

「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」

平成 26 年度採択研究代表者

H27 年度
実績報告書

角田 達彦

東京医科歯科大学 難治疾患研究所
教授

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

- ① 研究代表者:角田 達彦 (東京医科歯科大学難治疾患研究所、教授)
- 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)「角田」グループ2

- ① 研究代表者:角田 達彦 (国立研究開発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)
- 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

(3)「野間」グループ

- ① 主たる共同研究者:野間 久史 (情報システム研究機構 統計数理研究所、助教)
- 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

(4)「松井」グループ

- ① 主たる共同研究者：松井 茂之（名古屋大学医学系研究科、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

(5)「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

(6)「越智」グループ

- ① 主たる共同研究者：越智 秀典（広島大学医歯薬保健学研究院(医)、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

(7)「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 27 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、臨床データの整理、公開されている医学・生物学・分子データベースとオミックスデータを統合して解析可能にするプロトタイプを、より深化し、以下の解析を行った。

まず、疾患に関わる未知の遺伝子を新たに発見するための、オミックスデータと臨床データとの間の関連解析を高い検出力で行う新しい枠組みである階層ベイズ混合モデル解析を、それまでの遺伝子発現データに加え、遺伝子多型データに対しても適用可能とし、大規模なシミュレーション実験によりその有効性を確定した。この方法を、リウマチ、肝炎のゲノムワイド関連解析データに適用したところ、多くの未検出の疾患関連遺伝子の存在を示唆した。これを受けて、関連遺伝子多型の新規検出のために、既存の多重検定・ベイズ基準の比較も含めた検討を開始した。また、多発性骨髄腫に対しても多くの薬剤反応関連遺伝子を新たに検出し、疾患機序解明と創薬シーズ探索のための機能解析を開始した。

それに加え、大規模なゲノムワイド関連研究における、疾患関連遺伝子多型の、より効率的な探索方法のストラテジーの検討のため、スーパーコンピュータを使用した大規模シミュレーション実験を実施し、各種有効な多重検定手法の比較・評価を実施した。その結果、現状のスタンダードである、Familywise Error Rate (FWER) による保守的なストラテジーの非効率性が示され、さまざまな限界・問題があることが明らかとなった。本研究では、それに代わる有効な研究計画・統計解析のストラテジーについて、詳細な検討を行った。

また、近年、集団内にごく低い頻度でしか存在しない、疾患に関連する稀な多様性 (rare variant) の検出が、疾患のメカニズム解明のために大きな関心が寄せられている。本研究では、一定の条件のもと、理論上、最大の検出力を有する、疾患関連多様性の検出のための検定手法を開発した。特に、既存の方法論は、特定の想定したモデルの仮定に強く依存した方法であったため、条件によって、疾患関連多様性の検出の効率に不安定性があったが、本研究によって開発された手法は、より緩やかな仮定のもとでの最適性が保証されることが示された。

次に、疾患の亜病態関連解析の方法論として、混合モデルに基づく二方向クラスタリング法を開発し、シミュレーションデータ、骨髄異形成症候群の遺伝子発現データへの適用を通してその有効性を確認した。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、前年度に調査したオミックスの性質をもとにし、分子データベースなどの外部からの医学・生物学・分子データベースを付加し、多層オミックスのクラスタリング手法を進展させた。その際の大規模データにおけるクラスタリング手法の基礎アルゴリズムの開発も新規に行った (*IEEE Transactions on Biomedical Engineering*, 2016 Mar 24, オンライン)。そして前年度、医薬基盤研にて国立がん研究センターが収集した肺腺がん 200 症例多層オミックスデータの解析において培った解析方法を合わせてシステム化した。実データ解析として、国際がんゲノムコンソーシアム・全日本チームが産出した肝がん 300 症例の全ゲノム・多層オミックスデータ(ゲノム・データ、トランスクリプトーム・データ、メチローム・データ)に対して適用した結果、オミックスデータに基づく 6 種の亜病態の分類の新たな発見、および再発・予後などの臨床情報と有意に相関する分子マーカーの発見に成功し、

研究成果を **Nature Genetics** 誌において発表した (*Nature Genetics*, 48, 500-509 (2016))。

これらのフォローアップと検証のための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 27 年に行われた調査でのべおよそ 12,000 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新たにおよそ 500 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。