

## 145. タンパク 3 次構造を考慮した癌遺伝子探索

藤本 明洋

\*理化学研究所 統合生命医科学研究センター ゲノムシーケンス解析研究チーム

Key words : がんゲノム, 突然変異の機能的意義, タンパク質立体構造

### 緒 言

がんは、ゲノムの変異が原因となって起こる疾患であると考えられており、変異の解析はがん研究において最も重要な課題の一つである。近年のシーケンス技術の発展により、全ゲノム、全エクソーム配列決定が可能となり、盛んに研究されている。がんの発生・進展において直接的に重要な役割を果たす遺伝子は、ドライバー遺伝子と呼ばれ、ドライバー遺伝子を検出することにより、発がんのメカニズムの解明や治療標的の発見に結びつく成果が得られると期待されている。

一般に、ドライバー遺伝子は、変異数を統計的に検定することにより検出される。しかしながら、変異数に基づく解析では、低頻度のドライバー遺伝子の検出は困難であり、新たな解析法が必要となる。我々は、タンパク質の立体構造を考慮することにより、1次構造の解析では見逃されている変異の集積を検出できるのではないかと考え、立体構造を考慮する変異の集積を検定する方法（3D permutation 法）を開発した。<sup>1)</sup>

### 方 法

#### 1. データ

変異のデータは、TCGA データベース (<http://cancergenome.nih.gov>) より得た。また3次元構造情報は、PDB データベース (<http://www.rcsb.org/pdb/home/home.do>) より得た。

#### 2. 解析手法

3D permutation 法は、アミノ酸配列上の変異を3次元構造上の位置に変換し、変異の位置の偏りを検出する手法である。アミノ酸を変える変異（ミスセンス変異）3次元構造上の位置に変換するため、以下の解析を行うプログラムを作成した。

1) アラインメントソフトウェア Mafft を使用し、3次元構造情報が得られるアミノ酸配列（pdb ファイル内のアミノ酸配列）を、ヒトの遺伝子配列にアラインメントした<sup>2)</sup>。

2) アラインメントに基づいてゲノム上の位置を3次元構造上の位置に変換し、変異間の3次元構造内のユークリッド距離を総当たりで計算し平均値を求めた（距離の観察値）。

3) 得られた変異間の距離を検定するため、変異数と同じ数のアミノ酸をランダムに選出し、ユークリッド距離の平均値を計算した。この計算を1~100万回繰り返し、距離の平均値の分布を得た。

4) ランダムに選出された分布を元に、距離の観察値の P 値を得た。

以上の計算を、TCGA データベースよりダウンロードされた 21 癌腫のデータについて適用した。P 値の分布は、Q-Q plot で評価した。多重検定補正は、Benjamini と Hochberg の FDR 法で行った。

\*現所属：京都大学 大学院医学研究科 創薬医学講座

## 結果

P 値の分布の Q-Q plot において、一様分布からのずれはほぼ見られなかった。21 がん腫のうち、20 がん腫において、多重検定補正後に統計的有意水準に達する遺伝子が存在した (q-value < 0.1)。有意な 106 遺伝子のうち、31 遺伝子は複数のがん腫で有意であった。

既知のがん遺伝子、がん抑制遺伝子と 3D permutation 法で有意な遺伝子は、共通性が大きく、我々の解析が正しくドライバー遺伝子を同定していることが示唆された (図 1)。

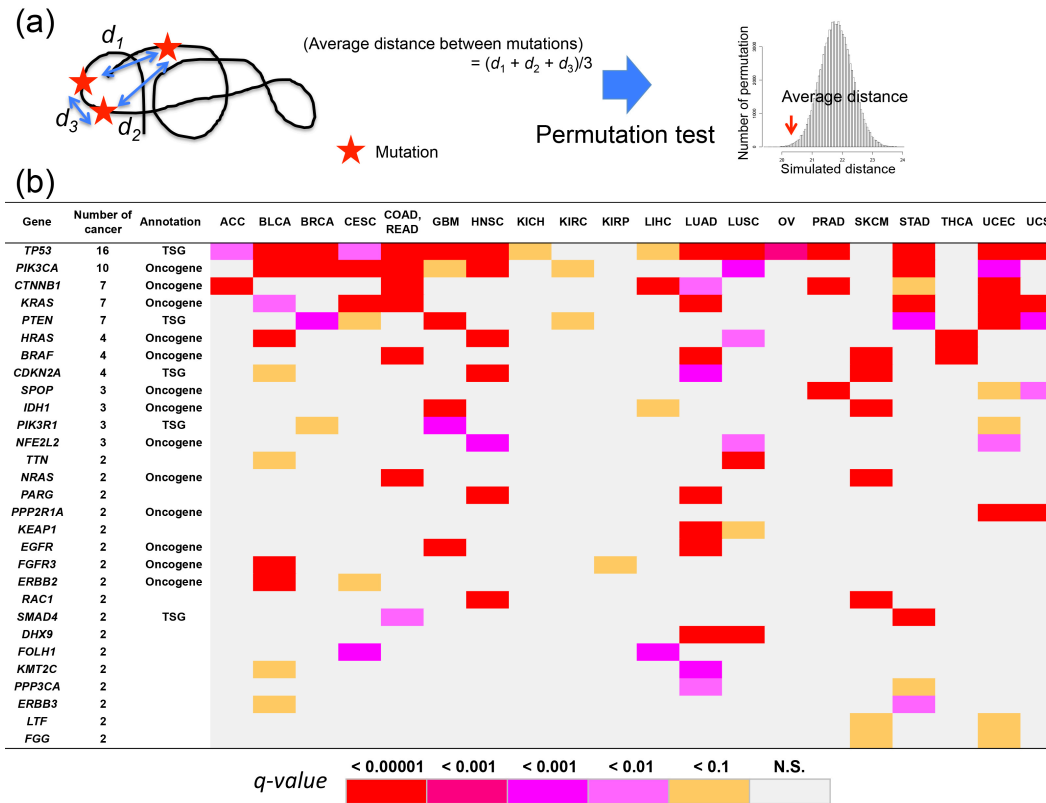


図 1. 3D permutation 法の概要と解析結果

(a) 3D permutation 法の概要。3次元構造を考慮して変異の集積の偏りを検定する。(b) TCGA 計画により公開された変異の解析の結果、複数のがん種で変異の位置が有意に偏っていた遺伝子の一覧。アノテーションは、Vogelstein らの総説に基づいた (Vogelstein *et al.* (2013) Science)。TSG はがん抑制遺伝子、Oncogene はがん遺伝子を示す。がん種名の略称は TCGA 計画に従った。

また、がん遺伝子では、変異間の距離ががん抑制遺伝子よりも小さく、変異が狭い範囲に集積していることが示唆された (図 2)。

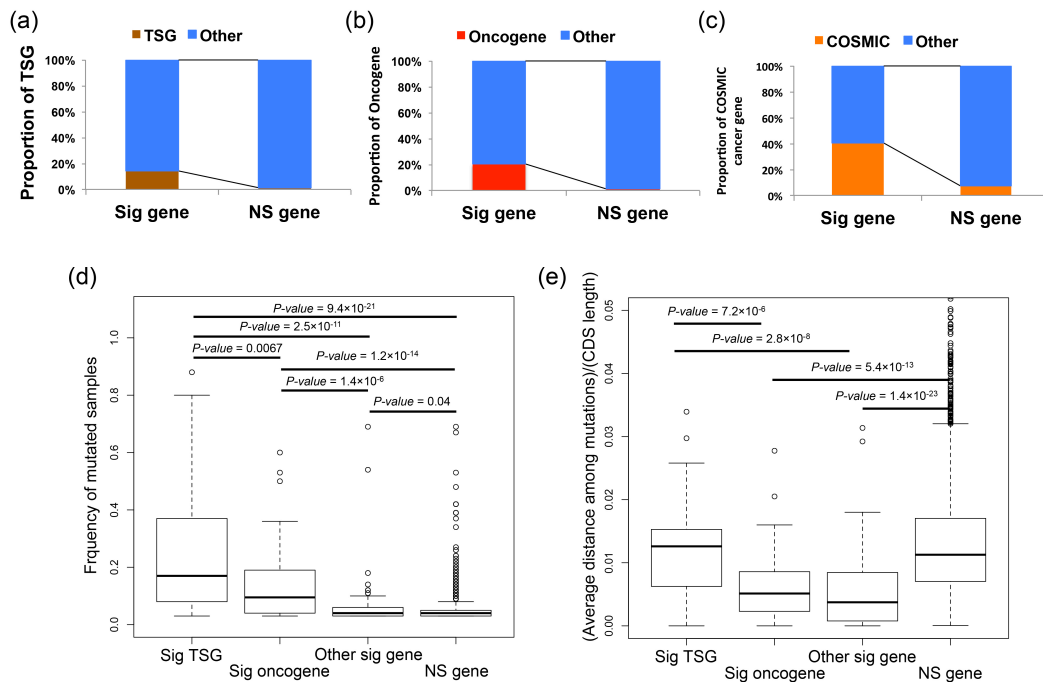


図2. 検出されたドライバー候補遺伝子の特性

(a) TSG (がん抑制遺伝子) の割合。(b) Oncogene (がん遺伝子) の割合。(c) COSMIC データベースのがん関連遺伝子の割合。(d) 変異の頻度。既知のがん抑制遺伝子とがん遺伝子は変異頻度が高かったが、既知の遺伝子以外のドライバー遺伝子候補の頻度は低く、頻度法では検出できない遺伝子を検出している可能性が示唆された。(e) 変異間の距離の分布。がん遺伝子の変異間の距離はがん抑制遺伝子に比べて変異が集積していることが示唆された。

ドライバー遺伝子を検出する他の手法と比較するため、MutSigCV 法や OncodriveCLUST 法との比較を行った。その結果、*FGFR3*, *HRAS*, *KEAPI* 等の遺伝子は 3D permutation 法のみで検出されており、3D permutation 法が既知の手法で見逃されているドライバー遺伝子を検出していることが示唆された。

12 遺伝子 (*TP53*, *PIK3CA*, *CTNBN1*, *KRAS*, *PTEN*, *HRAS*, *BRAF*, *CDKN2A*, *PIK3R1*, *NFE2L2*, *SPOP*, *IDHI*) は、変異の偏りが 3 がん腫以上において検出された。また、それらの中には、既知のホットスポット変異を取り除いても変異の集積が検出された遺伝子もあり、既知のホットスポット以外の変異も発がんにおいて重要な役割を持つ可能性があることが示唆された。

既知のがん関連遺伝子を除いて有意な遺伝子のパスウェイ解析を行ったところ、“axon guidance” や “semaphorin” に関連する遺伝子群が濃縮されていることが示唆された。

## 考 察

我々は、ドライバー遺伝子を検出するため、3次元構造上の変異の分布の偏りを検定する手法を開発した。この手法により、21 がん腫のゲノムデータを解析し、106 遺伝子に有意な変異の位置の偏りを検出した。Q-Q plot は、検定結果の P 値が一様分布に沿っており、我々の手法は偽陽性が少なく、変異の分布の偏りを検出していることが示唆された。また、がん抑制遺伝子においても変異の偏りが検出されており、がん抑制遺伝子においても変異のクラスターが存在することが示唆された。

我々の手法により検出されたドライバー遺伝子候補の約半数は既知のがん関連遺伝子ではなく、新規候補が検出された可能性がある。また、それらの発見に加え、既知のがん関連遺伝子においても興味深い知見が得られた。例えば、*BRAF* 遺伝子においては、変異はホットスポット周辺に集積していた。このことは、重要な既知のホットスポット周

辺部も機能的に重要であることを示唆している。これらの新規候補や集積変異の機能解析を行うことで、がんゲノム変異に新たな知見が得られる可能性は高いと考えられる。現在、*BRAF*上の変異の機能解析を行っているが、さらなる機能解析により新規創薬ターゲットの発見やゲノム医療への貢献したいと考えている。

## 文 献

- 1) Fujimoto A, Okada Y, Boroevich KA, Tsunoda T, Taniguchi H, Nakagawa H. Systematic analysis of mutation distribution in three dimensional protein structures identifies cancer driver genes. *Sci Rep.* 2016 May 26;6:26483. doi: 10.1038/srep26483. PubMed PMID: 27225414; PubMed Central PMCID: PMC4880911.
- 2) Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002 Jul 15;30(14):3059-66. PubMed PMID: 12136088; PubMed Central PMCID: PMC135756