

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 11 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24700272

研究課題名(和文) 癌の病態、発病機序をゲノムレベルで理解するための統計学的手法の開発

研究課題名(英文) Statistical method for understanding cancer genome development and pathogenesis

研究代表者

白石 友一 (Shiraishi, Yuichi)

東京大学・医科学研究所・助教

研究者番号：70516880

交付決定額(研究期間全体)：(直接経費) 3,500,000円、(間接経費) 1,050,000円

研究成果の概要(和文)：癌ゲノムにおける後天的点変異をこれまでよりも高感度に検出するための新たな統計手法，EBCallを開発した。ゲノムのポジションごとにシステムティックエラーが生じやすい個所があるという現象があるが，これまでの手法ではそうした情報を考慮せず，どのポジションにおいても同一の閾値を用いていた。これに対し，提案手法では，複数のコントロール検体からエラーの度合いを推定しつつ，変異コールの際に利用するというアプローチを採用し，経験ベイズ理論に基づいて定式化した。これにより，既存手法に比べて感度や正確性が上昇し，さらに，2%～10%ほどのサブクローンの検出が可能になることを実証することができた。

研究成果の概要(英文)：We developed a novel method, Empirical Bayesian mutation Calling (EBCall) for detecting somatic mutations. Unlike previous methods, the proposed method discriminates somatic mutations from sequencing errors based on an empirical Bayesian framework, where the model parameters are estimated using sequencing data from multiple non-paired normal samples. We demonstrated that our method not only outperforms several existing methods in the calling of mutations with moderate allele frequencies but also enables accurate calling of mutations with low allele frequencies (~10%) harboured within a minor tumour subpopulation, thus allowing for the deciphering of fine substructures within a tumour specimen.

研究分野：情報学基礎

科研費の分科・細目：情報学・統計科学

キーワード：ベイズ統計学 次世代シーケンス解析

## 1. 研究開始当初の背景

高速シーケンス技術の発展により、癌のゲノムを網羅的に調べ、点変異をゲノムワイドに同定することが可能になった。シーケンサーのデータのポテンシャルは非常に高く、点変異だけではなく、コピー数異常、数百～数千塩基の比較的大きな増幅や欠失、転座も同時に調べることができる。また、ゲノム領域の1%に満たないが、タンパク質をコードする重要な領域であるエキソン領域のみをシーケンスする技術も発達し、これにより安価で多検体の癌ゲノムを調べることが可能になった。

## 2. 研究の目的

### (1) シークエンスデータ解析のためのプラットフォーム開発

近年の大量シーケンス技術の発展により、癌に生じているゲノム変異を網羅的に検出することが原理的に可能になった。しかし、データが膨大となるために、大量のシーケンスデータ解析のためにはスーパーコンピュータが必要となることや、シーケンスエラーやアラインメントエラーを回避しつつ、正確かつ高感度に変異を検出するための統計的方法論の整備が進んでいないことなどの問題点があった。

### (2) 癌の進化のメカニズムを探るためのサブクローン変異検出

我々の体内の細胞は、化学物質やウイルスによる炎症などの刺激によりランダムに生じた変異を蓄積している、ゲノムとしてヘテロな構造を有する細胞集団である。特に、「癌は原癌遺伝子の変異により優位性を獲得した細胞がクローン増幅を起こす」ことが繰り返されることにより生じるが、その過程で複数の競争的に進化するクローン集団に分岐する。主要なクローン集団に隠れて生じているサブクローンと、癌の薬剤耐性獲得や、再発、転移の関係性を明らかにすることは癌研究における長年の課題であった。

現代の超並列大量シーケンサーの技術革新は、様々な種類のゲノム変異を網羅的に検出することに加えて、癌ゲノムにおけるサブクローン構造の推定を可能にした。ここでキーとなっている考え方はそれぞれのシーケンスリードは、一つ一つ別々の細胞からランダムに得られたものであり、ある変異を有しているリードの割合（以下、アレル比率と呼ぶ）が当該変異の細胞集団における割合を反映しているということである。検出された変

異をアレル比率による分類、すなわちクラスタリングを行うことで、サブクローン集団の有無や個数を調べることが可能になり、例えば、ある種の乳がんにおいては、「癌細胞中のクローンの個数と予後の悪さに相関がある」ということなど、癌ゲノムの heterogeneity 構造と癌の進化メカニズム、分類等に関して多くの新規知見が得られようとしている。

一方で、詳細にサブクローン構造を調べるためには、アレル比率が低い変異をできるだけ高感度で検出することが必要である。しかし、変異のアレル比率が低くなるにつれてシーケンスエラーとの区別が難しくなるために、既存の変異検出法ではアレル比率が10%以下の変異の検出は難しいと考えられていた。

## 3. 研究の方法

東京大学医学部附属病院（現京都大学医学系研究科）小川誠司研究室に常駐し、小川研究室の研究者と毎日毎晩のディスカッションを重ね、実用的な癌ゲノムシーケンス解析のために必要なものを現場にて身近に感じつつ、研究を遂行した。

## 4. 研究成果

### (1) シークエンスデータ解析のためのプラットフォーム開発

多数検体の円滑なデータ解析を可能にするために、東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピュータ上にて、癌エキソンシーケンスデータ解析パイプラインの構築を行った。解析パイプラインは、高速アラインメントツールを用いたショートリードのアラインメント、その後の局所アラインメント、フィッシャーの正確検定を基にした手法による効率的な変異候補の同定、既存のデータベースの情報や多型情報などのアノテーション付け等の一連の流れを、並列化により高速に行うことが可能である。さらに、開発したプラットフォームを、「Genomon-exome」として平成24年4月に公開した。

また、RNA シークエンスデータから融合遺伝子を検出する新たなプラットフォーム「Genomon-fusion」を開発し（右下図参照）、平成24年9月に公開した。提案手法は、トランスクリプト配列とゲノム配列の2段階のアラインメントや、soft clipping リードの探索による1塩基単位でのbreak point 検出、偽陽性を除くための様々なフィルターを備えており、既存の手法に比べ感度、正答率の

点で勝っていることが実験的に確認されている。

## (2) 癌の進化のメカニズムを探るためのサブクローン変異検出

癌ゲノムにおける後天的点変異をこれまでよりも高感度に検出するための新たな統計手法、EBCallを開発した。ゲノムのポジションごとにシステマティックエラーが生じやすい箇所があるという現象があるが、これまでの手法ではそうした情報を考慮せず、どのポジションにおいても同一の閾値を用いていた。これに対し、提案手法では、複数のコントロール検体からエラーの度合いを推定しつつ、変異コールの際に利用するというアプローチを採用し、経験ベイズ理論に基づいて定式化した。これにより、既存手法に比べて感度や正確性が上昇し、さらに、2%~10%ほどのサブクローンの検出が可能になることを実証することができた。

### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

{ 雑誌論文 } (計 7 件)

[1] Yoshida K\*, Sanada M\*, Shiraishi Y\*, Nowak D\*, Nagata Y\*, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S and Ogawa S, “Frequent pathway mutations of splicing machinery in myelodysplasia”, *Nature*, 478, pp.64-69, 2011 (\*第一著者と同等の寄与)

[2] Kunishima S, Okuno Y, Yoshida K, Shiraishi Y, Sanada M, Muramatsu H, Chiba K, Tanaka H, Miyazaki K, Sakai M, Ohtake M, Kobayashi R, Iguchi A, Niimi G, Otsu M, Takahashi Y, Miyano S, Saito H, Kojima S and Ogawa S, “ACTN1 mutations cause congenital macrothrombocytopenia”, *American Journal of Human Genetics*, 92(3), pp. 431-438, 2013.

[3] Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, Shiba N, Hayashi Y, Kume H, Homma Y, Sanada M, Ogawa S and Miyano S, “An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data”, *Nucleic Acids Research*, 41(7):e89, 2013.

[4] Sato Y\*, Yoshizato T\*, Shiraishi Y\*, Maekawa S\*, Okuno Y\*, Kamura T,

Shimamura T, Sato-Otsubo A, Nagae G, Suzuki H, Nagata Y, Yoshida K, Kon A, Suzuki Y, Chiba K, Tanaka H, Niida A, Fujimoto A, Tsunoda T, Morikawa T, Maeda D, Kume H, Sugano S, Fukayama M, Aburatani H, Sanada M, Miyano S, Homma Y and Ogawa S, “Integrated molecular analysis of clear-cell renal cell carcinoma”, *Nature Genetics*. 45(8), pp.860-7, 2013 (\*第一著者と同等の寄与).

[5] Sakaguchi H, Okuno Y, Muramatsu H, Yoshida K, Shiraishi Y, Takahashi M, Kon A, Sanada M, Chiba K, Tanaka H, Makishima H, Wang X, Xu Y, Doisaki S, Hama A, Nakanishi K, Takahashi Y, Yoshida N, Maciejewski JP, Miyano S, Ogawa S and Kojima S, “Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia”, *Nature Genetics*, 45(8), pp.937-41, 2013.

[6] Kon A, Shih LY, Minamino M, Sanada M, Shiraishi Y, Nagata Y, Yoshida K, Okuno Y, Bando M, Nakato R, Ishikawa S, Sato-Otsubo A, Nagae G, Nishimoto A, Haferlach C, Nowak D, Sato Y, Alpermann T, Nagasaki M, Shimamura T, Tanaka H, Chiba K, Yamamoto R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Nakamaki T, Ishiyama K, Nolte F, Hofmann WK, Miyawaki S, Chiba S, Mori H, Nakauchi H, Koeffler HP, Aburatani H, Haferlach T, Shirahige K, Miyano S and Ogawa S, “Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms”, *Nature Genetics*, 45(10), pp.1232-7, 2013.

[7] Yoshida K, Toki T, Okuno Y, Kanezaki R, Shiraishi Y, Sato-Otsubo A, Sanada M, Park MJ, Terui K, Suzuki H, Kon A, Nagata Y, Sato Y, Wang R, Shiba N, Chiba K, Tanaka H, Hama A, Muramatsu H, Hasegawa D, Nakamura K, Kanegane H, Tsukamoto K, Adachi S, Kawakami K, Kato K, Nishimura R, Izraeli S, Hayashi Y, Miyano S, Kojima S, Ito E, Ogawa S.

“The landscape of somatic mutations in Down syndrome-related myeloid disorders”, *Nature Genetics*, 45(11), pp.1293-9, 2013.

{ 学会発表 } (計 件)

{ 図書 } (計 件)

{ 産業財産権 }

出願状況 (計 件)

名称 :

発明者 :

権利者 :

種類：  
番号：  
出願年月日：  
国内外の別：

取得状況（計 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

Genomon-exome  
<http://genomon.hgc.jp/exome/>

Genomon-Fusion  
<http://genomon.hgc.jp/rna/>

## 6. 研究組織

### (1) 研究代表者

白石友一 (Shiraishi Yuichi)  
東京大学・医科学研究所・助教  
研究者番号：70516880

### (2) 研究分担者

( )

研究者番号：

### (3) 連携研究者

( )

研究者番号：