プレスリリース

大学共同利用機関法人 情報・システム研究機構 国立遺伝学研究所
大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用基盤施設 Joint Support-Center for Data Science Research (ROIS-DS)
地方独立行政法人 静岡県立病院機構 静岡県立総合病院 Shizuoka General Hospital

2025年8月25日

本件の情報公開は、既に解禁されています

ゲノム構造変異とリピート変異を配列識別して高精度に検出するソフトウェアTRsvを開発

■ 概要

個人間のゲノム配列の違いは、病気の罹りやすさを含めた様々な形質の違いを表しています。配列の違いを生み出すものの中で繰り返し変異（繰り返し配列のコピー数変異）は神経筋疾患や量的形質の要因となり、構造変異・インデル（欠失や挿入など）は神経発達障害や癌などの疾患要因となっています。しかし、これらを正確に区別して検出する解析手法はこれまでに存在しませんでした。

本研究では、ロングリードデータを用いて繰り返し変異と構造変異・インデルを配列識別によって正確に区別して同定するソフトウェア（TRsv）を開発しました。TRsvは既存のツールと比較してより高い繰り返し変異検出精度・感度を示し、繰り返し変異と構造変異をより高い精度で識別しました。さらに、160人のロングリード全ゲノムシーケンスデータを用いた解析において、TRsvは遺伝子発現、疾患、量的形質に関連する繰り返し変異を実際に検出できることを証明しました。今後ロングリードの活用が増す中で、TRsvはゲノムの繰り返し変異、構造変異・インデルを検出するツールとして広く活用されることが期待されます。

本成果は、情報・システム研究機構国立遺伝学研究所、同機構データサイエンス共同利用基盤施設、静岡県立総合病院、理化学研究所によるものです。

■ 成果掲載誌

本研究成果は、国際科学雑誌「Genome Biology」に2025年8月20日（日本時間）に掲載されました。

論文タイトル:

TRsv: simultaneous detection of tandem repeat variations, structural variations, and short indels using long read sequencing data.

（ロングリードを用いて繰り返し変異、構造変異、インデルを区別して検出するソフトウェアTRsvの開発）

著者: Shunichi Kosugi*, Chikashi Terao （小杉俊一*、寺尾知可史） *責任著者

DOI: 10.1186/s13059-025-03718-z

■ 研究の詳細

● 研究の背景

ゲノム（注1）の構造変異（SV）（注2）は、欠失、挿入、重複、逆位等の50 bp以上のゲノム配列変異の総称であり（50 bp未満の欠失と挿入はインデルと呼ばれます）、ヒトゲノムには個人あたり24,000～26,000のSVが存

在します。その大きなサイズから、遺伝子の機能や発現に大きな影響を与え、多くの疾患の原因となっています。ゲノムには、短いDNA配列の繰り返しからなるタンデム繰り返し（TR）領域があり、ヒトゲノムの約3%を占めますが、SVの60%以上はTR領域の繰り返し変異（繰り返し配列の増減）（注3）です。TR繰り返し変異とそれ以外のSVは、その配列の特性上、遺伝子に与える影響が異なるため、原因となる疾患のタイプも異なることが多くなります。このため、これらを正確に区別して同定する必要がありますが、TR領域に観察されるSV（挿入）には繰り返しを持たないものや異なる繰り返し単位を含むものがあり、正確にこれらを区別して同定する技術が求められていました。

- **本研究の成果**

情報・システム研究機構国立遺伝学研究所 先端ゲノミクス推進センター 特任准教授、同機構データサイエンス共同利用基盤施設 ゲノムデータ解析支援センター 特任准教授および静岡県立総合病院 研究員（研究当時）の小杉俊一は、個人のゲノムに存在する構造変異、インデルや繰り返し変異を正確かつ効率的に検出する情報解析ツールTRsvを開発しました。TRsvは、ロングリードデータを用いて、TR領域外のSVとインデルを検出、判別（挿入配列に含まれる重複や繰り返し、転移因子等の同定）するだけでなく、TR領域内に存在する挿入配列の配列組成を精査することによってTR繰り返し変異であるか否かを判別します（図1）。従来のツールは、繰り返し変異であるか否かに関わらず単にSVを検出するか、またはTR領域内の変異を検出するだけにとどまるものでしたが、TRsvはゲノム上のSVを隈なく同定し、SVがどのようなタイプの変異であるかの注釈付けを行うことを特徴とします。さらに本研究では、TRsvを用いて160人のロングリード全ゲノム配列データからTR繰り返し変異を含むSVを検出し、TR繰り返し変異のサイズが多くの疾患に関わる遺伝子の発現量と相関していることを明らかにしました（図2）。

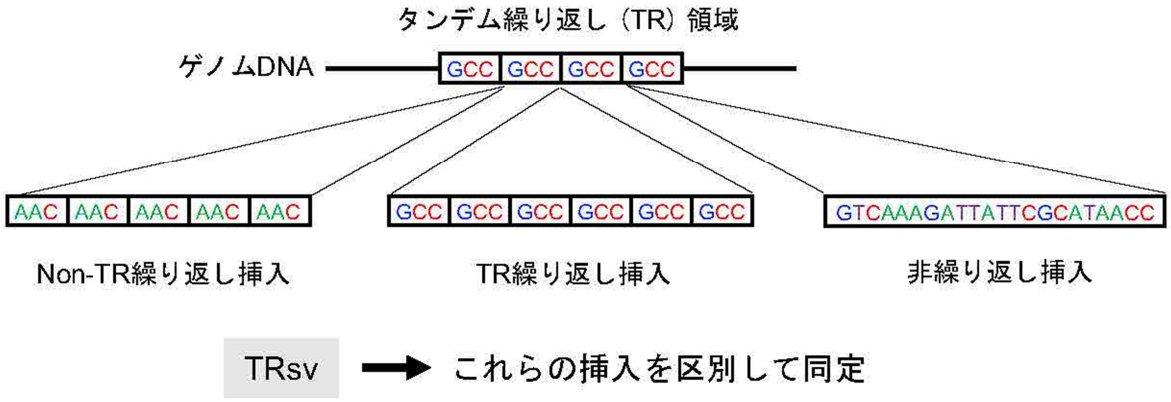

図1: TRsvはタンデム繰り返し領域で観察される異なるタイプの挿入を検出する

タンデム繰り返し（TR）領域内では、TR領域の繰り返し単位（図の例ではGCC）と同じ繰り返し単位からなるTR繰り返し挿入が観察されることが多いが、異なる繰り返し単位からなる挿入（Non-TR繰り返し挿入）や、繰り返しを持たない挿入（非繰り返し挿入）がしばしば観察される。TRsvは、これらの異なるタイプの挿入を区別して同定する。

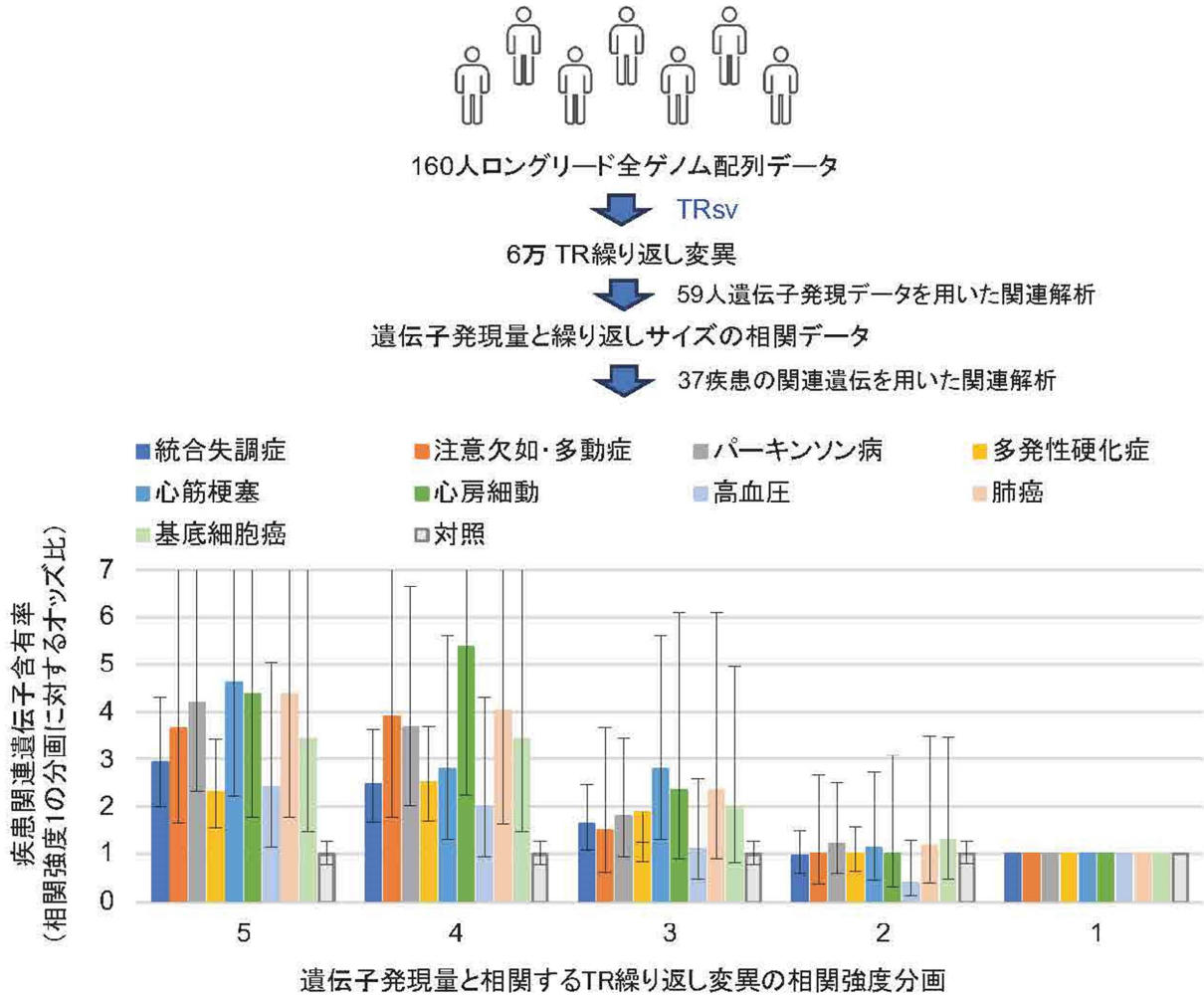

図2: 疾患関連遺伝子とTR繰り返し変異の関連がTRsvを用いて明らかにされた

160人のロングリード全ゲノム配列データからTRsvを用いて検出された6万個のTR繰り返し変異、および対応する59人分の遺伝子発現データを用いて、遺伝子発現量とTR繰り返し変異のサイズの相関が調べられた。その相関強度が強い順に5から1まで相関遺伝子セットを5等分に分画した(横軸)。各分画に含まれる疾患関連遺伝子の割合が計算され、相関強度1の分画の疾患関連遺伝子含有率に対する各分画の含有率の比(オッズ比)が各疾患毎にプロットされた(縦軸)。対照として、ヒト2万遺伝子から無作為に選別された800遺伝子セットのオッズ比が示される。バー上の黒線は標準誤差を示す。この結果は、多くの疾患に関わる遺伝子の発現がTR繰り返し変異によって制御されることを示すと共に、疾患の発症がTR繰り返し変異によって影響を受けることを示唆している。

- **今後の期待**

開発されたTRsvは、個人のゲノムに存在する構造変異、インデルや繰り返し変異を正確かつ効率的に検出することが出来るため、疾患の原因となる変異の同定を促進し、疾患の診断、治療、病気のなりやすさの推定などに役立てることができます。また、TRsvはヒト以外の生物にも適用可能なため、作物の形質に関わる変異を同定することに役立てることができます。

■ 用語解説

(1) ゲノム: 個々の生物の細胞1つに含まれるDNA配列のセットを言う。ヒトの場合、総計3.1ギガ塩基(3.1 x 109 bp)の並びを持った配列が父方由来と母方由来のそれぞれ23本の染色体に分かれた状態で、細胞中

の核内に存在する。ゲノムDNA配列には、RNAやタンパク質を作る基となる配列情報単位：遺伝子がヒトの場合2万以上存在する。

(2) 構造変異（SV: Structural variation）： 50塩基（bp）以上の長さを持ったゲノム配列変異。配列の欠失や挿入、重複などいくつかのタイプがある。50塩基未満の欠失と挿入の総称をインデル（indel）、1塩基のみの変換を一塩基多型（SNP）と呼ぶ。SVはSNVやindelに比べて数は少ないが、サイズが大きいため、遺伝子の発現（転写）や機能に大きな影響を及ぼす可能性を有する。

(3) タンデム繰り返し（TR: tandem repeat）変異： ゲノムには多くの繰り返し配列を含むが、繰り返し領域の中で1 bpから数十bpの比較的短く同方向の繰り返し単位を持った繰り返し領域をタンデム繰り返し領域と呼ぶ。TR領域は変異し易い性質があり、繰り返し単位毎の欠失や挿入が起こることが多い。特にCGGなどの3 bpの繰り返し単位を持ったTR領域において長い繰り返し単位の挿入が起こり、短いタンパク質に翻訳される場合があり、これがリピート伸長病と呼ばれる神経筋疾患を引き起こすことが知られる。

■ 研究体制と支援

本研究成果は、情報・システム研究機構国立遺伝学研究所 先端ゲノミクス推進センター 特任准教授、同機構データサイエンス共同利用基盤施設 ゲノムデータ解析支援センター 特任准教授および静岡県立総合病院 研究員（研究当時）の小杉俊一と、 静岡県立総合病院 免疫研究部長 、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 チームリーダーの寺尾知可史との共同研究成果です。

本研究は、日本学術振興会（JSPS）科研費（JP17K07264, JP21K06130）の支援を受け行われたものです。

■ 問い合わせ先

＜研究に関すること＞

● 情報・システム研究機構　データサイエンス共同利用基盤施設 ゲノムデータ解析支援センター
特任准教授　小杉俊一
メール: shunichi.kosugi@nig.ac.jp

＜報道担当＞

● 情報・システム研究機構 国立遺伝学研究所　広報室
メール: prkoho@nig.ac.jp

● 情報・システム研究機構 本部広報室
メール: koho@rois.ac.jp

● 静岡県立総合病院　総務課
メール: sougou-soumu@shizuoka-pho.jp

配付先
文部科学記者会、科学記者会、三島記者クラブ、静岡県庁記者クラブ