

審査の結果の要旨

氏名 ナガイ ルイス アウグスト エイジ

高等真核生物は、様々に分化した細胞種から形成されているが、見かけ上大きく異なるそれらの細胞も基本的には同じゲノム情報に基づいて形作られたものであり、ゲノムにコードされた遺伝子がうまく使い分けられた結果であると考えられている。しかし、その詳細なメカニズムについては、いわゆる多数の転写因子群や、エピジェネティックな情報の関与が言われているものの、まだ十分に理解できているとは言えない。一方、近年のいわゆる次世代シーケンサー技術の進歩により、多数の細胞種におけるトランスクリプトームデータ

(RNA-seq 解析)に加えて、いくつかの細胞種における染色体 DNA の大まかな三次元構造データ (Hi-C 解析など) も得られるようになり、細胞種によるクロマチン構造の変化が、それらの細胞種特異的遺伝子発現の調節に大きく関わっているのではないかとされるようになった。特に、クロマチン DNA が TAD (Topologically Associating Domain) と呼ばれるループ構造をとっているといわれ、そのループの付け根に比較的多く見られる CTCF というタンパク質が、エンハンサーの作用を制限するサイレンサーと呼ばれる領域と深く関わっていることから、TAD 構造の変化が TAD ループ内に存在するエンハンサーのターゲット遺伝子調節に関わっていることが示唆されている。しかし、TAD 構造検出のもととなる Hi-C 解析はいまだ発展途上の技術であり、誰にでもできるものではない上に、検出された TAD の情報にも不確定性があり、複数の情報源から得られた TAD の比較を行う作業は自明ではない。本研究は、公共の複数データから異なる細胞種における TAD や A/B コンパートメントとよばれるクロマチン構造を比較し、その違いが遺伝子発現の違いとどの程度関わっているのかを情報解析したものである。

本論文は 6 章からなるが、基本的には一つの研究について述べられており、その主要な結果は第 3 章に記載されている。すなわち、申請者は ES 細胞、プロ B 細胞、びまん性大細胞型 B 細胞性リンパ腫の 3 種類のマウス細胞について、それぞれ異なる文献から Hi-C および RNA-seq 生データをダウンロードし

て、共通のプロトコルに基づきデータ処理を行った。その結果、主成分分析によって定義される A/B コンパートメントというレベルでは、3 種の細胞はおおよそ 8 割から 9 割程度が共通していることが示された。また、プロ B 細胞とリンパ腫において、領域を 100kb の長さで区切って、領域ごとの比較をしたところ、両者の間で、コンパートメントの種類が違っている領域が、それぞれ約 1 割程度同定できた。予想通り、A コンパートメントに変化した領域の遺伝子は発現量が増加する傾向を示し、B コンパートメントに変化した領域の遺伝子は発現量が減少傾向を示した（ただし、ランダム遺伝子と比べた有意性は微妙なレベルであった）。GO タームの濃縮度に基づく機能解析を行ったところ、プロ B 細胞には免疫応答に関連する遺伝子が特徴的で、リンパ腫細胞では。たとえば T 細胞の細胞死の負制御と関わる遺伝子が顕著であった。さらに、3 種類の細胞間で TAD 構造の位置を比較したところ、おおよそ 8 割程度は 2 者の間の比較で共通しているという結果が得られた。TAD 境界領域には、がん化と関連する遺伝子が多くみられ、発現量も TAD 内部と比べて多い傾向がみられた。さらに、やや別の角度の解析として、染色体間の相互作用を Hi-C データから細胞種ごとに検出したところ、リンパ腫細胞の一部には特徴的な相互作用が見られ、転座とのかかわりについて、今後より詳細な解析が必要であることを報告している。

以上の研究は、複数の研究室から得られた外部データ（しかも発展途上の実験技術によるもの）を比較するという本質的な困難さに起因する限界が存在し、研究そのものの粗さは否定できない。特にがん細胞は多様であり、リンパ腫と一口にいてもクロマチン構造変化も多様であり、トランスクリプトームデータとの対応にも若干の疑問が残る。その反面、本研究には世界的に興味の持たれている重要テーマに先駆的に取り組み、大枠としては妥当な結果を得たという意義が認められる。また本論文の内容は、パク・ソンジュン、中井謙太との共同研究であるが、論文提出者が主体となって分析及び検証を行なったもので、論文提出者の寄与が十分であると判断する。

したがって、博士（医科学）の学位を授与できると認める。

以上 1,850 字