

特集「新型コロナウイルス治療薬の研究を推進」より

02 特集

新型コロナウイルス治療薬の研究を推進

12 原酒

「理研史の生き字引」を目指して

05 特集

**AIを活用して一人一人に
健康な生活を提供する**

08 研究最前線

がんの未知なる特徴をAIが発見

理事長からの メッセージ



われわれ人類は今、新型コロナウイルスという見えない敵と対峙し、生存を懸けて戦っています。理化学研究所はその叡智を結集して新型コロナウイルスを克服する術を人類にもたらしべく、特別プロジェクトを立ち上げました。

およそ百年前、わが国の繁栄の礎となるべく設立された理研は、自然科学におけるあらゆる分野で研究成果を積み重ねてきました。巨大な生命科学系プロジェクトを担うことで蓄積してきた、免疫学・遺伝学・構造生物学をはじめとした「知見」と、近年急速に発展している計算科学やAI、さらにそれを活かしたAI創薬などの多彩な「技術」を理研は有しています。これらを総動員し、本プロジェクトに注力する決意です。

人類はこれまで、科学技術を駆使して不可能を可能とし、多くの困難を乗り越えてきました。人類生存の危機に瀕した今、まさに科学技術の真価が問われています。理研は、国内外の研究機関や大学、企業とも力を合わせ、新型コロナウイルスの克服に貢献します。

2020年4月21日

理化学研究所 理事長 松本 紘

新型コロナウイルス感染症 (COVID-19) が人類の脅威となっている。

創薬・医療技術基盤プログラム (DMP) は2020年3月、

新型コロナウイルス感染症の治療薬の開発に向けた特別プロジェクトを開始した。

理研は、この人類共通の緊急課題の克服にどのように貢献するのか。COVID-19特別プロジェクトの概要を紹介する。

新型コロナウイルス治療薬の研究を推進

■ 治療薬の標的タンパク質

創薬・医療技術基盤プログラム (DMP) は2010年の設置以来、国内の創薬・医療技術シーズの探索、理研の創薬基盤技術の開発、創薬経験者から成るマネジメントチームの三つの機能を柱に、医薬品や医療技術への橋渡しを目指して研究プロジェクトを展開している。特に、進行性骨化性線維異形成症など製薬企業が取り組みにくい希少疾患・難病に対する創薬、iPS細胞による網膜再生や人工アジュバントベクター細胞 (aAVC) によるがん免疫療法などの実用化に力を入れてきた。

「今般、DMPは、10年間かけて整備してきた“アカデミア発創薬”の体制を活かし、国内外の研究機関と連携して新型コロナウイルス感染症の治療薬の研究開発を全力で加速させるため、COVID-19特別プロジェクトを開始します」。DMPを率いる後藤俊男プログラムディレクター (PD) はそう決意を語る。

ウイルスがヒト細胞に侵入すると、ウイルスを構成するタンパク質が宿主のヒト細胞上の受容体と接触し、酵素など特定のタンパク質が働いて増殖する。それらのタンパク質に結合して働きを阻害するものが治療薬となる。まず、新型コロナウイルスが侵入・増殖する過程で、タンパク質がどのように働いているかを簡単に説明しよう。

①ウイルス表面には「スパイクタンパク質」と呼ばれる突起があり、その突起でヒト細胞表面の受容体「ACE2」に結合して前駆体をつくる。その前駆体はヒトが持つ分解酵素「TMPRSS2」が働くことで形を変えて、ウイルスの侵入が始まる。

②細胞内に侵入したウイルスは自らのRNAの遺伝情報に従って、長いタンパク質 (ポリタンパク質) をつくる。ウイルス由来の「メインプロテアーゼ」と「パパイン様プロテアーゼ」やヒトが持つ分解酵素が、ポリタンパク質を十数種類の断片に切断し、それらが新しいウイルスを形づくるタンパク質や増殖に必要な機能を果たす酵素となる。

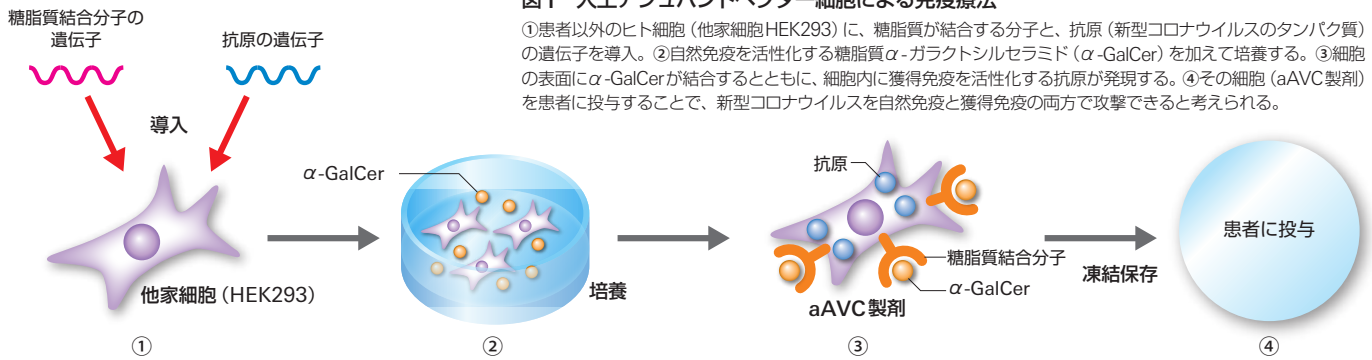
③ウイルスのRNAを「RNA依存性RNAポリメラーゼ」という酵素を用いて複写する。

④細胞内で新たにつくられたタンパク質とRNAが集まって新しいウイルスが大量にできて細胞外へ放出され、別のヒト細胞に次々と侵入していく。このように増殖していくことで、COVID-19の症状が表れる。

特別プロジェクトでは、4月時点でウイルス侵入・増殖過程で働く「スパイクタンパク質」「TMPRSS2」「メインプロテアーゼ」「パパイン様プロテアーゼ」「RNA依存性RNAポリメラーゼ」に加えて、ウイルス表面にある「エンベロープタンパク質」を治療薬の主な標的タンパク質として想定し、研究を開始した。

■ 既存薬から探索し、いち早く臨床に役立てる

「理研ではすでに、数々のライフサイエンス系プロジェクトにより培った標的タンパク質の構造解析、分子シミュレーションやAI (人工知能) など、医薬品設計において世界でも最先端の技術を確立しています。それらを特別プロジェクトに活用することで、治療薬の研究開発を国際協力で加速させていきます」と後藤PD。



標的タンパク質の働きを阻害する治療薬の開発戦略は短期と中長期の2本立てだ。

「短期」は2020年から3年以内。ほかの疾患の治療薬として現在販売されている薬や臨床試験薬が、COVID-19に効果があるかどうかを調べる。すでに体内での動態や副作用の表れ方が分かっているものなら、いち早くCOVID-19の臨床に役立てることができる。

そのためにまず、8,600点にも及ぶ既存薬や臨床試験薬およびそれらに形が似た周辺化合物について、標的タンパク質と結合してタンパク質の機能に作用を及ぼすものがあるかどうか、コンピュータによるシミュレーションで探索する「インシリコスクリーニング」を始めている。そこで探し出した化合物について試験管内で標的タンパク質と作用させる実験を行う。さらに外部機関とも連携して、培養細胞やラットなどを使った実験を進め、治療薬として有望なものを絞り込む。選び出した既存薬や臨床試験薬、周辺化合物について、開発元の製薬企業や医薬品医療機器総合機構と連携し、臨床試験へと開発を進めていく。

■ 変異体にも効果がある新しい治療薬を！

3～5年先を見据えた「中長期」では、新しい治療薬候補の開発を進める。新型コロナウイルスの感染者数がいったん減ったとしても、ウイルスが変異して、インフルエンザのように毎年あるいは数年置きに流行する恐れがある。そこで、変異体にも効果を発揮し、有効性・安全性の高い治療薬が求められている。

2003年に流行したSARSもコロナウイルスの一種だ。SARSと新型コロナウイルスが持つタンパク質を比較すると、メインプロテアーゼとRNA依存性RNAポリメラーゼは、タンパク質

を構成するアミノ酸配列が95%一致し、変異しにくいと考えられる。また、分解酵素TMPRSS2は宿主であるヒトのタンパク質でありウイルス変異の影響を受けにくい。このようなタンパク質を標的に、重要な機能を持ち、かつ形が変わりにくい部位に強く結合して作用する化合物を分子シミュレーションやAIを駆使して設計することで、変異体にも効果を持つ治療薬を開発できるだろう。

■ 独自の免疫療法や抗体医薬の開発を進める

「理研は免疫による疾患制御でも世界最先端の技術を持ち、DMPは人工アジュバントベクター細胞（aAVC）によるがん免疫療法の実用化にも力を入れてきました。その技術を新型コロナウイルスに応用できないかと考えています」と後藤PD。

免疫システムには、ウイルスやがん細胞などのあらゆる異物に素早く反応する「自然免疫」と、特定の異物（抗原）を認識して攻撃する「獲得免疫」がある。aAVCは、自然免疫を活性化する糖脂質と、獲得免疫を活性化する抗原を搭載した細胞だ。これまでは、がん免疫療法用のaAVCの開発が進められてきたが、ウイルス表面にあるスパイクタンパク質やエンベロープタンパク質などを抗原として搭載したaAVCを患者に投与することで、新型コロナウイルスを自然免疫と獲得免疫の両方に攻撃させることも可能だと考えられる（図1）。

さらに抗体医薬の開発も進める。免疫系は抗原に対し、生体内で抗体というタンパク質をつくり異物を攻撃するが、その抗体を体外でつくり投与するのだ。標的としてスパイクタンパク質やTMPRSS2が考えられる。スパイクタンパク質の抗体医薬はさまざまな研究機関が開発を進めているが、理研の分子シミュレーションやAIの技術を駆使して抗体の最適化設計を目指す。

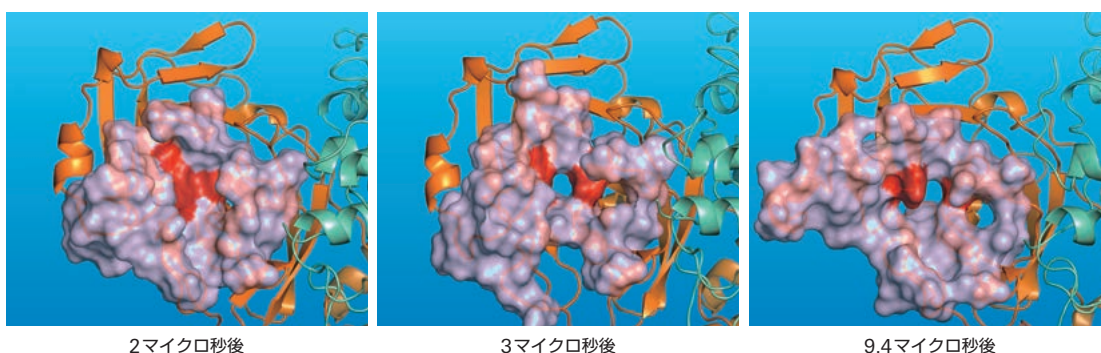


図2 新型コロナウイルスのメインプロテアーゼの分子動力学シミュレーション
メインプロテアーゼの酵素機能に重要な活性中心の拡大図。活性中心の形が計算開始から時間とともに揺らいていることが分かる。本誌表紙の画像は10マイクロ秒後のメインプロテアーゼ全体像。

新型コロナウイルス対策のためスパコン「富岳」を試行利用

2020年4月7日、国難ともいえる新型コロナウイルス感染症（COVID-19）への対策に貢献する成果をいち早く創出するため、スーパーコンピュータ「富岳」の一部ノードを優先して供出することを決定した。

理研が主体となって開発・整備中の「富岳」は、2019年12月より順次搬入を進め、現在は2021年度の共用開始を目指した開発・整備の途上にある。世界最先端・最高アプリケーション性能を目指して開発されているわが国のフラッグシップスパコンであり、その最重要ミッションの一つが、国民の

安全・安心を強大な計算の力で守ることだ。世界中で甚大な被害をもたらしているCOVID-19の診断・治療・感染拡大防止などに奮闘する科学に対し、稼働準備中とはいえ十分な即戦力となり得る「富岳」の能力を大幅に前倒して速やかに提供することで、一日も早いパンデミックの終結に貢献する。

理研が窓口となって研究課題を募集し、被害軽減への貢献が期待でき、かつ、速やかに着手できる研究開発に焦点を絞って文部科学省が決定し、実施していく。具体的には、治療薬候補同定、ウイルス表面のタンパク

質動的構造予測、パンデミック現象および対策のシミュレーション解析などから始動。実施課題で得られた成果は、関係機関などとも連携しつつ、国内外に広く公開する。



スーパーコンピュータ「富岳」（開発・整備中）

ウイルスのタンパク質を標的とする治療薬は、そのタンパク質が変異すると効かなくなる恐れがある。その点、ウイルス侵入で働くTMPRSS2はヒトが持つタンパク質なので変異することはない、その阻害剤は新型コロナウイルスの変異体にも効果を持つ可能性がある。すでにナファモスタットという急性肺炎などの治療薬がTMPRSS2を阻害して新型コロナウイルスの侵入を阻止することが報告されている。しかしTMPRSS2以外の十数種類の分解酵素も阻害するため副作用の懸念がある。TMPRSS2だけに結合して阻害できる抗体が開発できれば、より安全で強い薬効を持つ治療薬となり得ると考えられる。

■ ウイルス増殖に必須なタンパク質のシミュレーションを公開

特別プロジェクトに参加する生命機能科学研究センター（BDR）計算分子設計研究チームは、2019年に完成したばかり

の分子動力学シミュレーション専用計算機「MDGRAPE-4A」を用いて新型コロナウイルスのメインプロテアーゼの計算を行い、2020年3月17日にデータを公開した。

細胞内は水分で満たされており、その中でタンパク質は常に形を変化させながら揺らいでいる。分子動力学シミュレーションは、細胞内に近い状況でタンパク質と治療薬候補分子の相互作用の動態を解析する手法だ。一般的に、その相互作用は1万分の1秒以内に起きる。今回のシミュレーションは、メインプロテアーゼの10万分の1秒（10マイクロ秒）間の揺らぎを再現したものだ（表紙、図2）。候補分子が結合する直前のメインプロテアーゼの動態と、そこに候補分子が結合したときの強さを推定する研究に寄与する。今後、同チームは候補分子とメインプロテアーゼが結合した状態や、ほかの標的タンパク質のシミュレーションなども検討していく予定である。

シミュレーションによって、標的タンパク質と治療薬候補分子の挙動が解明されると、次はそれらの間にかかる力（相互作用）を正確に予測することが重要になる。BDR制御分子設計研究チームと星薬科大学の共同研究グループは、標的タンパク質のうち、すでに結晶構造が公開されている65個について、日本発の量子化学計算手法であるフラグメント分子軌道法（FMO法）を用いて計算し、標的タンパク質を構成する各アミノ酸と治療薬候補分子の間に働く、静電相互作用、分散力相互作用などのエネルギー値を精密に計算した（図3）。それらの量子化学による相互作用エネルギー値は、世界中の創薬研究者が自由に利用できるFMOデータベースにて4月17日に公開した。

本誌刊行時点には、研究をさらに進展させ、その成果を公開しているはずだ。理研における新型コロナウイルス克服のための取り組みは、いっそう幅を広げ、また共同研究・連携の輪を拡大しながら、科学技術の総力を結集して進められている。

（構成：立山 晃／フotonクリエイト）

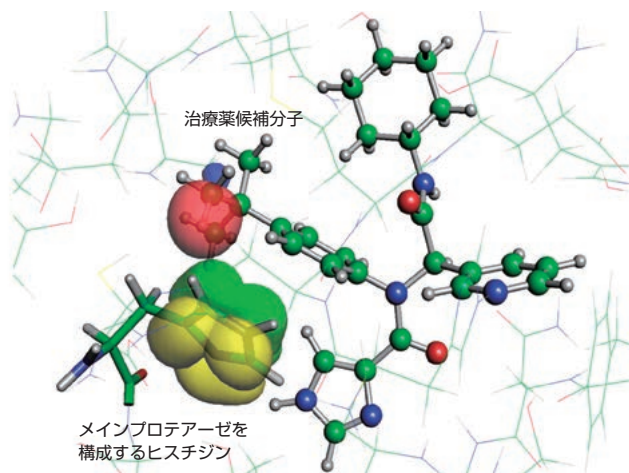


図3 FMO法によるメインプロテアーゼと治療薬候補分子の相互作用

メインプロテアーゼを構成するヒスチジンというアミノ酸の電子軌道（黄・緑）と、候補分子の電子軌道（赤）が接近して引力が働くことが分かった。このような計算により、候補分子の構造上のどのような特徴が標的タンパク質との強い結合に寄与するかを解明し、より強く結合することで高い効果を持つ治療薬を設計することができる。

医科学イノベーション推進プログラム (MIH) では、健康や病気に関するビッグデータをAI (人工知能) や数学を用いて解析することで、個別化医療の実現を目指している。2016年の発足以降、理研がハブ (中核拠点) となり医療機関や大学、企業と連携して展開してきた取り組み、そこから生まれた成果、そして今後について、小安重夫プログラムディレクター (PD)、桜田一洋 副PD、奥野恭史 副PDに聞いた。

AIを活用して一人一人に健康な生活を提供する

■ 個別化医療の実現を目指して

——理研は特定国立研究開発法人として、イノベーションをけん引するハブの役割を果たすことが求められています。医科学分野のイノベーションハブの構築を目指すMIHの発足には、どのような背景があったのでしょうか。

小安: 私は、2013年から統合生命医学研究センター (IMS) のセンター長を務めていました。IMSは、大規模なデータ解析をもとにゲノムと病気のなりやすさの関連を明らかにする研究と、免疫学を中心としたメカニズム研究を組み合わせることで、どうして病気になるのか、なぜ人によって病気のなりやすさが違うのかを理解し、一人一人に合った予防や治療を行う個別化医療につなげることを目指してきました。それには医療ビッグデータのさらなる活用が必要であり、ビッグデータを解析するためのAIの開発も必要となります。従来の研究分野や手法では限界があり、その壁を打ち破るためにも組織を超えた連携の場が必要だと考えるようになりました。

ちょうど科学技術振興機構 (JST) の「イノベーションハブ構築支援事業」の募集があったことから、「高精度の予測に基づく予防医療の実現に向けた疾患ビッグデータ主導型イノベーションハブ」を提案しました。2015年度のフィージビリティスタディ採択を経て、2016年度に本採択されました。その課題を実施するために発足したのが、MIHです。

——グループやチームが追加され、発足当初より活動が広がってきているようですね。

小安: 文部科学省は、膨大で高品質な研究データを集積し産学官で共有して活用することで新たな価値の創出につなげるデータプラットフォーム拠点の構築を進めています。理研は「健康・医療データプラットフォーム形成事業」を担当することになり、2017年4月からMIHで実施しています。さらに2018年10月からは、文部科学省の「Society 5.0実現化研究拠点支援事業」に採択された大阪大学の「ライフデザイン・イノベーション研究拠点」にも中核機関の一つとして参画しています。

MIHが目指すのは、個別化医療の実現です。そのために三つの事業を通して、個人の健康と医療に関わる情報を収集し利用できるプラットフォームを構築し、病気の正確な分類や変化していく体の状態について高精度な予測ができるAI、創薬のためのAIなどを開発してきました。MIHがハブとなり、大学や病院や企業と強力な連携体制を組んでいるのも大きな特徴です。

■ 卵巣がんの予後を血液検査データから予測

——これまでの成果を紹介していただけませんか。

桜田: ディープ・フェノタイピングの解析について大きな進展がありました。フェノタイプ (表現型) とは生物に表れている形質のことで、遺伝子発現や形態、行動、生化学的・物理的性質などがあります。多数のフェノタイプを計測することをディープ・フェノタイピングと呼び、生物の理解には不可欠だとして最近注目されています。しかし、例えば15項目のフェノタイプの計測データは15次元空間の点として表現されますが、15次元空間を視覚的に捉えて理解するのは困難です。そこで私たちは、情報の幾何学的な解釈を用いてディープ・フェノタイピングを解析する方法を開発しました。この方法によって15項目の計測データを2次元や3次元で表現できます。2~3次元であれば視覚的に理解可能で、そこからパターンや新しい分類を見つけることもできます。例えば、川上英良さん (健康データ数理



左から、小安重夫プログラムディレクター、桜田一洋 副プログラムディレクター、奥野恭史 副プログラムディレクター (2020年2月28日に取材)

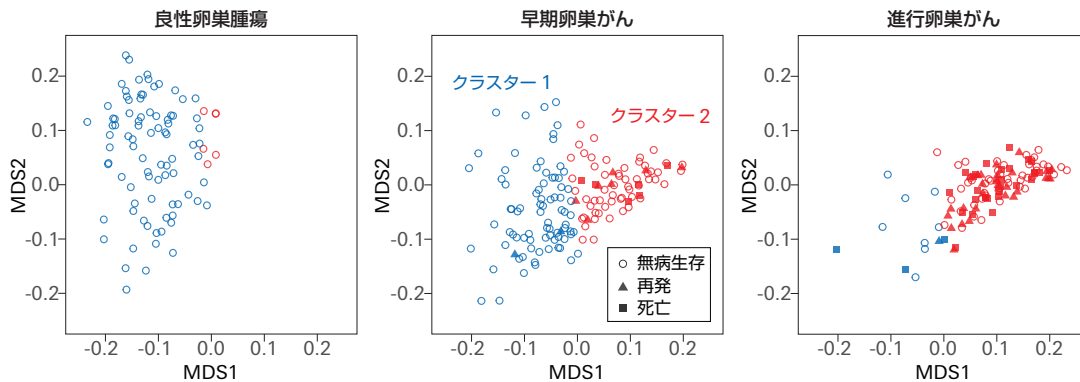


図1 術前血液検査データに基づく卵巣腫瘍の分類

卵巣腫瘍患者の術前血液検査32項目のデータを、教師なし機械学習と幾何学を用いて解析した。多次元尺度法 (MDS) によって似ている患者が近くなるように2次元で表すと、良性卵巣腫瘍と進行卵巣がんの分布が分かれた。早期卵巣がんは、良性卵巣腫瘍の分布 (クラスター1) に近い患者と、進行卵巣がんの分布 (クラスター2) に近い患者がいる。予後を調べると、クラスター1の患者はほとんど再発がなかったのに対し、クラスター2の患者は再発・死亡率が高かった。

推論チーム チームリーダー) は、卵巣腫瘍の特性を手術の前に分類し、予後を予測することに成功しています。

卵巣腫瘍には良性と悪性があり、悪性の卵巣腫瘍は転移の有無などによって早期卵巣がんと進行卵巣がんに分けられます。早期がんでも、腫瘍の切除後に再発することがあります。患者さんが一番知りたいのは再発するかどうかでしょう。現在の分類では、患者さんが知りたいことに答えられていないのです。

川上さんは、卵巣腫瘍の患者さんに対して手術前に行った血液検査の32項目のデータを、教師なし機械学習と幾何学を用いて解析しました。2次元分布で表現すると、良性腫瘍と進行がんははっきり分かれ、さらに早期がんは、良性腫瘍寄りに配置されるものと、進行がん寄りに配置されるものに分かれた (図1)。患者さんの予後を調べると、良性腫瘍寄りの方は再発率がほとんどなかったのに対して、進行がん寄りの方は再発率と死亡率が高くなっていました。予後に関連する特性を手術前に、しかも血液検査のデータだけで得られることから、適切な治療戦略を立てることができると期待されています。

健康から糖尿病へ至る道筋が見えた

桜田：石川哲朗さん (健康データ数理推論チーム 特別研究員) もまた、機械学習に幾何学的な解釈を組み合わせて、健康状態を可視化することに成功しました。具体的には、健診データに基づいて糖尿病の発症と関連している血糖値や血圧など8個のリスク因子の組み合わせをエネルギーランドスケープモデルで表現したのです。エネルギーランドスケープモデルとは、エ

ネルギーの高低を山や谷に見立てた地形図のようなもので、物理学や生化学で使われています。山は出現頻度が低いこと、谷は出現頻度が高いことを示しています。解析の結果、谷は7個の領域に分けられました (図2)。これは、無数にあるリスク因子の組み合わせがたった7個のパターンに分類できることを意味しています。さらに健診データを追跡すると、5年以内に糖尿病を発症した割合は、7個の領域ごとに大きく違っていました。この結果から、糖尿病を発症しやすいリスク因子の組み合わせや発症しにくいリスク因子の組み合わせが分かり、その中からこれまで知られていないパターンが浮かび上がってきました。

しかも、一人ずつ時系列で見えていくと、糖尿病の発症までの状態をたどったかを追跡できます。そうした情報に基づけば、あなたは肥満になると糖尿病を発症しやすいから摂取カロリーを抑えた方がいい、あなたは肝機能が低下すると糖尿病を発症しやすいから飲酒を控えた方がいい、といったように、一人一人に対して発症までのルートを予測し、予防策の提案ができるようになります。

このほか、機械学習の中で特に注目されているディープラーニング (深層学習) を用いたディープ・フェノタイピングの技術開発も進めています。清田 純さんが率いる健康データ深層学習チームは、40項目の臨床時系列データから6時間後に敗血症を発症するかどうかを予測するという国際コンペティションに参加し、101チーム中10位という成果を上げました。敗血症は、急速に進行して死に至ることもあるため、とりわけ早期発見が重要です。しかし初期段階では診断が難しいのです。臨床デー

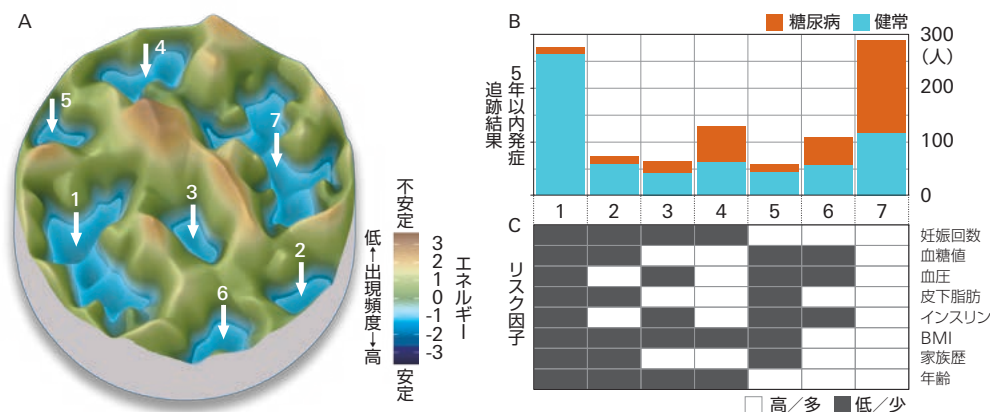


図2 健康状態を可視化するエネルギーランドスケープモデル

健診データをもとに糖尿病の8個のリスク因子の高低 (多少) を、機械学習に幾何学的な解釈を組み合わせて解析し、エネルギーランドスケープモデルで表現した。水色から青はエネルギーが低く谷に相当し、不安定で出現頻度が高いこと、緑から茶色はエネルギーが高く山に相当し、不安定で出現頻度が低いことを示している。地形はエネルギーの谷底 (局所安定点) を囲む7個の領域に分かれている (A)。各領域で5年以内に糖尿病を発症する割合 (B) は、リスク因子の組み合わせ (C) によって違うことが分かった。

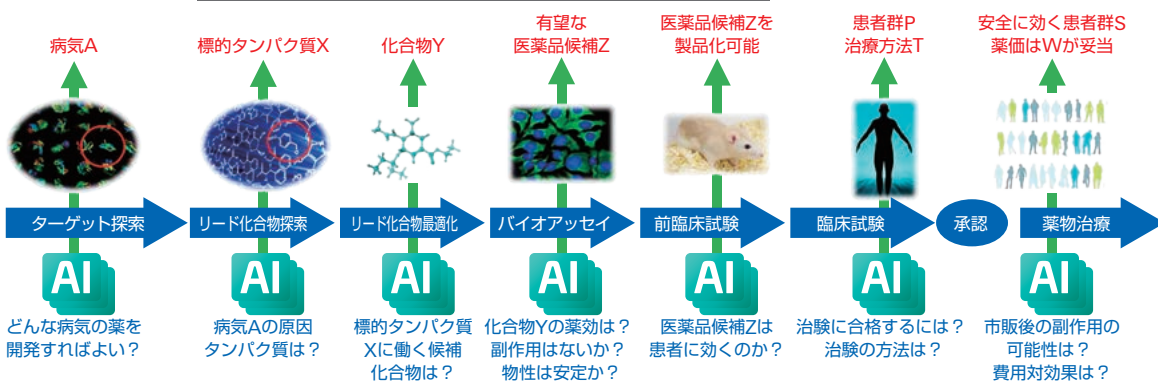


図3 創薬丸ごとAI化のイメージ

IT、ライフサイエンス関連の企業やアカデミアがコンソーシアムを形成し、医薬品開発プロセスの全てをカバーする約30種類のAIを開発することを目指している。MIHは、メディシナルケミストリー領域のAI開発を担当している。

タから発症を予測できるようになれば、治療をいち早く始められ重症化を防げると期待されています。

複雑性と個性を持つ生命を理解し、正しく分類したり予測したりするには、従来とはまったく異なる概念と手法、つまり新しい科学が必要です。川上さんと清田さんは、医学と数学、両方の知識を持っています。だからこそ新しい手法を開発し、画期的な解析ができるのです。彼らのもとには、大学院で数学を専攻した後に医学の分野に入ったり、数学と医学の知識を持つ若い研究者が集まっているので、今後も楽しみです。

■ AIが化学構造を見る目を持った

— MIHでは創薬のためのAIの開発にも力を入れています。

奥野：一つの薬を開発するのに、10年以上、1,000億円以上もの費用がかかります。市販まで到達できる成功確率は2万5千分の1以下といわれています。その状況を変えるにはAIの活用が不可欠だと考え、私は2016年にライフ インテリジェンス コンソーシアム (LINC) を立ち上げ、代表を務めています。LINCには、IT関連企業、ライフサイエンス関連企業、関連領域のアカデミアなど約110の企業・団体が参加していて、理研のMIHも主要メンバーです。現在は、LINCのメンバーで分担して、医薬品開発プロセスの全てをカバーする約30種類のAIの開発を進めています (図3)。これほどたくさんのAIを連結して医薬品を開発しようという試みは、世界でも類を見ません。医薬品開発プロセスを丸ごとAI化できれば、開発期間と費用を抑え、成功確率が大幅に向上すると期待されます。

— MIHが担当している創薬AIはどのようなものですか。

奥野：メディシナルケミストリー (創薬化学) という領域のAI開発を担当しています。どの病気の薬をつくるべきかが決まると、その病気の原因タンパク質に作用する化合物を探索します。化合物の探索から物性の確認、化学構造の最適化、合成までがメディシナルケミストリーです。タンパク質の立体構造や化合物との結合状態を予測して候補化合物を探索するAI、候補化合物の安全性や物性を予測するAI、最適な化学構造を設計するAI、候補化合物の合成経路を自動で提案するAIなどの開発を行います。

特に画期的なのが、化学構造を学習するAIです。化学を勉

強した人は化学構造を見るだけで、水に溶けやすい、毒性があるといった、その化合物の物性が分かります。そうした化学構造と物性の情報をAIに学習させてあるので、未知の化合物について化学構造だけから物性を予測したり、希望する物性を持つ化学構造を探索したりできます。アイコンを選んでだけで化学構造を設計できるインターフェースも開発したので、プログラムが書けない人でも使うことができ、すでに製薬企業や化学メーカーで利用されています。

■ 医科学を超えて

— 個別化医療を実現するために、今後、必要なことは何ですか。

桜田：MIH発足のきっかけとなったプロジェクト「高精度の予測に基づく予防医療の実現に向けた疾患ビッグデータ主導型イノベーションハブ」は2020年3月で終了しましたが、疾患データベースのプラットフォームの構築やディープ・フェノタイピングの解析手法の開発など、予想以上の成果を上げることができました。次にやらなければいけないのは、遺伝学とディープ・フェノタイピングの統合です。それにより、遺伝要因と環境要因の両方を踏まえて未来を予測できるようになります。また、AIによる予測を使うには、その信頼性の評価も必要です。国際標準となっているシステムによる評価を行いながら技術開発を進めていくつもりです。

奥野：創薬AIで開発した医薬品を社会に送り出すことです。しかし、出口を急ぐばかりではなく、難しくまだ誰もやっていないけれども実現すれば創薬を大きく変えるようなAIの開発に、腰を据えて取り組むことも重要です。医薬品の開発プロセスには生物学、化学、物理学、医学などあらゆる研究分野が入っていますから、開発した創薬AIはほかの分野の基礎研究でも大いに役立つでしょう。

小安：創薬AIに限らずMIHで開発している技術は、どれも普遍的に使える技術です。MIHは、すでに医科学の研究領域を超えて広がっています。MIHの理研の中での位置付けを見直し、数学・情報系のコアに据えることで、さまざまな研究分野と連携し、貢献できるのではないかと考えて、よい形を模索しているところです。

(取材・構成：鈴木志乃/フotonクリエイト、撮影：STUDIO CAC、奥野竹男)

革新知能統合研究センター（AIP）病事情報学チームの山本陽一郎チームリーダー（TL）は、病理画像に含まれるがんの特徴をAI（人工知能）が自ら学習して発見する技術を開発。AIが発見したがんの特徴の中には、これまで知られていないものが含まれていた。AIが見つけた新しい特徴も使うと、がんの再発予測の精度が向上する。医療AIの新しい時代の幕開けとなる成果を紹介しよう。

がんの未知なる特徴をAIが発見

■ 病理学とAI

病理医は、患者から採取した組織や細胞を顕微鏡で観察し、細胞の形やその並び方などから、どの疾患かを診断する。例えば、がんが疑われる場合、X線画像、血液、遺伝子などさまざまな検査が行われるが、病理医による診断が確定診断となる。

病理専門医の資格を持つ山本TLは、こう語る。「病理医の診断によって治療方針が決まるので、病理医は大きな責任を負っています。一口にがん細胞と言っても形や並び方は部位や患者さんごとに違うため、病理診断には広い知識と経験が必要です。また、診断にはただ一つの答えが必ずある、というわけではありません。熟練した病理医の間でも診断が異なることもあります。病理学をさらに発展させ、さまざまな患者さんの役に

立ちたい。その思いが、私の研究の出発点です」

山本TLは、病理専門医資格を取得後、米国のハーバード大学や大規模な総合病院として有名なミネソタ州のメイヨー・クリニックで、がん治療に対する数理モデルの開発を行い、病理と数理を統合する研究に取り組んだ。2018年に理研AIPで研究室を立ち上げ、病理にAIを活用することで、疾患の発症メカニズムの解明や新規治療法の発見、患者ごとに最適な治療方法を選択するシステムなどの開発を目指した研究を行っている。「数理モデル開発の経験は、私の医療AI研究の基礎になっています」

■ AIのブラックボックスに挑む

現在のAI技術の中心となっているディープラーニング（深層学習）は、機

械学習の手法の一つである。また、機械学習には、大きく分けて教師あり学習と教師なし学習がある。医療では、教師あり学習のディープラーニングが多く使われている。「教師あり学習のためには、たくさんの病理画像1枚1枚に、がんか否か、がんであれば悪性度など診断情報を付けなければならない、それは大変な作業です。また、新たな分類を見つけることができないという限界があります」

さらに山本TLは、「AIを医療に利用するには、人の命を扱うという医療の特殊性を考えていく必要がある」と指摘する。「今までの医療AIでは、判断根拠が医学的にブラックボックスとなっており、なぜその結果になったのかという根拠を医師が直接的に理解することが困難でした。根拠が分からなければ、医師は患者さんに自信を持って診断を説明する

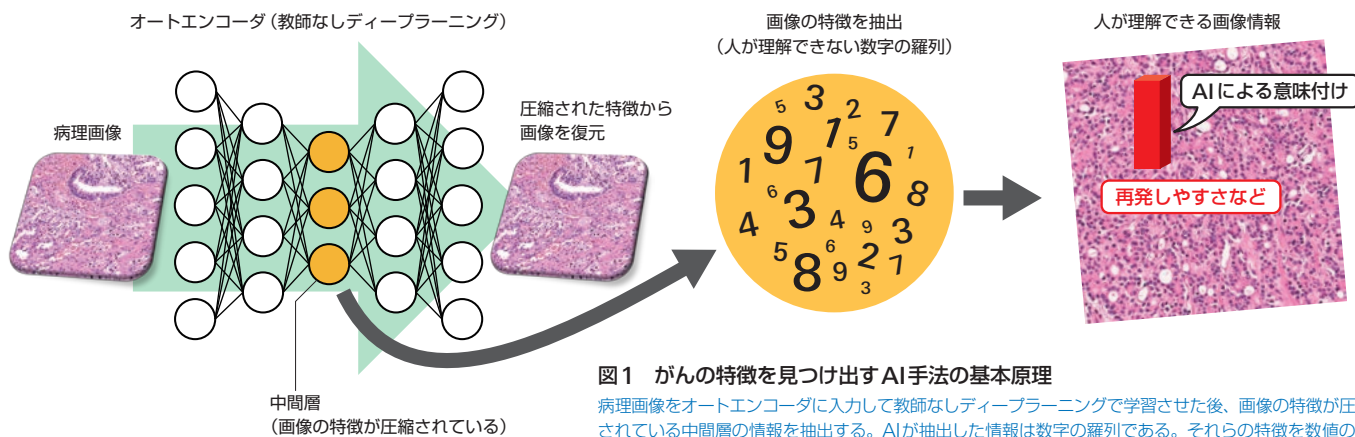


図1 がんの特徴を見つけ出すAI手法の基本原則

病理画像をオートエンコーダに入力して教師なしディープラーニングで学習させた後、画像の特徴が圧縮されている中間層の情報を抽出する。AIが抽出した情報は数字の羅列である。それらの特徴を数値の規則性を手掛かりに、新しく開発した技術を用いて人間が理解できる高解像度の画像情報として提示する。その後、画像に対してAIが最適な意味付けを行う。

山本陽一朗（やまもと・よういちろう）

革新知能統合研究センター
 目的指向基盤技術研究グループ
 病理情報学チーム
 チームリーダー

東北大学医学部医学科卒業。東北大学大学院医学系研究科博士課程修了。博士（医学）。米国ハーバード大学、メイヨー・クリニック、ドイツ・ハイデルベルク大学にて研究に従事。信州大学医学部特任准教授、理研革新知能統合研究センター コニットリーダーなどを経て、2018年より現職。



ことができず、また患者さんも納得できないでしょう。医療においてAIを信頼して使うためには、AIが導き出した結果を人間が理解できる情報として取り出せるようにすることが重要と考えています」

説明可能なAIを求める傾向は、医療以外の分野でも強まっている。「AIは私たちが気付いていない未知の特徴や規則を使って結果を出している可能性もあります。ブラックボックスを開けることができれば、病理学分野にとどまらない医学のさらなる発展につながる新たな知識の獲得も期待できます」

■ 人間が理解できる情報を取り出す

医療AIを取り巻くさまざまな問題を解決するため、山本TLらは教師なし学習のディープラーニングをベースとして病理画像から人間が理解できる形で高精度の情報を取り出すAI技術を開発した（図1）。

このAI技術には複数の手法が組み合わされている。その中の一つがオートエンコーダである。オートエンコーダは、例えば入力画像を圧縮して情報を減らし、その後、圧縮した情報から画像を復元する手法だ。たくさんの画像を使い、入力画像と復元画像が同じになるように学習を繰り返していく。すると、AI自身が画像の重要な特徴を見つけ、その特徴を残して画像を圧縮することで、入力画像に近い画像を復元できるようになる。今回は、低倍率の画像と高倍率の画像に対する2段階のオートエンコーダを使用している。

入力画像から復元画像を出力するこ

とに、どういう意味があるのだろうか。山本TLは、「今回の研究で重要なのは、オートエンコーダの出力ではなく、中間層です。中間層には、AIが見つけた入力画像の特徴が圧縮されています。その情報を取り出して使うのです」と答える。

しかし、オートエンコーダの中間層から取り出した画像の特徴は数字の羅列で表現されていて、人間は直接的には理解できない。「数字の羅列が持つ情報を言葉に変換するのも一つの方法ですが、画像診断のプロにとっては、言葉より画像を用いた表現の方が、よりよく情報を理解できます。そこで、中間層から取り出した情報を、数値の規則性を手掛かりに、高解像度の画像として出力することにしました。このとき、特徴を表現した人工の画像ではなく、特徴が最も顕著に表れている入力画像をAIが選んで出力するように設計しています。こうすることで、AIが見つけた入力画像の

特徴を人間が理解できる情報として、しかも病理医が見慣れている本物の病理画像として取り出すことができるようになりました」

さらに、入力した病理画像に対応する患者の手術から再発までの期間の情報を独自の手法で与えることで、AIが見つけた特徴ごとにがんの再発のしやすさを求められるようにした。

山本TLらは、この新しいAI技術を前立腺の病理画像に適用してみた。

■ 再発に関わる新しい特徴をAIが発見

「日本医科大学泌尿器科の木村 剛 准教授の協力で、前立腺を丸ごと病理標本として観察できる全割・全包埋という世界的にも貴重な病理標本を使用することができました」と山本TL（図2）。前立腺は男性の膀胱の下に位置し、尿道を取り囲んでいる。日本における男性のがんの部位別罹患数で、前立腺は第1位である（厚生労働省「全国がん登録 罹

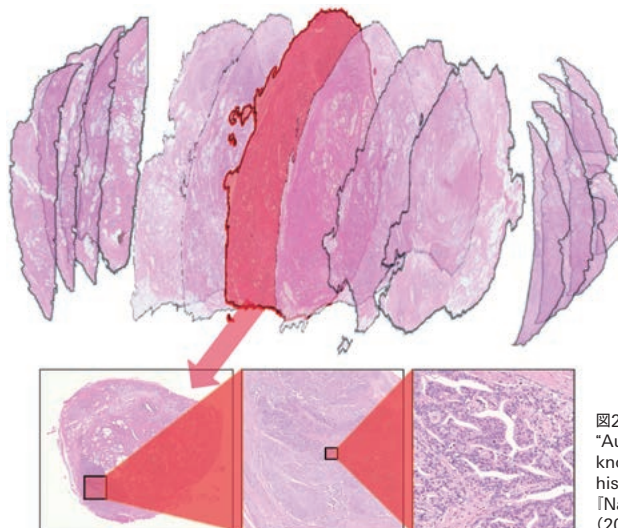


図2 前立腺の病理画像

摘出された臓器全体をパラフィンなどで固め、全領域を薄く切ってスライドガラスに貼り付けた全割・全包埋の病理標本をデジタル化した。病理画像は1枚100億画素以上の超高精細で、拡大していくと細胞一個一個を識別できる。

図2および図4：
 “Automated acquisition of explainable knowledge from unannotated histopathology images”
 [Nature Communications]
 (2019年12月18日付) より

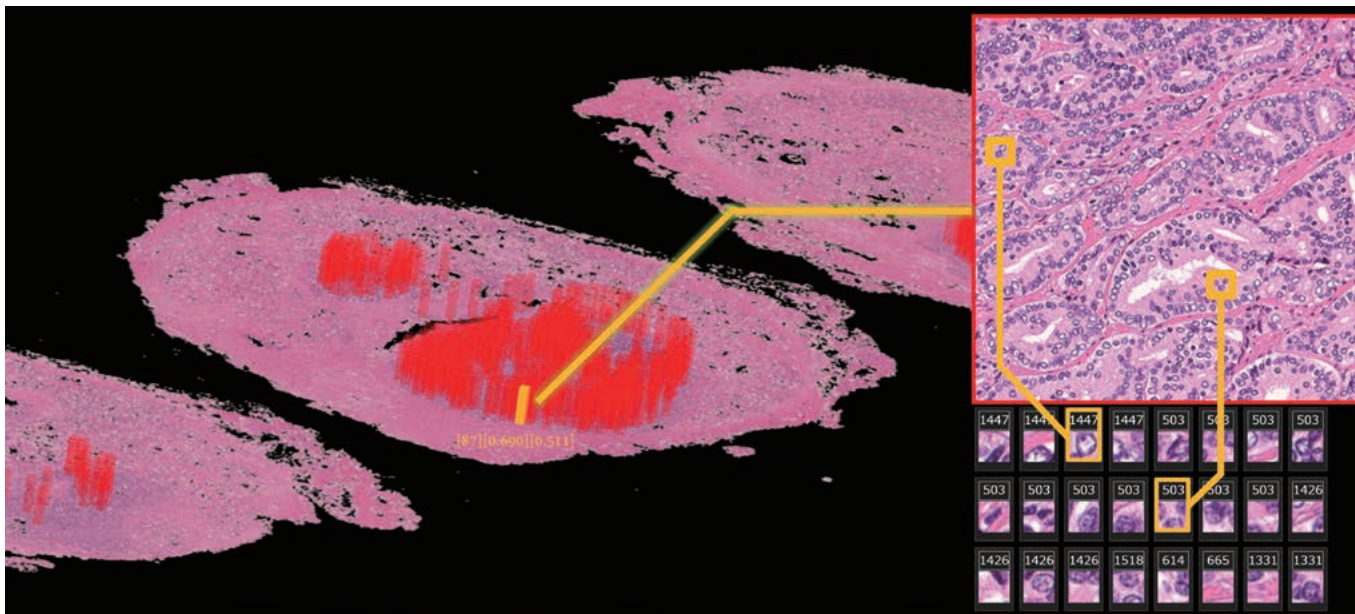


図3 AIが発見したがんの特徴

左は、前立腺病理標本の連続切片から作成した3D病理画像である。AIが再発しやすいと導き出した領域を赤と黄色で示している。右上は黄色い領域に対応する低倍率の病理画像。右下は高倍率の病理画像で、画像の特徴を強く表している細胞像が提示されている。

患数・率 報告」2017年)。病理標本は、摘出された組織をパラフィンなどで固め、薄く切った切片をスライドガラスに貼り付けてある。それをAIで扱うことができるように、デジタル画像化した。

そして、100人分の病理画像を約11億枚に分割し、開発したAI技術に適用したところ、再発しやすい特徴を持つ病理画像を自動で抽出させることに成功した(図3)。「結果を見て驚きました。再発しやすいとAIが導き出した病理画像の中に、前立腺がんの診断指標であるグリソンスコアで悪性度が高いとされるパターンが含まれていたのです」

グリソンスコアは、診断精度の高さが認められ世界中の病院で使用されている指標だ。病理医は、患者から採取した前立腺がん組織の細胞の形や並び方を顕微鏡で観察し、グリソンスコアのパターン1~5のどれに相当するかを判定

する。数値が大きいほど悪性度が高く、最も多い病変のパターンと2番目に多い病変のパターンを判定し、その数値を合計してグリソンスコアを算出する。グリソンスコアが高いほど、がんの悪性度が高く再発しやすい。

「入力した画像が病理標本であることも、がんの患者さんのものであることも、AIは知りません。AIは約11億枚の画像の特徴を独自に抽出しただけです。にもかかわらず、AIは自力で、がんの特徴を見つけたのです」

驚きはそれだけではない。「AIが再発しやすいと導き出した病理画像の中に悪性度が高いグリソンパターンにはないものがありました。それは間質の画像でした」。臓器の中でその臓器の機能に関わる部分を実質という。実質を支える部分が間質である。「がんの多くは実質で発生するので、病理医は実質を中心に

観察します。ところがAIは、がん細胞が存在しない間質にも再発のしやすさに影響する特徴があることを導き出したのです」

AIが再発しやすいと導き出した間質の画像は、通常より細胞密度が高かった(図4)。山本TLは、間質の細胞密度がどのようなプロセスで再発のしやすさに影響するのかは今後の研究が必要とした上で、「病理医も重要視していなかったがんの再発に関連する特徴を、AIが自ら学習して見つけ、さらにそれを人間が理解できる情報として取り出すことができた意義は大きい」と語る。

■ がんの再発の予測精度が向上

さらに山本TLらの研究は続く。「AIが見つけたがんの新たな特徴が再発予測に役立つかどうかを確かめました」。まず、日本医科大学病院の1万3,188枚の病理画像で検証した。AIが見つけたがんの特徴を用いて1年後に再発するかどうかを予測し、患者の予後情報をもとに求めた予測精度は、0.820となった(図5)。この値はAUCと呼ばれる機械学習で用いられる評価指標で、1に近いほど精度が高いことを意味する。同じ画像について病理医が従来のグリソンスコアを使用して予測したAUC値は、0.744

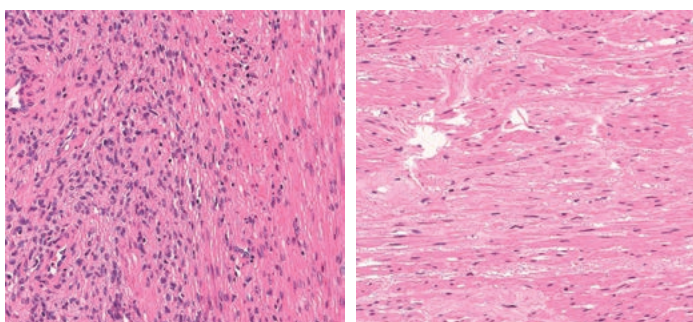


図4 再発に関連する間質の特徴

左は、再発しやすい特徴としてAIに検出された間質。右は、再発しにくい特徴としてAIに検出された間質。左の間質の方が細胞密度が高い。

だった。

グリソンスコアを用いた診断には、愛知医科大学病院病理診断科の都築豊徳教授の協力を仰いだ。「グリソンスコアは病理医によって数値が変わることもあるため、比較対照するグリソンスコアを誰が付けたのかも重要なポイントとなります。都築先生は最新版のWHO（世界保健機関）腫瘍分類の策定にも関わっている、前立腺がんの病理診断のエキスパートです。今回、都築先生のグリソンスコアと比較できたため、AIが見つけた新たながんの特徴を使うと病理医が従来の基準を使うより高い精度で再発予測ができることを、信頼性の高いデータで示すことができました」

一方で、ディープラーニングには過学習という問題も指摘されている。学習に使った画像を覚え過ぎてしまい、汎用性が低下してしまう現象のことだ。そのため、学習用の画像を取得した病院とは異なる病院でAIを使用するときに、予測精度が落ちることがあり、しばしば医療AIの実応用に向けての問題となっている。そこで、聖マリアンナ医科大学病院

と愛知医科大学病院の病理画像2,276枚を用いた検証も行った。AIが見つけた特徴を使用した再発予測のAUC値は0.845で、日本医科大学病院の病理画像の場合とほぼ同等だった。「今回、AIが見つけたがんの特徴は、学習画像を取得した病院とは別の病院の病理画像にも使える、汎用性の高いものであることが示されました。これは医療AIの実応用に向けて非常に重要な結果です」と山本TLは解説する。

さらに、AIが見つけた特徴と病理医が用いているグリソンスコアの両方を使うと予測精度がさらに向上した。「しばしば“医師 対 AI”という見方をされがちです。しかし、医師と競うのではなくチームを組むことが、医療におけるAIの上手な使い方だと、私は考えています」

■ 一歩先のことをやる

2019年12月にこれらの成果を発表すると、多数の新聞やテレビ、雑誌、ウェブメディアなどで取り上げられた。日本だけではない。アメリカや中国、ロシア

など海外メディアでも広く報道された。基となった論文は、英国の科学雑誌『Nature Communications』で2019年に最も多く読まれた物理分野の論文「Top 50 Physics Articles」の第5位に選ばれている。「多くの大学や病院の協力と、大量のデータを扱うことができるAIPのスーパーコンピュータ『RAIDEN』があったからこそ得られた成果です」と山本TLは語る。

AIが見つけたがんの特徴は、病理画像アトラスとして数十ページにわたって論文に掲載されている。病理画像アトラスとは、病理画像と解説が集められた図表集のことである。「がんの新しい特徴についての知識が多くの病理医に共有され、研究されていくことで、病理学のさらなる発展に貢献できればうれしいです」と山本TL。がんの再発予測の精度が向上すれば、一人一人に合った治療の選択ができるようになる。また、がんの新しい特徴の発見は、新たながん治療法の開発につながる可能性がある。

今回は前立腺がんの再発予測を対象としたが、山本TLらが開発したAI技術は、ほかのがんの診断や再発予測、また、さまざまな画像から新たな知識を獲得するための自動解析手法として、医療以外でも役立つと期待されている。

現在と未来の患者さんの治療に役立つ研究を——山本TLの信条だ。「理研にいますので、私たちは一歩先のことをやるべきと考えています。そして、その成果を患者さんに届ける責務があります」と山本TLは力強く語る。

(取材・執筆：鈴木志乃／フotonクリエイト)

学習画像と同じ施設の画像による検証	異なる施設の画像による検証
0.744	医師が従来の基準(グリソンスコア)を使用 0.721
0.820	AIが新たながんの特徴を使用 0.845
0.842	医師+AI全ての特徴を使用 0.889

図5 前立腺がんの再発についての予測精度の比較

数値はAUCという指標で、1に近いほど予測精度が高い。学習に用いた日本医科大学病院の病理画像だけでなく、異なる施設（聖マリアンナ医科大学病院と愛知医科大学病院）の病理画像でも、医師が従来のグリソンスコアを使用した予測より、AIが発見した新しい特徴を使用した予測の方が精度が高い。グリソンスコアとAIが発見した特徴を合わせて使うと、予測精度がさらに向上する。

「理研史の生き字引」を 目指して

三輪紫都香 みわ・しづか

広報室 事務業務員

記念史料室は理研と理研に関わる科学史についての史料を収集・展示しています。私は2018年に広報室に着任し、学芸員として記念史料室の運営に携わっています。創立から100年以上の歴史を持つ理研では数多くの発見や発明がなされ、今日の私たちの生活の一部となっている成果も多くあります。また、理研の沿革には調べても調べても尽きることのない大小さまざまな物語があります。記念史料室はそんな歴史の一握りをすくい上げご紹介することで、理研の魅力をお伝える役目を果たしています。

歴史学は史料から過去の事象を明らかにし、それを未来に活かす学問です。事実を明らかにするために史料の精査・検討はとても重要な作業です。私は近代日本の文化政策について研究してきましたが、史料を文化財として保存することの大切さを研究の中で実感し、学芸員という職業を選択しました。理研の史料は着任するまで扱ったことはありませんでしたが、これまでの経験をもとに試行錯誤を繰り返しながら業務に取り組んでいます。

史料はその当時につくられたものですから、時代をさかのぼるほどに当然古くなります。緩急の差はありますが、どんな史料も劣化します。貴重な史料は少しでもその劣化を遅らせなければなりません。しかし、誰の目にも触れさせずにしまい込んでいるのでは意味がありません。学芸員の仕事は史料の保存と活用の間でバランスを取り、適切な環境下で最大限史料を利用に供することであると思っています。そのために、史料一つ一つの状態をきちんと把握すること、日々の変化に向き合うことを心掛けています。

理研が所蔵する史料は大正時代から現在に至るまでの組織文書や、研究者が大切に保管してきた研究試料や実験器具、原稿など多岐にわたります。その中には理研に在籍していた朝永振一郎や湯川秀樹の直筆の履歴書といった、ほかでは見られない貴重なものも。さらに、理研の歴



写真1・筆者近影(左)。「理研史の生き字引」たちと共に。



写真2・弓を引く一連の動作の一つ、「会」の状態にある筆者。弓を引き分けて上下左右に伸び合い、発射の機を熟させている。その後「離れ」で放ち、「残心」で矢の行く末を見つめる。

史だけでなく、科学史や美術史、軍事史などさまざまな視点からの検証を要するものなど、数え上げれば切りがありません。必要とされる史料をご要望に応じて速やかに紹介できたときなどに、この仕事のやりがいを感じます。

一方で私はもともと理系の分野は得意ではないので、研究者の方々が取り組まれている研究内容は概要ですらなかなか理解できません。見学者への解説は記念史料室に在籍している「理研史の生き字引」の方々に頼っているのが現状です(写真1)。しかし、理研を「知りたい」という気持ちに答えられるよう、私からも正確な情報を提供できるようになりたいですし、その先に自分の研究活動も展開していくと思っています。

昨年亡くなった私の祖父は「苦手だと思ふことでも、周りの3倍努力すれば人並みくらいにはなれる」とよく私の背中を押してくれました。3倍の努力はなかなか難しいですが、その言葉を信じて学生時代から続けている弓道は私の心の支えになっています(写真2)。同じように、日々の積み重ねによって少しずつ新しい知識を自分のものにしていけたらと思っています。次の「理研史の生き字引」を目指して、これからも頑張ります。



理研の活動をご支援ください。

理研の研究の充実、さらなる発展は、法人や個人の皆さまからのご寄附で支えられています。

●問合せ先

理研 外部資金室 寄附金担当

Tel : 048-462-4955

Email : kifuf-info@riken.jp



<https://www.riken.jp/support/>