京都大学	博士（ 医 学 ）	氏名	Zheng Cheng
論文題目	Delving into gene-set multiplex networks facilitated by a k-nearest neighbor-based measure of similarity （k-最近傍法に基づく類似性尺度による、遺伝子セットの多重ネットワーク解析）		

（論文内容の要旨）

Motivation:

Gene sets are collections of genes annotated by their functional implications, which are often used for interpreting target genes via gene-set enrichment analysis (GSEA). However, the interpretation usually does not involve contextual and relational attributes of gene sets, and the role of enrichment may be exaggerated, potentially leading to a biased emphasis on less meaningful terms. To find alternative ways of prioritizing gene sets, it is worth knowing more about the diversity and interactions of gene sets in their communities.

Method:

By trial and error, simulation and comparison studies were designed to identify an optimal measure applied to single-cell RNA-seq datasets to estimate inter-gene-set similarities. Then, gene-set multiplex networks can be built based on the ideal similarity measure where each layer represents a biological context, e.g., a cell population, and the structural attributes of gene sets in the multiplex networks can be computed at different scales. Both high-level properties, e.g., relations between contexts, and low-level features, e.g., structural coefficients of gene sets, were further studied; attention was paid to the distinction between the semantic (Jaccard) network and the multiplex network. Moreover, the structural coefficients were used to guide the reorganization and prioritization of enriched GO-BP terms in a relevant GSEA study.

Results:

A k-nearest neighbor-based measure outperformed other candidates by its fast running time and flexibility of being both linear and nonlinear measures. The multiplex network of gene sets built on this measure captured the hierarchical relations among biological contexts and successfully preserved the dynamic patterns of proximity between families of GO-CC gene sets, which contrasted the static view by the traditional Jaccard network. The heterogeneous structural coefficients helped reorder the enriched terms in a relevant GSEA where the differentially expressed genes of the CD4+ T cell population compared to other immune cells were the target. Gene sets with a fine-tuned level of multiplex centrality and normalized enrichment scores dominantly prioritized terms related to leukocyte differentiation and T cell activation.

Conclusion:

This gene-set multiplex initiative explored the altering patterns of gene sets' connections across different biological conditions. The findings of gene-set contextual relations are valuable for strengthening the current framework of prioritizing gene sets in GSEA.

（論文審査の結果の要旨）

Gene Set Enrichment Analysis(GSEA)は、生物学的概念を記述した遺伝子オントロジーなどに分類された既知の遺伝子セットの情報を用いて、研究対象の解析で得られた遺伝子群と既知の遺伝子セットの関連を網羅的に検定する手法であるが、遺伝子セットに属する遺伝子に重複があるなどの理由で多数の遺伝子セットが検出され、結果の解釈が困難であった。

申請者は、網羅的転写物解析の情報を用いて、GSEA で検出された遺伝子セット間の類似性を計算し、遺伝子セットをノードとするネットワークを構築しネットワーク構造上での各遺伝子セットの集積度を推定する手法を開発した。この手法をナイーヴ T 細胞、メモリーT 細胞、メモリーB 細胞、NK 細胞、マクロファージの 5 種の免疫細胞の 1 細胞 RNA-seq データに適用した結果、メモリーT 細胞と関連するとされた上位の遺伝子オントロジーの間に冗長性が認められ、これらをグループ化し情報を圧縮することで、GSEA 結果の可読性を高めることが期待できる。また、白血球分化や T 細胞活性化がネットワーク構造の中で重要な遺伝子セットとして上位で検出された。本手法は、冗長かつ膨大な GSEA 結果から今後調査すべき遺伝子セットをより明確にすることで、網羅的転写物解析の結果の生物学的意味付けを容易にさせるという点で、ゲノム情報科学およびゲノム医学の研究に寄与するところが多い。

したがって、本論文は博士（医学）の学位論文として価値あるものと認める。なお、本学位授与申請者は、令和 6 年 2 月 16 日実施の論文内容とそれに関連した試問を受け、合格と認められたものである。