

# 審査の結果の要旨

氏名 バシヤ グティエレス ユセフ

本論文は、4章で構成されているが、最初と最後の章は、全体の前書きとまとめであり、実質的には二つの章から成る。両者ともに、ゲノム塩基配列中に存在する転写制御領域のコンピュータ解析に関する内容で、そのうち第2章では、テキストマイニングの一手法として近年注目されているトピックモデルの方法論を、バイオインフォマティクスでは古典的な共通モチーフ抽出問題に応用した研究について述べられている。また、第3章では、共発現遺伝子群の制御領域にみられる共通構造を検出する問題を取りあげ、具体的にはヒト小脳特異的発現遺伝子群の上流プロモーター領域と既知エンハンサー領域に頻出するモチーフの検出とその共起と進化的保存について述べられている。

論文提出者らは以前、遺伝的アルゴリズムを使って、共通モチーフ抽出を行うアルゴリズムを研究・発表してきた (Basha et al., *Lecture Notes in Comput. Sci.* **9043**, 438-449 (2015))。そこで効果を発揮したのが、遺伝的アルゴリズム中で用いた適合度関数に統計的評価尺度 (Mann-Whitney U 検定等) を利用することであった。しかしそれでも得られたモチーフ候補に適切な信頼度を付与することが困難であったため、偽陽性の絞り込みが難しかった。今回の研究で、そこにトピックモデルにおける perplexity を計算し、その値が適切な範囲のものだけ出力することとした。得られた方法の性能は、2005年に発表された様々なモチーフ抽出アルゴリズムの性能を検証した Tompa らの論文 (*Nature Biotech.*) と同じ枠組みで評価した。モチーフ抽出アルゴリズムの性能評価はいろいろな角度からなされるべきものなので、トピックモデルを応用した今回の方法が必ずしも以前に開発した遺伝的アルゴリズムによる手法を明瞭に凌駕しているとは言えない結果になったが、従来法と比べると、論文提出者らの方法が全体的にはかなり良好な結果を示し、少なくともトピックモデルの考え方や手法がモチーフ抽出問題に応用可能で、今後の新たな発展の可能性を示せたものと判断できる。本章に記載された研究内容はすでに査読付きジャーナルにて発表されている (Gutierrez & Nakai, *BMC Bioinformatics*, 2016)。

続いて、第3章では、小脳特異的発現遺伝子群の制御領域にみられるモチーフの検出について述べられている。論文提出者が所属する研究室では、これまで様々な生物種の様々な組織に対する特異的発現制御プロモーター領域の構造を研究してきたが、データの不足などの問題から、エンハンサー領域の情報を扱っていなかった。今回は、少数ながら既知のエンハンサー領域に関する情報を導入した点が一番の特徴であると言える。データ取得という便宜的な条件から、今回ヒトの小脳に関する遺伝子を取りあげたが、小脳の機能については従来考えられていたよりも多くの機能に関わっているという報告もあり、その研究自体興味深いものである。既存の複数のデータベース情報を組み

合わせ、小脳（と後脳）、コントロール（それ以外の脳）でそれぞれ発現する遺伝子群 200 余ずつとエンハンサーの情報を得て、エンハンサー（プロモーター）活性の中心部に存在してプロモーター（エンハンサー）領域に特異的に頻出するモチーフの抽出を行い、それぞれ 7 つの小脳特異的モチーフを得た。うち 3 つは既知モチーフと一致した。そのうちで既知 3 個を含む 6 モチーフの共起がみられたが、エンハンサー特異的に共起している対はなかった。また、これら 6 モチーフはある程度哺乳類で保存されていることを確認した。興味深いことに 3 つの既知モチーフはいずれも老化との関連が指摘されていた。今後のより詳細な研究が待たれる。今回用いた方法は基本的にはほかの組織特異的発現情報にも適用可能である。審査委員会では、本章で述べられた研究はまだ十分に尽くされていないのではないかという声もあったが、一定の水準には達しているという合意がなされた。

全般に審査委員会ではその研究内容について、好意的に評価されたが、最初に提出された論文では、形式上の不備があったり、内容について説明が不十分であったりする点が指摘されたので、指導者の責任のもとに改訂を行って、最終版を提出することとなった。なお、本論文の内容（第 2 章および第 3 章）は、中井謙太の指導のもとに行われたものであるが、論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。

従って、博士（科学）の学位を授与できるものと認める。

以上 1908 字