

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



large-scale data analysis and artificial intelligence technology

Unravelling origin of cancer and diversity by

Newsletter

No. 4

2022年1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

クローン性造血の臨床予後への影響を解明 — 遺伝子変異とコピー数異常の統合的な知見 —

課題参加者 京都大学医学研究科腫瘍生物学講座 教授 小川誠司
Karolinska Institutet, Visiting Professor
Department of Medicine/Center for Hematology and Regenerative
Medicine



血液がんの前がん病変であるクローン性造血という状態は、血液がんだけでなく心血管疾患のリスク上昇とも関連し、大きく注目されています。過去の研究で、クローン性造血は遺伝子変異とコピー数異常(CNA)という2種類のゲノム異常で引き起こされることが報告されていました。しかし、両方の異常を同時に調べた研究はなく、その関係性は不明のままです。

本研究課題のメンバーである京都大学大学院医学研究科・腫瘍生物学講座 小川誠司 教授、佐伯龍之介 同博士課程学生、理化学研究所・生命医科学研究センター・基盤技術開発研究チーム 桃沢幸秀 チームリーダー、東京大学医科学研究所・人

癌病因遺伝子分野 村上善則 教授、および本研究課題代表者である東京医科歯科大学M&Dデータ科学センター 宮野悟センター長らを中心とする研究チームは、約1万人の健常者でクローン性造血の統合解析を実施しました。スーパーコンピュータを駆使して、膨大なシーケンスデータを解析しました。その結果、

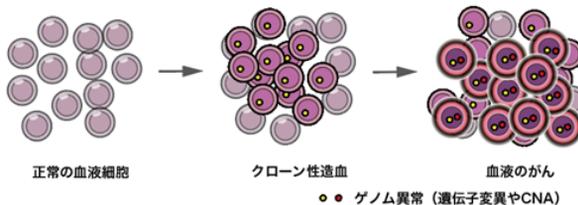
- (1) クローン性造血では遺伝子変異とCNAが高頻度に共存すること、
- (2) 両者が共存すると血液腫瘍・心血管疾患のリスクが有意に上昇すること、

などを解明しました。この研究は、がんの起源と多様性の一面を真正面から解明し

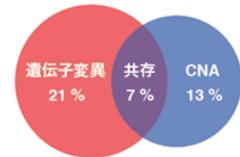
たものといえます。そして、その成果は、クローン性造血に基づく疾患の予後予測の重要な指標となる事が期待されます。

本研究のなかで強く認識されたことは、計算能力だけでなく、強靱なファイルI/O機能を有した大規模データ解析に適したスーパーコンピュータが、がんの起源と多様性の解明、さらにはがんの理解には不可欠であるという事実です。また、人海戦術的な数理解析の限界も見えており、人工知能技術などによりデータの理解を支援する必要性も生じています。

血液がんの前がん病変としてのクローン性造血

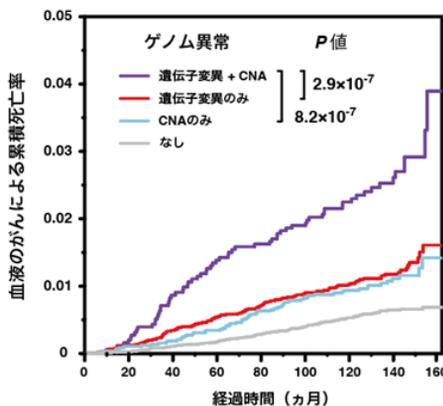


遺伝子変異とCNAの有意な共存



(健常者11,234人中における割合)

血液がんによる死亡率の上昇



心血管疾患による死亡率の上昇

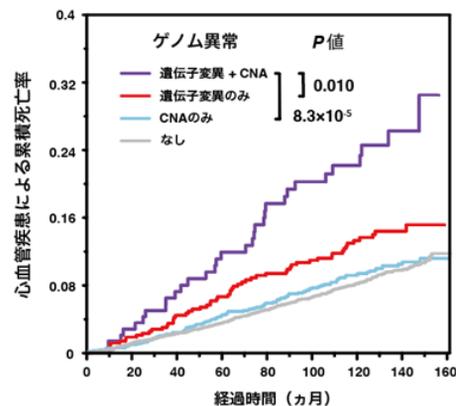


図1: 本研究成果をまとめたものです。

1. この研究の背景

次世代シーケンサーによる大規模な遺伝子解析により、血液のがん(急性骨髄性白血病や骨髄異形成症候群)は特定のゲノム異常によって引き起こされることを、本研究グループは過去に報告しました。しかし、近年の研究によってゲノム異常は健常者の血液でも一定の頻度で検出されることが判明し、「クローン性造血」として注目を集めています。クローン性造血は血液がんの前がん病変と考えられていますが、興味深いことに動脈硬化の進展にも関わっていることが報告されており、高齢者における主要な予後規定因子として認識され始めています。クローン性造血で検出される主なゲノム異常は、大きく遺

伝子変異(特定の遺伝子の配列が一部変化することで、クローン性造血では、*DNMT3A*、*TET2*などの遺伝子で最も高頻度に遺伝子変異が現れています)とコピー数異常(ヒトゲノムを構成する染色体の本数が増加・減少する現象)の2種類に分類されます。しかし、既存の研究で両者は別々に検討されてきたため、クローン性造血における両者の関係性については十分な理解が得られていませんでした。

2. 解析した検体と見えてきたゲノム異常

クローン性造血における遺伝子変異とコピー数異常の関係性を明らかにするため、11,234人の健常者の末梢血サンプルを用いて、遺伝子変異とコピー数異常の統合解析を実施しました。

その結果、60歳以上の健常者の40%でいずれかの異常が検出されました(図2、左上)。また、被験者の7%(陽性例の16%)では両方の病変を認め、統計学的に有意な共存を示すことから両者の協調的な関係が示唆されました(図2、左下)。

特に、*DNMT3A*、*TET2*、*JAK2*、*TP53*などの遺伝子では遺伝子変異とコピー数異常が頻繁に共存し、遺伝子の両アレル異常(本来2つ存在する遺伝子のペアの両方ともに異常を来した状態)を来していると考えられました(図2、右)。

これらの遺伝子の両アレル異常は血液がんでは既知の所見ですが、今回の研究によりクローン性造血の段階から存在することが判明しました。

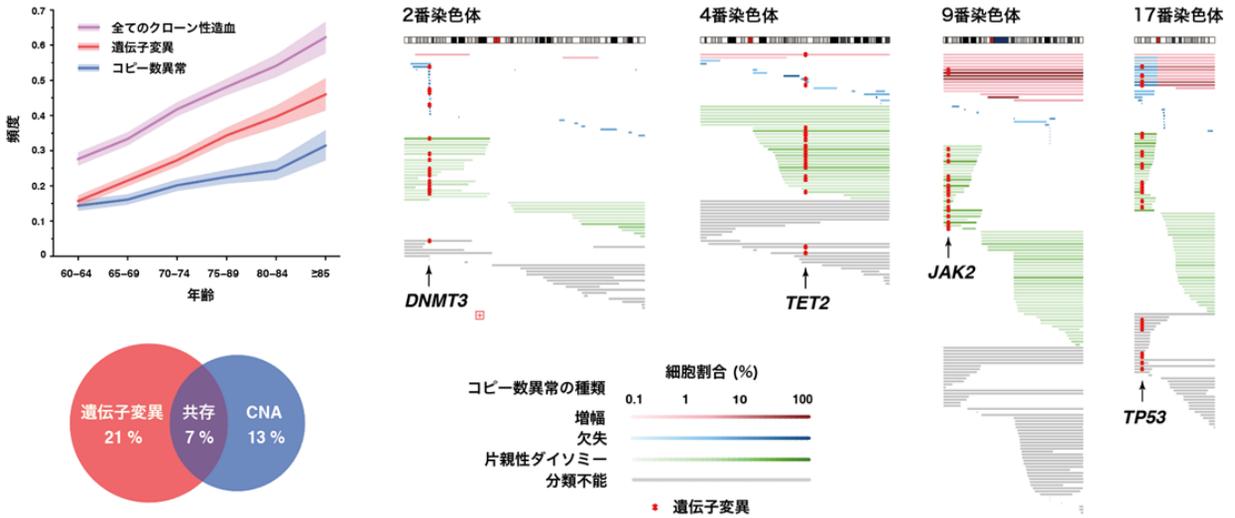


図2: クローン性造血における遺伝子変異とコピー数異常の共存関係

血液がんのリスクについては、過去の報告と一致して遺伝子変異あるいはコピー数異常を有する例では有意なリスクの上昇を認めましたが、血液がんの中でも骨髄系腫瘍ではリンパ系腫瘍に比べて強い

リスクの上昇を認めました(図3、左)。また遺伝子変異とコピー数異常が共存している例では、単独の場合に比べて血液がんによる死亡率が上昇し、両者が協調的に血液がんの発症に関わっていることが

示唆されました(図3、中央)。特に、両者が同一の遺伝子(*DNMT3A*、*TET2*、*TP53*、*JAK2*など)を標的として両アレル異常を起こしている場合には、さらにリスクが上昇することが分かりました(図3、右)。

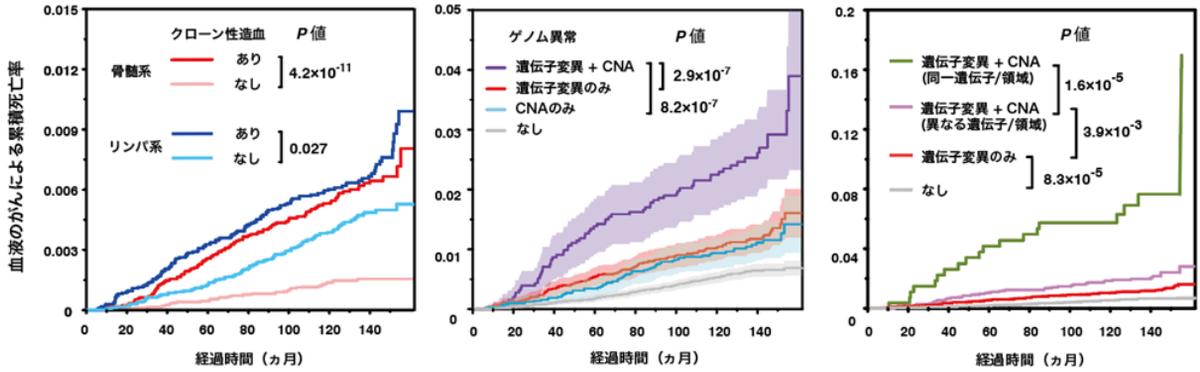


図3: クローン性造血が血液がんのリスクに与える影響

心血管疾患については、既存の報告の通り遺伝子変異が検出された例では有意にリスクが上昇していました。加えて、遺伝子変異が検出された例では高血圧を高頻度に認め(図4、左)、クローン性造血が動脈硬化を促進するという定説をさらにサポートする結果となりました。また、遺伝

子変異とコピー数異常が共存する例では、単独の場合に比べて心血管疾患による死亡率が上昇し、心血管疾患の発症についても2種類の病変が協調的に作用していることが類推されました(図4、右)。

これらの結果から、遺伝子変異とコピー数異常は血液がん・心血管疾患のリスクに

協調的に作用していて、クローン性造血について正確な理解を得るためには両者を統合的に評価することが重要であると考えられました。

高血圧に対するオッズ比 (多変量解析)		P 値
遺伝子変異	1.13 (1.03 - 1.25)	0.026
BMI (+5)	1.42 (1.30 - 1.56)	2.6×10^{-14}
年齢 (+10 年)	1.38 (1.26 - 1.51)	2.6×10^{-12}
赤血球数 (+50)	1.20 (1.06 - 1.36)	0.0037
ヘモグロビン (+1 g/dL)	1.14 (1.00 - 1.29)	0.043
ヘマトクリット (+1%)	0.94 (0.89 - 0.98)	0.010
血小板数 (+10 ⁹ /uL)	1.11 (1.02 - 1.21)	0.016
LDH (+100 IU/L)	1.06 (1.01 - 1.11)	0.025
クレアチニン (+1.0mg/dL)	1.25 (1.17 - 1.33)	7.5×10^{-11}
コレステロール (+50mg/dL)	1.14 (1.06 - 1.23)	6.6×10^{-4}
血糖 (+50mg/dL)	1.10 (1.04 - 1.16)	5.6×10^{-4}

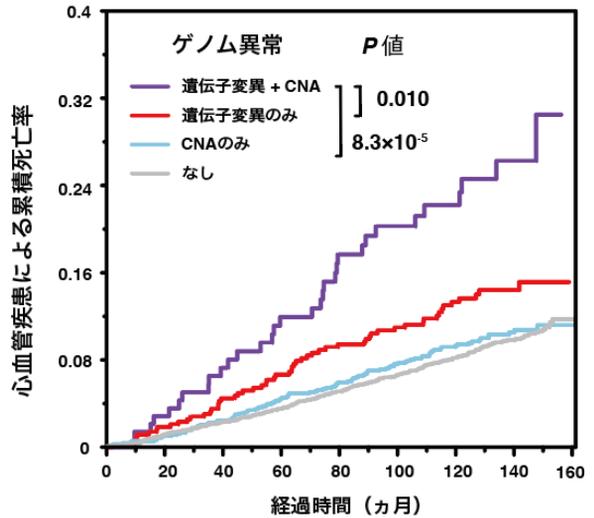


図4:クローン性造血が心血管疾患のリスクに与える影響

3. 波及効果と今後の予定

本研究の成果は、血液がんの起源を理解するための手がかりを与えるのみならず、クローン性造血に基づく臨床予後の予測を実現する上で重要な指標となることが期待されます。

4. 研究者からのコメント

遺伝子変異とコピー数異常の両者を含むクローン性造血の高精度かつ網羅的なゲ

ノム解析は、本研究で初めて実現いたしました。その結果、クローン性造血の中でも特にどのような場合にリスクが高くなるのか、詳細な検討を行うことができました。本研究の成果が、今後の疾患の予防や治療の発展に貢献できればと考えています。(京都大学医学研究科 腫瘍生物学講座・教授 小川 誠司、同博士過程学生 佐伯 龍之介)。

5. 本成果は次の論文に発表されました

Saiki R, Momozawa Y, Nannya Y, Nakagawa MM, Ochi Y, Yoshizato T, Terao C, Kuroda Y, Shiraiishi Y, Chiba K, Tanaka H, Niida A, Imoto S, Matsuda K, Morisaki T, Murakami Y, Kamatani Y, Matsuda S, Kubo M, Miyano S, Makishima H, Ogawa S. Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. (参加者:ポールド)

Nature Medicine. 2021 Jul;27(7):1239-1249. doi: 10.1038/s41591-021-01411-9. Epub 2021 Jul 8.

編集者メモ

- 2022年3月3日(木)10:00-12:00:「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」シンポジウム(仮題)をオンラインで開催予定。
- 2022年1月:田中洋子さんが東京医科歯科大学M&Dデータ科学センター統合解析分野の特任研究員になりました。田中さんは、2009年より、国際がんゲノムコンソーシアムでがんの全ゲノムシーケンスデータの解析をパソコンでやってこられました。
- 2021年10月28-29日:第8回HPCIシステ

- ム利用研究課題成果報告会 (hp200138)にオンライン参加しました。
- 2021年10月:小川弥穂先生が東京医科歯科大学M&Dデータ科学センター統合解析分野の特任助教に東京大学より着任しました。専門は血液内科です。東京大学医科学研究所では、がんゲノム医療にたずさわってきました。南谷先生と連携しています。
- 2021年9月:長谷川嵩矩先生が東京医科歯科大学M&Dデータ科学センター統合解析分野の准教授に昇進しました。

- 2021年4月:南谷泰仁先生が、東京大学医科学研究所先端医療研究センターに異動しました。京都大学大学院医学研究科腫瘍生物学講座も兼務しています。
- 2021年4月:(株)富士通研究所が富士通株式会社富士通研究所となりました。連携参加者の岡本青史さん(東京医科歯科大学M&Dデータ科学センター客員教授)は富士通研究所フェローになりました。丸橋弘治さん(東京医科歯科大学M&Dデータ科学センター客員准よ教授)は富士通研究所プロジェクトマネージャになりました。

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データ科学センター 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
富士通株式会社 富士通研究所フェロー
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 准教授 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データ科学センター
統合解析分野 特任教授 宮野悟
准教授 長谷川嵩矩
助教 伊東聰
特任助教 角田将典
特任助教 小川弥穂
特任研究員 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
特定教授 南谷泰仁
助教 越智陽太郎
他11名

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)
富士通株式会社 富士通研究所フェロー
岡本青史
富士通研究所 プロジェクトマネージャ
丸橋弘治

東京大学医科学研究所
ヒゲノム解析センター 准教授 片山琴絵
先端医療研究センター 教授 南谷泰仁
(2022年1月1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)。

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
課題番号: JPMXP1020200102

課題ID
2021年度: hp210167
2020年度: hp200138

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. 4
発行日★2022年1月4日
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>