

そこで、ワシントン大学 Vikram Agarwal 博士（現：米国サノフィ社）、Jay Shendure 教授、カリフォルニア大学サンフランシスコ校 Nadav Ahituv 教授、京都大学高等研究院ヒト生物学高等研究拠点（WPI-ASHBi）井上 詞貴 特定准教授、張 子聡 同研究員らの国際共同研究グループは、大規模並列レポーター解析法 (lentiMPRA)³ を応用し、ヒトゲノム中のシス制御エレメントを網羅的に同定しました。さらに、この解析で得られた大規模データを機械学習⁴ 訓練に用いることで、任意の塩基配列情報からその制御活性の予測を可能としました。これらの成果により、これまで困難であった個人差や疾患が生じる分子メカニズムの理解につながることを期待されます。

本成果は、2025 年 1 月 15 日 16:00（グリニッジ標準時、日本時間 1 月 16 日午前 1 時）に国際学術誌「Nature」にオンラインで掲載されました。

<研究の背景>

ヒトゲノム中には、タンパク質の設計図である遺伝子以外にも、さまざまな機能を持った領域が存在しています。例えば、ある遺伝子の発現量は、すぐ上流にあるプロモーター配列と、その働きを活性化するエンハンサー配列、さらにそこに結合する転写因子⁶によって制御されています。この遺伝子発現の制御によって、細胞の“特異性”がもたらされます。プロモーターとエンハンサーは、「シス制御エレメント」と呼ばれる配列で、特にエンハンサーは疾患や進化との関連が示唆されています。しかしながら、約 30 億塩基対存在するゲノム中に無数に存在するシス制御エレメントの活性を網羅的に定量することは非常に困難でした。

<研究手法・成果>

lentiMPRA (lentivirus-based Massively Parallel Reporter Assay)は、「転写バーコード」(15bp のランダムな塩基配列)を利用することで、数万種類のシス制御エレメントの活性を同時に測定できる技術です。本研究ではこの技術を応用し、ヒトゲノム中のシス制御エレメント“候補”領域の転写活性化能を網羅的に解析し、活性を持つ領域の同定と、その活性化能を定量しました。解析では、3種類のヒト培養細胞株(肝細胞、血球系細胞、iPS細胞⁵)を用い、ヒトゲノム中の全てのシス制御エレメントをカバーすると考えられる約 68 万の候補配列について、その配列特異性と細胞特異性を網羅的に評価しました。その結果、シス制御エレメントの中でも、プロモーターは、配列の方向に依存するが、細胞特異性は低く、一方、エンハンサーは、配列の方向には依存しないが、細胞特異性が高いということを明らかとしました。

さらに、この lentiMPRA 解析で得た大規模データを学習する機械学習モデル「MPRALegNet」を開発しました。MPRALegNetにより、任意のシス制御エレメント内の変異や多型が、その活性にどのように影響するのかを予測できるようになり、その活性を担う転写因子の結合モチーフの決定を可能としました。その結果、例えば肝細胞では HNF4 モチーフ、血球系細胞では GATA モチーフが細胞特異性に重要であることがわかりました。本研究成果から、細胞特異性を決定する分子メカニズムやその制御機構についてさらに理解が進むことが期待されます。

<展望>

ヒトのゲノム配列は個人個人で少しずつ異なっており、多くの多型⁷が存在します。本研究で確立したシス制御エレメントの同定・定量およびその予測技術を、このような多型の機能解析に応用することで、疾患や個人差、進化が生じる分子メカニズムの理解に繋がると期待されます。

<研究プロジェクトについて>

本研究は、Vikram Agarwal 博士(研究開始当時ワシントン大学、現：米国サノフィ社)、井上 詞貴 博士(研究開始当時カリフォルニア大学サンフランシスコ校 (UCSF)、現：京都大学高等研究院 ヒト生物学高等研究拠点 (WPI-ASHBi) 特定准教授)、Nadav Ahituv 教授 (UCSF) および Jay Shendure 教授 (ワシントン大学) らの共同研究グループによる成果で

あり、NIH（米国立衛生研究所）および NHGRI（米国立ヒトゲノム研究所）等の支援を受けて行われました。

<研究者のコメント>

近年、ヒトゲノムのほぼ全てが解読されていますが、明らかとなった DNA 配列にどのような機能情報が隠されているのかは、未だ解明されていません。今回の論文は、ゲノムの DNA 配列情報とその機能情報を直接結びつける成果であり、ヒト疾患や進化など様々な生命現象の理解に利用されることを期待しています。（井上 詞貴）

<用語解説>

1. シス制御エレメント: エンハンサーやプロモーターなど、遺伝子の転写を制御する塩基配列。
2. [ENCODE](#): The Encyclopedia of DNA Elements。米国立ヒトゲノム研究所 (NHGRI) が支援し、世界 5 カ国が参加するヒトゲノム解析プロジェクト。
3. lentiMPRA (lentivirus-based Massively Parallel Reporter Assay): 大規模並列レポーター解析法。シス制御エレメントの活性により転写されるランダム 15 塩基配列 (バーコード) を次世代シーケンサーにより定量することで、数万種類のシス制御エレメントの活性を同時に測定できる技術。
4. 機械学習: コンピュータが大量のデータから学習し、データの背景にある規則性やパターンを発見する方法。
5. iPS 細胞: 人工多能性幹細胞。体細胞へ少数の遺伝子を導入することで、多くの細胞に分化できる多分化能と自己複製能を持たせた細胞。
6. 転写因子: シス制御エレメントなど特定の DNA 配列に結合し、遺伝子の転写を制御するタンパク質の一群。
7. 多型: SNP (一塩基多型、single-nucleotide polymorphism) 等、同じ生物種において個体間で異なっている塩基配列。

<論文書誌情報>

Agarwal, V*, Inoue, F*, Schubach, M., Penzar, D., Martin, B. K., Dash, P. M., Keukeleire, P., Zhang, Z., Sohota, A., Zhao, J., Georgakopoulos-Soares, I., Noble, W. S., Yardımcı, G. G., Kulakovskiy, I. V., Kircher, M., Shendure, J., & Ahituv, N. (2025). Massively parallel characterization of transcriptional regulatory elements. *Nature*. DOI: [10.1038/s41586-024-08430-9](https://doi.org/10.1038/s41586-024-08430-9)

**These authors contributed equally to this work*