

**厚生労働科学研究費補助金**  
**(政策科学総合研究事業 (臨床研究等 ICT 基盤構築・人工知能実装研究事業))**  
**分担研究報告書**

課題名 : 新薬創出を加速する症例データベースの構築・拡充/創薬ターゲット推定アルゴリズムの開発に関する研究

研究分担者名 : 奥野 恭史

国立大学法人京都大学大学院 医学研究科 人間健康科学系専攻 教授

**研究要旨**

**創薬標的を探索する AI 開発**

**薬剤反応性の識別、バイオマーカー推定、分子メカニズムの解明と創薬標的探索を行う AI の開発**

医薬品開発において、近年国内外を問わず創薬ターゲットの枯渇が問題となっている。現在残されているのは高難易度の創薬ターゲットのみであるがために、新薬の研究開発には多額の費用が必要となっており、これが高薬価、ひいては医療費の高騰の要因となっている。

更に、臨床試験段階で期待していた薬効が得られず開発が中断する例が増えていることも問題点として挙げられる。特に医薬品開発の 70~80%が Phase2 で中止となっており、この約 60%が、薬効が得られなかったことが原因との報告がある。つまり、「動物では効くが、ヒトでは効かなかった」という事案が多発している。これは現在の創薬研究開発スキームの限界であると考えられる。

このような現状を打開する解決策として、人工知能 (AI; Artificial intelligence) が注目されている。AI のパフォーマンスと可能性に創薬・医療・ヘルスケア分野が大きな期待を寄せており、今後国際競争が激化することが必至である。

これらの現状を背景に、本事業では、「創薬ターゲットの枯渇問題」を克服すべく、動物からではなくヒトの情報から創薬ターゲット分子を探索する AI の開発実装を目的とする。つまり、臨床情報 (=電子カルテを始めとする診療情報+オミックスデータ) を収集・利用して創薬標的を探索する AI 手法の開発をおこなう。本事業では、対象疾患として難病指定の IPF (特発性肺線維症) を含む間質性肺炎及び部位別がん死亡者数 1 位である肺がんを選択し、これらの臨床情報収集とそれを支援する基盤構築、異種かつ大量のデータを統合して創薬標的候補となる生体分子群を自動的に抽出する AI 手法の開発、創薬標的候補の実験的検証に有益となる基盤構築を包括的に遂行する。当該年度は、i)データ収集: IPF を含む間質性肺炎で 300 検体の追加及び肺がんデータ収集の欠損補填、ii)データ解析: これまでに開発した各種 AI による対象疾患患者臨床情報の解析とそれによる創薬標的候補の提示、iii)結果解釈・仮説創出: 文献からの自動知識抽出を介した未知の関係性推論技術の開発及びデータウェアハウス TargetMine を用いた ii)データ解析の結果解釈、iv)糖鎖解析により由来臓器を推定する基盤の構築、v)オープンプラットフォームの基盤構築と創薬支援を志向した AI の開発・搭載、国内運用開始に向けた体制構築を目標とする。

**A. 研究目的**

患者個別の疾患タイプ、薬剤反応性等について、識別・分類、分子メカニズムの解明、バイオマーカー推定と創薬標的分子探索を可能にする AI 技術を開発する。特に、生命科学や医療における様々な情報を創薬に応用する上で問題となる以下の課題の克服を目指す。

「克服すべき技術的な課題」

- ・全ゲノム配列 (超次元データ)、マルチオミックスのデータ処理
- ・薬剤反応性などの患者層別化・個別化
- ・多種多様情報 (実験データ、臨床データ、文献データ等) の統合
- ・分子-パスウェイ-細胞-臓器-動物-ヒトのマルチスケールモデル
- ・AI モデルへの時間的・空間的・定量的概念の導入

## B. 研究方法

### ① DNN や機械学習による薬剤反応性の識別とバイオマーカー推定

限られたデータを有効活用し、効率的に薬剤反応性の予測やバイオマーカーの推定を行うためには、ゲノムやオミクスデータ、薬剤（化合物）、臨床情報の情報を組み合わせた超高次元のデータを扱う必要がある。そのため本年度では、これらのデータを組み合わせて利用する方法として、昨年度は、オミクスデータ、化合物、臨床情報などの複数種類のデータを入力するマルチモーダルアプローチが有効であることが分かった。今年度では、前年度で開発した実装を元に、さらに改良したモデルとして、グラフデータとサンプルデータを同時に利用可能なマルチモーダルニューラルネットワークを構築し、実際に公共で利用可能なデータを用いてモデル構築を行い、これら二つのアプローチを薬剤反応性識別やバイオマーカーの推定に利用する場合にどのような長所や短所があるかといったことを明らかにする。

本年度では、公開ベンチマークデータを用いて、上記の既存手法に関する調査及び実装を行い、これらのアプローチの優位性について評価を行った。

### ② GCN によるネットワーク構造を考慮した薬剤反応性の識別、バイオマーカー推定、創薬ターゲット分子探索

限られたコストで効率的に実験を行い、得られた実測データを効果的に活用する方法の一つは、文献や既存の公共データベースの知識を利用することである。文献や既存の公共データベースにある網羅的な情報（文献空間の情報）を実験によって得られた実測空間の情報を統合した機械学習の方法を開発することで、少ない実測データからの効率的な予測を実現する。これらの異なる性質を持つ情報を柔軟に表現する方法として、本研究課題ではグラフ表現（知識グラフ）を利用する。現状のシステムではこれらの文献や既存の知識の表現として、知識グラフを用いており、その知識グラフの学習のために GCN(Graph convolutional network)を用いている。GCN では、ネットワーク構造を教師データとして学習し、複雑なネットワークの中から新たな関係性や特性を推定できる技術であり、文献や種々のデータベースに登録される膨大かつ多種多様なデータからの知識発見に有力であると期待されている。今年度では GCN によるネットワークの予測に関するソフトウェアのツール化を行った。ツール化により、様々なデータに容易に適用可能にし、実データの分析において、試行錯誤が効率よく可能になることが期待できる。

### ③ ベイジアンネットワークによる分子メカニズムの解明、バイオマーカー推定、創薬ターゲット分子探索

ベイジアンネットワークを用いた研究のうち、今年度は本事業で収集した臨床情報を用いたベイジアンネットワークによる分子メカニズム解明、バイオマーカー推定、創薬ターゲット分子探索方法の確立を目指す。前年度においてトランスクリプトーム・データを用い、患者毎の薬剤への反応の違いや患者毎の分子ネットワークの違いを、ベイジアンネットワークを用いて解析する方法を確立することができた。今年度はこれらを発展させることで臨床情報を含む多種多様な情報を用いた分子メカニズムの解明、創薬ターゲット分子探索が目標で、同時にシミュレーション技術と組み合わせたによるバイオマーカー探索技術の研究開発を押し進める。

#### （倫理面への配慮）

本研究は、医薬基盤・健康・栄養研究所ならびに本学において倫理審査、承認を得た後、ヒトゲノム・遺伝子解析研究に関する倫理指針及び、人を対象とする医学系研究に関する倫理指針に従って遂行した。

## C. 研究結果

### ① DNN や機械学習による薬剤反応性の識別とバイオマーカー推定

限られたデータを有効活用し、効率的に薬剤反応性の予測やバイオマーカーの推定を行うためには、ゲノムやオミクスデータ、薬剤（化合物）、臨床情報の情報を組み合わせたマルチモーダル情報を扱うことが重要である。昨年度に献相当のグラフデータとして公開データの Reactome を利用し、実測データ相当のデータとして公開データ GDS C・CCLE を用いたモデルを構築してきたが、より応用範囲を広げた調査を行った。Pathway Commons(ver12)と TCGA (The Cancer Genome Atlas) の遺伝子発現データと臨床データを利用して巨大なデータへの適用可能性の評価を行った。1, 3, 5 年内の死亡予測タスクを行い、グラフを用いたマルチモーダルな推定が高い性能を示すことが分かった(図 1)。

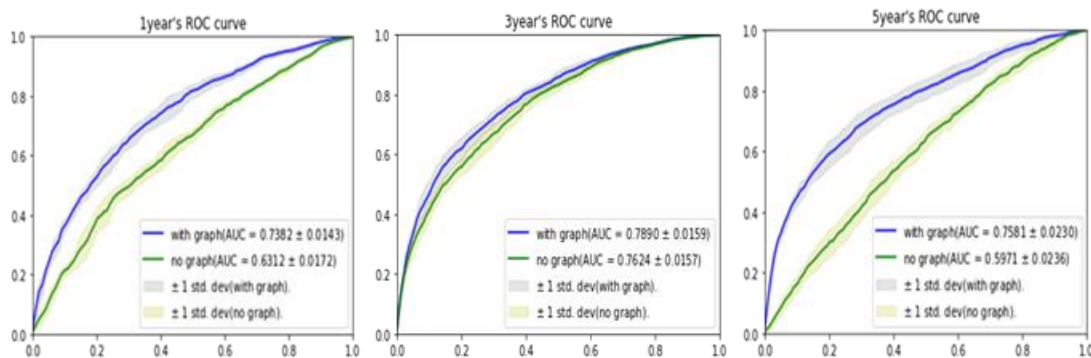


図 1. 1, 3, 5 年内の死亡予測タスクの ROC : 青 : 提案手法であるグラフを用いたマルチモーダルニューラルネットワーク、緑 : グラフを用いないシンプルなマルチモーダルニューラルネットワーク

### ② GCN によるネットワーク構造を考慮した薬剤反応性の識別、バイオマーカー推定、創薬ターゲット分子探索

昨年度までに遺伝子・薬剤・疾患の文献等から構築したグラフから新たな関係を予測する手法と予測結果を説明する部分を可視化する手法を開発し、GCN を用いた手法が他の手法よりも精度よく予測を行うことができることを示した (図 3-1)。本四半期では、前期に行った GCN によるネットワークの予測に関するソフトウェアの公開作業を行い、github にてオープンソースとして公開した

([https://github.com/clinfo/kGCN/tree/master/sample\\_kg/network\\_prediction](https://github.com/clinfo/kGCN/tree/master/sample_kg/network_prediction))。

また、このツールを用いた結果をプレプリントとして公開した (<https://arxiv.org/abs/2104.03871>)。



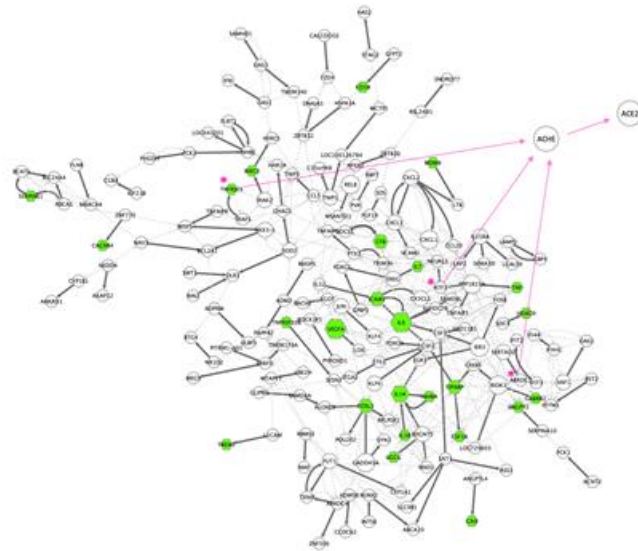
図 3.2 (左図)公開したソフトウェア (右図) 公開したソフトウェアを用いて新規エッジの予測成功率 (Enrichment score)

### ③ ベイジアンネットワークによる分子メカニズムの解明、バイオマーカー推定、創薬ターゲット分子探索

#### (1) SARS-CoV-2/COVID-19 データを用いた遺伝子発現ネットワーク解析法の検証

公開 SARS-CoV-2/COVID-19 データを用いた遺伝子ネットワーク解析法の検証を行い、その論文を出版した。本研究の大半は昨年度中に実施し論文は投稿済みであったがその査読結果が 4 月 5 日に到着した。結果

は査読者への対応が必要というものであったため、本年度の実施内容として査読結果への対応を行い4月26日に修正版を再投稿した。その結果 *Scientific Reports* 誌に受理され、5月27日付で公開された (Tanaka et al., *Scientific Reports* 11, 11241, 2021. DOI: 10.1038/s41598-021-90556-1)。査読者からの質問として、SARS-CoV-2 スパイクタンパクと結合する遺伝子として知られている ACE2 を制御していると思われる遺伝子が何であるかと問われ、同定したサブネットワークを調査したところ直接ではないが、ACHE という遺伝子を介し、同定したサブネットワーク中に含まれる3遺伝子(TNFRSF9, ATF3, ARRDC3)が ACE2 の上流(親の親)として存在することがわかった(図3.1)。このうち ATF3 については、最近になって他の文献でも、ACE2 を制御する遺伝子の候補として報告もあり(Liu X et al. *Front. Physiol.* 11, 540591, 2021)、ネットワーク推定及び解析がうまくいっていることを示した結果と言える。



Supplementary Figure S3: The schematic network illustration of ACE2 involvement in immune defense system of host cells. The network (188 nodes and 420 edges) was generated by the ΔECV-extracted edges (black) shared with A549, H1BE and Calu-3 cells in response to SARS-CoV-2 infection. The nodes (green) represent the known drug target genes. The edges (magenta), and nodes (marked magenta) represent potential regulatory signalings to ACE2 via ACHE. The dotted edges represent the basal edges (gray).

図 3.1 ACE2 上流解析結果 (論文の図 S3 より)

## (2) 遺伝子ネットワーク解析による新規サブタイプ同定法の研究

サンプル毎の遺伝子ネットワーク解析が可能な遺伝子ネットワーク推定法によるがんサブタイプ分類法の研究を行い、論文発表を行った。本研究も前述同様、昨年度からの継続している研究であるが、本年度に投稿を行い、論文出版に必要な査読の対応を行った。

査読での修正の要求としては、TCGA の breast cancer での追加の解析を査読者2名から要求されたため、それを本年度に行った。本論文の新規ポイントは、マルチオミクスのデータで予後差のあるサブタイプの同定が、マルチオミクスデータ対応のクラスタリングアルゴリズムを使ってもできないデータについて、遺伝子ネットワークにより実現可能である点である。Breast cancer データは、元々サブタイプが明確にあり、遺伝子発現データのみで綺麗にそれが分かれることが予想された。追加の解析で、実際に遺伝子発現データでのクラスタリングにより、3タイプに別れ、予後差が見出された(図3.3 b & c)。遺伝子ネットワークを利用した提案手法でもほぼ同等の結果となったが(図3.2 & 図3.3 a)、log-rank test は  $p=0.05$  ちょうどで統計的有意差は棄却された。しかし、生存時間曲線のパターンでは明確に分かれており、ほぼ遺伝子発現データのみと同じ結果であるため、この解析結果を追加し、またその他指摘された細かい修正を行い再投稿し、*Scientific Reports* 誌に受理された。(Nakazawa et al., *Scientific Reports* 11, 23653, 2021. DOI: 10.1038/s41598-021-02394-w)

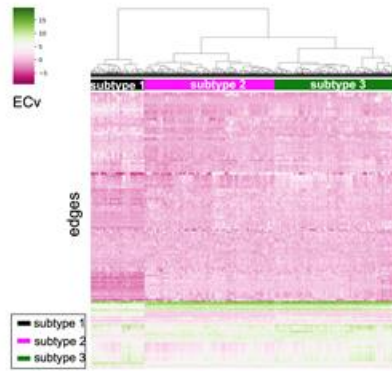


図 3.2 遺伝子ネットワークによる TCGA Breast Cancer データ分類結果

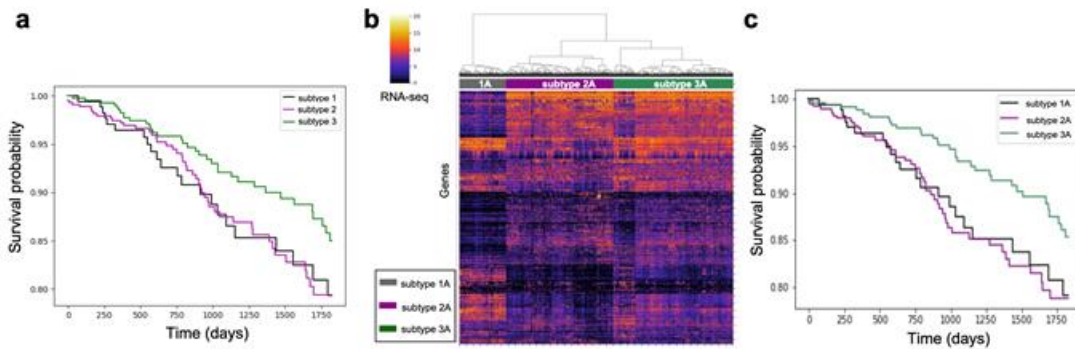


図 3.3 (a) 遺伝子ネットワークによるサブタイプ別生存時間曲線、(b) 発現データのみによる分類結果、(c) 発現データのみによるサブタイプ別生存時間曲線

### (3) MCMC による遺伝子発現介入シミュレーション法の確立

ベイジアンネットワークで構築した遺伝子発現ネットワークから創薬標的分子を探索するためのシミュレーション技術の研究開発を行い、その評価を行った。具体的には本技術は MCMC (マルコフ連鎖モンテカルロ法) と呼ばれるサンプリング技術を用いたシミュレーションの応用である。これを用いると、データから推定したベイジアンネットワークモデルを用いて、そこから発生しうる仮想的な「サンプル」をシミュレーションすることができる。ベイジアンネットワークは確率変数間の構造がネットワークとして陽に表現されることからモデルへの介入・操作が容易にできる、という特徴がある。これを利用することで、仮想的に特定の遺伝子をノックダウンした場合や、薬剤により特定の遺伝子の活性に介入した際の細胞内の遺伝子発現制御システムのモデル (確率構造) をシミュレートすることができると期待している。MCMC は一般的に高次元モデルでも高速にサンプリングが可能である、という特徴があるが、ベイジアンネットワークを用いた遺伝子ネットワークモデルは 20,000 変数からなる確率モデルであり、既存の MCMC をそのまま適用することはできない。またモデルへの介入方法も未確立である。

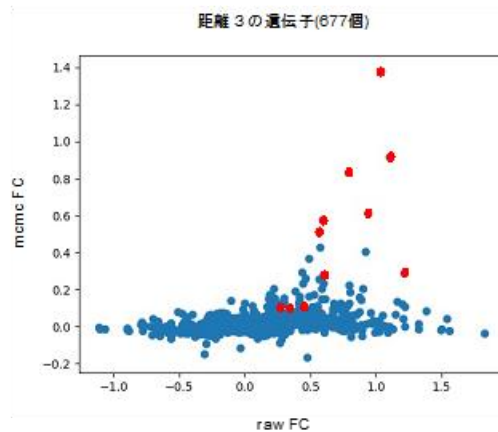


図 3.4 ノックダウン・シミュレーション結果評価。各点がノックダウン (KD) した遺伝子周辺の遺伝子の fold change で横軸が実データ、縦軸がシミュレーションでの値である。赤点は KD 遺伝子直下の遺伝子で、ほぼ全ての直下遺伝子で正しくシミュレート (=予測) できていることがわかる。

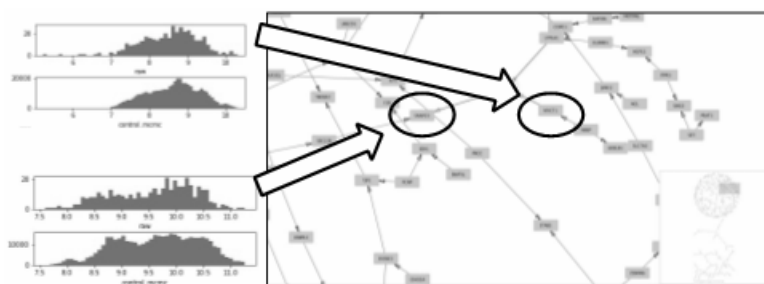


図 3.5 介入なしでのシミュレーションの検証。二つの遺伝子の遺伝子発現のヒストグラムを例として提示している。並べているヒストグラムのうち、上が 400KD の実際の遺伝子発現のヒストグラム、下が MCMC による発現シミュレーション結果。

本研究では、NNSR 法によって推定された 20,000 変数のベイジアンネットワークから DAG (非巡回有向グラフ) をサンプリングすることにより、高速にシミュレーションを行う方法を考案した。また介入シミュレーションの実現方法として、介入点 (遺伝子) 上流のネットワーク上のエッジを切断し、モデルパラメータを再推定し、介入点周辺の DAG をサンプリングすることで介入シミュレーションを実現する方法を考案した。これらの方法の検証を既存の 400 遺伝子をノックダウンした遺伝子発現データセット (400KD) を用いて実施した。

まず介入がない場合の検証を行った。その例を図 3.1 に示す。二峰性があるような複雑な分布も再現できることが確認できた。次にノックダウンのシミュレーションの評価を行った。図 3.2 は開発した技術によるシミュレーション結果と実際のデータとの比較である。もっともノックダウン効率が良かった遺伝子をノックダウンした際の、その周辺の遺伝子の発現量変化をプロットしたもので、それぞれの点が 1 つの遺伝子の実データ (横軸) とシミュレーション (縦軸)、それぞれ場合の fold change である。従って、理想的には  $y = x$  の直前上に点が打たれ、ノックダウン遺伝子の下流直下は特に第一象限内に来ることが望ましい。図中の赤点はベイジアンネットワーク上でノックダウン遺伝子直下 (距離 1) の位置に予測された遺伝子の fold change で、直下遺伝子については全て正しくシミュレーション (=予測) ができていることを表している。

方法論はある程度確立したため現在手動で行っている大半の手順について、これを大規模に自動化して行うためのコーディング作業を外注により実施した。

現在はこれまでに研究成果を論文にまとめている段階である。上述の通りシミュレーション技術はおおむねうまく行っているといえるが、ベースとなる遺伝子ネットワークの能力に強く依存していることもわかり、再来年、条件の検討を進めるとともに、これまでに開発したサブネットワーク同定法などと組み合わせ、同定したサブネットワーク中の遺伝子群に対して網羅的にシミュレーションを行うことにより、創薬標的遺伝子の絞り込みが可能かどうか検証する予定である。

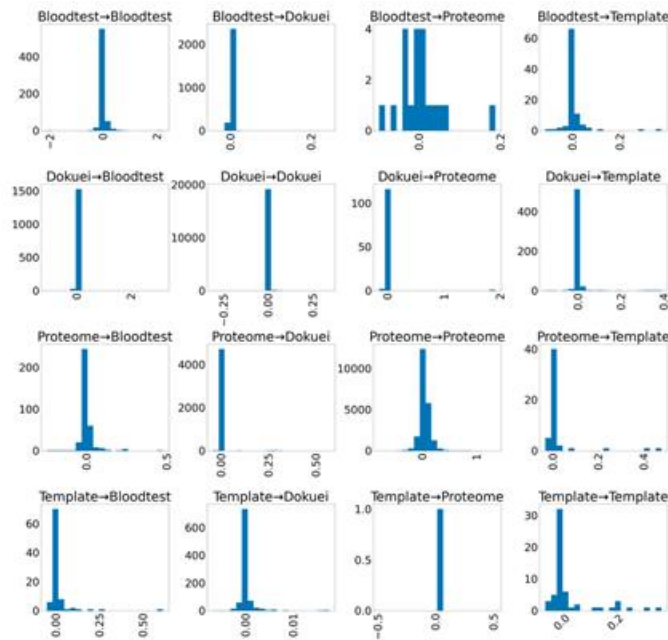


図 3.6 IPF エクソソーム+臨床データから推定したネットワークによる枝の種類別 ECv 分布 (IPF vs control)。

#### (4) 内部データを用いたマルチオミクスネットワーク解析法の研究

マルチオミクスデータを用いた遺伝子ネットワーク解析法の検討のため、内部データを用いて、新たに方法論の検証を始めた。本研究のために新規メンバーを加え必要なトレーニングを行った。内部データは 2 種類あり、1 つは遺伝子発現データでは区別がつかない疾患についての少数の患者由来遺伝子発現データ及びメチル化データ。もう 1 つは 10 患者に対して 5 種類の薬剤を投与した際の増殖率及びゲノム、トランスクリプトーム、プロテーム、メチロームデータである。現在本研究室で確立した遺伝子ネットワーク解析法は遺伝子発現データのみに基づくものであるが、まずは前者のデータを用いてメチル化データを組み合わせる方法の検討を行った。他プロジェクトの内部データのため結果を示せないが、遺伝子ノードとメチル化ノードを合わせてネットワーク推定するという単純な方法でも既存知識と整合している結果が得られており、今後より精緻な方法の検討を行う予定である。後者のデータはゲノムデータ (エクソーム・シーケンズ) の解析から着手し、PCR の特定部位の変異とゲノムデータの一致を確認した段階である。またトランスクリプトーム (遺伝子発現データ) の前処理を完了し、遺伝子発現差解析を行った。現時点では発言差のある遺伝子リストが得られただけの状態であり、今後同定した遺伝子の確認とネットワーク解析を行う。

#### (5) IPF オミクスデータを用いた創薬ターゲット探索

IPF プロテオーム及び臨床データの利用が可能となったためそのデータを用いたネットワーク解析を進めた。まず始めにデータの前処理を行い、プロテームデータの各タンパク、血液検査、臨床データなど全 5125 項目・533 サンプルの入力データセットを完成させ、ネットワーク推定を行った。最終的には IPF 関連サブネットワークをこれまでに開発した個別サンプルごとの枝評価法 ( $\Delta ECv$  法) により抽出することが最初の目標であるが、今回、異種変数を、同時にネットワーク推定を行う、ということを手手法初めて試みているため、まず異種変数間での ECv (Edge Contribution value: 枝貢献量) の分布を確認した (図 3.6)。サンプルのうち、IPF のラベルのついたものは読影所見項目に含まれる疾患ラベルを用いて付与した IPF 疑いを含む 6 サンプルである。 $\Delta ECv$  法は各枝の ECv の差を、閾値を用いて抽出する方法である。遺伝子発現データによる遺伝子ネットワークでは 2-fold change に相当する 1.0 を閾値として使うことが基本である。図の通り、プロテオーム変数間 (Proteome→Proteome) でかなり小さい値になっており、異種変数間で分布が異なることがわかった。 $\Delta ECv$  の最大値は 0.75 であった。したがって、異種変数間の枝ごとに閾値を調整する必要があることがわかった。閾値 0.5 固定で、健常人と IPF ラベル患者間でサブネットワークを抽出したところ 140 変数、

388 枝残り、現在その解析を進めているところである。今後、閾値の変数種ごとの調整や、プロテオームのみのネットワークの構築なども進め、さらなる解析と IPF 重要サブネットワークの同定を行う予定である。

#### D. 考察

##### ① DNN や機械学習による薬剤反応性の識別とバイオマーカー推定

従来の Omics 情報及び臨床情報を用いたマルチモーダルな予測において、さらに知識グラフを加えることは予測において多くの場合で有用に働くことが分かった。一方で、この予測結果がなぜうまく行ったのかなど、より詳細な予測の内訳については、疾患ごとの分析や予測根拠の可視化などさらなる分析が必要である。

##### ② GCN によるネットワーク構造を考慮した薬剤反応性の識別、バイオマーカー推定、創薬ターゲット分子探索

整備したソフトウェアを広く利用してもらいフィードバックを得ることで、より現実的なタスクに対応できるようになり、これらの結果や知見を課題①と共有し、課題①において新たな手法開発へと結びつけることができた。現在は文献情報（知識グラフ）に関しては公開されているものをそのまま利用しているが、今後、使用する知識に関してもより精緻なものやより適したものを検討し、利用していきたい。

##### ③ ベイジアンネットワークによる分子メカニズムの解明、バイオマーカー推定、創薬ターゲット分子探索

今年度は、前年度までに開発したサンプル毎に推定・解析可能なネットワーク推定技術の応用を進め、論文としての発表を行うことができた。またシミュレーション技術の開発を進めた。内部データ及び IPF データの利用準備が整ったため今年度からの新たな挑戦として異種変数を混合させたデータを用いたネットワーク解析技術の研究開発に着手することができた。現時点では解析ソフトウェアの実行が確認でき、初期結果の簡易的な解析で有望な結果が得られた段階であり、検証すべきあるいは確認すべき条件・仮説が多いが、今後それらを進め最終的には利用しているデータによる創薬ターゲット候補探索を実現させたい。

#### E. 結論

##### ① DNN や機械学習による薬剤反応性の識別とバイオマーカー推定

昨年度の検討を元に、文献情報や複数の特徴量を用いてさらに精度の高い予測が可能なモデルの構築に成功した。今後、この手法のより広いアプリケーションへの適用及び、さらに様々なデータへの適用を行いたい。

##### ② GCN によるネットワーク構造を考慮した薬剤反応性の識別、バイオマーカー推定、創薬ターゲット分子探索

昨年度の検討を元に、グラフ関連ツールの整備を進めることができた。特に、ユーザからのフィードバックにより、インターフェースの改良や機能追加へと結びつけることができた。今後も、継続的に改良を続けるとともに、グラフの前処理技術や選択方法など、より効果的なグラフの活用方法のさらなる検討を進めたい。

##### ③ ベイジアンネットワークによる分子メカニズムの解明、バイオマーカー推定、創薬ターゲット分子探索

ベイジアンネットワークを用いたシミュレーション技術の研究開発に一定の成果が得られた。今後サブネットワーク抽出技術と組み合わせ創薬ターゲット候補の自動抽出を試みたい。またトランスクリプトームデータ以外のオミクスデータや臨床データの併用方法の検討も進み、その確立及びこれまでの技術の更なる精緻化を進めたい。

#### F. 研究発表

##### 1. 論文発表

- 1) Harada Y, Sato A, Araki M, Matsumoto S, Isaka Y, Sagae Y, Abe T, Aoyagi Y, Sueoka E, Okuno Y, Kimura S & Sueoka-Aragane N. Integrated approach to functional analysis of an ERBB2 variant of unknown significance detected by a cancer gene panel test, Cellular Oncology, 2022.
- 2) Sato N, Uchino E, Okuno Y. Artificial Intelligence in Kidney Pathology. In: Lidströmer N., Ashrafian H. (eds) Artificial Intelligence in Medicine. Springer, Cham., pp 1-11, 2021.
- 3) Uchino E, Sato N, Okuno Y. Artificial Intelligence in Predicting Kidney Function and Acute Kidney Injury, Artificial Intelligence in Medicine, pp 1-17, 2021/8.
- 4) Terayama K, Sumita M, Katouda M, Tsuda K, Okuno Y. Efficient Search for Energetically Favorable



Molecular Conformations against Metastable States via Gray-Box Optimizatio. Journal of Chemical Theory and Computation, 17:5419-5427, 2021.

- 5) Nakazawa MA, Tamada Y, Tanaka Y, Ikeguchi M, Higashihara K, Okuno Y. Novel cancer subtyping method based on patient-specific gene regulatory network. Scientific Reports, 11:23653, 2021.
- 6) Iwata H, Kojima R, Okuno Y, AIM in Pharmacology and Drug Discovery. Artificial Intelligence in Medicine. 1-9, 2021.
- 7) Ma B, Terayama K, Matsumoto S, Isaka Y, Sasakura Y, Iwata H, Araki M, Okuno Y. Structure-Based de Novo Molecular Generator Combined with Artificial Intelligence and Docking Simulations. Journal of Chemical Information and Modeling. 61(7): 3304-3313, 2021.
- 8) Matsumoto S, Taniguchi-Tamura H, Araki M, Kawamura T, Miyamoto R, Tsuda C, Shima F, Kumasaka T, Okuno Y, Kataoka T. Oncogenic mutations Q61L and Q61H confer active form-like structural features to the inactive state (state 1) conformation of H-Ras protein. Biochemical and Biophysical Research Communications. 565:85-90, 2021.
- 9) Kamada M, Okuno Y. AIM in genomic basis of medicine: applications. Artificial Intelligence in Medicine. 1-10, 2021.
- 10) Sato N, Uchino E, Kojima R, Hiragi S, Yanagita M, Okuno Y. Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. Computer Methods and Programs in Biomedicine. 206:106129, 2021.
- 11) Chiba S, Lim KRQ, Sheri N, Anwar S, Erkut E, Shah MH, Aslesh T, Woo S, Sheikh O, Maruyama R, Takano H, Kunitake K, Duddy W, Okuno Y, Aoki Y, Yokota T. eSkip-Finder: a machine learning-based web application and database to identify the optimal sequences of antisense oligonucleotides for exon skipping. Nucleic Acids Research. gkab442, 2021.
- 12) Tanaka, Y, Higashihara, K, Nakazawa, M.A, Yamashita F, Tamada Y, Okuno Y. Dynamic changes in gene-to-gene regulatory networks in response to SARS-CoV-2 infection. Scientific Reports. 11(1):11241, 2021.
- 13) Nojima S, Terayama K, Shimoura S, Hijiki S, Nonomura N, Morii E, Okuno Y, Fujita K. A deep learning system to diagnose the malignant potential of urothelial carcinoma cells in cytology specimens. Cancer Cytopathology, 2021.
- 14) Nakamura K, Kojima R, Uchino E, Ono K, Yanagita M, Murashita K, Itoh K, Nakaji S, Okuno Y. Health improvement framework for actionable treatment planning using a surrogate Bayesian model. Nature Communications. 12(1):3088, 2021.
- 15) Araki M, Matsumoto S, Bekker G.J, Isaka Y, Sagae Y, Kamiya N, Okuno Y. Exploring ligand binding pathways on proteins using hypersound-accelerated molecular dynamics. Nature Communications. 12(1):2793, 2021.
- 16) Yoshizawa T, Uchibori K, Araki M, Matsumoto S, Ma B, Kanada R, Seto Y, Oh-hara T, Koike S, Ariyasu R, Kitazono S, Ninomiya H, Takeuchi K, Yanagitani N, Takagi S, Kishi K, Fujita N, Okuno Y, Nishio M, Katayama R. Microsecond-timescale MD simulation of EGFR minor mutation predicts the structural flexibility of EGFR kinase core that reflects EGFR inhibitor sensitivity. npj Precision Oncology. 5(1):32, 2021.
- 17) Furuta H, Araki M, Masago K, Sagae Y, Fujita S, Seto K, Shimizu J, Horio Y, Sasaki E, Hosoda W, Katayama R, Okuno Y, Hida T, Novel resistance mechanisms including L1196Q, P1094H, and R1248\_D1249 insertion in three patients with NSCLC after ALK tyrosine kinase inhibitor treatment. Journal of Thoracic Oncology. 16(3):477-482, 2021.

## 2. 学会発表

- 1) “Introduction to HPC- and AI-driven Drug Development Platform Division”, 226th R-CCS Café, online, 2022/2/10
- 2) 「AI・ビッグデータが拓く医療の未来」, 宮城県臨床細胞学会 共催セミナー, オンライン, 2022/2/6
- 3) “Differential privacy under incalculable sensitivity” Mimoto T, Hashimoto M, Yokoyama H, Nakamura T, Isohara T, Kojima R, Hasegawa A and Okuno Y, , 2022 6th International Conference on Cryptography, Security and Privacy (CSP 2022), online, China, 2022/1/15. 【オンライン開催】

- 4) 「Precision Medicineにおけるオミクスデータ×AIの可能性」 トミーデジタルバイオロジー株式会社 ウェビナー, 2022/1/13, WEB 講演
- 5) 「がん分子標的治療における AI・スーパーコンピュータの可能性」, 第25回日本肝がん分子標的治療研究会, オンライン, 2022/ 1/ 8, ホテル日航福岡(WEB 登壇)
- 6) 「AI創薬の現状と未来 Today and future of AI-based drug development」 SEMI テクノロジーシンポ, 東京ビッグサイト (東展示棟・会議棟) 及びオンライン, 2021/12/17
- 7) 「COVID-19 で加速する創薬デジタルトランスフォーメーション」 ” Drug Discovery Digital Transformation Accelerated by COVID-19” , 第3回 ファーマラボ EXPO, 幕張メッセ, 2021/12/10
- 8) 「ウィズコロナ時代の社会と情報科学」, 第13回 総合科学を考えるセミナー, 東北大学/オンラインによる同時開催, 2021/12/4
- 9) 「AI・シミュレーションが拓く医療・創薬の未来」, 山口大学大学院医学系研究科主催第4回シンポジウム「人工知能・システム医学による難治性疾患への新たな挑戦」, 2021/11/27, オンライン
- 10) 特別講演会①「ビッグデータ・AI が拓くヘルスケアの未来」, 次世代ヘルスケアプロジェクト 2021 (日本能率協会), 2021/11/25, 録画配信
- 11) 基調講演「AI・シミュレーションから見る創薬モダリティの New normal—COVID-19 を超えて」, 日本薬物動態学会 第14回ショートコース, 2021/11/16日, Web開催 (Zoom ライブ配信)
- 12) 特別企画セミナー「医療・創薬における AI・データ利活用の現状と可能性」ATR オープンハウス 2021 科学技術が描く明るい未来社会 ～大阪・関西万博に向けて～, 2021/11/11, オンライン開催
- 13) 「COVID-19 から創薬 DX を考える Thinking about DX in drug development from COVID-19」 CBI 学会 2021 年大会, 2021/10/26 オンライン開催
- 14) 「ビッグデータ・AI が拓く医療の未来」 ”Future perspectives in medicine driven by big data and AI technologies”第39回日本骨代謝学会学術集会, 2021/10/10. オンライン開催 事前収録
- 15) 「AIによる薬効・ADMET 予測と企業間連合学習」 生命医薬情報学連合大会(IIBMP 2021), 2021/9/29. オンライン開催
- 16) 「ビッグデータ・人工知能が拓く医療の未来」公益社団法人医療・病院管理研究協会 医療における人工知能の活用と影響 (2021.08.27) オンライン開催
- 17) 「ビッグデータ・AI が拓く創薬・医療の未来」2021年 KEC セミナー 『AI・ビッグデータによる進化のスパイラル』 (2021.07.02) オンライン開催
- 18) 「AI・スーパーコンピュータによるデータ駆動型創薬の可能性」第15回ケミカルバイオロジー (2021.06.22) オンライン開催
- 19) 「AI がもたらす創薬研究へのインパクト」第68回 日本実験動物学会総会 (2021.05.20) オンライン開催
- 20) ”Mutation-induced drug sensitivity prediction for computer-assisted personalized medicine” Araki M and Okuno Y 2021 American Chemical Society Spring National Meeting (April 5, 2021 - April 16, 2021) オンライン開催

G. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

- 1) Life Intelligence Consortium (LINC) 代表