

転写因子が結合する塩基配列の新たな基盤データを構築

ヒトの遺伝子発現を制御する転写因子が結合する塩基配列の基盤データ「MOCCS プロファイル」を新たに構築し、転写因子が細胞の種類ごとに特異的な結合配列を持つことを明らかにしました。また、これを応用し、遺伝的変異が転写因子の DNA 結合に与える影響を評価する方法を確立しました。

ヒトの身体を構成する多種多様な細胞の特徴は、遺伝子発現の違いによって現れます。このような遺伝子発現の制御は、ゲノム上で特異的な塩基配列と結合する転写因子によって成り立っており、細胞の種類ごとに転写因子が結合する配列（転写因子結合配列）を明らかにすることは、それぞれの遺伝子発現の制御メカニズムの解明に重要です。しかしながら、これまで、転写因子の種類や細胞の種類に横断的な共通性や多様性といった、転写因子結合配列の全体像は明らかになっていませんでした。

本研究では、大規模なヒト転写因子の結合部位に関するデータを用いて、転写因子結合配列の新たな基盤データ「MOCCS プロファイル」を構築し、転写因子および細胞の種類横断的に、転写因子結合配列の解析を行いました。その結果、解析した約半数の転写因子は、細胞の種類ごとに特異的な結合配列を持つことが明らかとなりました。さらに、MOCCS プロファイルを応用して、一塩基多型（SNP）が転写因子の DNA 結合に与える影響を予測する指標を開発し、転写因子・細胞型の観点から疾患関連 SNP が転写因子結合に与える影響を適切に評価できることを示しました。

今回構築した MOCCS プロファイルは、エピゲノムのデータ等と組み合わせて、細胞型特異的な遺伝子発現制御メカニズムの理解につなげたり、がん細胞に生じた体細胞変異が転写因子の結合に与える影響度の評価など、多方面での活用が期待されます。

研究代表者

筑波大学医学医療系
尾崎 遼 准教授

研究の背景

ヒトのゲノム^{注1)}は約 30 億塩基対から構成され、数万種類の遺伝子の情報や遺伝子発現^{注2)}（遺伝子のオン・オフ）の制御に関する情報を含んだ、人体の設計図としての役割を担っています。ヒトの身体は形態や機能の異なるさまざまな細胞で構成されており、細胞の種類ごとに特異的な遺伝子発現パターンがあります。この遺伝子発現の制御を担うのが転写因子というタンパク質で、ゲノム中の特定の塩基配列に結合することで遺伝子発現を制御します。この際、転写因子が結合する塩基配列は「転写因子結合配列」と呼ばれます。ヒトでは転写因子が約 1000 種類あると考えられており、また、それぞれの転写因子が異なる転写因子結合配列を持つため、細胞の種類ごとの転写因子結合配列を明らかにすることは、それぞれの細胞が持つ遺伝子発現制御メカニズムの解明につながると期待されます。しかしながら、これまで、転写因子結合配列の細胞横断的な多様性の調査は、大規模には行われていませんでした。

研究内容と成果

本研究では、ヒトの転写因子の結合部位に関する ChIP-seq^{注3)}データベース ChIP-Atlas に収録されているデータを用いて、転写因子結合配列の新たな基盤データを構築しました。具体的には、大規模なヒト転写因子 ChIP-seq データセットに対して、転写因子結合配列を特定する手法 MOCCS 法^{注5)}を適用し、塩基の部分配列 k -mer^{注4)}ごとの結合特異性スコアをまとめた「MOCCS プロファイル」を、473 種類の転写因子と 398 種類の細胞型（20 種類の組織型）について求めました。ここでの組織型は、血液系・循環器系・肝臓などの臓器レベルの分類を指し、細胞型には組織型の亜分類で、実験で用いられる細胞株などが含まれます。これを用いて、転写因子結合配列を転写因子・細胞の種類横断的に解析したところ、解析した転写因子のうち約半数は、細胞の種類ごとに特異的な結合配列を持つことが明らかとなりました。このことは、同じ転写因子であっても、細胞の種類ごとに異なる結合配列を結合する傾向があることを示します。

さらに、MOCCS プロファイルから、転写因子の結合部位に存在する一塩基多型（SNP）^{注6)}が転写因子結合に与える影響を予測する「 Δ MOCCS2score」という指標を開発しました。この指標を用いて、クローン病^{注7)}に関連する GWAS-SNP^{注8)}が転写因子結合に与える影響を評価すると、すでに知られている転写因子 FOS や JUN が、影響度の高い転写因子として挙がりました。また、全身性エリテマトーデス^{注9)}についても、既存の知見と比較して妥当な評価結果が得られました。これらの結果は、MOCCS プロファイルおよび Δ MOCCS2score が、細胞の種類ごとの転写因子を解析するツールとして有効であることを示しています（参考図）。

今後の展開

本研究で開発した MOCCS プロファイルおよび Δ MOCCS2score は、多方面における応用の可能性を秘めています。例えば、MOCCS プロファイルをエピゲノムデータ等と組み合わせることで、細胞種特異的な遺伝子発現制御機構の理解につながったり、 Δ MOCCS2score を用いて、がんなどの体細胞変異が転写因子結合に与える影響度を評価するなどの用途が考えられます。

参考図

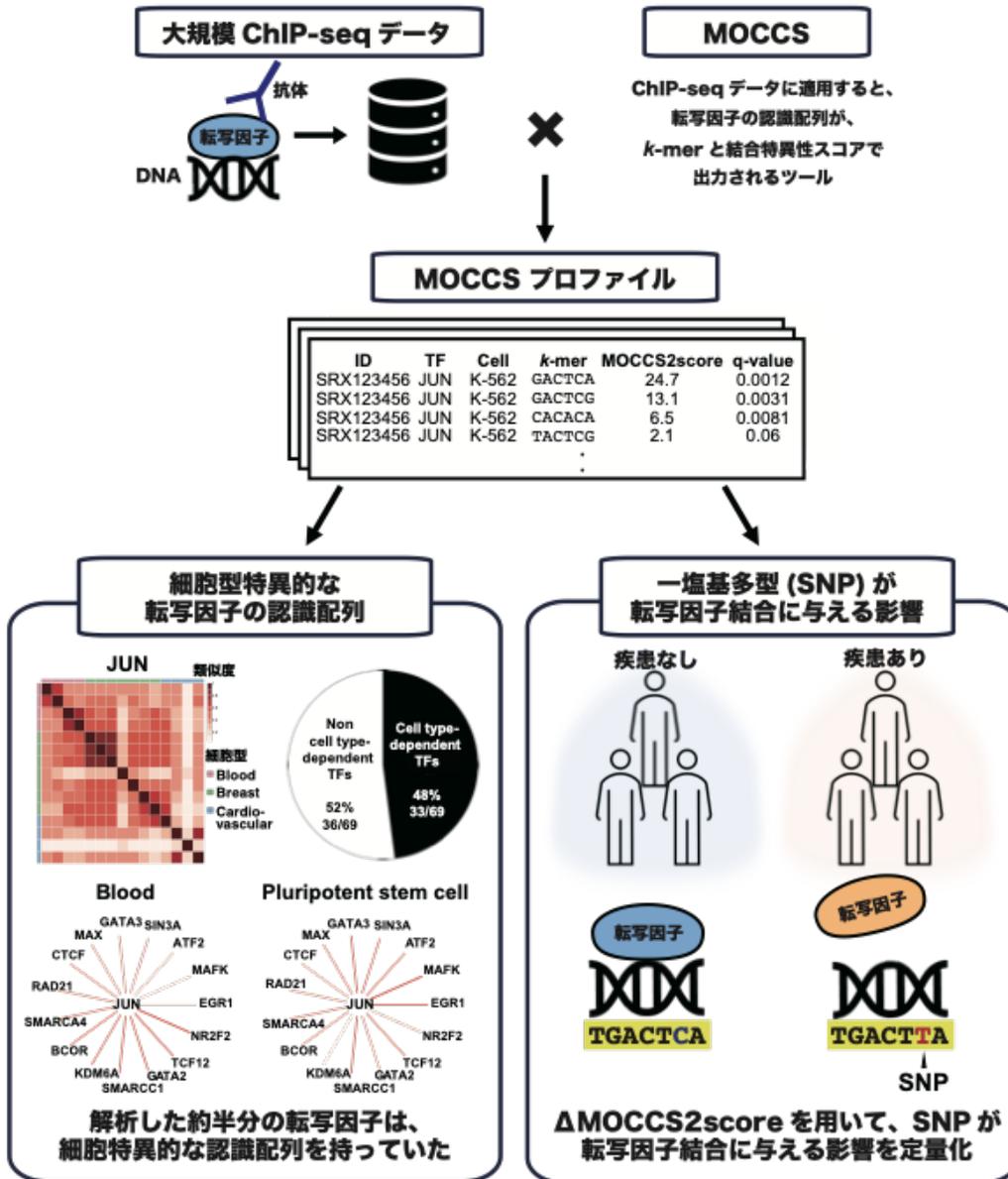


図 本研究の概要

データベースに収集されたヒトの転写因子 ChIP-seq データに MOCCS 法を適用し、「MOCCS プロファイル」を構築した。MOCCS プロファイルは、ヒトの転写因子の結合する配列が *k*-mer と結合特異性スコア (MOCCS2score) で表された、転写因子結合配列の新たな基盤データである。これを解析し、細胞特異的な転写因子の結合配列を明らかにした。さらに、一塩基多型 (SNP) が転写因子結合に与える影響を予測する指標 ΔMOCCS2score を開発し、これが疾患関連 SNP に対して有効であることを示した。

用語解説

注1) ゲノム

遺伝子を始めとする遺伝情報全体。ゲノム情報は、DNA に含まれるアデニン (A)、チミン (T)、グアニン (G)、シトシン (C) の4種類の塩基で構成されている。

注2) 遺伝子発現

遺伝子から mRNA が合成され、さらに mRNA からタンパク質が作られること。生物の身体を構成する細胞が生命活動を行うためには、さまざまな種類のタンパク質を作る必要がある。異なる形態や機能を

持つ細胞は、異なる遺伝子発現を示す。

注3) ChIP-seq (chromatin immunoprecipitation sequencing)

クロマチン免疫沈降法 (ChIP: chromatin immunoprecipitation) と、次世代シーケンサーを組み合わせた実験技術で、転写因子がゲノム上で結合する部位を特定できる。

注4) k -mer

ここでは長さ k 塩基の DNA 塩基配列のこと。例えば 6-mer は長さ 6 塩基の塩基配列を意味する。

注5) MOCCS (Motif Centrality Analysis of ChIP-Seq) 法

さまざまな ChIP-Seq のデータから転写因子結合配列を正確に抽出できるバイオインフォマティクス手法。(Ozaki et. al, 2016)

注6) 一塩基多型 (SNP: single nucleotide polymorphism)

ヒトのゲノム配列は個人間で異なっており、その違いを多型 (polymorphism) といい、多型の中で、1 塩基だけ違いのある多型を一塩基多型 (SNP) という。

注7) クローン病

若年者に発生しやすい炎症性腸疾患 (大腸及び小腸の粘膜に、慢性の炎症や潰瘍を引き起こす原因不明の疾患) の一つ。原因として遺伝的要因等が示唆されているものの、完全に解明されていない。

注8) GWAS-SNP

GWAS (ゲノムワイド関連解析: genome-wide association study) で同定された SNP。GWAS は、世界で広く使われる遺伝統計解析手法の一つで、ヒトゲノム全体の SNP 一箇所ずつについて「疾患を発症する人」と「疾患を発症しない人」とで保有する変異の頻度を調べる。

注9) 全身性エリテマトーデス (SLE: systemic lupus erythematosus)

全身性の炎症性病変を特徴とする自己免疫疾患。原因は完全には解明されていないが、遺伝的要因と環境因子の両方が関わっていることが示唆されている。

研究資金

本研究は、JSPS 科学研究費補助金 (JP19K24361, JP20K19915, JP19H03696, JP19K20394, 22K17992)、AMED ムーンショット型研究開発事業 (A3I03313) の支援を受けて実施されました。

掲載論文

【題名】 Transcription factor-binding k -mer analysis clarifies the cell type dependency of binding specificities and cis-regulatory SNPs in humans.

(転写因子結合 k -mer 解析は、転写因子結合配列の細胞型特異性と、ヒトの一塩基多型の影響を明らかにする)

【著者名】 田原沙絵子 (筑波大学医学群医学類)、土屋貴穂 (筑波大学医学医療系バイオインフォマティクス 助教)、松本拓高 (長崎大学情報データ科学部 准教授)、尾崎遼 (筑波大学医学医療系バイオインフォマティクス 准教授)

【掲載誌】 *BMC Genomics*

【掲載日】 2023 年 10 月 07 日

【DOI】 10.1186/s12864-023-09692-9

問合わせ先

【研究に関すること】

尾崎 遼（おざき はるか）

筑波大学 医学医療系バイオインフォマティクス／人工知能科学センター 准教授

URL: <https://sites.google.com/view/ozakilab-jp>

【取材・報道に関すること】

筑波大学広報局

TEL: 029-853-2040

E-mail: kohositu@un.tsukuba.ac.jp