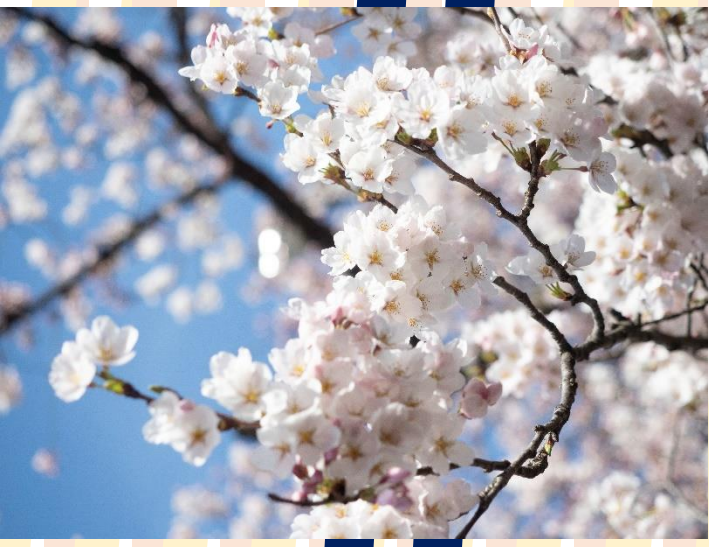


Unravelling origin of cancer and diversity by large-scale data analysis and artificial intelligence technology

大規模データ解析と人工知能技術による がんの起源と多様性の解明

文部科学省「富岳」成果創出加速プログラム

Newsletter

No. 1

2021年3月

「富岳」成果創出加速プログラム

大規模データ解析と人工知能技術による がんの起源と多様性の解明 ー１年を振り返って

課題代表者

東京医科歯科大学M&Dデータ科学センター 特任教授・センター長 宮野 悟

この課題の目標は、ゲノムシーケンスデータ解析パイプラインGenomon及びネットワーク解析アプリケーションとネットワーク深層学習システムDeep Tensor等により、正常組織においてどのように遺伝子変異クローンが生じるのか、遺伝子変異ないしその組み合わせがどのように細胞の表現型を決定するのか、さらには、その多様性・複雑性のために研究が進んでいないゲノムの構造異常が発がんにどう関わるのか、について解明することです。

本研究開発に先立ち、大規模シーケンスによる解析で、小川誠司等は、加齢に伴う正常組織の遺伝子異常とがん化のメカニズムを解明しています(***Nature***. 565(7739):312-317, 2019)（図１）。食道がんにおいて遺伝子変異が乳児期から獲得され、加齢とともに増加し70歳以上では全食道面積の40～80％が、がん遺伝子の変異をもった細胞で置き換わっていることを明らかにしました。がんがなぜ高齢者に生じ、「飲酒」や「喫煙」によって促進されるのかを解明し、がんが生ずる初期のメカニズムの解明の突破口を創った研究です。また、がんの早期診断、予防、がんの死亡率の低減に資する大きな社会的意義があります。

さらに、小川等は、潰瘍性大腸炎による上皮再構築メカニズムと発がんとの関係を解明しています(***Nature***. 577(7789):260-265, 2020)（図2）。潰瘍性大腸炎に着目し、長期間の炎症に暴露された大腸粘膜と、これを背景として発症する大腸がんの大規模なゲノム解析を通じて、潰瘍性大腸炎の長期罹患患者の大腸上皮は、大腸がんで認められる遺伝子変異の他に、炎症に関わるIL-17シグナル経路の遺伝子変異を獲得した細胞が増加し、直腸では全体の50～80%の面積がこれらの遺伝子変異を有する細胞によって置き換わっていることを証明しました。また、潰瘍性大腸炎患者の大腸粘膜で高頻度に認められるIL-17シグナル経路の遺伝子のうち、*NFKBIZ*（IκBζ）や*ZC3H12A*（Regnase-1）の変異は大腸がんではほとんど認められないことに着目し、これらの遺伝子に変異を獲得した上皮細胞は発がんしがたいことも証明しました。

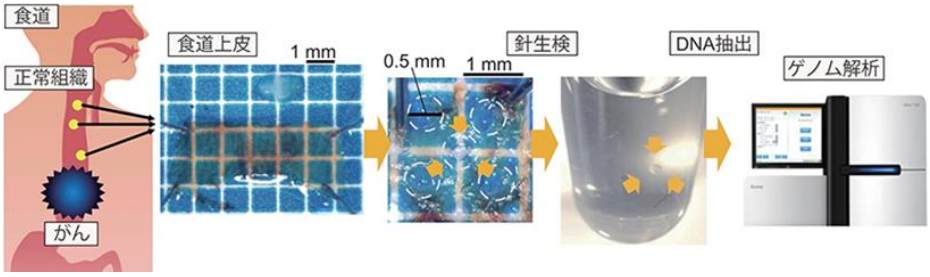


図1：小川誠司 医学研究科教授、横山顕礼 医学部附属病院特定助教、垣内伸之 医学研究科研究員、吉里哲一 同助教、武藤学 同教授、宮野悟 東京大学教授らの研究グループのプレスリリースより。
https://www.kyoto-u.ac.jp/ja/research-news/2019-01-08

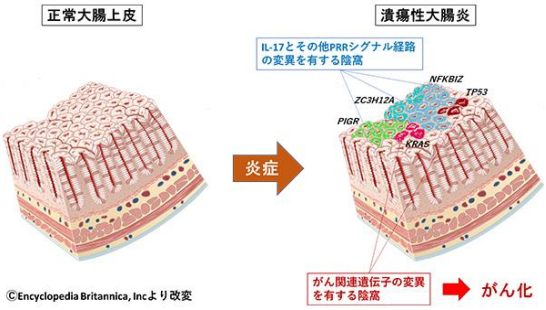


図2：小川誠司 医学研究科教授（兼・高等研究院ヒト生物学高等研究拠点（ASHBi）主任研究者）、垣内伸之 同助教（兼・同研究者）、妹尾浩 同教授、宮野悟 東京大学教授らの研究グループのプレスリリースより。
https://www.kyoto-u.ac.jp/ja/research-news/2019-12-20

また、だいぶ前の研究になりますが小川等は、骨髄異形成症候群から急性白血病を起こすクローン進化の遺伝子異常パターン解明しています(***Nature Genetics***. 49(2):204-212, 2017)（図3）。骨髄異形成症候群(MDS)が急性骨髄性白血病(AML)へ進行する際に、(a) 2つのタイプのゲノム異常が細胞に蓄積し異常クローンが増大すること、(b) 低リスクMDSに「タイプ 2 ゲノム異常」が起こると白血病リスクの高い状態に移行し、「タイプ 1 ゲノム異常」が起こると急性白血病が起こることを2,250例データからGenomonで解析し、解明しています。この急性白血病進行への2ステップのクローン進化を解明しました。この成果により、MDSの慢性状態で急性白血病の兆しを検出しその発症を早く予測し、早期の治療開始につなげ、治療効果を高めることが期待されます。

また、片岡圭亮等は、定説破るがん遺伝子の「複数変異」から新たな発がん機構を発見しています(***Nature***. 582(7810): 95-99, 2020)（図４）。6万例、150がん種以上を超える大規模ながんゲノムデータについて、スパコン＋Genomonで遺伝子解析を行い、同一がん遺伝子内における複数変異同士が相乗的にがん化を促進するという新たな発がん機構を発見しています。複数変異は、単独では比較的低頻度で機能的に弱い変異に集積しており、これまで意義不明であった変異が生じる理由が説明可能となったといえます。

こうした新事実が明らかになっていくなか、がんの理解が変わり始めました。同時に、その理解のためには大規模データ解析ができる計算環境が不可欠になっています。また、大規模データ解析の結果でてくる解析結果（ネットワークや変異のリスト、変異間の相互関係）も大規模で、その解釈は局所的にしかできなくなっています。上述の片岡等の研究では特定のがん遺伝子に着目して解釈をしています。これからは数理モデリングと人工知能技術などを活用することが求められています。

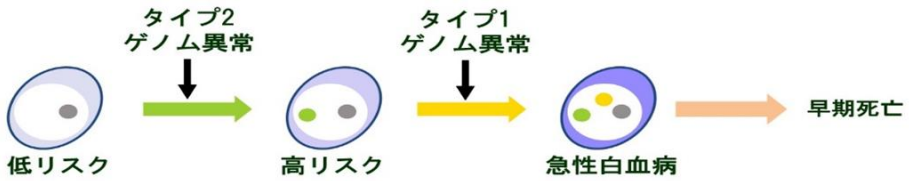


図3：京都大学腫瘍生物学講座 小川誠司 教授、東京大学医科学研究所ヒトゲノム解析センター 宮野悟 教授、独国ミュンヘン白血病研究所 Torsten Haferlach 教授、米国クリーブランドクリニックJaroslaw P. Maciejewski 教授を中心とする国際共同研究チームのプレスリリースより。
https://www.ims.u-tokyo.ac.jp/imsut/content/900004285.pdf

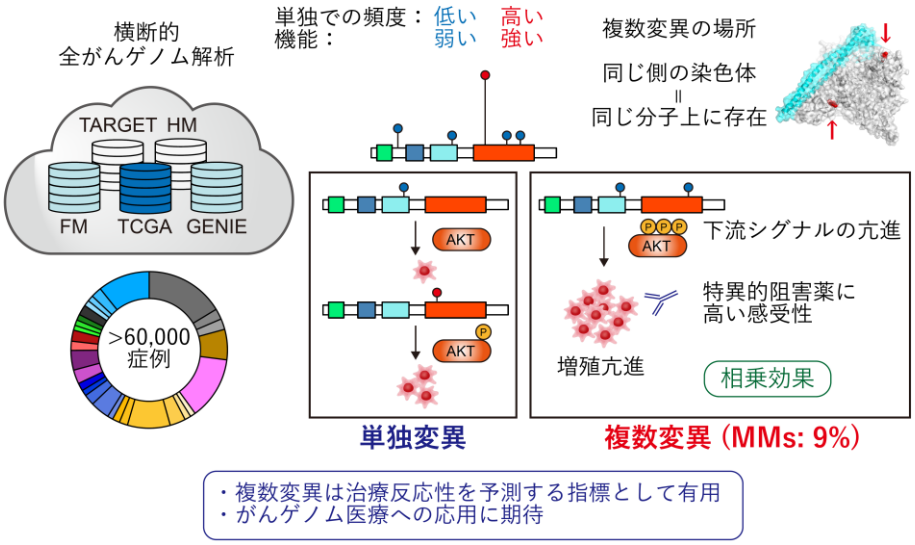


図4：国立研究開発法人国立がん研究センター（理事長：中釜 斉、東京都中央区） 分子腫瘍学分野 斎藤優樹任意研修生、古屋淳史主任研究員、片岡圭亮分野長らの研究グループ、京都大学大学院医学研究科 奥野恭史教授、東京大学医科学研究所 宮野悟教授らとの共同プレスリリースより。
https://www.ncc.go.jp/jp/information/pr_release/2020/0408/index.html

人工知能というとDeep Learningと考える人達が多いかもしれません。これは、画像や信号処理では威力を発揮しています。しかし、精度はよいがなぜか人間にわからないブラックボックスです。人間が解釈不能な大規模な解析結果は、説明可能なAI (XAI)でブレークスルーし、高度予測性と解釈性を同時に実現することが求められています。

本研究では富士通研究所人工知能研究所で開発された、ネットワークを入力とするDeep Tensorをベースに説明可能な人工知能技術をつくってきました。Deep Tensorは「富岳」に実装しました。

ネットワーク推定技術は、「次世代スパコンプロジェクト」時代から開発してきたSiGNシリーズを「富岳」に実装しました。解析例として上皮間葉転換（上皮系の細胞が間葉系の細胞に変わる現象で、がんの転移や薬剤耐性の獲得などに関わっていると考えられているものです）というがんのメカニズムと多様性に関して、これまでに開発した技術で、「最近の10年以上に渡る分子生物学的な結果がSanger Instituteの10年前のデータから一挙に出現」するなど、今後の応用に大きな期待感がでてきています（図5）。

一方、ゲノムシーケンスデータ解析パイプラインGenomonの「富岳」での実行は、モグラたたきのようにあとからあとから「富岳」のシステム側で問題がでてきており、理化学研究所の石川裕先生方と富士通といっしょになって解決を目指しているところです。

研究体制は、東京医科歯科大学M&Dデータ科学センターのチーム、京都大学医学研究科腫瘍生物学講座の小川誠司教授のチーム、愛知県がんセンター研究所システム解析学分野の山口類分野長、そして連携機関として富士通研究所人工知能研究所と東京大学医科学研究所ヒトゲノム解析センターからなります。図6の作戦で攻めています。

2020年4月、COVID-19の感染が拡大する中、この「富岳」成果創出加速プログラムは開始しました。課題代表者が、東京大学医科学研究所から東京医科歯科大学に新設されたM&Dデータ科学センターに異動したこともあり、研究環境の整備に多くの時間と労力を使いました。Face to Faceのミーティングが難しいためZoomによるミーティングで研究打ち合わせや外部有識者のご助言などをいただきながらプロジェクトを進めてきました。ニュースレター発行も先送りされ続け、この1年を総括する形でこの第一号の発行にたどり着いたところです。

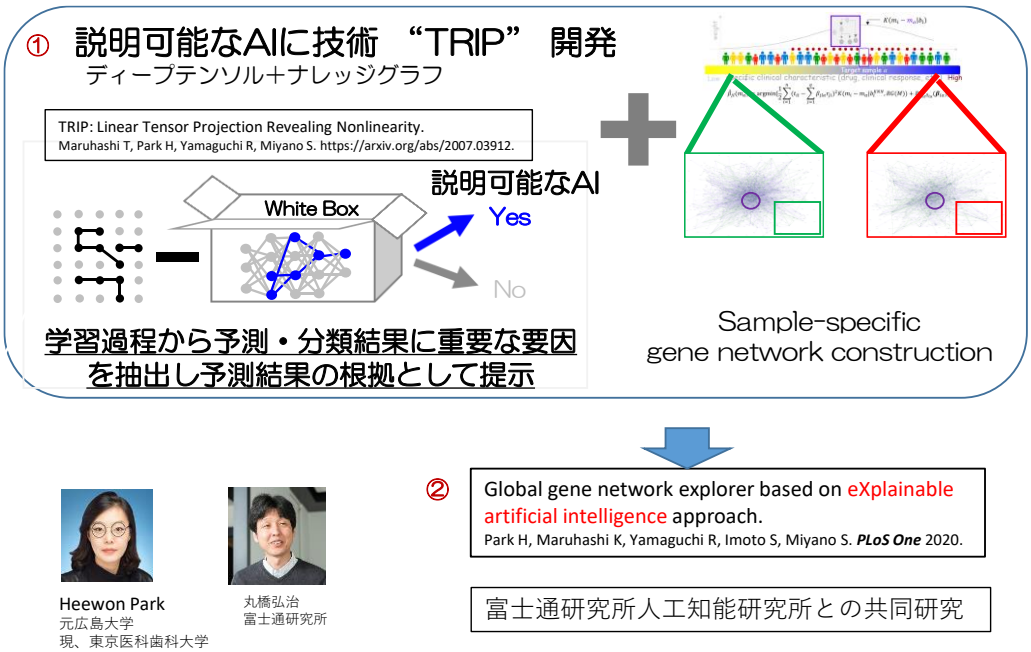


図5：Deep Tensor+TRIPという説明可能なAIを開発し、10年以上に渡る分子生物学的な結果が10年前のデータから一挙に出現

1. 代表機関

国立大学法人東京医科歯科大学

研究課題開発責任者

M&Dﾃﾞｰﾀ科学ｾﾝﾀｰ　特任教授　宮野悟

2。協力機関

国立大学法人京都大学

協力機関代表者

大学院医学研究科　教授　小川誠司

愛知県がんセンター

協力機関代表者

研究所ｼｽﾃﾑ解析分野　分野長　山口類

3. 連携機関

(株)富士通研究所

連携機関連絡担当者

人工知能研究所　所長　岡本青史

国立大学法人東京大学

連携機関連絡担当者

医科学研究所　助教　片山琴絵

Karolinska Institutet

連携機関連絡担当者

Department of Medicine/Center for Hematology and Regenerative Medicine

Visiting Professor 小川誠司

4. 事業参加者

事業参加者(代表機関)

東京医科歯科大学M&Dﾃﾞｰﾀ科学ｾﾝﾀｰ

統合解析分野　特任教授　宮野悟

講師　長谷川嵩矩

特任助教　伊東聰

特任助教 角田将典

研究支援者 田中洋子

AI技術開発分野　教授　Heewon Park

事業協力者(協力機関)

京都大学大学院医学研究科

腫瘍生物学講座　教授　小川誠司

特定准教授　南谷泰仁

助教　越智陽太郎

愛知県がんセンター研究所

ｼｽﾃﾑ解析分野　分野長　山口類

連携参加者(連携機関)

(株)富士通研究所　人工知能研究所

所長　岡本青史

主任研究員　丸橋弘治

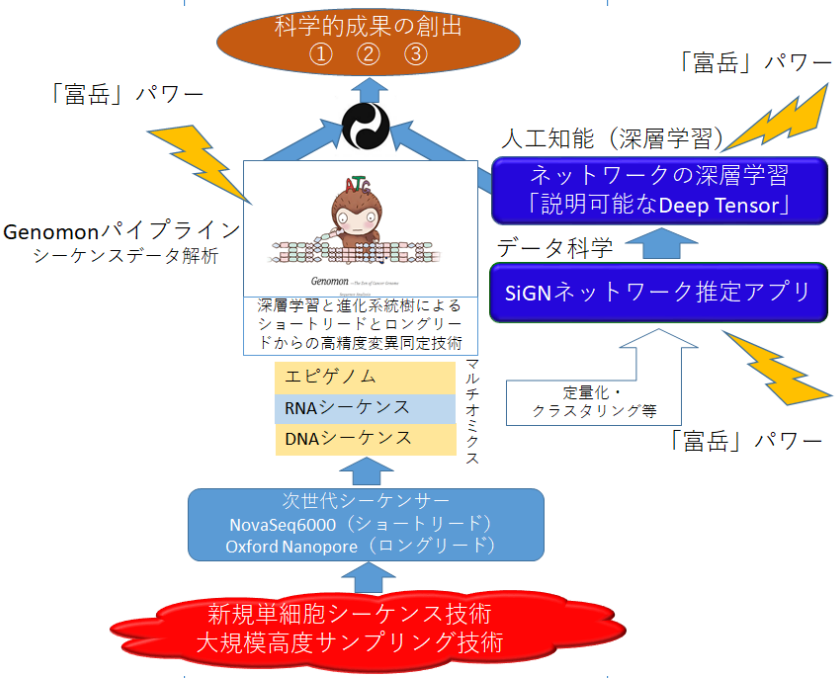
東京大学医科学研究所

ヒトゲノム解析センター　助教　片山琴絵

(2021年3月22日時点)

スーパーコンピュータ「富岳」成果創出加速プログラムについて（理化学研究所計算科学研究センターより抜粋）
https://www.r-ccs.riken.jp/jp/fugaku/promoting-researches

スーパーコンピュータ「富岳」成果創出加速プログラムは、「富岳」を用いた成果を早期に創出することを目的として文部科学省が設置しました。①人類の普遍的課題への挑戦と未来開拓、②国民の生命・財産を守る取組の強化、③産業競争力の強化、④研究基盤の4領域について課題の選定が行われ、19課題が選定されています。選定された課題は、「富岳」の計算資源を優先的に無償で使用し、速やかな成果創出を目指します（2020～2022年度）。



① 発がんのごく初期のクローン選択の過程とそのメカニズム
② がんがその後、どのようにして、さらなる遺伝子変異を獲得しつつ高度な多様性を獲得し、浸潤や転移、治療抵抗性や再発を来すのか
③ これまで同定された多数の遺伝子変異がどのように腫瘍の病型、その他の臨床像を含めた表現系を規定しているのか

図6：各技術要素のつながりから成果創出へ

Information

文部科学省「富岳」成果創出加速プログラム
課題名：大規模データ解析と人工知能技術によるがんの起源と多様性の解明
ニュースレター No.1
発行日★2021年3月22日（初版）
課題代表者★宮野 悟

- 東京医科歯科大学M&Dデータ科学センター　統合解析分野
- 郵便物宛先：　〒113-8510　東京都文京区湯島1-5-45
- 宅配便宛先：　〒101-0062　東京都千代田区神田駿河台2-3-10
  駿河台キャンパス22号館5階
- TEL: 03-5280-8620 FAX: 03-5280-8632
- E-mail: mdsc.dsc@tmd.ac.jp
- HP: https://odcla.mddsc.jp