

令和 6 年 5 月 30 日現在

機関番号：14301

研究種目：若手研究

研究期間：2022～2023

課題番号：22K15646

研究課題名（和文）スプライシング関連タンDEMリピートが遺伝性疾患の病態に果たす役割の解明

研究課題名（英文）Comprehensive identification of tandem repeats associated with splicing in humans

研究代表者

浜中 耕平（HAMANAKA, Kohei）

京都大学・高等研究院・特定助教

研究者番号：20801129

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：我々は、800人以上のドナーの全ゲノムシーケンシングと全身の組織のRNAシーケンシングを用いて、スプライシング関連タンDEMリピートのカタログを作製した。これらの中に、脊髄小脳変性症6、脊髄小脳変性症12という2つのリピート伸長病の原因座位（各々CACNA1A、PPP2R2B遺伝子）が含まれていた。このCACNA1AとPPP2R2Bのリピートについて、一般集団とリピート伸長病患者を比較した結果、一般集団でも軽度ながら患者と同様のスプライシング変化を起こしていた。この結果は、一般集団のデータを用いてリピート伸長病の病理メカニズムを予測することが可能であることを示唆している。

研究成果の学術的意義や社会的意義

本研究で作成した、スプライシングと関連するリピートの網羅的なカタログは、将来的に希少な神経疾患のメカニズムの理解に役立つと思われる。また、遺伝性がある頻度の高い疾患（自己免疫疾患など）のメカニズムの理解にも役立つ可能性が考えられる。

研究成果の概要（英文）：Tandem repeats (TRs) are one of the largest sources of polymorphism, and their length is associated with gene regulation. Although previous studies reported several tandem repeats regulating gene splicing in cis (spl-TRs), no large-scale study has been conducted. In this study, we established a genome-wide catalog of 9537 spl-TRs with a total of 58,290 significant TR-splicing associations across 49 tissues (false discovery rate 5%) by using Genotype-Tissue Expression (GTEx) Project data. Regression models explaining splicing variation by using spl-TRs and other flanking variants suggest that at least some of the spl-TRs directly modulate splicing. In our catalog, two spl-TRs are known loci for repeat expansion diseases, spinocerebellar ataxia 6 (SCA6) and 12 (SCA12). Splicing alterations by these spl-TRs were compatible with those observed in SCA6 and SCA12. Thus, our comprehensive spl-TR catalog may help elucidate the pathomechanism of genetic diseases.

研究分野：疾患遺伝学

キーワード：タンDEMリピート スプライシング リピート伸長病

1. 研究開始当初の背景

同じ塩基配列が連続して繰り返されているゲノム領域、タンデムリピート、の長さは、ゲノムの機能に影響し、遺伝子のスプライシングをローカル (in cis) に調節する例が報告されている (スプライシング量の形質座位、splicing quantitative trait loci [sQTL])。しかし、この sQTL タンデムリピートをゲノムワイドに同定した研究はない。そのため、sQTL タンデムリピートの情報が、リピート伸長病などの遺伝性疾患のメカニズムの理解に役立てられてこなかった。

2. 研究の目的

- (1) 全身の各組織における sQTL タンデムリピートの網羅的なカタログを作る。
- (2) sQTL タンデムリピートがスプライシングを変化させるメカニズムを明らかにする。
- (3) sQTL タンデムリピートとリピート伸長病の関わりを網羅的に明らかにする。

3. 研究の方法

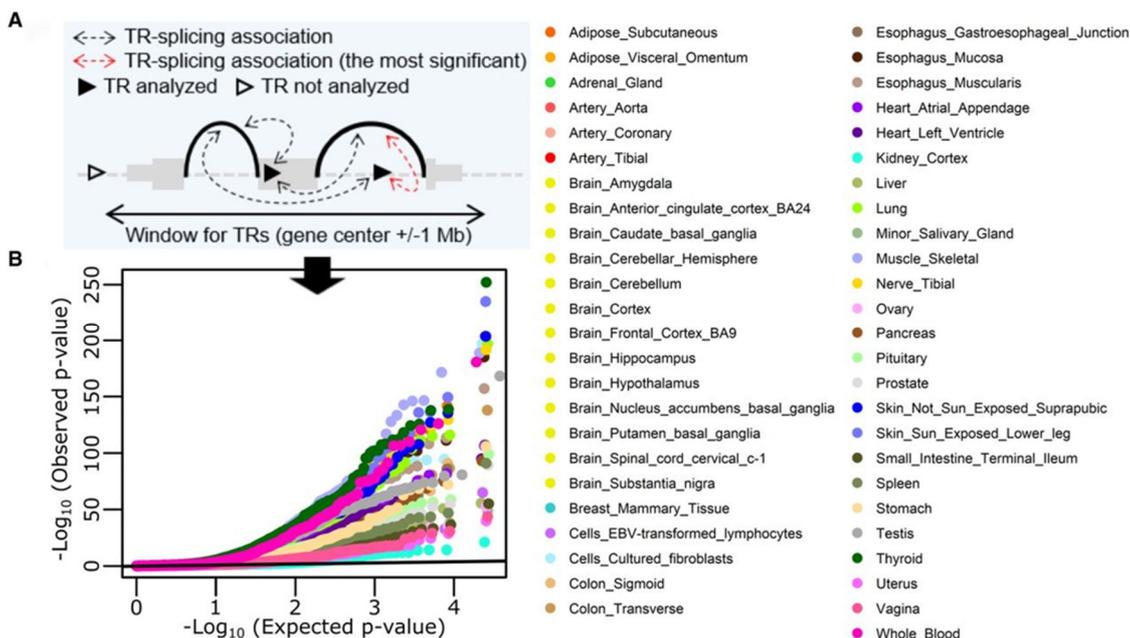
(1) Genotype-Tissue Expression (GTEx) コンソーシアムが生み出した、800 人以上のドナーの全ゲノムシーケンシング (whole genome sequencing、WGS) と全身の組織の RNA シーケンシング (RNA-sequencing、RNA-seq) を用いて sQTL タンデムリピートのカタログを作る。具体的には、The database of Genotypes and Phenotypes (dbGaP) と GTEx ポータルから以下をダウンロードする： 838 人分の WGS データ、49 組織の RNA-seq によるスプライシング量データ、49 組織の共変数データ。次に、この WGS データをタンデムリピートジェノタイピングツールである GangSTR を用いて、全ゲノムのタンデムリピート長を得る。スプライシング量をこのタンデムリピート長と共変数で説明するように回帰分析することで関連を調べる。

(2) sQTL タンデムリピートの長さや配列のスプライシングに対する影響を、SpliceAI (ゲノム配列からスプライシングのパターンを予測する深層学習モデル) を用いたインシリココンピュータジェネシスで調べる。

(3) sQTL タンデムリピートのカタログを用いて、リピート伸長病の原因リピートとスプライシングの関連を調べる。具体的には、カタログの中で、既知のリピート伸長病の原因である 52 個のタンデムリピートと一致するものを探す。次に、当該疾患の細胞・組織の RNA-seq を行い、スプライシング異常の有無を調べる。

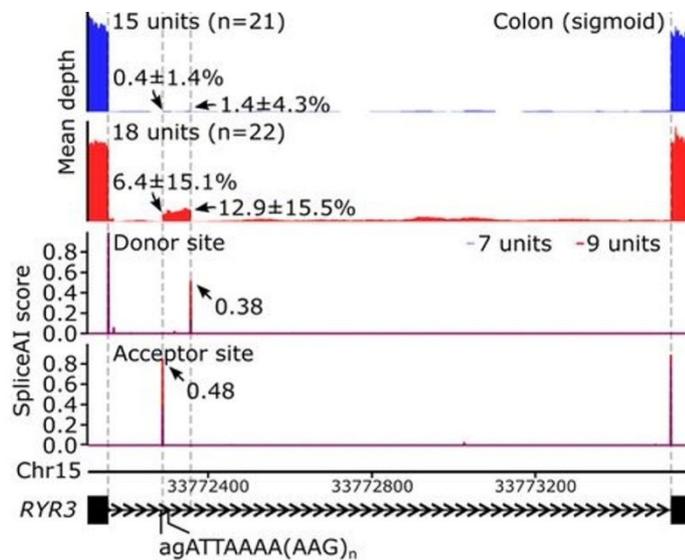
4. 研究成果

以上のようにして、我々は 49 組織にわたって sQTL タンデムリピートのカタログを作製した (11,626 個、偽陽性率 5%)。



次に、sQTL タンデムリピートがスプライシングをどのようなメカニズムで変化させるかについて調べた。

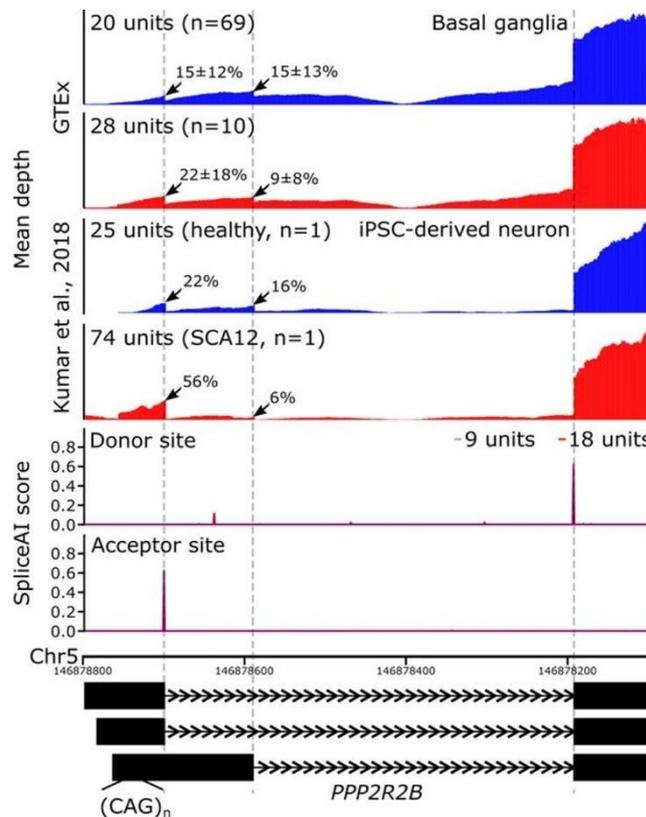
(1) リpeat伸長がエクソニックスプライシングエンハンサーの機能を強める例：*RYR3* 遺伝子の例では、エクソンイントロン接合部の近傍のAAGリpeatがスプライシング変化と関連していた。この因果関係は、ゲノム配列からスプライシングのパターンを予測する深層学習モデルである SpliceAI によってサポートされた。また、このリpeatが位置するエクソンと周辺配列を模したミニジーンベクターによる *in vitro* 系で、リpeat長とスプライシングの変化の相関を確認した。AAG の繰り返しは典型的なエクソニックスプライシングエンハンサーとして知られているため、この *RYR3*



の例では AAG の繰り返しが長くなるにつれてエクソニックスプライシングエンハンサーとしてより強く機能したことが、リpeat長とスプライシングの相関の原因と考えられた。

(2) リpeat伸長によりアクセプターサイトとブランチポイントの間の距離が適正な範囲から外れる例：*NARS2* 遺伝子の例では、GTTTTT リpeatがアクセプターサイトとブランチポイントの間に位置しており、伸長するにつれてこの間の距離が大きくなる関係にあった。本例も上記の *RYR3* と同様に、SpliceAI とミニジーン実験により、そのリpeat長とスプライシング変化の相関が確認された。この SpliceAI におけるインシリコミュタジェネシスでは、リpeat伸長によるスプライシングの変化は、その中の配列 (GTTTTT) に依存していなかった。そのため、このスプライシングの変化はアクセプターサイトとブランチポイントの間の距離の変化が原因である可能性が考えられた。

最後に、sQTL タンデムリpeatとリpeat伸長病の関連について調べた。我々のカタログの中に、脊髄小脳変性症 (Spinocerebellar ataxia, SCA) 6 と脊髄小脳変性症 12 という2つのリpeat伸長病の原因座位であるリpeat (各々 *CACNA1A* 遺伝子と *PPP2R2B* 遺伝子のタンデムリpeat) が含まれていた。これらのリpeatについて、GTEx プロジェクトの一般集団とリpeat伸長病患者を比較した結果、一般集団でもリpeatが長くなれば、軽度ながら患者と同様のスプライシング変化を起こしていた。例えば、右図において、GTEx の一般集団において、*PPP2R2B* の CAG リpeatが 20 ユニットと 28 ユニットの時に、複数のドナーサイトの使用割合に変化があり、これは Kumar et al. のデータの 74 ユニットを持つ SCA12 患者の RNA-seq でみられる変化と同じ傾向である。また、SCA12 については、先行研究におけるスプライシング異常の実験的エビデンスが十分でなかったため、本研究で同様のスプライシング変化の傾向が一般集団でも確認されたことは大きな価値がある。これらの結果は、一般集団のデータを用いてリpeat伸長病の病理メカニズムを予測することが可能であることを示唆している。



このように、sQTL タンデムリpeatのカタログは将来的に様々なリpeat伸長病の病態メカニズムの理解に役立つことが期待される。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Kohei Hamanaka, Daisuke Yamaguchi, Eriko Koshimizu, Kei Watase, Kaoru Mogushi, Kinya Ishikawa, Hidehiro Mizusawa, Naomi Tsuchida, Yuri Uchiyama, Atsushi Fujita, Kazuharu Misawa, Takeshi Mizuguchi, Satoko Miyatake, Naomichi Matsumoto	4. 巻 33
2. 論文標題 Genome-wide identification of tandem repeats associated with splicing variation across 49 tissues in humans	5. 発行年 2023年
3. 雑誌名 Genome Research	6. 最初と最後の頁 435-447
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/gr.277335.122	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 濱中 耕平
2. 発表標題 スプライシング関連タンDEMリピートのゲノムワイドな同定
3. 学会等名 NGS EXPO 2023
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------