

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。2018 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

肺がん統合データベース構築に向けて、2018 年度は肺がんに関する下記のデータを収集した。

ゲノム(エクソーム):600 例
ゲノム(核・SNP アレイ):200 例
エピゲノム(DNA メチル化アレイ):191 例
トランスクリプトーム(RNA-seq): 600 例
トランスクリプトーム(マイクロアレイ): 79 例
プロテオーム:165 例
メタボローム:51 例
臨床情報:600 例

今後ヒストン修飾解析のデータなども拡充し、世界最大規模の AI 解析を志向した肺がんデータベースの構築を行う。

(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

本研究課題の最重要事項の一つとして、ゲノムのみならずエピジェネティクス情報も合わせた、機械学習・深層学習技術を用いた統合的な解析を行うことが挙げられる。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を行ううえにおいて、精度の高いエピジェネティクス解析を行う必要があるが、これまで技術的な困難さが問題となっていた。そこで我々は汎用ヒト型ロボット

LabDroid を活用した次世代型 ChIP-seq 解析法の確立に成功し、精度の高いヒストン修飾解析及び転写因子結合部位を同定する基盤を構築した。またオープンクロマチン領域を同定することを目的に、ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) 法を用いて解析する基盤も構築した。

(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、最重要事項の一つとして取り組んできた。2018 年度においては、国立がん研究センター中央病院内に vendor neutral archive (VNA) システムを設置し、様々なモダリティのデータを統合した後に、閉鎖された専用 LAN を介して匿名化データとして研究所に移動するシステムを構築した。さらに、放射線画像に効率的に annotation 付けを行うことを目的として、AI も活用した annotation 付けプラットフォームの開発にも成功した。

(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

これまでゲノムデータ・エピゲノムデータは入力する parameter (P) に対して、sample number (N) が極端に低いという新 NP 問題と呼ばれる問題が存在し、これまで深層学習技術の活用が困難と考えられていた。そこで我々は 2018 年度においては、(a) オートエンコーダを中核とした深層学習技術と生物統計解析技術を誘導させたマルチオミックス解析データ解析手法の確立及び、(b) 入力に関して誤差逆伝播法に基づく学習可能なスケール因子も導入した改変型 Diet Networks に基づく解析手法の確立に取り組み、肺がんの予後予測及び病理型に基づく肺がん分類の高精度の予測に成功した。

§ 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 (株式会社 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