

令和 3 年 6 月 22 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K08564

研究課題名(和文) 乳癌のRNAseq解析手法の最適化による有用性の高い変異の同定

研究課題名(英文) Optimization of the RNAseq analysis to identify valuable variation of the breast cancer

研究代表者

坂東 裕子 (Bando, Hiroko)

筑波大学・医学医療系・准教授

研究者番号：00400680

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：癌固有の遺伝子変異から生じるネオアンチゲンは抗腫瘍免疫応答の中心的役割を担うと考えられている。一般的にネオアンチゲンの予測には腫瘍の全エクソンシーケンス(WES)とRNAシーケンス(RNA-seq)、正常サンプルのWESデータを用いるが、本研究では乳癌組織のRNA-seqと正常部のWESを用いてネオアンチゲンを予測する手法を確立した。本手法は発現している体細胞変異のみが検出されるため効率的なネオアンチゲンの予測手法であり、腫瘍のWESの省略することでコスト削減にも繋がるため費用対効果の高い手法であることが明らかとなった。

研究成果の学術的意義や社会的意義

近年の次世代シーケンス技術、バイオインフォマティクスの進歩によりネオアンチゲン予測手法の精度は向上しているものの、偽陽性率が依然として高いことや、複数のシーケンスデータを必要とすることからコスト面でも課題がある。ネオアンチゲンの予測は、免疫療法の治療ターゲットの探索や免疫チェックポイント阻害薬などの治療効果予測において重要である。腫瘍部のWESを省略する本手法は、より効率的で費用対効果の高い手法になる可能性があり、実臨床におけるネオアンチゲン予測手法として有用であると考えられる。

研究成果の概要(英文)：Neoantigens have attracted attention as biomarkers or therapeutic targets. However, accurate prediction of neoantigens is still challenging, in terms of its accuracy and cost. Variant detection using RNA sequencing (RNA seq) data has been reported to be a low accuracy but cost effective tool, but the feasibility of RNA seq data for neoantigen prediction has not been fully examined. In the present study, we used whole exome sequencing (WES) and RNA seq data of tumor and matched normal samples from six breast cancer patients to evaluate the utility of RNA seq data instead of WES data in variant calling to detect neoantigen candidates. We found that the method combining tumor RNA seq data and normal WES data (Combination method) could detect neoantigen candidates that have higher expression and rich variant transcripts, and this method may be an efficient and cost effective strategy alternative to the conventional method (DNA method).

研究分野：乳癌、薬物療法、遺伝子

キーワード：ネオアンチゲン 乳癌 RNAシーケンス 全エクソンシーケンス バイオインフォマティクス

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

癌固有の遺伝子変異から生じるネオアンチゲンは抗腫瘍免疫応答の中心的役割を担うと考えられている。一般的にネオアンチゲンの予測には腫瘍の全エクソンシーケンス (WES) と RNA シーケンス (RNA-seq) 正常サンプルの WES データを用いる。近年の次世代シーケンス技術、バイオインフォマティクスの進歩によりネオアンチゲン予測手法の精度は向上しているものの、偽陽性率が依然として高いことや、複数のシーケンスデータを必要とすることからコスト面でも課題が多い。ネオアンチゲンの予測は、免疫療法の治療ターゲットの探索や免疫チェックポイント阻害薬などの治療効果予測において重要であり、より効率的で費用対効果の高い予測手法の開発が求められている。

2. 研究の目的

本研究の目的はより効率的で費用対効果の高いネオアンチゲン予測手法を開発することである。今回我々は遺伝子発現情報のみならず塩基配列情報も有する RNA-sequencing (RNA-seq) の特性に着目し、Whole-exome sequencing (WES) データの代わりに RNA-seq データを用いて体細胞変異解析を行うことによりネオ抗原を予測することが可能かについて検討を行った。

3. 研究の方法

筑波大学附属病院で治療を行なった乳癌患者 6 人の乳癌組織と正常検体 (血液または正常乳腺組織) の NGS データを用いて、以下の 3 つ手法で解析を行なった。

腫瘍と正常共に WES データを用いる従来手法 (DNA method: Dm)

腫瘍と正常共に RNA-seq データを用いる手法 (RNA method: Rm)

腫瘍の RNA-seq データと正常の WES データを組み合わせる手法 (Combination method: Cm)

(1) マッピングおよびデータの前処理にはハイスループットシーケンスデータの変異解析において現在最もスタンダードなツールである Genome Analysis Toolkit (GATK) の推奨するワークフロー (GATK Best Practices) を用いた。WES データのマッピングには BWA-MEM (version 0.7.17) を、RNA-seq データのマッピングには STAR (version 2.7.3a) を使用した。

(2) Dm の体細胞変異解析には Mutect2 (GATK) を使用した。Rm、Cm では VarScan2 (version 2.4.4) を用いて体細胞変異を行った後、RNA-seq データを用いた体細胞変異解析でよく見られる偽陽性をフィルタリングするため、RNA-seq 用に開発されたソフトウェアである SNPiR のフィルターを適用した。遺伝子変異解析で得られた variant call format (VCF) ファイルは Ensembl Variant Effect Predictor (VEP version 99) を用いてアノテーションした。多型の高度な領域におけるアライメントはエラーが発生しやすく、特に免疫グロブリン (*IGH@*: 14q32.33、*IGK@*: 2p11.2、*IGL@*: 22q11.2) や *HLA* 遺伝子 (6p21) の体細胞変異解析には特殊な解析ツールを要するため、これらの遺伝子に位置する変異はその後の解析から除外した。

(3) ネオ抗原予測には pVACseq (version 1.5.9) を用いた。カバレッジと variant allele frequency (VAF) に関するフィルターを除去した以外は既定のパラメーターを使用して、HLA クラス I (A、B または C) に結合する 7~11mer のエピトープを予測し、複数の結合予測アルゴリズムを用いて IC50 の中央値が 500 nM 以下のエピトープをネオ抗原候補とした。各患者の HLA Class I アレルは正常 DNA FASTQ ファイルから HLA-HD (version 1.2.0.1) を用いて決定した。

(4) 各手法で特異的/共通のネオ抗原候補についてカバレッジや発現量、塩基置換の種類などの特徴を比較検討し、Rm や Cm のネオ抗原予測手法としての有用性を検証した。

4. 研究成果

(1) 各手法における体細胞変異解析の結果

アノテーション後、免疫グロブリンおよび *HLA* 遺伝子に位置する変異を除去し、最終的に残った体細胞変異数は Dm 3443 個、Cm 401 個、Rm 54 個であった。手法共通の体細胞変異数は少なく、全ての手法で共通して検出された体細胞変異は 5 個のみであった (図 1)。

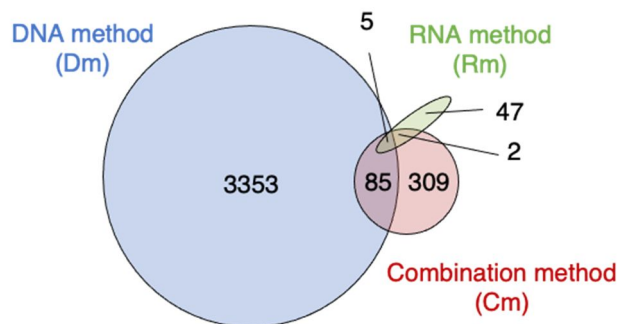


図1 各手法における体細胞変異解析の結果

(2) RNA method で検出された体細胞変異の特徴

Rm と Cm の違いは正常サンプルが RNA か DNA かであるため、はじめに Rm と Cm の結果を比較した。まず、Rm と Cm それぞれ固有のバリエーションについて、正常組織の RNA-seq のカバレッジの中央値を検証したところ、Rm 4.0 (四分位間範囲 [IQR], 3.0-5.0)、Cm 0 (0-1.0) ($P < 0.001$, Mann-Whitney U 検定) と Cm で有意に低く、Cm 固有のバリエーションのうち 71% (281/394) は正常 RNA のカバレッジが 0 であった。Cm 固有のバリエーションの多くが、正常 RNA のカバレッジが存在しない、または低い場合 Rm では検出されなかったと考えられた。

続いて Rm 固有のバリエーションが Cm で検出されなかった原因を検証した。Rm 固有のバリエーションのうち 68% (32/47) は正常 DNA にも変異アレルが検出され、Cm では生殖細胞系列変異あるいはヘテロ接合性喪失 (Loss of heterozygosity : LOH) に分類されていた。これは、生殖細胞系列変異であるにも関わらず、アレル特異的な発現やカバレッジの低さにより正常 RNA に変異アレルが認められなかったため、体細胞変異として誤って検出されたことが原因であった。Rm 固有のバリエーションの 30% (14/47) は、正常 DNA のカバレッジが低かったため Cm では検出されず、残りの 1 個は Cm ではアーチファクトと分類された。Rm 固有の体細胞変異の大多数 (45/47) が dbSNP (build146) に登録されており、Rm では生殖細胞系列変異を体細胞変異として誤って検出するリスクが高いことが明らかとなった。

(3) DNA method と Combination method で検出された体細胞変異の特徴

Dm と Cm の結果を比較すると、Dm 固有の体細胞変異は 3353 個、Cm 固有の変異は 311 個、共通の変異は 90 個であった。各手法固有のバリエーションが他の手法で検出されなかった原因を探索するため、Dm 固有のバリエーションの腫瘍 RNA におけるカバレッジと VAF、Cm 固有のバリエーションの腫瘍 DNA におけるカバレッジと VAF を検証したところ、Dm 固有のバリエーションの多くは腫瘍 RNA でのカバレッジが低く、74% (2471/3353) がカバレッジ 0 であった。それに対して Cm 固有のバリエーションは腫瘍 DNA のカバレッジの中央値 57.0 (IQR, 5.5-168.5) と高値であったが、74% (231/311) は腫瘍 DNA に変異アレルを認めなかった。つまり、Dm 固有のバリエーションの大部分は腫瘍 RNA にカバレッジがないことにより Cm で見落とされ、Cm 固有のバリエーションは腫瘍 DNA に変異アレルを認めないことが原因で生じることが明らかとなった。

次に、Dm 固有、Cm 固有、共通のバリエーションにおける塩基置換の内容について検討した。すでに SNPiR によるフィルタリングで既知の RNA 編集については除外していたにも関わらず、ヒトの RNA 編集の主な塩基置換のパターンである A-to-G と T-to-C の割合は、Cm 固有のバリエーションにおいて他の 2 群よりも有意に高いことが明らかとなった (vs. Dm 固有 $P < 0.001$, vs. 共通 $P = 0.003$, Bonferroni 補正)。

バリエーションを含む遺伝子の、腫瘍組織における遺伝子発現量を各手法固有のバリエーションと手法共通のバリエーションで比較したところ、Dm 固有、Cm 固有、共通のバリエーションの transcripts per million (TPM) の中央値はそれぞれ 4.8 (IQR, 0.07-22.3)、21.9 (7.5-51.8)、38.7 (17.9-102.5) であり、Kruskal-Wallis 検定で 3 群間に有意差が認められた ($P < 0.001$)。多重比較検定では共通のバリエーションの遺伝子発現量が最も高く、Dm 固有バリエーションが最も低いことが示された (いずれも $P < 0.001$, Mann-Whitney U 検定 (Bonferroni 補正))。

(4) DNA method と Combination method で検出されたネオ抗原候補の比較

Dm 固有のネオ抗原候補は 154 個、Cm 固有のネオ抗原候補は 28 個、両手法で検出されたネオ抗原候補は 26 個であった。

IGV を用いて目視で確認したところ、大多数の Dm 固有のネオ抗原候補は、腫瘍 RNA のカバレッジが 0 か ($n=29$)、腫瘍 RNA 中に変異アレルが存在しないか少ない ($n=97$) ことが原因で Cm では検出されなかった。Dm 固有のネオ抗原候補のうち 12 個は indel (2 個の insertion と 10 個の deletion) によるものであったが、うち 8 個の deletion は Cm でソフトクリップ (リードの 5' または 3' 末端の、リファレンス配列に一致しない塩基配列) と誤認され、indel として検出されなかった。次に、28 個の Cm 固有のネオ抗原候補の特徴を検討した。2 個のネオ抗原候補は、正常部・腫瘍部の WES のリードにバリエーションが集簇しており (clustered-event) これはマッピングエラーによる偽陽性と考えられた。この 2 個は後の解析から除外した。26 個のネオ抗原候補のうち、1 個は DNA にも変異アレルを有していたにも関わらず (変異アレルカバレッジ 6、VAF 4.2%) Dm では体細胞変異として検出されなかった (図 2)。腫瘍 DNA に変異アレルリードを有さなかった残りの 25 個のネオ抗原候補のうち 5 個は、A-to-G または T-to-C 変異であった。

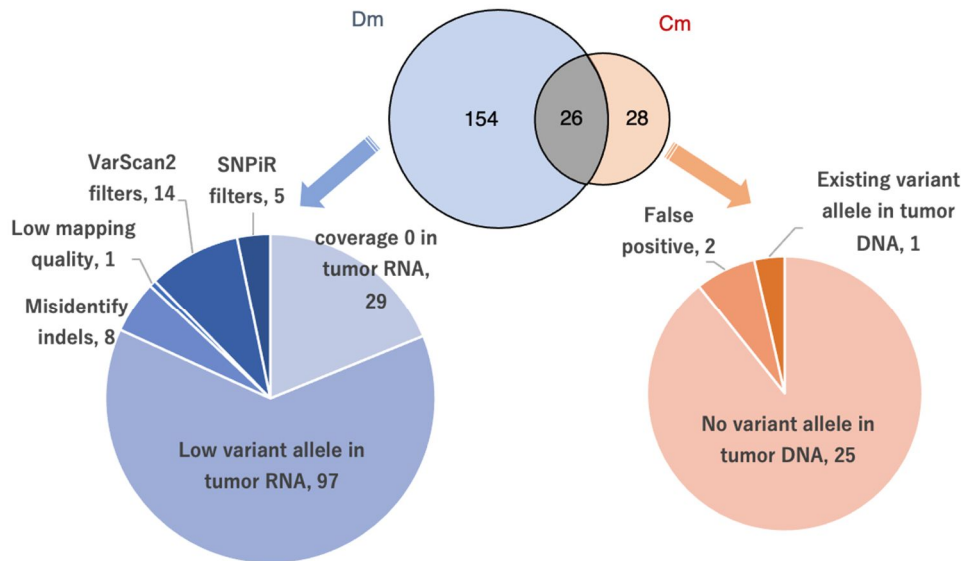


図2 各手法固有のネオ抗原候補の特徴

続いて手法固有および共通のネオ抗原候補の遺伝子発現レベルを比較した。Dm 固有、Cm 固有、共通のネオ抗原候補の TPM の中央値はそれぞれ 6.0 (IQR、2.7-16.7)、11.6 (6.7-23.4)、22.0 (9.6-44.3) であり、3 群間には統計学的有意差が認められた ($P < 0.001$ 、Kruskal-Wallis 検定)。共通のネオ抗原候補は Dm 固有のネオ抗原候補よりも遺伝子発現量が高かったが ($P < 0.001$ 、Mann-Whitney U 検定 (Bonferroni 補正))、Cm 固有のネオ抗原候補と Dm 固有 ($P = 0.29$)、共通のネオ抗原候補 ($P = 0.23$) との間には有意差は認められなかった。

ネオ抗原候補としては腫瘍組織に該当遺伝子が発現しているだけでなく、変異アレルが十分に発現していることも必要な条件である。腫瘍 RNA における変異アレルのカバレッジの中央値は、Dm 固有のネオ抗原候補で 0 (IQR、0-1.0)、Cm 固有のネオ抗原候補で 2.0 (1.3-3.0)、共通のネオ抗原候補で 5.5 (3.3-9.8) であり、3 群間には Kruskal-Wallis 検定で有意差が認められ ($P < 0.001$)、多重比較検定では共通のネオ抗原候補の変異アレルカバレッジが最も高く、Dm 固有のネオ抗原候補が最も低かった。最後に、腫瘍 RNA における変異アレルリード数と遺伝子発現レベルを組み合わせて評価を行った。遺伝子発現が非常に高い ($TPM > 1000$) Dm 固有のネオ抗原候補 ($n=5$) はすべてミトコンドリア DNA に由来するものであり、このうち 3 個は腫瘍 RNA に豊富な変異アレルリードを有していたが、Cm では複数部位にマッピングされるリードとして BLAT でフィルターされており、マッピングエラーの可能性が有ると考えられた。この 3 個のネオ抗原候補を除き、遺伝子発現量、変異転写産物の十分なネオ抗原候補 ($TPM > 10$ 、および腫瘍 RNA の変異アレルリード > 5) に限って検討すると、68% (13/19) が Cm においても同定されることが明らかとなった (図 3)。

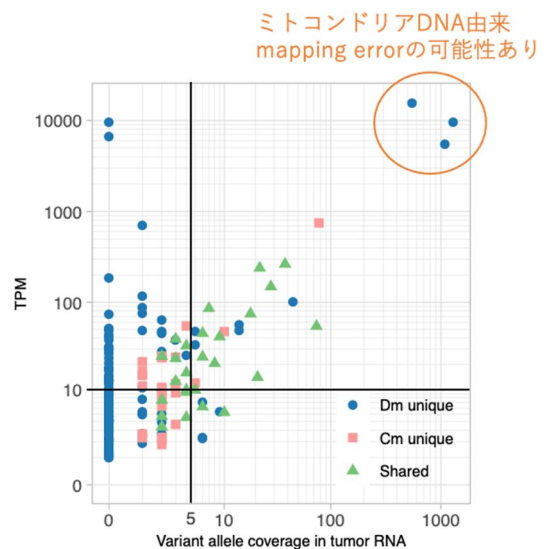


図3 ネオ抗原候補の腫瘍RNAにおける変異アレルリード数 (x軸) と遺伝子発現量 (y軸) の分布

(5) 結論

ネオ抗原候補の条件として、遺伝子発現があることに加え、変異アレルが転写されることが挙げられる。よって変異転写産物がない体細胞変異を見落とすことはネオ抗原探索という点においては許容されると考える。発現量・変異転写産物が十分あるネオ抗原候補に関しては Cm でも同定することができ、かつ従来法では検出できなかった新規 RNA editing site などによるネオ抗原候補も検出できる可能性が示唆された。腫瘍部の WES を省略することでコストを抑えることもできるため、Cm は従来法に代わる効率的かつ cost-effective なネオ抗原予測手法であると考えられる。

引用文献

Hashimoto S, Noguchi E, Bando H, et al. Neoantigen prediction in human breast cancer using RNA sequencing data. Cancer Sci. 2020.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔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 橋本幸枝、野口恵美子、宮寺浩子、坂東裕子、原尚人
2. 発表標題 Discrepancy of mutation allele frequency between whole exome sequencing and RNA-seq in breast cancer: a case report
3. 学会等名 日本人類遺伝学会 第64回大会
4. 発表年 2019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
研究分担者	野口 恵美子 (Noguchi Emiko) (40344882)	筑波大学・医学医療系・教授 (12102)	

	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
研究協力者	橋本 幸枝 (Hashimoto Sachie)	筑波大学・大学院・大学院生 (12102)	
研究協力者	宮寺 浩子 (Miyadera Hiroko)	筑波大学・医学医療系・助教 (12102)	

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
---------	---------