

令和 2 年 6 月 8 日現在

機関番号：12602

研究種目：若手研究

研究期間：2018～2019

課題番号：18K15305

研究課題名（和文）データベースを活用したがん体細胞バリエーションの大規模機能解析

研究課題名（英文）Massive identification of somatic variants in exonic splicing enhancers using public databases

研究代表者

谷本 幸介 (Tanimoto, Kousuke)

東京医科歯科大学・難治疾患研究所・助教

研究者番号：60611613

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：公共データベースThe Cancer Genome Atlasの32のがん種にわたる9,635症例の体細胞バリエーションとRNA-seqデータを用いて、Exonic Splicing Enhancer(ESE)の機能を阻害する体細胞バリエーション(ESE-disrupting variants)を網羅的に特定することを試みた。その結果、646のESE-disrupting variantsが特定された。既存のツールでは病原性が高くないと推測されるバリエーションであっても、ESE機能を阻害することでエクソンスキップを引き起こし、タンパク質機能に大きな影響を与える可能性が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

ゲノム医療においてVariants of Uncertain Significanceと呼ばれる意義不明バリエーションの存在が課題の一つであるが、本研究はスプライシング制御という観点から体細胞バリエーションの意義を明らかにすることができたため、ゲノム医療推進の一助になると期待される。

研究成果の概要（英文）：Using somatic variants and RNA-seq data from 9,635 patients across 32 The Cancer Genome Atlas (TCGA) projects, we identified 646 exonic splicing enhancer (ESE)-disrupting variants. Existing bioinformatics tools could not predict the pathogenicity of ESE-disrupting variants identified in this study, although these variants could cause exon skipping.

研究分野：ゲノム科学

キーワード：スプライシング がん体細胞バリエーション TCGA データベース

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

シーケンス技術の進展により、あらゆるがん種における体細胞バリエントが解析され、各がん種におけるゲノム変化の全体像が明らかになっている(Vogelstein, B. et al. *Science*, 2013)。大量のがん体細胞バリエント情報が入手できるようになった現在、それぞれの体細胞バリエントの機能を明らかにしていくことがこれからの課題である。体細胞バリエントのなかでもアミノ酸変化を伴う非同義置換については、アミノ酸変化がタンパク質の機能に与える影響を予測するツールが多数開発され、いくつかの遺伝子については分子生物学実験による検証もなされている。しかし、現在報告されている全ての体細胞バリエントの機能を実験的に検証することは現実的には不可能であり、その意義が明らかでないバリエントが多数残されているのが現状である。

2. 研究の目的

本研究では公共データベースの大規模データを活用することにより、スプライシング制御に影響する可能性のあるがん体細胞バリエントを、同義置換および非同義置換を問わず多症例かつゲノムワイドに明らかにすることを試みる。

3. 研究の方法

公共データベース The Cancer Genome Atlas (TCGA) の 32 のがん種にわたる 9,635 症例の体細胞バリエントと RNA-seq データを用いて、Exonic Splicing Enhancer (ESE) の機能を阻害する体細胞バリエント(以下 ESE-disrupting variant とする)をバイオインフォマティクス的手法により網羅的に特定することを試みた。解析の workflow を図 1 に示す。

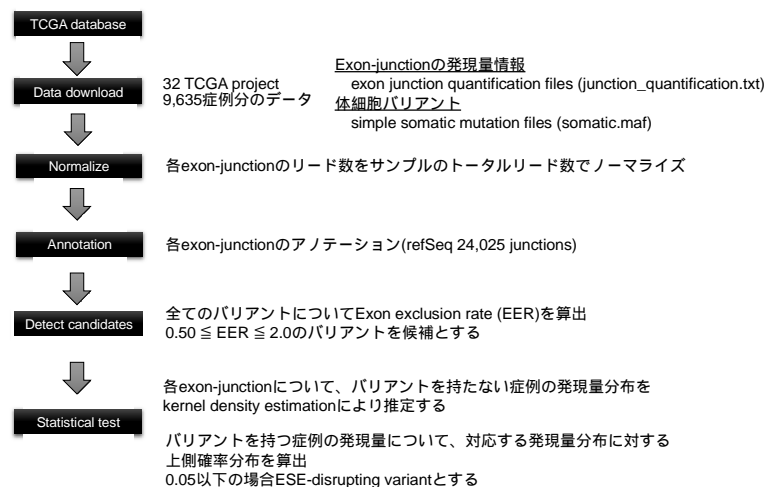


図 1 解析の workflow

ESE-disrupting variant を検出するため、バリエントを持つエクソンがスキップされる度合いを示す指標として Exon Exclusion Rate(EER)を定義し、候補の絞り込みに用いた(図 2)。

Exon Exclusion Rate (EER)の定義

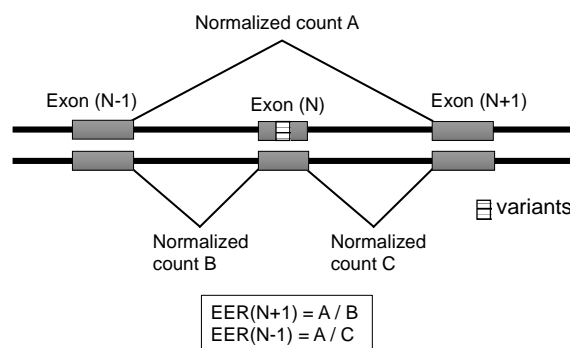


図 2 Exon Exclusion Rate(EER)の定義

候補となったバリエントが有意にエクソンスキップを起こしているかどうかを検定するため、バリエントを持たない症例群の発現量分布を kernel density estimation により推定し、その

分布に対しての上側確率密度が 0.05 以下である場合、そのバリエントを ESE-disrupting variant とした(図 3)。

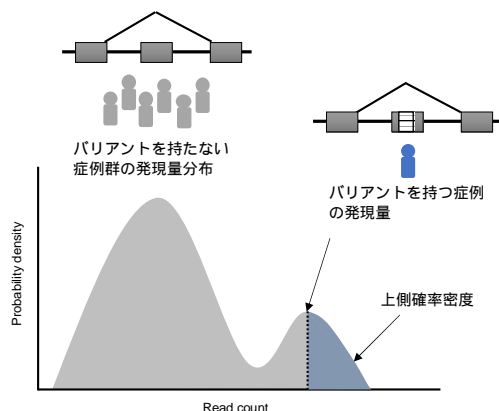


図 3 候補バリエントの検定方法

4. 研究成果

TCGA に収載されている 156,794 の体細胞バリエント(32 プロジェクト、9,635 症例)を上述の方法で解析した結果、646 の ESE-disrupting variant が特定された(表 1)。がん種による検出率の違いは見られなかった。

表 1 解析バリエント数と解析結果

TCGA project		症例数	バリエント数	ESE-disrupting variants	percentage
副腎皮質がん	ACC	79	538	0	0
褐色細胞腫および傍神経節腫	PCPG	179	336	2	0.60
胆管がん	CHOL	36	308	2	0.65
膀胱尿路上皮がん	BLCA	408	5,084	27	0.53
グリオブラストーマ	GBM	154	4,600	21	0.46
ローグレードグリオーマ	LGG	516	2,228	5	0.22
乳がん	BRCA	1,092	9,079	34	0.37
子宮頸部扁平上皮癌、子宮頸がん	CESC	305	3,742	17	0.45
結腸腺がん	COAD	431	11,507	38	0.33
直腸腺がん	READ	151	2,555	12	0.47
食道がん	ESCA	182	2,638	11	0.42
ぶどう膜黒色腫	UVM	80	158	0	0
頭頸部扁平上皮がん	HNSC	520	4,068	11	0.27
嫌色素性細胞	KICH	66	258	1	0.39
腎明細胞がん	KIRC	531	1,227	5	0.41
乳頭状腎細胞がん	KIRP	290	1,202	2	0.17
肝細胞がん	LIHC	371	38,398	94	0.24
肺腺がん	LUAD	516	7,204	27	0.37
肺扁平上皮がん	LUSC	500	9,198	40	0.43
びまん性大細胞型B細胞性リンパ腫	DLBC	48	773	3	0.39
漿液性嚢胞腺癌	OV	304	1,822	9	0.49
膵臓がん	PAAD	177	2,696	8	0.30
中皮腫	MESO	87	175	0	0
前立腺がん	PRAD	497	1,451	7	0.48
皮膚黒色腫	SKCM	104	2,229	16	0.72
肉腫	SARC	258	1,854	11	0.59
胃がん	STAD	379	7,299	38	0.52
精巣胚細胞腫瘍	TGCT	147	344	0	0
胸腺腫	THYM	120	1,048	2	0.19
甲状腺がん	THCA	505	494	0	0
子宮体がん	UCEC	545	31,875	201	0.63
子宮癌肉腫	UCS	57	406	2	0.49
total		9,635	156,794	646	0.41

特定された ESE-disrupting variant の偽陽性率を検定するために、permutation test を施行した。1000 回の施行で検出された ESE-disrupting variant の平均は 6.493 であったことから、本手法の偽陽性率は 1%程度であると考えられる(図 4)。

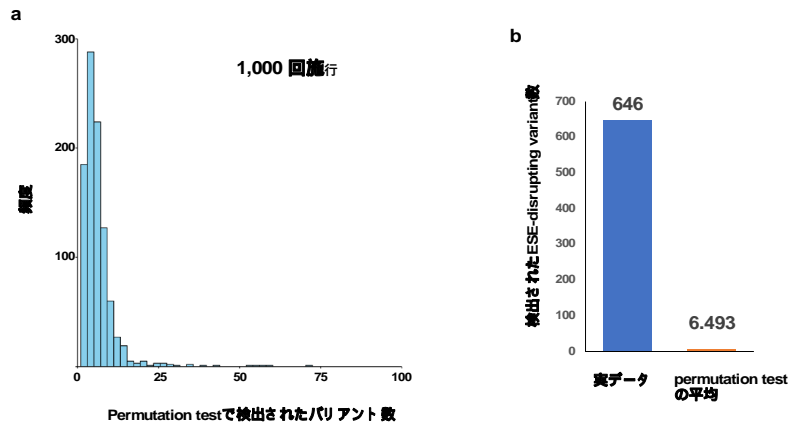


図 4 permutation test 結果 (a)検出されたバリエーション数の分布 (b)実データとの比較

図 5 に検出された ESE-disrupting variant の一例を示す。サンプル ID A2HD の *APMAP* は exon 8 の chr20:24949636 に G>A のバリエーションが存在し、EER(N-1)および EER(N+1)はそれぞれ 0.71 と 0.83 であった(図 5a)。Exon8 をスキップするリード数は 195 であり、これは同一がん種で *APMAP* の exon8 にバリエーションを持たない症例群の発現量分布からは大きく外れた数値であった(図 5b)。ESE finder による既知の ESE binding motif の探索結果から、このバリエーションは SRSF2 の結合モチーフを変化させることで結合を阻害している可能性が示唆された(図 5c)。このバリエーション周辺を阻害するモルフォリノオリゴを作製し、HeLa 細胞株および HEK293 細胞株に処理したところ、exon8 がスキップされた転写産物が生じることが実験的に確認された(図 5d)。

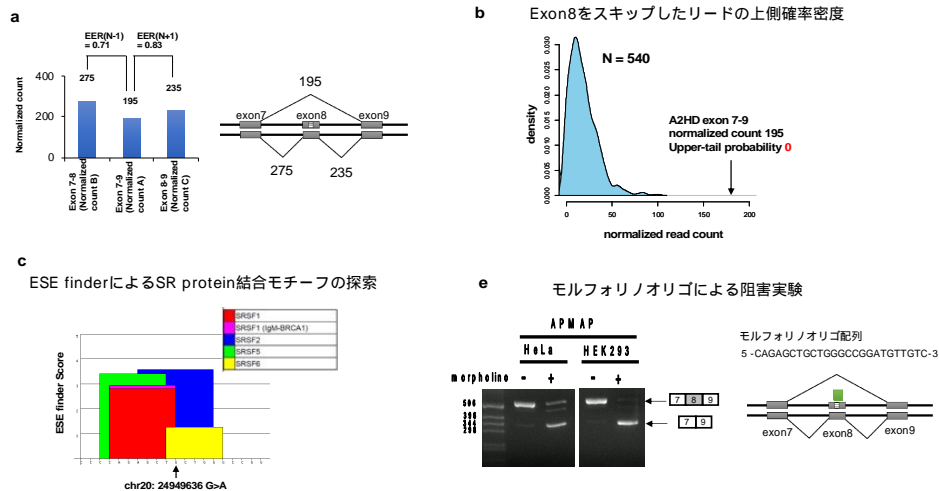


図 5 検出された *APMAP* の ESE-disrupting variant

本研究で ESE-disrupting variant として特定された 646 バリエーションのタイプを調べたところ、missense バリエーションが最も多く (51%)、その次に多いのはアミノ酸変化を伴わない synonymous (silent) バリエーションであった (18%) (図 6a)。バリエーションの病原性を予測するためのバイオインフォマティクスツールである REVEL を用い、特定した ESE-disrupting variant のスコアを算出した結果、大部分は病原性が低いと判断されるスコアであった(図 6b)。これらの結果より、従来の手法では病原性が低いものとして見過ごされかねないバリエーションであっても、ESE 機能を阻害することでエクソンスキップを引き起こし、タンパク質機能に大きな影響を与える可能性があることが示唆された。

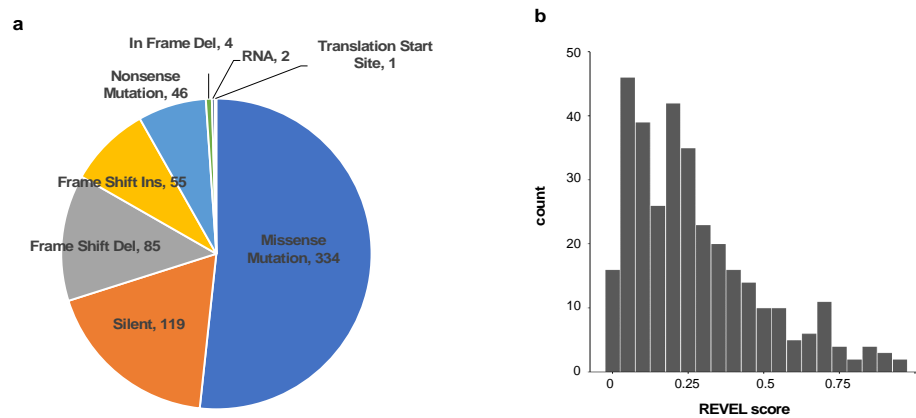


図6 特定された ESE-disrupting variant のタイプと REVEL ツールによる病原性スコア

Nonsense バリエントを有する転写産物は、バリエントの位置が最も 3' 側のエクソン-エクソンジャンクションの 50-55 塩基下流までであれば、通常は NMD (nonsense mediated mRNA decay) により除去されるため、病原性に影響は少ないものと考えられる。しかし、遺伝子ごとの発現量データを比較したところ、ESE-disrupting variant と判断された nonsense バリエントは、そうでないものに比べて該当する遺伝子自体の発現量が低下していない傾向が見られた(図7)。これは、ESE-disrupting variant がエクソンスキップを引き起こしたことにより、本来は NMD で除去されるはずの転写産物が除去されていない可能性があることを示唆している。

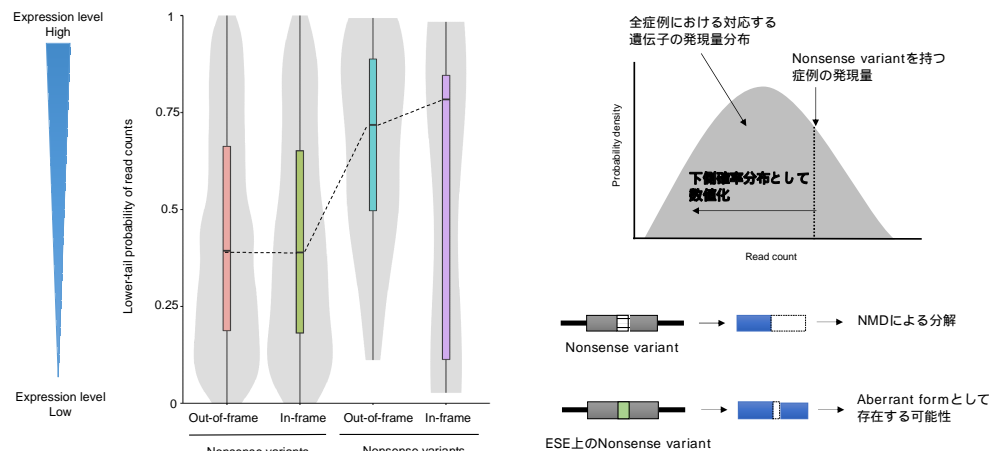


図7 nonsense バリエントを有する遺伝子の発現量分布

本研究では TCGA に登録されている症例の RNA-seq と体細胞バリエント情報の統合解析を行うことにより、ESE を阻害してエクソンスキップを生じさせる可能性のあるバリエントを特定することができた。ゲノム医療において variant of uncertain significance (VUS) と呼ばれる意義不明変異の存在が問題となっているが、本手法はスプライシング制御の観点から VUS の解消に寄与できるものと期待される。

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 1件）

1. 著者名 Kousuke Tanimoto, Tomoki Muramatsu, Johji Inazawa	4. 巻 8
2. 論文標題 Massive computational identification of somatic variants in exonic splicing enhancers using The Cancer Genome Atlas	5. 発行年 2019年
3. 雑誌名 Cancer Medicine	6. 最初と最後の頁 7372-7384
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cam4.2619	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
---------------------------	-----------------------	----