

戦略的創造研究推進事業 CREST
研究領域「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」
研究課題「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」

研究終了報告書

研究期間 2014年10月～2021年3月

研究代表者:角田 達彦
(東京大学大学院理学系研究科、教授)

§ 1 研究実施の概要

(1) 実施概要

【全体の考え方】 患者オミックス・臨床・分子の医療ビッグデータの統合解析による個別化医療の根幹の確立を目指した。特に当分野に内在する高次元性と潜在的ヘテロ性による解決の困難性の克服のための学術的体系を確立すべく、ビッグデータ統合、効率的バイオマーカー探索や疾患亜病態分類を含む個人病態・治療反応性の semi-supervised 手法等による予測、臨床試験計画の順の方針に沿って研究を進めた。

【解析スキームと一般化フレームワーク】 1年追加支援による実施内容とし、解析スキームを深めた。まず、各患者から得られるオミックスデータと臨床データが得られるとする。そして時間的には後にはなるが、臨床データの一部として、各患者の予後や治療奏効データも得られる。臨床の場では、直接の治療奏効も、予後も、両方が治療の評価の対象となる。これらのオミックス・臨床データと予後・治療奏効データとを突き合わせ、予後・治療奏効を予測するためのバイオマーカーを抽出し、予後・治療奏効の予測モデルを構築する。このモデルに対してさらなる臨床データを用いてフィードバックし、モデルを洗練化する。新規の患者に対しては、オミックス・臨床データ(予後・治療奏効以外)をもとに、予後・治療奏効(可能なら副作用も)を治療法ごとに予測する。この結果から患者個人にとって最適な治療法を選択する個人化医療を実現することを目標とする。この際に問題になることは、背景となる疾患や薬剤応答メカニズムが均一でないこと、またそれらが必ずしも一対一対応しないこと、そしてオミックス・臨床データが検体数に比べてはるかに高次元になることである。したがって、トップダウンアプローチによってバイオマーカー探索し予測モデルを構築することは難しいと判断した。そこで本プロジェクトでは、特徴抽出手法や医学・生物学的知見を用いてボトムアップに個人ごとのオミックス・臨床データから特徴を計算し、それらと治療効果をトップダウンに統合する・重み付ける、というセミスーパーバイズドな方法論が適っていると判断した。その応用対象に非依存な部分を体系化し、一般化した統合フレームワークを確立したことが本プロジェクトの最大の成果の一つである。そして実際のデータを解析する際には疾患の特性に合わせ、応用対象に依存した部分が生じるため、データ処理や統計解析の各段階でそれらの個別性を扱う必要が生じる。これらにより、フレームワークは下記の手法の(a)~(c)からなり、この順に解析を進める。また(d)は臨床試験との接続のための設計思想であり、実証に欠かせない。

【手法開発】 (a) **ビッグデータ標準化と統合の手法**: 角田 G が次世代シーケンサーデータから中間サイズの断片の挿入や欠失を予測する方法、加藤 G が FFPE 試料に適用できる実臨床シーケンス高精度解析アルゴリズムを開発し、企業技術移転と各機関公認、がんゲノム医療検査で保険適用承認のもと、がんゲノム医療を推進した。また野間 G が個人ゲノムデータの秘匿性と公開キーの組合せで機微変数が知られるリスクの評価方法論を構築した。

(b) **疾患・薬剤応答因子(バイオマーカー)の探索の手法**: 松井 G がセミパラメトリック階層混合モデル SP-HMM と経験ベイズにより、表現型に関わるオミックス変数の割合と効果サイズ分布や患者亜集団背後の関連構造を推定する新しい枠組みと、統計学と生物学的な二段階マーカー選択統計法を、また野間 G がオミックスランダム化臨床試験での治療反応性予測のためベイズ階層混合モデルによる交互作用検出方法と最適ランキング・新たな指標、疾患関連の稀な多様性を検出する3段階階層混合モデルを開発した。両 G で理論上最大検出力の OPD による検定手法を導いた。続き角田 G と加藤 G が再発・予後・治療奏効の臨床情報とオミックスを解析する semi-supervised 手法、種間オミックスを統合し因子を探索する手法、患者個人の個々の遺伝子発現を予測し遺伝子レベルの関連解析を行う方法を開発した。

(c) **疾患亜病態分類と個人病態・治療反応性予測の手法**: 松井 G が、疾患内分子レベル異質性モデル、入れ子型二方向混合モデルとその推定に基づく判別・予測法、疾患異質性の高次元交互作用モデリング、非線形特徴量の抽出が可能な Lasso を開発した。併せ角田

G が、超高次元少数例データをタンパク質間相互作用を用いて情報量を増しクラスタ化する多層オミックス潜在構造分析および semi-supervised 手法を提案した。

1 年追加支援による実施内容とし、その枠組みをさらに深層学習を用いて拡張した。

予測では、松井 G が臨床データと多数のオミックスマーカー変数の周辺分布情報に基づく予後関連マーカー検出・予後予測解析をパッケージ化、野間 G がランダム化臨床試験・観察研究の双方で利用可能な個人治療効果の推定に勾配ブースティング木を実装した手法や深層学習による個別治療効果推定の方法、最適治療方針の決定理論的定式化から個人ごとの治療効果を階層ベイズモデルに基づき推定する方法を開発した。個人化医療の基盤とし、角田 G と松井 G が個人ごと薬剤ごとに予後を予測する方法を開発した。

(d) **臨床試験計画デザインの手法**: 松井 G が、薬剤効果予測因子の検証に向けて効率的な治療効果予測マーカーを用いた第三相臨床試験のデザインの手法と、多様な治療効果プロファイルの複数のマーカー規定患者集団を想定した試験デザインの評価基準を開発した。

【実データ解析の成果】

(i) **がん多層オミックスの解析、予後・薬剤奏効予測**: 角田 G、加藤 G、越智・三木 G が連携し、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 300 例の肝がん多層オミックスデータを上記手法により解析し、バイオマーカーを発見、亜病態分類に成功、低再発症例に特徴的な新規がん関連遺伝子発見と繊維化、治療との関係を解明した。また血中の ctDNA を調べ、がん診療での薬剤選択や予後予測にて通常腫瘍マーカーよりリアルタイム・高精度に病勢や腫瘍の性質がわかることを報告した。続き角田 G は乳がんでも、患者の長期フォロー結果と遺伝子発現から新規マーカー群を発見、複合因子効果をスコア化し再発高リスク患者や長期生存期間患者の予測に成功し、日本人・欧米人両方での有効性を証明した。さらに加藤 G が、希少がん、消化管間質腫瘍の多層オミックスから薬剤耐性機序を解明した。また国際 ICGC と TCGA の多層オミックスを用いた横断解析を実施し、各がん種に対し予後マーカーを発見、HLA、患者の遺伝子多型、non-coding RNA が予後に関係することを新たに発見した。そして国際がんゲノムコンソーシアムの国際連携により、がん種横断的な全ゲノム網羅的データ解析・機能解析を行い、発症や予後に関わる変異機能の集大成をまとめた。1 年追加支援によりさらに、角田 G が胃がんのがん免疫学的な側面の共同研究を行い、症例の免疫学的なヘテロ性とそれぞれに対する治療法の指針を見出した。

(ii) **がん多層オミックス臨床試験研究との連携**: 松井 G が、SP-HMM を多発性骨髄腫の治療薬サリドマイドの効果検証ランダム化臨床試験に適用し、その効果予測遺伝子群を発見した。また、分子レベル疾患異質性モデリングを、骨髄異形成症候群、多発性骨髄腫の病型分類に適用、サブ集団関連構造の同時分布の解析方法を多発性骨髄腫の治療薬サリドマイド、大腸がんの治療薬 S1、二型糖尿病の治療薬メトホルミンに適用し、効果予測解析を行った。続き野間 G が OPD を使って米国大規模ランダム化比較試験の解析を行い、新規の有意な関連要因を検出した。また角田 G と加藤 G が、各臨床試験研究グループと連携し、がん化学療法、がん免疫療法の奏効に関わるバイオマーカー解析と予測モデル構築を進めた。

(iii) **関節リウマチ等バイオバンクデータ等の多層オミックス**: 角田 G が開発した遺伝子発現予測による関連解析を行う方法を、角田・猪狩 G の関節リウマチ症例に適用し、新たなリウマチ関連遺伝子と、CD4⁺T 細胞での TNF サイトカインパスウェイの関与を発見した。また同発症に関わる miRNA の同定に成功、関節破壊度では HLA アレルが疾患活動性とは独立して関わることを発見した。そしてプレジジョン医療のため、関節リウマチの分子標的薬 TNF 阻害薬の治療反応性関連遺伝子探索を大規模国際共同研究により行い、治療反応性関連遺伝子群を発見、複数因子の組み合わせをスコア化して高精度な予測モデルを構築した。続き、大規模な国際協力のもと喘息のゲノムワイド関連統合解析を行い、新しい喘息関連遺伝子群を発見、eQTL により、それらが免疫学的に関わることを解明した。同様に、独自に開発した種間オミックス解析手法を用い、アルツハイマー患者の脳のアβ蓄積の原因となる新規遺伝子群を発見、認知症前向きコホートの血清 miRNA 発現を用い、世界で初めて認知症でのアルツハイマー病・血管性認知症・レビー小体型認知症の亜病態を区別するバイオマーカーを発見し

リスク予測モデルを構築した。このアルツハイマー研究に関し、さらに1年追加支援により、認知低下の症状からアルツハイマーに進展する予測モデルをマルチオミクスにより構築し発表した。

肝がんを含む肝疾患に関し、松井 G の階層混合モデリング等の手法を越智・三木 G が HCV の持続感染データに適用し、宿主遺伝要因とウイルス側因子との間の相互作用が HCV 持続感染成立に寄与するリスク、それらが抗ウイルス治療に対するウイルスの応答性を調べ、また肝炎の GWAS データに適用し、多くの未検出の肝疾患関連遺伝子の存在を見出した。続き越智・三木 G が、C 型肝炎根治後にも依然として残る肝がん発症リスクや非ウイルス性の肝がんの遺伝的要因や臨床的リスク因子を解明した。最後に、松井 G、角田 G、猪狩 G、越智・三木 G が連携し、関連構造推定の枠組みを用いて、リウマチ、精神疾患等の多様な疾患間の多遺伝子構造の相違を定量的に評価し、アレル頻度や遺伝子発現調整領域への顕著な依存を発見、関連遺伝子多型の発見に将来必要な症例数を予測した。

(2) 顕著な成果

<優れた基礎研究としての成果>

1.

概要： 独自開発のがん多層オミクス解析手法を肝がん 300 症例の全ゲノム・オミクスデータに適用し、6 種の亜病態の分類法、再発・予後などの臨床情報と関連する分子マーカー、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、そして肝臓の繊維化との密接な関係を発見した (*Nature Genetics*, **48**, 500-509 (2016))。かつ国際がんゲノムコンソーシアムに参画し、この肝がんを含む 2,658 のがん種横断的症例の全ゲノム網羅的機能解析を行い、非コード領域 RNA の関与などを発見した (*Nature*, **578**, 102-111 (2020))。

2.

概要： 健康者から得られた 5 種類の免疫細胞と末分画末梢血の eQTL データと公共エピゲノムデータを組み合わせることで、個体内の遺伝子発現を予測し、(従来のゲノムワイド関連解析とは異なり) 遺伝子レベルでの関連解析と細胞特異的なパスウェイの予測が可能となった。これを関節リウマチに適用し、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*, **49**, 1120-1125 (2017))。多くのリスクアレルが制御している生活習慣病に対して、細胞特異的なパスウェイの解析に成功した先駆的研究である。

3.

概要： 患者のケアを改善し、個別化医療を推進するには、個別化生存死亡予測モデルの開発が必要である。ゲノム情報と大規模メタ分析データセットが利用可能になり、従来の生存予測を超えた解析の必要性が臨床の場に広がった。そこで我々は、台湾のグループと連携し、虚弱コピュラモデルを高次元の遺伝因子にも拡張し、遺伝的要因と動的な腫瘍進行状態と予測式を考慮した個別化死亡リスク予測式を開発 (*Statistical Methods in Medical Research*, **27**, 2842-2858 (2018))、パッケージ化し、卵巣がん患者の個別化データで有効性を実証した。

<科学技術イノベーションに大きく寄与する成果>

1.

概要： 多発性骨髄腫のランダム化臨床試験において、サリドマイド投与により延命する患者を分類する新しいマーカー解析法を、セミパラメトリック階層混合モデルを用いて開発し、681 個の効果予測遺伝子のスクリーニングに成功した (*Biometrics*, **74**, 313-320 (2018))。開発した階層混合モデル解析は、オミクスデータと臨床データの背後にある関連構造の柔軟な推

定と、視覚化を可能とし、リウマチ、統合失調症などの多遺伝的疾患のゲノムワイド関連解析や二型糖尿病の薬剤効果予測マーカー解析など、既に多くの適用事例が集積しており、オミックスデータの汎用的解析技術としての確立が期待される。

2.

概要：本邦における81例の早期乳がん患者の10年以上の長期追跡データを用いて、予後予測のシステムを新たに開発した。検証セットとして、公共で得られる420例の患者データを用いたところ、追跡開始時の遺伝子発現を用いることで、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**, 6, 1627-1638 (2017))。現在、MammaPrint®を始めとする海外で開発された予後予測システムが国内で多く用いられているが、日本人における予測精度の限界が指摘されており、待望されている本邦発の予後予測システムの開発に繋がるものと期待される。

3.

概要：臨床現場での治療の礎として、FFPE組織検体の臨床シーケンス用の高精度なSNV/indel, fusion, CNA 検出アルゴリズムを開発した (**Genome Medicine**, 10, 44 (2018))。国立がん研究センターのがんゲノム医療臨床研究 TOPGEAR と MASTER KEY プロジェクトにて実際に使用、企業技術移転、がんゲノム中核一連携病院体制の先進医療 B の遺伝子検査「個別化医療に向けたマルチプレックス遺伝子パネル検査研究」にて、実際に使用された。また日本初のがんゲノム医療機器として厚生労働省所管 PMDA に公的に承認された。

< 代表的な論文 >

1. Fujimoto A+, Furuta M+, Totoki Y+, Tsunoda T+, Kato M+ (+: co-first), Shiraishi Y, Tanaka H, Taniguchi H, Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Wardell CP, Hayami S, Nakamura T, Aikata H, Arihiro K, Boroevich KA, Abe T, Nakano K, Maejima K, Sasaki-Oku A, Ohsawa A, Shibuya T, Nakamura H, Hama N, Hosoda F, Arai Y, Ohashi S, Urushidate T, Nagae G, Yamamoto S, Ueda H, Tatsuno K, Ojima H, Hiraoka N, Okusaka T, Kubo M, Marubashi S, Yamada T, Hirano S, Yamamoto M, Ohdan H, Shimada K, Ishikawa O, Yamaue H, Chayama K, Miyano S, Aburatani H, Shibata T, Nakagawa H. Whole genome mutational landscape and characterization of non-coding and structural mutations in liver cancer. **Nature Genetics**, 48, 500-509 (2016).

概要：ビッグデータ解析による肝がん300症例の亜病態分類から、再発率の極端に低いクラスとその変異パターンを世界で初めて発見した、臨床的にも大変有意義な論文である。独自に開発したがん多層オミックス解析手法を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に適用し、6種の亜病態の分類法を新たに発見し、さらに、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、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。その研究成果を発表し、実臨床での応用方法を示した。

2. Ishigaki K, Kochi Y, Suzuki A, Tsuchida Y, Tsuchiya H, Sumitomo S, Yamaguchi K, Nagafuchi Y, Nakachi S, Kato R, Sakurai K, Shoda H, Ikari K, Taniguchi A, Yamanaka H, Miya F, Tsunoda T, Okada Y, Momozawa Y, Kamatani Y, Yamada R, Kubo M, Fujio K, Yamamoto K. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. **Nature Genetics**, 49, 1120-1125 (2017).

概要：健康者から得られた5種類の免疫細胞と未分画末梢血のeQTLデータと公共エピゲノムデータを組み合わせることで、個体内の遺伝子発現を予測し、(従来のゲノムワイド関連解析とは異なり) 遺伝子レベルでの関連解析と細胞特異的なパスウェイの予測が可能となった。これを関節リウマチに適用し、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。多くのリスクアレルが細胞特異的な遺伝子発現を制御している生活習慣病に対して、細胞特異的なパスウェイへのポリジーン効果の解析に成功した先駆的研究である。

3. Matsui S, Noma H, Qu P, Sakai Y, Matsui K, Heuck C, Crowley J. Multi-subgroup gene screening using semi-parametric hierarchical mixture models and the optimal discovery procedure: Application to a randomized clinical trial in multiple myeloma. *Biometrics*, **74**, 313-320 (2018).

概要:多発性骨髄腫のランダム化臨床試験で、サリドマイド投与により延命する患者を分類する新しいマーカー解析法を、セミパラメトリック階層混合モデルを用いて開発し、681個の効果予測遺伝子のスクリーニングに成功した。開発した階層混合モデル解析は、オミクスデータと臨床データの背後にある関連構造の柔軟な推定と、視覚化を可能とし、リウマチ、統合失調症などの多遺伝的疾患のゲノムワイド関連解析や二型糖尿病の薬剤効果予測マーカー解析など、既に多くの適用事例が集積しており、オミクスデータの汎用的解析技術としての確立が期待される。

§2 研究実施体制

(1) 研究チームの体制について

① 「角田」G1

- ・研究代表者:角田 達彦 (東京大学大学院理学系研究科、教授)
- ・研究項目
- ・臨床・ゲノム・オミ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・多層オミクス・コホートサンプルの解析

② 「角田」G2

- ・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)
- ・研究項目
- ・臨床・ゲノム・オミ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・多層オミクス・コホートサンプルの解析

③ 「角田」G3

- ・研究代表者:角田 達彦 (理化学研究所 生命医科学研究センター、チームリーダー)
- ・研究項目
- ・臨床・ゲノム・オミ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・多層オミクス・コホートサンプルの解析

④ 「野間」G

- ・主たる共同研究者:野間 修平 (統計数理研究所、准教授)、野間 久史 (統計数理研究所、准教授)
- ・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
- ・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
- ・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

⑤ 「松井」G

- ・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)
- ・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

⑥ 「猪狩」G

- ・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)
- ・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

⑦ 「三木」G

- ・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・診療准教授)、三木 大樹 (広島大学医歯薬保健学研究院(医)、講師)
- ・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

⑧ 「加藤」G

- ・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)
- ・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)国内外の研究者や産業界等との連携によるネットワーク形成の状況について

- ① 肝がんの全ゲノム・オミックスと化学療法を含む治療法奏効のデータ取得と解析に関し、広島大学病院 茶山一彰教授、理化学研究所 生命医科学研究センター 中川英刀チームリーダー、東京医科歯科大学 肝胆膵外科の田邊 稔教授と連携を行なった。
- ② 肺腺がんのオミックスと化学療法のデータの取得、化学療法奏効マーカー同定と奏効予測システム構築に関し、国立がん研究センター研究所 吉田輝彦分野長、河野隆志分野長、白石航也研究員と密に連携した。現在もその連携を続けている。
- ③ 大腸がんのオミックスと化学療法のデータに関し、東京医科歯科大学 植竹宏之教授との連携を行なった。当 CREST プロジェクトの松井 G (松井茂之教授)も既に植竹教授と大腸がん遺伝子発現と治療有効性の研究を行っていたが、さらに角田 G もゲノムシーケンスを加えたがん多層オミックスと化学療法を突き合わせた解析を行った。現在もその連携を続けている。
- ④ 卵巣がんのオミックスと化学療法のデータに関し、ヘルシンキ大学・カロリンスカ研究所 Olli Kallioniemi 教授、Dr. Simon Anders, Dr. Caroline Heckman、東京医科歯科大学 稲澤譲治教授、慶應義塾大学医学部産婦人科 青木大輔教授、岡山大学大学院医歯薬学総合研究科 平沢 晃教授と連携した。
- ⑤ がん免疫は、第 4 のがん治療として世界で認識され始め、小野薬品のオブジーゴなどが大変良い効果を示すことがわかってきた。しかし、一人の患者が支払う額は大変高価である。また現時点で有効な患者は全体の 2~3 割にとどまることから、選択のためのマーカーと、他のがん免疫治療薬の開発の重要性が認識されている。がん免疫に関し、東京大学医学部附属病院 垣見和宏特任教授と連携し、免疫プロファイルやがんオミックス解析を行う計画をたて

た。さらに1年追加支援で、この連携により胃がんの免疫学的個性と治療法への示唆について論文発表した。現在もその連携を続けている。

- ⑥ 角田・加藤 G は、世界中の研究者との連携体制であり、様々ながん種横断的にオミックス解析をするプロジェクトある国際がんゲノムコンソーシアムに参画した。これは世界中の 37 カ国が連携したプロジェクトであり、2,658 例のがん種横断的な全ゲノム・オミックスの網羅的解析を行い、2020 年 2 月 5 日に *Nature* 6 報を含むトップジャーナル 22 報の論文を出版し成果報告を行ったものである。
- ⑦ 腸内細菌に関して、東京工業大学生命理工学院 山田拓司准教授(データさきがけ)と、角田 G、猪狩 G との共同研究を進めた。
- ⑧ 非線形特徴選択の手法開発に関して、理化学研究所革新知能統合研究センター 山田誠ユニット長(社会さきがけ、現在 京都大学大学院情報学研究科 知能情報学専攻 集合知システム 准教授)との共同研究を行なった。これにより cancer outlier などの病態およびマーカーのヘテロ性の解明の研究が強力に推進された。
- ⑨ 松井 G は、台湾の研究グループと共同で、多数のオミックスマーカーを探索的に選択して予後予測を行う方法とパッケージを開発した。現在もその連携を続けている。
- ⑩ 角田 G は、オミックスデータを深層学習に適用する方法論の開発のため、フィジーの南太平洋大学のグループと共同研究を行った。現在もその連携を続けている。
- ⑪ 1 年追加支援で、臨床試験研究グループとの連携とし、JCOG-BBJ 東大医科研の松田浩一教授、がん研究会の阿江啓介部長・船内雄生副医長、国立がん研究センター中央病院の中谷文彦医長と、肉腫を中心としたがんの予後および治療奏効の研究を進めた。