

論文内容の要旨

申請者氏名 CHEN XING

本研究は、脊椎動物の発生過程における二つの重要な対象、神経管と卵母細胞を対象に、バイオインフォマティクスを用いて遺伝子制御ネットワーク (gene regulatory network; GRN) の推定をおこなった。神経管は中枢神経系の発達の前駆体であり、Sonic hedgehog (Shh) によって誘導される GRN の下で、腹側において Shh の濃度依存的に区別されたドメインに分かれる。それぞれのドメインにおいて前駆細胞は多様な神経細胞タイプに分化する。しかし腹側の神経管では Shh の濃度に依存したドメインに細分化され、そこから多様なタイプの神経細胞が分化する。Shh によって誘導される GRN が生み出す分化パターンの複雑さに比べ、その GRN を構成する既知遺伝子の種類は現在のところ 10 個しかない。これは、GRN に関わる未知遺伝子およびパスウェイが存在する可能性を意味する。一方で、健康な卵母細胞の放出は個体の発達全体の基盤である。卵母細胞の成熟は受精前の様々な段階で異なるシグナル伝達経路によって制御されるが、その遺伝子と経路に関してはまだほとんど知られていない。候補遺伝子の数が多いため、バイオインフォマティクスの技術を用いて、候補遺伝子とパスウェイを推定できれば、実験による検証が効率的に行える。

Shh で誘導される神経管 GRN における未解明の遺伝子を推定するため、4 種の Shh 濃度で刺激後の 4 時点で計測した前駆細胞のトランスクリプトームデータを解析した。発現時系列のパターンで遺伝子をクラスタリングし、各クラス (遺伝子セット) ベースでネットワークの数理モデルを構築および発現時系列にフィッティングした。その際、パスウェイの多くをゼロにする LASSO 回帰と、既知のパスウェイをゼロにしない制約を導入し、既知の遺伝子セットから追加するに相応しいものをランキングした。卵母細胞の解析では、機械学習で用いられる特徴選択アルゴリズム (Boruta) を用いて発達段階を区別するのに役立つ遺伝子を抽出した。

神経管 GRN の解析では、交差検証にもとづき 9 種の既知遺伝子セットを学習データにして、残りの 1 つの既知遺伝子セットを上位にランキングすることに成功した。その後、10 種の既知遺伝子セットから、特定の遺伝子が GRN に関与する可能性を見出した。卵母細胞の解析においては、241 の特徴遺伝子を同定した。これらの遺伝子は卵母細胞の発達過程に関連し、一般的な発現解析ツールでは検出されなかった遺伝子を推定した。これらの遺伝子を用いたときの発達ステージ予測は 0.92 の精度を得た。これらの結果は、脊椎動物の発生において重要な二つの過程における GRN の解明に貢献し、神経管発達と卵母細胞の生成に関わる未知の遺伝子と経路を特定に寄与するものである。

やむを得ない事由 [図書出版, 学術雑誌等への掲載, 特許・実用新案出願, 個人情報等の保護, その他 ()] により本要旨を非公表とする。

【※該当する事由に○印をすること】

論文審査結果の要旨

申請者氏名 CHEN XING

発達生物学の遺伝子ネットワーク (GRN) を解明するためには、鍵となる遺伝子、および遺伝子間のパスウェイの同定を行う。しかし、遺伝子の種類が多岐にわたるため、それらの同定は困難である。情報が限られている状況下での候補遺伝子の一つ一つの実験的検証は困難である。したがって、既知の知見から未知の遺伝子とパスウェイを推定する必要があるが、ネットワークの複雑さのために研究者の想像力でそれらを推定することは不可能である。本研究の意義は、定量データと数理的な手法を用いてこれらの課題に取り組むことにある。GRNを推定する数理的な先行研究は数多くあるが、ほとんどが交差検証に基づく妥当性評価を行っていない。交差検証は近年のデータサイエンスで極めて重要な手続きであり、既知のデータを学習用とテスト用に分け、学習データからテストデータの予測力を持って妥当性を評価する。先行研究が交差検証を行わなかったのは、全ての遺伝子を数理モデルに組み込むアプローチが交差検証に適さないためである。一方で本研究は、既知の遺伝子とパスウェイに基づくネットワークから始めて、追加すべき遺伝子とパスウェイのランキングを生成する。この方法論であれば、既知の10個の関連遺伝子のうち9個で学習したあと、テストの10個目の遺伝子が高いランキングで推定されるかどうかを検証できる。本研究は神経管に関連する既知遺伝子10個について交差検証を行い、テストとして抜かれた遺伝子が高いランキングで推定されることを確認した。その際のモデルパラメータに基づいて11個目の遺伝子とパスウェイを推定した。本研究はデータサイエンスにおいて重要な交差検証を用いて、その方法論の妥当性を示しており、大きな新規性と結果の重要性を持っている。本研究のもう一つの新規性は、Shh濃度4条件で4時点のトランスクリプトームデータを、ネットワークの不変性に基づいてクラスタリングしたことである。神経管細胞内の遺伝子は、Shhのいかなる濃度に対しても、分子レベルの反応親和性、すなわちパスウェイを変えることはない。各遺伝子の発現量だけが変化し、活性化および不活性化の制御関係に基づいてシグナル伝達のフローが変わるだけで、分子間の相互作用関係自体は変化しない。本研究はこの事実に基づいて、Shh濃度ごとに各遺伝子が加わるクラスターを区別し、遺伝子群を300以上の集合に分割した。このような分割を行った研究は過去に存在していない。本研究はデータ処理の新規性を示し、単一濃度条件によるGRN推定の問題点も明らかにしている。以上の点から、本論文はGRN推定において学術的、応用的に重要な貢献をしている。よって審査委員一同は、本論文が博士(工学)の学位論文として価値あるものと認めた。

やむを得ない事由[図書出版, 学術雑誌等への掲載, 特許・実用新案出願, 個人情報等の保護, その他 ()]により本要旨を非公表とする。

【※該当する事由に○印をすること】