

学 位 論 文 の 要 旨

論文題目

オープンデータを活用した低酸素応答新規パスウェイの探索手法の開発

(Development of exploration methods for novel hypoxia response pathways using open data)

広島大学大学院統合生命科学研究科

生命医科学プログラム

学生番号 D205302

氏 名 小野 擁子

生命科学研究には、よく知られている分野や遺伝子ほどよく研究される出版バイアスが存在する。例えば、疾患との関連性の有無などの情報が既知の遺伝子ほど研究対象になりやすい。一方で、昨今のマイクロアレイ技術やハイスループットシーケンシング技術の発展により、研究者の興味対象の如何に関わらず一度の実験で生体内の数万種類の遺伝子発現プロファイルを網羅的に取得できるようになった。これらの網羅的な遺伝子発現情報を含む論文を投稿する際には、その再現性の担保のため、オープンデータベースに登録することが論文誌掲載の条件とされているのが一般的である。National Center for Biotechnology Information (NCBI)の gene expression omnibus (GEO) や European Bioinformatics Institute (EBI)の ArrayExpress に代表されるオープンデータベースには、2022年現在ヒト遺伝子発現データはおよそ6万6千を超えるデータシリーズが登録されており、これらの遺伝子発現プロファイルデータを元にしたデータドリブンな解析により、遺伝子の注目度の影響を受けずに知見を見出すことが期待できる。

低酸素刺激応答は好気性生物において生命維持に重要な機能を担う。低酸素刺激下では転写因子である Hypoxia inducible factor -1 (HIF-1) が下流の遺伝子発現を制御し、低酸素に対応するための厳密な分子機構が備わっていることが明らかになっている。低酸素に対する生体反応の研究は1990年台の HIF-1 の発見により劇的に進み、そのメカニズムを明らかにしたウィリアム・ケリン教授、ピーター・ラトクリフ教授、グレッグ・セメンザ教授は2019年にノーベル生理学・医学賞を受賞している。ノーベル賞を受賞するほどの研究分野であることから公共遺伝子発現データベースへの多くの低酸素刺激関連データの登録が期待できることと、低酸素応答研究においても出版バイアスの存在は否定できないことから、低酸素を研究対象とし低酸素応答新規パスウェイの探索手法の確立を本研究の目的とした。

本研究は新規低酸素応答遺伝子の探索と、non-coding RNA (ncRNA)を含めた転写産物の低酸素応答の評価方法構築の大きく二つの研究にて構成される。

新規低酸素応答遺伝子の探索では、まずはオープンデータベースである GEO および低酸素に関わる RNA sequencing (RNA-Seq) データのメタデータをマニュアルキュレーションにより精査したのちに解析対象のデータセットを選定し、Sequence Read Archive (SRA)から配列データを取得した。低酸素、通常酸素の条件のおよそ500ペアとなるサンプルの遺伝子発現プロファイルを元に、遺伝子ごとに Hypoxia-Normoxia (HN) -score を算出した。高い HN-score を示す上位100遺伝子を元にエンリッチメント解析を行い、想定通りに低酸素応答関連遺伝子群が濃縮されていることを確認した。

加えて、SRA で公開されている既報の Chromatin Immunoprecipitation sequencing (ChIP-Seq) データを元にした解析ツールである ChIP-Atlas を用いて、高い HN-score を示す上位 100 遺伝子の発現制御には HIF-1 を構成する遺伝子である HIF1A が関与していることを示した。

次に NCBI が提供している gene2pubmed と呼ばれる、gene ID とその遺伝子が研究報告されている PubMed ID が記載されたデータセットを用いて新規低酸素応答遺伝子を探索した。その結果、G Protein-Coupled Receptor 146 をはじめとするいくつかの遺伝子が、データドリブンな解析では低酸素応答遺伝子と判定されるにも関わらず今まで低酸素応答遺伝子として注目されていなかったことを明らかにした。

上記の研究では、HN-score の下位 100 遺伝子を元にしたエンリッチメント解析にて ncRNA metabolic process に関わる遺伝子の発現抑制が示された。しかしながら低酸素条件下での ncRNA metabolic process の詳細メカニズムについては明らかになっていない。新規低酸素応答遺伝子の探索の研究と同様にデータドリブンに ncRNA を含めた転写産物の低酸素応答の評価をすることが必要と考えた。

エンリッチメント解析に用いられた Gene Ontology を元に精査をしたところ、これらの発現が抑制された遺伝子群は ncRNA metabolic process の中でも ribosomal RNA (rRNA) processing に関わることが明らかになった。rRNA はリボソームを構成してタンパク質を合成する機能を持つ。リボソームと細胞質 long non-coding RNA (lncRNA) との関係性や、ncRNA が rRNA のサイレンシングに寄与することを示した報告もあることから、ncRNA を含めたリファレンスを活用して転写産物の低酸素応答性を網羅的に評価した。FANTOM CAGE Associated Transcriptome (FANTOM-CAT) は FANTOM5 Cap Analysis of Gene Expression (CAGE) のデータを用いた信頼性の高い 5' 末端をもつヒト lncRNA も含まれている転写産物のカタログである。このカタログをリファレンスとして活用することにより、ヒトのコーディング遺伝子以外も含めた遺伝子発現情報を活用することが可能と考えた。

転写産物の網羅的な低酸素応答性評価の結果、低酸素応答によりミトコンドリア DNA 由来の転写産物の発現抑制が顕著であること、低酸素応答遺伝子群のアンチセンスに着目した解析では、大半の転写産物は低酸素応答遺伝子群と同様の発現制御パターンを示す一方で、センス-アンチセンスで異なる発現制御がなされている遺伝子群があることを明らかにした。

以上、本研究ではデータドリブンに新規低酸素応答遺伝子を見出し、ncRNA を含めた転写産物の低酸素応答の評価方法を提示した。これらの研究は、低酸素応答研究のみならず新規パスウェイのデータドリブンな探索手法の提示の点で貢献したと考える。