

2024 年 6 月 3 日

国立大学法人東京大学医科学研究所

がんゲノムにおける構造異常の病原性を高精度に予測し、根拠を説明する人工知能を開発～全ゲノムに基づくゲノム医療への貢献に期待～

発表のポイント

- ◆がん患者の全ゲノム解析データから、構造異常の病原性を世界トップレベルの精度で予測し、その根拠をがんの発生メカニズムに基づいて説明できる「説明可能な人工知能（XAI）」を世界で初めて開発しました。
- ◆本 XAI 技術により、がん患者のゲノムに見つかった構造異常の臨床的な解釈が進み、ゲノム医療の発展が期待されます。

概要

東京大学医科学研究所附属ヒトゲノム解析センター健康医療インテリジェンス分野の井元清哉教授、東京大学医科学研究所附属先端医療研究センター造血病態制御学分野の南谷泰仁教授、富士通株式会社（以下、富士通）の共同研究グループは、がんゲノムにおける融合遺伝子（注 1）等の構造異常（注 2）の病原性を高精度に予測し、予測根拠も説明できる人工知能（Explainable AI: XAI）（注 3）を開発しました。本研究で開発した XAI 技術は、既存の手法以上の精度で融合遺伝子の病原性を予測できることがわかりました。さらに、XAI 技術が提示した予測根拠は、構造異常ががんを引き起こす実際のメカニズムと整合することが確認されました。本研究成果は、がん患者の全ゲノム解析から新たに見つかった融合遺伝子の病原性を、素早くかつ正確に判定して、患者一人ひとりに最適な治療方針の決定を支援するゲノム医療の実現に向けた重要な一歩になると考えられます。

本研究成果は 2024 年 5 月 17 日（金）、国際科学誌「*Cancers*」オンライン版で公開されました。

背景

近年、全ゲノム解析技術の進展により、がんの診断と治療における遺伝情報の重要性が増しています。本邦においては、2019 年に策定された「全ゲノム解析等実行計画」に基づき、全ゲノム情報を用いたゲノム医療が推進されています。全ゲノム解析では、一塩基変異（注 4）のみならず、大きな染色体異常にも繋がる構造異常も網羅的に同定できることがメリットの 1 つです。遺伝子の構造異常は、がんの発症や進展に深く関わることが知られています。しかしながら、同定された構造異常が、がんの原因となるドライバー変異（注 5）なのか、それとも病気に関係しないパッセンジャー変異なのかを判断することは容易ではありません。特に、融合遺伝子等の構造異常は判断が難しく、場合によっては数多くの関連文献を調査する必要があるため、専門家でも多大な労力と時間を要します。また、医療応用に向けては、予測精度の高さだけでなく予測根拠の説明も重要です。そこで、高精度に構造異常の病原性を予測し、その根拠も説明できる AI 技術がゲノム医療の発展のために求められています。

研究成果

本研究では、がんにおける融合遺伝子を伴う構造異常の病原性を高精度に予測し、予測した根拠を説明できる説明可能な人工知能（XAI）技術を世界で初めて開発しました。この XAI 技術は、知識グラフ（注 6）と呼ばれる大規模なデータベースと、深層学習という高度な機械学習（注 7）技術を組み合わせた独自のアルゴリズムにより実現されています（図 1）。

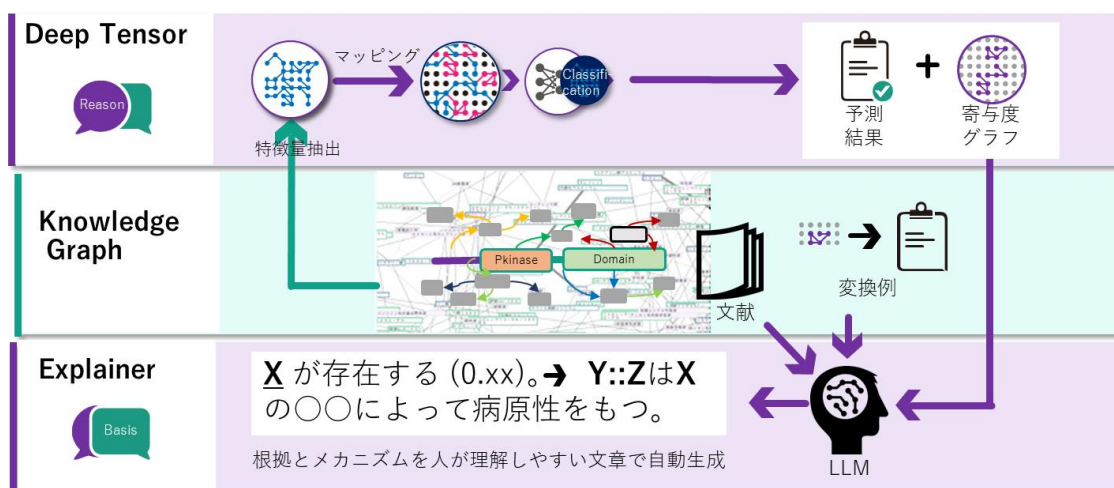


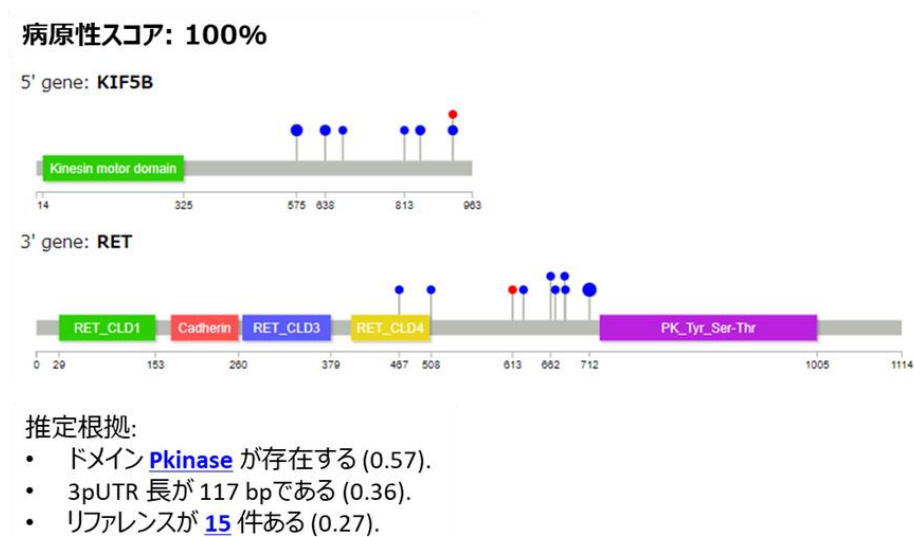
図 1. 本研究で開発した XAI システムの概要図

がんゲノムの構造異常の病原性予測機能（上部）と、文献やデータベースの情報とその関連性を蓄積した知識グラフ（中央部）、予測の根拠を説明する機能（下部）から構成される。病原性予測では知識グラフから抽出した特徴量を用いて深層学習モデルで予測を行う。説明機能では、予測に寄与した知識グラフ上の特徴量を基に、大規模言語モデル（LLM）を用いて人が理解しやすい説明文を自動生成する。

本研究では、遺伝子やタンパク質、ゲノム変異に関する文献やデータベースの情報とその関連性を知識グラフ化することで、AI が膨大な情報を自由に取り出せる形式で蓄積し、計算対象としてデータ化しました。そして、過去の文献情報から病原性の有無が分かっている融合遺伝子の情報を教師データとして用い、知識グラフから抽出した様々な特徴量を入力として深層学習モデルを構築しました。このモデルを用いて、新しく見つかった融合遺伝子に対して、病原性の有無を評価させたところ、これまでの手法と同等以上の高い精度でその病原性を予測できることが分かりました。本技術は、富士通が有する知識グラフを駆使した最先端の XAI 技術と、東大医科研のがんゲノム医療での豊富な研究実績にもとづく深い知見とを巧みに組み合わせることで実現しました。

さらに、この XAI 技術の大きな特徴は、予測の根拠を人が解釈可能な形式で提示できる点にあります。具体的には、病原性予測に寄与した知識グラフ上の特徴量を抽出し、自然言語処理技術を用いて、人が理解可能な文章で説明を生成します。また、いくつかの融合遺伝子の予測事例について、XAI が生成した予測根拠の説明内容を実際に分析したところ、これまで知られているがんを引き起こすメカニズムと整合していることが確認され、XAI の推論が妥当であることが確認されました。例えば、非小細胞肺癌でよく見られる *KIF5B::RET* 融合遺伝子では、RET のキナーゼドメインが活性化することが病原性に関わると考えられています（図 2）。実際に、本技術では *KIF5B::RET* の病原性を予測する際の根拠として、1) RET のキナーゼドメインの存在と、2) 過去に 15 報の文献報告が存在する点を特徴量として重要視していました。さら

にその予測根拠として、「RET のキナーゼドメインが KIF5B との融合により活性化し、がん化に関わる細胞生存や細胞増殖シグナルを生じると考えられる」といった説明が自動生成されました。このように、医師が納得できる形で予測根拠が示されることで、ゲノム医療など高度な説明責任が問われる場面においても、XAI の予測結果を安心して意思決定に役立てられるようになります。



要約: *KIF5B::RET* 融合の病原性は、KIF5B との融合によって RET キナーゼドメインが恒常的に活性化されることに起因しています。これにより、細胞増殖と生存を促進する経路を通じた持続的なシグナル伝達が行われ、発癌に寄与します。

重要点: *KIF5B::RET* 融合は RET チロシンキナーゼドメインの恒常的な活性化を引き起こし、非小細胞肺癌（NSCLC）でのがん原性シグナル伝達と細胞増殖を促進します。これにより、そのキナーゼ活性を阻害する分子療法の標的となる可能性があります。

図 2. *KIF5B::RET* での病原性予測の具体例

各遺伝子のタンパク質構造、予測に寄与した推定根拠、およびキナーゼドメインが病原性に与える機序に関する要約と重要点を示す説明文が示されている。

今後への期待

本研究で開発した XAI 技術により、がん患者の全ゲノム解析から新たに見つかった融合遺伝子の病原性を素早く、かつ正確に判定できるようになります。この成果は、患者一人ひとりに最適な治療方針の決定に役立つだけでなく、新規の創薬ターゲットの発見にも貢献すると期待されます。さらに、研究グループが開発した XAI の説明生成技術は、ゲノム医療にとどまらず、判断の「根拠」が求められる他分野での応用も期待できます。

現在、日本では全ゲノム解析等実行計画のもと、がんや難病等の全ゲノム解析が急速に普及しつつあります。本研究の XAI 技術と蓄積されたゲノムビッグデータを組み合わせることで、ゲノム医療の発展がさらに加速すると考えられます。本技術を応用すれば、将来的にゲノム解

析で得られたデータをその場で即座に解釈し、高精度な治療提案を行うシステムの実現も考えられます。

資金情報

本研究は、富士通株式会社との共同研究として行われました。

発表者

東京大学医科学研究所

附属ヒトゲノム解析センター健康医療インテリジェンス分野

井元 清哉 教授

附属先端医療研究センター造血病態制御学分野

南谷 泰仁 教授

小川 弥穂 特任研究員（東京大学医学部附属病院ゲノム診療部特任助教兼任）

高森 弘之 特任研究員

福島 英人 大学院生

附属病院血液・腫瘍内科

横山 和明 准教授

論文情報

〈雑誌〉Cancers (Basel)

〈題名〉Pathogenicity prediction of gene fusion in structural variations: a knowledge graph-infused explainable artificial intelligence (XAI) framework

〈著者〉Katsuhiko Murakami, Shin-Ichiro Tago, Sho Takishita, Hiroaki Morikawa, Rikuhiko Kojima, Kazuaki Yokoyama, Miho Ogawa, Hidehito Fukushima, Hiroyuki Takamori, Yasuhito Nannya, Seiya Imoto, Masaru Fuji

〈DOI〉: 10.3390/cancers16101915

〈URL〉: <https://doi.org/10.3390/cancers16101915>

用語解説

（注1）融合遺伝子：二つの異なる遺伝子が融合して形成された新規の遺伝子。がんの進行や転移において重要な役割を果たし、がん細胞の増殖を促進するドライバー変異として機能するものがある。

（注2）構造異常：DNAの大規模な変化を指し、DNAの一部が削除、複製、挿入、逆位、または転座すること。これらの変異はがんの発症や進行に関連することがある。

（注3）XAI：説明可能な人工知能。AIの予測結果の根拠を明示することで、ユーザーがその結果を理解しやすくする技術。

（注4）一塩基変異：DNA配列中の単一のヌクレオチド（塩基）の変化。多くのがん関連遺伝子変異の基本形であり、がんの診断や治療において重要な指標となる。

（注5）ドライバー変異：がんの発生や進行を直接引き起こす遺伝子変異。これに対して、がんの進行等に寄与しない変異は「パッセンジャー変異」と呼ばれる。

（注6）知識グラフ：データ間の関係性を示すネットワーク構造（AI分野ではグラフと呼ばれる）のデータベース。医療やゲノム解析においては、遺伝子や疾患、治療法の間に関連性をモデル化するために使用される。

（注7）機械学習：コンピュータがデータからパターンや規則を学習し、予測や判断を行うための数学的手法。XAIの基盤となる技術。

問合せ先

〈本件に関する問合せ〉

国立大学法人東京大学医科学研究所 附属ヒトゲノム解析センター

教授・センター長 井元 清哉（いもと せいや）

https://www.ims.u-tokyo.ac.jp/imsut/jp/lab/hgclink/page_00072.html

〈報道に関する問合せ〉

国立大学法人東京大学医科学研究所 プロジェクトコーディネーター室（広報）

<https://www.ims.u-tokyo.ac.jp/>