

# 遺伝子発現量データを利用した表現型の違いに影響を及ぼす 遺伝子セットの検出方法の開発

土多隆雄†1 大林武†1 木下賢吾†1†2†3

マイクロアレイ解析では遺伝子単独の発現量変化に対して統計学的仮説検定を行うことで、表現型の違いに影響を及ぼす遺伝子を検出することが多い。しかし、実際には遺伝子は複数で相互に関わりあいながら機能することから、遺伝子単独ではなく複数の遺伝子による発現パターン変化にも着目する必要があるといえる。そこで、本研究では複数の遺伝子による発現パターン変化を検出する手法を新たに開発し、肺がんに関する遺伝子発現量データを用いて先行研究で提案されている手法との比較解析を行った。その結果、先行研究で提案されている手法では検出されない発現パターン変化を示す遺伝子ペアを検出することができた。

## 1. はじめに

マイクロアレイ解析では、表現型によって異なる発現をする遺伝子を検出するために各々の発現量変化に対して統計学的仮説検定を行うことが多い。しかし、実際遺伝子は複数で機能することから、表現型に影響を与える遺伝子の検出には遺伝子単独ではなく複数の遺伝子による発現パターン変化に着目する必要がある(図1)。

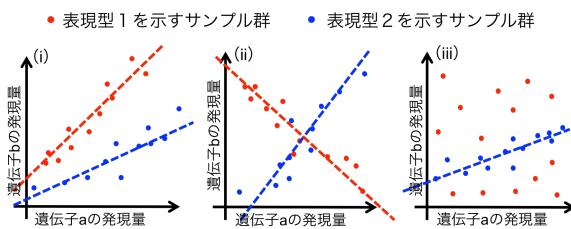


図1: 遺伝子ペアによる発現パターン変化

図1のいずれの発現パターンも遺伝子aもしくはbのどちらか一方にのみ着目すると表現型1と表現型2で発現量の分布に有意な差はないが、遺伝子ペアに着目するとその発現パターンが表現型によって明らかに異なる。

先行研究では図1-(i)(ii)のような発現パターン変化を示す遺伝子ペアの検出を可能にしている[1]。しかし、その手法では図1-(i)(ii)のような2つの表現型の両方に強い相関がみられる遺伝子ペアしか検出されず、図1-(iii)のような片方の表現型のみみられる相関関係がもう一方の表現型では失われる発現パターン変化を示す遺伝子ペアは検出されない。また、先行研究の手法では相関係数を用いているため、その使用は遺伝子ペアに限られ3遺伝子以上の発現パターン変化を評価できない。

そこで、本研究では先行研究で想定されていない発現パターン変化を示す遺伝子ペアの検出ならびに3遺伝子以上で発現パターン変化の評価を行う。

## 2. 解析手法と使用したデータ

### 2.1 解析手法

先行研究では、相関係数によって図1-(i)(ii)のような発現パターン変化を示す遺伝子ペアをそれぞれ評価式  $S_1 = |\rho_1 + \rho_2 - \alpha\rho|$ ,  $S_2 = |\rho_1 - \rho_2|$  ( $\rho$ :全サンプルの相関係数、 $\rho_1$ :表現型1の相関係数、 $\rho_2$ :表現型2の相関係数、 $\alpha=1.5$ ) によって検出する。この評価式の値が高い遺伝子ペアほど発現パターン変化を示す。そして、本研究で図1-(iii)を検出する評価式  $S_3 = ||\rho_1| - |\rho_2||$  を新たに定義する。 $S_3$ も先行研究と同様に相関係数を用いる。

また、これらの評価式は3遺伝子以上の発現パターンを評価することができないため、3遺伝子以上にも適用できる評価式  $S_4$  を次のように定義する。

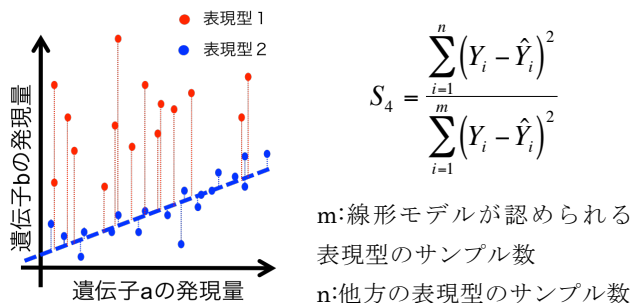


図2: 評価式  $S_4$  による発現パターン変化の検出

評価式  $S_4$  は図2のように比較する2つの表現型のどちらかに認められる線形回帰モデル ( $Y=aX+b$ ) に対し残差平方和を求め、比をとったものである。図2の場合、表現型2に回帰モデルが認められるので、そのモデルに対する表現型1の残差平方和を表現型2の残差平方和で除算する。回帰モデルの認定は決定係数から求められるF値によって行い、有意水準  $\alpha=0.05$  で仮説「回帰係数 a,b がともにゼロである」が棄却された場合のみ上記線形モデルを認める。どちらの表現型にも回帰モデルが認められる場合、よりF値が小さい表現型を採用し、どちらにも回帰モデルが認められない場合は  $S_4$  の値を0とする。

### 2.2 使用したデータ

解析には日本で癌死亡者数が最も多い肺がんに関するマ

†1 東北大学大学院情報科学研究科  
†2 東北大学加齢医学研究所  
†3 東北大学東北メディカル・バンク機構

マイクロアレイデータ (GDS3257:Cigarette Smoking effect on lung adenocarcinoma) を使用する[2]. この発現量データは 58 のがん細胞サンプル、49 の正常細胞サンプルとなり、それをそれぞれ表現型 1、表現型 2 とする. また、解析の対象となる遺伝子は、ベンゾピレンの代謝に関わる 45 遺伝子とヒトの転写因子 1,558 遺伝子のうち、プローブに対応付けられる 1,011 遺伝子である. ベンゾピレンは肺がんの形成に関わるとされるタバコの煙に含まれる. 1,011 遺伝子に対応するプローブは 1,680 プローブあり、1,409,180 プローブペアについて解析する. (同じ遺伝子のプローブペアは解析から除く.)

### 3. 結果

#### 3.1 評価式 S<sub>1</sub>,S<sub>2</sub>,S<sub>3</sub>,S<sub>4</sub> による遺伝子ペアの解析結果

各評価式 S<sub>1</sub>,S<sub>2</sub>,S<sub>3</sub>,S<sub>4</sub> で検出された代表的な遺伝子ペアをそれぞれ図 3-(i)(ii)(iii)(iv)に示す.

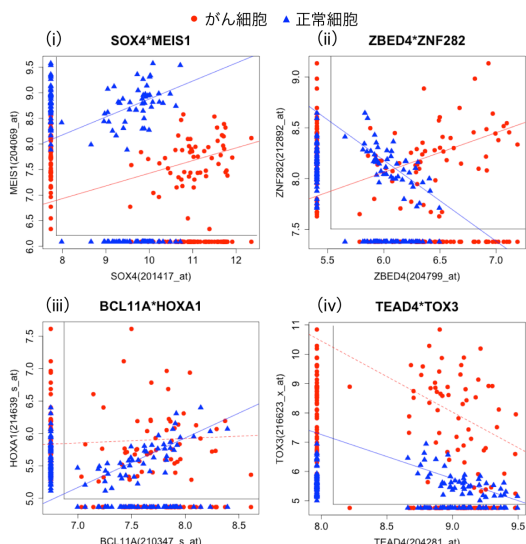


図 3 : 各評価式によって検出された遺伝子ペア

図 3 より、それぞれの評価式で想定している発現パターンを示す遺伝子ペアを検出していることがわかる. 特に今回新たに定義した評価式 S<sub>3</sub> で得られた図 3-(iii)の遺伝子ペアは正常細胞では相関関係がみられるが、がん細胞ではその関係が失われていることがわかる. 評価式 S<sub>4</sub> で得られた図 4-(iv)の遺伝子ペアは正常細胞では相関関係がみられるが、がん細胞では相関関係が弱くなっており、また、発現量の和が大きくなっている.

さらに、評価式 S<sub>3</sub> についてスコアが高い上位 100 の遺伝子ペアの関係図を図 4 に示す. 図 4 にて濃い水色で示されている遺伝子は単独の発現量変化についての Wilcoxon 検定で検出される遺伝子であり、薄い水色で示されている遺伝子は評価式 S<sub>3</sub> によって検出された遺伝子を示す. 図 4 より、HOXA1 は単独の発現量変化についての検定では検出されないが、多くの遺伝子に関わり発現パターンを変化させていることがわかる. この遺伝子は胎児や幼生の発生

に関わる遺伝子だが、肺がん細胞ではその発現量が減少していることや、細気管支肺がん細胞では過剰にメチル化されていることが報告されている. さらに、図中の ASCL1 は肺の上皮細胞や神経内分泌細胞の分化に関わる遺伝子であり、肺がん細胞ではその発現が変化し、肺がんの種類によってもその発現パターンが異なることが知られている. この遺伝子は評価式 S<sub>4</sub> でも検出されている.

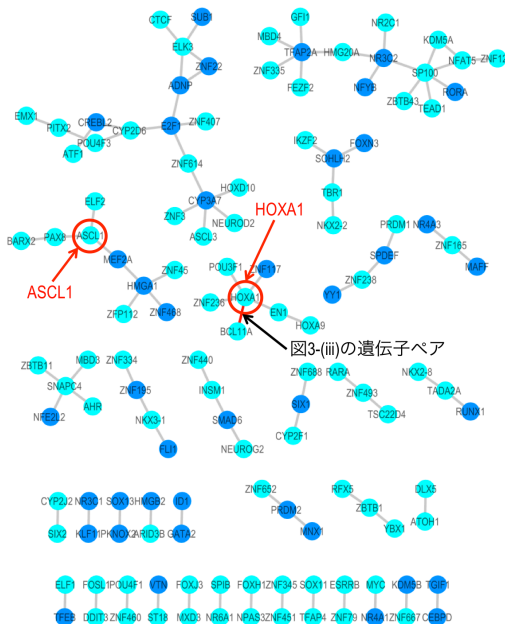


図 4 : 評価式 S<sub>3</sub> で検出された遺伝子ペアの関係図

### 4. 結論

各評価式による解析の結果、それぞれ想定する発現パターン変化を示す遺伝子ペアが検出された. S<sub>1</sub> ではスコアが高い上位 100 の遺伝子ペアに含まれる遺伝子は全て Wilcoxon 検定で検出できる遺伝子であったが、S<sub>3</sub>,S<sub>4</sub> では Wilcoxon 検定で検出されない遺伝子を多く検出していた. また、評価式 S<sub>4</sub> は原理的には S<sub>1</sub>,S<sub>2</sub>,S<sub>3</sub> のそれぞれで想定する発現パターン変化を示す遺伝子ペアを検出できるが、今回解析した発現量データでは S<sub>1</sub>,S<sub>3</sub> で得られる遺伝子ペアをよく検出していた. さらに、各評価式について得られた遺伝子ペアの関係図から、多くの遺伝子と関わり発現パターンを変化させる遺伝子が見つかった. それらの遺伝子は既に肺がんとの関連性を示唆されているものが多かったが、ほとんど先行研究がなされていない遺伝子もあった.

最後に、3 遺伝子発現パターンの解析については、今後遺伝子ペアでは検出されないものの、この解析によって検出された遺伝子があるか検討する必要がある.

#### 参考文献

- 1) M.Dettling, et al., Searching for differentially expressed gene combinations, Genome Biology, Vol.6, No.10, pp.R88 (2005)
- 2) M.Landi, et al., Gene Expression Signature of Cigarette Smoking and Its Role in Lung Adenocarcinoma Development and Survival, PLOS ONE, Vol.3, No.2, pp.e1651 (2008)