

**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））**  
**総括研究年度終了報告書**

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  
 研究代表者名 : 夏目やよい

国立研究開発法人医薬基盤・健康・栄養研究所 AI健康・医薬研究センター  
 バイオインフォマティクスプロジェクト プロジェクトリーダー

**研究要旨**

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。

令和 4 年度は、肺がん手術検体及びバイオプシー検体のオミックス解析を行うとともに、神奈川県立循環器呼吸器疾患センターで IPF を含む間質性肺炎患者の臨床情報（オミックスデータ及びそれと紐づけられた診療情報）を収集・完了した。本事業で開発された患者層別化 AI の改良版を用い、IPF の特徴に該当する診療情報の項目に紐づけられる肺組織転写産物をデータ駆動的に 84 個見出した。大阪大学コホートの患者層別化 AI 解析でこれまでに見出された IPF 関連タンパク質のパスウェイについて、その活性化をメタボローム解析により確認した。さらに、本事業の成果を広く共有するための基盤であるオープンプラットフォーム「峰」に AI ガジェットを 8 つ追加し、利用申請受付を開始した。

**A. 研究目的**

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・利用して創薬ターゲットを探索する 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ターゲット候補となる生体分子群を自動的に抽出する AI 手法の開発を行う。また、本事業で作成される IPF/肺がんの疾患統合データベース、機能分子を特定するための AI 及び知識ベース等を多くの研究者等に利用してもらうための環境 (オープンプラットフォーム) の構築を目指す。

**B. 研究方法**

**1. 臨床情報の収集と機械可読表現の開発 :**

肺がん手術検体及びバイオプシー検体のオミックス解析を行うとともに、大阪大学医学部呼吸器・免疫内科バイオバンク、大阪大学医学部附属病院医療情報部大阪大学病院バイオバンク及び神奈川県立循環器呼吸器疾患センターで IPF を含む間質性肺炎患者の臨床情報を収集した。

肺がん手術検体・バイオプシー検体のマルチオミックスデータ収集

国立がん研究センターにおいて収集された肺がん手術検体及びバイオプシー検体 (検体採取・処理・保存

は国立がん研究センターにおけるプロトコルに則って行われた)を用いて、下記のオミックスデータを収集した。オミックスデータ取得はタカラバイオ株式会社への外部委託により実施した。

- ・ 全ゲノム解析
- ・ 全エクソーム解析
- ・ DNA メチル化解析
- ・ ヒストン修飾 (ChIP-seq) 解析
- ・ トランスクリプトーム (RNA-seq)

#### 神奈川県立循環器呼吸器病センターコホートにおける臨床情報収集

神奈川県立循環器呼吸器疾患センターにおいて、下記の臨床情報(診療情報+オミックスデータまたは肺組織や血液)を収集した。診療情報及び血液は同意を得られた対象疾患患者全員より取得し、採血のタイミングから最も近い診療記録と紐づけた。肺組織は、診断の上でクライオバイオプシーが必要と判断された場合及び外科手術検体が得られた場合においてのみ収集した。クライオバイオプシーを収集する場合、1患者より数カ所からクライオバイオプシーの採取を行い、縦に半割して一方を病理診断、もう一方をオミックスデータ取得に用いた。採取した肺組織及び血液は、下記のプロトコルで処理・保存し、1ヶ月に1回当所に輸送(三井倉庫ホールディングス株式会社に委託)し、オミックスデータ取得は、ゲノム、RNA-seq 及びメチロームはタカラバイオ株式会社、miRNA-seq は株式会社 DNA チップ研究所、画像特徴量抽出はザイオソフト株式会社に委託した。

#### 収集した診療情報

- ・ 診察記録 (97 項目)
- ・ 所見
  - 超音波 (5 項目)
  - 気管支鏡 (19 項目)
  - 外科生検 (16 項目)
  - CT 画像検査 (19 項目)
- ・ 検査
  - 血液+尿 (103 項目)
  - 生理機能 (35 項目)
  - CT 画像
  - 肺組織
- ・ クライオバイオプシー:採取後チューブに回収し、速やかに-80℃で保存した。
- ・ 外科手術検体 (VATS):採取後チューブに回収し、速やかに-80℃で保存した。
- ・ 上記と紐づけられた病理所見
- ・ 上記と紐づけられた病理画像からの特徴量
- ・ 血液
  - 血清 (プロテオーム解析用):採血後、室温で静かに5回転倒混和し、室温で30分間静置した後、スイング式ローターを用いて、室温、1,300×gで10分遠心した。上清をチューブに250μLずつ分注して速やかに-80℃で保存した。
  - 血漿 (miRNA トランスクリプトーム解析用):採血後、室温で静かに10回転倒混和し、スイング式ローターを用いて、室温、1,300×gで10分遠心した。上清をチューブに250μLずつ分注して速やかに-80℃で保存した。
  - 末梢血単核細胞 (PBMC、全ゲノム解析用):採血後、室温で静かに5回転倒混和し、スイング式ローターを用いて室温、1,600×gで20分遠心した。まとめて採血する場合、最初の検体から最後の検体の時間が1時間以内になるようにした。遠心後、ゲルバリアの上にある細胞層を乱さないようにして血漿層を約半分吸引し、パストゥールピペットを用いて、細胞層全量をチューブに回収して速やかに-80℃で保存した。

#### 肺組織からの核酸 (DNA・RNA) 同時抽出

肺組織サンプル (56 症例、80 検体) について、DNA・RNA 同時抽出キット (NucleoSpin RNA、NucleoSpin RNA/DNA Buffer Set) を用いて核酸抽出を行った。抽出された核酸はバイオアナライザーを用いて RNA 分画における DNA 混入の有無を確認し、RNA 分画は NucleoSpin RNA/DNA Buffer Set のマニュアル通りに DNase 処

理を行なった。

### 肺組織からのマルチオミクスデータ収集

#### i. 全ゲノム解析 (depth 30)

##### (1) TruSeq DNA ライブラリの作成

アコースティックソルビライザーCovaris(コバリス社)を用いて物理的に数百bpに断片化したDNAの両末端を平滑化とリン酸化処理をした後、3'-dA突出末端処理を行い、Index付きアダプターを連結した。その後、アガロースゲル電気泳動によりアダプターを連結したDNAをサイズ選別し、それを鋳型とし、PCRによる増幅を行い、シーケンスライブラリとした。ライブラリの品質はAgilent2100 BioAnalyzerを用いて測定した。

##### (2) シーケンス解析

Illumina社のNovaSeq6000を用いて150base長ペアエンドシーケンスを行い、シーケンサー付属ソフト(NovaSeq Control Software v1.6.、Real Time Analysis (RTA) v3.4.、bcl2fastq2 v2.20)を用いて塩基配列(リード配列)を得た。タグ配列に基づき塩基配列(リード配列)を検体ごとに分類してfastqファイルを取得した。

#### ii. DNAメチル化解析

Illumina社のInfinium MethylationEPIC BeadChipを用いて、Infinium HD Methylation Protocol Guide, Manual Protocol (15019519 v01)に従い、ゲノムDNAを用いてメチル化解析を実施した。

##### (1) Bisulfite 処理

ゲノムDNA(250ng)をEZ DNA Methylation Kit(ZYMO RESEARCH社)を用いてBisulfite変換し、精製、回収を行った。

##### (2) 全ゲノム増幅、断片化、及び精製

Bisulfite処理したDNAはアルカリ処理後に酵素による全ゲノム増幅を行なった。その後、酵素による断片化、イソプロパノール沈殿による精製の後、バッファーに再懸濁した。これを熱変性させ、Infinium MethylationEPIC BeadChipにアプライして、48°Cで約23時間ハイブリダイゼーションを行なった。

##### (3) 一塩基伸長反応とスキャニング

ハイブリダイゼーション後、BeadChipをバッファーで洗浄し、一塩基伸長反応によってプローブ末端に一塩基の標識ヌクレオチドを導入、ハイブリダイズしたDNAを変性・除去し、取り込ませた標識ヌクレオチドに対する蛍光色素標識抗体染色を行い、蛍光イメージを取得した。

##### (4) データ解析

取得した蛍光イメージデータをGenomeStudio/Methylation Moduleを用いて、Background Subtraction及びNormalizaionを実施して解析を行なった。

#### iii. トランスクリプトーム (RNA-seq)

##### (1) TruSeq Stranded mRNA ライブラリの作製

検体よりPolyA+RNAを単離し、断片化により得られるRNAを鋳型とした逆転写反応により一本鎖cDNAを合成した。これを鋳型として、dUTPを取り込ませた二本鎖cDNAを合成した。得られた二本鎖cDNAの両末端を平滑化・リン酸化処理した後、3'-dA突出処理を行い、Index付きアダプターを連結した。アダプターを連結した二本鎖cDNAを鋳型とし、dUTPを持つ鎖を選択的に増幅しないポリメラーゼによりPCR増幅を行い、シーケンスライブラリとした。ライブラリの品質はAgilent2100 BioAnalyzerを用いて測定した。

##### (2) シーケンス解析

Illumina社のNovaSeq6000を用いて150base長ペアエンドシーケンスを行い、シーケンサー付属ソフト(NovaSeq Control Software v1.6.、Real Time Analysis (RTA) v3.4.、bcl2fastq2 v2.20)を用いて塩基配列(リード配列)を得た。タグ配列に基づき塩基配列(リード配列)を検体ごとに分類してfastqファイルを取得した。

## 血液からのマルチオミクスデータ収集

### i. 全ゲノム解析 (depth 30)

肺組織を取得していない症例においては、PBMC (末梢血単核細胞) を用いた全ゲノム解析を実施した。プロトコルは、「肺組織からのマルチオミクスデータ収集 1. 全ゲノム解析」と同一である。

### ii. トランスクリプトーム (miRNA-seq)

miRNA トランスクリプトーム解析には血漿を用いた。RNA 抽出は QIAGEN miRNeasy micro kit を用いた。各サンプルに QIAzol 700 $\mu$ l を添加して RNA 抽出後、14 $\mu$ l の nuclease free water で溶出した。ライブラリ調製キットは QIAseq miRNA Library Kit を使用し、input 量は 5 $\mu$ l とした。こうして調製したライブラリの quality check はバイオアナライザーと qPCR (2nM サンプルを使用) にておこなった。miRNA-seq は Nextseq 75bp シングルエンドで行い、1 サンプルあたり 2000 万リード以上取得することとした。RNA 抽出から miRNA-seq までの一連の作業は株式会社 DNA チップ研究所に委託した。

### iii. プロテオーム解析 (DIA)

血清 200 $\mu$ l から MagCapture isolation kit (Fujifilm Wako) を用いて細胞外小胞を精製した。細胞外小胞中のタンパク質をトリス (2-カルボキシエチル) ホスフィンによる還元・ヨードアセトアミドによるアルキル化を行い、トリプシン消化、脱塩を行った。前処理したサンプルを Data independent acquisition (DIA) 法を用いた LC-MS/MS 解析を実施した。データ解析は、DIA 解析ソフト Spectranout を用いて実施し、欠損値には run-wise imputation を行った。Quality control として、15 検体毎に市販血清 1 検体を加えて、サンプル調製からデータ解析までの品質保証を行った。また質量分析の Quality control として HeLa 細胞消化物の DIA 解析を実施した。

## 2. 統合データ解析の手法開発とそれを用いた新規創薬ターゲット候補の同定：

本事業において理研 AIP との共同研究により開発された患者層別化 AI は、高次元データに対して計算負荷が大きいという欠点があった。しかしオミクスデータは一般的に高次元である傾向があるため、この問題に対処すべくアルゴリズムの改良を行なった (以後、改良版患者層別化 AI と表記)。診療情報の欠損値は、健常者の代表値を用いて補填してから解析に用いた。

## 改良版患者層別化 AI を用いた臨床情報の解析

改良版患者層別化 AI を用いた解析では、紐付けるデータ 2 種類を行列の形で入力する必要がある。そこで上記臨床情報の解析時には、基本患者情報などからなる診療情報とオミクスデータ (当該年度には肺組織 RNA-seq データを使用) の組み合わせで入力し、出力された結果の中から特に IPF の特徴が含まれる患者層別化ルールについて更に分析を進めている。

## バイオインフォマティクス解析による出力結果解釈

上記解析によって見出された生体分子に関して、TargetMine や IPA (Ingenuity Pathway Analysis, QIAGEN 社) を用いた既存知識との照合による分析を行い、関与が疑われるパスウェイの推定や肺線維化に寄与する因子の絞り込みを行なった。

## 血清、尿中代謝物の網羅的測定 (メタボローム)

大阪大学より入手した血清、尿のうち、IPF 確定診断 10 例および健常者 10 例について、キャピラリー電気泳動・フーリエ変換型質量分析計 (CE-FTMS) を用いて、糖リン酸、有機酸、アミノ酸、アミン等のイオン性代謝物を網羅的に測定した。また液体クロマトグラフィー・飛行時間型質量分析計 (LC-TOFMS) を用いて遊離脂肪酸、アシルカルニチン、胆汁酸、ポリフェノール類等の中性代謝物を網羅的に測定した。なお、分析対象物質の物性に応じて、ポジティブモードとネガティブモードの 2 つのメソッドにて実施した。測定データは、検出ピークを標準データベース (約 1300 物質) と照合することにより帰属するとともに、糖代謝、アミノ酸代謝に関わる 110 物質に関しては濃度定量を行った。本測定は、ヒューマン・メタボローム・テクノロジー株式会社 (山形) に委託して実施した。

### 3. オープンプラットフォームの構築：

本事業で作成される機能分子を特定するためのAI及び知識ベース等を多くの研究者等に利用してもらうための環境（オープンプラットフォーム、以後OPF）の構築を進めてきた。

#### OPF システム基盤の構築

特定非営利活動法人 システム・バイオロジー研究機構（SBI）との連携により、OPF システム基盤を構築した（<https://www.nibiohn.go.jp/mine/>）。SBIが開発したプラットフォームであるGaruda

（<http://www.garuda-alliance.org/>）で用いられている技術（SBIが特許保有）により、OSやプログラム言語に依存せずデータベースやAIを相互連結させることができるように設計・構築した。OPFへのアクセスはアカウント発行により制御し、AIは「ガジェット」として逐次追加可能な仕様となっている。

#### （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分担研究機関において倫理審査、承認を得た後、ヒトゲノム・遺伝子解析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. 臨床情報の収集：

肺がん手術検体及びバイオプシー検体のオミックス解析の研究結果全般の詳細については、研究分担者（浜本隆二）による報告書を参照されたい。

神奈川県循環器呼吸器病センターより受理した検体を用いて、下記のマルチオミックスデータを取得した。

肺組織またはPBMCを対象に下記のおミックスデータを取得した。

- ・ 全ゲノム解析（depth 30）：72 症例

肺組織を対象に下記のおミックスデータを取得した。

- ・ DNAメチル化解析：45 症例
- ・ トランスクリプトーム（RNA-seq）：46 症例

血漿を対象に下記のおミックスデータを取得した。

- ・ トランスクリプトーム（miRNA-seq）：247 症例

神奈川県循環器呼吸器病センターの臨床情報収集全般の詳細については、研究分担者（小倉高志）による報告書を参照されたい。

### 2. 統合データ解析の手法開発とそれを用いた新規創薬ターゲット候補の同定：

改良版患者層別化AIの開発により、より高次元のおミックスデータに対しても適用可能となった。神奈川県立循環器呼吸器病センターコホートにおいて得られた肺組織RNA-seqデータ299症例及びそれと紐づけられる診療情報を用いて改良版患者層別化AIを用いた解析を実施したところ、診療情報レベルでIPFの特徴と紐づけられた転写産物が84つ検出された。従来の患者層別化AIではこれほど多くの生体分子を一群にまとめるためには計算負荷が大きく検出が困難であったが、その問題点を改善できていることを確認することができた。さらに、この解析で見出されたIPF関連転写産物の分析を現在進めている。

大阪大学コホートの臨床情報を患者層別化AIの入力データとして用いた解析の結果、患者基本情報を含む診療情報・胸部CT画像の読影所見・血液検査でIPF患者に認められる特徴と紐づけられるタンパク質を複数個見出すことにこれまで成功している。これらのタンパク質が関連するパスウェイの活性状態について検証すべく、メタボローム解析を実施してIPF患者（10例）と健常者（10例）の比較を行なった。その結果、IPFにおいて上記パスウェイが活性化している可能性を支持する結果が得られた。

### 3. オープンプラットフォームの構築：

#### OPF システム基盤の構築

OPF システム基盤を医薬基盤・健康・栄養研究所サーバー上に構築し、下記のAIガジェットを追加搭載し

た。国内・国外を対象に、利用申請の受付を開始した。

これまでに搭載されていた AI ガジェット：

kGCN Network Prediction  
Multiomics\_Analyzer  
SFMEDM  
NamedEntityRecognizer  
EntityLinker  
TargetMine  
SubsetBinder  
molenc  
JaMIE  
Modified\_Diet\_Network  
Vanishing Ranking Kernel

当該年度に追加搭載された AI ガジェット：

RPPA  
Semantic Search  
HeaRT  
PathoGN  
INGOR  
INGOR ECv  
INGOR RC  
INGOR Network

#### D. 考察

当該年度の成果は、分担研究者との共同研究に関連するものを除くと大きく分けて4つである。すなわち、①改良版患者層別化AIの開発、②神奈川県立循環器呼吸器病センターコホートのデータ収集完了、③大阪大学コホートの患者層別化AI解析によって見出されたIPF関連タンパク質が含まれるパスウェイの活性化の検証、④OPFへのAIガジェット追加と利用受付開始である。①の成果より、高次元オミックスデータの解析に対応可能となったことから、本事業で収集した神奈川県立循環器呼吸器病センターコホートデータに対して本AIを適用した解析を実施する。本事業では、IPFを含む間質性肺炎患者のデータを1500症例収集することを最終マイルストーンとしていたが、②の通りその目標を達成し、IPFの創薬標的探索や疾患発症メカニズムの解明に重要な情報源となるデータベースを構築することができた。③の成果は、患者層別化AIによって得られる結果が生物学的に妥当であることを示しており、本事業の解析パイプラインの有用性やこれによって見出された創薬標的候補の有望性を支持するものであるとも言える。そして、本事業の成果を共有するためのプラットフォームの運用開始を達成した。峰を介したデータ共有についてはやむを得ず見送る判断をするに至ったが、本事業後も継続して状況把握と成果共有の実現に努める。

#### E. 結論

当該年度研究計画の達成状況は概ね良好である。2つ目のコホートである神奈川県立循環器呼吸器病センターコホートのデータ収集が完了し、そのデータ解析に用いる改良版患者層別化AIの開発も計画通りに達成した。本事業の成果を広く共有するためのプラットフォーム「峰」が完成し、AIガジェット19つを搭載した状態で運用を開始した。

#### F. 研究発表

##### 1. 論文発表

- 1) Gupta S., Vundavilli H., Allendes Osorio R. S., Itoh M., Mohsen A., Datta A., Mizuguchi K., Tripathi L., Integrative Network Modeling Highlights the Crucial Roles of Rho-GDI Signaling

- Pathway in the Progression of Non-Small Cell Lung Cancer, *IEEE Journal of Biomedical and Health Informatics*, 26(9) :4785-4793, 2022
- 2) Mohsen A., Chen Y., Allendes Osorio R. S., Higuchi C., Mizuguchi K., Snaq: A Dynamic Snakemake Pipeline for Microbiome Data Analysis With QIIME2, *Frontiers in Bioinformatics*, 2, 2022
  - 3) Chen Y., Allendes Osorio R. S., Mizuguchi K., TargetMine 2022: A new vision into drug target analysis, *Bioinformatics*, 38(18): 4454-4456, 2022
  - 4) Hosomi K., Saito M., Park J., Murakami H., Shibata N., Ando M., Nagatake T., Konishi K., Ohno H., Tanisawa K., Mohsen A., Chen Y., Kawashima H., Natsume-Kitatani Y., Oka Y., Shimizu H., Furuta M., Tojima Y., Sawane K., Saika A., Kondo S., Yonejima Y., Takeyama H., Matsutani A., Mizuguchi K., Miyachi M., Kunisawa J., Oral administration of *Blautia wexlerae* ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota, *Nature Communications*, 13(1):4477, 2022
  - 5) Futami Y., Takeda Y., Koba T., Narumi R., Nojima Y., Ito M., Nakayama M., Ishida M., Yoshimura H., Naito Y., Fukushima K., Takimoto T., Eda Hiro R., Matsuki T., Nojima S., Hirata H., Koyama S., Iwahori K., Nagatomo I., Shirai Y., Suga Y., Satoh S., Futami S., Miyake K., Shiroyama T., Inoue Y., Adachi J., Tomonaga T., Ueda K., Kumanogoh A. Identification of CD14 and lipopolysaccharide-binding protein as novel biomarkers for sarcoidosis using proteomics of serum extracellular vesicles, *International Immunology*, 34 (6): 327-340, 2022
  - 6) Hosoe Y., Miyanoiri Y., Re S., Ochi S., Asahina Y., Kawakami T., Kuroda M., Mizuguchi K., Oda M. Structural dynamics of the N-terminal SH2 domain of PI3K in its free and CD28-bound states, *The FEBS Journal*, Oct. 25, 2022
  - 7) Otaki M., Hirane N., Natsume-Kitatani Y., Nogami Itoh M., Shindo M., Kurebayashi Y., Nishimura S. Mouse tissue glycome atlas 2022 highlights inter-organ variation in major N-glycan profiles, *Scientific Reports*, 12(1):17804, 2022
  - 8) Kawasaki T., Takeda Y., Eda Hiro R., Shirai Y., Nogami-Itoh M., Matsuki T., Kida H., Enomoto T., Hara R., Noda Y., Adachi Y., Niitsu T., Amiya S., Yamaguchi Y., Murakami T., Kato Y., Morita T., Yoshimura H., Yamamoto M., Nakatsubo D., Miyake K., Shiroyama T., Hirata H., Adachi J., Okada Y., Kumanogoh A. Next-generation proteomics of serum extracellular vesicles combined with single-cell RNA sequencing identifies MACROH2A1 associated with refractory COVID-19, *Inflammation and Regeneration*, 42(1):53, 2022
  - 9) Hioki K., Hayashi T., Natsume-Kitatani Y., Kobiyama K., Temizoz B., Negishi H., Kawakami H., Fuchino H., Kuroda E., Coban C., Kawahara N, m Ishii H. J., Machine-learning-assisted screening of herbal medicine extracts as vaccine adjuvants, *Frontiers in Immunology*, 13, 2022
  - 10) Watanabe R., Kawata T., Ueda S., Shinbo T., Higashimori M., Natsume-Kitatani Y., Mizuguchi K., Prediction of the Contribution Ratio of a Target Metabolic Enzyme to Clearance from Chemical Structure Information, *Molecular Pharmaceutics*, 20(1): 419-426, 2023
  - 11) Sohrab, M. G., Duong, K. N., Masami, I., Topić, G., Natsume-Kitatani, Y., Kuroda, M., Nogami-Itoh M., Takamura, H. (2022, November). BiomedCurator: Data Curation for Biomedical Literature, In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing: System Demonstrations*, PP 63-71, 2022
  - 12) 第1章3節 生命情報科学からのAI創薬  
夏目やよい、水口賢司  
革新的AI創薬 ～医療ビッグデータ、人工知能をもたらす創薬研究の未来像～ 2022年7月発刊  
エヌティーエス, ISBN 978-4-86043-788-6, 2022.7
  - 13) 第3章1節 新薬創出を加速する人工知能の開発 -臨床情報を活用した創薬標的探索  
夏目やよい  
革新的AI創薬 ～医療ビッグデータ、人工知能をもたらす創薬研究の未来像～ 2022年7月発刊  
エヌティーエス, ISBN 978-4-86043-788-6, 2022.7
  - 14) 第5章7節 サブセット・バインディングによる患者層別化AIの開発  
上田修功、夏目やよい  
革新的AI創薬 ～医療ビッグデータ、人工知能をもたらす創薬研究の未来像～ 2022年7月発刊  
エヌティーエス, ISBN 978-4-86043-788-6, 2022.7

- 15) 中村恵宣, 北村英也, 小倉 高志, 夏目やよい, 水口賢司, 官民研究開発投資拡大プログラム (PRISM) で構築する特発性肺線維症に対する創薬標的探索プラットフォームについて, MEDCHEM NEWS, 32 巻 3 号, 2022. 8. 1
- 16) 夏目やよい, 機械学習によって加速される次世代アジュバント開発, 別冊「医学のあゆみ」ワクチン設計のサイエンス, 2022. 10. 22
- 17) ロドルフォアジェンデスオソリオ, 夏目やよい, 機械学習を用いたアジュバント開発の新潮流, ・月間フラインケミカル 12 月号, 2022. 12. 15
- 18) 長尾知生子, 鎌田真由美, 中津井雅彦, 深川明子, 片山俊明, 川島秀一, 水口賢司, 安倍理加, 医薬品関連文書の利活用に向けたインタビューフォームの構造化の提案, 医薬品情報学, 24 巻 4 号, 2023 年

## 2. 学会発表

- 1) 夏目やよい, 診察情報とオミックスデータを用いたデータ駆動的な患者層別化と創薬標的探索 ～特発性肺線維症を例として～, 第 4 回メディカル AI 学会学術集会, 仙台, 2022. 6. 10
- 2) 夏目やよい, 臨床情報を用いたデータ駆動的な創薬標的探索 FRONTIERS OF ENGINEERING JAPAN, 長崎, 2022. 6. 24
- 3) 夏目やよい, 診療情報とオミックスデータを用いたデータ駆動的創薬標的探索 ～特発性肺線維症に対する挑戦～, 第 49 回日本毒性学会学術年会, 札幌, 2022. 7. 2
- 4) 夏目やよい, 新薬創出を加速させる AI の開発, 第 81 回日本癌学会学術総会, 横浜, 2022. 9. 29
- 5) 夏目やよい, 「診療情報と生体分子の測定データから薬の標的を探し出すために」 ABCI User Group ウェビナー, Web 開催, 2022. 10. 14
- 6) 夏目やよい, CBI 学会『データサイエンス/バイオインフォマティクス』座長, CBI 学会 2022 年大会, 2022. 10. 25, 東京
- 7) 夏目やよい, 診療情報の利活用とデータ駆動的創薬標的探索, MD-DSC 研究会, Web 開催, 2022. 11. 4
- 8) 夏目やよい, 臨床情報の活用によるデータ駆動的な創薬標的探索, 医療・創薬・ヘルスケアにおける DX -LINC の紹介と兵庫県、神戸市企業との連携に向けて-, 神戸, 2022. 11. 14
- 9) 夏目やよい, 診療情報とオミックスデータを用いたデータ駆動的な患者層別化と創薬標的探索, 第 6 回医学 AI セミナー特別レクチャー, 仙台, 2022. 11. 17
- 10) 夏目やよい, トランスクリプトームデータを用いた毒性発現メカニズムの推定, 第 5 回医薬品毒性機序研究会, 東京, 2022. 12. 9
- 11) 夏目やよい, 次世代創薬を支えるデータ駆動型研究の最前線, 彩都産学官連携フォーラム 2023, 大阪, 2023. 1. 18
- 12) 夏目やよい, 診療情報とオミックスデータを紐付けるクリニカルトランスオミクス DX, 創薬薬理フォーラム, 東京, 2023. 1. 20
- 13) 夏目やよい, 「診療情報とオミックスデータを用いたデータ駆動的な患者層別化と創薬標的探索」, 東京女子医大女性医療人キャリア形成センター, Web 開催, 2023. 2. 13
- 14) 夏目やよい, R&D をデータ駆動型で支援する「創薬ターゲット探索プラットフォーム」の価値提供, IQVIA ジャパンメディアセミナー, Web 開催, 2023. 2. 14
- 15) 夏目やよい, データ駆動的な患者層別化と創薬標的探索, 徳島大学先端酵素学研究所 2022 年度共同利用・共同研究拠点成果報告会, 徳島, 2023. 3. 10
- 16) 夏目やよい, 「PRISM 創薬 AI・IPF プロジェクトの成果」、オープンプラットフォーム<峰> 「新薬創出を加速する症例データベースの構築・拡充/創薬ターゲットの推定アルゴリズムの開発」, 令和 4 年度成果報告会, 東京, 2023. 3. 13
- 17) 夏目やよい, 異種データを紐付ける層別化アルゴリズム subset binding を用いたデータ駆動的な創薬標的探索, 日本薬学会第 143 年会, 札幌, 2023. 3. 26
- 18) 原侘奈, 武田吉人, 足立淳, 夏目やよい, 伊藤眞里, 熊ノ郷淳, エクソソームの次世代プロテオミクスによる線維化性過敏性肺炎の新規バイオマーカー開発, 第 62 回日本呼吸器学会学術講演会, 京都, 2022. 4. 22
- 19) 夏目やよい, 伊藤眞里, 松村泰志, 武田吉人, 足立淳, 熊ノ郷淳, 水口賢司, 上田修功, 患者層別化 AI による特発性肺線維症の創薬標的提示に向けて, 第 62 回日本呼吸器学会学術講演会, 京都, 2022. 4. 23
- 20) 榎本貴俊, 武田吉人, 網屋沙織, 足立雄一, 新津敬之, 原侘奈, 野田成美, 白井雄也, 足立淳, 伊藤眞里, 夏目やよい, 熊ノ郷淳, 次世代プロテオミクスにより進行性線維化を伴う間質性肺疾患の新規バイオマーカー探索 (PRISM), 第 62 回日本呼吸器学会学術講演会, 京都, 2022. 4. 23



- 21) 白井雄也, 武田吉人, 榎本貴俊, 足立雄一, 網屋沙織, 野田成美, 菅泰彦, 白山敬之, 三宅浩太郎, 平田陽彦, 足立淳, 夏目やよい, 伊藤眞里, 熊ノ郷淳, 「PRISM」データから見えてきた新たな線維化バイオマーカー“PRISM”data reveal fibrotic biomarkers, 第62回日本呼吸器学会学術講演会, 京都, 2022. 4. 23
- 22) 網屋沙織, 武田吉人, 榎本貴俊, 白井雄也, 白山敬之, 三宅浩太郎, 平田陽彦, 足立淳, 伊藤眞里, 夏目やよい, 熊ノ郷淳, エクソソームの次世代プロテオミクスによるサルコイドーシスの新規バイオマーカー探索 (PRISM) 第62回日本呼吸器学会学術講演会, 京都, 2022. 4. 23
- 23) 吉村華子, 武田吉人, 足立淳, 伊藤眞里, 榎本貴俊, 原伶奈, 白井雄也, 三宅浩太郎, 白山敬之, 平田陽彦, 川崎孝裕, 熊ノ郷淳, 次世代プロテオミクスによる気管支喘息 T2 炎症の新規 BM 開発, 第62回日本呼吸器学会学術講演会, 京都, 2022. 4. 23
- 24) 安部祐子, 武田吉人, 福島清春, 白山敬之, 平田陽彦, 足立淳, 伊藤眞里, 横井崇, 南俊行, 栗林康造, 木島貴志, 熊ノ郷淳, エクソソームの次世代プロテオミクスによる悪性胸膜中脾腫の新規バイオマーカー探索, 第62回日本呼吸器学会学術講演会, 京都, 2022. 4. 24
- 25) 夏目やよい, 診察情報とオミックスデータを用いたデータ駆動的な患者層別化と創薬標的探索, 「新薬創出を加速する症例データベースの構築・拡充/創薬ターゲットの推定アルゴリズムの開発」令和3年度成果報告会, Web 開催, 2022. 5. 12
- 26) 夏目やよい, 伊藤眞里, 松村泰志, 武田吉人, 足立淳, 熊ノ郷淳, 水口賢司, 上田修功, 診療情報と紐付けられたプロテオームデータによる特発性肺線維症を対象としたデータ駆動的創薬標的探索, 第70回質量分析総合討論会, 福岡, 2022. 6. 22
- 27) 夏目やよい, 診療情報とオミックスデータを用いたデータ駆動的創薬標的探索, 第1回 NIBIOHN 定例研究発表会, 大阪, 2022. 7. 5
- 28) 吉村華子, 武田吉人, 榎本貴俊, 原伶奈, 菅泰彦, 川崎貴裕, 三宅浩太郎, 白山敬之, 平田陽彦, 小山正平, 長友泉, 岩堀幸太, 足立淳, 伊藤眞里, 夏目やよい, 熊ノ郷淳, 次世代プロテオミクスによる気管支喘息 T2 炎症の新規 BM 開発, 第43回日本炎症・再生医学会, 淡路島, 2022. 7. 6
- 29) 夏目やよい, 新薬創出を加速する人工知能の開発, 第5回さきがけ終了領域研究会, 千葉, 2022. 8. 25
- 30) 伊藤眞里, 足立淳, 武田吉人, 黒田正孝, 朝長毅, 武田理宏, 松村泰志, 夏目やよい, 水口賢司, 熊ノ郷淳 「新薬創出を加速する人工知能の開発」特発性肺線維症患者血清中エクソソーム内プロテオームと診療情報を用いるデータ駆動的創薬標的の探索, 第9回日本細胞外小胞学会学術集会, 東京, 2022. 10. 25
- 31) Itoh-Nogami M., Natsume-Kitatani Y., Kuroda M., Adachi J., Chen Y., Mizuguchi K., Takeda Y., Kumanogoh A., Ueda N., Artificial Intelligence to Accelerate New Drug Discovery: Target identification and drug discovery by data-driven approach and experimental validation in Idiopathic Pulmonary Fibrosis (IPF), CBI 学会 2022 年大会, 東京, 2022. 10. 25 (Like! Poster Award 受賞)
- 32) Mohsen A., Chen Y-A., Allendes Osorio R. S., Higuchi C., Mizuguchi K., Optimizing and automating QIIME2 Microbiome data analysis using the Snakemake pipeline (Snaq), IHMC2022, 神戸, 2022. 11. 8
- 33) 渡邊怜子, 河田敏生, 上田真也, 新保拓未, 東森光雄, 夏目やよい, 水口賢司, 化学構造情報を用いた代謝酵素のクリアランスへの寄与率予測モデルの構築, 日本薬物動態学会第37回年会, 横浜, 2022. 11. 8
- 34) 樋口千洋, 黒田正孝, 伊藤眞里, 長尾知生子, 水口賢司, 夏目やよい, PRISM 成果利用システム「峰」の構築, 第42回医療情報学連合大会 第23回日本医療情報学会学術大会, 北海道, 2022. 11. 20
- 35) 黒田正孝, 伊藤眞里, 夏目やよい, 鎌田英世, 深川明子, 武田吉人, 武田理宏, 松村泰志, 北村 英也, 丹羽崇, 岩澤多恵, 高橋陽子, 荒牧英治, 黒橋禎夫, 水口賢司, 小倉 高志, 熊ノ郷淳, 臨床データの匿名加工情報作成に向けた検討-新薬創出を加速する人工知能の開発におけるデータ利用促進に向けて-, 第42回 医療情報学連合大会 第23回日本医療情報学会学術大会, 北海道, 2022. 11. 20
- 36) 伊藤眞里, 特発性肺線維症臨床情報収集と創薬標的探索手法の構築, 第42回 医療情報学連合大会 第23回日本医療情報学会学術大会, 札幌, 2022. 11. 20
- 37) 伊藤眞里, デジタルヘルスデータ基盤構築にむけて, LINC SHOWCASE 2022Autumn, 大阪, 2022. 11. 21
- 38) 藤原 大, 陳 怡安, 夏目 やよい, 水口 賢司, 統合創薬研究データウェアハウスによる疾患標的探索モデルの構築, 日本薬学会第143年会, 札幌, 2023. 3. 27
- 39) Natsume-Kitatani Y., Inference of mechanisms of toxicity from omics data, The Open Tox 2022 Virtual Conference, Web 開催, 2022. 9. 12

- 40) Natsume-Kitatani Y, Itoh MN, Takeda Y, Kuroda M, Hirata H, Miyake K, Shiroyama T, Shirai Y, Noda Y, Adachi Y, Enomoto T, Amiya S, Adachi J, Narumi R, Muraoka S, Tomonaga T, Kurohashi S, Cheng F, Tanaka R, Yada S, Aramaki E, Wakamiya S, Chen YA, Higuchi C, Nojima Y, Fujiwara T, Nagao C, Takeda T, Matsumura Y, Mizuguchi K, Kumanogoh A, Ueda N. Data-driven patient stratification and drug target discovery by detecting paired itemsets from medical information and omics data, ISMB2022, アメリカ, 2022.7.11~12
- 41) Natsume-Kitatani Y., Aisaki K., Kitajima S., Kanno J., Comparative study of dynamic changes in gene expression profiles induced by PPARα ligands, ECCB2022, スペイン、バルセロナ, 2022.9.18
- 42) Allendes Osorio R. S., Nystroem-Persson J.T., Kosugi Y., Mizuguchi K., Natsume-Kitatani Y., Panomicon, allowing heterogeneous multi-omics analysis on the web ECCB2022, スペイン、バルセロナ, 2022.9.18

G. 知的財産権の出願・登録状況（予定を含む。）

1. 特許取得

国際出願番号：PCT/JP2023/11960

発明の名称：特発性肺線維症の治療または予防剤

発明者：夏目やよい、伊藤真里、黒田正孝、水口賢司、足立淳、朝長毅、熊ノ郷淳、武田吉人、上田修功

出願人：国立研究開発法人医薬基盤・健康・栄養研究所

国立大学法人大阪大学

国立研究開発法人理化学研究所

出願日：2022年3月25日

出願番号：特願 2023-58620

発明の名称：オミックスデータとそのメタデータの格納と高速検索を可能にするデータベース形式

発明者：夏目やよい、水口賢司、

出願人：国立研究開発法人医薬基盤・健康・栄養研究所

出願日：2023年3月31日

2. 実用新案登録

なし

3. その他

なし