

「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進  
のための次世代アプリケーション技術の創出・高度化」

H26 年度  
実績報告書

平成 26 年度採択研究代表者

角田 達彦

独立行政法人 理化学研究所 統合生命医科学研究センター  
グループディレクター

医学・医療における臨床・全ゲノム・オミックスのビッグデータの解析に基づく  
疾患の原因探索・亜病態分類とリスク予測

## § 1. 研究実施体制

### (1)「角田」グループ

- ① 研究代表者:角田 達彦 ((独)理化学研究所 統合生命医科学研究センター、グループディレクター)
- ② 研究項目
  - 臨床・ゲノム・オミックス情報を用いたビッグデータ解析基盤の開発とバンク・多層オミックス・コホートサンプルの解析

### (2)「間野」グループ

- ① 主たる共同研究者:間野 修平 (情報システム研究機構 統計数理研究所、准教授)
- ② 研究項目
  - 各種ゲノムデータ解析に対する階層混合モデリングと経験ベイズ解析の方法論の開発
  - 個人ゲノムの秘匿性確保
  - 統計数理研究所内での研究グループの体制構築

### (3)「松井」グループ

- ① 主たる共同研究者: 松井 茂之 (名古屋大学大学院医学系研究科、教授)
- ② 研究項目
  - オミックスデータの解析体制の構築と解析項目の同定
  - 階層混合モデル解析に関する基礎的検討

(4)「猪狩」グループ

- ① 主たる共同研究者:猪狩 勝則 (東京女子医科大学膠原病リウマチ痛風センター、准教授)
- ② 研究項目
  - リウマチ患者の大規模前向き観察研究 IORRA コホートの調査項目とゲノムデータの整備、提供、解析と結果の医学的考察

(5)「越智」グループ

- ① 主たる共同研究者:越智 秀典 (広島大学大学院医歯薬保健学研究院、講師)
- ② 研究項目
  - 肝疾患臨床データベースの統一化と診療情報収集登録作業の効率化推進

(6)「加藤」グループ

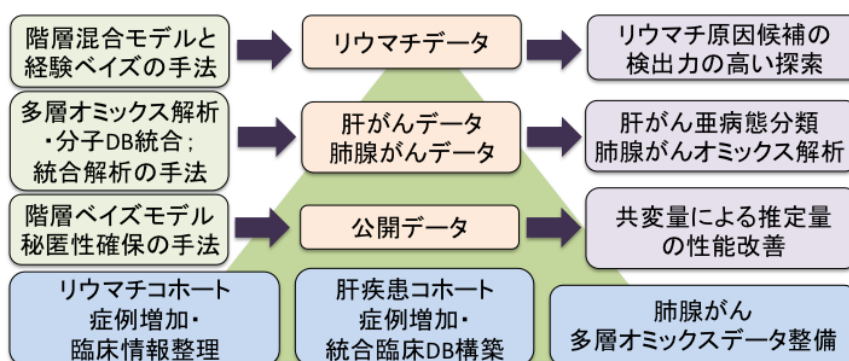
- ① 主たる共同研究者:加藤 護 (国立がん研究センター研究所、部門長)
- ② 研究項目
  - がん関連遺伝子のオミックス的性質の調査

## § 2. 研究実施の概要

平成 26 年度は、バイオバンク、多層オミックス、肝炎、リウマチのそれぞれの研究から提供されるゲノムデータを解析するための体制を構築し、臨床データ・オミックスデータの整理、外部の分子データベースと統合、そして解析を可能にするプロトタイプを作成し、以下の解析を開始した。

まず、手法開発と評価について述べる。検出力の高い探索方法として、階層ベイズモデルを用いた遺伝子発現・多型データなどの大規模データの統計解析方法の開発を行った。特に、治療反応予測や予後リスク診断法の開発における方法論的研究を行い、過去に行われた大規模臨床研究への適用などを通じてその有用性を評価した。さらに、階層混合モデル解析という汎用的なゲノムデータ解析の枠組みについて、遺伝子多型データなどを想定した基礎的検討を行った。そしてその手法をリウマチデータに適用することを開始した。同時に、階層ベイズモデルを用いた大規模データの有効な統計解析方法を、ケース・コホートデザインについて検討した。全検体の共変量を使うことで推定量の性能を大きく改善できることを示し、公開されているデータに適用した。さらに、多因子解析や亜病態分類のための多層オミックス解析の方法論と、オミックスのクラスタリング手法を進展させ、分子データベースによるネットワークモジュールを付加することにより、サンプル間の結束性を高め、亜病態分類を可能にする手法の開発を行った。そしてその有効性と情報量の増加、そして医学上の有用性を、300 例の肝がん全ゲノム・多層オミックスデータにて検討した。

実データの収集に関し、関節リウマチの IORRA コホートデータベースの整備拡充と DNA サンプルの追加収集については、平成 26 年 10 月-12 月に



行われた調査で 5,867 名からのデータ収集に成功した。また同コホート参加者からはこれまでに約 2500 名からインフォームドコンセント取得の下で DNA 収集を行っているが、平成 26 年 10 月-平成 26 年 3 月に新たに 128 名の DNA サンプルの追加収集を行った。また、肝疾患データに関しては、既存の疾患別臨床データベースから新規開発した統合臨床データベースへの移行を完了した。

多層オミックスデータの解析に関し、医薬基盤研「保健医療分野における基礎研究推進事業」にて国立がん研究センターが収集した肺がん 200 症例多層オミックスデータ(エキソーム、トランスクリプトーム、メチローム、プロテオーム、臨床情報)に対し、データクリーニングを行い、多層オミックス解析を実施した。具体的には、まず、エキソーム・シークエンスデータ、トランスクリプトーム・シークエンスデータに対し、研究代表者が開発したインハウス・プログラム等を用いて、SNV/indel コール、CNA コール、発現量コールを行い、さらにメチローム、プロテオーム、臨床情報を併せて解析できるよう紐付け ID のクリーニング、遺伝子やゲノム領域のデータ間の統一を行った。