

「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」

H28 年度
実績報告書

平成 26 年度採択研究代表者

角田 達彦

東京医科歯科大学 難治疾患研究所
教授

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

(2)「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、准教授)
- ② 研究項目
 - ゲノムビッグデータの統計解析における疾患関連遺伝子探索のための方法論・ストラテジーの開発と評価
 - 稀な遺伝的変異の疾患発症リスクの評価
 - 個人ゲノムの秘匿性確保

(3)「松井」グループ

- ① 主たる共同研究者：松井 茂之（名古屋大学医学系研究科、教授）
- ② 研究項目
 - オミックスデータを用いた関連遺伝子の探索と予測解析
 - 階層混合モデルと経験ベイズ解析の開発と各種疾患への適用
 - 疾患の遺伝的異質性の解析

(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

(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
 - がんオミックスデータを活用したリスク予測マーカーの探索と臨床応用の可能性

§ 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、特に化学療法などの治療法や予後などとの関係を解明し患者さんへの還元の道筋を明らかにすべく、以下の研究を行った。

まず、前年度に開発したセミパラメトリック階層混合モデル (semi-parametric hierarchical mixture model; SP-HMM) を用いて、疾患の多遺伝的構造を定量的に評価し、疾患間の相違のみならず、アレル頻度や遺伝子発現調整領域 (eQTL) への顕著な依存を明らかにした。併せて、関連遺伝子が未だ報告されていない大鬱病に対して、関連遺伝子多型の発見に必要な症例数を予測した。さらに、多次元 SP-HMM 解析を開発し、多発性骨髄腫に対するサリドマイドや二型糖尿病に対するインスリン抵抗性改善薬剤メトフォルミン等の薬剤効果予測解析に適用し、薬剤効果予測遺伝子を同定した (*Biometrics*, published online: 12 MAY 2017, doi: 10.1111/biom.12716)。疾患の遺伝的異質性の解析では、がん固有の局所的遺伝子発現異常 (cancer outliers) を捉える混合モデル解析を開発し、がんの分類と判別解析への適用を開始した。

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。そこで、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。そして、オミックスビッグデータ解析のための基本的な解析手法として、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を新規に行った。これは最尤法を用いた、複雑な症例の混合にも対応する階層クラスタリング手法であり、例題として、日本人集団の地域差によるゲノム背景の違いがより明確になった (*IEEE Transactions on Biomedical Engineering*, 64, 112-122 (2017))。同時に、ボトムアップに、1 事例ずつクラスタに入れるかどうかを最尤法で決めることを繰り返す手法の提案も行い、より複雑な混合の場合にも対応可能となった (*BMC Bioinformatics*, 17(1), 319 (2016))。

一方、がんの亜病態を分類するためのマーカー選択の方法はまだ未発達であることから、

遺伝子マスキングという手法を独自に開発、提案し、高次元オミックスデータの一例としてマイクロアレー遺伝子発現データでの分類の精度向上を達成した (*BMC Medical Genomics*, 9(Suppl 3), 74 (2016))。

続き、次世代シーケンサーデータの実験解析技術として、現在用いられている手法を比較し、疾患遺伝子探索のためのエクソーム解析の指針を見出したことから、各種講演や総説に招待され、成果の公表を行った。それと同時に、実験手法で、エクソーム解析の新たな実験手法の提案、開発と成果の公表を行った。

さて、がん多層オミックスデータを用い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、再発の起こりにくい症例と、それを特徴付ける変異のある、新規がん関連遺伝子 *MACROD2*、そして繊維化との関係を発見した。その研究成果を *Nature Genetics* 誌において発表し、実臨床での応用方法を示した (*Nature Genetics*, 48, 500-509 (2016))。

また、AstraZeneca 社主催の世界的なコンペティションである DREAM challenge で、(1) 薬剤併用による効果の予測のコンテストと、(2) 疾患モジュール同定コンテストに参加し、トップの成績を収め、受賞した。その手法の詳細を本 CREST プロジェクトの成果として招待講演およびポスターで発表した。

そして患者の予測に関する研究成果として、乳がんの予後 (Overall Survival と Disease Free Survival) の予測の確立と精度向上を達成した。これは 81 例の患者の 10 年以上の長期にわたるフォローの貴重な結果であり、検証セットでは公共で得られる 420 例ほどの患者データを追加で用いた。その結果、遺伝子発現による新規マーカーが発見でき、患者の予後の予測精度の向上に多大な寄与が見られる成果が得られた (*Cancer Medicine*, published online: 24 MAY 2017, DOI: 10.1002/cam4.1092)。これにより、通常用いられている遺伝子セットによる MammaPrint をさらに改善できると期待がもてる。

これらに加え、現在、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化学療法の奏効データをもとに、バイオマーカー探索と症例の亜病態分類を行っている。

また、生活習慣病のビッグデータ解析の代表的な成果として、関節リウマチの発症リスクの上昇要因を見出すため、細胞特異的なパスウェイへのポリジーン効果を用いることに成功した。近年、生活習慣病のリスクアレルの多くが、細胞特異的に遺伝子発現を制御していることがわかりつつあったことから、今回、免疫関連の生活習慣病の原因となる遺伝子やパスウェイを同定するため、日本人の健康なボランティアから得られた、5 種類の免疫細胞と末梢血の eQTL 解析を行った。得られた eQTL データに加え、公共エピゲノムデータを使うことで、患者個人の個々の遺伝子発現を予測し、これまでのゲノムワイド関連解析とは異なり、遺伝子レベルの関連解析を行い、細胞特異的なパスウェイの動きの予測を行えるようにした。これを関節リウマチに用い、新たなリウマチ関連遺伝子とサイトカインパスウェイの発見に成功した (*Nature Genetics*, published online: 29 MAY 2017, doi:10.1038/ng.3885)。

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。