

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年5月21日現在

機関番号：15501

研究種目：基盤研究(C)

研究期間：2009～2011

課題番号：21500173

研究課題名（和文） 少数サンプルを対象とした統計的パターン認識とその肝臓診断への応用

研究課題名（英文） Statistical Pattern Recognition in Small Sample Size Situations and its Application to Detection of Hepatocellular Carcinoma

研究代表者

浜本 義彦 (HAMAMOTO YOSHIHIKO)

山口大学・大学院医学系研究科・教授

研究者番号：90198820

研究成果の概要（和文）：本研究ではC型肝炎ウイルスに起因する肝臓癌を対象とした癌診断システムの開発を目的とする。このため、超高次元で少数サンプルの状況下においても実行可能な特徴選択法を確立し、それに基づいて肝臓癌の診断に役立つ標的因子群としてメチル化遺伝子 SPINT2, SRD5A2 を発見した。さらにこれら2つの新規マーカーと AFP, PIVKA-II を用いて肝臓癌の診断システムを開発し、4施設から提供されたテストサンプル 258 症例に対して、血液検査で感度 73.2%、特異度 87.7%という高精度で肝臓癌を診断できたことから、臨床応用の見通しを得た。

研究成果の概要（英文）：The purpose of this study is to develop a novel blood test system for detecting hepatocellular carcinoma (HCC) associated with hepatitis C virus (HCV) infection. In order to develop the novel test system, we must establish a feature selection method to find effective bio-makers in small training sample size situations, compared with very high dimensions. We found 2 methylated genes (SPINT2 and SRD5A2). With use of 2 methylated genes, AFP, and PIVKA-II, we designed the blood test system for detecting HCC, which achieved high abilities (73.2% sensitivity and 87.7% specificity) for 258 test samples provided from 4 institutes. This suggests that our blood test system might lead to clinical applications.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,800,000	540,000	2,340,000
2010年度	1,100,000	330,000	1,430,000
2011年度	600,000	180,000	780,000
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

研究分野：情報工学

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：癌, パターン認識, 診断

1. 研究開始当初の背景

癌の克服は国民的課題と言っても過言ではない。そのために早期発見、早期治療が必要であるが、患者によって癌は異なり、癌の難治性はその多様性にある。このため癌研究においては、多様性の中に潜む普遍的な標的因子群を発見することが最も重要となっている。

癌研究において、20世紀末からの分子生物学の急速な発展による分子のアプローチが、これまでの医学を飛躍的に進歩させるものと期待された。実際、遺伝子発現データを用いるトランスクリプトーム解析と蛋白質データを用いるプロテオーム解析が国内外で癌への応用を目的として盛んに研究された。しかし、癌は複雑であり、これらはいずれも癌の側面を断片的にしか捉えておらず、その限界が明らかとなった。現状では、遺伝子発現データや蛋白質データなどの分子データは、いずれも個々では決め手に欠けている。

しかし、個々の分子データには癌情報の片鱗しかないが、異種の分子データを組合せる、さらには医学に立脚した臨床データをも組合せて互いに補完しあうことで、癌を網羅的に捉える可能性が高まる。このような考えのもとで、分子データと臨床データを統合した疾患データベースの構築が、国立がんセンター研究所や産業技術総合研究所で進められている。

一方、これを診断・予測システムを設計する立場から見ると、分子データと臨床データの両方を対象とすることは、患者を記述する因子の数、すなわち次元数の増大を意味する。さらに問題なのは、臨床の裏づけがあるという意味で信頼性の確保されたサンプル数が少ないことである。従って、癌の診断・予測システムを設計するには、候補となる因子が数千から数万の超高次元に対して、サンプルが数十からせいぜい数百しかないという極めて困難なパターン認識問題を解かなければならない。疾患データベースの構築とともに並行して、この困難なパターン認識問題を解決することは、癌研究において車の両輪のごとく重要であると考えられる。

2. 研究の目的

本研究では、超高次元で少数サンプルの状況下においても実行可能な特徴選択法を確立し、それに基づいて癌の多様性の中に潜む普遍的な、肝癌の診断に役立つ標的因子群を発見する。さらに標的因子群を用いて肝癌の診断システムを開発することを目的とする。

3. 研究の方法

(1) 特徴選択

図に全体像を示す。実サンプル集合から仮想化処理により複数の仮想サンプル集合を生成する。次に、仮想サンプル集合毎に Fisher 比により有力な候補因子を独立に選択する。統合処理では、各有力候補因子集合に対して、仮想サンプル集合が変わっても有効な因子の組合せを標的因子群として選択する。本特徴選択法では、仮想化処理と統合処理が重要な役割を果たす。

まず仮想化処理において、複数の仮想サンプル集合を、実サンプル集合から復元抽出か非復元抽出により求めるか、あるいは実サンプルにノイズを付加する、または実サンプルの局所平均化などにより求める。

統合処理では、有力候補因子集合に含まれる因子に注目する。そして、これら因子の組合せに関して仮想サンプル集合上で特徴評価関数値の平均をとり、この平均値が大きい因子の組合せを標的因子群として選択する。

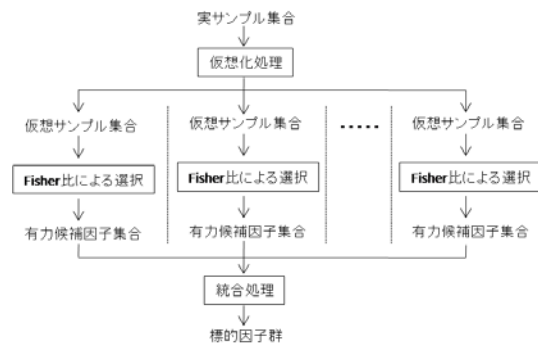


図 仮想サンプル変動を考慮に入れた特徴選択法

(2) Fisher 識別器の高精化

Fisher 線形識別器は、以下のように設計される。各クラスに対して訓練サンプル集合 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ が与えられているものとする。この訓練サンプル集合を用いて、以下の式により各クラスの平均ベクトルと共分散行列を推定する。

$$\text{平均ベクトル} \quad \boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$$

$$\text{共分散行列} \quad \boldsymbol{\Sigma} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$$

Fisher 線形識別器は

$$f_1(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \left[\frac{1}{2}\boldsymbol{\Sigma}_1 + \frac{1}{2}\boldsymbol{\Sigma}_2 \right]^{-1} (\mathbf{x} - \boldsymbol{\mu}_1)$$

$$f_2(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \left[\frac{1}{2}\boldsymbol{\Sigma}_1 + \frac{1}{2}\boldsymbol{\Sigma}_2 \right]^{-1} (\mathbf{x} - \boldsymbol{\mu}_2)$$

を用いて、線形識別関数

$$f_1(\mathbf{x}) - f_2(\mathbf{x}) + \alpha = 0 \quad \alpha : \text{Cut off}$$

で表すことができる。この α の最適化が重要となる。本研究では、 α の候補値を予め複数定め、感度と特異度を評価基準とした組合せ最適化問題を解くことにより α の最適化を行った。

次に、入力データの変換を行う。これは、入力データが広い範囲で分布し、これを一定の範囲内に収めることが必要となった理由から行うものである。具体的には、べき変換により入力データの変換を行うが、このときべき変換のパラメータの最適化が問題となり、カットオフ値 α と同様にしてべき変換のパラメータの最適化を行った。

4. 研究成果

肝癌診断用のマーカーとして、新たにメチル化遺伝子 SPINT2, SRD5A2 を発見した。更に既存の腫瘍マーカーである AFP、PIVKA-II と SPINT2, SRD5A2 を組み合わせて高精度化された Fisher 線形識別器により、テストサンプル 258 症例に対して、血液検査だけで肝癌を感度 73.2%、特異度 87.7% という高精度で診断することができた。

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 4 件)

- ① N. Iizuka, M. Oka, I. Sakaida, T. Moribe, T. Miura, N. Kimura, S. Tamatsukuri, H. Ishitsuka, K. Uchida, S. Terai, S. Yamashita, K. Okita, K. Sakata, Y. Karino, J. Toyota, E. Ando, T. Ide, M. Sata, R. Tsunedomi, M. Tsutsui, M. Iida, Y. Tokuhisa, K. Sakamoto, T. Tamesa, Y. Fujita, Y. Hamamoto, Efficient detection of hepatocellular carcinoma by a hybrid blood test of epigenetic and classical protein markers, Clinica Chimica Acta, 査読有, 412, pp.152-158(2011).
- ② M. Tsutsui, N. Iizuka, T. Moribe, T. Miura, N. Kimura, S. Tamatsukuri, H. Ishitsuka, Y. Fujita, Y. Hamamoto, R. Tsunedomi, M. Iida, Y. Tokuhisa, K. Sakamoto, T. Tamesa, I. Sakaida and M. Oka, Methylated cyclin D2 gene circulating in the blood as a prognosis predictor of hepatocellular carcinoma, Clinica Chimica Acta, 査読有, 411, pp.516-520(2010).
- ③ N. Kimura, T. Moribe, N. Iizuka, T. Miura, S. Tamatsukuri, H. Ishitsuka, Y. Hamamoto and M. Oka, Rapid and quantitative

detection of CpG-methylation status using TaqMan PCR combined with methyl-binding-domain

polypeptide, Clinical Biochemistry, 査読有, 42, pp.1113-1122, (2009).

- ④ T. Moribe, N. Iizuka, T. Miura, N. Kimura, S. Tamatsukuri, H. Ishitsuka, Y. Hamamoto, K. Sakamoto, T. Tamesa and M. Oka, Methylation of multiple genes as molecular markers for diagnosis of a small well-differentiated hepatocellular carcinoma, Int. J. Cancer, 査読有, 125, pp.388-397, (2009).

[学会発表] (計 2 件)

- ① 飯塚徳男、岡 正朗、恒富亮一、藤田悠介、浜本義彦、森部豊輝、血清 CIMP と肝細胞癌の予後、第 70 回日本癌学会学術総会、名古屋国際会議場、名古屋、p.192、2011.10.3.
- ② Y. Fujita, Y. Hamamoto, N. Iizuka, M. Oka, T. Moribe, An optimized Fisher linear classifier for detection of hepatocellular carcinoma, 8th AACR-JCA Joint Conference Cancer Genomics, Epigenomics and the Development of Novel Therapeutics, Hilton Waikoloa Village, Hawaii (米国), February 6, 2010.

[図書] (計 1 件)

- ① 浜本義彦、統計的パターン認識入門、森北出版、2009、122.

[産業財産権]

○出願状況 (計 1 件)

名称：AFP および PIVKA-II の測定値を特徴値とした識別関数を利用する、肝臓癌および慢性肝疾患の病態進行度の検出方法

発明者：木村直紀、森部豊輝、玉造 滋、

飯塚徳男、岡正朗、浜本義彦

権利者：エフ ホフマンシーラ ロシュ アーゲー

種類：特許

番号：特願 2009-094252

出願年月日：平成 21 年 4 月 8 日

国内外の別：国内

[その他]

ホームページ等

<http://www.ir.csse.yamaguchi-u.ac.jp/>

6. 研究組織

(1)研究代表者

浜本 義彦 (HAMAMOTO YOSHIHIKO)

山口大学・大学院医学系研究科・教授

研究者番号：90198820

(2)研究分担者

飯塚 徳男 (IIZUKA NORIO)

山口大学・医学部附属病院・准教授

研究者番号：80332807

藤田 悠介(FUJITA YUSUKE)

山口大学・大学院医学系研究科・助教

研究者番号：40509527