

**PRESS RELEASE**

理化学研究所  
 東京大学  
 日本医療研究開発機構  
 静岡県立総合病院  
 静岡県立大学

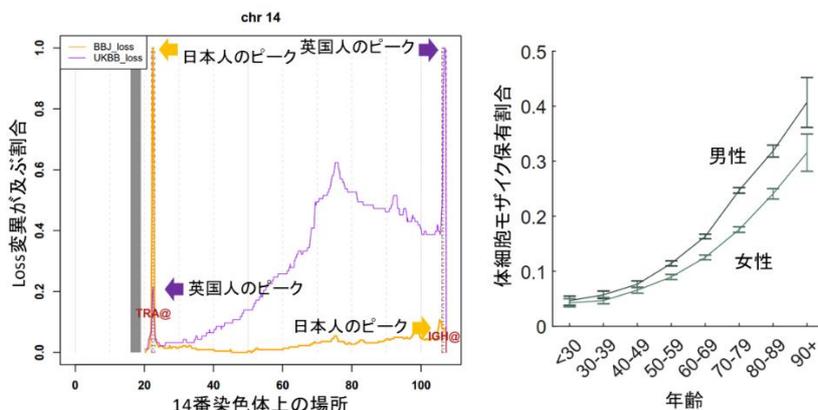
**日本人特有の白血病発症メカニズムの解明へ**  
**ーバイオバンク・ジャパンデータベースの活用による成果ー**

理化学研究所生命医学研究センターゲノム解析応用研究チームの寺尾知可史チームリーダー（静岡県立総合病院免疫研究部長、静岡県立大学特任教授）、鎌谷洋一郎客員主管研究員（東京大学大学院新領域創成科学研究科教授）らの国際共同研究グループ<sup>※</sup>は、血中で後天的な DNA 変異を持つ白血球がクローン性に増殖する<sup>[1]</sup>ことで、生まれながらの DNA 配列と変異した配列が混ざって見える現象（体細胞モザイク<sup>[2]</sup>）を解析し、加齢に伴う DNA 変異と体細胞モザイク出現はほぼ不可避であること、体細胞モザイクが白血病をはじめとするがん化メカニズムに影響を与えることを明らかにしました。さらに、体細胞モザイクは全死亡率の10%の上昇と関連することも分かりました。

今回、国際共同研究グループは、日本最大級の DNA データベースであるバイオバンク・ジャパン（BBJ）<sup>[3]</sup>の登録者約 18 万人の DNA マイクロアレイ<sup>[4]</sup>のデータを解析し、従来、生まれつきの変異の同定にのみ使われてきたデータの中から後天的 DNA 変異の存在を表す体細胞モザイクを検出しました。そして、体細胞モザイク出現に関連する遺伝的多型<sup>[5]</sup>を同定し、その分子機構を明らかにしました。また、血液悪性腫瘍では、がん化を反映する変化が発症前の段階から起きていることを見いだしました。さらに、日本人に多い白血病発症に関連する変異を同定し、イギリス人に多い白血病発症に関連する変異が日本人にはあまり見られないことも明らかにしました。

今後、これらの知見に加えて、新たに全ゲノムシーケンス解析<sup>[6]</sup>を行い情報を集約させることにより、日本人に特有の機構を含め、老化やがん化メカニズムのさらなる詳細な解明や、生命予後の予測を可能とする臨床医学の発展につながると期待できます。

本研究は、科学雑誌『Nature』の掲載に先立ち、オンライン版（6月24日付：日本時間6月25日）に掲載されました。



日本人特異的体細胞モザイク（左）と加齢に伴う体細胞モザイク保有割合の上昇（右）

## ※国際共同研究グループ

### 理化学研究所

#### 生命医科学研究センター

##### ゲノム解析応用研究チーム

- チームリーダー 寺尾 知可史 (てらお ちかし)  
(静岡県立総合病院 臨床研究部 免疫研究部長、  
静岡県立大学 薬学部 ゲノム病態解析講座 特任教授)  
客員主管研究員 鎌谷 洋一郎 (かまたに よういちろう)  
(東京大学 大学院新領域創成科学研究科  
メディカル情報生命専攻 複雑形質ゲノム解析分野 教授)  
客員研究員 秋山 雅人 (あきやま まさと)  
(九州大学 医学部 眼科 特任講師)  
客員研究員 石垣 和慶 (いしがき かずよし)  
(ハーバード大学 ポスドクフェロー)

##### 自己免疫疾患研究チーム

- チームリーダー 山本 一彦 (やまもと かずひこ)  
副チームリーダー 鈴木 亜香里 (すずき あかり)

##### 基盤技術開発研究チーム

- チームリーダー 桃沢 幸秀 (ももざわ ゆきひで)

#### 統合生命医科学研究センター (研究当時)

- 副センター長 久保 充明 (くぼ みちあき)

#### 東京大学 大学院新領域創成科学研究科 メディカル情報生命専攻

##### クリニカルシーケンス分野

- 教授 松田 浩一 (まつだ こういち)

#### 東京大学 医科学研究所

##### 癌・細胞増殖部門 人癌病因遺伝子分野

- 教授 村上 善則 (むらかみ よしのり)

#### ブリガムアンドウィミンズ病院 遺伝学分野

- 助教 ポールー・ロー (Po-Ru Loh)

#### ハーバード大学

- 教授 スティーブン・マッカロール (Steven A Maccarroll)

## 研究支援

本研究は、日本医療研究開発機構 (AMED) のオーダーメイド医療の実現プログラム「疾患関連遺伝子等の探索を効率化するための遺伝子多型情報の高度化 (研究開発代表者: 久保充明 (当時))」の支援のもと行われました。本研究で使用したサンプルは、「オーダーメイド医療の実現プログラム」において収集されたものです。

## 1. 背景

ヒトの体を構成する細胞は、一つの受精卵が細胞分裂を繰り返してできたもので、細胞分裂のたびにゲノム DNA が複製されます。その過程で DNA に障害が起きることがあり、多くの場合は修復機構が働きますが、なかには後天的な DNA 変異として残るものが生じます。悪性腫瘍は、このような後天的 DNA 変異が生

じた結果、先天的な要素と相まって、正常な機能を失った細胞がクローン性に増殖する疾患であることが分かっています。

また、特に悪性腫瘍患者ではない一般的な高齢者の血液中で、特定の遺伝子に後天的 DNA 変異が生じた、とりわけ異常もないであろう細胞のクローン性増殖がしばしば見られます。以前より、これらの細胞の増殖と、その後の血液悪性腫瘍や心臓疾患発症リスクとの関連が指摘されていました。

一方、後天的 DNA 変異は、染色体ごとあるいは染色体の幅広い領域で生じることもあります。このような変異のある細胞とない細胞が存在する状態を「体細胞モザイク」と呼びますが、体細胞モザイクの疾病に対する影響については、2018年に英国の UK バイオバンク (UKB) <sup>[7]</sup>が報告したのみで<sup>注1)</sup>、詳しいことは分かっていませんでした。さらに、こうした研究の多くは欧州系集団で行われており、非欧州系集団における研究が非常に少ないことが、近年懸念されています<sup>注2)</sup>。ゲノム配列のパターンは集団により異なるため、日本人集団で後天的 DNA 変異を調べることは、日本や東アジアにおける疾患の理解や、今後の予防・治療法への応用につながると期待されます。

注1) Loh PR et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature*. 2018 Jul;559(7714):350-355.

注2) Sirugo G et al. The Missing Diversity in Human Genetic Studies. *Cell* 2019; 177: 26-31.

## 2. 研究手法と成果

国際共同研究グループは、現在データ数が最も多い遺伝子データである DNA マイクロアレイのデータから体細胞モザイクを同定する手法を確立しました。検出できる体細胞モザイクは、次の3種類です (図1)。

1. 染色体の一部または全体が欠損する状態 (loss)
2. 染色体の一部または全体が2本とも片親由来となる状態 (copy-neutral loss of heterozygosity ; CN-LOH)
3. 染色体の一部または全体がコピーとして増える状態 (gain、2コピーが3コピー、あるいはそれ以上になる)

そして、バイオバンク・ジャパン (BBJ) の登録者約18万人のDNAマイクロアレイの常染色体データを解析し、3.3万個の体細胞モザイクの同定に成功しました。この同定の割合 (約18.3%) は、2018年のUKBデータの場合の4倍以上です。また、3種類の体細胞モザイクのうち、CN-LOHの数が最も多いことが分かりました。

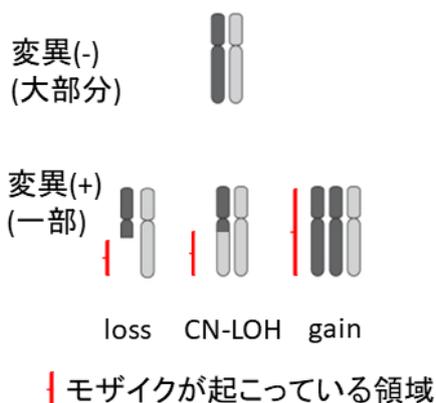


図1 体細胞モザイクの種類

本研究で同定可能な3種類の体細胞モザイク。赤の部分が、体細胞モザイクが生じている遺伝子領域。lossは染色体の一部または全体が欠損する状態、CN-LOHは染色体の一部または全体が2本とも片親由来となる状態、gainは染色体の一部または全体がコピーとして増える状態である。

また、加齢に伴って体細胞モザイクの保有割合が上昇し、90歳以上では約35%の人が体細胞モザイクを保有することが分かりました(図2)。この結果は、体細胞モザイクの保有が珍しいことではなく、加齢によって一般的な集団においても保有割合の上昇がほぼ不可避であることを強く示唆しています。UKBのデータでは高齢の登録者が少なく、この結果は日本人のバイオバンクだからこそ分かったといえます。また、女性よりも男性の方が(図2)、非喫煙者よりも喫煙者の方が、体細胞モザイク保有割合が高いことも分かりました。

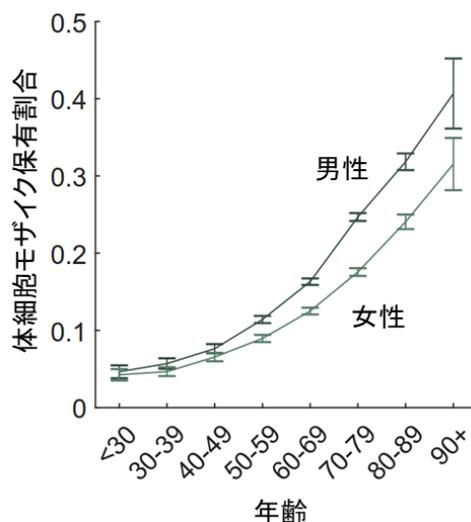


図2 加齢に伴う体細胞モザイク保有割合の上昇

加齢に伴って体細胞モザイクの保有割合が上昇するが、一貫して男性の方が高い。90歳以上の男性では、約41%の人が体細胞モザイクを保有する。

次に、体細胞モザイクがどの遺伝子に、あるいはどの遺伝子を含んだ領域に、またどの染色体に生じているかを解析し、2020年のイギリス人データと比較しました。その結果、白血球の一部のリンパ球であるT細胞<sup>[8]</sup>の受容体領域(TCR<sup>[8]</sup>)とB細胞<sup>[8]</sup>の受容体領域(BCR<sup>[8]</sup>、免疫グロブリン)において、日本人ではTCRの体細胞モザイク(loss)が多く、BCRの体細胞モザイク(loss)が少ないことが分かりました(図3)。

TCRとBCRは、T細胞とB細胞がさまざまな抗原と結合・認識できるように、遺伝子組換えが起こる領域です。そのため、もしリンパ球細胞のみを解析し、各細胞レベルでこれらの遺伝子領域に注目した場合、T細胞でTCRに、B細胞でBCRに構造変化(この場合はloss)が生じることは当然のことです。しかし、今回の解析は白血球全体のDNAを解析したものであり、構造変化の同定は、それらの構造変化を持つ細胞の検出可能レベル以上の増殖を意味します。すなわち、日本人ではT細胞の一部がクローン性に増殖し、検出可能レベル以上に達した人が多く見られた一方で、B細胞の一部が検出可能レベル以上にクローン性に増殖した人は少なかったといえます。

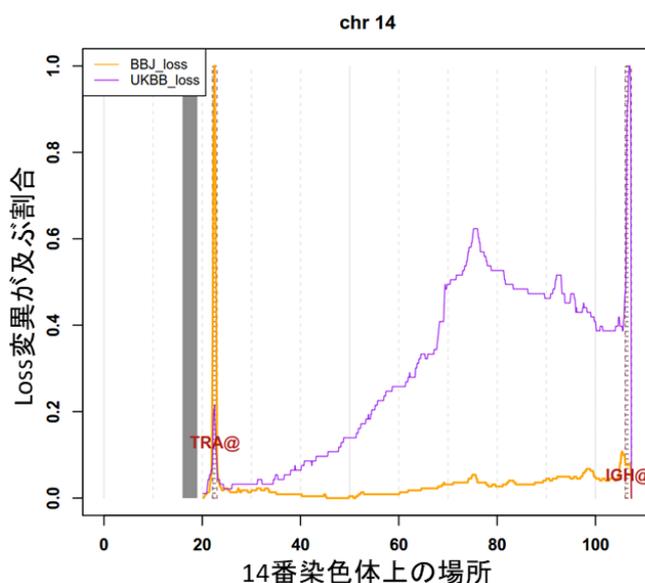


図3 人種特異的なT細胞およびB細胞の増殖

T細胞の受容体領域TCR(TRA@)、B細胞の受容体領域BCR(IGH@)における、人種特異的な体細胞モザイク(loss)の割合。TCRは日本人(黄色線)、BCRはイギリス人(紫線)でlossが多いことが分かった。

白血病に代表される血液悪性腫瘍の内訳には、人種による特徴があります<sup>注3-4)</sup>。日本人ではT細胞系の悪性腫瘍が多く、B細胞系の悪性腫瘍は少ないこと、欧州人では、慢性リンパ球性白血病(CLL)というB細胞系の血液悪性腫瘍が最も多いことが分かっています。さらに、2018年のUKBデータから、特定の体細胞モザイクを持つと、将来のCLL発症リスクが大きく上昇することが示されています。したがって、今回明らかになった体細胞モザイクのTCRとBCRにお

る人種差は、これら人種差のある血液悪性腫瘍の前段階を反映している可能性があります。

さらに、染色体レベルでも、血液悪性腫瘍の人種差を支持する結果を得ました。各染色体における体細胞モザイクの相対的な割合は、日本人とイギリス人で非常に似ていましたが、一部日本人に少なく、イギリス人に多い体細胞モザイク（あるいはその逆）が見られました（図4）。特に、日本人に少なくイギリス人に多い三つの体細胞モザイク（13番染色体長腕 loss、13番染色体長腕 CN-LOH、12番染色体 gain）は、いずれも CLL 患者に見られる構造変化であり、さらに2018年のUKBデータでCLL発症リスクを大きく上昇させる体細胞モザイクでした。このことは、血液悪性腫瘍発症の数年前から責任細胞のクローン性増殖があり、その前がん段階ともいふべき状態の時点で、その責任細胞は血中で検出可能であること、さらに前がん段階で人種差が既に存在していることを意味しています。

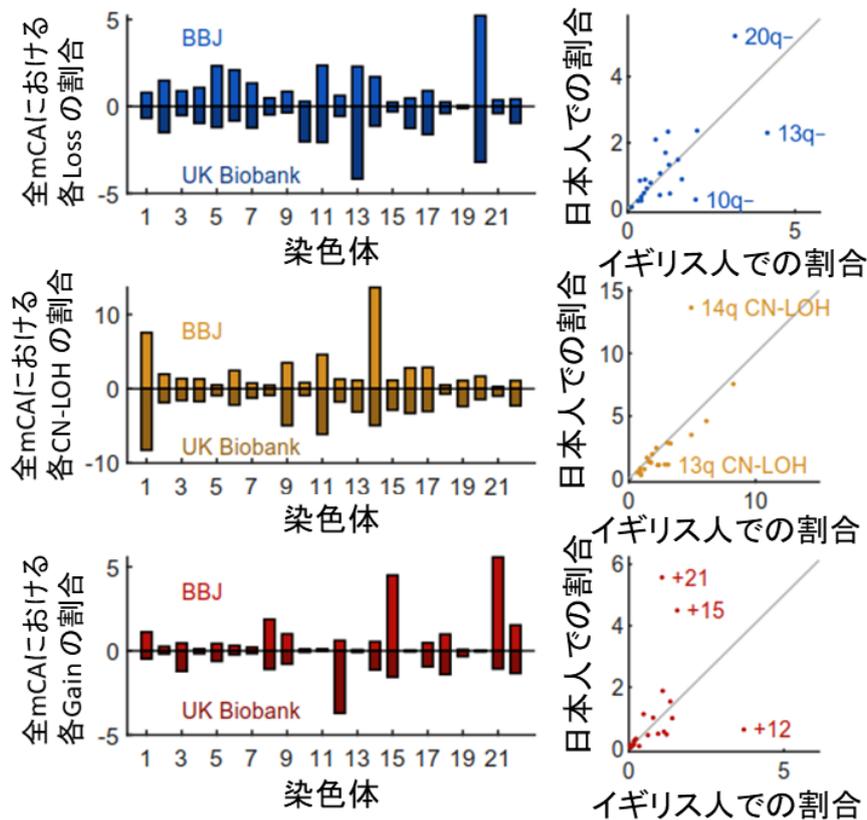


図4 日本人とイギリス人の体細胞モザイクの相対割合の相違

日本人(BBJ)とイギリス人(UKB)で見られた各染色体における体細胞モザイク(mCA)の割合(百分率)。Loss(上段)、CN-LOH(中段)、gain(下段)ともに両人種で良く一致しているが、一部には人種特異的な増減が見られる。イギリス人に多くみられる13番染色体長腕のloss(13q-)、13番染色体長腕のCN-LOH(13qCN-LOH)12番染色体のgain(+12)は慢性リンパ球性白血病(CLL)リスクである。

- 注3) Iwanaga, M. *et al*, Adult T-cell leukemia: a review of epidemiological evidence. *Front Microbiol.* 2012; 3: 322.  
注4) Tamura, K. *et al*, Chronic lymphocytic leukemia (CLL) is rare, but the proportion of T-CLL is high in Japan. *Eur. J. Haematol.* 2001; 67, 152-157

また、この局所的な loss は、血液悪性腫瘍以外のさまざまな種類のがんに関連する遺伝子を含む領域に多く生じていることも分かりました（図5）。そのうちの一部は、日本人にのみ見られる loss で、今回 BBJ データを解析したからこそ明らかになったものです。このことは、がん化メカニズムに人種差があることと、今後のさらなる体細胞モザイクの研究で血液悪性腫瘍以外のがんについても、人種間の疫学的な違い、特に日本人に固有の因子を説明できる可能性があることを意味しています。

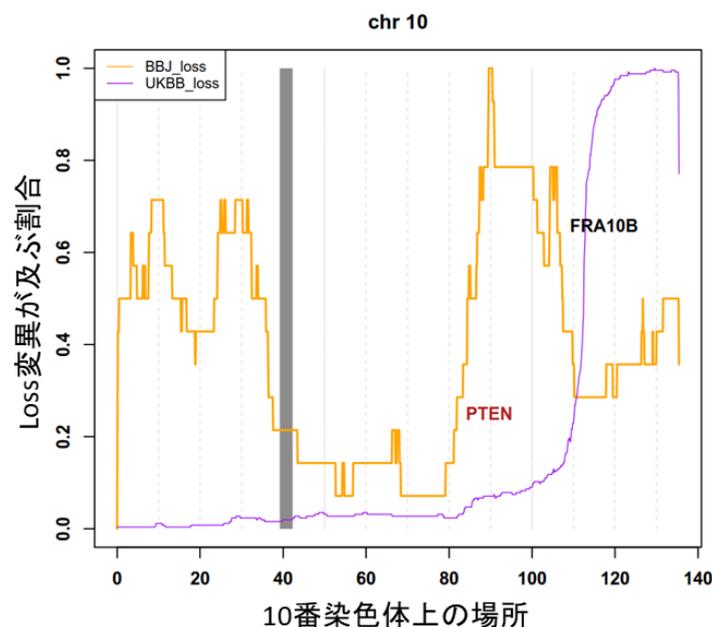


図5 がん関連遺伝子における人種特異的 loss

10 番染色体に存在する *PTEN* 遺伝子は、さまざまな悪性腫瘍に関わるがん関連遺伝子である。この領域に日本人（黄色線）では loss が生じているが、イギリス人（紫線）ではそのような傾向は見られない。

次に、後天的 DNA 変異である体細胞モザイクと関連する、生まれながらの遺伝的多型（以下、多型）を解析したところ、体細胞モザイクと非常に強く関連する多型があることを見いだしました。これまで、DNA マイクロアレイデータを用いたゲノムワイド関連解析（GWAS）<sup>[9]</sup>では、発症リスクの指標であるオッズ比<sup>[10]</sup>が 2 を超えるものは非常に少なかったため、個人が発症リスクに関わる多型を持っていても、その個人の疾患リスクを推定することは困難でした。

今回の解析では、オッズ比 90 以上に達する多型を含む合計 10 の遺伝子領域に存在する 13 の体細胞モザイク関連多型を同定しました。これらは、2018 年の UKB データと同じ領域、本研究と 2020 年のイギリス人データで同時に初めて分かった領域、さらに日本人独自の領域を含んでいました。興味深いことに、イギリス人と日本人において同じ遺伝子領域が体細胞モザイクと関連する中で、

一方の人種ではほぼ遺伝的多型性が無い稀な多型が関連する領域が複数見られました。

これは、人類がアフリカから移動して両人種が分かれた後、同じ遺伝子領域にそれぞれの人種で生じた多型が同じ体細胞モザイクの出現に関わっていることを意味しており、異なる多型による同じ遺伝子を介した同じメカニズムが存在することを示唆しています。特に関連の強い領域は *NBS1* 遺伝子と *MRE11* 遺伝子であり、これらの遺伝子がコードするタンパク質は、DNA 二重鎖切断の修復に関わる MRN 複合体<sup>[11]</sup>を構成する要素でした。すなわち、DNA 二重鎖切断という細胞にとって重篤な変化に対する修復機構の異常が、体細胞モザイクを起こす原因であることが強く示唆されました。

また、これら人種特異的な稀な関連多型を同定するにあたり、日本人の全ゲノムシーケンス解析の結果を組み込んだ参照配列<sup>[12]</sup>が有効でした<sup>注5)</sup>。全ゲノムシーケンス解析によって日本人特異的な稀な多型を同定し、18万人という大きな DNA マイクロアレイデータを用いて多型データを推定したことが、今回の結果につながりました。

最後に、BBJ データの中で生存調査データのある14万人のうち、BBJ登録時点でがん罹患していない約8.7万人を対象に、体細胞モザイクと将来の死亡率の関連を調べました。その結果、体細胞モザイクを持つと全死亡率が10%上昇することが分かりました(図6a)。疾患別の死亡率では、血液悪性腫瘍、特に白血病の死亡率の上昇と強く関連していました(ハザード比<sup>[10]</sup>4.7、図6a)。個別の体細胞モザイクとの関係を解析したところ、白血病による死亡率を80倍以上にするCN-LOHや(図6b)、全死亡率を40%以上上昇させるCN-LOHを同定しました。また、白血病死亡を除いても全死亡率の上昇はほとんど影響を受けなかったことから、体細胞モザイクの存在は血液悪性腫瘍以外の死因とも広く関わっていることが示されました。

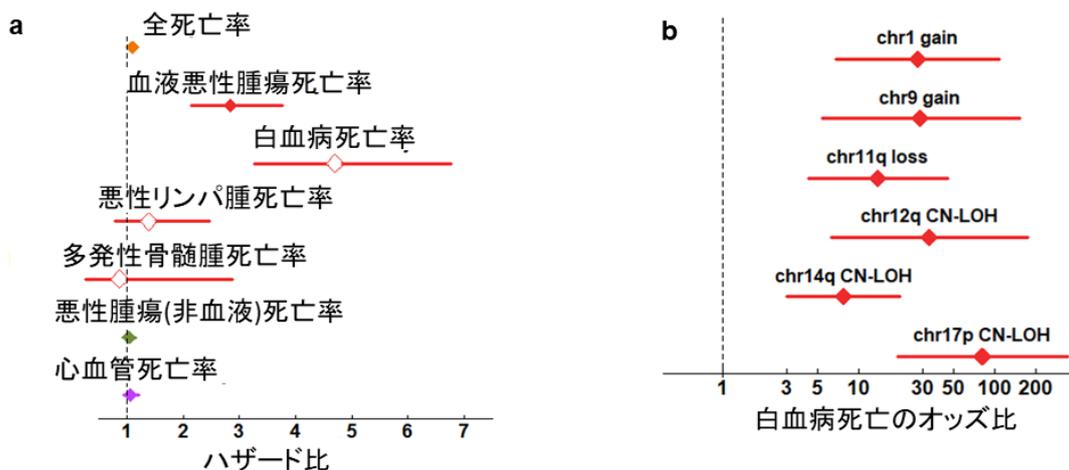


図6 体細胞モザイクと死亡率の関連

- a: 体細胞モザイクは全死亡率を10%上昇させ(橙菱形)、特に白血病死亡率を4.7倍に上昇させる。  
b: 体細胞モザイクの種類によっては、白血病死亡率の高い上昇と関連するものがある。特に、17番染色体短腕のCN-LOH (chr17p CN-LOH)は白血病死亡率を80倍以上上昇させる。

注5) Akiyama *et al.* Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat Commun.* 2019 Sep 27;10(1):4393.

### 3. 今後の期待

今回の研究では、日本人 18 万人における白血球の体細胞モザイクの地図を描くことができました。解析結果から、DNA 二重鎖切断の修復機構の異常により体細胞モザイクが生じることが強く示唆されました。体細胞モザイクと白血病死亡率の上昇との関連は、がんの発生機構の解明に向けた基礎医学の進歩に貢献すると期待できます。

また、加齢に伴う体細胞モザイク保有割合の大きな上昇は、老化に伴う DNA 二重鎖切断の修復機構の異常を示唆しています。この結果は、体細胞モザイクと全死亡率上昇の関連と合わせて、老化メカニズムの解明と疾患治療法の開発に貢献すると期待できます。

DNA マイクロアレイデータは、全世界で数百万人以上の規模で存在し、これまでその情報は生まれつきの変異の同定にのみ使われてきました。今回、このデータには後天的な染色体変化の情報が含まれていることが明らかになり、そのメカニズムの一端を解明できました。人種特異的関連多型の同定には、全ゲノムシーケンス解析の結果を利用することの有効性も示されました。今後、既に利用可能な DNA マイクロアレイ情報を集約させるほか、新しい大規模アレイデータを作り出し、日本人大規模全ゲノムシーケンス解析データと組み合わせる解析すれば、日本人に特有な機構を含む詳しい体細胞モザイクの機構が解明され、老化やがん化のメカニズムの解明、体細胞モザイク発生と老化・がん化を予測する臨床医学の発展につながると期待できます。

### 4. 論文情報

<タイトル>

Chromosomal alterations among age-related hematopoietic clones in Japan.

<著者名>

Terao C, Suzuki A, Momozawa Y, Akiyama M, Ishigaki K, Yamamoto K, Matsuda K, Murakami Y, McCarroll SA, Kubo M, Loh PR, Kamatani Y.

<雑誌>

*Nature*

<DOI>

10.1038/s41586-020-2426-2

## 5. 補足説明

### [1] クローン性に増殖する

ある細胞が何回も増殖すること。自分のクローンを作るように増殖することから、このように表現されることが多い。

### [2] 体細胞モザイク

体細胞（生殖細胞以外の細胞一般）において、後天的な変異が生じることによって、変異が無い体細胞と変異がある体細胞が混ざった状態（モザイク）となること。今回は特に後天的な変異が一塩基レベルではなく、染色体レベルあるいはその一部（数 Mb）に及ぶものを体細胞モザイクと表現している

### [3] バイオバンク・ジャパン（BBJ）

日本人集団 27 万人を対象とした生体試料のバイオバンクで、東京大学医科学研究所内に設置されている。理化学研究所が実験を行って取得した約 20 万人のゲノムデータを保有する。オーダーメイド医療の実現プログラムを通じて実施され、ゲノム DNA や血清サンプルを臨床情報とともに収集し、研究者へのデータ提供や分譲を行っている。

### [4] DNA マイクロアレイ

基板の上に、遺伝的多型（主に SNP）に相補的な probe を搭載したビーズを高密度に配置し、数十万～数百万の遺伝的多型を検出するための分析器具。

### [5] 遺伝的多型

遺伝的多型とは、ある集団において、一つの遺伝的座位に、二つかそれ以上の頻度の高い異なるアレルが存在する状態をいう。

### [6] 全ゲノムシーケンス解析

次世代シーケンサーを使って、個人（約 30 億塩基）やがんの全ゲノム情報を解読し、配列の違いや変化を同定すること。データが大量になるため、スーパーコンピュータを使って情報解析を行うのが一般的である。タンパク質をコードする 1～2% の範囲のエクソンだけでなく、遺伝子の発現を制御するゲノム領域の変異やさまざまな構造異常（大きなゲノム配列異常）も検出可能である。

### [7] UK バイオバンク（UKB）

イギリスで構築されているバイオバンクであり、50 万人規模の疾患罹患情報、臨床情報、遺伝情報などから構成される。

### [8] T 細胞、TCR、B 細胞、BCR

白血球の中で、獲得免疫と呼ばれる、外来抗原を認識して抗体を産生したり、免疫記憶をつかさどるリンパ球のサブセットに「T 細胞」と「B 細胞」がある。T 細胞の受容体を「TCR」、B 細胞の受容体を「BCR」と呼ぶ。TCR、BCR とともに、無数の外来抗原に限られた遺伝子数で対応するため、組換えが起こって無数の多型性を作り出すことが知られている。TCR、BCR はそれぞれ T-Cell Receptor、B-Cell Receptor の略。

#### [9] ゲノムワイド関連解析 (GWAS)

形質に対する遺伝的関連を知るための手法であり、SNP を用いて解析するものが一般的である。形質(疾患のある/なし や 量的形質)を目的変数、SNP の量的情報や各種共変量を説明変数にしてモデル化し、SNP の関連を評価する。GWAS は Genome-Wide Association Study の略。

#### [10] オッズ比、ハザード比

オッズ比は、発症リスクの大きさの指標。基準とする集団のリスクに対して、発症するリスクが何倍になるかを表す。ハザード比は、発症までの時間も加味したリスクの上昇を表す指標。

#### [11] MRN 複合体

MRE11、NBS1、RAD50 のタンパク質からなる複合体。DNA の二重鎖切断が起こると、その修復に決定的な役割を果たす。

#### [12] 参照配列

全ゲノム領域の遺伝子多型の遺伝子型を推測する (imputation) 際に用いる、DNA 全ゲノムシーケンスデータをもとにした配列のこと。

## 6. 発表者・機関窓口

理化学研究所 生命医科学研究センター ゲノム解析応用研究チーム  
チームリーダー 寺尾 知可史 (てらお ちかし)  
(静岡県立総合病院 臨床研究部 免疫研究部長、  
静岡県立大学 薬学部 ゲノム病態解析講座 特任教授)  
客員主管研究員 鎌谷 洋一郎 (かまたに よういちろう)  
(東京大学 大学院新領域創成科学研究科 教授)

#### <機関窓口>

理化学研究所 広報室 報道担当  
<https://www.riken.jp/>

東京大学 大学院新領域創成科学研究科 広報室  
<https://www.k.u-tokyo.ac.jp/>

静岡県立総合病院 事務部総務課  
<http://www.shizuoka-pho.jp/sogo/>

#### <AMED の事業に関すること>

日本医療研究開発機構  
ゲノム・データ基盤事業部 ゲノム医療基盤研究開発課  
<https://www.amed.go.jp/>