

論文審査の結果の要旨

氏名 若栗 浩幸

本論文は、大きく4章からなる。第1章では完全長 cDNA データの大量情報を整備するための枠組みの開発について述べられている。著者は完全長 cDNA データを効率よくゲノムにマップする方法に加え、転写開始点、選択的プロモーター、多生物種間におけるプロモーター領域を抽出するための方法など統一的な枠組みの開発を行った。いずれの方法も簡潔でありながら、十分に正しい結果を反映する方法が導かれ、ヒトのみならず様々な生物種の cDNA 情報整備への基盤となった。

第2章では公共データベース遺伝子情報(モデル遺伝子)の、完全長 cDNA を使った修正について述べられている。第1章の解析手法をアピコンプレクサ原虫 6 生物種に適用し、モデル遺伝子データの正しさの評価を行った。結果として、いずれの生物種もモデル遺伝子は 5'端や 3'端に必要な以上に長く定義されている傾向があることを示し cDNA 情報を用いて遺伝子情報をさらに修正することが必要不可欠であることを示した。

第3章では完全長 cDNA を使った TSS と 5'-UTR 情報の解析及び分子進化解析について述べられている。TSS と 5'UTR の情報解析では *Plasmodium* 属では、TSS 位置のばらつきが他の生物種に比べかなり大きいことや、Ribosome に関する遺伝子が Cp を除く多くの生物種で 5'UTR が有意に短い傾向にあることを明らかにした。6 生物種共通のオルソログ遺伝子を使った分子進化系統樹解析ではネズミマラリア(Py、Pb)の共通先祖と熱帯熱(Pf)、三日熱マラリア(Pv)の3つが非常に近い時期に分岐した可能性を示した。

第4章では次世代シーケンサーを使った3つの解析(cDNA-Seq, TSS-Seq, RNA-Seq)について述べられている。cDNA-Seq (cDNA ショットガンシーケンシング)では、効率的なアセンブル手法を開発することで高い頻度での完全長配列の決定が行えることを可能とした。TSS-Seq (オリゴキャッピング法と次世代シーケンサーを組み合わせる TSS 位置を一度に解析する方法)では、Pf のデータを用いて転写開始点上流に Poly T 様の興味深い配列を見出した。RNA-Seq 解析は準備的ではあるが、ホストと寄生虫が混ざった状態の RNA を取得したデータが収集されつつある。これらのデータを用いてホスト寄生虫間の相互的なトランスクリプトーム像を明らかにすることを目指し、大量データを意義付けするためのデータ解析手法の探索を開始している。

このように本論文では、大量の完全長 cDNA という新種のデータに対し、統一的なデータ整備方法を開発し、解析をスムーズに前進させるための基盤を作った。またこの基盤を元にモデル遺伝子をより正しく修正できることや、新しい知見が得られることを実証した。本研究により組織的かつ横断的な cDNA のデータ解析が可能となり、トランスクリプトーム全体像の解明に向けて大きく貢献するものであると考えられるために、博士(生命科学)を授与できると認める。