

浜本 隆二

国立がん研究センターがん分子修飾制御学分野
分野長

人工知能技術を活用した革新的ながん創薬システムの開発

§ 1. 研究成果の概要

本研究課題は、(1)病院のリアルワールド・データを元にした臨床的表現型の確立、(2)薬剤投与に伴う *in vitro* マルチオミックス・プロファイルの解析、(3)病院からのデータ収集のスキームと統合データベースの構築、(4)新規の創薬標的タンパク質の発見を目的としたマルチオミックス・データ解析法の確立を 4 つの技術的課題として取り組んでいる。2019 年度の研究成果の詳細を下記に示す。

(1) 病院のリアルワールド・データを元にした臨床的表現型の確立

2019 年度においては、世界最大規模の肺がん統合データベースを構築した。データの詳細は下記に示す。

診療情報: 1569 例

ゲノム(全エクソーム): 1569 例

トランスクリプトーム(RNA-seq): 1494 例

エピゲノム(DNA メチル化アレイ): 430 例

ゲノム(全ゲノム): 272 例

エピゲノム(ChIP-seq): 100 例

今後は ChIP-seq 解析を中心にオミックス解析の症例数を増やしていき、世界最大規模の肺がんデータベースの拡充に取り組んでいく。

(2) 薬剤投与に伴う *in vitro* マルチオミックス・プロファイルの解析

(3) 病院からのデータ収集のスキームと統合データベースの構築

本研究課題を推進する上において、効率的に病院からのデータを研究所に移行させ統合デー

データベースとして拡充していくシステムを構築することは、重要事項の一つとして取り組んできた。2019 年度においては、自然言語処理による診療データ構造化に取り組み、「肺がん取扱い規約」に準拠した肺がん画像所見の知識モデルへのマッピングを行った。現代のがん診療では、診断、治療方針の決定、治療後評価などのあらゆる過程で放射線画像による評価が行われている。自然言語処理より抽出した臨床情報と画像特徴の組み合わせにより、個々の疾患の治療経過のより精緻な層別化が得られる可能性がある。今後は今回構築した知識モデルをアノテーション・プラットフォームと共有することで、放射線画像を起点にした診療情報の構造化が自動的に行われるシステムの構築を目指す。

(4) 新規の創薬標的タンパク質の発見を目的としたマルチオミックス・データ解析法の確立

様々なオミックスデータを統合してマルチモーダルに解析することを目的に、AI 技術を活用した2種類の解析プラットフォームの構築に取り組んだ。1 つ目は、DNA メチル化データ、トランスクリプトームデータなどのマルチオミックス・データを、オートエンコーダを用いて次元圧縮した後に、生存に関連する特徴量を Cox-PH モデルを用いて抽出するプラットフォームである。このプラットフォームを用いて、予後良好群と予後不良群の分類に成功した。2 つ目は、非負値行列因子分解(NMF)に基づく解析手法である、Pathway-Level Information Extractor (PLIER) に、転移学習を組み合わせた手法である MultiPLIER を基に構築したプラットフォームである。本プラットフォームを用いて、ターゲットドメインデータ(TD)として、肺腺癌 DNA メチル化データの解析を行った結果、無再発生存期間に寄与する、潜在特徴ベクトルの取得に成功した。

§ 2. 研究実施体制

(1) 浜本グループ

- ① 研究代表者: 浜本 隆二 (国立がん研究センターがん分子修飾制御学分野 分野長)
- ② 研究項目
 - ・研究計画全体の統括
 - ・肺がん統合データベース及び主要解析パイプラインの構築

(2) 瀬々グループ

- ① 主たる共同研究者: 瀬々 潤 (産業技術総合研究所人工知能研究センター 招聘研究員)
- ② 研究項目
 - ・肺がん統合データベースの構築
 - ・機械学習技術全般を用いたデータ解析

(3) 岡野原グループ

- ① 主たる共同研究者: 岡野原 大輔 ((株)Preferred Networks 取締役副社長)
- ② 研究項目
 - ・肺がん統合データベースの構築
 - ・深層学習技術を用いたデータ解析