

令和 3 年 6 月 1 日現在

機関番号：12601

研究種目：若手研究

研究期間：2019～2020

課題番号：19K16792

研究課題名（和文）長鎖シーケンサーによる全長RNA解読が可能にするがんの新規ネオ抗原評価法の開発

研究課題名（英文）Identification of novel neoantigen candidates in cancers by full-length cDNA sequencing using a long read sequencer

研究代表者

鈴木 絢子（Suzuki, Ayako）

東京大学・大学院新領域創成科学研究科・特任准教授

研究者番号：00770348

交付決定額（研究期間全体）：（直接経費） 4,700,000円

研究成果の概要（和文）：本研究は、MinIONを駆使して肺がん細胞の全長cDNA解読を行い、がん細胞における異常転写産物の全長構造の同定およびそれらを由来とする新規ネオ抗原候補の探索を行うものである。肺がん細胞株および非小細胞肺がん臨床検体を用いて、肺がんの全長転写産物カタログを作成し、異常転写産物の同定を行った。また、NMDやRNAスプライシング因子のステータスと異常転写産物の関係性を示した。さらに、得られた異常転写産物から翻訳される予測ペプチドについて、HLA結合親和性や免疫原性を評価した。本研究によって全長mRNA配列から明らかとなる異常スプライスバリエーションの解析が極めて重要であることが示唆された。

研究成果の学術的意義や社会的意義

本研究では、ロングリードシーケンシング技術を用いて、がん細胞における転写産物の全長構造を明らかにしており、さらに異常mRNAから産生されるネオ抗原候補を検出する手法を構築している。これまで正しく評価ができていなかったフレームシフト変異や異常スプライシングパターンに由来するネオ抗原候補を新たに検出することができる、新規ゲノム解析技術を駆使した手法を構築している面から、ゲノムおよび腫瘍診断分野における学術的意義は大きいと考えている。また、TMBに加えて免疫チェックポイント阻害療法に対する新たなマーカー候補を提案することができており、社会的意義も十分にあると考えている。

研究成果の概要（英文）：In this study, we conducted full-length cDNA sequencing of lung cancers using a long read sequencer MinION to identify aberrant transcript variants and potential neoantigen candidates. We first constructed a catalogue of full-length cDNA sequences from lung cancer cell lines and non-small cell lung cancer clinical samples and identified aberrant splicing variants, such as exon skipping, intron retention and their combinations. We also evaluated whether nonsense-mediated mRNA decay (NMD) and RNA splicing factors affected accumulation of aberrant transcripts and found that UPF1 and SF3B1 knockdown resulted in increasing aberrant splicing patterns in A549 cells. We next estimated HLA binding affinity and antigenicity of peptides which were predicted from the detected aberrant splicing variants using NetMHCpan and experimental assays, such as ELISpot. The results indicate that precise identification of full-length mRNA sequences are important for detection of neoantigens.

研究分野：がんゲノム科学

キーワード：ロングリードシーケンシング 全長cDNAシーケンシング 肺がん ネオ抗原

### 1. 研究開始当初の背景

がん組織において、免疫細胞は、ゲノム変異に由来したがんネオ抗原を目印にがん細胞を非自己として認識することが知られている。一方、がん細胞は PD-L1 等の免疫チェックポイント (Immune Checkpoint; IC) 因子を発現し、免疫細胞の IC 受容体 (PD-1 等) と結合することで免疫疲弊を引き起こし、免疫細胞の攻撃から逃れる。近年、これら IC 因子を標的とした抗体を用いて、免疫細胞の再活性化によりがん細胞を排除する IC 阻害療法が盛んに行われている。最近の研究で、がんネオ抗原のもととなるゲノム変異の多いがん患者には特に効果が高いことが明らかとなり、IC 阻害療法の効果予測マーカーとして注目されている。一方で、変異の量 (Tumor Mutation Burden; TMB) だけでは IC 阻害療法に対する効果を完全には予測することができず、効果予測手法の開発は注目度の高い研究である。

そこで、本研究では、TMB の算出に主に使用されるミスセンス変異の量だけでなく、mRNA の全長構造を考慮したうえで、フレームシフト変異やスプライシング異常に着目し、ネオ抗原のもととなりうる新規通常転写産物の同定を行うこととした。このようなアミノ酸配列に大きな変化を伴う異常 mRNA に由来したネオ抗原は、抗原性が高いのではないかと推測される。がん細胞における異常 mRNA の全体像を明らかにすることは、IC 阻害療法の効果を予測するうえで非常に重要である可能性がある。

がん細胞の異常 mRNA を網羅的に同定するためには、ナノポアシーケンサーを用いた全長 cDNA-seq 解析手法を活用することを考えた。ナノポアシーケンサー MinION は、数十 kb におよぶ長いシーケンスリードを解読できる長鎖シーケンサーである。1 ランで数百万リードを解読することができるため、cDNA の全長配列を網羅的に同定することができる。本研究グループではこれまで、さまざまな細胞種の mRNA の量およびスプライスパターン、全長構造について MinION を用いて詳細に解析してきた (Seki M et al. 2018 *DNA Res*)。

また、一部のがんでは、RNA スプライシング関連因子や NMD (nonsense-mediated mRNA decay) といった RNA 品質管理機構関連因子に変異を有することが報告されている (Seiler M et al. 2018 *Cell Reports*)。例えば、スプライシング関連因子である *U2AF1* 遺伝子では、血液腫瘍や肺腺癌において変異が頻出することが知られており (TCGA 2014 *Nature*)、スプライシングパターンに影響を及ぼすと言われている。このような RNA の品質に関連する因子における変異は、異常 mRNA の蓄積に寄与すると考えられるため、こうした因子のゲノム変異ステータスも本研究にとって重要な情報である。

実際にどの程度、どのような状態で異常 mRNA ががん細胞に蓄積しているのか、その詳細は明らかとなっていない。そこで、本研究では、MinION を用いてがん細胞の全長 cDNA-seq 解析を行うことによって、異常 mRNA の全長構造およびその量を同定し、さらにそれらがネオ抗原となりうる可能性を評価する手法を開発する。

### 2. 研究の目的

本研究では、がん細胞における異常 mRNA の全長配列を、ナノポアシーケンサー MinION を用いて同定し、それらを計測する手法を開発することを目指す。また、異常 mRNA の蓄積を引き起こしうる NMD やスプライシング関連因子のゲノム変異ステータスにも着目し、異常 mRNA のパターンおよび量との関係を明らかにする。異常 mRNA の量をサロゲートするような因子 (例えば、NMD 因子のゲノム変異等) の同定も目指す。また、得られた異常 mRNA がネオ抗原になりうるか、異常 mRNA 全長配列から予測される新規ペプチド配列に対する評価も実施する。

最終的には、TMB の測定に加えて、mRNA の全長解読を行うことにより IC 阻害剤の効果予測能を増大させることを目指す。

### 3. 研究の方法

本研究では、下記の (1) ~ (5) を実施する (図 1)。

#### (1) 肺腺癌細胞株の全長 cDNA シーケンシ解析

まず、豊富な多層オミクス情報 (Suzuki A et al. 2014 *NAR*) が付随した肺腺癌細胞株から RNA を抽出し、ナノポアシーケンサー MinION による全長 cDNA シーケンシ解析を実施する。

Minimap2 (Li H. 2018 *Bioinformatics*) を用いて、得られた全長 cDNA シーケンシデータをヒトゲノム参照配列にマッピングを行う。RefSeq や GENCODE に登録されている転写産物をリファレンスとし、未知の転写産

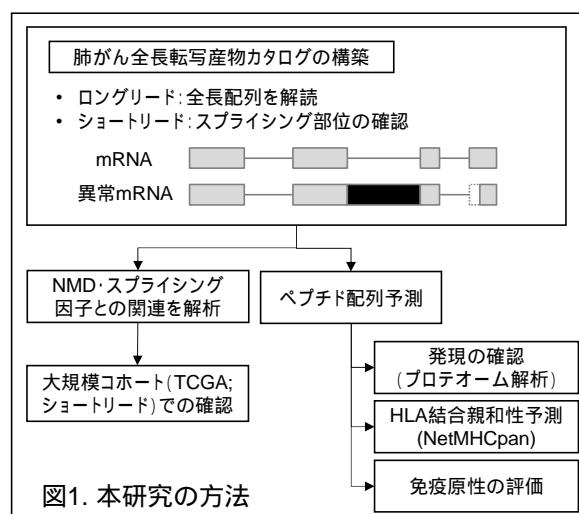


図1. 本研究の方法

物バリエーションを異常 mRNA として同定する。MinION から得られたシーケンス情報はその塩基精度が 90%程度と低いので、新たに見出されたスプライシング部位は、取得済みのショートリード RNA-seq データで確認し、より正確な全長構造を同定する。

#### (2) 異常 mRNA 蓄積に関連する因子の探索

NMD や RNA スプライシングに関わる因子に異常が生じると、異常 mRNA の排除機構が正常に機能せず、異常 mRNA が蓄積することが予想される。本研究では、NMD および RNA スプライシング関連因子におけるゲノム変異ステータスを解析し、異常 mRNA スプライスバリエーションの数とどのような関係性があるか調べる。

#### (3) 肺がん臨床検体の全長 cDNA シーケンス解析

非小細胞肺がん検体について、全長 cDNA シーケンス解析を行う。細胞株の解析で構築した解析パイプラインを駆使してスプライスバリエーションを検出する。非がん部で検出されない、もしくは、低発現であるものを異常スプライスバリエーションとして抽出する。

#### (4) 異常 mRNA 由来のペプチド配列予測と評価

得られた異常 mRNA 全長配列より、翻訳されるアミノ酸配列の予測を行う。この際に、点変異 (ミスセンス・ナンセンス変異、インフレーム・フレームシフト挿入および欠失) の存在も考慮して配列を予測する。一部の細胞株に関しては、プロテオーム解析を実施し、異常 mRNA 由来のペプチドが実際に発現しているかどうか確認する。また、NetMHCpan を用いて、HLA との結合親和性を予測する。さらに、ペプチドが細胞障害性 T 細胞 (CTL) を誘導できるか、その免疫原性を調べる。

#### (5) TCGA の解析

(1) および (3) で得られた異常スプライスバリエーションを合わせて、肺がん異常 mRNA のカタログとし、そのスプライシング部位が The Cancer Genome Atlas (TCGA) の RNA-seq データで検出できるか確認する。また、NMD やスプライシング因子の変異と、異常スプライシング部位の数の関係性を解析し、異常 mRNA 量をサロゲートする因子を同定する。

### 4. 研究成果

#### (1) 肺腺癌細胞株および非小細胞肺癌臨床検体の全長 cDNA 配列カタログの構築

肺腺癌細胞株の全長 cDNA をナノポアシーケンサー MinION で解読し、全長 cDNA シーケンスデータを取得した。取得したデータは DNA Data Bank of Japan (DDBJ) に登録している。また、得られたシーケンスデータは、多くの mRNA の全長をカバーするのに十分なリード長であることが確認できた。

得られた全長 cDNA 配列から RefSeq に登録のないスプライシングバリエーションを抽出したところ、平均 476 個 (323-725 個) のスプライスバリエーションが検出された (図 2)。これらのスプライスバリエーションの中には、exon skipping や intron retention などが含まれていた。さらに、これらの異常がコンベネーションとして一つの転写産物の中に検出される例が見られ、

これらは従来のショートリード解析ではその組み合わせを検出できないことから、ロングリード技術による全長 cDNA の解読を実施することがきわめて重要であることが示された。

また、構築した解析パイプラインを駆使して、非小細胞肺がん臨床検体の全長 cDNA シーケンスデータから、同様に異常スプライスバリエーションを検出した。156-252 個の異常スプライスバリエーションが検出された。

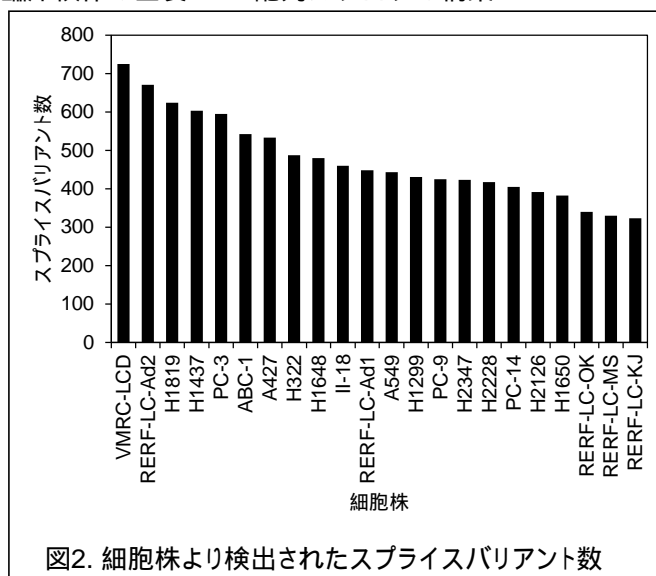


図2. 細胞株より検出されたスプライスバリエーション数

#### (2) NMD およびスプライシング関連因子の、異常 mRNA 蓄積に対する影響の評価

異常スプライスバリエーションの蓄積に寄与すると考えられる NMD やスプライシング関連因子の解析を実施した。VMRC-LCD 細胞株では、NMD 複合体の主要因子である *UPF1* 遺伝子に変異が検出され、細胞株の中で最も多くのスプライスバリエーションを有していた (図 2)。NMD は、今回着目しているスプライシング異常やナンセンス変異、フレームシフト変異などにより生じた異常な終止コドンを認識し、mRNA を分解する品質管理機構であり、VMRC-LCD に見られるような NMD 因子の機能喪失は異常 mRNA 蓄積の原因になりうると考えている。そこで、*UPF1* が異常スプライスバ

リアントの蓄積に寄与するかどうかを評価するために *UPF1* 遺伝子のノックダウン実験を実施した。*UPF1* 遺伝子に変異を有しない A549 細胞株にて *UPF1* の発現を siRNA によりノックダウンしたところ、異常スプライスバリエーションの割合が増加した。

また、RNA スプライシング因子である *SF3B1* 遺伝子についても同様の解析を実施した。*SF3B1* 遺伝子はさまざまながん種において変異が報告されている。A549 細胞株を用いて、*SF3B1* をノックダウンしたところ、exon skipping が生じた異常バリエーションの割合が増加した。

これらの結果から、NMD や RNA スプライシング因子の機能異常が、がん細胞中の異常 mRNA の産生と蓄積に寄与していることが示唆され、これら遺伝子の変異や発現ステータスを解析することは、異常 mRNA のパターンを理解するうえで重要であると考えられた。

#### (3) 異常 mRNA 由来のペプチド配列からの新規ネオ抗原候補の同定と評価

検出された異常スプライスバリエーションから翻訳されると考えられるペプチドについて、実際に翻訳され、ネオ抗原となりうるかを評価した。まず、一部の細胞株で、京都大学の協力のもと、プロテオーム解析 (LC/MS/MS) を実施した。その結果、異常 mRNA 由来と考えられるペプチド配列がいくつか検出された。また、フレームシフト変異や異常スプライスバリエーション由来のペプチドからは、ミスセンス変異やインフレーム変異と比較して、より多くの異常ペプチド配列候補が同定された。

次に、これらペプチドにおける HLA との結合親和性を調べるために、NetMHC というツールを用いて解析を実施した。その結果、フレームシフト変異や異常スプライスバリエーション由来のペプチドは、親和性スコアが高い傾向を示した。

さらに、予測されたネオ抗原候補ペプチドの免疫原性をさらに詳しく評価した。ネオ抗原は HLA によって抗原提示細胞に提示されて認識され、CTL を誘導する。そこで、抽出したネオ抗原候補ペプチドが、CTL を誘導できるかを評価するために、国立がん研究センターの協力のもと、ペプチドの接種実験を 3 回行った HLA-A24 トランスジェニックマウスの脾臓細胞を抽出し、ELISpot アッセイを実施した。フレームシフトもしくは異常スプライスバリエーション由来の候補ペプチドを 17 種類選抜して実験に使用した。実験の結果、8 種類のペプチドで反応が見られた。

これらの結果から、異常スプライスバリエーションから翻訳されるペプチドについて、その一部はタンパク質として発現し、抗原提示され、免疫応答を活性化しうることが示唆された。

#### (4) 大規模公共データでの検証

本研究で細胞株および臨床検体の解析より得られた異常スプライスバリエーションがどの程度の症例で発現しているのか、また、それらの量に影響を及ぼすと考えられる NMD およびスプライシング関連因子の異常はそれらと関係性があるのかを評価するために、より大規模なデータセット (TCGA) を用いた検証を実施した。

まず、本研究にて細胞株と臨床検体データから抽出した異常スプライスバリエーション情報から、肺がんの異常 mRNA におけるスプライシング部位のカタログを作製した。次に、TCGA の肺腺がん検体 436 症例 (TCGA-LUAD) において、異常スプライシング部位が検出されるかどうか、RNA-seq データを解析して検証した。TCGA で取得されている RNA-seq はショートリードであるため、異常スプライシング部位をカバーしたリードを抽出し、評価に使用した。まあ、非がん部では発現していないがん部特異的異常スプライシング部位を抽出するため、非がん部検体由来の TCGA RNA-seq データや GTEx データベースにて検出される配列は除去した。

さらに、NMD および RNA スプライシング関連因子のゲノム変異ステータスも解析した。その結果、NMD 関連因子に変異を持つ症例では、異常スプライシング部位の数が多いことが分かった。他のがん種では有意な関係性が見られなかったことから、他がん種への応用の際には、別途がん種ごとの異常スプライスバリエーションのカタログ作製が必要となることが示唆された。

これらのことから、(2) でも示唆されていたように、NMD 関連因子の変異ステータスが異常スプライスバリエーション数のサロゲートマーカーとなりうることが考えられた。

本研究では、ロングリードシーケンシング技術を駆使し、がん細胞中の異常 mRNA の全長構造を網羅的に同定する手法を開発した。また、NMD やスプライシング関連因子の異常が、異常 mRNA の蓄積に寄与していることを示唆した。さらに、異常 mRNA から翻訳されうるペプチド配列は、一部ではあるが、実際にタンパク質として発現し、HLA と結合、免疫細胞に誘導しうることを示した。これらの結果は、異常 mRNA が、がん組織における免疫応答にとって重要な、ネオ抗原候補となりうることを示している。

本研究成果については、Genome Biology 誌より下記の論文を発表した。

Oka M, Xu L, Suzuki T, Yoshikawa T, Sakamoto H, Uemura H, Yoshizawa AC, Suzuki Y, Nakatsura T, Ishihama Y, Suzuki A, Seki M. Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer. 2021 *Genome Biology* 22(1):9.

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 1件）

1. 著者名 Oka Miho, Xu Liu, Suzuki Toshihiro, Yoshikawa Toshiaki, Sakamoto Hiromi, Uemura Hayato, Yoshizawa Akiyasu C., Suzuki Yutaka, Nakatsura Tetsuya, Ishihama Yasushi, Suzuki Ayako, Seki Masahide	4. 巻 22
2. 論文標題 Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer	5. 発行年 2021年
3. 雑誌名 Genome Biology	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s13059-020-02240-8	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計0件

〔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
---------	---------