

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2007-37409
(P2007-37409A)

(43) 公開日 平成19年2月15日(2007.2.15)

(51) Int. Cl.		F I			テーマコード (参考)
C 1 2 N 15/09 (2006.01)		C 1 2 N	15/00	Z N A A	4 B O 2 4
C 1 2 Q 1/68 (2006.01)		C 1 2 Q	1/68	A	4 B O 6 3

審査請求 未請求 請求項の数 11 O L (全 20 頁)

<p>(21) 出願番号 特願2005-222083 (P2005-222083)</p> <p>(22) 出願日 平成17年7月29日 (2005.7.29)</p>	<p>(71) 出願人 000207827 大鵬薬品工業株式会社 東京都千代田区神田錦町1-27</p> <p>(71) 出願人 504013775 学校法人 埼玉医科大学 埼玉県入間郡毛呂山町毛呂本郷38</p> <p>(74) 代理人 110000084 特許業務法人アルガ特許事務所</p> <p>(74) 代理人 100068700 弁理士 有賀 三幸</p> <p>(74) 代理人 100077562 弁理士 高野 登志雄</p> <p>(74) 代理人 100096736 弁理士 中嶋 俊夫</p>
--	---

最終頁に続く

(54) 【発明の名称】 抗癌剤投与後の大腸癌患者の予後予測方法

(57) 【要約】 (修正有)

【課題】 抗癌剤投与開始後の大腸癌患者の予後を精度よく予測するための遺伝子群及びこれを用いた大腸癌患者の予後予測方法を提供すること。

【解決手段】 抗癌剤の投与を受けた大腸癌患者の予後を予測するための遺伝子群であって、以下の(1)~(8)で示される予後と正相関する遺伝子又はこれらに相当する遺伝子、及び(9)~(16)で示される予後と負相関する遺伝子又はこれらに相当する遺伝子から選ばれる2以上の遺伝子からなる遺伝子群。(1) TM4SF6遺伝子、(2) ADRM1遺伝子、(3) ITGB4BP遺伝子、(4) FLOT2遺伝子、(5) PHB遺伝子、(6) GSTM1遺伝子、(7) BCAP31遺伝子、(8) STIP1遺伝子、(9) TRIM8遺伝子、(10) MST1R遺伝子、(11) TCF12遺伝子、(12) PDE3B遺伝子、(13) CCNG2遺伝子、(14) SAT遺伝子、(15) ALOX5遺伝子、(16) MAN2A2遺伝子

【選択図】 なし

【特許請求の範囲】

【請求項 1】

抗癌剤の投与を受けた大腸癌患者の予後を予測するための遺伝子群であって、以下の(1)～(8)で示される予後と正相関する遺伝子又はこれらに相当する遺伝子、及び(9)～(16)で示される予後と負相関する遺伝子又はこれらに相当する遺伝子から選ばれる2以上の遺伝子からなる遺伝子群。

(1) TM4SF6遺伝子、(2) ADRM1遺伝子、(3) ITGB4BP遺伝子、(4) FLOT2遺伝子、(5) PHB遺伝子、(6) GSTM1遺伝子、(7) BCAP31遺伝子、(8) STIP1遺伝子、(9) TRIM8遺伝子、(10) MST1R遺伝子、(11) TCF12遺伝子、(12) PDE3B遺伝子、(13) CCNG2遺伝子、(14) SAT遺伝子、(15) ALOX5遺伝子、(16) MAN2A2遺伝子

10

【請求項 2】

(1)～(8)で示される予後と正相関する遺伝子又はこれらに相当する遺伝子から選ばれる遺伝子数と、(9)～(16)で示される予後と負相関する遺伝子又はこれらに相当する遺伝子から選ばれる遺伝子数とが同数である請求項1記載の遺伝子群。

【請求項 3】

(1)～(8)で示される予後と正相関する遺伝子及び(9)～(16)で示される予後と負相関する遺伝子のそれぞれにおいて、予後との相関が最も強い遺伝子から相関の強い順番に選ばれた1以上の遺伝子又はこれらに相当する遺伝子、からなる請求項1又は2記載の遺伝子群。

【請求項 4】

予測精度が90%以上である請求項1～3のいずれか1項記載の遺伝子群。

20

【請求項 5】

以下の(1)～(6)で示される予後と正相関する遺伝子又はこれらに相当する遺伝子、及び(9)～(14)で示される予後と負相関する遺伝子又はこれらに相当する遺伝子、からなる請求項1又は2記載の遺伝子群。

(1) TM4SF6遺伝子、(2) ADRM1遺伝子、(3) ITGB4BP遺伝子、(4) FLOT2遺伝子、(5) PHB遺伝子、(6) GSTM1遺伝子、(9) TRIM8遺伝子、(10) MST1R遺伝子、(11) TCF12遺伝子、(12) PDE3B遺伝子、(13) CCNG2遺伝子、(14) SAT遺伝子

【請求項 6】

以下の(1)～(7)で示される予後と正相関する遺伝子又はこれらに相当する遺伝子、及び(9)～(15)で示される予後と負相関する遺伝子又はこれらに相当する遺伝子、からなる請求項1又は2記載の遺伝子群。

30

(1) TM4SF6遺伝子、(2) ADRM1遺伝子、(3) ITGB4BP遺伝子、(4) FLOT2遺伝子、(5) PHB遺伝子、(6) GSTM1遺伝子、(7) BCAP31遺伝子、(9) TRIM8遺伝子、(10) MST1R遺伝子、(11) TCF12遺伝子、(12) PDE3B遺伝子、(13) CCNG2遺伝子、(14) SAT遺伝子、(15) ALOX5遺伝子

【請求項 7】

以下の(1)～(8)で示される予後と正相関する遺伝子又はこれらに相当する遺伝子、及び(9)～(16)で示される予後と負相関する遺伝子又はこれらに相当する遺伝子、からなる請求項1又は2記載の遺伝子群。

40

(1) TM4SF6遺伝子、(2) ADRM1遺伝子、(3) ITGB4BP遺伝子、(4) FLOT2遺伝子、(5) PHB遺伝子、(6) GSTM1遺伝子、(7) BCAP31遺伝子、(8) STIP1遺伝子、(9) TRIM8遺伝子、(10) MST1R遺伝子、(11) TCF12遺伝子、(12) PDE3B遺伝子、(13) CCNG2遺伝子、(14) SAT遺伝子、(15) ALOX5遺伝子、(16) MAN2A2遺伝子

【請求項 8】

抗癌剤が5-フルオロウラシル系抗癌剤である請求項1～7のいずれか1項記載の遺伝子群。

【請求項 9】

抗癌剤が5-フルオロウラシルとロイコボリンの併用である請求項1～7のいずれか1項記載の遺伝子群。

50

【請求項10】

請求項1～9のいずれか1項記載の遺伝子群の各遺伝子の発現量を指標とすることを特徴とする抗癌剤の投与を受けた大腸癌患者の予後予測方法。

【請求項11】

以下の工程を含む、請求項10記載の予後予測方法。

1) 大腸癌患者の検体から、請求項1～9のいずれか1項記載の遺伝子群の発現量を測定する。

2) 各遺伝子の発現量を予後予測モデルに適用して得られる長期生存確率Pが0.5以上のとき、該患者の予後が良好と判定し、0.5未満のとき、該患者の予後が不良と判定する。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、抗癌剤投与を受けた癌患者の予後予測に有用な遺伝子群、及びこれを用いた大腸癌患者の予後予測方法に関するものである。

【背景技術】

【0002】

わが国の癌の罹患率・死亡率はともに増加傾向にあり改善が求められている。とりわけ大腸癌における増加は顕著であり、現在において死亡率は男性で第4位、女性では第2位、全体で第3位であるものの、2015年の癌罹患患者数の推計では、大腸癌（結腸癌＋直腸癌）患者は約17万人に及び、胃癌、肺癌を抜いて第1位となると予測されている。

20

【0003】

癌の治療法としては抗癌剤による方法が代表的であり、例えば、大腸癌に対しては、5-フルオロウラシル（5-FU）やロイコボリン（LV）が知られている。5-FUはウラシルの代謝拮抗剤であり、LVは5-FUの抗腫瘍効果増強剤である。5-FUとLVを組み合わせること（「5-FU/LV」と称す）によってより強い抗腫瘍効果が発揮され、5-FU/LVは大腸癌における標準療法として使用されてきた。抗腫瘍作用は、5-FUの活性代謝物であるFUTP（フルオロウリジン三リン酸）によるRNAの機能障害とFdUMP（フルオロデオキシウリジン一リン酸）によるDNAの合成阻害の、2つの機序によって引き起こされることが知られている。

【0004】

しかしながら、同じ臓器に生じた癌でも、個々の癌細胞でも、それぞれの患者によって、あるいは癌が進行していくに従って、個々の癌細胞の性質は大きく異なるため、必ずしも全て患者の癌に同一の抗癌剤が効くわけではない。したがって、投与前に抗癌剤に対する感受性・抵抗性を把握し、抗癌剤の有効性を予測できることが望ましい。

30

【0005】

近年、癌組織における特定の遺伝子の発現量を調べることにより、抗癌剤投与後の患者の予後を予測する試みが行われている。例えば、5-FU/LV投与を受けた患者においては、遺伝子DPD（ジヒドロピリミジンデヒドロゲナーゼ）、TS（チミジル酸生成酵素）、TP（チミジン・ホスホリラーゼ）の発現量がその後の生存時間と相関があることが示され（非特許文献1）、遺伝子DCCの欠失がその後の生存時間と相関があることが示されている（非特許文献2）。また、抗癌剤UFT/LV投与を受けた大腸癌患者においては、遺伝子OPRT（オロチン酸ホスホリボシルトランスフェラーゼ）の発現量と予後との間に相関があることが示されている（非特許文献3）。さらに、PHB遺伝子が骨肉腫において（非特許文献4）、GSTM1遺伝子が乳癌において（特許文献1）、それぞれ予後と正相関することが知られており、MST1R遺伝子が肺癌において（非特許文献5）、ALOX5遺伝子及びMAN2A2遺伝子が腎細胞癌において（特許文献2）、それぞれ予後と負相関することが知られている。しかしながら、これらの遺伝子を用いた予後予測の精度は満足できるものではない。

40

【特許文献1】国際公開第03/078662号パンフレット

【特許文献2】国際公開第2004/097052号パンフレット

【非特許文献1】Clin Cancer Res. 2000 Apr; 6(4):1322-7

50

【非特許文献2】J Clin Oncol. 2004 Sep 15; 22(18): 3758-65

【非特許文献3】Br J Cancer. 2003 Oct 20; 89(8): 1486-92

【非特許文献4】International Journal of Cancer (2003), 105(5), 636-643

【非特許文献5】Cancer Epidemiol Biomarkers Prev.2003 Sep;12(9):905-10.

【発明の開示】

【発明が解決しようとする課題】

【0006】

本発明の目的は、抗癌剤投与開始後の大腸癌患者の予後を精度よく予測するための遺伝子群及びこれを用いた大腸癌患者の予後予測方法を提供することにある。

【課題を解決するための手段】

【0007】

本発明者らは、5-FU/LVで治療を行った転移性大腸癌の症例について大腸癌組織における遺伝子発現データを取得し、統計解析を行うことにより、患者の予後と正相関する遺伝子8種類と負相関する遺伝子8種類の遺伝子が互いに予測精度を高める関係にあり、当該遺伝子の2以上を用いることにより、実用可能な程度にまで精度よく抗癌剤投与開始後の患者の長期生存確率を予測できることを見出した。

【0008】

すなわち、本発明は、抗癌剤の投与を受けた大腸癌患者の予後を予測するための遺伝子群であって、以下の(1)~(8)で示される予後と正相関する遺伝子又はこれらに相当する遺伝子、及び(9)~(16)で示される予後と負相関する遺伝子又はこれらに相当する遺伝子から選ばれる2以上の遺伝子からなる遺伝子群を提供するものである。

(1) TM4SF6遺伝子、(2) ADRM1遺伝子、(3) ITGB4BP遺伝子、(4) FLOT2遺伝子、(5) PHB遺伝子、(6) GSTM1遺伝子、(7) BCAP31遺伝子、(8) STIP1遺伝子、(9) TRIM8遺伝子、(10) MST1R遺伝子、(11) TCF12遺伝子、(12) PDE3B遺伝子、(13) CCNG2遺伝子、(14) SAT遺伝子、(15) ALOX5遺伝子、(16) MAN2A2遺伝子。

【0009】

また、本発明は、上記遺伝子群の各遺伝子の発現量を指標とすることを特徴とする抗癌剤の投与を受けた大腸癌患者の予後予測方法を提供するものである。

【発明の効果】

【0010】

本発明によれば、大腸癌患者の抗癌剤投与開始後の生存時間の長短を精度よく、且つ簡便に予測することができる。また、この予測結果に基づき、当該薬剤を用いるか否かの選択をすることもできる。

【発明を実施するための最良の形態】

【0011】

本発明における抗癌剤としては、癌細胞増殖抑制作用をもつ化合物(例えば、代謝拮抗剤、アルキル化剤、抗癌抗生物質、植物由来抗癌剤、免疫賦活剤、サイトカイン等)であれば特に限定されず、市販の抗癌剤のほか開発中の抗癌剤であってもよいが、代謝拮抗剤が好ましく、特に5-フルオロウラシル系抗癌剤が好ましい。

5-フルオロウラシル系抗癌剤としては、特に限定されないが、5-FU又は生体内において5-FUに代謝され抗腫瘍効果を発揮する化合物、例えばテガフル、テガフル・ウラシル配合剤(UFT)、テガフル・ギメラシル・オテラシルカリウム配合剤(TS-1)、カルモフル、カペシタピン、ドキシフルリジン等又は5-フルオロデオキシウリジン及びその誘導体等が挙げられる。ここで、5-FUとはウラシル代謝拮抗作用を有する抗癌剤であることが知られているものである。また、本発明の抗癌剤投与には、単独投与の他、複数の薬剤を併用投与する場合も包含され、例えば、5-フルオロウラシル系抗癌剤と併用することができる薬剤としては、ロイコボリン(LV)、イリノテカン、オキサリプラチン等が挙げられ、これらを1以上併用することができる。これらを併用することにより、5-フルオロウラシル系抗癌剤の抗腫瘍効果を増強することができる。本発明の癌患者の予後予測にお

10

20

30

40

50

いては、5-FUとLVの併用（5-FU/LV）投与の場合に特に適する。

【0012】

本発明における遺伝子群は、（１）TM4SF6遺伝子、（２）ADRM1遺伝子、（３）ITGB4BP遺伝子、（４）FLOT2遺伝子、（５）PHB遺伝子、（６）GSTM1遺伝子、（７）BCAP31遺伝子及び（８）STIP1遺伝子からなる予後と正相関する遺伝子又はこれらに相当する遺伝子、及び（９）TRIM8遺伝子、（１０）MST1R遺伝子、（１１）TCF12遺伝子、（１２）PDE3B遺伝子、（１３）CCNG2遺伝子、（１４）SAT遺伝子、（１５）ALOX5遺伝子及び（１６）MAN2A2遺伝子からなる予後と負相関する遺伝子又はこれらに相当する遺伝子から選ばれる２以上の遺伝子よりなるものである。

上記遺伝子群は、転移性大腸癌症例を対象とした、抗癌剤投与における患者の生存期間と遺伝子発現量とのSpearman順位相関係数を求めることにより、生存期間との相関があった遺伝子群のうち生存期間と正相関する遺伝子及び負相関する遺伝子のそれぞれにおいて最も相関の強い遺伝子から相関の強い順番に選ばれたものである。これらの遺伝子のGenBankアクセッション番号、機能、相関係数及びこれらの遺伝子の塩基配列を示した配列表における配列番号を下記表１に示す。（１）TM4SF6遺伝子、（２）ADRM1遺伝子、（３）ITGB4BP遺伝子、（４）FLOT2遺伝子、（５）PHB遺伝子、（６）GSTM1遺伝子、（７）BCAP31遺伝子及び（８）STIP1遺伝子は大腸癌患者の予後と正相関する遺伝子であり、（９）TRIM8遺伝子、（１０）MST1R遺伝子、（１１）TCF12遺伝子、（１２）PDE3B遺伝子、（１３）CCNG2遺伝子、（１４）SAT遺伝子、（１５）ALOX5遺伝子及び（１６）MAN2A2遺伝子は予後と負相関する遺伝子である。

本発明にかかる予後と正相関する遺伝子のうち、PHB遺伝子が骨肉腫において（International Journal of Cancer (2003), 105(5), 636-643）、GSTM1遺伝子が乳癌において（国際公開第03/078662号）、それぞれ予後と正相関することは公知である。さらに、本発明にかかる予後と負相関する遺伝子のうち、MST1R遺伝子が肺癌において（Cancer Epidemiol Biomarkers Prev.2003 Sep;12(9):905-10.）、ALOX5遺伝子及びMAN2A2遺伝子が腎細胞癌において（国際公開第2004/097052号）、それぞれ予後と負相関することも公知である。

しかし、本発明遺伝子群が、大腸癌において予後に関与していることは知られておらず、また、癌腫が違えば、予測に関わる遺伝子群が異なることが知られている。

【0013】

10

20

30

」程度の条件、さらに厳格には「 $0.1 \times \text{SSC}$ 、 $0.1\% \text{SDS}$ 、 65°C 」程度の条件が挙げられる。

【0016】

上記遺伝子の取得は、公知の方法により得られた染色体DNAと、適当なプライマーとを用いてPCR法により行うことができる。本発明の遺伝子の配列は公知であるため、かかるPCR法の採用に際して使用されるプライマーは、かかる配列に基づいて適宜設定することができ、常法に従って合成することができる。尚、増幅させたDNA断片の単離精製は、例えばゲル電気泳動法等により行うことができる。

【0017】

尚、上記遺伝子群は、大腸癌患者の予後を予測するために有用な遺伝子を含むものであるが、大腸癌患者の予後予測に当たっては、当該遺伝子群にさらに別の遺伝子を加えることにより予測精度が上がる場合は、当該遺伝子を加えて予後予測を行うことができる。

10

【0018】

上記遺伝子群の各遺伝子の発現量を調べ、これを指標とすることによって、抗癌剤投与を受けた大腸癌患者の予後が良好であるかどうか、すなわち抗癌剤投与開始後の生存時間が長いかどうかについて予測することができる。

【0019】

上記遺伝子群の各遺伝子の発現量を調べるためには、患者の大腸癌の原発巣切除物などから抽出されたRNAをもとに逆転写酵素などの酵素によってcDNAを調整したのち、かかるcDNAにしめる各遺伝子のcDNA量を測定すればよい。また、同様に抽出されたタンパク質にしめる各遺伝子の翻訳産物量を測定すればよい。翻訳産物としては、タンパク質、ポリペ

20

【0020】

各遺伝子のcDNA量又は翻訳産物量の測定は、それぞれそのプローブ、抗体などを用いた公知の方法により行うことができる。ここで測定の便宜として、cDNAに適当な標識を付加することが好ましい。かかる標識の付加は公知の方法により行うことができる。ここで、標識としては、酵素、ビオチン、蛍光物質、ハプテン、抗原、抗体、放射性物質、発光団が挙げられ、酵素としては、アルカリフォスファターゼ、ペルオキシダーゼなどが挙げられ、蛍光物質としては、FITC、6-FAM、HEX、TET、TAMRA、テキサスレッド、Cy3、Cy5などが挙げられ、ハプテンとしては、ビオチン、ジゴキシゲニンなどが挙げられる。放射性物質としては、 ^{32}P 、 ^{35}S などが挙げられ、発光団としては、ルテニウムなどが挙げられるが、標識としてはビオチンが好ましい。かかる標識由来のシグナルの強度を測定することにより、当該遺伝子発現量を反映した数値を求めることができる。また、cDNA量の測定を行うに際しては、かかるcDNAを鋳型として、*in vitro* transcriptionなどの方法によって増幅したcRNAの量を指標としてもよい。

30

【0021】

また、かかる測定に際しては、複数のプローブが固定されたマイクロアレイを用いれば、容易にかつ網羅的に行うことができる。マイクロアレイとしては、GeneChip(登録商標)プローブアレイが好ましい。

【0022】

発現量を調べる2以上の遺伝子の選択は、用いる遺伝子によって予後予測精度が変わりうるので、精度がより高くなるような遺伝子又はその組み合わせを選択することにより行うのが好ましい。かかる選択は、選択された遺伝子の発現量から予後予測モデルを作成し、この予後予測モデルがどの程度の精度で予後を予測できるか評価することにより行うことができる。ここで、予測精度は、例えば、leave-one-out cross validation(CV)法、hold-out methodなどにより計算することができる。すなわち、本発明の遺伝子群としては、作成された予後予測モデルについてleave-one-out CV法によって計算される予測精度が、90%以上、好ましくは92%以上、より好ましくは94%以上となる遺伝子群が好ましい。

40

【0023】

なお、leave-one-out CV法とは、標本サイズ n のデータセットに対して、 $n - 1$ 個を使

50

ってモデルを構築し、残しておいた1個のデータに対してモデルを適用して結果を判定するという作業を順番にn回(データの数)実施することにより、モデルの予測精度を調べる方法である。また、hold-out methodとは、標本をランダムに2分割し、一方(training set)をモデル構成に使い、そこで得たモデルを他方(testing set)のデータに対して適用し、モデルの良さを検討することにより、モデルの予測精度を調べることができる方法である。

【0024】

ここで、患者の遺伝子発現量を予後予測モデルに適用した際の予後予測の当否判定は、予後予測モデルによって長期生存確率Pが $P > 0.5$ と計算された患者が、実際に長期生存者(ここでは、生存時間が、集団の中央値よりも長い患者)であったとき、予測が当たったとし、該患者が、実際に長期生存者でなかったとき、予測が外れたとすることにより行うことができる。

10

【0025】

また、後記実施例表2に示すとおり、前述した表1記載の予後と正相関する遺伝子及び予後と負相関する遺伝子のそれぞれにおいて、表1に示される相関係数の絶対値により把握される予後との相関が最も強い遺伝子から相関の強い順番に同数ずつ遺伝子を選んだ場合、後記実施例の予後予測モデル作成に使用したデータセットに対し80%以上の予測精度を有する予後予測モデルを作成することができた。さらに、前述した表1記載の予後と正相関する遺伝子及び予後と負相関する遺伝子のそれぞれにおいて、最も相関の強い遺伝子から相関の強い順番に6以上の遺伝子を選び、12、14又は16の遺伝子を選んだ場合、後記実施例の予後予測モデル作成に使用したデータセットに対し90%以上の予測精度を有し、かつ平均二乗誤差が0.01未満である予後予測モデルを作成することができた。特に、前述した表1記載の本発明の遺伝子の全て、すなわち表1記載の16の遺伝子を全て用いた場合、後記実施例の予後予測モデル作成に使用したデータセットに対し100%の予測精度を有し、かつ平均二乗誤差が0.0004である予後予測モデルを作成することができた。このことは、かかる12、14又は16の遺伝子は、高い予測精度を実現できる好適な組み合わせであることを意味すると考えられる。したがって、本発明の遺伝子群としては、かかる12、14又は16の遺伝子からなる遺伝子群が好ましく、かかる16の遺伝子からなる遺伝子群がより好ましい。

20

【0026】

また、後記実施例に示すとおり、本発明の遺伝子群を構成する16の遺伝子に、これらの次に相関が高い遺伝子をさらに加えて予後予測モデルを作成した場合、特に150以上の遺伝子を用いて予後予測モデルを作成したとき、かかる予後予測モデルの予測精度は、むしろ下がることもあることが明らかになった(図1)。これは、遺伝子数を増やすことにより、ノイズなどの影響が無視できなくなったため、又は、予測精度が不十分である遺伝子をも加えたためであると考えられる。これに対し、選ばれる遺伝子を表1記載の16の遺伝子から選んで用いる場合であれば、用いる遺伝子数を増やすことによって、概して予後予測精度が向上することが明らかになった(図1)。したがって、かかる16の遺伝子は、他の遺伝子に比べると互いに予測精度を下げる影響が少ない点で、有用な組み合わせであると考えられる。

30

40

【0027】

なお、後記比較例1に示すとおり、本発明の遺伝子群を構成する16の遺伝子について、単独の遺伝子を用いて予後予測モデルを作成した場合、かかる予後予測モデルの予測精度は十分でない。これに対し、2以上の遺伝子、特に後記実施例2に示すとおり、上記16の遺伝子を用いた場合には、実用可能な程度にまで精度よく抗癌剤投与開始後の患者の長期生存確率を予測できる。これは、これら16遺伝子の組み合わせが好ましく、互いに予測精度を上げる関係にあるためであると考えられる。

【0028】

さらに、後記実施例3に示すとおり、前述した本発明の遺伝子群から遺伝子を選ぶ方法としては、正相関のものと負相関のものとを同数ずつ選ぶ方法が、より高い予測精度が実

50

現できる点で好ましく、また、これに加えて、予後と正相関する遺伝子及び予後と負相関する遺伝子のそれぞれにおいて、表1に示される抗癌剤投与後の生存時間と発現量とのSpearman順位相関係数の絶対値が最も大きいものから順に選ぶという方法がより好ましい。

【0029】

すなわち、本発明の遺伝子群から2以上の遺伝子を選ぶ方法としては、「予後と正相関する遺伝子又はこれらに相当する遺伝子から選ばれる遺伝子数と、予後と負相関する遺伝子又はこれらに相当する遺伝子から選ばれる遺伝子数とが同数であるように選ぶ方法」が好ましく、「予後と正相関する遺伝子及び予後と負相関する遺伝子のそれぞれにおいて、予後との相関が最も強い遺伝子から相関の強い順番に選ばれる1以上の遺伝子又はこれらに相当する遺伝子を選ぶ方法」が好ましく、これら2つの方法をあわせて行う方法がより好ましい。

10

ここで「予後と正相関する遺伝子及び予後と負相関する遺伝子のそれぞれにおいて、予後との相関が最も強い遺伝子から相関の強い順番に選ばれた」遺伝子とは、例えば、上位から3の遺伝子を選ぶ場合、(1)TM4SF6遺伝子、(2)ADRM1遺伝子、(3)ITGB4BP遺伝子、(9)TRIM8遺伝子、(10)MST1R遺伝子、(11)TCF12遺伝子を示す。

【0030】

かくして調べた各遺伝子の発現量を、当該遺伝子についての他の患者における発現量のデータなど、適当な対照と比較することにより、抗癌剤投与を受けた大腸癌患者の予後が良好であるかどうかを予測することができる。

【0031】

ここで適当な対照と比較する方法としては、具体的には、予後が良い患者と悪い患者の遺伝子発現量のデータを用いて統計解析処理すること、例えば、かかるデータを反映する予後予測モデルを作り、かかる予後予測モデルに、予測が目的となる患者の遺伝子発現量を当てはめ、当該被験者の予後が良好か否かを数値として算出することにより行うことが、判断の客観性・汎用性・容易性などの点から好ましい。

20

【0032】

ここで、患者のデータを反映する予後予測モデルは、公知の方法を用いて患者のデータの次元を縮減し、適当なモデルに適用することにより作ることができる。すなわち、抗癌剤投与開始後の患者の生存時間が知られている大腸癌患者の原発巣切除物凍結標本を集め、これらにつき、各遺伝子の発現量を調べる。そして、かくして得られた発現量と、抗癌剤投与後の生存時間とのデータの次元を、PCA法、Partial Least Squares(PLS)法などの公知の方法により縮減し、ロジスティックモデルや比例ハザードモデルなどの公知のモデルに適用することにより生存時間の長短が各遺伝子の発現量の関数となるような予後予測モデルを構築する。これにより、患者のデータを反映した予後予測モデルを作ることができる。なお、ロジスティックモデルを用いれば、長期生存となる確率を求める予後予測モデルとして構築することができる。

30

【0033】

ここで、PLS法とは、部分的最小二乗法とも称される方法であり、主成分分析(PCA法)と並んで、解析しようとしている多次元のデータを、低次元のデータに縮約する手法であるが、PCA法は、予測変数の情報をできるだけ損失しないよう新しい総合的な変数を作成することを目的としているのに対し、PLS法は、予測変数を説明するという目標に加え、応答の変動を説明するという目標の2つが、互いのバランスを取りながら最適な変数を作成する手法である(宮下芳勝・佐々木慎一:コンピューターケミストリーシリーズ3 ケモメトリックス、pp55~71、共立出版株式会社)。

40

【0034】

また、ロジスティックモデルとは、例えば、説明変数がk個($X_1 \sim X_k$)の場合に、以下の式で表されるモデルであり、目的変数(P)の範囲を、0~1の範囲に収めるために用いることができるモデルである。ここでbは、各説明変数に対する重みを示す係数である。(浜田知久馬:学会・論文発表のための統計学、pp136~146、真興交易医書出版部)

【0035】

50

【数 1】

$$P = 1 / \{1 + \exp(-b_0 - b_1 X_1 - b_2 X_2 - \dots - b_k X_k)\}$$

【0036】

後記実施例 4 に示すとおり、Gene Spring6.1に準備されているclass predictionアルゴリズムにより作成した予後予測モデルよりもPLS法とロジスティックモデルを組み合わせで作成した予後予測モデルのほうが、より高い予測精度を実現できることが明らかになった。よって、ここで用いる予後予測モデルは、PLS法とロジスティックモデルを組み合わせで作成するのが好ましい。

10

【0037】

かくして作成される予後予測モデルに対し、予後予測をしたい患者の各遺伝子発現量を適用すれば、長期生存となる確率など、抗癌剤投与後の患者の生存時間を反映する値が計算される。かかる値を指標として、患者の予後が良好であるかどうかにつき、予測することができる。より具体的には、ロジスティックモデルを用いて予後予測モデルを作成する場合においては、例えば、計算される長期生存確率 P が $P = 0.5$ のときに、当該患者の予後が良好と判定し、 $P < 0.5$ のときには、当該患者の予後が不良と判定することによって、予測することができる。

【0038】

さらに具体的に、表1記載の16の遺伝子を全て用いる場合であって、PLS法とロジスティックモデルにより予後予測モデルを作成する場合には、後記実施例の予後予測モデル作成に使用したデータセットをもとにして算出された以下に示す式を用いた以下のような工程を行うことにより、大腸癌患者の予後予測をすることができる。すなわち、まず、予後予測をしたい患者の各遺伝子発現量をノーマライズ後、自然対数変換したのち、表6記載の各遺伝子の平均、標準偏差を用いて、中心化・尺度化した X_0 を求める。

20

【0039】

【数 2】

$$X_0 = [x_1, x_2, \dots, x_{16}]$$

30

【0040】

〔但し、ここで、式中は、それぞれ、順に、TM4SF6遺伝子、ADRM1遺伝子、ITGB4BP遺伝子、FLOT2遺伝子、PHB遺伝子、GSTM1遺伝子、BCAP31遺伝子、STIP1遺伝子、TRIM8遺伝子、MST1R遺伝子、TCF12遺伝子、PDE3B遺伝子、CCNG2遺伝子、SAT遺伝子、ALOX5遺伝子、MAN2A2遺伝子の発現量から、各遺伝子の平均、標準偏差を用いて、中心化・尺度化した値を示す〕

【0041】

次に、得られた X_0 を次に示す式に当てはめてることにより、PLS法による新たな説明変数 X_{score1} 、 X_{score2} を算出する。なお、以下の式は、変数の次元が2次元のものであるが、かかる次元数が良好であることは、cross-validation法により算出される。

40

【0042】

【数 3】

$$X_{score1} = X_0 w_1$$

$$X_{score2} = (X_0 - X_{score1} p_1) w_2$$

【0043】

〔但し、ここで、

【数 4】

$$\begin{aligned}
 w_1 &= [0.26036, 0.26874, 0.24653, 0.23596, 0.23023, 0.25175, 0.24084, 0.22662, \\
 &\quad -0.31794, -0.27827, -0.24637, -0.25469, -0.26489, -0.25081, -0.23869, -0.24471]' \\
 w_2 &= [0.08216, -0.16333, -0.30195, -0.34767, -0.29498, 0.32136, -0.21580, -0.19323, \\
 &\quad -0.00043, -0.56019, 0.03296, -0.16423, -0.07470, -0.03236, 0.20348, -0.37876]' \\
 p_1 &= [0.23630, 0.28910, 0.29293, 0.29105, 0.27592, 0.18441, 0.27174, 0.25390, \\
 &\quad -0.30675, -0.16649, -0.24376, -0.21587, -0.24202, -0.23615, -0.26742, -0.16716]'
 \end{aligned}$$

を示す。]

10

【0044】

そして、かくして求められる説明変数 X_{SCORE1} 、 X_{SCORE2} を用いて、次に示す式にあてはめることにより、長期生存確率 (P) を求める。

【0045】

【数 5】

$$P = 1 / \{1 + \exp(-b_0 - b_1 X_{score1} - b_2 X_{score2})\}$$

【0046】

〔但し、ここで、

20

【数 6】

$$b_0 = -0.28871, b_1 = 6.26918, b_2 = 4.69156$$

を示す。]

【0047】

得られた長期生存確率 (P) が、 $P \geq 0.5$ のときに、当該患者の予後が良好と判定し、 $P < 0.5$ のときには、当該患者の予後が不良と判定する。以上の工程により、この場合における抗癌剤の投与を受けた大腸癌患者の予後予測をすることができる。

【実施例】

30

【0048】

以下に実施例、比較例及び参考例を示し、本発明をさらに詳しく説明する。しかしながら、本発明はこれら実施例に制限されるものではない。

【0049】

実施例 1 転移性大腸がん症例を対象にした遺伝子発現による 5-FU/LV 治療の予後予測

1. GeneChip (登録商標) システムによる遺伝子発現量の測定

転移性大腸がん患者 38 症例の原発巣切除凍結標本より、RNeasy Mini Kit (QIAGEN) を用いて total RNA を抽出した。total RNA $8 \mu\text{g}$ を逆転写酵素とオリゴ T7-(dT)₂₄ プライマーを用いて cDNA に変換し、さらに cDNA を in vitro transcription で増幅しビオチン標識された cRNA を合成した。このビオチン標識された cRNA を 94 (35 分) の断片化バッファー内でインキュベートし断片化し、チップにハイブリダイズした。ハイブリダイズしたビオチン標識 cRNA を Affymetrix 社 Fluidics Stasion 400 を用いて Streptavidin-Phycoerythrin で染色し洗浄した。そのチップを Agilent GeneArray Scanner でスキャンし、Affymetrix 社 Micro Array Suite 5.0 を用いてシグナル強度を計算した。尚、チップとしては、8,500 以上のヒトシーケンスに対応するプローブセットが搭載されている Human Genome FOCUS Array を使用した。

40

【0050】

実施例 2 mRNA 発現量の解析

(1) 遺伝子発現量の Normalize

生物学的に意味のない負の値を引き上げるために、0.01 より小さい値を 0.01 にデータ変

50

換した。次に、チップ間のばらつきを揃え、遺伝子発現パターンによる解析を行うために、各チップごと及び各遺伝子ごとのシグナル強度の中央値で補正した。このようにして得られたNormalized値を自然対数変換により正規化し、以後の解析に使用した。

【 0 0 5 1 】

(2) Filtering

信頼性の低い遺伝子を解析から除外するために、フラグ情報が19個以上（サンプル数の半分以上）の遺伝子を解析から除外した。

【 0 0 5 2 】

(3) 相関解析

5-FU/LV治療開始後の生存時間の中央値（275.5日）を算出し、その中央値より生存時間が短い症例を短期生存者(0)、生存時間が長い症例を長期生存者(1)とした。予後（短期生存者(0)/長期生存者(1)）と各遺伝子の発現量とのSpearmanの順位相関係数を計算し、その有意確率P値が0.05未満である遺伝子を予後と有意な相関がある遺伝子とした。その結果、予後と正相関の遺伝子が182遺伝子、負相関の遺伝子が134遺伝子の合計316遺伝子が抽出された。

10

【 0 0 5 3 】

(4) 予後と関連する遺伝子

予後と関連する遺伝子（Spearmanの順位相関係数における有意確率P値が0.05未満の遺伝子）の中には、ピリミジン代謝パスウェイに関連する遺伝子として、ピリミジン代謝酵素であるDPYD、UMPS、CAD、NT5C2やRNAポリメラーゼPOLR2C、POLR2F、POLR2G、POLR3K、ロイコボリンの代謝に関連する遺伝子であるATICが含まれていた。また、予後だけでなく腫瘍縮小効果（RECIST(Response Evaluation Criteria in Solid Tumors)ガイドラインによる効果判定基準によって、PD(progressive disease)及びNC(no change)を効果なし、PR(partial response)及びCR(complete response)を効果ありとした場合における効果）に関連する遺伝子が46個含まれていた。

20

【 0 0 5 4 】

(5) 予後予測モデルの作成

予後と相関の高い上位 p^* 個（正相関の遺伝子 $p^*/2$ 個と負相関の遺伝子 $p^*/2$ 個）の遺伝子を用い、PLS法によって次元を縮約し新しい総合的な変数を作成した。次に、これらの新しい変数を用いて予後に対するロジスティックモデルを作成した。このロジスティックモデルを用いて長期生存者となる確率Pを計算し、 $P \geq 0.5$ ならば長期生存者、 $P < 0.5$ ならば短期生存者と判定する。さらに、このロジスティックモデルを用いて、leave-one-out CV法により38症例の予後を予測し、その予測精度と平均二乗誤差(MSE)を計算した。（図1）

30

【 0 0 5 5 】

(6) 予後予測遺伝子の決定

予測に使用する遺伝子数 p^* を2、4、・・・、268と変化させ、予測精度が最も高く、かつ平均二乗誤差の最も小さい時の遺伝子セットを予後予測の最適遺伝子とした。その結果、 $p^*=16$ の時、予測精度が100%で平均二乗誤差が0.0004となった（表2）。このことから、正の相関の高い上位8個と、負の相関が高い上位8個とを用いることにより、精度の高いモデルを作ることができることが明らかになった。

40

【 0 0 5 6 】

【表 2】

No of genes	No of sample	No of misclass	Correct percentage	MSE
2	38	3	92.1	0.0590
4	38	7	81.6	0.0228
6	38	6	84.2	0.0230
8	38	4	89.5	0.0039
10	38	6	84.2	0.0109
12	38	2	94.7	0.0004
14	38	2	94.7	0.0071
16	38	0	100	0.0004
18	38	1	97.4	0.0029
20	38	1	97.4	0.0073
22	38	3	92.1	0.0073
24	38	3	92.1	0.0223
26	38	3	92.1	0.0184
28	38	3	92.1	0.0077
30	38	4	89.5	0.0057
32	38	3	92.1	0.0078
34	38	4	89.5	0.0040
36	38	4	89.5	0.0034
38	38	3	92.1	0.0021
40	38	4	89.5	0.0090
42	38	2	94.7	0.0051
44	38	3	92.1	0.0119
46	38	2	94.7	0.0152
48	38	3	92.1	0.0101
50	38	4	89.5	0.0136

10

20

30

【0057】

(7) 予後予測遺伝子

上で最大の予測精度を実現する遺伝子として特定された16遺伝子は、予後にのみ関連する9個の遺伝子（BCAP31、CCNG2、ITGB4BP、PDE3B、SAT、STIP1、TCF12、TM4FSF6、TRIM8）と予後と効果の両方に関連する7個の遺伝子（ADRM1、ALOX5、FLOT2、GSTM1、MAN2A2、MST1R、PHB）であった（前記表1）。

【0058】

(8) hold-out methodによる検証

さらに、38症例のデータをランダムにtraining set（短期生存者10例、長期生存者10例、合計20例）とtesting set（短期生存者9例、長期生存者9例、合計18例）に2分割し、training setを用いてモデルを構成し、そこで得られたモデルを用いてtesting setのデータに適用しモデルの良さを検証した。100回のランダム化によって得られた予測精度は、平均が94.5%で標準偏差が7.4%であり、抽出した16遺伝子を用いた予測モデルによって、精度よく予後を予測できることが検証された。

40

【0059】

実施例 3

正負関係なく、Spearman順位相関係数の絶対値が高いものから順に遺伝子を採用する手

50

法との比較を行った。なお、これに対し、前記の実施例では、モデル構築に際して、正相関のものと負相関のものとを同数ずつ採用している。その結果、相関の強い順に採用する手法での予測精度は、16遺伝子のときに最大で97.4%にすぎなかった（図2及び表3）。このことから、予測モデルを作成する際に用いられる遺伝子は、正相関、負相関のものをそれぞれ同数ずつ採用する方法により選ぶことがより有用であることが明らかになった。

【0060】

【表3】

No of genes	No of sample	No of misclass	Correct percentage	MSE
1	38	11	71.1	0.0604
2	38	3	92.1	0.0590
3	38	5	86.8	0.0252
4	38	7	81.6	0.0228
5	38	5	86.8	0.0239
6	38	4	89.5	0.0124
7	38	2	94.7	0.0041
8	38	4	89.5	0.0039
9	38	4	89.5	0.0177
10	38	2	94.7	0.0055
11	38	3	92.1	0.0067
12	38	3	92.1	0.0032
13	38	2	94.7	0.0072
14	38	2	94.7	0.0055
15	38	2	94.7	0.0007
16	38	1	97.4	0.0009
17	38	1	97.4	0.0013
18	38	1	97.4	0.0013
19	38	2	94.7	0.0032
20	38	2	94.7	0.0014
21	38	4	89.5	0.0161
22	38	3	92.1	0.0149
23	38	2	94.7	0.0057
24	38	1	97.4	0.0104

10

20

30

40

【0061】

実施例4

PLS法の代わりに、市販のDNAマイクロアレイ・ソフトウェアGeneSpring 6.1(Silicon Genetics社製)を用いて本発明と同様のモデルを作成した。なお、この手法は、Golub氏の論文(Science, v286, pp531-537 (1999))を元にSilicon Genetics社にて改良を加えられたアルゴリズムである。GeneSpringを用いた場合、leave-one-out CV法により得られた予測精度について、最大の予測精度を実現するためには、13遺伝子を使用したときに、予測精度は最大で76.3%であった(図3及び表4)。このことから、予測モデルを作成する際に用いる方法としては、GeneSpring6.1のアルゴリズムよりもPLS法の方が優れていることが明らかになった。

50

【 0 0 6 2 】

【 表 4 】

No of genes	No of sample	No of misclass	Correct percentage
2	38	13	65.8
4	38	15	60.5
6	38	19	50.0
8	38	14	63.2
10	38	13	65.8
11	38	10	73.7
12	38	11	71.1
13	38	9	76.3
14	38	10	73.7
15	38	12	68.4
16	38	12	68.4
17	38	12	68.4
18	38	12	68.4
19	38	14	63.2
20	38	13	65.8
25	38	17	55.3
30	38	18	52.6
40	38	20	47.4
50	38	18	52.6

10

20

【 0 0 6 3 】

比較例 1

30

上記 16 遺伝子のうち、単独遺伝子を用いて作成したロジスティックモデルの予測精度を検討した。すると、上記 16 遺伝子を単独で用いて検出する場合は、以下の表 5 に示すように、65.8~76.3%の予測精度しか得られず、単独遺伝子での予後予測は予測精度が不十分であることが明らかになったのに対し、上記実施例 2 に示すように、16 の遺伝子を用いた場合には、100%という高い予測精度を実現することができることが明らかになった。

【 0 0 6 4 】

【表 5】

	予測精度 (%)
TM4SF6 遺伝子	73.7
ADRM1 遺伝子	76.3
ITGB4BP 遺伝子	71.1
FLOT2 遺伝子	71.1
PHB 遺伝子	68.4
GSTM1 遺伝子	68.4
BCAP31 遺伝子	68.4
STIP1 遺伝子	68.4
TRIM8 遺伝子	71.1
MST1R 遺伝子	68.4
TCF12 遺伝子	73.7
PDE3B 遺伝子	73.7
CCNG2 遺伝子	73.7
SAT 遺伝子	68.4
ALOX5 遺伝子	65.8
MAN2A2 遺伝子	68.4

10

20

【0065】

5. 参考例 1

上記実施例で各遺伝子の発現量を検討した38症例について、各遺伝子発現量 (Normalized値を自然対数変換した値) の平均値及び標準偏差を以下の表6に示す。以下の表6記載のデータを用いれば、発現量から患者の生存時間の長短を調べるモデルを作成することができ、また、かくして作成されるモデルの予測精度についても調べることができる。

【0066】

30

【表 6】

遺伝子名	平均値	標準偏差
TM4SF6	-0.08039	0.47295
ADRM1	0.02853	0.31154
FLOT2	0.02578	0.25697
ITGB4BP	-0.02070	0.35919
PHB	0.03435	0.40250
GSTM1	0.06743	0.39998
BCAP31	0.00325	0.37229
STIP1	0.04533	0.33520
TRIM8	-0.03301	0.33249
MST1R	-0.00123	0.29352
TCF12	0.05862	0.44798
PDE3B	-0.01900	0.42797
CCNG2	0.00899	0.40385
SAT	-0.02522	0.36424
ALOX5	0.05937	0.70460
MAN2A2	-0.05468	0.48284

10

20

【図面の簡単な説明】

【0067】

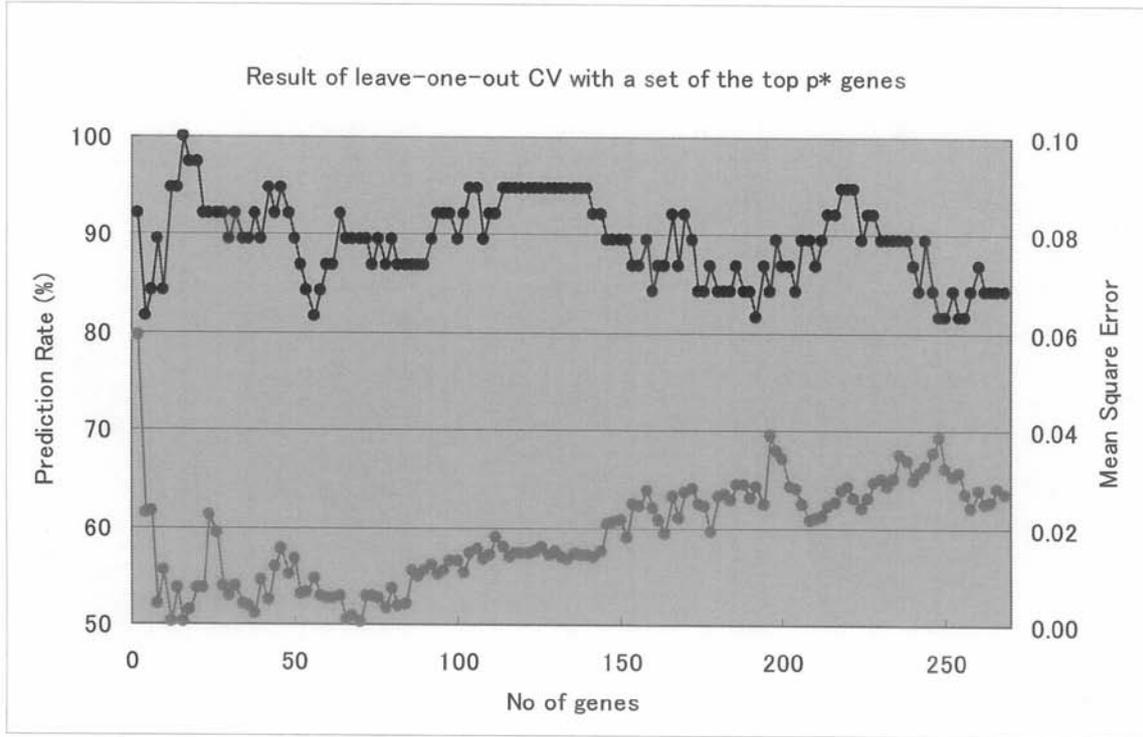
【図 1】本発明の一例として構築された予後予測モデルの予測精度および平均二乗誤差を、予後予測モデルに用いられる遺伝子数とともに示した図である。

【図 2】Gene Springなるソフトウェアを用いて構築した予後予測モデルの予測精度及び平均二乗誤差を、予後予測モデルに用いられる遺伝子数とともに示した図である。

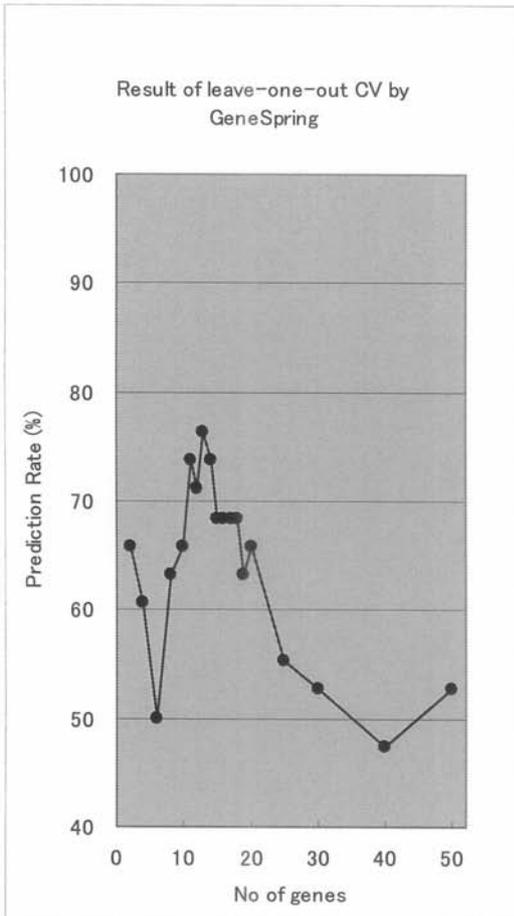
30

【図 3】相関の高いものから順に遺伝子の選択をして構築された予後予測モデルの予測精度及び平均二乗誤差を、予後予測モデルに用いられる遺伝子数とともに示した図である。

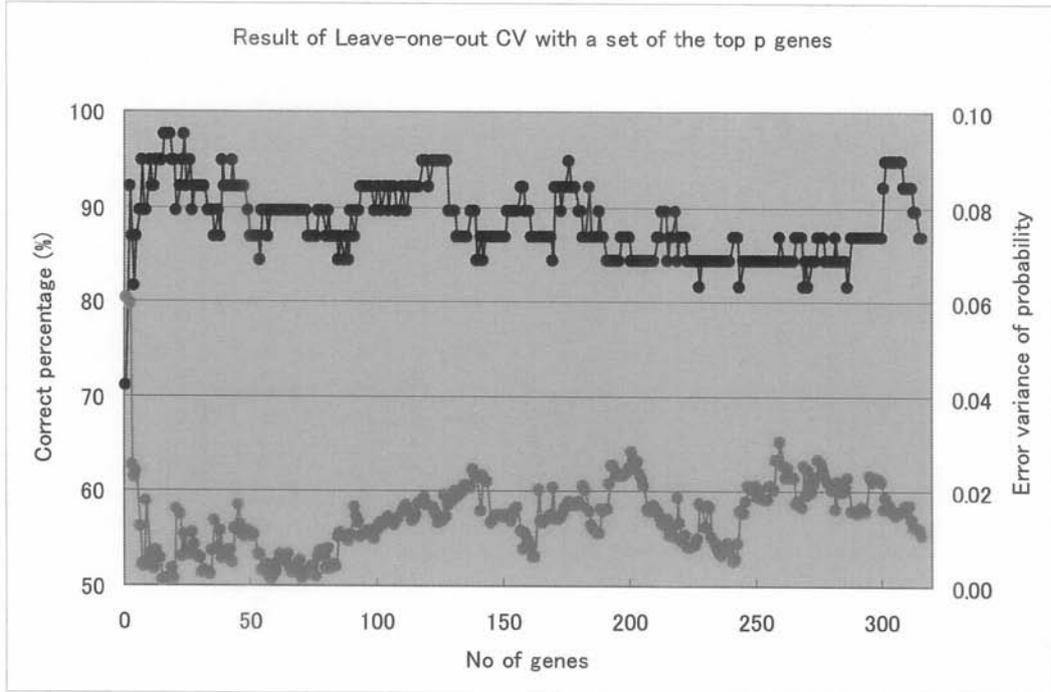
【 図 1 】



【 図 2 】



【 図 3 】



【 配列表 】

[2007037409000001.app](#)

フロントページの続き

- (74)代理人 100117156
弁理士 村田 正樹
- (74)代理人 100111028
弁理士 山本 博人
- (74)代理人 100089048
弁理士 浅野 康隆
- (74)代理人 100101317
弁理士 的場 ひろみ
- (74)代理人 100121153
弁理士 守屋 嘉高
- (74)代理人 100134935
弁理士 大野 詩木
- (74)代理人 100130683
弁理士 松田 政広
- (72)発明者 岡山 佳弘
徳島県徳島市川内町平石夷野 2 2 4 - 2 大鵬薬品工業株式会社内
- (72)発明者 市川 度
埼玉県入間郡毛呂山町毛呂本郷 3 8
- (72)発明者 平山 廉三
埼玉県入間郡毛呂山町毛呂本郷 3 8
- Fターム(参考) 4B024 AA12 CA01 HA12
4B063 QA19 QQ53 QR08 QR42 QR55 QR62 QS25 QS28 QX01