

組織特異的なメチル化と遺伝子型の関係

Relationship between tissue specific methylation and genotype

物理学専攻 木下 遼一

Department of Physics, Ryoichi Kinoshita

0. 概要

遺伝子型の発現にメチル化がどのように影響するかは不明です。私の研究では、30人の患者からの三種類の組織、すなわち、腫瘍、腫瘍に隣接する正常組織、血液のDNAの遺伝子型とメチル化の両方に主成分分析を適用しています。メチル化は遺伝子型とメチル化の両方について血液、腫瘍に隣接する正常組織、腫瘍の順序で頻度が増加していくことが一般的であることが判明しました。大幅に頻度が増加している遺伝子型とメチル化遺伝子の間にはオーバーラップが見られ、少なくとも、一部の部位は、メチル化の増大は遺伝子の多型頻度の増大と相関関係があることがわかりました。

1. 研究背景・目的

近年、実験技術の進展に伴い、メチル化、ヒストン修飾の様ないわゆるエピジェネティックマーカー (Epigenetic Marker) の観測が盛んに行われている。いまだ、エピジェネティックマーカーが遺伝子発現にいかなる影響をどのように与えているかは完全には解っていないが、なんらかの影響をあたえうる、という点については大方の意見として合意がとれていると思われる。実際、遺伝子発現量ではなく、エピジェネティックマーカーを用いた方がカテゴリ間差異をより判別しやすい、という報告も出てきている。例えば、Yangら (参考文献[1]) はDNAのメチル化パターンを30人の被験者に対して種々の臓器ごとに観測し、メチル化の方が、ゲノム配列より臓器依存性が高いということを報告している。具体的には、「主成分分析を用いた場合、メチル化では最初の2つの主成分が張る空間で臓器ごとにクラスター化がはっきりと観測できるが同じことを遺伝子発現を行うのは困難である」と報告している。

本研究発表でも、Yangらと同様のサンプルを用いて解析を行った。なお、彼らの主成分分析の使い方は本研究発表での主成分分析の使い方とは根本的に異なっている。

・エピジェネティック修飾による遺伝情報の発現制御

生物の誕生や生命現象の営みは、DNA上の遺伝子に書き込まれた遺伝情報に基づいて行われる。ヒトの健康の維持・増進や変調にも遺伝情報の働きが関与する。「エピジェネティクス」は、「塩基配列に依存しない遺伝子機能の調節機構」である。具体的には、遺伝子の働きのON/OFFの調節が、DNAの塩基配列自身ではなく、主としてDNA塩基へのメチル化修飾、DNAが巻きついているヒストンタンパクへのメチル化、アセチル化修飾などによって行われるという仕組みである。

・DNAメチル化

動物のゲノムDNAでは3'-CG-5'の並び(CpG配列)のシトシンがメチル化修飾を受ける。CpG配列は、多くのゲノムDNA領域では散在的に存在するが、例外的に遺伝子の転写開始点上流には高頻度に存在する領域があり、CpGアイランドと呼ばれている。CpGアイランドのDNAメチル化は転写活性と相関して

おり、転写活性の高い遺伝子では低メチル化状態に、逆に転写が抑制されている遺伝子では高メチル化状態になっている。

2. 材料及び方法

・ 研究に用いたデータ

NCBIのGEO^{*1}より:[アクセッション番号:GSE20123]

30人の食道扁平上皮癌(ESCC)の患者の血液(B)、腫瘍に隣接する正常組織(N)、腫瘍(T)の3種類の細胞の遺伝子多型頻度とメチル化のパターン。Nspチップ(SNP総数262,264個ずつ)とStyチップ(SNP総数238,304個ずつ)の異なる2種類のSNPアレイで分析。

これらのデータは中華人民共和国山西省太原市の山西がん病院で、1998年から2001年の間にESCCと診断され、根治的な外科的切除の候補と考えられた患者である。静脈血(10mL)を、手術前に各患者から採取し、全血中から子孫に受け継がれるDNAは抽出し、標準的なフェノール/クロロホルム法を用いて精製した。ESCCの腫瘍と隣接正常組織は手術の時に解剖し、使用するまで液体窒素中で保存した。

Affymetrixマッピング500Kアレイセット(NspアレイとStyアレイ)を使用して、タイピングとメチル化実験を行った。2つのアレイの違いは、ゲノムDNAが消化されるのに用いた制限酵素がNsp IまたはSty Iの違いである。

*1 <http://www.ncbi.nlm.nih.gov/geo/>

・ 主成分分析

相関のある何種類かの変数の情報を、互いに無相関な少数個の総合特性値に要約する方法である。

$$\text{データ行列 } X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

Xの行ベクトルに対応する確率変数ベクトル $x = (x_1, x_2, \dots, x_p)^T$

m個の総合特性値 z_1, z_2, \dots, z_m を、p次元ベクトル x の線型結合として考える。

$$\begin{cases} z_1 = l_1^T x = l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ \vdots \\ z_m = l_m^T x = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \end{cases} ; \quad l_k = (l_{k1}, l_{k2}, \dots, l_{kp})^T \text{ は未知定数ベクトル}$$

- ・ 総合特性値 z_1, z_2, \dots, z_m : 第1主成分、第2主成分、...、第m主成分
- ・ 係数 $l_k = (l_{k1}, l_{k2}, \dots, l_{kp})^T$: 第k主成分に対する主成分負荷量 $k = (1, 2, \dots, m)$

3. 結果

Nspチップの場合、遺伝子多型頻度とメチル化パターンの両方にPCAを適用した結果、通常の間接(N)を介して血液(B)から腫瘍(T)へと多型頻度/メチル化の増大に第二主成分が対応しているのを発見した。第一主成分は、組織特異性を欠いていたので無視しました。その後、遺伝子多型頻度およびメチル化の第二主成分の結果上位300個のプローブを拾い上げている。

各プローブはのサンプル、Nのサンプル、Tのサンプルそれぞれ30個ずつで構成される90個のサンプルの遺伝子型頻度とメチル化パターンを持っているので、3種類のサンプルセットのペア、

すなわち B 対 T、B 対 N、および T 対 N の t 検定を適用した。300 個のプローブの中から遺伝子発現（メチル化）パターンで $P < 0.05/N$ を満足する少なくとも 1 つの P 値がある。ここでは $N = 262339$ は配列中のプローブの数である。このように PCA がその遺伝子多型頻度とメチル化パターンは組織間で大幅に異なっているプローブを拾った。

次に、これらのセットを比較した。トップ 300 個の遺伝子多型頻度のセットは、 $P < 1 \times 10^{-22}$ でメチル化と共通のプローブを持つと判明した。これは、遺伝子多型頻度が直接メチル化に関連していることを意味する。

50	SNP_A-1825620	SNP_A-2309865	SNP_A-4233167		
100	SNP_A-1984943	SNP_A-2172952	SNP_A-2234716		
150	SNP_A-1988914	SNP_A-2040111	SNP_A-2089983	SNP_A-4195285	SNP_A-4199352
200	SNP_A-2276203	SNP_A-4196078	SNP_A-4226834	SNP_A-4229534	
250	SNP_A-1980533	SNP_A-1886593	SNP_A-2042678	SNP_A-2143521	SNP_A-1880907
	SNP_A-1989613	SNP_A-2142865	SNP_A-4193660	SNP_A-1845324	SNP_A-1852621
300	SNP_A-2053247	SNP_A-1911642	SNP_A-2221049	SNP_A-4213049	SNP_A-4228665
	SNP_A-4236336	SNP_A-2287632	SNP_A-2043441		

また、直接遺伝子多型頻度とメチル化を比較しました。B と N と T 間の相関係数が選択されたプローブで有意に大きいことがわかった。

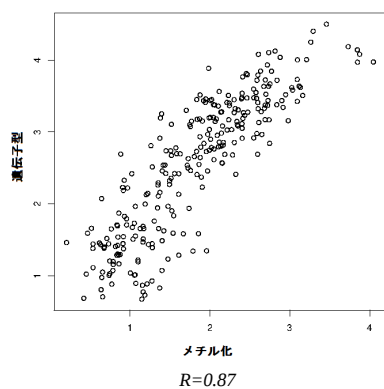


図1:主成分上位50の有意差が一致したSNPの相関図

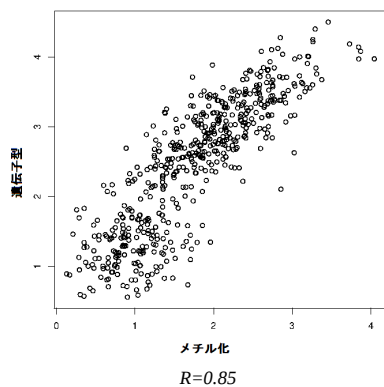


図2:主成分上位100の有意差が一致したSNPの相関図

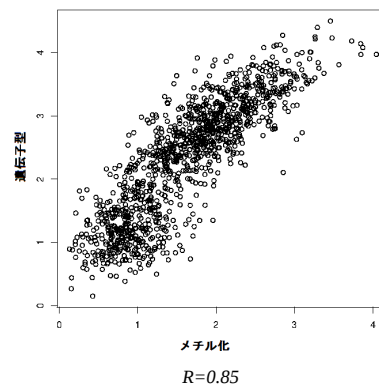


図3:主成分上位150で有意差が一致したSNPの相関図

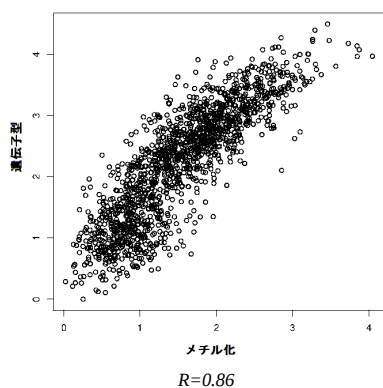


図4:主成分上位200で有意差が一致したSNPの相関図

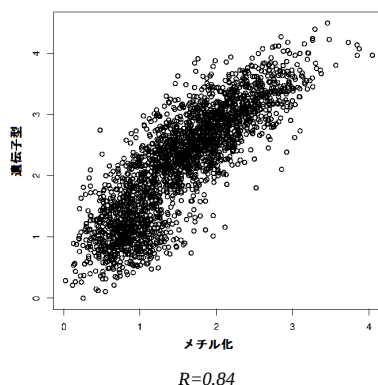


図5:主成分上位250で有意差が一致したSNPの相関図

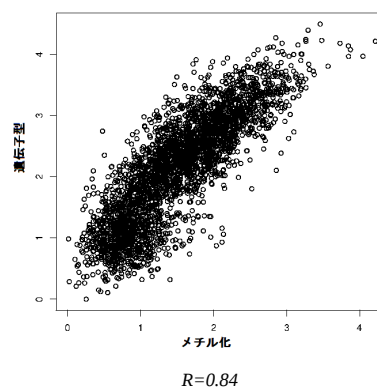


図6:主成分上位300で有意差が一致したSNPの相関図

相関係数の 0.84 ~ 0.86 程の値をとり、高い相関関係にあるといえる。

Nsp で絞った SNP35 個に関連する遺伝子がどのような働きをするのか調べた。調べ方として絞った SNP の関連遺伝子の RefSeq(Reference Sequence) を NCBI や Wikipedia*² を活用し、遺伝子名(遺伝子シンボル)を調べる。その後、NCBI や Gendoo*³ で遺伝子名を調べ、その遺伝子がどのような性質をもつのか、どのような症状をもたらすのかを調べた。結果、22 個の遺伝子の性質がわかった。その中で、腫瘍に関わる遺伝子は 14 個あった。

*² <http://www.wikipedia.org/>

*³ <http://gendoo.dbcls.jp/>

SNP ID	SNP_A-1825620	SNP_A-4233167	SNP_A-1984943	SNP_A-2234716	SNP_A-4199352
遺伝子名	TOMM7	RICTOR	TGFB2	F2RL3	SAMD12
SNP ID	SNP_A-2089983	SNP_A-4196078	SNP_A-1886593	SNP_A-2042678	SNP_A-2143521
遺伝子名	ZMIZ1	C7orf62	SEMA3E	ZBED2	RPL14
SNP ID	SNP_A-1880907	SNP_A-1845324	SNP_A-1911642	SNP_A-1845324	
遺伝子名	NXPH1	KLHL6	RND3	PLA2G4A	

4. 結論

主成分分析を行い、特異的だと思われる SNP の中で、遺伝子多型頻度とメチル化パターンで組織間の有意差の関係が共通する SNP には相関関係が見られた。また、この SNP に関連する遺伝子には腫瘍に関わる遺伝子が多く見つかった。

これらのことを踏まえ、メチル化された遺伝子は発癌に関わる可能性がある。そのため、メチル化量が高い組織は遺伝子発現量が高くなっているのだと考える。

参考文献

[1] Yang HH, Hu N, Wang C, Ding T, Dunn BK, Goldstein AM, Taylor PR, Lee MP.: Influence of genetic background and tissue types on global DNA methylation patterns, PLoS One, vol.5, No. 2, e9355. doi: 10.1371/journal.pone.0009355. (2010)

[2] Ryoichi Kinoshita and Y-h. Taguchi : Tissue specific methylation and genotype, IPSJ SIG Technical Report, Vol.2011-BIO-25 No.39, pp.1-2 (2011)

[3] Ryoichi Kinoshita and Y-h. Taguchi : Tissue specific methylation and genotype, Bioinformatics and Biomedicine (BIBM) vol.2, pp.953-955 (2011)