

## 日本人集団ゲノムにおける、未検出のゲノムの欠失の発見と機能的意義の解明

### 1. 発表者：

藤本 明洋（東京大学大学院医学系研究科国際保健学専攻人類遺伝学分野 教授）

### 2. 発表のポイント：

- ◆人類集団には、さまざまな遺伝的多様性（注1）が存在し、疾患のリスクや表現型の個人差に関わっていることが知られている。しかしながら、これまでの研究の多くは、一塩基の違いや短い挿入・欠失（注2）に注目しており、それ以外の多様性の研究はほとんど行われていない。
- ◆これまであまり解析されていない30-5000塩基の欠失を正確に検出する方法を開発し、日本人集団の全ゲノムシーケンズデータ（注3）から約4000個の欠失の個人差を検出した。また、それらのうち4%程度が遺伝子発現の個人差と関連しており、機能的であることが示唆された。
- ◆本研究で開発した方法を疾患研究に適用することで、疾患のリスクに関わる欠失の新たな発見へと繋がることが期待される。

### 3. 発表概要：

東京大学大学院医学系研究科国際保健学専攻人類遺伝学分野の Wong 特任研究員、藤本教授らのグループは、30-5000塩基の欠失を正確に検出する方法を考案した（図1a）。この方法を用いて、先行研究において次世代シーケンサーで決定された174人の日本人ゲノムを解析し、日本人集団で多型的な（個人差が大きい）約4000個の欠失を発見した（図1b）。また、長鎖シーケンサー（Oxford Nanopore シーケンサー、注4）を用いて全ゲノムシーケンズを行い、その結果と比較したところ、一致率が97%であり、今回の方法の精度が高いことが示唆された。

そこで、欠失と遺伝子発現量の関連について調べたところ、約4%の欠失が遺伝子発現の個人差（注5）と関連していた。また、遺伝子発現の個人差に関連する欠失をそれ以外の欠失と比較したところ、関連する欠失は遺伝子発現を調整するスーパーエンハンサー領域やプロモーター領域に多かった。遺伝子発現に関連する欠失のうち2つ（イントロンの中にある欠失と約1万塩基離れた遺伝子の発現量に関連する欠失）を選び、ゲノム編集技術 CRISPR-Cas9 法（注6）を用いて細胞株に導入したところ、遺伝子発現量が変化し、これらの欠失が機能的であることが強く示唆された（図2）。

本研究は、これまで検出が難しかった欠失を高い精度で検出する方法を構築し、欠失の個人差の中には遺伝子発現量に直接影響するものが存在することを示した。この方法を疾患研究に適用することで、新たな疾患関連遺伝子の発見につながると考えられる。

### 4. 発表内容：

人類集団には、さまざまな遺伝的多様性が存在し、疾患のリスクや表現型の個人差に関わっていることが知られている。遺伝的多様性の中では、一塩基多様性が最も数が多いが、そのほかにも DNA の長さの違い（塩基配列の挿入・欠失）や構造異常（塩基配列の大規模な順序変化）などが知られている。近年の塩基配列決定技術の発展により、多くの個体の配列が決定されている。しかしながら、シーケンズデータの特性や解析技術の困難さにより、一塩基多様

性や短い挿入欠失以外のゲノムの違いは、それほど調べられていない。また、それらの遺伝的多様性に機能的意義があるのか否かもよくわかっていなかった。

東京大学大学院医学系研究科国際保健学専攻人類遺伝学分野の Wong 特任研究員、藤本教授らのグループは、30-5000 塩基の欠失を正確に検出する方法を考案した (図 1 a)。この方法を用いて、先行研究において次世代シーケンサーで決定された 174 人の日本人ゲノムを解析し、日本人集団で多型的な (個人差が大きい) 約 4000 個の欠失を発見した (図 1 b)。また、長鎖シーケンサー (Oxford Nanopore 社) を用いて全ゲノムシーケンスを行い、その結果と比較したところ、一致率が 97% であり、今回の方法の精度が高いことが示唆された。

そこで、すでに公開されている日本人集団の血球系の細胞株の遺伝子発現の個人差のデータを用いて、欠失と遺伝子発現量の関連について調べたところ、約 4% の欠失が遺伝子発現の個人差と関連していた。また、遺伝子発現の個人差に関連する欠失をそれ以外の欠失と比較したところ、関連する欠失は遺伝子発現を調整するスーパーエンハンサー領域やプロモーター領域に多く、ヘテロクロマチン領域 (ゲノム内の遺伝子発現が抑制されている位置) に少なかった。さらに、ゲノム散在し機能的意義が不明のトランスポゾンの中にも遺伝子発現の個人差に関わるものがあり、トランスポゾンの欠失にも機能的なものが存在することが示唆された。発現に関連する欠失とそれ以外の欠失には長さの差はなく、欠失する領域の特性が重要であることが示唆された。遺伝子発現に関連する欠失のうち 2 つ (イントロンの中にある欠失と約 1 万塩基離れた遺伝子の発現量に関連する欠失) を選び、ゲノム編集技術 CRISPR-Cas9 法を用いて細胞株に導入したところ、遺伝子発現量が変化し、これらの欠失が機能的であることが強く示唆された (図 2)。

本研究は、これまで検出が難しかった欠失の検出法を構築し、欠失の個人差の中には遺伝子発現量に直接影響するものが存在することを示した。また、ゲノム多様性研究への長鎖シーケンシング技術の有用性を示している。この方法を疾患研究に適用することで、新たな疾患関連遺伝子が発見につながると考えられる。

なお、本研究は、日本医療研究開発機構 (AMED) ゲノム医療実現推進プラットフォーム事業「先端ゲノム研究開発」プロジェクトにおいて行われた研究である。

## 5. 発表雑誌:

雑誌名: 「*Genome Medicine*」 (オンライン版: 7 月 25 日)

論文タイトル: Identification of intermediate-sized deletions and inference of their impact on gene expression in a human population

著者: Jing Hao Wong, Daichi Shigemizu, Yukiko Yoshii, Shintaro Akiyama, Azusa Tanaka, Hidewaki Nakagawa, Shu Narumiya, Akihiro Fujimoto

アブストラクト URL: <https://genomemedicine.biomedcentral.com>

## 6. 問い合わせ先:

東京大学大学院医学系研究科人類遺伝学教室

教授 藤本 明洋 (ふじもと あきひろ)

Tel: 03-5841-3692 E-mail: [afujimoto@m.u-tokyo.ac.jp](mailto:afujimoto@m.u-tokyo.ac.jp)

## 7. 用語解説：

### 注1 遺伝的多様性

遺伝物質である DNA の個人間の違いのこと。ATGC の 4 種類の塩基が連なっている DNA 配列には、塩基の置換、塩基の挿入、塩基の欠失や、塩基の順序の違いなどが存在することが知られている。遺伝的差異は、ヒト一人当たり約 300-400 万個存在すると考えられている。これらの遺伝的多様性はヒトの表現型の違い（皮膚色、髪質など）や病気のなりやすさなどにも関係している。

### 注2 挿入・欠失

DNA に置ける塩基配列の欠損や異なる塩基の挿入のこと。塩基の長さの変化を引き起こす。遺伝子領域に生じた場合、アミノ酸配列に大きな影響を与えることもある。本研究では、30-5000 塩基の欠失を対象とした研究を行なった。

### 注3 全ゲノムシーケンス

全ゲノムの塩基配列を決定すること。シーケンサーと呼ばれる塩基配列決定装置を用いて行われる。約十年前に超並列シーケンサー（いわゆる次世代シーケンサー）が登場し、ヒトの全ゲノムシーケンスが比較的容易になった。遺伝多様性、がんや難病の研究のために、多くのゲノムシーケンス研究が行われている。次世代シーケンサーのデータは膨大であり、正確な遺伝的多様性を検出と、それらの機能的意義の解明には、まだ課題が多い。

### 注4 長鎖シーケンサー

長い塩基配列を直接決定できるシーケンサーのこと。現在広く使われている次世代シーケンサーは、読み取り長が 100-300 塩基程度であるが、長鎖シーケンサーは非常に長い塩基配列（10 万塩基以上）が直接決定できる。そのため、次世代シーケンサーが苦手とする塩基配列が複雑な領域（相互によく似ている、個人差が極めて大きいなど）の解析に力を発揮すると考えられている。本研究では、現在利用可能な 2 種類の長鎖シーケンサーのうち、Oxford Nanopore 社のシーケンサーを用いた。エラー率の高さや実験のコストの高さ、データ解析の難しさなどからそれほど普及していないが、有用な技術であり、今後、活用が広まってゆくと考えられる。本研究では、次世代シーケンサーと組み合わせることで、従来検出が難しいと言われていた欠失を多数同定した。

### 注5 遺伝子発現の個人差

遺伝子発現量（細胞中の mRNA の量）の個人差のこと。体質の違いのように、遺伝子の量も個体間で異なっていることが知られている。最近の研究で、遺伝子発現量の個人差が疾患のリスクに影響することが示されている。

### 注6 CRISPR-Cas9 法

ゲノム編集技術の一つ。ゲノム上の任意の領域を切断することができる。切断されたゲノム領域は細胞内で修復されるが、その際に、塩基の変化が引き起こされることがある。このことを用いて、ゲノムに変異が導入される。本研究では、2 箇所を切断し欠失を導入した。

## 8. 添付資料：

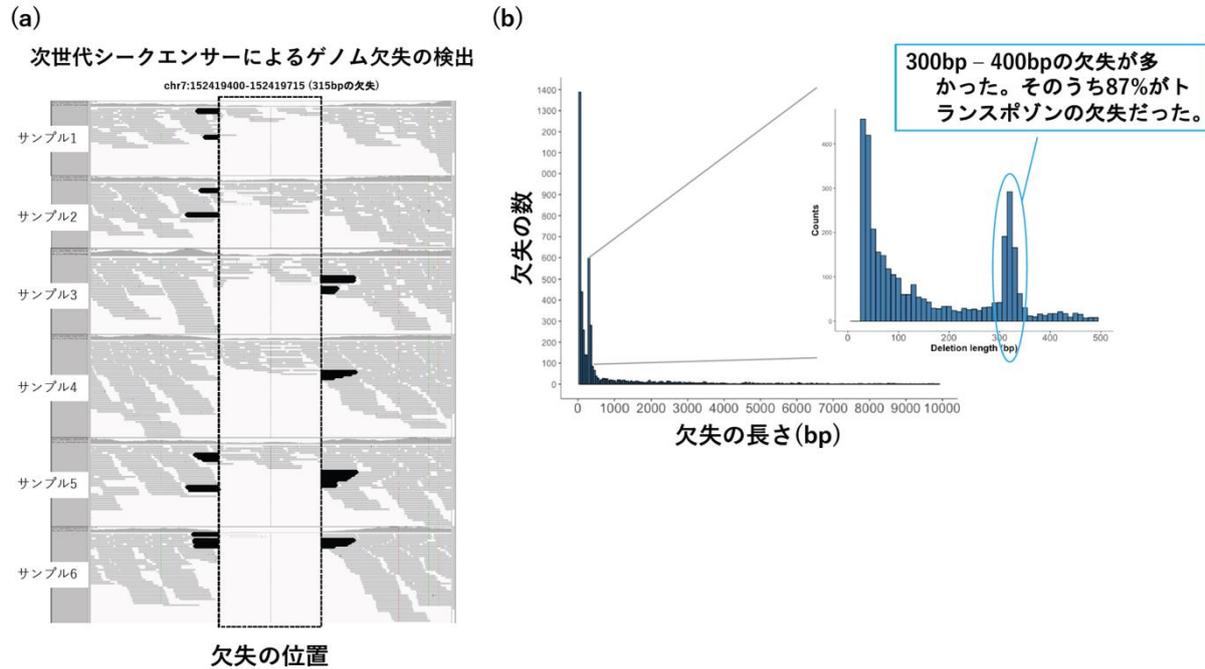


図1 欠失の検出

(a)次世代シーケンサーのデータからの欠失の検出 6 サンプルの例を示す。欠失を示唆するリード（シーケンサーが出力する塩基配列）は黒、それ以外は灰色。複数のサンプルでリードの本数が減っている箇所があり、欠失であると考えられる。(b)検出された欠失のサイズの分布。短い欠失が多い傾向がある。300-400bp の欠失も多く、それらの 87%がトランスポゾンを単位とする欠失であった。

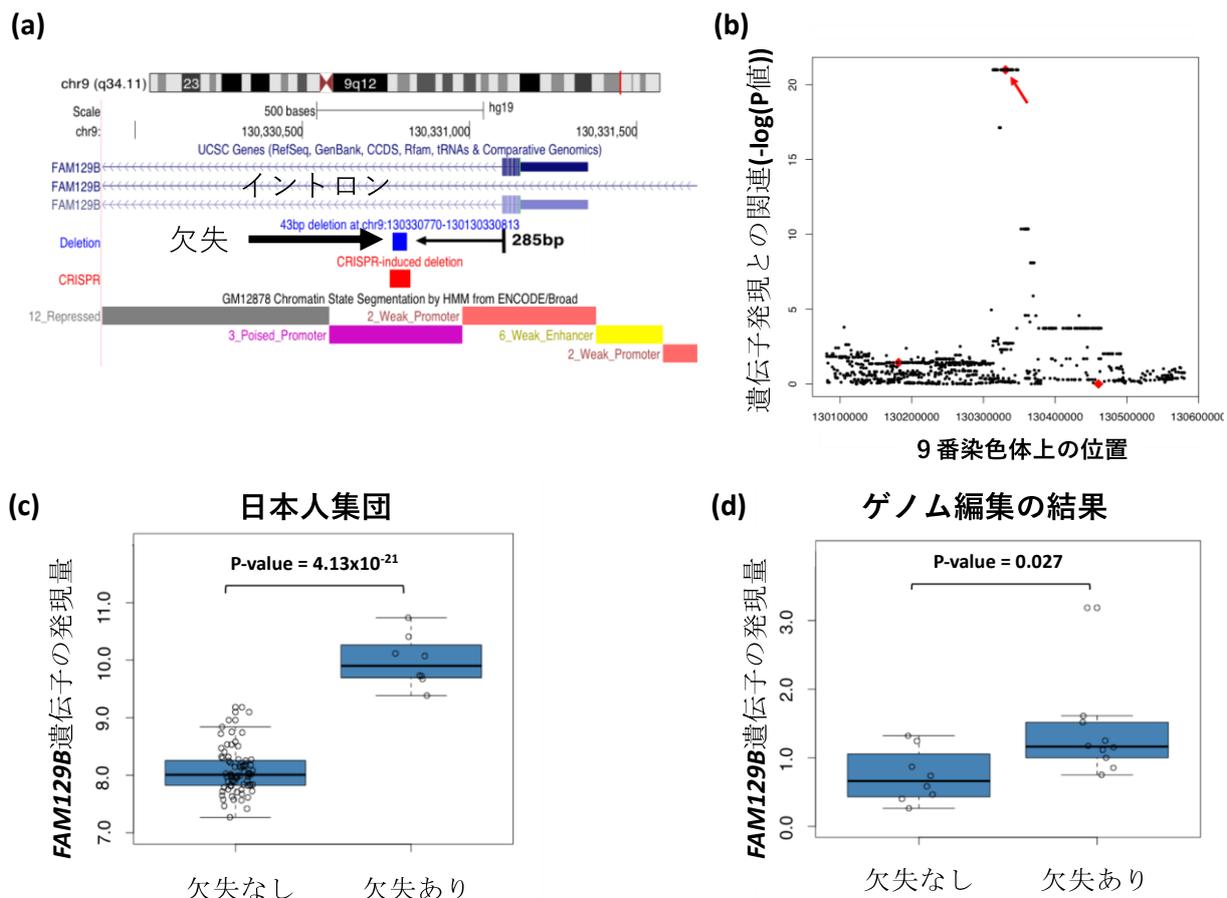


図2 欠失と遺伝子発現の関連の例

(a) *FAM129B* 遺伝子の欠失の例。イントロンの転写因子結合部位に 43bp の欠失が検出された。(b) 遺伝子発現量との関連解析の結果。黒が一塩基多型 (SNP)、赤が欠失。x 軸にゲノム上の位置、y 軸に関連の強さ (P 値) を示す。(c) 日本人集団における欠失の有無と発現量の個人差。欠失を持つ個体は発現量が高かった。(d)ゲノム編集による欠失の導入の結果。日本人集団の結果 ((c) と同様の結果が得られており、欠失が発現量の変化を引き起こすことが強く示唆された。