

審査の結果の要旨

氏名早川 仁

ゲノムやトランスクリプトームを始めとする生体情報と疾患との関連は重要視されている。これらの情報はオミクスデータと呼ばれている。この中でも、疾患を特徴づけたり、予後に関わったりする遺伝子変異はすでに診断、治療の場面で利用されつつある。一方で、大量のオミクスデータは意義が未だに不明なものも多い。研究目的で測定されたオミクスデータの数多くがデータベースで公開されており、それらを二次利用することでも研究が行われている。また、これまでの研究で得られた知見をデータベースにまとめてコンピュータで利用可能な形にしたデータベースの一つに、**Kyoto Encyclopedia Genes and Genomes (KEGG)**がある。KEGGに収められた情報に、生体分子の相互作用に関するパスウェイについてまとめられたものがある。このKEGGパスウェイは、複数の遺伝子の共発現などを利用して遺伝子の機能に関する解析を行う事にも研究利用されている。**Gene set enrichment analysis**は遺伝子機能の解析のために利用される遺伝子セットの解析手法の一つであり、生物学的に意味のある遺伝子セットを利用して複数の遺伝子の解析をおこなうことで遺伝子グループの有意な変動に対して高い検出力を示すとされている。例えば、KEGGパスウェイに基づいた遺伝子セットを用いることで、サンプル集団におけるマイクロアレイや次世代シーケンサで得られた遺伝子発現量を用いて、特定のフェノタイプに特異的に発現しているパスウェイを推定することができる。しかし、GSEAには批判もある。GSEAは変数とフェノタイプの一方向的な関係性しか検出できず、逆相関したり、複雑な関連がある場合は検出できない。本研究ではGSEAによる遺伝子発現量の変動の検出に関する制約を改善するために、深層学習を用いて、パスウェイの分析を試みている。

本研究で利用している深層学習の一手法であるグラフ畳み込みニューラルネットワーク(GCN)は、グラフネットワークを利用したグラフ畳み込みを行うことで、互いに関連し合う変数から特徴量を取り出すことが可能である。KEGGパスウェイに示される遺伝子間の関係性をグラフネットワークに表現して、グラフ畳み込みを利用して取り出したパスウェイの特徴量を利用してフェノタイプの分類を行うGCNを構築した。このGCNは、内部にKEGGパスウェイにそれぞれ対応するグラフネットワークを複数持ち、遺伝子の発現量を入力としてフェノタイプの分類予測を目的変数に取る。遺伝子発現量をもとにフェノタイプの予測を行う過程で、パスウェイに相当するグラフネットワークから取り出した特徴量が予測にどれほど寄与したかをShapley additive explanation(SHAP)によって明らかにすることで、フェノタイプに特徴的なパスウェイを明らかにしている。GCN

からフェノタイプに特徴的なパスウェイを選択できることを検証するために、悪性リンパ腫のびまん性大細胞型B細胞リンパ腫（DLBCL）で、遺伝子発現パターンをもとに分類されるサブタイプである **germinal-center B-cell-like (GCB)**タイプと **activated B-cell-like(ABC)**タイプを分類した。DLBCL の遺伝子発現量とフェノタイプをもとに学習した GCN モデルは、2つのフェノタイプを高い精度で分類することが可能であった。さらに、この GCN モデルの予測を SHAP によって解析することで、それぞれのフェノタイプの予測に強く寄与するパスウェイを明らかにしている。これらのパスウェイは GSEA で検出された、フェノタイプ特異的に発現する上位のパスウェイとよく一致が見られた。また、GSEA では高い重要度で検出されなかったものの、これまでに行われた研究から DLBCL のサブタイプと強く関連することが知られている **B cell receptor signaling pathway** などのパスウェイも高い寄与度を示すことが検出された。この結果から、まず、GCN モデルは、GSEA で検出される遺伝子発現パターンとフェノタイプの重要な関連を検出したことが示唆された。さらに、GSEA では検出が困難であった関連も検出しうることが示唆された。GCN では、高発現とフェノタイプの関連にとどまらず、遺伝子発現量の高低あるいは複雑な発現パターンとフェノタイプの関連をフェノタイプの予測に利用しうるため、GSEA で検出されない重要なパスウェイも示すことができたと考えられる。DLBCL をサンプルデータとした検証により、本研究で用いた DLBCL のサブタイプを分類する GCN に SHAP を適用することにより、フェノタイプと関連の強いパスウェイを選択することが可能であることが示唆された。

本研究の新規性としては、GCN と SHAP を組み合わせてパスウェイの分析を行うことができるようにするネットワークの構築と、悪性リンパ腫のサンプルにおける実装によって検証をしたことにある。本研究のネットワークの検証に関しては、これまでの研究から十分に検証が可能と考えられる悪性リンパ腫に対する実証を行い、目的を概ね示すことができていると考える。本研究の手法は、遺伝子パスウェイの分析をする際の一つの選択肢となる。例えば、他の種類の悪性腫瘍サンプルに対して、GSEA の結果に加えて解析を行い、新たな知見が得られる可能性がある。また、本研究は公開データベースを用いて十分に検証可能なものであった。このため公開データベースを二次利用した研究の有用性を支持するものであり、疾患データベースの拡充を促し様々な医療情報の二次的利用の可能性を期待させ、データベースの構築を促すものである。よって本論文は博士（医学）の学位請求論文として合格と認められる。