

マイクロアレイ解析における遺伝子の絞り込み方法の検討

新井誠人¹、関 直彦¹、森 幹人²、島田英昭²、落合武徳²

¹ 千葉大学大学院医学研究院 機能ゲノム学

² 千葉大学大学院医学研究院 先端応用外科

要 旨

マイクロアレイを用いた遺伝子発現解析は、我々も含め多くの研究者が行ってきた。特に、癌組織をサンプルとして、その予後因子や転移因子などを検索する研究はこれまでも数多くなされてきた。しかし、その鍵となる遺伝子の選出し方については、確立された方法がなく、各研究機関によってばらばらである。本研究では、先に報告された胃癌組織をサンプルとしたマイクロアレイ解析データを基に、転移に関与する遺伝子の絞り込みを行った。様々なステージの胃癌組織 16 例を約 4600 遺伝子搭載の in house cDNA マイクロアレイ解析を行い、このデータを解析対象とした。(1) マイクロアレイ上の blank スポットの intensity を計測し、その中心値を含む 80% のデータを使用し、平均値、標準偏差を計算し、各スポットの原 intensity より、blank スポットの平均値と 2 倍の標準偏差値を引き、正であるものを解析対象スポットとした。(2) リンパ節転移がある群とない群で臨床サンプルをグループ分けし、並び替え検定と交差検証解析を行った。(3) (2) の解析で、仮説検定のための有意水準の数値 (p 値) が得られ、p 値が 0.01 以下になった遺伝子に対して、臨床サンプルを一例ずつ抜いて同じ解析を行い、全サンプルを入れ替えても、p 値が 0.01 以下であり、さらに 2 群の平均の差が 2 倍以上である遺伝子を、グループ間の判別に使用可能な遺伝子とした。リンパ節転移のある、なし群の判別に有意な遺伝子として、(2) の解析で 228 遺伝子が選出された。さらに (3) の解析を加え、99 遺伝子を選出した。さらに、発現レベルが一定以上必要と考え、最終的に 29 遺伝子が、リンパ節転移の有無の判別の候補遺伝子群として選出された。臨床検体をサンプルとしたマイクロアレイ解析を基に、統計学的手法を用いて、リンパ節転移判別遺伝子群を同定した。これら遺伝子群は、腫瘍の大きさや他臓器への転移などの因子との関連は薄く、独立した予測因子である可能性があり、今後の胃癌の治療方針決定に役立つ可能性がある。

研究目的

1. マイクロアレイ解析の難しさ

マイクロアレイによる臨床検体を対象とした解析はこれまで多くの報告がある。特に悪性腫瘍のマイクロアレイ解析は、様々な臓器を対象として、予後予測などに使用可能かどうか検討されている。一部には、実際の臨床現場に、これまでにない新しい情報を提供する試みが始まっている。しかしながら、それぞれの研究機関で、別々の臓器保存方法、RNA 抽出方法、異なった種類のマイクロアレイ、実験手技の差、病期分類の違いなど、その解析方法は全く異なったものといってもよい印象すらある。さらに、マイクロアレイから得られた情報をどのように処理するかについても、議論の余地がある。いわゆるバイオインフォマティクスの進歩により、多くのマイクロアレイ解析方法が示されているが、ゴールドスタンダードとなるものはまだ無いのが現状である。そもそも

マイクロアレイ解析自体が、やや不安定な結果をもたらす可能性が高く、これが混乱に拍車をかけている (1-5)。

2. 胃癌症例のリンパ節転移の問題点

胃癌は、検診等の普及によって減少傾向にはあるが、日本人の死亡原因の依然として上位である。根本的な胃癌治療には、手術による病変の完全切除が最も効果的であるが、リンパ節や他臓器への転移を認めた場合の予後は不良である。リンパ節への転移が認められるか否かによって、治療方針の変更を余儀なくされる。また、術後の経過観察中に転移を認めることもある。したがって、手術検体もしくは術前内視鏡検査時の検体からの情報でリンパ節転移の将来にわたっての有無が予測できれば、患者も医師も最善の治療を選択する重要な情報となることは間違いない。現在、病理所見によって、ある程度情報は得られるが、十分なものとはいえない。

方法

1. 対象とするマイクロアレイ解析

肝臓、胃由来の in house cDNA microarray 千葉大学付属病院にて手術を受けた 20 症例を対象サンプルとし、約 4600 個の遺伝子を搭載した肝臓、胃由来の in house cDNA マイクロアレイを使用し、duplicate で解析した。そのうち、リンパ節転移に関与する可能性がある遺伝子として S100A11 という遺伝子に注目し、既に報告した (6)。そのうち 6 症例はバックグラウンドの intensity が高いため、解析不適とし、16 例 (マイクロアレイ解析は 28 回) を再解析対象とした。

2. バックグラウンド値の処理

マイクロアレイ上の Blank スポット全ての intensity データより、平均値、標準偏差値を測定した。各スポットの intensity より、バックグラウンド値の平均値と標準偏差の 2 倍を引き算し、正であるスポットを解析可能とした。全サンプルで、intensity が正であるスポットを、解析対象とした。

3. 並び替え検定と交差検証

はじめに、それぞれの遺伝子についてリンパ節転移ある、なし群で intensity の平均値を求め、実際の intensity がどちらの群に近いかによって、いずれの群に所属する遺伝子かを仮に決定する。その上で、2 群を判別する遺伝子群を選択する。並び替え検定 (Permutation tests) と交差検証 (Cross-validation) を繰り返し行った。仮説検定のための有意水準の数値 (p 値) が得られ、p 値が 0.01 以下になった遺伝子に対して、臨床サンプルを一例ずつ抜いて同じ解析を行い、全サンプルを入れ替えても、p 値が 0.01 以下であり、さらに 2 群の平均の差が 2 倍以上である遺伝子を、グループ間の判別に使用可能な遺伝子とした (7)。

4. T 因子、M 因子との比較

方法 3 で選び出された遺伝子群が、リンパ節転移あるなしの因子 (N 因子) 以外の従来用いられてきた臨床的な因子とどのような関係があるかを検討した。胃癌取り扱い規約に基づき、各臨床例を、T、M 因子の stage を決定し、選び出された遺伝子群の発現量を検討した。

結果

約 4600 遺伝子全スポットに対して、方法 (2) の解析を行った。全 16 症例において、十分な発現が確認され、解析対象となった遺伝子は 2127 個であった。リンパ節転移のある、なし群の判別に有意な遺伝子として、(2)

の解析で 228 遺伝子が選び出された。さらに並び替え検定 (Permutation tests) と交差検証を行い、リンパ節転移あるなし群を有意に判別する遺伝子 ($P < 0.01$) であり、かつ両群での平均値の差が 2 倍以上ある遺伝子として、99 遺伝子を選び出した。さらに、マイクロアレイ解析において、スポットの intensity が低い場合は、解析困難であることが多く、intensity 値が 1000 以上のものが今後の解析に耐えうるものと判断し、最終的に 29 遺伝子が、リンパ節転移の有無の判別の候補遺伝子群として選び出された (表 1、図 1)。29 遺伝子の中で、リンパ節転移群では、発現が上昇し、転移陰性群では発現が不変もしくは低下している遺伝子は 7 個認めた (表 2 *マーク)。この 7 個の遺伝子の発現について、T、M、N 因子について検討したところ、N 因子では、群間で明らかな発現差を認めたが、T 因子、M 因子との関連性は乏しかった (図 2-a, -b, -c)。

考察

マイクロアレイの解析方法については、近年統計学的手法を取り入れた解析が広く行われ、一定の評価を受けている。しかし、絶対的な解析方法というのではなく、目的や用途、また実験のスケールに応じて、最も適した解析方法を選択している。これまで我々は、マイクロアレイを用いた実験を行い、既に報告してきた。解析方法としては、いわゆるクラスター解析を中心に行ってきたが、臨床サンプルを用いた実験の場合、予後診断、病型診断など、2 つの群を分ける主因となる遺伝子群を検出することが必要になってきた。今研究で、株式会社ダイナコムと協力して、市販ソフトでは不可能な解析方法を試みた。その結果、29 個の遺伝子群を胃癌のリンパ節の有無を判別する遺伝子群として抽出した。興味深いことに、この 29 個の遺伝子群のうち、発現量の差が大きい 7 つの遺伝子についてみると、腫瘍の大きさや他臓器転移などのいわゆる T 因子、M 因子といわれる癌の進行度を表す因子とは発現量との関連が薄かったことである。これまで、リンパ節転移は、癌が浸潤、成長し起こるものであり、リンパ節転移を伴う癌に特異的な遺伝子が存在することは、我々も疑問視していた。しかしながら、今回の結果から、手術標本もしくは術前の内視鏡検査から得られる標本を使用し、遺伝子発現解析を行うことによって、リンパ節転移の有無の予測に役立つ可能性があり、患者にとって、治療法の選択にあたる上で、重要な情報提供となるかもしれない。本研究では、前向き試験を行っていないが、今後症例を積み重ねることによって、よりよいマイクロアレイ解析のアルゴリズム作製の手がかりとなるかもしれない。

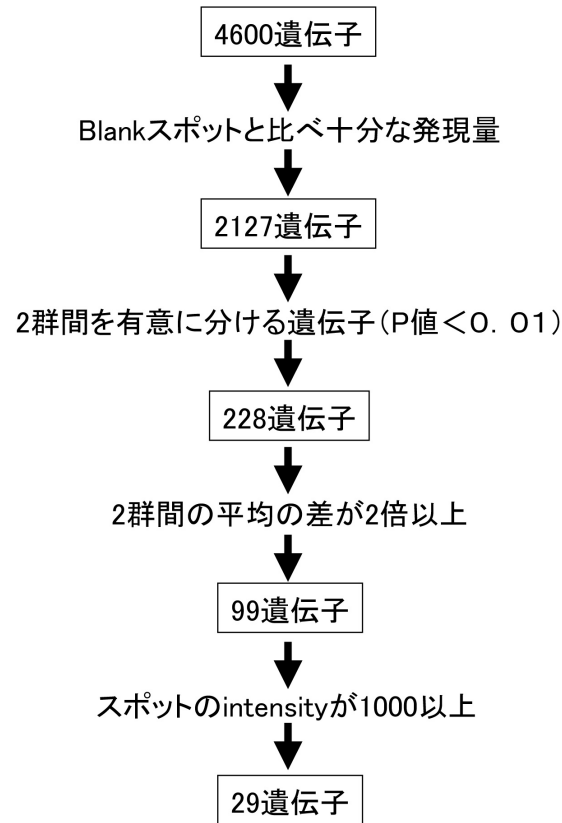


図 1. マイクロアレイ解析を基に、統計学的手法を用いた、リンパ節転移の有無判別遺伝子群の選り出し手順

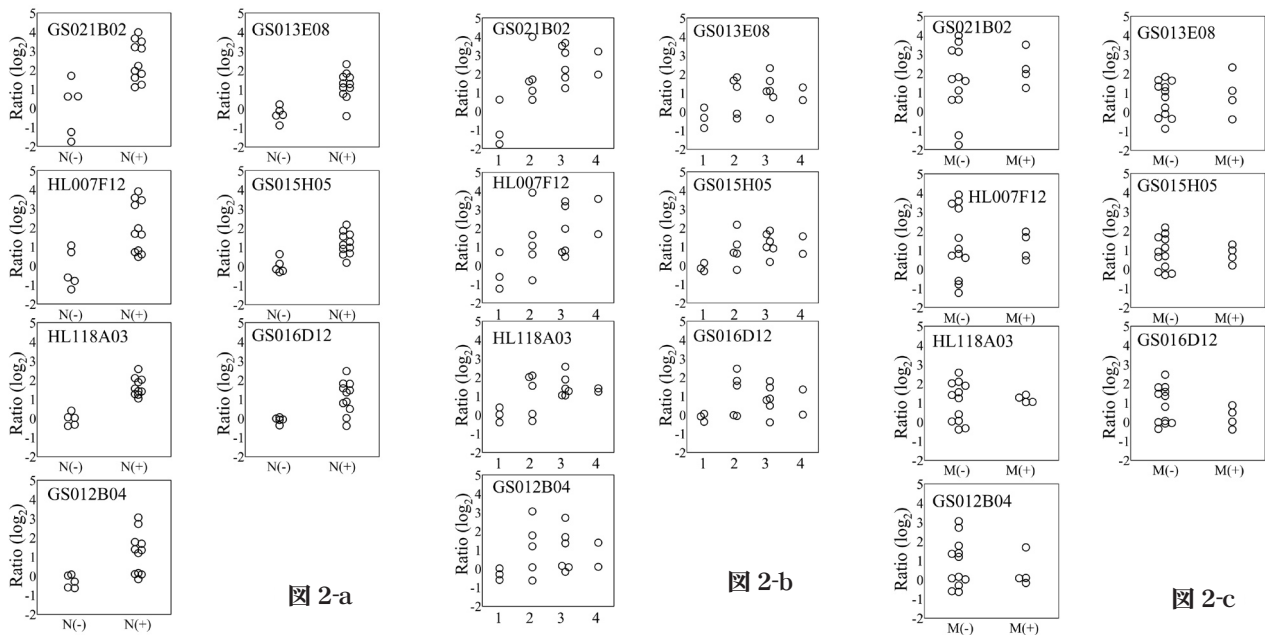


図 2. マイクロアレイ解析によって選り出された 7 つの遺伝子群の発現比

(a) リンパ節転移あり (N(+)), なし (N(-)) 群による分類。(b) T 因子 (1 から 4) による分類。x 軸の値は、T 因子を示す。(c) M 因子による分類。他臓器に転移があり (M(+)), なし (M(-)) 群

表 1. リンパ節転移の有無を判別する遺伝子セット

遺伝子名は、未発表のため、番号のみの表示とした。P 値は、各遺伝子が各群を判別可能かを表しており、0.01 以下を有意としている。リンパ節転移がある群を N(+), 無い群を N(-) とした。マイクロアレイでの結果の比は、全て log2 換算してある。

Gene Name	P 値	N(+)群の平均比 (log2)	N(-)群の平均比 (log2)	群間の変化量 (log2)
GS021B02*	0.0041	2.49	-0.02	2.51
HL007F12*	0.0091	2.00	-0.17	2.17
HL118A03*	0.0007	1.59	-0.05	1.64
GS012B04*	0.0046	1.21	-0.29	1.50
GS013E08*	0.0037	1.21	-0.29	1.49
GS015H05*	0.0025	1.18	0.02	1.17
GS016D12*	0.0007	1.11	-0.09	1.21
GS034E02	0.0009	0.90	-0.12	1.03
GS015A01	0.0054	0.87	-0.28	1.14
GS018B08	0.0075	0.81	-0.53	1.34
GS012F01	0.0047	0.76	-0.53	1.30
GS016A08	0.0024	0.74	-0.58	1.32
GS008C06	0.0095	0.70	-0.46	1.15
GS012D01	0.0009	0.47	-0.77	1.24
GS018F09	0.0098	0.47	-0.63	1.10
GS015D03	0.0012	0.38	-0.76	1.14
GS019E06	0.0015	0.26	-0.75	1.01
GS024C12	0.0090	-0.40	0.84	1.24
HL122D03	0.0043	-0.40	0.70	1.10
GS018G02	0.0073	-0.50	0.60	1.10
GS012A09	0.0033	0.04	-0.98	1.03
GS015G05	0.0018	0.13	-1.01	1.15
GS013A08	0.0008	0.16	-1.05	1.21
GS033B05	0.0017	-0.11	-1.12	1.01
HL116H03	0.0006	0.00	-1.13	1.13
HL087G12	0.0051	-0.05	-1.20	1.15
GS025A04	0.0006	-0.23	-1.30	1.07
HL121H07	0.0025	-0.08	-1.33	1.24
GS017C04	0.0018	-0.01	-1.39	1.39

文献

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999;286:531-537.
2. Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, Fukayama M, Kodama T, et al. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res*. 2002;62:233-240.
3. Kihara C, Tsunoda T, Tanaka T, Yamana H, Furukawa Y, Ono K, Kitahara O, Zembutsu H, Yanagawa R, Hirata K, Takagi T, Nakamura T. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res*. 2001;61:6474-6479.
4. Tahara E. Molecular biology of gastric cancer. *World J Surg*. 1995;19:484-488; discussion 489-490.
5. Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, Westerga J, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene*. 2003;22:1872-1879.
6. Mori M, Shimada H, Gunji Y, Matsubara H, Hayashi H, Nimura Y, Kato M, Takiguchi M, Ochiai T, Seki N. S100A11 gene identified by in-house cDNA microarray as an accurate predictor of lymph node metastases of gastric cancer. *Oncol Rep*. 2004;11:1287-1293.
7. I.S. Kohane, A.T. Kho, A.J. Butte 著 (星田有人 訳) . 統合ゲノミクスのためのマイクロアレイデータアナリシス . シュプリンガー・フェアラーク社