

分子プロファイルによる固形癌の分類と臨床応用への展開

谷口 一也

奈良先端科学技術大学院大学

バイオサイエンス研究科 細胞増殖学講座

(竹家 達夫 教授)

平成18年1月29日提出

1. 序論

- 1.1. 癌研究の現状
 - 1.1.1 癌関連遺伝子と癌
 - 1.1.2 癌と発現プロファイリング解析
 - 1.1.3 癌の分子標的薬
- 1.2. 本論文の構成

2. ATAC-PCR による甲状腺濾胞腺腫と濾胞癌の分類

- 2.1. 序論
- 2.2. 材料と方法
 - 2.2.1. 3'末端 cDNA ライブラリーによる遺伝子探索
 - 2.2.2. ATAC-PCR による転写物の定量
 - 2.2.3. 統計解析
- 2.3. 結果
 - 2.3.1. 甲状腺特異的な遺伝子発現情報の取得
 - 2.3.2. ATAC-PCR のデータの正規化と補正
 - 2.3.3. 濾胞腺腫と濾胞癌を分類する分類器の構築
 - 2.3.4. テストサンプルによる予測器の精度確認
 - 2.3.5. 全 85 症例を用いた解析
- 2.4. 考察

3. 二値分類器の最適な組み合わせによる遺伝子発現プロファイルからの癌サブクラス識別法

- 3.1. 序論
 - 3.1.1. 癌の遺伝子発現プロファイリングと分類器
 - 3.1.2. 学習と分類器
 - 3.1.3. 教師あり多クラス分類器
- 3.2. 材料と方法
 - 3.2.1. ATAC-PCR による甲状腺組織 168 症例の転写物の定量
 - 3.2.2. 他の多クラス分類器との性能比較に用いたデータセット
 - 3.2.3. Correlation ratio と統計解析
 - 3.2.4. 多クラス識別法の確立
- 3.3. 結果
 - 3.3.1. 組織型ごとの遺伝子発現パターン
 - 3.3.2. 各二値分類器の性能評価
 - 3.3.3. 多クラス分類器による 4 種の甲状腺組織の分類
 - 3.3.4. 他の多クラス分類器との比較
 - 3.3.5. 多クラス分類器による各クラスへの予測率

3.4. 考察

- 3.4.1. 新規多クラス分類器の特徴
- 3.4.2. 他の多クラス分類器との比較
- 3.4.3. 新規多クラス分類器の課題と展望
- 3.4.4. 癌の分類問題に関する発現プロファイリングの臨床応用

4 . *EGFR* 遺伝子変異の定量的評価によるゲフィチニブ感受性解析

4.1. 序論

4.2. 材料と方法

- 4.2.1. サンプル cDNA の調製
- 4.2.2. ダイレクトシーケンスによる変異の同定
- 4.2.3. サブクローニングによる変異 mRNA 比率の解析
- 4.2.4. 組織内の癌細胞割合の測定
- 4.2.5. 統計解析

4.3. 結果

- 4.3.1. ダイレクトシーケンスによる変異の同定
- 4.3.2. 変異 mRNA 比率とゲフィチニブの相関

4.4. 考察

- 4.4.1. *EGFR* 遺伝子変異とゲフィチニブ感受性
- 4.4.2. *EGFR* 遺伝子変異の定量解析

5 .非小細胞肺癌における *EGFR* 遺伝子変異の組織内多様性による抗癌剤感受性への影響

5.1. 序論

5.2. 材料と方法

- 5.2.1. サンプル cDNA の調製
- 5.2.2. GenomeDNA の濃度測定
- 5.2.3. 蛍光色素による蛍光強度の差異
- 5.2.4. SNaPshot による変異の定量解析
- 5.2.5. 統計解析

5.3. 結果

- 5.3.1. 微小サンプル中の遺伝子変異定量測定系の確立
- 5.3.2. *EGFR* 遺伝子変異の組織内多様性

5.4. 考察

6 . 結語

7 . 謝辞

8. 参考文献

1. 序論

本論文は、甲状腺癌と非小細胞肺癌組織を用いて遺伝子発現プロファイリングや遺伝子変異等の分子プロファイルによる分類を行い、組織型や抗癌剤感受性などの臨床指標との関連を調べた研究報告である。本章では、本研究の位置づけを明確化するために、癌に関する遺伝子変異、発現プロファイル研究、分子標的治療の現状について述べる。続いて、論文の構成を述べ、第2章以降の研究内容への導入とする。

1.1. 癌研究の現状

1.1.1 癌関連遺伝子と癌

1960年代には癌はウイルスにより引き起こされると考えられていたが、1976年に癌遺伝子 *src* (Levinson AD.et.al.1978)、1986年に癌抑制遺伝子 *Rb* (Friend SH.et.al.1986) が発見されて以来、癌は様々な原因で遺伝子に突然変異が起こり発生することが明らかとなった。それ以来、様々な癌関連遺伝子が同定され、これらの遺伝子において、点変異、欠失、挿入、重複、染色体レベルでの欠失や転座、増幅などが起こっていることが明らかとなった。加えて、これらの遺伝子の一次構造の変化のみならず、DNAのメチレーションやヒストンのアセチル化、RNAスプライシングの異常、遺伝子発現量の変化など遺伝子の構造変化を伴わない異常も発癌の重要な原因であることも明らかになった。代表的な癌抑制遺伝子である *p53* は多くの癌組織で変異が確認されている。他の癌関連遺伝子と比較しても最も多くの症例に変異を持つ *p53* であるが、変異頻度が高いとされる卵巣癌でも半分以下の症例にしか変異は確認されていない (IARC TP53 Mutation Database)。また、癌には100種類以上の癌関連遺伝子の報告があり、各組織、各個人で癌の変異は様々であることが分かっている。また、遺伝子変異により癌の特徴の一端を知ることはできるが、現状では直接治療に結びつくような分類はほとんどできていない。

1.1.2 癌と発現プロファイリング解析

2001年にヒトゲノムの全塩基配列が決定され (International Human Genome Sequencing Consortium.2001)、疾患に関連する研究にもゲノムワイドな手法が用いられるようになった。mRNAやタンパク質の発現量を一度の解析で数千から数万の遺伝子に関して比較することができる遺伝子発現プロファイル解析は代表的な手法の一つであり、特にマイクロアレイやDNAチップを用いた癌に関連する研究は数多くなされている。個々の遺伝子の機能を分子生物学的手法で解析していくことの重要性は変わらないが、癌細胞や癌組織の発現プロファイルには、細胞や組織が持つ極めて詳細な情報が含まれているため、組織の分子分類、治療標的遺伝子の探索、抗癌剤感受性、薬剤代謝等、研究のさまざまな段階で利用されている (Gerhold.et.al.2002)。その中でも、分子分類により形態学的特徴だけでは捉えることのできない癌の特性をより正確に捉えることは臨床的な意義も高く、大腸癌 (Muro S.et.al.2003,Alon U.et.al.1999,Zou TT.et.al.2002)、肺癌 (Beer DG.et.al.2002,Garber ME.et.al.2002)、乳癌 (Iwao

K.et.al.2002,Veer LJ.et.al.2002) 脳腫瘍 (Pomeroy SL.et.al.2002,Macdonald TJ.et.al.2001) など数多くの種類の癌について、分類がなされてきた。しかし、マイクロアレイ自体の再現性の問題や統計手法の誤った用いられ方から、多くの文献では分類精度を高く評価しすぎていたことが明らかとなった。そのため、遺伝子発現プロファイリングが臨床応用に至った例はない。

1.1.3 癌の分子標的薬

分子生物学の進展に伴い、癌遺伝子や癌抑制遺伝子等、癌に関連する遺伝子は正常細胞の増殖、分化、固体の発生、免疫系、神経系などにおいて細胞接着やシグナル伝達、転写因子として関わっていることが明らかになった。近年、これらの知見を生かした分子標的治療が開発されている。これまでの抗癌剤は、DNA 合成と細胞周期に対して作用するものがほとんどであった。これらの抗癌剤は癌細胞が活発に増殖する性質を用いて癌細胞に選択的な毒性を示すものであるため、癌細胞以外の分裂速度の速い細胞、すなわち骨髄中の幼若造血細胞、口腔粘膜細胞、消化管上皮細胞、毛嚢といった細胞も同時に障害を受ける。そのため、従来の抗癌剤は骨髄抑制(白血球減少、血小板減少)、消化器障害、脱毛などの副作用を比較的共通して持っていた。一方、分子標的薬は癌細胞で特異的に発現している、あるいは過剰発現している分子を標的とすることから、これまでの抗癌剤に比べて癌細胞に対してより選択的に作用することができると考えられている。これにより、従来の抗癌剤がその作用機序のために抱え続けてきた、副作用の発現を抑えることが可能になる。しかし、分子標的薬にもターゲット分子の機能抑制による副作用や、分子標的薬が想定された分子以外の分子と相互作用することによって未知の作用を示す可能性がある。これまで開発された分子標的薬としては、乳癌に対する Trastuzumab、リンパ肉腫に対する Rituximab、慢性骨髄性白血病、Gastrointestinal stromal tumor に対する Imatinib、多発性骨髄腫に対する Bortezomib があり、これらは標準的治療の一環として認識されている。また、非小細胞肺癌に対するゲフィチニブ、Erlotinib は一部の患者には高い効果を示すことが明らかとなっている。現在でも、血管新生阻害剤や、チロシンキナーゼ阻害剤は多数開発されており臨床試験が行われている。しかし、期待されていた効果を示さないばかりか、副作用により臨床に耐えない場合が多く出てきている。そこで、より患者を絞ったオーダーメイド医療の実現のために各分子標的薬の感受性試験の確立が待たれている。

1.2.本論文の構成

本論文では、遺伝子発現プロファイリングや遺伝子変異に基づく分子プロファイルにより癌の分類を行い、臨床応用への展開を考察した。特に、分子標的薬の登場により癌の分子分類は重要度が高くなっている。本章では、研究の背景にある癌の遺伝子発現プロファイルと分子標的薬の現状を述べた。

以降 2 章では、甲状腺癌の組織分類の分野で最も問題となっている甲状腺濾胞腺腫

と濾胞癌の分類を遺伝子発現プロファイリングにより試みた。本研究室独自の手法である Adaptor-tagged competitive PCR (ATAC-PCR)(Kato K.1997) により外科手術後の腫瘍サンプルから得られた total RNA を材料として発現プロファイルの情報を得て、Weighted voting アルゴリズム (Chomczynski P.et.al.1987) により分類器を構築した。また、Leave-one-out cross-validation (LOO 法) により分類器を評価し、さらに、追加症例により分類器の精度を確認した。

3 章では、甲状腺癌の組織分類を 4 種の多クラス分類へ拡張するために、新たに構築した多クラス分類器を示し、4 種の甲状腺組織の分類を試みた。また、既存の多クラス分類器の問題点を明らかにし、今回開発した新規多クラス分類器との性能比較を行った。

4 章では、肺癌の分子標的薬ゲフィチニブと *EGFR* 遺伝子変異との関連性について肺癌組織内の変異 mRNA の比率を定量的に測定し、その比率をもとに肺癌を分類しゲフィチニブ感受性に関して議論した。ゲフィチニブは非小細胞肺癌に用いられる *EGFR* 遺伝子を標的とした分子標的薬である。*EGFR* 遺伝子に変異があるとゲフィチニブが奏効することが明らかとなり、注目されているが、*EGFR* 遺伝子に変異がある症例においてもゲフィチニブ効果が見られない症例がある。そこで、変異 mRNA の比率とゲフィチニブの奏効率の関連を解析した。その結果、変異 mRNA の比率が低いとゲフィチニブの効果が低くなっているという結果が得られた。

5 章では、変異 mRNA の比率が低い原因を調べることにより、ゲフィチニブ感受性と相関がある理由を明らかにした。肺癌組織内の多領域を切り出し、*EGFR* 遺伝子変異を同定することにより、組織内の *EGFR* 遺伝子変異多様性を明らかにすることができた。つまり、*EGFR* 遺伝子変異を持つ肺癌の中に正常の *EGFR* 遺伝子のみからなる癌細胞が混在している症例が確認され、これらの症例さらなる分類を試みた。

6 章では、個別化医療の実現へ向けた研究に対する本研究の位置づけを行い、本研究の結果を今後の展望も含めて考察した。

2. ATAC-PCR による甲状腺濾胞腺腫と濾胞癌の分類

2.1. 序論

固形癌患者の治療方針の決定に直結する組織型は主に病理診断によって決定される。現在の病理組織学的な手法には多くの問題が残されており、鑑別診断が困難な場合が存在する。ゲノムワイドな遺伝子発現情報の重要な利用法の一つとなっているのが癌の病理診断である。数千から数万の遺伝子発現情報は、組織が示す特異的な表現型と関わりを持つ大量の分子生物学的マーカーとして利用することができる。

日本における甲状腺癌の有病率は 1000 人に 1 人の割合であり、そのうち乳頭癌が 85~90%、濾胞癌が 5~10%を占めている。これらの術後の予後は、他臓器癌に比べると良好で、10 年生存率は 95%程度である (Sherman SI.et.al.2003)。しかし、濾胞癌に関して、術前に良性腫瘍の濾胞腺腫と区別できないという臨床的な課題がある。現在、甲状腺癌の診断は、臨床所見、超音波診断、穿刺吸引細胞診 (FNA)、CT、シンチグラフィ等により行われている。FNA による乳頭癌の診断精度は 95%と高い精度で診断可能である (Mazzaferri EL.et.al.1993)。しかし、濾胞癌と濾胞腺腫の区別については、被膜浸潤、脈管浸潤等が診断基準となっており、術後の組織を調べなくては、診断が確定しないことが多く、診断目的で手術が行われることも多い。事実、濾胞性の腫瘍として手術が行われた場合に濾胞癌であった割合は 10~20%であった (Baloch ZW.et.al.2002, Greaves TS.et.al.2000)。このような不必要な外科手術を減少させるために新たな診断方法の確立が求められている。これに対して、分子マーカーの探索がなされ、*galectin-3* (Takano T.et.al.2003)、*hepatocyte growth factor receptor* (Ippolito A.et.al.2001)、*Pax8/peroxisome proliferator-activated receptor γ* 等の候補遺伝子が報告されているが、臨床応用には至っていない。また、近年マイクロアレイや SAGE を用いた発現プロファイル解析がなされているが (Barden CB.et.al.2003, Takano T.et.al.2004, Cerutti JM.et.al.2004)、症例数や診断精度の問題があり確立した診断法には至っていない。

そこで、本研究では、甲状腺濾胞癌と濾胞腺腫を区別する遺伝子群の同定を目的として、ヒト甲状腺癌の遺伝子発現プロファイル解析を行った。遺伝子発現プロファイルリングの解析手法としては一般に、DNA のハイブリダイゼーションを基盤技術とした DNA チップやマイクロアレイを利用した方法が広く用いられているが、再現性が低い、質の高い RNA を必要とする、発現量の低い遺伝子は検出しにくい (Holland MJ.et.al.2002) など技術的な課題が残されている。ATAC-PCR は発現量の低い遺伝子でも検出可能であり、術前の生検組織からでも解析が可能であるため、本研究では発現解析の手法として ATAC-PCR を用いて解析を行った。

2.2. 材料と方法

2.2.1. 3'末端 cDNA ライブラリーによる遺伝子探索

大阪大学腫瘍外科より提供された原発性甲状腺腫瘍(乳頭癌 5 症例、濾胞癌 4 症例、

濾胞腺腫 5 症例)、非癌部(正常甲状腺)5 症例、それぞれから TRIzol (Invitrogen) を用い、AGPC 法(Chomczynski P.et.al.1987)により total RNA を抽出した。

抽出したそれぞれの total RNA を組織型別に等量ずつ混合し、cDNA 合成に用いた。調製した total RNA 3 μ g にベクタープライマー(ベクター pUC119 を *Pst* 制限酵素処理し、ターミナルトランスフェラーゼを用いてポリ T を付加したもの) 0.5 μ g、DW を加えて合計 11 μ l とし、70 °C で 3 分間変性させた後、42 °C で 3 分間保温し、ベクタープライマーに total RNA を結合させた。次に、0.1M DTT(Invitrogen) 2 μ l、5 \times 1st.strand Buffer (Invitrogen) 4 μ l、10mM dNTP (Amersham Pharmacia Biotech) 1 μ l、RNAGuard Ribonuclease Inhibitor (33units/ μ l、Amersham Pharmacia Biotech) 1 μ l、を加え、42 °C で 2 分間保温した後、逆転写酵素 Superscript (200units/ μ l、Invitrogen) 1 μ l を加え、全量 20 μ l として 42 °C で 60 分間保温し、1st.strand cDNA を合成した。

次に、この反応液に、10 \times *E.coli* DNA ligase Buffer(Invitrogen) 16 μ l、10mM dNTP 3 μ l、0.1M DTT 6 μ l、*E.coli* DNA ligase (Invitrogen) 2 μ l、*E.coli* DNA polymerase (Invitrogen) 4 μ l、*E.coli* DNA RNase H (Invitrogen) 1 μ l、DW 98 μ l を加え、全量 130 μ l として、16 °C で 90 分反応させ、2 本鎖 cDNA を合成した。その後、70 °C で 20 分間保温し、酵素を失活させた。

合成した cDNA 40 μ l を別のチューブに移し、10 \times L Buffer(TaKaRa) 5 μ l、1M KCl 10 μ l、制限酵素 *Mbo* (TaKaRa) 1 μ l、*Bam*H (TaKaRa) 1 μ l、DW 45 μ l を加え、全量 102 μ l として 37 °C において 60 分間保温し、切断した。さらに、反応液 20 μ l を新しいチューブへ移し、10 \times *E.coli* DNA ligase Buffer 10 μ l、*E.coli* DNA ligase 2 μ l、DW 70 μ l を加え、全量 102 μ l とし、16 °C で 12 時間保温し、合成した cDNA 側の *Mbo* サイトと pUC119 ベクター側の *Bam*H サイトの連結によるセルフライゲーションを行った。

乳頭癌、濾胞癌、濾胞腺腫、正常甲状腺由来の 3'末端 cDNA ライブラリーをそれぞれ Electroporation 法により上記のライゲーション溶液 1 μ l あたり ELECTROMAX DH12S コンピテント Cells (Invitrogen) 30 μ l に形質転換を行い、SOC 液体培地 1ml を加え、37 °C で 60 分間振とう培養し、アンピシリン (50 μ g/ml) を含む LB プレートにプレーティングし、37 °C で 16 時間培養した。生育したコロニーは、乳頭癌のライブラリーからランダムに 2255 クローン、濾胞癌、濾胞腺腫、正常甲状腺のライブラリーからはそれぞれ、ランダムに 2304 クローン選択し、M13-4 プライマー (5'-GTTTTTCCCAGTCACGAC-3') を用いて、シーケンスを行った。

2.2.2 ATAC-PCR による転写物の定量

大阪大学腫瘍外科、及び隈病院(神戸市)から提供された外科切除後の原発性甲状腺腫瘍濾胞癌 22 症例、濾胞腺腫 45 症例を学習サンプルとして用いた。また、濾胞癌 6 症例、濾胞腺腫 12 症例を新たに追加しテストサンプルとして用いた。2.2.1 と同様に RNA 抽出を行い、抽出した RNA の O.D. (260nm) 値に基づいて定量した RNA 5 μ g に、ビオチン付加オリゴ d(T)₁₈ を 15pmol、を加え、DW で合計 11.5 μ l とした。その後、70 °C で 5 分間変性させた後、氷中で急冷した。次に、0.1M DTT 2 μ l、5 \times 1st.strand

Buffer 4 μ l、20mM dNTP 0.5 μ l、RNAguard Ribonuclease Inhibitor 1 μ l、を加え、42 で 2 分間保温した後、逆転写酵素 Superscript 1 μ l を加え、全量 20 μ l として、42 で 60 分間保温し、1st.strand cDNA を合成した。

次に、この反応液に、10 \times E.coli DNA ligase Buffer 10 μ l、20mM dNTP 2 μ l、0.1M DTT 2 μ l、E.coli DNA ligase 2 μ l、E.coli DNA polymerase 4 μ l、E.coli DNA RNase H 1 μ l、DW 82 μ l を加え、全量 123 μ l として、16 で 120 分反応させ、2 本鎖 cDNA を合成した。その後、70 で 20 分間保温し、酵素を失活させた。

さらに 10 \times K Buffer (TaKaRa) 35 μ l、Mbo 5 μ l、DW 187 μ l を加え、37 において 60 分間保温し、制限酵素 Mbo により cDNA を切断した。次に、フェノール・クロロホルム・イソアミルアルコール (Invitrogen) にて処理し、エタノール沈殿させた。75%エタノールにて溶液を精製し、0.1 \times TE Buffer (pH8.0) 600 μ l に溶出した。

精製した溶液 100 μ l に、10 \times T4 DNA ligase Buffer (Promega) 15 μ l、MB アダプター 30 μ l、T4 DNA ligase (3units/ μ l、Promega) 5 μ l を加え、16 で 12 時間保温して、cDNA の Mbo 切断部位 (5'-GATC-3') に MB アダプターの突出末端 (5'-GATC-3') を付加した。使用した 7 種の MB アダプターの塩基配列を図 2 - 1 に示す。

C1S-FAM (HEX or VIC) Adaptor primer 22mer

5'-6FAM[GTACATATTGTCGTTAGAACGC]-3'

MB1 S,L(23mer,27mer)

5'-GTACATATTGTCGTTAGAACGCG-3'

3'-CATGTATAACAGCAATCTTGCGCCTAG-5'

MB2 S,L(26mer,30mer)

5'-GTACATATTGTCGTTAGAACGCGACT-3'

3'-CATGTATAACAGCAATCTTGCGCTGACTAG-5'

MB3 S,L(29mer,32mer)

5'-GTACATATTGTCGTTAGAACGCGCATACT-3'

3'-CATGTATAACAGCAATCTTGCGCGTATGACTAG-5'

MB4 S,L(32mer,35mer)

5'-GTACATATTGTCGTTAGAACGCGATCCATACT-3'

3'-CATGTATAACAGCAATCTTGCGCTAGGTATGACTAG-5'

MB5 S,L(35mer,38mer)

5'-GTACATATTGTCGTTAGAACGCGTCAATCCATACT-3'

3'-CATGTATAACAGCAATCTTGCGCAGTTAGGTATGACTAG-5'

MB6 S,L(38mer,41mer)

5'-GTACATATTGTCGTTAGAACGCGTACTCAATCCATACT-3'

3'-CATGTATAACAGCAATCTTGCGCATGAGTTAGGTATGACTAG-5'

MB7 S,L(41mer,44mer)

5'-GTACATATTGTCGTTAGAACGCGCTATACTCAATCCATACT-3'

3'-CATGTATAACAGCAATCTTGCGCGATATGAGTTAGGTATGACTAG-5'

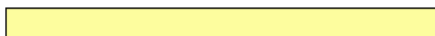
 : 共通配列

図 2 - 1: アダプタープライマー 及び、MB アダプター

MB アダプターは MB1 から MB7 までの長さの 3bp ずつ異なる合成オリゴを用い、MB1 と MB2 にコントロール由来の cDNA を、MB3 から MB7 をそれぞれサンプル由来の cDNA に付加した。また、乳頭癌 4 症例、濾胞癌 1 症例、濾胞腺腫 3 症例、正常甲状腺 4 症例の totalRNA 混合し、サンプルと同様に cDNA 合成を行い、コントロールとして用いた。7 種のアダプターを付加した cDNA 溶液 150 μ l に 5M NaCl 60 μ l を加え、MB1 から MB7 までを 10:2:5:5:5:5:5 の比になるように混合した。

この混合溶液 592 μ l に、Dynabeads M-280 Streptavidin (10mg/ml、DYNAL) を 148 μ l 加え、室温で 5 分おきに攪拌しながら 20 分間放置し、表面の avidin と oligo d (T) 末端の biotin とを共有結合させた。磁力で回収したビーズを 600 μ l の DW で洗浄し、840 μ l の DW を加え、PCR の鑄型とした。

2.2.3 PCR による cDNA 断片の増幅

10 \times PCR Buffer(Perkin Elmer Applied Biosystems)216 μ l、20mM dNTP 24 μ l、FAM(HEX or VIC) で蛍光標識したアダプタープライマー (10pmol/ μ l、Figure 1) 24 μ l、Ampli Taq Gold (Perkin Elmer Applied Biosystems) 40 μ l、DW776 μ l を混合し、このうち 1050 μ l を 2.2.2 の鑄型溶液と混合し、384 well のプレートに 4.4 μ l ずつ分注した。さらに、0.6 μ l の遺伝子特異的プライマーを加え、Gene Amp PCR System 9700 Thermal Cycler (Perkin Elmer Applied Biosystems) を用いて、PCR 反応を行った。具体的な反応条件を図 2 - 2 に示す。

1. 95 $^{\circ}$ C \times 10分
2. 94 $^{\circ}$ C \times 30秒 65 $^{\circ}$ C \times 30秒 72 $^{\circ}$ C \times 45秒 \rightarrow 5サイクル
3. 94 $^{\circ}$ C \times 30秒 60 $^{\circ}$ C \times 30秒 72 $^{\circ}$ C \times 45秒 \rightarrow 5サイクル
4. 94 $^{\circ}$ C \times 30秒 55 $^{\circ}$ C \times 30秒 72 $^{\circ}$ C \times 45秒 \rightarrow 10サイクル
5. 94 $^{\circ}$ C \times 30秒 50 $^{\circ}$ C \times 30秒 72 $^{\circ}$ C \times 45秒 \rightarrow 30サイクル
6. 72 $^{\circ}$ C \times 30分
7. 4 $^{\circ}$ C \times ∞

図 2 - 2 : ATAC-PCR に用いた PCR の条件

DNA Sequence アナライザ 3100 または 3730 (Perkin Elmer Applied Biosystems) を用い、3100 の場合は、約 600 倍、3730 の場合は約 3000 倍に DW で希釈し、PCR 産物にサイズマーカーとして、3100 の場合は GENESCAN-500ROX (3730 の場合は LIZ、Perkin Elmer Applied Biosystems) を加え、蛍光シグナルの強度とフラグメントサイズを測定した。図 2 - 3 に ATAC-PCR 法の概略を示す。

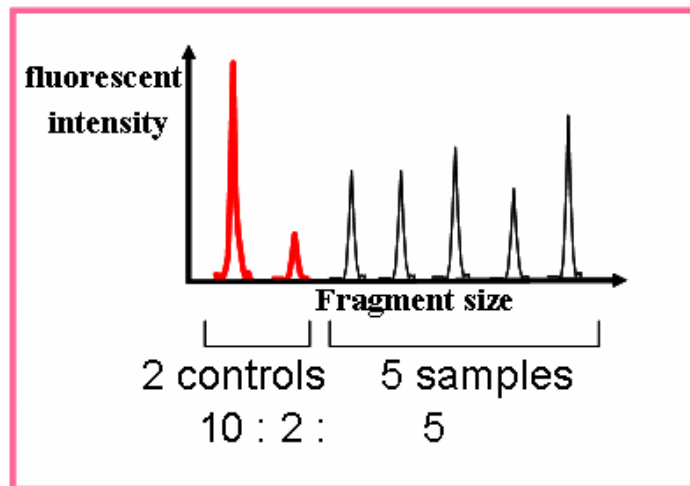
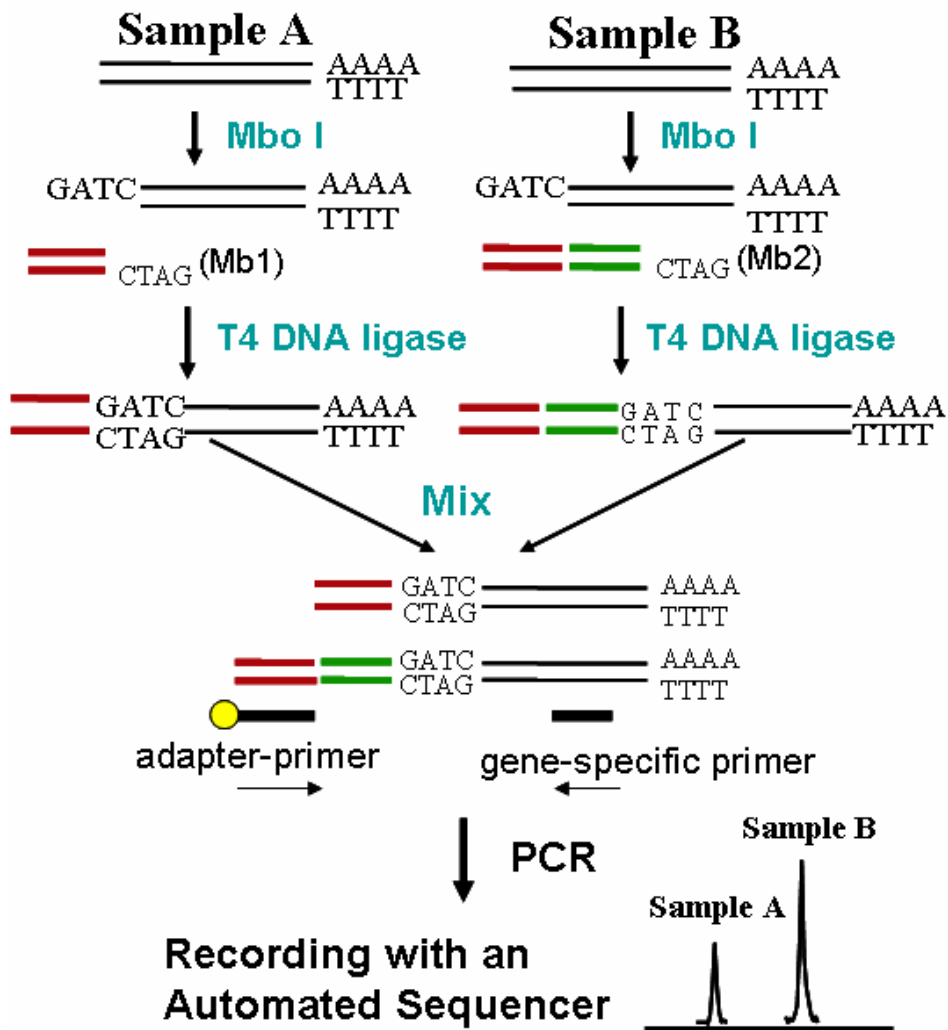


図 2 - 3 : ATAC-PCR 法の概略

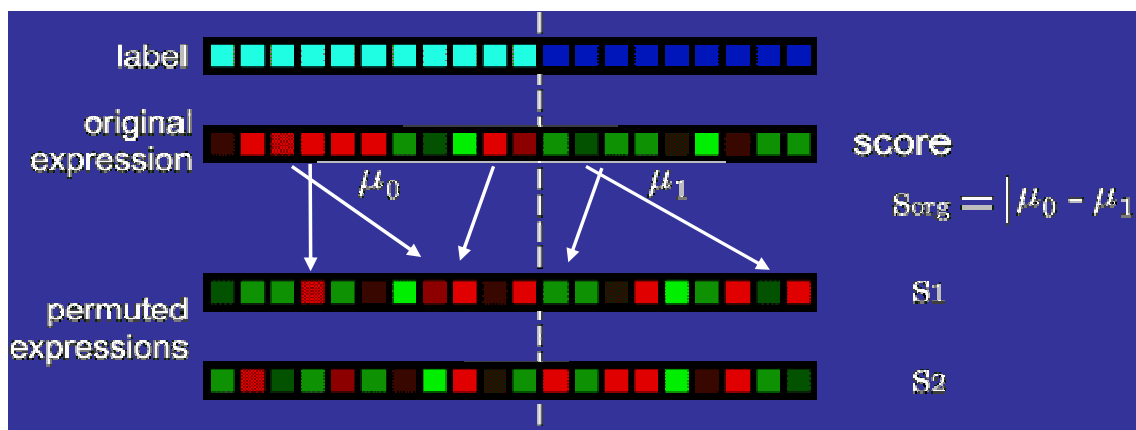
コントロールには、乳頭癌、濾胞癌、濾胞腺腫、正常甲状腺を混合したものを用い、MB1、MB2 を付加し、サンプルには MB3 から MB7 のアダプターを付加した。

2.2.3 統計解析

ATAC-PCR により得られたデータマトリクスの解析には、特徴抽出のために教師なし学習である階層的クラスタリング、及び主成分分析を用いた。また、濾胞腺腫と濾胞癌を区別する遺伝子の選択のためには、Permutation test、及び、教師あり学習である Weighted voting アルゴリズムを用いた。

今回行った、階層的クラスタリングはマイクロアレイデータの解析に一般的に用いられており、解析ソフトを利用することで、比較的簡単に解析を行うことができる。解析には、Gene Explore ver2.0 (Applied Math) を用い、クラスター分析の手法としては、Ward 法、遺伝子間の距離の算出にはユークリッド平方距離を用いた。

Permutation test は母集団の特性を規定しない、ノンパラメトリック法の一つであり、比較したい 2 群間において発現に差のある遺伝子を選択することができる。まず、2 群間の平均値の差を計算しておき、次に、無作為にサンプルを入れ替え、入れ替えた時の 2 群間の平均値の差を得る。これを 5 万回繰り返し、もとの 2 群間の平均値の差よりも差が大きい試行の割合を求める。この割合は p-value として算出され、差が大きい遺伝子ほど値が小さくなるので、濾胞腺腫と濾胞癌との間で遺伝子発現に差のある遺伝子の選択のために Permutation test を用い $P = 0.01$ 、または $P = 0.05$ の遺伝子を選択した。図 2 - 4 に Permutation test の概要を示す。



$$N_s : \text{the number of score } S_x < S_{org} \quad N_L : \text{the number of score } S_x \geq S_{org}$$

$$p \text{ value} = N_L / (N_s + N_L)$$

図 2 - 4 : Permutation test の概要

甲状腺濾胞腺腫と濾胞癌とを分類する遺伝子群の選択と予測アルゴリズムの構築のために Weighted voting アルゴリズムを用いた。まず、学習検体において、濾胞腺腫と濾胞癌とで発現の差が大きい順に Signal-to-noise ratio (SNR) を指標として順位付けした。次に、学習検体のデータより算出したある遺伝子 a に関する SNR (S_a) を重みとして用い、さらに、テストしたい検体の遺伝子 a に関する発現レベル X_a と、学習検体のうちの濾胞腺腫、濾胞癌の遺伝子 a に関するそれぞれの発現レベルの平均 μ_{1a} 、 μ_{2a} を用いて、遺伝子 a の重み V_a を算出した。 V_a は濾胞腺腫 (正) または濾胞

癌（負）へ投票することになり、正の vote の合計と負の vote の合計を合わせた総計の正負でテスト検体が濾胞腺腫であるのか、濾胞癌であるのかを判定した。最終的な総計の値は、Prediction strength（PS）として算出した。

次に、Weighted voting アルゴリズムを用いて構築したアルゴリズムを癌研究においてよく用いられる LOO 法にて評価した。この方法は、1 検体をテスト用に除き、残りの学習検体で構築したアルゴリズムを用いて、テスト検体を判定し、この操作を検体の数だけ行うものである。本研究では、濾胞腺腫と濾胞癌の計 67 症例を用いて解析を行ったため、アルゴリズムの構築と除いておいた症例の判定という操作を 67 回繰り返し行った。図 2 - 5 に SNR、図 2 - 6 に Weighted voting アルゴリズム、図 2 - 7 に LOO 法の概略を示す。

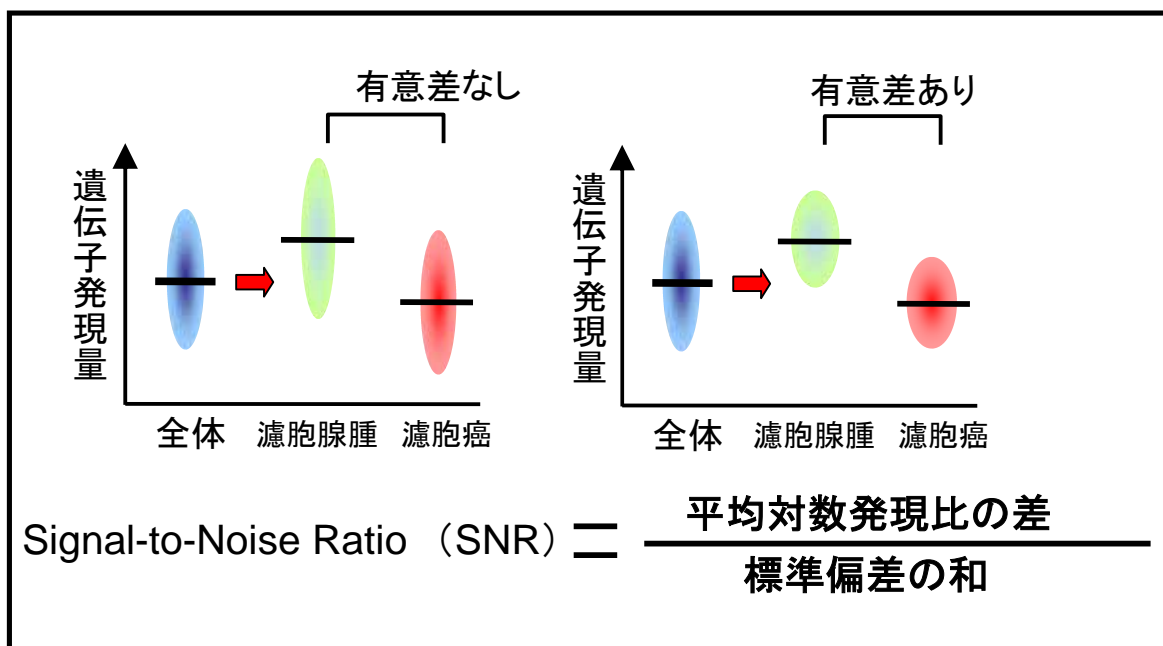


図 2 - 5 : Signal-to-noise ratio の概略

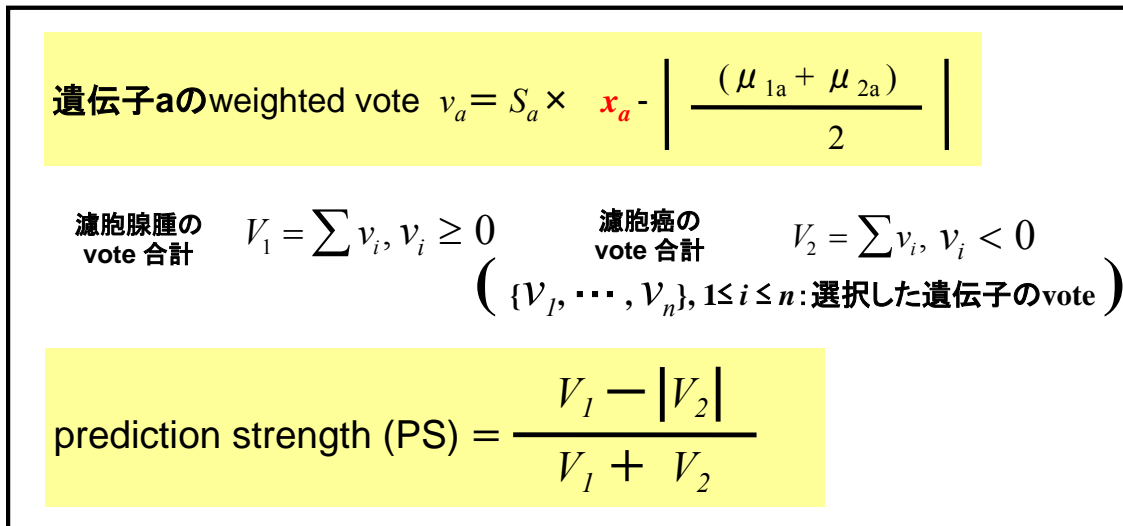


図 2 - 6 : Weighted voting アルゴリズムの概略

- Sa : 遺伝子 a に関する SNR
- Xa : 遺伝子 a の発現レベル
- μ_{1a} : 濾胞腺腫の相対発現比の平均
- μ_{2a} : 濾胞癌の相対発現比の平均

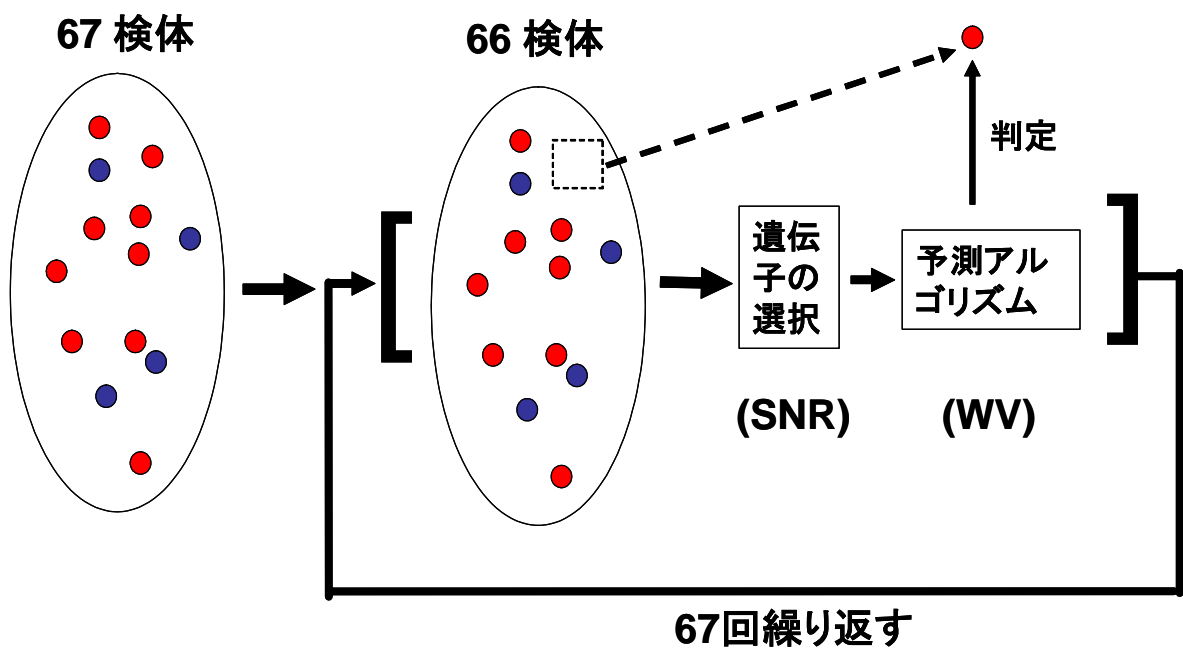


図 2 - 7 : LOO 法の概略

2.3. 結果

2.3.1 甲状腺特異的な遺伝子発現情報の取得

乳頭癌、濾胞癌、濾胞腺腫、正常甲状腺それぞれの3'末端 cDNA ライブラリーのランダムシーケンスにより、計 9157 クローンの配列情報が得られた。得られた配列情報からクローニングサイト (GATC) とポリ A を検索し、インサート cDNA の配列を選択し、不明瞭配列 (N) が 6 つ以上含まれる場合、クローニングサイト側から数えて 6 つになった時点で、5'側の配列を削除した。次に、インサートの塩基数が 20 塩基以下のもの、不明瞭配列が配列全体の 5%以上を占めるものは削除した。さらに、ミトコンドリアゲノムに相同性が見られるもの、リピート配列を除くと、乳頭癌、濾胞癌、濾胞腺腫、正常甲状腺、においてそれぞれ、1370 クローン、1556 クローン、1539 クローン、1525 クローン、となった。得られたクローン内で、相同性検索を行うとともに、Refseq、及び Genbank のデータベースに検索をかけ、アノテーション付けを行った結果 3300 遺伝子が甲状腺腫瘍、正常甲状腺に発現していることが確認された。

3'末端 cDNA ライブラリーにおいて甲状腺にて発現が確認された 3300 遺伝子のうち、得られたクローンが 1 つのみで、Refseq、及び Genbank のデータベースのどちらにも登録のない未知遺伝子を除いた、2435 遺伝子を選択した。ただし、52 遺伝子に関しては、ポリ A から MboI サイトまでの長さが 20bp 以下でプライマーを作成できなかった。次に、文献により甲状腺癌に関連のある遺伝子を 133 遺伝子追加し、計 2516 遺伝子を測定遺伝子とし、遺伝子特異的プライマーを作成した。作成にあたり、5'末端に GT 配列を付加することで、Taq ポリメラーゼによるターミナルトランスフェラーゼ活性の影響を抑えた。

2.3.2 ATAC-PCR のデータの正規化と補正

濾胞癌 22 症例、濾胞腺腫 45 症例、全 67 症例、2516 遺伝子に対する遺伝子発現情報を得た。得られた PCR 産物の PCR 強度を、Gene Mapper (Perkin Elmer Applied Biosystems) により解析した。次に、Perl のスクリプトを用い、必要なデータのみを取得し、Gene Mapper に示されたデータと照合した。

次に、コントロールの値をもとにサンプルの相対発現比を以下のようなキャリブレーションより算出した。

- ・ 10x control = 0 のときは欠測値とした。
- ・ 2x control = 0 のとき、サンプルの値が 10x control を超えるとき、 $10x\ control < 2x\ control$ のときは 10x control と原点とを結ぶ直線をキャリブレーションカーブとした。
- ・ そのほかの場合には、10x control と 2x control の中心を通る直線をキャリブレーションカーブとした。

各症例の誤差を補正するために、症例ごとに全 2516 遺伝子の発現比の中央値を求め、この値を 1 として補正を行った。次に、コントロールに対し、発現比が 20 以上

または 0.05 以下の値のものは、20 または 0.05 とした。さらに、常用対数変換を行った。その後、欠測値を 0 に置換した。また、欠測値の多い遺伝子を解析対象から外す目的で、10x control の蛍光強度の平均値の大きい上位 2000 遺伝子を選択し解析に用いた。

2.3.3 濾胞腺腫と濾胞癌を分類する分類器の構築

濾胞癌と濾胞腺腫とを分類する遺伝子群の選択と予測アルゴリズムの構築を行った。まず、SNR をもとに遺伝子を順位付けし、次に、Weighted voting アルゴリズムを用いて予測アルゴリズムを作成した。さらに、LOO 法によりその予測精度を計算したところ、60 遺伝子を用いたとき予測精度が最も高く、90.0%であった(図 2 - 8)。13 遺伝子を用いた場合にも 90.0%であったが、これは LOO 法特有の誤りであり、少なすぎる遺伝子の場合、特に accuracy curve が安定していない場合には、この遺伝子数を選択すべきではない。以上により最適遺伝子数を 60 遺伝子と決定した。図 2 - 8 では、順位の高いものから横軸で示されている数の遺伝子を用いて予測アルゴリズムを作成し、その精度を LOO 法により算出している。

次に、遺伝子の決定のために 67 症例全体を用いて SNR を計算し、上位 60 遺伝子を選択した(表 2 - 1)。また、この 60 遺伝子を用いて階層的クラスタリング分析を行った(図 2 - 9)。60 遺伝子により濾胞癌と濾胞腺腫とがよく分類されていることが分かる。また、26 遺伝子は濾胞癌で発現が高く、34 遺伝子は濾胞腺腫で発現が高くなっている。

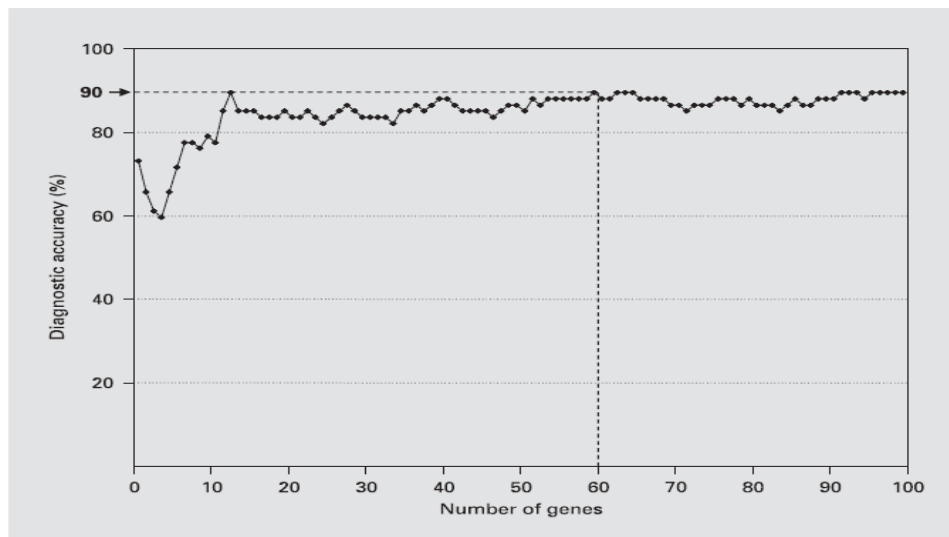


図 2 - 8 : LOO 法による予測アルゴリズムの精度

横軸に遺伝子数、縦軸に予測精度をパーセントで示した。

表 2 - 1 : SNR に基づいた発現の異なる上位 60 遺伝子

Rank	GS No.	Accession No.	SNR	p value	Symbol	Description
Upregulated in FTC						
1	GS5099	NM_022664	0.73643	0.00001	ECM1	extracellular matrix protein 1
2	GS4507	NM_031322	0.73128	0.00001	LRP4	low-density-lipoprotein receptor-related protein 4
3	LT38	NM_000674	0.61750	0.00004	ADORA	adenosine A1 receptor
4	LT96	NM_022845	0.52816	0.00004	CBFB	core-binding factor, β -subunit
5	GS4784	NM_174882	0.47105	0.00019	CRB3	crumbs homolog 3 (Drosophila)
6	GS4582	NM_033252	0.46885	0.00024	B3GNT1	β -1,3-N-acetylglucosaminyltransferase 1
7	GS15425	NM_003016	0.45887	0.00116	SFRS2	splicing factor, arginine/serine-rich 2
8	LT1	NM_053056	0.42545	0.00174	CCND1	cyclin D1
9	GS6886	No hit	0.42287	0.00180		
10	GS15657	NM_004591	0.41787	0.00262	CCL20	chemokine (C-C motif) ligand 20
11	LT36	NM_002276	0.41704	0.00274	KRT19	keratin 19
12	GS14806	NM_006765	0.39405	0.00275	TUSC3	tumor suppressor candidate 3
13	GS15543	NM_024640	0.38572	0.00279	FLJ23476	ischemia/reperfusion inducible protein
14	GS4	NM_001273	0.38388	0.00298	CHD4	chromodomain helicase DNA binding protein 4
15	GS1008	NM_002107	0.37136	0.00316	H3F3A	H3 histone, family 3A
16	GS3506	NM_032413	0.36947	0.00371	NMES1	normal mucosa of esophagus-specific 1
17	GS2562	No hit	0.36941	0.00445		
18	LT20	NM_053056	0.36723	0.00478	CCND1	cyclin D1
19	GS2646	No hit	0.36623	0.00507		
20	GS14772	NM_004494	0.36367	0.00665	HDGF	hepatoma-derived growth factor
21	LT114	NM_000368	0.36360	0.00795	TSC1	tuberous sclerosis 1
22	LT35	NM_001935	0.35698	0.00964	DPP4	dipeptidylpeptidase 4
23	GS818	No hit	0.35456	0.00974		
24	GS6094	NM_000598	0.35261	0.01114	IGFBP3	insulin-like growth factor binding protein 3
25	GS3474	NM_001235	0.35096	0.01153	SERPINH1	serine proteinase inhibitor
26	GS311	NM_001642	0.35058	0.01589	APLP2	amyloid β (A4) precursor-like protein 2
Upregulated in FTA						
1	GS9988	NM_025113	0.78044	0.00000	C13orf18	chromosome 13 open reading frame 18
2	GS6135	NM_002084	0.64222	0.00009	GPX3	glutathione peroxidase 3
3	GS3295	NM_003226	0.55029	0.00011	TFF3	TFF3 (intestinal)
4	LT92	NM_016135	0.52663	0.00020	ETV7	ets variant gene 7
5	GS3623	No hit	0.49355	0.00023		
6	GS1670	NM_001677	0.49067	0.00031	ATP1B1	ATPase, Na ⁺ /K ⁺ transporting, β ₁ polypeptide
7	GS5892	NM_016339	0.48347	0.00038	RAPGEFL1	RAP guanine-nucleotide-exchange factor -like 1
8	GS1200	NM_001268	0.48314	0.00045	CHC1L	chromosome condensation 1-like
9	GS1416	NM_001003	0.47533	0.00059	RPLP1	ribosomal protein, large, P1
10	GS14741	NM_175722	0.47269	0.00063	TPO	thyroid peroxidase
11	GS110	NM_000982	0.47108	0.00067	RPL21	ribosomal protein L21
12	GS7007	NM_021134	0.46330	0.00074	MRPL23	mitochondrial ribosomal protein L23
13	GS13671	NM_005952	0.46107	0.00093	MT1X	metallothionein 1X
14	GS2744	NM_003102	0.45617	0.00135	SOD3	superoxide dismutase 3, extracellular
15	GS598	NM_033022	0.44989	0.00166	RPS24	ribosomal protein S24
16	LT22	NM_003235	0.44383	0.00182	TG	thyroglobulin
17	GS866	NM_001019	0.43642	0.00183	RPS15A	ribosomal protein S15a
18	GS415	NM_000994	0.42773	0.00253	RPL32	ribosomal protein L32
19	GS1774	NM_001312	0.40742	0.00301	CRIP2	cysteine-rich protein 2
20	GS2700	NM_138992	0.40043	0.00309	BACE2	β -site APP-cleaving enzyme 2
21	GS6487	NM_022977	0.39065	0.00388	ACSL4	acyl-CoA synthetase long-chain family member 4
22	GS2158	NM_002415	0.39026	0.00423	MIF	macrophage migration inhibitory factor
23	GS3745	NM_005564	0.38999	0.00620	LCN2	lipocalin 2
24	GS15458	NM_005853	0.37970	0.00622	IRX5	iroquois homeobox protein 5
25	GS2131	NM_005953	0.37610	0.00667	MT2A	metallothionein 2A
26	GS14713	NM_003235	0.37250	0.00682	TG	thyroglobulin
27	GS919	NM_015921	0.36808	0.00803	C6orf82	chromosome 6 open reading frame 82
28	GS13773	NM_014366	0.36663	0.00984	NS	nucleostemin
29	GS3802	NM_018049	0.36543	0.01043	PLEKHJ1	pleckstrin homology domain
30	GS3434	NM_182471	0.36494	0.01333	PKM2	pyruvate kinase, muscle
31	GS1186	NM_005138	0.36448	0.01388	SCO	SCO cytochrome oxidase-deficient homolog 2
32	GS1736	NM_052932	0.36015	0.01449	PORIMIN	pro-oncosis receptor inducing membrane injury gene
33	GS4084	NM_002085	0.35340	0.01524	GPX4	glutathione peroxidase 4
34	GS3200	NM_199185	0.35081	0.01682	NPM1	nucleophosmin

SNR = Signal to noise ratio.

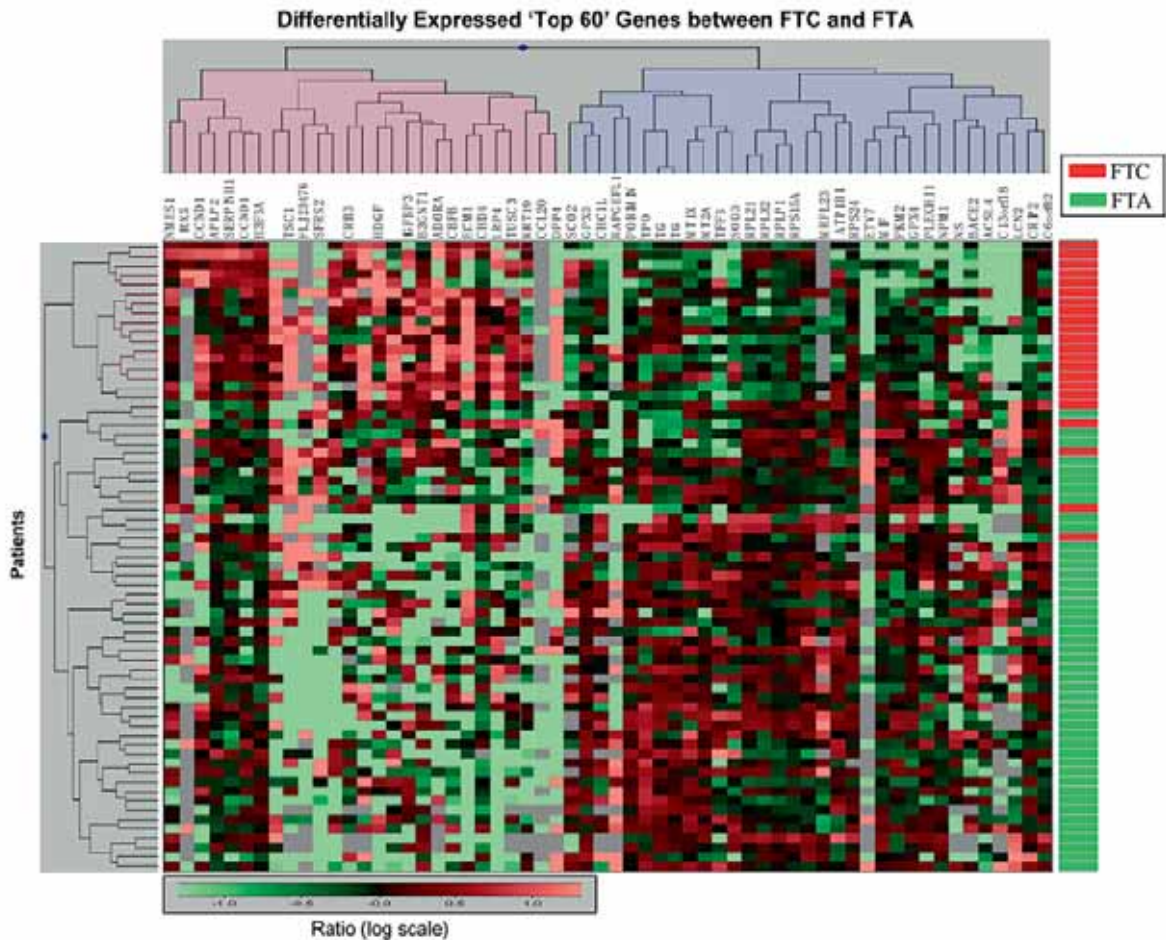


図 2 - 9 : 同定された 60 遺伝子を用いた階層的クラスタリング解析

2.3.4 テストサンプルによる予測器の精度確認

次に、テストサンプル 18 症例を用いて構築した予測器の精度を確認した。用いた遺伝子数は 60 遺伝子であり、それぞれに学習サンプルで求められた重みが付けてある。解析の結果、13 症例は濾胞腺腫に予測され ($PS \geq 0$)、5 症例は濾胞癌に予測された ($PS < 0$) (表 2 - 2)。濾胞腺腫に予測された 13 症例のうち 11 症例は FA であった (negative predict value=84.6%)。また、濾胞癌に予測された 5 症例のうち 4 症例は FC であった (positive predict value=80.0%)。特異性、感度、予測精度はそれぞれ、91.6% (11/12)、66.7% (4/6)、83.3% であった。

表 2 - 2 : テストサンプルの濾胞癌と濾胞腺腫の分類

	Histopathological diagnosis	
	FTA	FTC
Prediction strength ^a		
High (≥ 0)	11	2
Low (< 0)	1	4

2.3.5 全 85 症例を用いた解析

テストサンプルに用いた 18 サンプルに対しても 2516 遺伝子全ての発現プロファイリング情報を得て、濾胞腺腫 58 症例、濾胞癌 27 症例、計 85 症例を用いて Permutation test を行った。5 万回の Random Permutation Test の結果、濾胞癌と濾胞腺腫とで有意に発現が異なる遺伝子を $P < 0.01$ で 82 遺伝子、 $P < 0.05$ で 226 遺伝子を同定することができた。 $P < 0.01$ の 82 遺伝子を表 2 - 3 に示す。テストサンプルを用いて解析された SNR の上位 60 遺伝子のうち 29 遺伝子は $P < 0.01$ であった。

表 2 - 3 :濾胞癌と濾胞腺腫とで発現の異なる遺伝子

	p-value	Acc.No.	Symbol	Definition
1	0	NM_197955	NMES1	Normal mucosa of esophagus specific 1
2	0	NM_003226	TFF3	Trefoil factor 3 (intestinal)
3	0	NM_022664	ECM1	Extracellular matrix protein 1, transcript variant 2
4	0.0000	AB011540	LRP4	Low density lipoprotein receptor-related protein 4
5	0.0001	NM_000674	ADORA1	Adenosine A1 receptor
6	0.0001	NM_002107	H3F3A	H3 histone, family 3A
7	0.0002	NM_016322	RAB14	RAB14, member RAS oncogene family
8	0.0002	NM_006595	API5	Apoptosis inhibitor 5
9	0.0002	NM_153649	TPM3	Tropomyosin 3
10	0.0003	BC052961		cDNA clone IMAGE:4838650, partial cds
11	0.0003	NM_003016	SFRS2	Splicing factor, arginine/serine-rich 2
12	0.0003	NM_022845	CBFB	Core-binding factor, beta subunit, transcript variant 1
13	0.0005	NM_024591	FLJ11749	Hypothetical protein FLJ11749
14	0.0005	NM_001312	CRIP2	Cysteine-rich protein 2
15	0.0006	NM_014220	TM4SF1	Transmembrane 4 superfamily member 1
16	0.0006	NM_032327	ZDHHC16	Zinc finger, DHHC domain containing 16, transcript variant 1
17	0.0006	NM_174881	CRB3	Crumbs homolog 3 (Drosophila), transcript variant 3
18	0.0006	NM_053056	CCND1	Cyclin D1 (PRAD1: parathyroid adenomatosis 1)
19	0.0006	NM_002999	SDC4	Syndecan 4 (amphiglycan, ryudocan)
20	0.0007	NM_000593	TAP1	Transporter 1, ATP-binding cassette, sub-family
21	0.0007	NM_001268	CHC1L	Chromosome condensation 1-like
22	0.0007	NM_145690	YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein
23	0.0007	NM_002823	PTMA	Prothymosin, alpha (gene sequence 28)
24	0.0007	NM_024045	DDX50	DEAD (Asp-Glu-Ala-Asp) box polypeptide 50
25	0.0008	AK021966		cDNA FLJ11904 fis, clone HEMBB1000048
26	0.0009	NM_001935	DPP4	Dipeptidylpeptidase 4 (CD26, adenosine deaminase complexing protein 2)
27	0.0009	NM_032255	EPST11	Epithelial stromal interaction 1 (breast)
28	0.0009	NM_024845	FLJ14154	Hypothetical protein FLJ14154
29	0.0012	NM_001273	CHD4	Chromodomain helicase DNA binding protein 4
30	0.0014	NM_003186	TAGLN	Transgelin
31	0.0014	NM_003235	TG	Thyroglobulin
32	0.0015	NM_025113	C13orf18	Chromosome 13 open reading frame 18
33	0.0015	NM_004199	P4HA2	Procollagen-proline, 2-oxoglutarate 4-dioxygenase, alpha polypeptide II
34	0.0017	No hit		
35	0.0019	NM_005952	MT1X	Metallothionein 1X
36	0.0022	NM_014873	KIAA0205	KIAA0205 gene product
37	0.0029	NM_182931	MLL5	Myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, Drosophila)
38	0.0030	NM_000362	TIMP3	Tissue inhibitor of metalloproteinase 3
39	0.0030	NM_004999	MYO6	Myosin VI
40	0.0031	NM_000313	PROS1	Protein S (alpha)
41	0.0033	NM_015922	H105E3	NAD(P) dependent steroid dehydrogenase-like
42	0.0034	NM_148416	A2LP	Ataxin 2 related protein, transcript variant E
43	0.0034	NM_016339	Link-GEF1	Link guanine nucleotide exchange factor II
44	0.0036	NM_000295	SERPINA1	Serine (or cysteine) proteinase inhibitor, clade A member 1
45	0.0037	NM_001833	CLTA	Clathrin, light polypeptide (Lca), transcript variant nonbrain
46	0.0037	NM_004143	CITED1	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 1
47	0.0038	NM_015455	KIAA1194	KIAA1194
48	0.0039	NM_006765	N33	Putative prostate cancer tumor suppressor, transcript variant 1,
49	0.0040	NM_001642	APLP2	Amyloid beta (A4) precursor-like protein 2
50	0.0043	NM_182471	PKM2	Pyruvate kinase, muscle, transcript variant 3

	p-value	Acc.No.	Symbol	Definition
51	0.0043	NM_178148	SLC35B2	Solute carrier family 35, member B2
52	0.0045	NM_006432	NPC2	Niemann-Pick disease, type C2
53	0.0047	NM_002773	PRSS8	Prostasin preproprotein
54	0.0047	NM_000982	RPL21	Ribosomal protein L21.
55	0.0050	NM_002276	KRT19	Keratin 19 (KRT19)
56	0.0053	AK026518		cDNA: FLJ22865 fis, clone KAT02171.
57	0.0058	NM_005517	HMG2	High-mobility group nucleosomal binding domain 2
58	0.0059	NM_177543	PPAP2C	Phosphatidic acid phosphatase type 2C isoform 3
59	0.0063	No hit		
60	0.0065	NM_002467	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
61	0.0066	NM_181642	SPINT1	Serine protease inhibitor, Kunitz type 1, transcript variant 1
62	0.0067	NM_030782	CRR9	Cisplatin resistance related protein CRR9p
63	0.0072	NM_014365	H11	Protein kinase H11
64	0.0073	NM_016946	F11R	F11 receptor, transcript variant 1
65	0.0074	NM_004448	ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
66	0.0075	NM_001551	IGBP1	Immunoglobulin (CD79A) binding protein 1
67	0.0078	NM_033281	MRPS36	Mitochondrial ribosomal protein S36, nuclear gene encoding mitochondrial protein
68	0.0080	NM_004896	VPS26	Vacuolar protein sorting 26 (yeast) (VPS26), mRNA.
69	0.0082	NM_004960	FUS	Fusion (involved in t(12;16) in malignant
70	0.0086	NM_006243	PPP2R5A	alpha isoform of regulatory subunit B56
71	0.0087	NM_006330	LYPLA1	Lysophospholipase I
72	0.0088	NM_001948	DUT	dUTP pyrophosphatase
73	0.0089	NM_031450	P5326	Hypothetical protein p5326
74	0.0090	NM_005566	LDHA	Lactate dehydrogenase A
75	0.0091	NM_022977	FACL4	Fatty-acid-Coenzyme A ligase, long-chain 4, transcript variant 2
76	0.0095	BX648120		cDNA DKFZp686L1440 (from clone DKFZp686L1440)
77	0.0097	NM_006640	MSF	MLL septin-like fusion
78	0.0099	NM_016946	F11R	F11 receptor , transcript variant 1
79	0.0099	NM_002222	ITPR1	Inositol 1,4,5-triphosphate receptor, type 1
80	0.0100	NM_145792	MGST1	Microsomal glutathione S-transferase 1, transcript variant 1a
81	0.0100	No hit		
82	0.0100	NM_175719	TPO	Thyroid peroxidase, transcript variant 2

2.4 考察

甲状腺濾胞腺腫と濾胞癌を術前に分類することは、甲状腺癌において最も大きな臨床課題の一つである。非常に進行した濾胞癌の場合には、超音波診断、及び穿刺吸引細胞診の結果などをもとに確定診断されることもあるが、濾胞癌の多くは、微小浸潤型であり、甲状腺濾胞癌を疑って手術した場合、癌である割合は最大 20%である。本研究では、濾胞癌と濾胞腺腫とを区別する遺伝子の同定のみならず、ATAC-PCRを用いた診断能の可能性を検討した。

これまでに行われた cDNA マイクロアレイや SAGE を用いた遺伝子発現プロファイリングによる解析では、23 症例 (Barden CB.et.al.2003) や 17 症例 (Cerutti JM.et.al.2004) と限られたサンプルでの解析であり、テストサンプルでの解析も 6 症例 (Barden CB.et.al.2003) のみである。マイクロアレイ等数万の遺伝子から予測器を構築した場合にはテストサンプルによる検証が重要であるが 6 症例のみの解析では確立された方法とは言えない。

また、ATAC-PCR は競合 PCR を基本にした実験であり、定量的に扱うことができる幅が広い。この ATAC-PCR を用いて、本研究室では、様々な癌の分類を確立している (Iwao K.et.al.2002, Kurokawa Y.et.al.2004, Kurokawa Y.et.al.2003, Muro S.et.al.2003, Motoori M.et.al.2005)。

本研究では、22 症例の濾胞癌と 45 症例の濾胞腺腫とを用いて、両者を区別する遺

伝子群の選択と予測アルゴリズムの構築を行った。SNR をもとに遺伝子を順位付けし、Weighted voting アルゴリズム を用いて予測アルゴリズムを作成した。作成したアルゴリズムの予測精度を、LOO 法により算出したところ、60 遺伝子を用いたときの予測精度が最も高く、90.0%であった。さらに、テストサンプル 18 症例を用いて予測精度を検証したところ、83.3%の精度であった。術前の FNA ではほとんど両者を分類することができないが、この予測器を用いた ATAC-PCR による診断が可能であることが示唆された。

SNR の上位 60 遺伝子には、*Trefoil factor 3* (TFF-3) (p-value 0.0000) が含まれていた。この遺伝子は、胃の粘膜に発現するタンパク質抵抗性を示す比較的分子数の小さいペプチドである。胃癌との関連を示唆する報告はあったが、近年まで、甲状腺癌との関連を示す報告はなされていなかった。最近の cDNA チップを用いた解析で発現が抑制されているデータが示され (Barden CB.et.al.2003) また、穿刺吸引細胞診による TFF-3 と *galectin-3* の mRNA の定量により濾胞腺腫と濾胞癌を分類できるとした報告がなされている (Takano T.et.al.2004) 。本研究においての TFF-3 の発現をみると、この遺伝子単独での解析を行った場合、対数変換後の数値で -1 を基準にし、-1 よりも発現が高かった症例を濾胞腺腫、低かった症例を濾胞癌と判断したときが一番良く分類できたのだが、そのとき 8 割の症例を正しく分類できた。しかし、疑陽性、疑陰性の症例がそれぞれ 58 症例中 9 症例、27 症例中 8 症例あり、現状では、この遺伝子だけでの診断をするには危険が大きいと判断される。また、*galectin-3* は 10 年以上前から濾胞癌のマーカーとして注目されている遺伝子であるが、その診断能力は、7 割程度であるという報告が近年なされており (Takano T.et.al.2003) 補助診断としては有効である可能性があるが、この遺伝子単独での診断は難しいと考えられる。

この他、同定した遺伝子には、甲状腺癌に関連が報告されている遺伝子として、甲状腺乳頭癌を対象とした DNA チップによる解析において発現が上昇している報告のある (Huang Y.et.al.2001) *Adenosine A1 receptor* (ADORA1) *Dipeptidylpeptidase 4* (DPP4) *Procollagen-proline, 2-oxoglutarate 4-dioxygenase* (P4HA2) の発現が濾胞癌で上昇していた。DPP4 (Tanaka T.et.al.1995) はこの報告以前にすでに乳頭癌において報告がなされていたが、ADORA1、P4HA2 に関しては、この報告のみである (Huang Y.et.al.2001) 。また、*Interleukin 6* (IL6) は、免疫染色を行った実験により IL6R と伴に濾胞癌で発現が減少していることが確認されており (Ruggeri RM.et.al.2002) 本研究の結果と一致する。*Normal mucosa of esophagus specific 1* (Zhou J.et.al.2002) が、本研究でも濾胞癌で発現が下がっており、また、転写因子で急性骨髄性白血病の癌遺伝子 *Core-binding factor beta* (Hajira A.et.al.1995) の濾胞癌での発現上昇がみられた。さらに、乳癌において転移や浸潤への関連が示唆されている、*Epithelial stromal interaction 1* (Zhou J.et.al.2002) 癌抑制遺伝子 *p53* により発現が制御されているとの報告がある *Transporter 1* (Zhu K.et.al.1999) がそれぞれ濾胞癌で発現が上昇していた。

Thyroglobulin に関しては血中のサイログロブリン濃度が高いほど悪性度が高いとされることがあるが、甲状腺濾胞癌と髄様癌の混合癌において組織内の mRNA は正

常と同じか、下がっていると報告がある (Novel M.et.al.1991)。本研究での測定結果でも発現が低下しており、注目に値する。

さらに、最近報告されたメタ解析の論文では、発現プロファイリングのデータを集め、12 個の遺伝子を同定している (Groffith OL.et.al.2006)。今回、5 万回の Random Permutation Test により $p < 0.01$ 以下の遺伝子として同定した TFF3、TPO、PROS1 はこの 12 遺伝子の中に入っていた。

このように、本研究で同定した 60 個の遺伝子群の中には、先行研究において、癌との関連が報告されている遺伝子が多数含まれており、その発現パターンも矛盾のないものであった。先行研究において、報告のない遺伝子も甲状腺癌の発癌過程や悪性度等と関係している遺伝子が含まれていると大いに予想される。

本研究において、甲状腺濾胞腺腫と濾胞癌とを区別する 60 遺伝子を同定した。これらの遺伝子群が実際に濾胞腺腫と濾胞癌とを分類可能であるのかは、更なる確認実験を繰り返す必要がある。また、Permutation Test で有意差のみられた遺伝子とともに、今後、機能解析を行っていく必要がある。今回解析した遺伝子発現プロファイリングのデータは、Cancer gene expression database (CGED) に登録、公開されている (Kato K.et.al.2005)。多くの研究者と情報を共有することは重要であり、これにより研究が進展していくことが期待される。

3. 二値分類器の最適な組み合わせによる遺伝子発現プロファイルからの癌サブクラス識別法

3.1. 序論

3.1.1 癌の遺伝子発現プロファイリングと分類器

前章において、甲状腺濾胞癌と濾胞腺腫を用いた ATAC-PCR による発現プロファイリング解析を行い、術前鑑別診断の可能性を述べた。こうした遺伝子発現プロファイルを利用したアプローチは、従来の組織病理学的な癌診断法で問題となっていた病理組織学者ごとの診断結果の違いや、形態的な類似性により悪性組織と正常組織の鑑別が困難な場合などへの解決法の一つとして期待されている。一方で、これまで報告されている固形癌の分類器は濾胞癌と濾胞腺腫の鑑別のように二クラス分類の手法が多い。しかし、甲状腺悪性腫瘍の約 7 割は乳頭癌であり、実際の臨床現場では多クラス分類の手法が必要となる。多クラス分類を教師あり学習アルゴリズムを適用して行った研究報告としては、4 種類の small round blue cell tumors (SRBCTs) の人工ニューラルネットワークによる分類 (Khan J.et.al.2001)、そして、multi-class support vector machine (MC-SVM) による 14 種類の成人悪性腫瘍の診断 (Ramaswaamy S.et.al.2001) などがある。これらの研究は、異なる組織に由来する腫瘍に関しては、分類アルゴリズムを用いてよく識別することができるという事実を明らかにした。この理由は主として、ある組織の遺伝子発現プロファイルが、別の組織のプロファイルとは大きく異なることによる。しかし、例えば家族性乳癌 (Hedenfalk H.et.al.2003) など、同一組織を起源とした複数の種類の腫瘍の識別は、異なる癌の表現型間に遺伝子発現パターンの類似性があるために鑑別が難しく、決定的な手法は存在しない。遺伝子発現プロファイルを用いた鑑別診断にはこのような困難な鑑別を行うことが期待されており、より高精度な多クラスパターン分類法が求められている。

3.1.2 学習と分類器

分類器は、クラスが既知であるサンプルの遺伝子発現プロファイルに基づいて構成される。分類器を構成するプロセスを学習と呼び、学習に用いるサンプルを学習データと呼ぶ。ここで、学習データに含まれず、かつ腫瘍ラベルが既知であるサンプルを考える。このサンプルの遺伝子発現プロファイルを分類器に送り、腫瘍ラベルを予測する。真の腫瘍ラベルと分類器により予測されたラベルを比較することで、分類器の誤り率を推定することができる。分類器には多種多様なものが考えられるが、以上のようにして誤り率を評価することによって、サンプルの識別に有効な分類器を選ぶことができる。

3.1.3 教師あり多クラス分類器

教師あり多クラスパターン識別法機械学習の分野においてこれまでに研究されてきたパターン分類器には、二値分類問題と多クラス分類問題の区別なく同様に使うこ

とが可能であるものと、基本的には二値分類器専用として開発されたものがある。前者としては K 最近傍法 (Duda RO.et.al.2000)、shrunken centroids algorithm (SC) (Tibshirani R.et.al.2002)、各クラスに属する入力の分布を多次元正規分布などのパラメトリック確率モデルで表現する方法 (Oba S.et.al.2000) などがある。後者は margin classifiers と呼ばれる SVM (Scholkopf B J.et.al.1995, Vanpik V.et.al.1998)、AdaBoost (Freund Y.et.al.1996) が代表例である。特に、多クラス分類に拡張された SVM を MC-SVM と呼ぶ。これらの手法を多クラス分類に拡張する場合通常用いられるのは、各クラスについて one-versus-the-rest (1-R) の 2 クラス問題を解く分類器を M 個作り M 個で投票をして推定クラスを決定する方法 (Scholkopf B J.et.al.1995) one-versus-one (1-1) を全て網羅したうえで投票をして推定クラスを決定する方法である (Hastie T.et.al.1998, Scholkopf B J.et.al.1999)。また 1-R と 1-1 との両方を用いて結果を示している研究もある (Ramaswaamy S.et.al.2001, Tax D.et.al.2002)。これらは理論的な根拠のないヒューリスティクスながら、SVM のような二値分類器自体が強力に働く問題であれば前者のどれよりも高い性能を示すことがある。しかし、1-1 と 1-R のどちらの組合せがどういう場合に良いか、など具体的な組合せ方については不明確なままだった。例えば (Li T.et.al.2004) は、公開遺伝子発現データセットを用い、MC-SVM として 1-R、1-1 の単純投票型(引き分けはランダム選択)及び、エラー訂正符号方式(ECOC) [71]のバリエーションである random coding(Dietterich TGet.al.1995)、exhaustive coding (Allwein EL.et.al.2000) を比較し、加えて、Naïve Bayes 法、K 最近傍法、決定木 J4.8 を比較した。彼らが見せた結果によれば、ほとんどの場合で MC-SVM が最高の性能を示したが、様々な二値分類器の集合の作り方のうちでどれが良いかは問題依存であり、常に最高性能を示すアルゴリズムは無かった。また、同様の研究として(Stantnikov A.et.al.2005) では、各種公開遺伝子発現量データセットに対して、各種 MC-SVM およびそれ以外の多クラス識別法を網羅的に適用することで性能評価を行い、その結果では、Weston and Watkins の方法 (Weston J.et.al.1998)、Crammer and Singer の方法 (Crammer K.et.al.2001) および、1R による単純投票を用いた MC-SVM が多くの場合で優れた性能を示したが、どの MC-SVM アルゴリズムが良いかは問題依存であった。

本研究では、こうした多クラス識別問題に対する新たな教師あり学習法を構築し、二値分類器集合による遺伝子発現プロファイルからの 4 種の甲状腺組織の分類を試みた。

3.2. 材料と方法

3.2.1 ATAC-PCR による甲状腺組織 168 症例の転写物の定量

大阪大学腫瘍外科及び、隈病院（神戸市）から提供された外科切除後の甲状腺腫瘍及び正常甲状腺（乳頭癌、濾胞癌、濾胞腺腫、正常甲状腺の計 4 種）、計 168 症例（学習サンプル 119 症例、テストサンプル 49 症例）に対し、2516 遺伝子の相対比発現量を高精度の解析が可能な ATAC-PCR 法により計測した。濾胞癌及び濾胞腺腫は前章と同一のサンプルであり、また、RNA 抽出、cDNA 合成から転写物のシーケンサーを用いた定量までの各実験も前章と同様である。

3.2.2 他の多クラス分類器との性能比較に用いたデータセット

他の多クラス分類器と性能を比較するために、学習サンプル 119 症例の甲状腺癌、食道癌（Kato K.et.al.2005）、GCM データセット（Ramaswamy S.et.al.2001）、SRBCT データセット（The SRBST data set, Khan J.et.al.2001）を用いた。食道癌は我々のグループが ATAC-PCR で測定した 141 症例のデータである。分化度で 3 種に分類され、poorly differentiated 14 例、moderately differentiated 97 例、well differentiated 30 例である。GCM データセットは 14 種の癌を含むデータセットである。ただし、二値分類器の増加を考慮して乳癌 10 例、前立腺癌 10 例、白血病 30 例、腎癌 11 例、中皮種 11 例の 5 種の癌に絞って解析した。SRBCT データセットは幼児期に発生する腫瘍 83 症例のデータセットである。内訳は、ユーイング腫瘍 29 例、横紋筋肉腫 25 例、バーキットリンパ腫 11 例、神経芽細胞腫 18 例である。

3.2.3 Correlation ratio と統計解析

学習サンプルを用いて、Correlation ratio（CR）に基づくスペクトル解析を行い、4 種の甲状腺組織が異なった遺伝子発現パターンを示しているかを調べた。CR は以下のように定義した。

$$R_i = \sqrt{\frac{\sum_{c \in C} n_c (\bar{x}_{ci} - \bar{x}_i)^2}{\sum_j (x_{ij} - \bar{x}_i)^2}} \quad (\text{式 3 - 1})$$

また、p-value は t 統計量をもとに算出し、q-value は Dr Storey が作成したソフトを用い算出した（q-value）。

3.2.4 多クラス識別法の確立

多クラス識別法は共同研究者である奈良先端科学技術大学院大学情報科学研究科の行縄博士らとともに確立された。ここには、新たに考案した識別法の概要を述べる。

今回考案した多クラス分類器は二値分類器の統合によるものである。各二値分類器は前章のものと同様に、SNR を重み付けの指標とした Weighted voting アルゴリズム

を用いて構築した。また、このための学習サンプルは 119 症例であり、各分類器を作成する時はこの 119 症例の中から適宜必要な組織型を選別し使用した。LOO 法やテストサンプルの判定に用いる遺伝子は、各分類器において分類遺伝子 ($p < 0.0001$) を決定し用いた。そして、各分類器の出力は Weighted voting アルゴリズムの prediction strength を一次元ロジスティック回帰で $[0,1]$ にしたものをを用いた。この過程が確率的であり、各二値分類器からの出力は 0 か 1 かではなく、PS に基づいて算出される。

ここで、4 種の甲状腺組織からの識別問題を二値分類問題として扱う場合、以下の 4 種の選択方法がある (図 3 - 1)。今回はこれら全ての方法で解析を行った。

Type 1-1

1 対 1 の分類問題として扱う場合である。PC、FC、FA、N の 4 種の分類の場合には、PC-FC、PC-FA、PC-N、FC-FA、FC-N、FA-N の 6 種の分類問題を考え、これらの全ての分類結果から答えを導き出す。

Type 1-R

1 対残りのクラスに分け、扱う場合である。同様に 4 種の甲状腺腫瘍を考える場合には、PC-[FC-FA-N]、FC-[PC-FA-N]、FA-[PC-FC-N]、N-[PC-FC-FA] の 4 種の分類問題を考え、これらの全ての分類結果から答えを導き出す。

Type 1-A

[1-1]、[1-R]を含み、さらに以下の分類[1-2]を扱う場合である。PC-[FC-FA]、PC-[FC-N]、PC-[FA-N]、FC-[PC-FA]、FC-[PC-N]、FC-[FA-N]、FA-[PC-FC]、FA-[PC-N]、FA-[FC-N]、N-[PC-FC]、N-[PC-FA]、N-[FC-FA]である。

Type A-A

[1-A]を含み、さらに以下の分類[2-2]を扱う場合である。[PC-FC]-[FA-N]、[PC-FA]-[FC-N]、[PC-N]-[FC-FA]である。

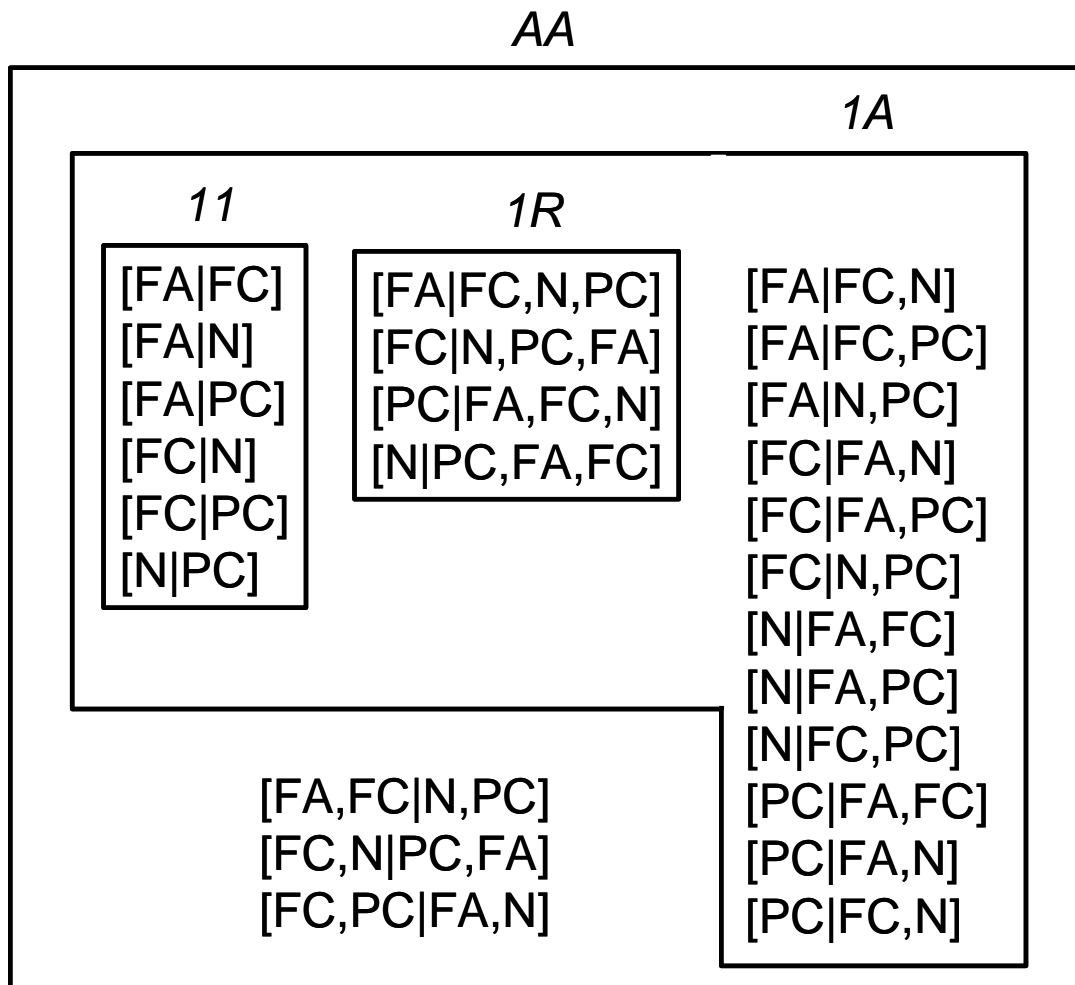


図 3 - 1 : 4 種の甲状腺癌組織分類で考えられうる二値分類

[1-A]は[1-1]、[1-R]を含み、[A-A]は[1-A]を含む全ての組み合わせである。[1-1]は 6 個、[1-R]は 4 個、[1-A]は 22 個、[A-A]は 25 個の二値分類の集合である。

次に、[1-R]や[1-2]などでは、複数のラベルを含む出力が行われた場合にその扱いをどのようにするかで、二つの方法が考えられる（図 3 - 2）。一つは Simple summation（SIS）であり、これは、複数のクラスラベルを含む出力は無視し、0 とする。もう一方は、Shared summation（SHS）であり、複数のクラスラベルを含む出力に関して均等に配分する方法である。

テストサンプル等の最終的な予測ラベルは、これまで述べてきた二値分類器の出力結果を単純に統合して求めた（図 3 - 2）。例えば、[A-A]では、25 個の二値分類器の結果を足し合わせて、一番数値の高い組織型であると予測されるのである。合計された数値は、各クラスの全合計が 1 となるように回帰させ、クラスの所属確立とした。

Simple summation (SIS)

$$\begin{aligned} q_{[1|23]} = 0.6 &\rightarrow p_1 \rightarrow p_1 + 0.6 \\ q_{[23|1]} = 0.4 &\rightarrow p_2 \rightarrow p_2 + 0 \\ &\rightarrow p_3 \rightarrow p_3 + 0 \end{aligned}$$

複数のクラスラベルを含む出力は破棄

Shared summation (SHS)

$$\begin{aligned} q_{[1|23]} = 0.6 &\rightarrow p_1 \rightarrow p_1 + 0.6 \\ q_{[23|1]} = 0.4 &\rightarrow p_2 \rightarrow p_2 + 0.2 \\ &\rightarrow p_3 \rightarrow p_3 + 0.2 \end{aligned}$$

複数のクラスラベルを含む出力は均等に配分

図 3 - 2 : Simple summation と Shared summation

3.3. 結果

3.3.1 組織型ごとの遺伝子発現パターン

まず、学習サンプルを用いて、CR に基づくスペクトル解析を行い、4 種の甲状腺組織が異なった遺伝子発現パターンを示しているかを調べた。CR は 2 クラス間の遺伝子発現の関連性を表す指標であり、高い CR 値は遺伝子発現の違いが大きいことを示す。遺伝子は CR 値が高い順に並べられ、測定された遺伝子発現に基づくオリジナルデータは、ランダムにラベルを入れ替えた (permuted) 場合と比較される。図 3 - 3a は FA と N の比較の結果を示している。オリジナルデータの CR は常に permuted の場合より高い。また、CR 全体で常に高いため、FA と N との違いは少数の遺伝子に限られたものではなく、グローバルな特徴であると考えられる。他の組織型のペアを用いた解析でも同様の結果であった。これらの結果は、4 種の組織型ごとに特有の遺伝子発現パターンを持っていることを示している。

次に、全てのペアで行った CR の結果を比較した (図 3 - 3b)。図 3 - 3b では、上位 250 遺伝子の結果を示してある。この範囲において、FA vs. N、FC vs. FA は他の比較に比べて特に低い CR 値を示していた。この結果は、医学・生物学的な知見に一致する。具体的には、FA と N は良性の組織であり、FA は正常甲状腺の生物学的な特徴が維持されているはずである。FA と FC は 1 章でも述べたように、病理組織学的には判断が難しく分類が一番困難な組織である。他の 4 つの比較はいずれも高い CR 値を示しており、上位 50 遺伝子ではどれも同様の傾向があるが、中位 (50 から 250 遺伝子) では、N vs. PC が FC vs. FA にやや近くなっていた。

CR スペクトルはグローバルな遺伝子発現の差を直接示している。さらに、この結果を p-value と q-value によって裏付けた。q-value は各遺伝子における偽陽性率を表しており、q-value が 0.01 なら 1%の偽陽性率である。p-value のカットオフを 0.001 とした時の遺伝子数と q-value は次の通りである。

- FA vs. N : 8 遺伝子、q-value 0.018
- FA vs. FC : 7 遺伝子、q-value 0.017
- N vs. PC : 44 遺伝子、q-value 0.0038
- FC vs. N : 25 遺伝子、q-value 0.0048
- PC vs. FA : 66 遺伝子、q-value 0.002
- PC vs. FC : 41 遺伝子、q-value 0.0026

これらの結果は CR 解析の結果と一致している。FA vs. N と FA vs. FC は有意差のある遺伝子が少なく q-value も高い、一方他のペアでは有意差のある遺伝子が多く q-value も低い。

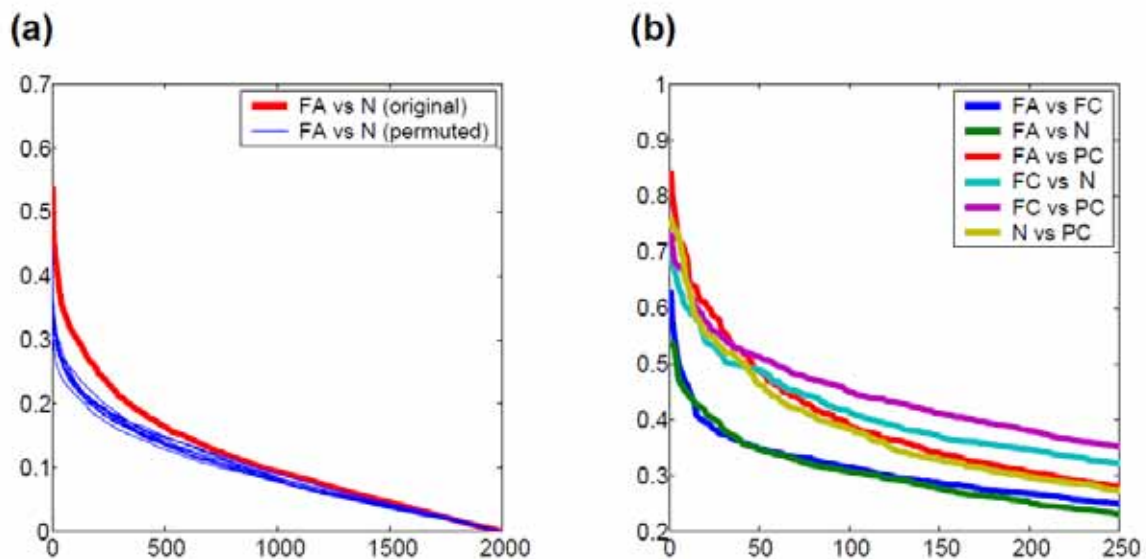


図 3 - 3 : Correlation ratio のスペクトラル解析

(a): FA と N の分類に関し CR 解析結果。赤線が測定された遺伝子発現プロファイルに基づく CR であり、青線が permuted の解析を 10 回行った際の結果を示してある。

(b): 各分類に対して CR 解析を行い、それらの結果をまとめて示してある。各分類とも permuted の結果は FA と N の解析の場合と同様であり、permuted の結果より高い CR を示した。

3.3.2 各二値分類器の性能評価

多クラス分類器で分類する前に、各二値分類器の性能を学習サンプルを用いて LOO 法により確かめた (図 3 - 4)。PC と他の組織型との分類は安定した Accuracy curve を示しており、よく分類できている。一方、FA vs. FC、FA vs. N の分類では、不安定な Accuracy curve であり、これらの結果は CR 解析や q-value の結果と同様であった。

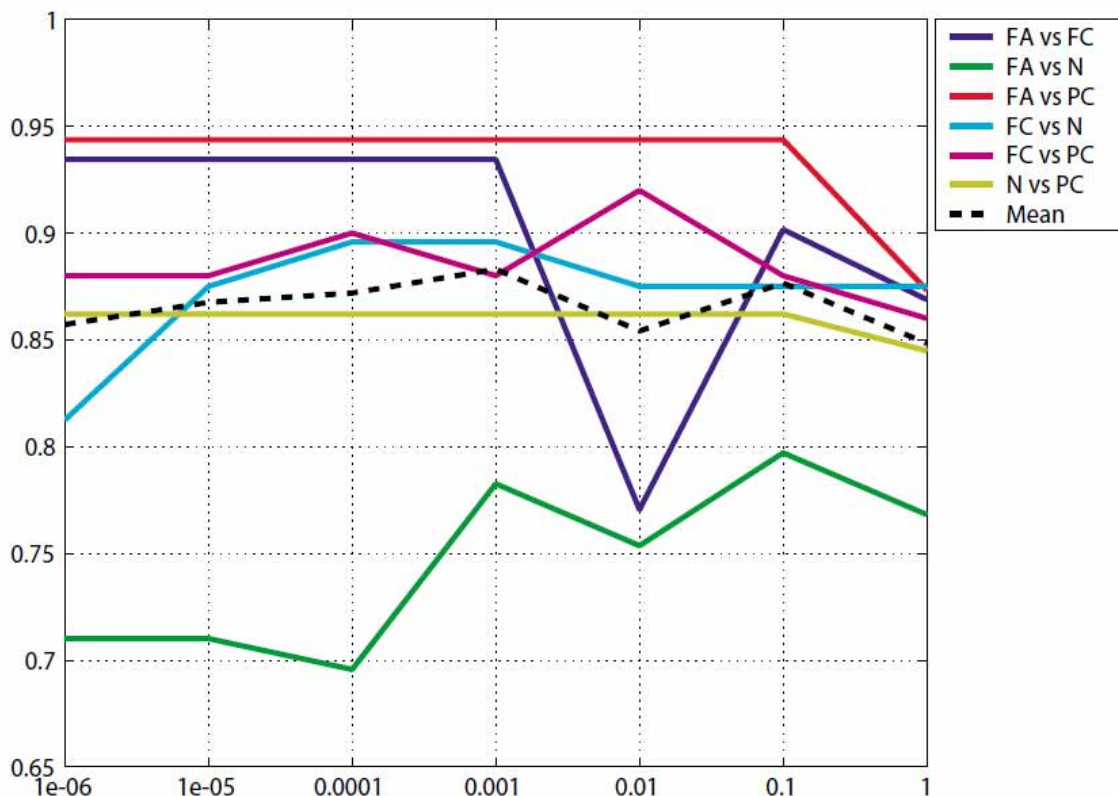


図 3 - 4 : 各二値分類の予測精度

横軸には p-value が示してあり、それぞれの p-value 以下の遺伝子を用いて予測精度を算出している。この予測精度を縦軸に示してあり、100%の予測精度を 1 とした。

3.3.3 多クラス分類器による 4 種の甲状腺組織の分類

学習サンプル 119 症例を用い、多クラス分類器を構築し、LOO 法によりその分類器を評価した。[1-A]や[A-A]の場合には、[1-1]や[1-R]の場合に比べ高い正答率であった。[1-A]-SIS、[1-A]-SHS の場合が正答率が最も高く 79.8%であった。その次は、[A-A]-SHS の 79.0%であった (表 3 - 1)。

[1-A]-SIS の詳細な解析を confusion matrix により行った (図 3 - 5 - a)。この解析より各組織型の症例がどの組織型に予測されたのかを知ることができる。FC が FA へ、FA が N へ誤って予測された症例が確認された。一方、PC の症例はほとんど正しく予測されていた。

次に、テストサンプル 49 症例を用いて予測器の精度を確認した。テストサンプルでは、学習サンプルでの性能が高かった[1-A]-SIS、[1-A]-SHS、[A-A]-SHS の 3 つに注

目すると、いずれの予測器を用いても正答率は 85.7%であり、最も精度が高かった。テストサンプルを用いた場合の[1-A]-SHS の confusion matrix を図 3 - 5 - b に示す。

表 3 - 1 : 多クラス分類器による甲状腺癌の組織の予測精度

: 学習サンプル 119 症例を用いた際のそれぞれの二値分類器の組み合わせ、及び SIS、SHS を用いた場合の予測精度

: テストサンプル 49 症例を用いた際の予測精度

① LOO-CV accuracy

	1R	1I	1A	AA
SIS	0.698	0.773	0.798	N/A
SHS	0.698	0.773	0.798	0.790

② independent test accuracy

	1R	1I	1A	AA
SIS	0.816	0.816	0.857	N/A
SHS	0.816	0.816	0.857	0.857

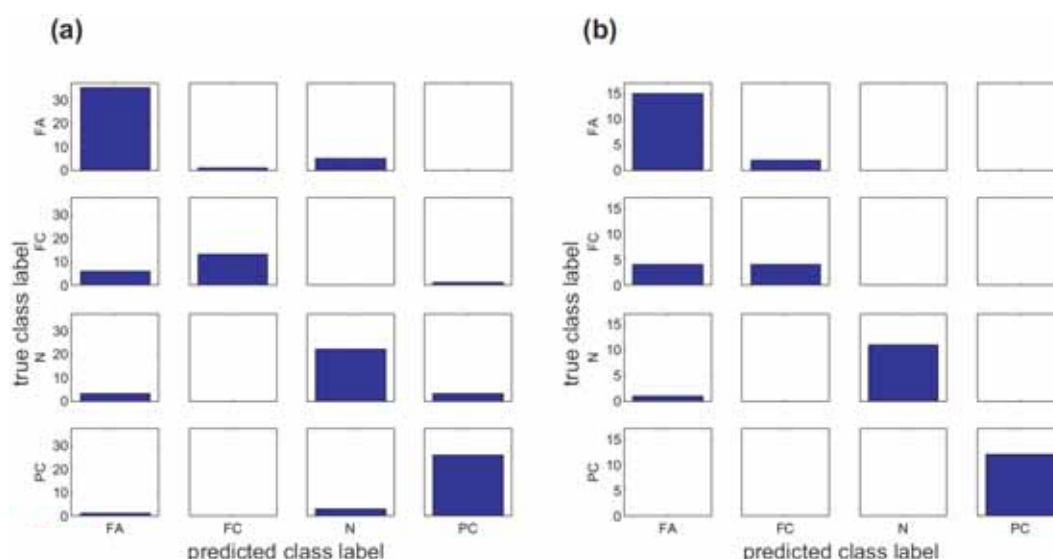


図 3 - 5 : Confusion matrix

(a): 学習サンプルを用いて[1-A]-SHS の解析を行った際の confusion matrix。横軸のラベルは予測されたクラスを示し、縦軸のラベルは真のクラスを示している。

(b): テストサンプルを用いて [1-A]-SHS の解析を行った際の confusion matrix。

3.3.4 他の多クラス分類器との比較

今回構築した二値分類器の統合による多クラス分類器作成手法を他の多クラス分類器と比較するために、学習サンプルの甲状腺癌、食道癌、GCM データセット、SRBCT データセットを用い、LOO 法により性能比較を行った。

まず、甲状腺癌の学習サンプルを用いて、分類器構築に用いる遺伝子数を p-value で 10^{-6} 、 10^{-5} 、 10^{-4} 、 10^{-3} 、 10^{-2} 、 10^{-1} 、の 6 通りの数で分類器の精度を検討した (表 3

- 2)。また、遺伝子数が少なすぎると結果が不安定になるために、最低 10 遺伝子として解析を行った。その結果、1-A-SHS/SIS が最も精度が高かった。また、1-1-SIS も同様に精度が高かったが、p-value で 1-A-SHS/SIS の時の 10 倍の遺伝子数であった。また、GCM の解析では、1-A-SHS が、食道癌の解析では、1-A-SHS/SIS が最も精度が高かった。これらのことから、[1-1]や[1-R]に比べ、多くの分類器を用いることができる[1-A]や[A-A]の方が分類精度を高くすることができることがわかった。

次に、他の分類器との比較であるが、MC-SVM と SC を用いて行った。MC-SVM を用いた解析では、遺伝子数を 2000、1024、512、256、128、64、32、16、8、4、2、1 で解析を行った。最も精度が高かった場合の正答率と遺伝子数を表 3 - 2 に示してある。いずれのデータセットを用いても新規分類法の方が高い精度を示していた。

SC を用いた解析では、shrinkage のパラメーターを 0.5 ごとに 0 から 6 まで変えて解析を行った。この場合にも、最も精度が高かった場合の正答率とパラメーターを表 3 - 2 に示した。また、SC を用いた解析でも、新規分類法の精度には及ばなかった。

表 3 - 2 : 他の学習セットを用いた解析、及び他の分類器との比較

それぞれの予測精度をパーセンテージで示した。MC-SVM の括弧内は用いた遺伝子数であり、SC 内は shrinkage のパラメーターであり、その他は遺伝子を決めるための p-value である。

	11-SIS	11-SHS	1R-SIS	1R-SHS	1A-SIS	1A-SHS	AA-SHS	MC-SVM	SC
thyroid	79.8 (10 ⁻³)	79.8 (10 ⁻³)	74.8 (10 ⁻³)	74.8 (10 ⁻³)	79.8 (10 ⁻¹)	79.8 (10 ⁻¹)	79.0 (10 ⁻¹)	74.8 (2000)	74.8 (0.5)
GCM	86.3 (10 ⁻³)	86.3 (10 ⁻³)	89.0 (10 ⁻³)	89.0 (10 ⁻³)	89.0 (10 ⁻⁴)	89.0 (10 ⁻⁴)	90.4 (10 ⁻⁴)	80.8 (32)	74.0 (0)
SRBCT	100	100	100	100	100	100	100	100 (32-2308)	100 (1-2)
esophageal	72.3 (10 ⁻¹)	72.3 (10 ⁻¹)	71.6 (10 ⁻¹)	71.6 (10 ⁻¹)	73.1 (10 ⁻¹)	73.8 (10 ⁻¹)	73.8 (10 ⁻¹)	71.7 (8)	68.8 (≥ 1)

3.3.5 多クラス分類器による各クラスへの予測率

Confusion matrix で示したとおり、新規多クラス分類器は各サンプルがどのクラスに予測されたのかを知ることができる。さらに、この分類器は最終的に予測されるクラスのみならず、各クラスへの予測率をそれぞれ知ることができる。このことにより、例えば FA と FC の中間的な性質を持つ集団が存在すればこのような新たなクラスを発見することも可能である。図 3 - 6 に甲状腺癌のそれぞれのサンプルがどのクラスにどの程度の予測率をもって予測されたのかを三次元上に示した。今回の解析では、明らかな新規集団の存在は確認できなかったが、FC のサンプルのうち FA に近い位置にプロットされていた。また、いくつかのサンプルは病理組織学的な診断とは異なるラベルへ予測されていることが分かるが、これらのサンプルは診断自体が間違っている可能性も考えられる。さらに、三次元上での距離が遠いサンプルは発現プロファイルが大きく異なっていることを示しており、クラス間の距離も知ることができた。

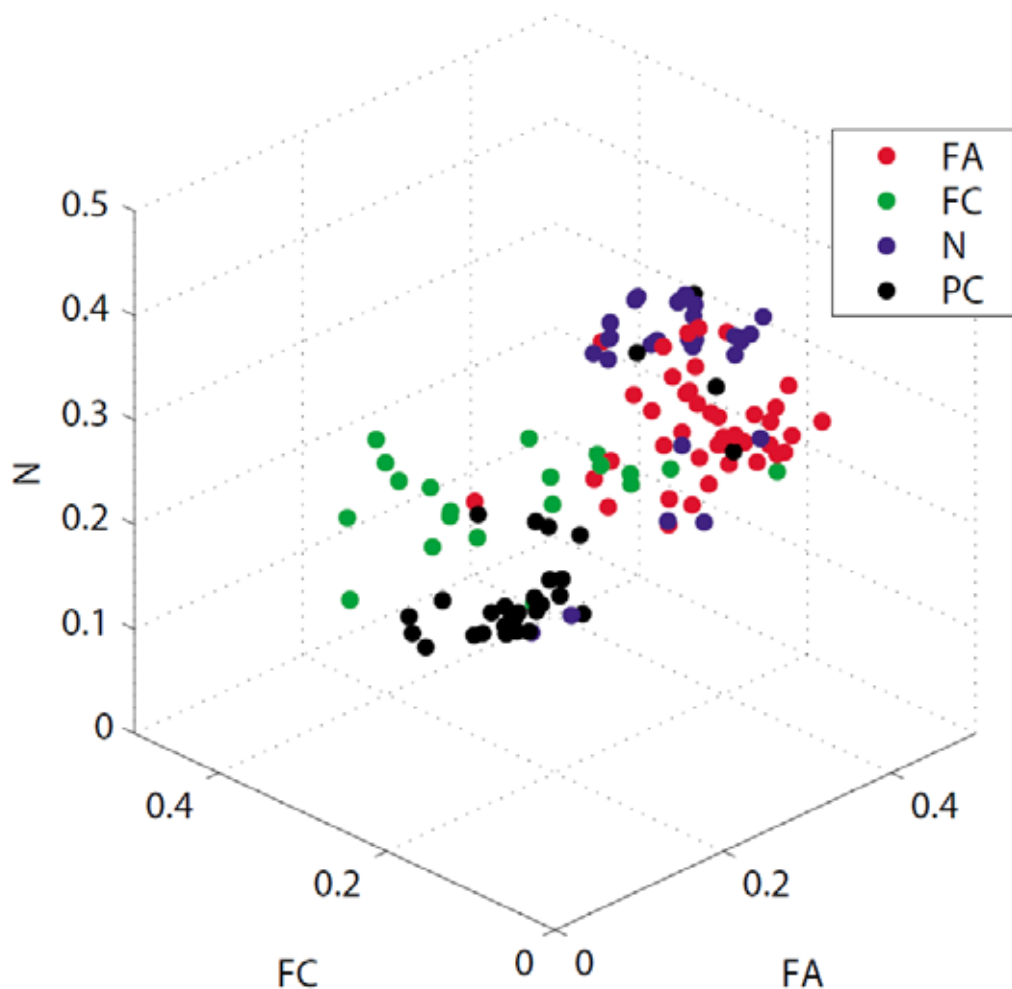


図 3 - 6 : 学習サンプルの予測率の三次元図

3.4. 考察

3.4.1 新規多クラス分類器の特徴

癌の組織診断は、WHO、各種癌学会において診断基準が設定されているにも関わらず、最終的には病理医の個人的な判断に委ねられることが多く、誤った診断結果が出たり、病理医によって異なった診断がなされたりすることも多い。この問題を解決するために、組織画像のコンピュータによる解析や遺伝子発現プロファイルを用いて診断する試みが行われている。遺伝子発現プロファイリングを用いた解析を行う際には、分類すべき組織型が多くの場合 3 種類以上であるので多クラス分類器が必要となる。本研究では、新たに構築した二値分類器集合による多クラス分類器により 4 種の甲状腺組織を 79 ~ 85% の精度で分類した。この分類器は病理組織学的な診断を補助する多クラス分類器として応用できる可能性を示した。

新規多クラス分類器では、これまでの二値分類器集合よりも二値分類器の数を増や

し、各二値分類器の予測精度の向上を図った。これまでの二値分類器集合が[1-1]や[1-R]に限られていたのに対し、我々は[1-A]や[A-A]を用いている。この場合、分類すべき組織型が多くなると新規多クラス分類器では二値分類器の数が多くなりすぎる点が問題であるが、通常癌の組織型の分類は7種類以下であるので重大な問題にはならない。

新規多クラス分類器のもう一つの特徴としては、各二値分類器の結果を確率的に統合したことである。これにより各クラスへの予測率が求められ、二値分類器では求めることができなかつた各サンプルのクラス予測率を知ることができる。また、クラス間の距離を二次元や三次元の予測図として知ることができ、主成分分析や多次元圧縮の手法よりも有用であることが示唆された。これは、単純な遺伝子発現プロファイルの情報を用いた主成分分析に比べ、分類に有用な遺伝子を用いていること、各遺伝子に重みが付けてあるためである。さらに、各サンプルの各クラスへの予測率を総合してプロットしてあるため、サンプルの予測率を詳細に知ることができる。例えば、FCのサンプルのうちいくつかのサンプルはFAに近い場所にプロットされていた。これらのサンプルは新たな組織型のグループとして区別できるものであるかもしれない。また、病理組織学的な診断が間違っており、FAの症例であるのかもしれない。これらのサンプルを病理組織学的に再検討することで、診断の信頼性を上げることが期待される。このように、新規多クラス分類器は新たな分類群を見つけ出す目的でも用いることができる。

3.4.2 他の多クラス分類器との比較

多クラス分類器は多クラス問題を二値分類問題の集合として捉える方法と二値分類手法をそのまま多クラス分類問題へ応用できる手法とがある。新規多クラス分類器は前者に分類される手法である。他の多クラス分類器の前者の代表として[1-R]を用いたMC-SVM法、後者の代表としてSC法を用いた。

MC-SVM法、SC法との比較の結果、新規多クラス分類器の性能が最も高いことが示された。また、RamaswamyらのMC-SVM法では、遺伝子数を先に決定して二値分類器を構築しているが、新規多クラス分類器の場合には遺伝子数をP値により決定し、用いた。これは、分類能のある遺伝子のみを用いて分類器を構築していることになり彼らの遺伝子選択法より優れていると考えられる。

SC法に関しては、MC-SVM法と同程度の精度で予測可能であったが、新規多クラス分類法には及ばなかつた。さらに、shrinkageのパラメーターに関して最も精度が高い場合のパラメーターを採用したため、学習サンプルにオーバーフィッティングしている可能性も考えられる。

3.4.3 新規多クラス分類器の課題と展望

新規多クラス分類器では、各クラスへの分類を二値分類器ごとに所属確率を求めてその合計により最終的な予測を行っている。しかし、各二値分類器には学習しやすい

ものやしにくいものがあるにも関わらず平等に扱われている。そこで、分類器ごとに重みを付けることにより更なる精度の向上を見込むことができ、行縄が WMAP 法を提案している (Yukinawa, 2006)。分類器ごとに重みを付けた結果、予測精度の向上を実現している。

3.4.4 癌の分類問題に関する発現プロファイリングの臨床応用

今回の解析では、異なる癌を予測した GCM データセット、SRBCT データセットを用いた解析では精度よく分類が可能であった。しかし、これらの臓器の異なる癌を発現プロファイリングを用いて診断する必然性は無く、甲状腺癌や食道癌などより困難な分類問題への応用が期待されている。本研究では、二値分類器の組み合わせによる多クラス識別器の構成のための確率モデルを提案し、それがヒューリスティクス投票法と並ぶ性能を実現できることを確認した。4種の甲状腺癌組織の分類では既存の多クラス分類器よりも高い精度で分類することができた。分類手法の進展により、組織病理学的に分類困難な組織型等の分類が可能であることを示すことが出来たといえる。発現プロファイリングのデータ取得に関する再現性の問題や更なる分類手法の進展により、組織病理学的な診断を補助できるようになることが期待される。

4 . EGFR 遺伝子変異の定量的評価によるゲフィチニブ感受性解析

4.1. 序論

肺癌は世界的に癌死の主要な原因となっており (Parkin DM.et.al.2005)、日本においても年間約 6 万人が死亡し、その数は増加傾向にある。また、手術法、化学療法等の進歩にも関わらず、肺癌患者の 5 年生存率は 10 ~ 15% とほとんど向上していない。WHO の組織分類では肺癌は分類不能癌を含め、10 種類の癌に大別している (Histological Typing of Lung and Pleural Tumours)。非小細胞肺癌は小細胞肺癌を除くすべての組織型の総称であり肺癌の約 8 割を占める。小細胞肺癌が急速な増殖と放射線化学療法への応答を特徴とするのに対し、非小細胞肺癌の放射線化学療法に奏効する例は少ない (Murren J.et.al.2001)。非小細胞肺癌には主に腺癌、扁平上皮癌、腺扁平上皮癌、大細胞癌の 4 つが含まれる。非小細胞肺癌は全肺癌のおよそ 80 ~ 85% を占めており (Devesa SS.et.al.1991)、ステージ から A の場合手術で完全に癌を取り除いても 5 割から 6 割の割合で再発し、死に至る (Scagliotti GV.et.al.2003, Arriagada R.et.al.2004)。一度再発すると予後は非常に悪く、再発後の平均生存期間は 11.5 ヶ月である (Ichinose Y.et.al.1994)。再発後の標準治療は確立しておらず、化学療法等の進展が期待されている。

一方、近年の分子生物学の進歩により、癌の発症や悪化の原因となるさまざまな変異遺伝子や変異タンパク質などが同定され、その分子を標的とした分子標的治療薬の開発が進んできた。従来の抗癌剤は主な作用点が DNA の合成阻害にあるため、正常細胞と癌細胞の両方の細胞を障害してしまう危険性があった。それに対し、分子標的治療薬では癌細胞に特異的に発現している、あるいは過剰発現している分子を標的に設計しているため、少ない副作用で高い効果を上げることが期待されていた。しかし現実には、分子標的薬にも重大な副作用がある場合があり、患者の選別が重要になってきている。

非小細胞肺癌の分子標的薬であるゲフィチニブは Epidermal growth factor receptor (EGFR, 上皮成長因子受容体) を標的としている。EGFR は Epidermal growth factor (EGF) などの増殖因子が結合する細胞外ドメインと、チロシンキナーゼ部位を有する細胞内ドメインをもつ細胞膜貫通型の受容体である。非活性化状態ではリン酸化を受けにくい一量体として存在しているが、EGF などの増殖因子が結合すると二量体を形成し、細胞内チロシンキナーゼドメインに ATP が結合することで起こる自己リン酸化を経て (Cohen S .et.al.1980)、癌増殖シグナルが核へと伝えられ、その結果、癌細胞の増殖、血管新生、浸潤及び転移の亢進や、アポトーシスの抑制などが起こる (Ciardiello F .et.al.2004)。非小細胞肺癌などの固形癌では、腫瘍細胞内の EGFR が過剰発現していることが知られており、病期の進行や転移といった癌細胞の活性化に重要な役割を果たすと考えられている (Sordella R .et.al.2004, Rush V .et.al.1993)。ゲフィチニブは、細胞内の EGFR チロシンキナーゼ部位に特異的に結合し、ATP の結合を競合的に阻害する。その結果、EGFR の自己リン酸化が抑制され、悪性化に関連す

るシグナルを遮断することにより抗腫瘍効果を発揮すると考えられている (Ciardiello F.et.al.2001, Janne PA.et.al.2004)。第 Ⅰ相臨床試験にて 10～19%の患者に対して抗腫瘍効果があることが明らかとなり (Fukuoka.et.al.2003, Kris GM.et.al.2003, Thatcher N.et.al.2005)。また、第 Ⅱ相臨床試験にて非喫煙者、女性、アジア人、腺癌の群でより効果があることが明らかとなった (Herbst RS.et.al.2004, Giaccone.et.al.2004)。その後、現在に至るまで他の抗癌剤との併用療法や大腸癌をはじめとした他の臓器での臨床試験が行われている。

ゲフィチニブは一部の患者においてのみ抗腫瘍効果を示したため、患者を区別する因子の同定が望まれていた。これに対し、2004年に非小細胞肺癌における *EGFR* 遺伝子変異の有無がゲフィチニブの抗腫瘍効果と関連するという報告がなされ (Lynch TJ.et.al.2004, Paez JG.et.al.2004)。その後多くの追試が行われた (Giaccone G.et.al.2005)。欧米からの研究報告では *EGFR* 遺伝子変異を持つ症例群におけるゲフィチニブの効果に関して生存期間の延長に関する結論は出ていないが、日本を含めアジアでは *EGFR* 遺伝子変異がゲフィチニブの効果予測因子であるという報告が多くなされており (Mitsudomi T.et.al.2005, Tokumo M.et.al.2006)。愛知県がんセンターでは効果予測の重要なマーカーとして用いられている。非小細胞肺癌における *EGFR* 遺伝子変異は、Exon18～24のチロシンキナーゼドメインに数多く存在し、日本人では全体の40%程度に確認されている (Kosaka T.et.al.2004)。また、*EGFR* 遺伝子変異は東洋人、非喫煙者、女性、腺癌に高率であり、臨床試験において効果が得られた群と一致する。しかし、*EGFR* に遺伝子変異がある症例でも約2割の症例では効果が無く、より精度の高い予測法が必要視されている。

過去に報告のある *EGFR* 遺伝子変異に関する研究では、ダイレクトシーケンス法を用いて遺伝子変異を探索し、変異の有無とゲフィチニブの抗腫瘍効果との相関を検討するものが主であった。ここで、抗腫瘍効果の指標である奏効率は腫瘍縮小量に基づく指標であり、無増悪期間はゲフィチニブ投与後に再び癌が増大するまでの期間であり、全生存期間はゲフィチニブ投与後の死亡までの期間である。本研究では、各症例の *EGFR* 遺伝子をクローニングし、複数のクローンのシーケンスから変異のある mRNA の比率を推定することにより、従来の定性的な解析に加え遺伝子変異の定量的評価を行い、ゲフィチニブに対する抗腫瘍効果との関係を検討した。

4.2. 材料と方法

4.2.1. サンプル cDNA の調製

大阪府立成人病センター提供された非小細胞肺癌 169 例 (扁平上皮癌 45 例、腺癌 111 例、大細胞癌 13 例、腺扁平上皮癌 1 例) から、TRIZOL (Invitrogen) を用いて AGPC 法 (Chomczynski P.et.al.1987) により total RNA を抽出した。

抽出した total RNA の O.D. (260nm) 値に基づいて定量した total RNA 3 μ g に Oligo d(T)₁₈ プライマー 15pmol、DW を加えて 11.5 μ l とし、70℃ で 5 分間変性させた後、氷上で冷却してプライマーと total RNA を結合させた。次に、0.1M DTT (Invitrogen)

2 μ l、5 \times 1st.strand Buffer (Invitrogen) 4 μ l、20mM dNTP (Invitrogen) 0.5 μ l、RNAguard RNase inhibitor (33units/ μ l, Amersham Bioscience) 1 μ l を加え 45 で 1 分間処理した後、逆転写酵素の Superscript (200units/ μ l, Invitrogen) 1 μ l を加え全量を 20 μ l として 42 60 分間保温し 1st strand cDNA 合成を行った。この反応液に 10 \times E.coli DNA ligase Buffer (Invitrogen) 10 μ l、20mM dNTP (Invitrogen) 2 μ l、0.1M DTT (Invitrogen) 2 μ l、E.coli DNA ligase (Invitrogen) 2 μ l、E.coli DNA RNaseH (Invitrogen) 1 μ l、E.coli polymerase (Invitrogen) 4 μ l、DW 82 μ l を加え全量を 123 μ l として 16 で 120 分間反応させ 2 本鎖 cDNA を合成した。その後 70 で 20 分間保温し、酵素を失活させた。

4.2.2. ダイレクトシーケンスによる変異の同定

cDNA を調製した 169 症例に関してシーケンスを行い、EGFR の遺伝子変異を同定した。合成した cDNA 5 μ l に EGFR チロシンキナーゼドメイン (Exon18~24 : 786bp) を網羅する範囲で設計した F-1 primer (10pmol/ μ l, 5'-CGA TCG GCC TTC ATG C-3') R-1primer(10pmol/ μ l, 5'-AAC TTT GGG CGA CTA TCT GC-3')、各 1 μ l、10 \times PCR buffer for KOD -Plus- (TAKARA) 5 μ l、2mM dNTPs (TAKARA) 5 μ l、25mM MgSO₄ (TAKARA) 2 μ l、KOD -Plus- DNA Polymerase (TAKARA) 1 μ l、DW 30 μ l を加え全量を 50 μ l とした。これを、Gene Amp PCR System 9700 Thermal Cycler (Applied Biosystems) を用いて、94 5 分で変性させた後、94 15 秒、62 30 秒、68 90 秒で 35cycle の PCR 反応を行った。反応後、Shrimp Alkaline Phosphatase (1unit/ μ l, usb corporation): Exonuclease (10unit/ μ l, usb corporation): TE buffer = 1 : 2 : 50 の割合で混合した溶液を 2.5 μ l ずつ加え 37 30 分、80 30 分処理し、酵素の失活や未反応のプライマーを除去した。次に、PCR 産物 1 μ l に、BigDye terminator v3.1 (Applied Biosystems) 0.5 μ l、5 \times sequencing buffer (Applied Biosystems) 0.75 μ l、sequence primer (1pmol/ μ l, 5'-AAC TTT GGG CGA CTA TCT GC-3') 1 μ l、DW1.75 μ l を加え全量を 5 μ l とし、96 10 秒、50 5 秒、60 2 分で 30cycle のシーケンス反応を行った。反応後、Sephadex G75 (GE ヘルスケア バイオサイエンス) を用いてカラム精製を行い、DNA Analyzer ABI 3730 (Applied Biosystems) を用いて塩基配列を決定した。

4.2.3. サブクローニングによる変異 mRNA 比率の解析

4.2.1. で調製した cDNA を用い、EGFR 遺伝子を Zero Blunt PCR cloning Kit (Invitrogen) によるサブクローニングを行い、ランダムにコロニーを 24 個または 96 個選択し、シーケンスを行った。このクローニングキットは、致死遺伝子 *ccdB* を持つ pCR-Blunt ベクターを使用した PCR 産物のクローニングシステムであり、形質転換体のみが増幅できるポジティブ選択法を使用しているため高いクローニング効率を得られる。

まず、cDNA を用いて 4.2.2. と同様に EGFR 遺伝子を PCR により増幅し、MinElute PCR Purification Kit (QIAGEN) を用いて反応溶液中に含まれる未反応のプライマー及び dNTP を除去した。次に、O.D.(260nm) 値に基づいて定量し、cDNA 溶液 150 μ g

を pCR-Blunt 50 μ g に加え、混合液と同等量(5 μ l)の DNA Ligation Kit (TAKARA) を加え 16 で 30 分間反応させ、vector DNA に連結させた。さらに、DH5 α Competent Cells (Invitrogen) 50 μ l にライゲーション反応液 2 μ l を加えヒートショック法により形質転換を行い、S.O.C Medium (Invitrogen) 450 μ l を加え、37 で 1 時間保温した。その後、Kanamycin (30 μ l/ml, SIGMA) を含む LB 寒天培地(1%Tryptone Peptone [DIFCO]、0.5%Yast Extract [DIFCO]、1%NaCl[ナカライテスク]、1.1%Apar Powder[和光純薬]) に 100 μ l ずつ撒き、37 で 12 時間培養した。

大腸菌が生育した LB 寒天プレート培地からコロニーを無作為に 24 個、または 96 個選択し、Kanamycin (30 μ l/ml, SIGMA) を加えた LB 液体培地 100 μ l に植菌し 37 で 10 時間培養した。培養後の上清 0.6 μ l を用いて 4.2.2.と同様の条件で 30cycle の PCR 反応を行い、インサート cDNA を増幅した。その後、4.2.2.と同様にシーケンスを行った。得られたデータは Clustal W を用いて、マルチプルアラインメントを行い正常配列 (NM_005228.3, National Center for Biotechnology Information) と比較した。また、J-SNP (A database of Japanese Single Nucleotide Polymorphisms) に登録のある変異は一塩基多型 (SNP) と判断し、後の解析から除去した。

4.2.4. 組織内の癌細胞割合の測定

癌組織中の癌細胞の割合を調べるために、ヘマトキシリン・エオジン染色 (HE 染色) した切片を用い、顕微鏡上で癌細胞と非癌細胞の数を測定した。症例は変異 mRNA 比率が高い群から 5 症例、低い群から 5 症例を無作為に選択し用いた。

4.2.5. 統計解析

各臨床所見で分類した場合の相関を検定する目的で、 χ 二乗検定、fisher's 正確確率検定を用いた。また、全生存期間と無増悪期間の 2 群間の比較のために Kaplan-Meier 法による生存曲線解析、及び Log-rank 検定を行った。以上の解析は統計解析ソフト SPSS 11.5J for windows を用いて行った。

4.3. 結果

4.3.1 ダイレクトシーケンスによる変異の同定

研究対象とした非小細胞肺癌 169 症例のダイレクトシーケンスにより、64 症例 (37.5 %) に *EGFR* チロシンキナーゼドメインの exon18 ~ 21 の範囲で変異が同定された。また、そのうち 7 症例には 2 個の変異が同定された。肺癌の組織型別に見ると、腺癌 111 症例中 62 症例 (56%)、扁平上皮癌 45 症例中 1 症例 (2%)、腺扁平上皮癌 1 症例中 1 症例 (100%) であったが、大細胞癌 13 症例には変異は認められなかった。詳細を図 4 - 1、表 4 - 1 に示す。

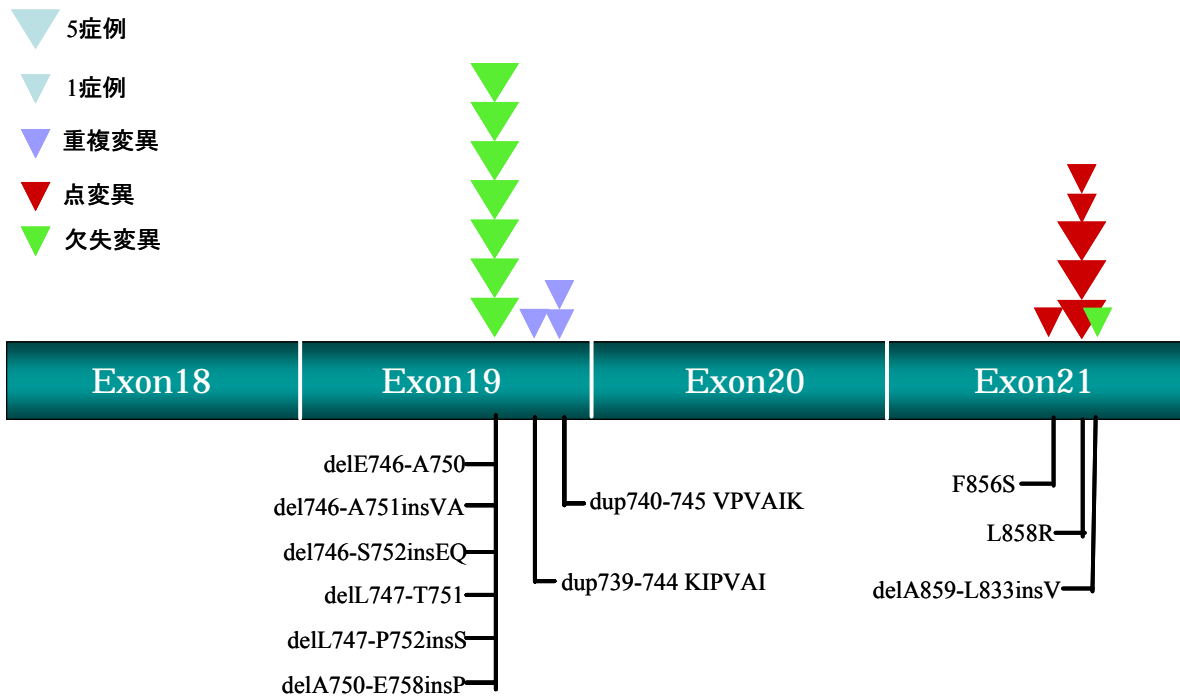
点変異 28 症例は exon18、19、21 内で認められ、28 症例中 20 症例と最も多く確認された点変異は、exon21 内の 2573 番目の塩基 G から T への置換で、858 番目のアミノ酸ロイシン(L) がアルギニン(R) に置換する変異 L858R であった(図 4 - 1.A)。

この 20 症例の中で 3 症例が二箇所にて点変異を持っており、他方がそれぞれ exon18 上の E709G、exon21 上の V834L、A871G (図 4 - 1.B) であった。次に、exon18 内における G719A が 3 症例に確認され、L861Q、二箇所にて点変異、その他の変異は 1 例ずつであった。欠失変異 37 症例のうち 35 症例は、exon19 内の 746-758 番アミノ酸周辺における 15 塩基、もしくは 18 塩基の in-frame の欠失であった (図 4 - 1.A)。32 症例中 22 症例と最も多く確認された変異は、2235・2236 番目の塩基から始まる 15 塩基の欠失による、746-750 番目のアミノ酸 ELREA の欠失変異であった。この 22 症例と 747-751 番目のアミノ酸 LREAT の欠失 3 症例は、5 アミノ酸の欠失のみであったが、他の 7 症例においては、欠失に加え 1~2 アミノ酸の挿入が認められた。747-752 番目のアミノ酸の欠失とセリン (S) の挿入が 3 症例、750-758 番目の欠失と塩基 CC の挿入によるプロリン (P) の挿入が 2 症例、746-751 番目のアミノ酸欠失と塩基 TTGCT の挿入によるバリンとアラニン (V, A) の挿入、746-752 番目のアミノ酸の欠失と塩基 GCAA の挿入によるグルタミン酸とグルタミン (E, Q) の挿入が各 1 症例ずつ認められた。重複変異の 3 症例は、exon19 内の 5~6 アミノ酸が重複しているというものであった。739-744 番目のアミノ酸の重複が 1 症例、740-745 番目のアミノ酸の重複が 2 症例であった。

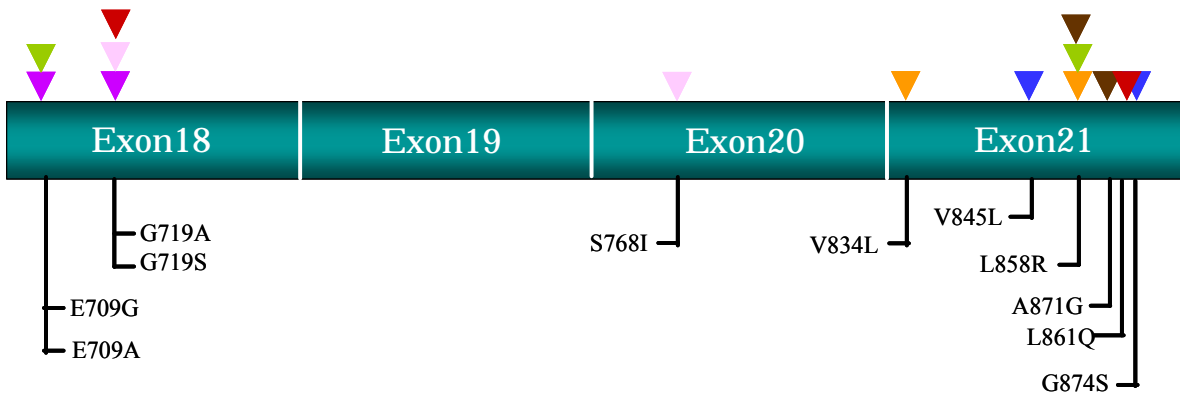
K-ras に関しても、シーケンスにより変異を同定し、7 症例に変異を認めただが、EGFR 遺伝子変異を持つ症例は K-ras の変異を持たなかった。また、今回の解析では、ゲフィチニブに抵抗性を示す変異と言われている T790M の点変異は同定されなかった。

ゲフィチニブ投与症例に限って、変異の有無に関してゲフィチニブ投与後の全生存期間を Log-rank 法で解析したところ、 $p < 0.0001$ と有意に変異を持つ症例の生存期間が延長しており、EGFR の遺伝子変異がゲフィチニブ感受性に関連していることが示された。

また、これまで、ゲフィチニブの抗腫瘍効果との関連が示された変異をゲフィチニブ関連変異として今後の解析はこれらの変異に関して行った。



A. 遺伝子変異が一箇所の症例



B. 遺伝子変異が二箇所の症例

同一症例内での遺伝子変異は同じ色のマーカーで表す

図 4 - 1 : EGFR における変異型と症例数の分布

表 4 - 1 : EGFR における遺伝子変異の種類

		Type of Mutations		No of patients
Gefitinib related mutations	Deletion of exon 19	del2235-2249	delE746-A750	19
		del2236-2250	delE746-A750	7
		del2240-2254	delL747-T751	3
		del2240-2257	delL747-P752 ins S	3
		del2247-2273	delE750-K758	1
		del2252-2275	delS752-I759	1
		del2235-2251 insAATTC	E746-T751 insIP	1
		del2238-2256 insGCAA	delE746-S752 ins EQ	1
	L858R	T2573G	L858R	20
	Other point mutations	T2582A	L861Q	2
		G2156C	G719A	2
		G2155A	G719S	1
	Gefitinib non-related mutations		A2126G	E709H
		dup2215-2232	dup739-744 KIPVAI	1
		dup2219-2236	dup740-745 VPVAIK	2
		G2248C	A750P	1
		A2267G	N756S	1
		G2303T	S768I	1
		C2364T	L788F	1
		G2500T	V834L	1
		T2534C	V845L	1
		del2576-2647	delA859-L883 insV	1
		T2567C	F856S	1
		C2612G	A871G	1
		G2620A	G874S	1

4.3.2 変異 mRNA 比率とゲフィチニブの相関

ダイレクトシーケンスの結果、変異の認められた 64 症例に関しては、cDNA 断片のクローニングを行ってから、24 クローンまたは、96 クローンの exon18 ~ 21 をシーケンスした。得られた解析結果から、(変異の認められたクローン数) / (シーケンスを行った 総クローン数) で変異 mRNA 比率を算出した。まず、臨床情報の揃っている 29 症例から 10 症例を選択し、それぞれ 96 クローンを無作為に選択して解析を行った。10 症例中 5 症例に遺伝子変異が認められ、それぞれの変異 mRNA の比率は 19%、24%、54%、61%、76%であった。残りの 54 症例においては、1 症例につき 24 クローンを無作為に選択して解析を行った。

ゲフィチニブ関連変異の定量的な解析の結果、変異 mRNA の比率は 0 ~ 20% : 3 症例、20 ~ 30% : 8 症例、30 ~ 40% : 8 症例、40 ~ 50% : 7 症例、50 ~ 60% : 9 症例、60 ~ 70% : 8 症例、70 ~ 80% : 9 症例、80 ~ 90% : 3 症例、90 ~ 100% : 2 症例であった (図 4 - 2)。また、この変異 mRNA 比率を変異型別に示した (図 4 - 2)。変異 mRNA 比率は 13% から 96% と広く分布し、平均 52.3%、中央値 52.2% であった。以上のように、

L858R、欠失変異に関わらず、変異 mRNA 比率が多様であることが明らかとなった。

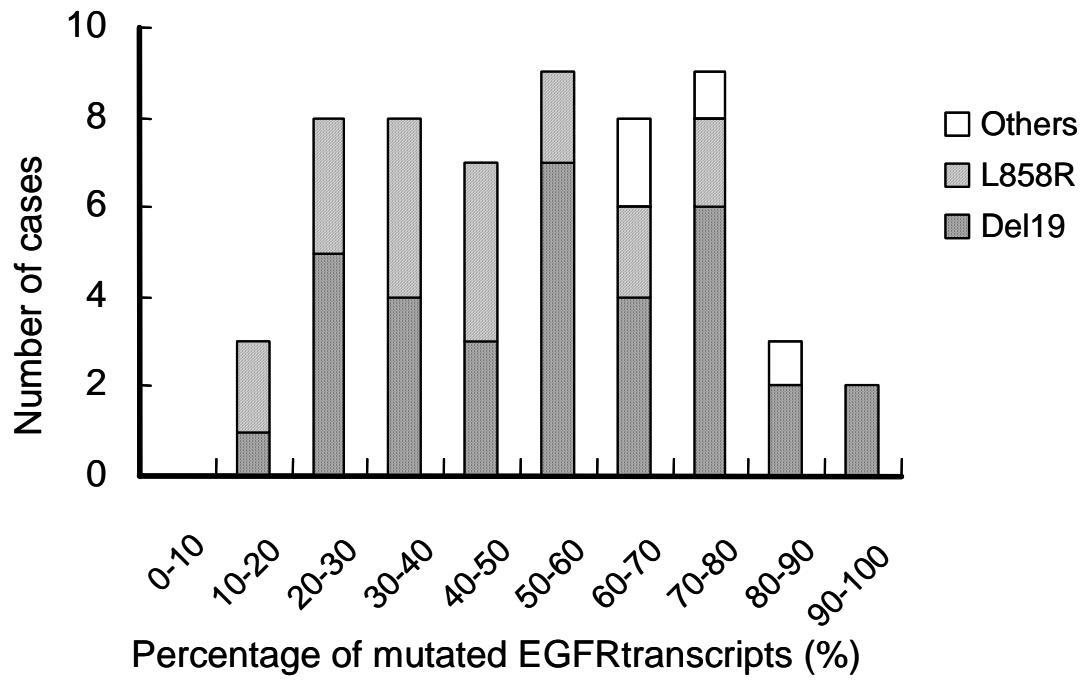


図 4 - 2 : 変異 mRNA の比率の分布

次に、ゲフィチニブ関連変異が認められた症例のうち変異 mRNA 比率が中央値である 52.2%を超える症例を High グループ、中央値以下の症例を Low グループとして、これらの 2 群と臨床情報に関連があるのかを調べた(表 4 - 2)。その結果、変異 mRNA 比率と性別、分化度、喫煙歴、年齢、ステージに関して有意な相関は認められなかった。ただし、性別、喫煙歴に関しては、それぞれ女性、非喫煙の群が高い変異 mRNA 比率を示す傾向があった。また、L858R の点変異と exon19 の欠失変異の術後生存期間を Kaplan-Meier 法による生存曲線解析、及び Log-rank 検定で解析したが、有意な差は認められなかった。

表 4 - 2 : 変異 mRNA 比率と臨床情報との関連

	Mutation rate		<i>p</i>
	H N=30	L N=31	
Age			
Median	59	65	0.3488
Mean (Range)	60.5 (38-80)	63.0 (38-79)	
Sex			0.0728
male	7	14	
Female	23	17	
Tumor size			0.8213
>31 mm	17	18	
<30 mm	13	13	
Nodal involvement			0.8998
Positive	15	16	
Negative	15	15	
Stage at operation			0.7099
I	14	13	
II	4	8	
III	12	9	
IV	0	1	
Surgical Radicality			0.7221
Complete Resection	27	27	
Incomplete Resection	3	4	
Histology			0.5344
Adenocarcinoma, por	7	6	
Adenocarcinoma, mod or well dif.	21	24	
Squamous cell carcinoma	1	0	
Others	1	1	
Smoking			0.0798
BI=0	25	20	
BI<200	3	4	
201<BI<1000	1	5	
1001<BI	1	2	

さらに、High グループと Low グループの術後生存期間に差が認められるかどうかを、ゲフィチニブ投与例を除いた症例を用いて解析した(図 4 - 3)。その結果、有意差は認められず($p=0.5811$)、ゲフィチニブ非投与下では 2 群に予後の差が無いことが示された。

また、この結果が正常細胞の混入によるものではないことを確認するために、組織中の癌細胞の割合を HE 染色した切片の組織画像を用いて調べ、それぞれの群の切片上の癌細胞の非癌細胞に対する割合を求めた。それぞれの群からランダムに 5 症例ずつ選択し、癌細胞の割合の平均を求めた(図 4 - 4)。その結果、癌細胞の割合に有意な差は認められず、正常細胞の混入が原因で変異 mRNA 比率に差が生じているのではないことが明らかとなった。

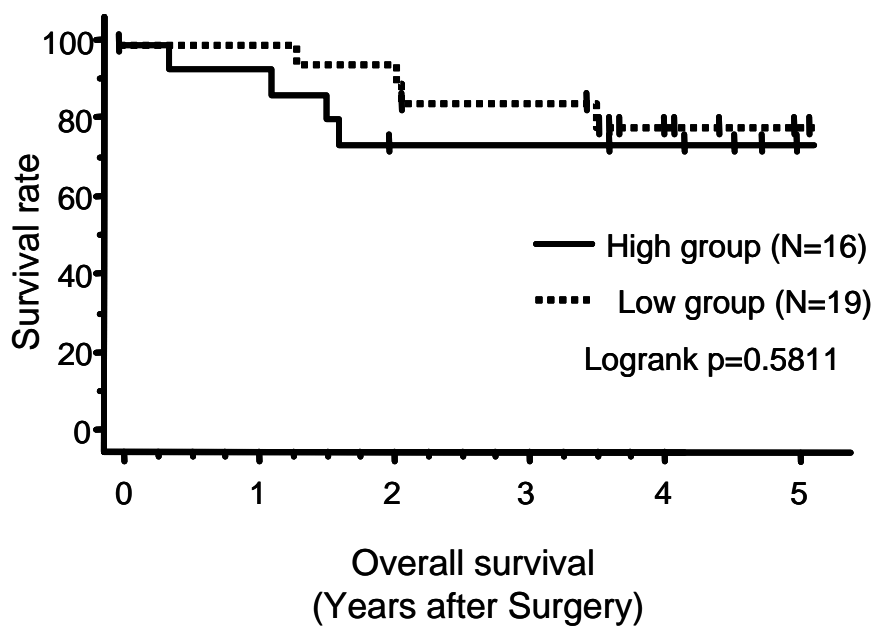


図 4 - 3 : *EGFR* 遺伝子変異を持つ症例の術後生存期間

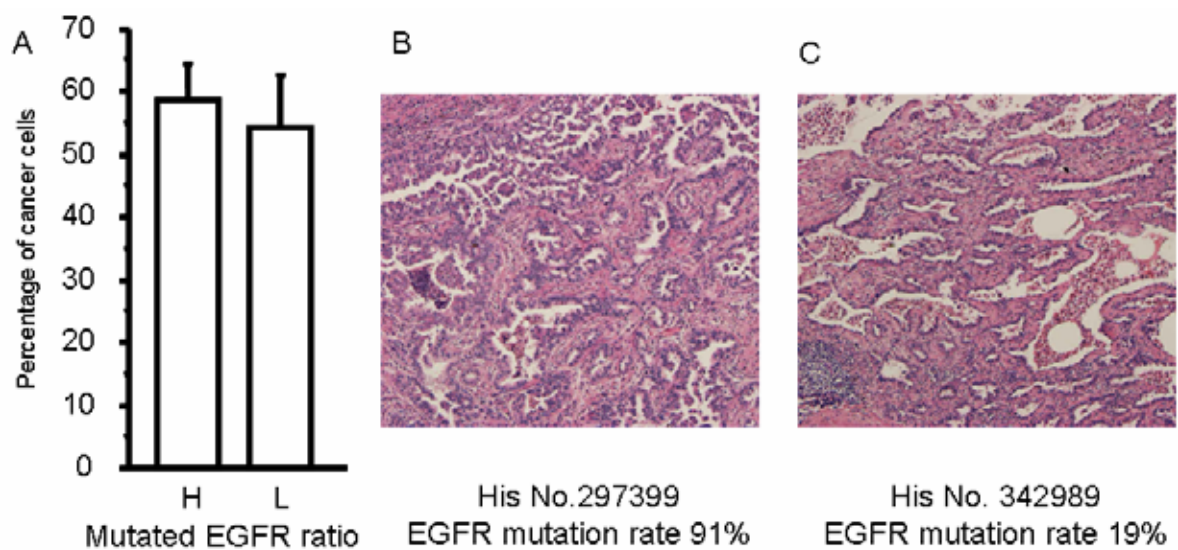


図 4 - 4 : HE 染色切片における癌細胞の割合

A: High グループと Low グループの癌細胞の割合の中央値を棒グラフとして示し、標準偏差をエラーバーとして示した。B: High グループの切片画像の一例。C: Low グループの切片画像一例。

次に、EGFR の変異とゲフィチニブの抗腫瘍効果との相関を検討するために、EGFR 遺伝子変異が同定され、ゲフィチニブが投与された 25 症例に関して解析を行った。変異 mRNA 比率の中央値は 60.0 であり、この中央値を基に症例を High グループと Low グループの 2 群に分けて解析を行った。2 群の臨床情報を表 4 - 3 に示す。この結果より、変異 mRNA 比率と前化学療法、転移臓器、Performance status、術後からゲフィチニブ投与までの期間には相関は認められなかった。各症例のゲフィチニブへの抗腫瘍効果を表 4 - 4 にまとめた。6 症例に関しては、抗腫瘍効果が測定不可能な病変であったが、そのうち 4 症例は Carcinoembryonic antigen レベルや骨転移病変の縮小が見られたため、Partial responder とした。他の 2 例は抗腫瘍効果が得られなかった。High グループの 13 症例、Low グループの 10 症例で抗腫瘍効果が得られた結果となった。fisher's 正確確率検定による解析では 2 群間に有意な差は認められなかった。

表 4 - 3 : 変異 mRNA 比率とゲフィチニブ投与症例の臨床情報との関連
 化学療法、及び放射線療法における括弧は両方の治療を受けたことを示す。

		EGFR Mutation	
		H	L
		N=13	N=12
Age	Median	53	63
	Mean (Range)	56.5 (38-74)	61.7 (38-74)
Sex	male	1	4
	Female	12	8
Stage at operation	I, II	7	6
	III	6	6
Surgical Radicality	Complete Resection	12	9
	Incomplete Resection	1	3
Histology	Adeno, por dif.	4	3
	Adeno, mod or well dif.	9	9
Smoking	BI=0	10	8
	1<BI<200	3	1
	201<BI	0	3
Prior chemotherapy	0,1	8	9
	>2	5	3
Number of metastatic organs	Single	10	7
	Multiple	3	5
Performance status	0 or 1	12	9
	2 or more	1	3
Period between surgery and Gefitinib	< One year	2	4
	One year or longer	11	8
Treatment after Gefitinib	No treatment	9	7
	Chemotherapy	3	3 (2)
	Radiotherapy	1	4 (2)

表 4 - 4 : 変異 mRNA 比率とゲフィチニブに対する奏効

	High mutation rate	Low mutation rate
Complete responders	2	2
Pertial responders	11	8
Non-responders	0	2

さらに High グループと Low グループのゲフィチニブに対する感受性を解析するために、ゲフィチニブ投与後の全生存期間、及び無増悪期間に関して Kaplan-Meier 法による生存曲線解析、及び Log-rank 検定を行った (図 4 - 5)。High グループの全生存期間の中央値は 45 ヶ月であり、Low グループの中央値は 20 ヶ月であった。また、両群を Log-rank 法により検定すると有意に High グループの生存期間が長かった ($p=0.0122$)。さらに、無増悪期間においても High グループの中央値が 22 ヶ月であるのに対し、Low グループでは 9 ヶ月であり、両群間に有意な差が認められた ($p=0.0294$)。

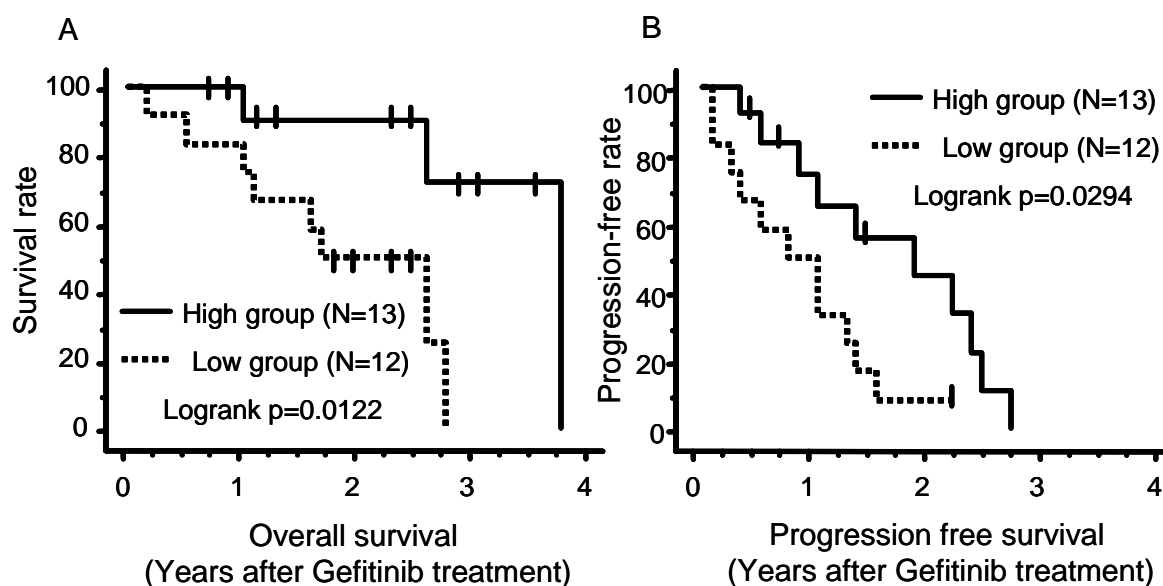


図 4 - 5 : ゲフィチニブ投与例における全生存期間及び無増悪期間

A: 全生存期間、B: 無増悪期間。

4.4 考察

4.4.1 EGFR 遺伝子変異とゲフィチニブ感受性

ダイレクトシーケンスにより 169 症例のうち 64 症例に EGFR 遺伝子変異が同定された。ゲフィチニブ投与症例に限って、EGFR 遺伝子変異の有無と全生存期間を解析すると有意に変異を持つ症例の生存期間が延長していた ($p<0.0001$)。ゲフィチ

ニブは、EGFR の ATP cleft と呼ばれる ATP 結合部位に結合して EGFR キナーゼ活性を阻害するが、肺癌における遺伝子変異はその ATP cleft に隣接した部分に起こっていることが明らかとなっている (Lynch TJ.et.al.2004)。ATP 結合部位に変異が起こると、ATP 結合部位の構造変化によってゲフィチニブが結合しやすくなるために、よりゲフィチニブ感受性が高くなるのではないかと考察されている。EGFR 遺伝子の TK ドメインにおける変異に関しては、EGFR 遺伝子を標的としているゲフィチニブの感受性に関連することが 2004 年に報告され (Lynch TJ.et.al.2004, Paez JG.et.al.2004)。今日では EGFR 遺伝子変異を持つ症例群でゲフィチニブ奏効率が高いことが確認されている (Mitsudomi T.et.al.2005, Tokumo M.et.al.2006)。しかしながら、EGFR 遺伝子変異がゲフィチニブ投与症例の生存期間を延長するのかどうかに関しては明らかにされていなかった。日本においては、EGFR 遺伝子変異がゲフィチニブ投与症例の生存期間を延長するという報告が多くなされているが、EGFR 遺伝子変異を持つ約 2 割の症例ではゲフィチニブが奏効していない。リンパ節転移や遠隔転移を伴った進行肺癌においても劇的に改善する症例が少なくないため、ゲフィチニブは有用な分子標的薬であると考えられるが、感受性の有無に関して患者のさらなる分類を行う必要がある。

本研究では、25 種類の EGFR 遺伝子変異を同定した。75 遺伝子変異のうち 56 変異 (75%) はホットスポットである exon19 の欠失変異が exon21 の L858R の点変異であった。また、新規の変異が 8 変異同定された。EGFR の TK ドメインにおける遺伝子変異には T790M の点変異のようにゲフィチニブ 抵抗性を示す変異も同定されている。そのため、変異 mRNA 比率とゲフィチニブとの相関に関する解析はゲフィチニブとの関連が示されている変異に限り行った。

4.4.2 EGFR 遺伝子変異の定量解析

これまでの報告では、ダイレクトシーケンスにより検体の EGFR 遺伝子の変異の有無を探索し、ゲフィチニブの抗腫瘍効果との相関を検討している。その結果、肺癌組織に EGFR 遺伝子変異が起こっている症例では、ゲフィチニブに感受性があるという報告がなされている (Mitsudomi T.et.al.2005, Tokumo M.et.al.2006, Huang SF.et.al.2004)。これに対して、本研究では、各症例の目的遺伝子のクローニングを行い、複数のクローンのシーケンスから変異のある mRNA の比率を推定するという方法を用いた。これにより新たに各症例における変異 mRNA 比率を明らかにし、従来の定性的な解析に加え、遺伝子変異の定量的評価とゲフィチニブに対する抗腫瘍効果との関係を検討した。具体的には、癌組織中に存在する EGFR の mRNA 中の変異 RNA の比率を変異 mRNA 比率として解析した。その結果、変異 mRNA 比率は 13% から 96% と組織間で様々であった。

ここで、変異 mRNA 比率が正常細胞の混入の結果である可能性があったため、各組織内の癌細胞の割合を組織から作成した HE 染色切片を用いて解析した。その結果、変異 mRNA 比率の低い Low グループと高い High グループ間で癌細胞の組織に占める割合に差が無かったため、変異 mRNA 比率のばらつきは他の要因に起因するものと

考えられた。この要因は3つ考えられる。1つ目の可能性は、各組織中に *EGFR* 遺伝子変異を持つ癌細胞と *EGFR* 遺伝子変異を持たない癌細胞が混在しているということである。肺癌を含め、癌組織は多様性に富んだ組織であり、組織学的に多様な癌細胞が混在している。これは、遺伝子変異の違いによることが示されている (Yatabe Y.et.al.2000)。また、例えば、*p53* の遺伝子変異に関して免疫組織化学解析により癌組織の中の一部の癌細胞のみが遺伝子変異を持っていることも明らかになっている (Bodner SM.et.al.1992,Higashiyama M.et.al.1998)。このように *EGFR* 遺伝子変異に関しても組織内の多様性があることが考えられる。この議論に関しては次章でさらなる解析を行った。2つ目の可能性は、細胞内の変異 *EGFR* 遺伝子のゲノムコピー数の違いが考えられる。癌細胞は染色体や遺伝子の増幅が頻繁に起こっているため、増幅によるゲノムコピー数の変化が変異 mRNA 比率の変化をもたらしていると考えられる。実際に *EGFR* 遺伝子の増幅は非小細胞肺癌でよく起こっている (Taron M.et.al.2005,Therasse P.et.al.2000)。さらに、*EGFR* 遺伝子の増幅がゲフィチニブの感受性に関連しているという報告もあり、細胞内の変異 *EGFR* 遺伝子の数も重要であると考えられる。3つ目の可能性は、*EGFR* 遺伝子発現の影響である。*EGFR* 遺伝子変異と mRNA レベルでの遺伝子発現、タンパクレベルでの遺伝子発現との相関は無いことが報告されているが (Dziadziszuko R.et.al.2006)、ヘテロ遺伝子変異であった場合に正常アレル由来の *EGFR* と変異アレル由来の *EGFR* とが mRNA レベル、またタンパクレベルで発現比率が異なってくるのが十分に考えられる。実際には、これら3つの要因が複雑に関連して、変異 mRNA 比率として観察されているものと考えられる。

変異 mRNA 比率と臨床情報との比較では、女性の群で高い変異 mRNA 比率を示し ($p=0.07$) またタバコ喫煙歴が無い群で高い変異 mRNA 比率を示した ($p=0.08$) が有意な差ではなかった。また、その他の臨床情報との関連は示されなかった。組織型に関しては、変異を持つ非小細胞肺癌のほとんどが肺腺癌であるため、他の組織型との比較は困難である。

変異 mRNA 比率とゲフィチニブ投与後の無増悪期間、また全生存期間を比較した結果、変異 mRNA 比率が高い群において有意に予後が良かった。変異 mRNA 比率はゲフィチニブ非投与下では生存期間の延長に関連していない。今回の結果は、変異 mRNA 比率が術後のゲフィチニブ感受性に関連していることを示している。特に、High グループにおいては全例奏効しており、変異 mRNA 比率が高い群へのゲフィチニブ投与は有効であると考えられる。

EGFR の遺伝子変異とゲフィチニブの奏功との関連性を検討する報告は数多くなされているが、変異 mRNA 比率を用いた遺伝子変異の定量的評価を行った例はない。本研究において、*EGFR* の遺伝子変異症例における変異 mRNA の比率は症例ごとに多様であった。また、抗腫瘍効果の得られた症例では変異 mRNA の比率は高い傾向を示し、変異遺伝子の定量的な評価もゲフィチニブの効果予測因子として有用であると考えられた。しかし、変異 mRNA 比率の多様性が何に起因するのか、また、抗腫

瘍効果に影響を及ぼす直接的な原因は不明なままである。そこで、次章において *EGFR* 遺伝子変異とゲフィチニブの抗腫瘍効果に関するさらなる解析を行った。

5. 非小細胞肺癌における *EGFR* 遺伝子変異の組織内多様性による抗癌剤感受性への影響

5.1. 序論

前章において *EGFR* 遺伝子の変異 mRNA 比率がゲフィチニブの感受性に関連することを示した。変異 mRNA 比率が低い症例ではゲフィチニブが奏効しない傾向にあった。本研究では、*EGFR* の遺伝子変異を持つ症例に関して、*EGFR* の遺伝子変異の組織内多様性に着目してゲフィチニブの抗腫瘍効果との関連を調べた。癌組織は一様なものではなく多様性に富んだ組織であり、遺伝子変異に関しても同一組織内に不均一性があることが明らかとなっている (Samowitz.et.al.1999)。例えば、Gonzalez-Graciaらは、大腸癌組織を用いて組織内の多様性について検討している。*BAX*、*transforming growth factor β receptor* に関して、正常の遺伝子を持つ癌細胞とヘテロの変異を持つ癌細胞とホモの変異を持つ癌細胞が同一組織中に混在していることを明らかにしている (Gonzalez-Garcia I.et.al.2006)。また、Maleyらは、食道癌組織の遺伝子変異に関する多様性を明らかにするとともに、癌の発生、進展に関連する様々な遺伝子変異の獲得状況を時系列で示している (Maley CC.et.al.2006)。このように、固形癌は、遺伝子変異の組み合わせが組織内で異なる癌細胞から成り立っていることが明らかにされている。遺伝子変異が異なる細胞は分子標的薬の感受性も異なると考えられる。*EGFR* の遺伝子に関して変異を持つ癌細胞と変異を持たない癌細胞が混在している症例があると仮定すると、混在例では正常 *EGFR* を持つ癌細胞がゲフィチニブ抵抗性を示し、結果として奏効しない。そこで、本研究では、*EGFR* の遺伝子変異の組織内多様性を明らかにし、ゲフィチニブの感受性との関連を調べた。

5.2. 材料と方法

5.2.1. サンプル DNA の抽出

大阪府立成人病センター提供された肺腺癌のうち Gefinitib 投与症例であり、*EGFR* 遺伝子変異を持つ 21 症例を解析に用いた。凍結組織を OTC コンパウンド (Sakura Finetechnical) で包埋後、液体窒素を用いて急速凍結した。その後、クライオスタット (Leica) を用いて、35 μ m の切片を切り出し、0.05% トルイジンブルー (Wako) による染色を行った。乾燥後速やかに Laser Micro Dissection (Leica) を用いて癌細胞 100 μ m 平方の小領域を各症例 60 箇所切り出した。切り出された細胞より QIAamp DNA micro kit (QIAGEN) を用いて GenomeDNA を 20 μ l の DW に抽出した。DNA 抽出までの解析の流れを図 5 - 1 に示す。

また、組織から確実に癌細胞のみを切り出すために症例ごとに HE 染色切片を作成し、組織像を確認した。

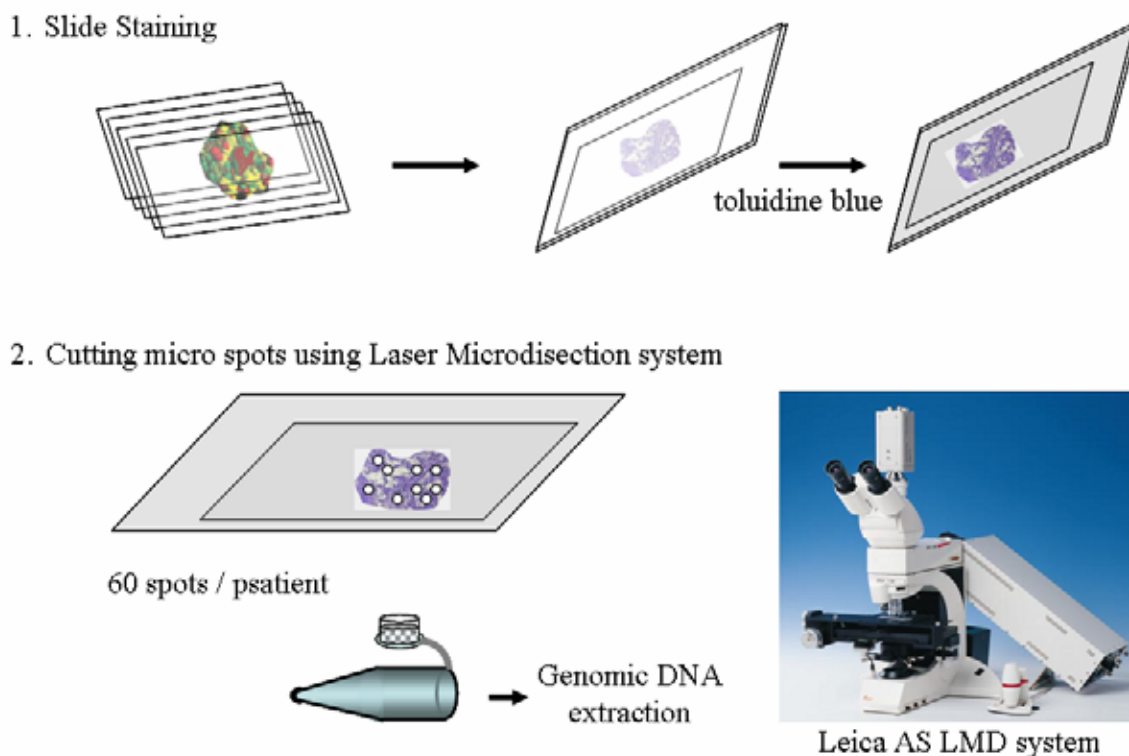


図 5 - 1 : 肺癌組織からの微小領域の切り出し、及び DNA 抽出

5.2.2. GenomeDNA の濃度測定

微小領域より抽出した GenomeDNA は濃度が薄いため光学濃度測定器では濃度を測定できない。そのため Realtime PCR 法によりハウスキーピング遺伝子である AldolaseA の増幅曲線を解析して濃度測定を行った。AldolaseA が含まれる 16p11.2 の染色体領域は肺癌においてゲノムの増幅、失欠頻度が低く、重要な癌遺伝子や癌抑制遺伝子が近傍に存在しない (Balsara BR.et.al.2002、Tonon G.et.al.2005)。検量線には肺癌組織から直接抽出した GenomeDNA の濃度を吸光高度計で計測し、希釈系列を作成して用いた。また、希釈系列の DNA 量は、50ng、5 ng、0.5 ng、0.05 ng で行った。微小領域より抽出したサンプル DNA を 10 サンプル用意し、それぞれサンプル DNA 5 μ l に SYBR Green master mix (TAKARA) 10 μ l、aldolase F primer 0.4 μ l、aldolase R primer 0.4 μ l、DW 4.2 μ l を加え、Light Cycler (Roche) を用いて、95 30 秒で変性させた後、95 5 秒、60 10 秒、72 15 秒で 47cycle の条件で Realtime PCR を行った。反応の結果得られた増幅曲線よりサンプル DNA の濃度を求めた。

5.2.3. 蛍光色素による蛍光強度の差異

PCR により増幅した断片中に存在する遺伝子変異は一塩基多型を検出する手法である SNaPshot kit (Applied Biosystems) を用いた。SNaPshot kit に含まれる ddNTP は塩基ごとに異なる蛍光色素で標識されており (表 5 - 1)、ddNTP が結合した後のさらなる塩基の伸長は起こらず、目的塩基の変異を検出することができる。

表 5 - 1 : 塩基と蛍光色素

ddNTP	Dye Label	Color
A	dR6G	Green
C	dTAMARA™	Black
G	dR110	Blue
T	dROX™	Red

蛍光色素ごとの蛍光強度による差を補正する目的で、SNaPshot kit における各色素の蛍光強度比を求めた。まず、増幅した *EGFR* 遺伝子（正常配列）を pCR Blunt ベクターに組み込み、回収したプラスミドを比較対照とした。次に、T から G（ 70 : 症例番号、T2573G : 変異）、G から A（ 110、del2236-2250）、T から C（ 2、del2240-2257）に変異した遺伝子を持つプラスミドをそれぞれ同様に回収した。正常配列を持つプラスミドと同量ずつ混合し、PCR 反応により *EGFR* 遺伝子を増幅した後、SNaPshot により蛍光強度を比較した。この反応には、実際測定する DNA 量の 100 倍量（5500 分子相当）のプラスミド DNA をテンプレートとした。これは、蛍光色素による差を調べるためのもので、DNA 量が少ないことによる実験誤差を抑えるためである。

5.2.4. SNaPshot による変異の定量解析

抽出された GenomeDNA 5 μ l を用いて *EGFR* 遺伝子内に同定された変異部位を含むエクソンを PCR により増幅した。Exon19 の PCR 反応は、exon19F primer（10pmol/ μ l, 5'-GCAATATCAGCCTTAGGTGCGGTC-3'）、exon19R primer（10pmol/ μ l, 5'-CATAGAAAGTGAACATTTAGGATGTG-3'）、各 0.4 μ l、10 \times PCR buffer for KOD -Plus- 1 μ l、2mM dNTPs 1 μ l、25mM MgSO₄ 0.4 μ l、KOD -Plus- DNA Polymerase 0.2 μ l、DW 2.2 μ l を加え全量を 10 μ l とした。94 2 分で変性させた後、94 15 秒、62 30 秒、68 90 秒で 25cycle の PCR 反応を行った。さらに、exon19Fint primer（10pmol/ μ l, 5'-CCTTAGGTGCGGCTCCACA GC-3'）、exon19Rint primer（10pmol/ μ l, 5'-CATTTAGGATGTGGAGATGAG C-3'）を用いて全量 50 μ l の反応量になるように試薬を加え、25cycle の nested-PCR を行った。Exon21 の PCR 反応は、exon21F primer（10pmol/ μ l, 5'-CTAACGTTCCGCCAGCCATAAGTCC-3'）、exon21R primer（10pmol/ μ l, 5'-GCTGCGAGCTCACCCAGAATGTCTGG-3'）、を用いて全量 10 μ l の反応量になるように試薬を加え、40cycle の条件で行った。

次に、増幅した cDNA 7.5 μ l に Shrimp Alkaline Phosphatase（TAKARA）0.5 μ l、Exonuclease（TAKARA）0.25 μ l、DW 0.75 μ l、Shrimp Alkaline Phosphatase buffer 1 μ l を加え、37 で 1 時間保温し、未反応の dNTP、プライマーの除去を行い、80 で 30 分保温し酵素を失活させた。次に、反応溶液を DW で 20 倍希釈し、この溶液 2 μ l に SNaPshot multiplex mix 2.5 μ l、SNaPshot 用に作成したプライマー 0.5 μ l を加え、Thermal

Cycler を用いて 96 15 秒、50 5 秒、60 30 秒のサイクル反応を 25cycle 行った。次に、Shrimp Alkaline Phosphatase (TAKARA) 0.5 μ l、Shrimp Alkaline Phosphatase buffer 0.5 μ l を加え、37 で 1 時間保温し、SNaPshot の反応を止め、80 で 30 分保温し酵素を失活させた。SNaPshot 反応後の溶液を DW で 10 倍希釈し、サイズマーカーである LIZ120 (Applied Biosystems) を加え、DNA Analyzer ABI 3100 (Applied Biosystems) でフラグメント解析を行い、正常配列のシグナル及び変異配列のシグナルを蛍光強度として検出した。SNaPshot 解析の流れを図 5 - 2 に示す。また、SNaPshot 解析に用いたプライマーを表 5 - 1 に示す。

また、SNaPshot の精度と再現性を調べる目的で 5.2.3 で用いた 3 種の変異を持つプラスミドを用いて解析を行った。まず、3 種のプラスミドをそれぞれ 0.1 f g / μ l に希釈し、正常配列の *EGFR* を持つプラスミドと 0%、25%、50%、75%、100% の割合で混合し、これを 2.6 μ l (55 分子相当) 用いて PCR 反応を行った。PCR は本実験と同様に 10 μ l の反応量で行った。次に、増幅された DNA 断片を用いて SNaPshot により解析を行い、各蛍光強度を解析した。

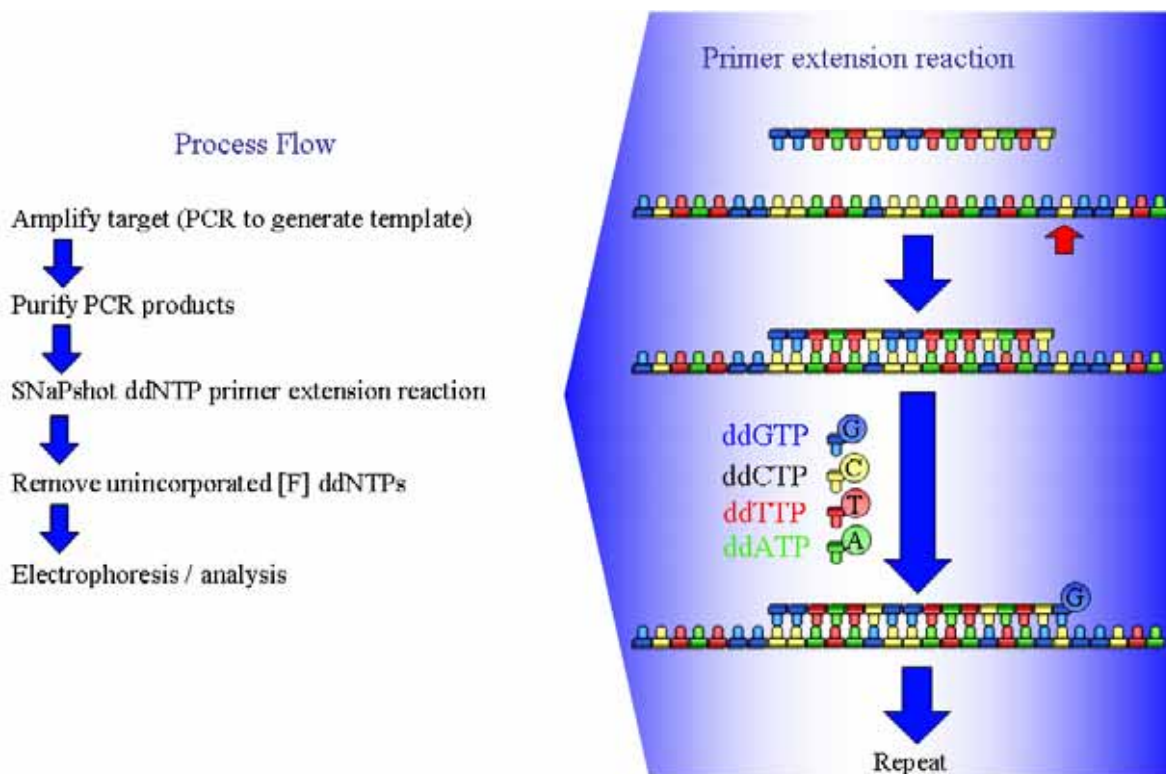


図 5 - 2 : SNaPshot 解析の流れ

表 5 - 1: SNaPshot プライマー

name	bp	sequence	正常 / 変異	検出部位 (bp)	アミノ酸変異
	20	CAAGATCACAGATTTTGGGC	T / G	2573	L858R
	20	AAATTCCTCGCTATCAAG	G / A	2236	del 746-750
	20	TCCCGTCGCTATCAAGGAAT	T / C	2240	del 747-751、753
	20	AAAATTCCTCGCTATCAA	G / A	2235	del 746-750
	20	CTCCACCGTGCAGCTCATCA	C / T	2369	T790M
	20	AATTCCTCGCTATCAAGG	A / C	2237	del 746-750、751
	20	ATTCCTCGCTATCAAGGA	A / T	2238	del 747-751、752
	20	TTCCCGTCGCTATCAAGGAA	T / G	2239	del 747-749
	20	TCAAAAAGATCAAAGTGCTG	G / T	2155	G719C
	20	CAAAAAGATCAAAGTGCTGG	G / C	2156	G719C
	20	AGATTTTGGGCTGGCCAAAC	T / A	2582	L861Q
	20	GACCTGGCAGCCAGGAACGT	A / G	2529	del L844-L883
	20	CTATCAAGGAATTAAGAGAA	G / C	2248	del A750-E758 ins P

5.2.5. 統計解析

統計解析は 4.2.5.と同様に行った。

5.3. 結果

5.3.1. 微小サンプル中の遺伝子変異定量測定系の確立

LMD により癌細胞 30 細胞から 50 細胞を含む微小部位を切り出し、DNA 抽出を行う。抽出した DNA 中の正常 *EGFR* と変異 *EGFR* を定量的に測定する系として SNaPshot

kit を用いた系として確立した。まず、LMD により切り出された微小領域から 10 スポット分の濃度測定を行った結果、PCR 反応に用いた 5 μ l 中の GenomeDNA 量は 1 スポット当り 177.8pg (\pm 41.0 SD : 標準偏差) であった。これは、GenomeDNA55 分子に相当する。また、全体で 20 μ l であるので、1 スポット当り 711.2pg、220 分子である。

次に、蛍光色素による蛍光強度の差異を検討した。この結果、Red (T) に対し Blue (G) が 2.1 倍、Black (C) が 0.67 倍、Blue (G) に対し Green (A) が 0.66 倍の蛍光強度であることがわかった。この強度比を是正する目的で、実験データの補正を行った。

次に、SNaPshot の精度と再現性を調べた (図 5 - 3)。T から G、G から A、T から C のいずれの場合にも検出精度は非常に高く、10 回の実験データの平均はおおよそ混合比と同一であった。また、実験間のばらつきも低く抑えられており、再現性の高いデータが得られた。

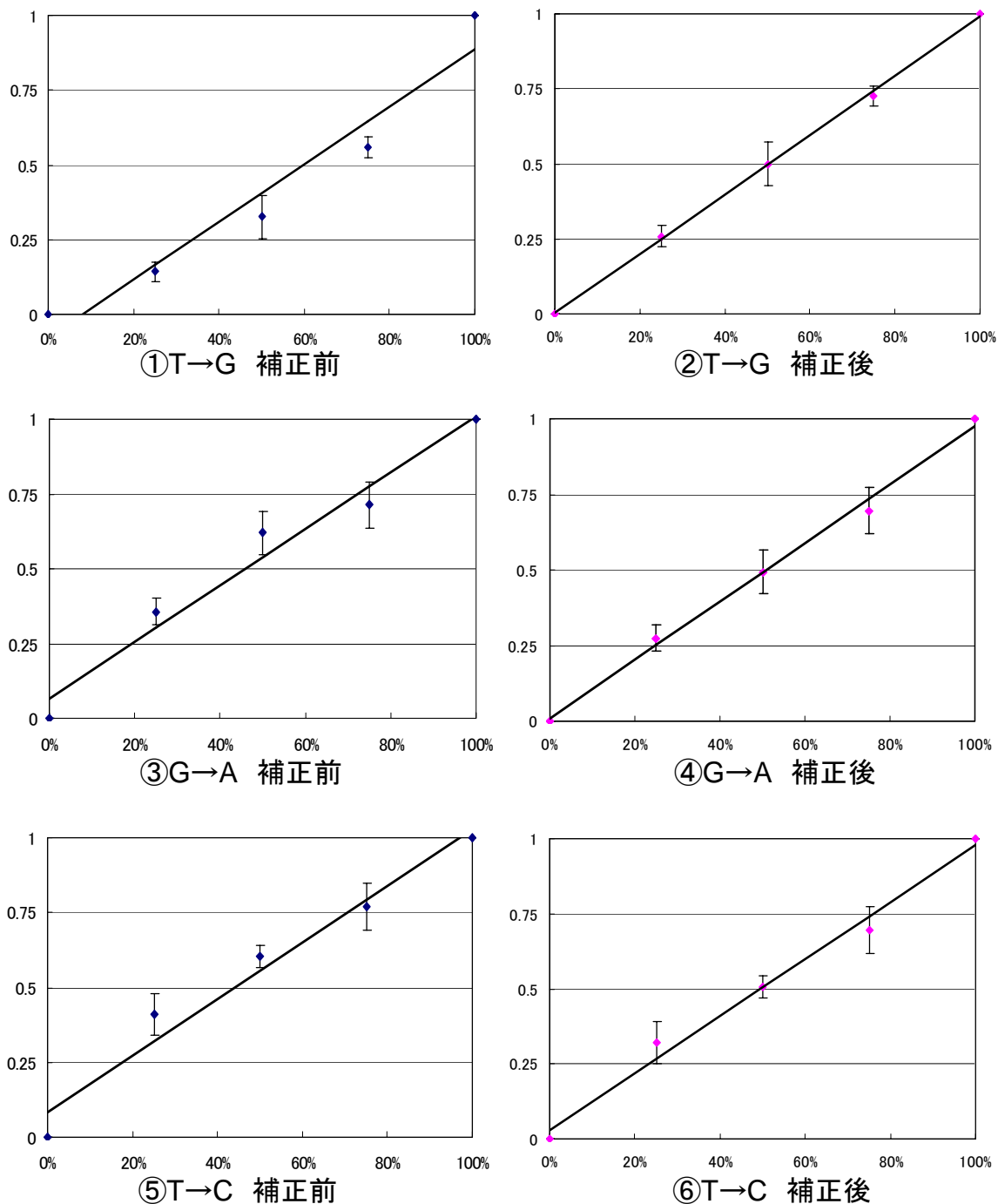


図 5 - 3 : SNaPshot データの補正

縦軸：正常配列に対する変異配列の蛍光強度比

横軸：混合プラスミド中の変異を持つプラスミドの割合

5.3.2. EGFR 遺伝子変異の組織内多様性

EGFR 遺伝子変異を持つ 21 症例の組織からそれぞれ 60 ヶ所を切り出し、EGFR の遺伝子変異を定量的に測定した結果、同一組織内に EGFR 遺伝子変異癌細胞と正常 EGFR 遺伝子のみからなる癌細胞の遺伝的背景の異なる 2 種類の癌細胞からなる組織

が確認された(図5-4)。IV-17の症例では、同一切片上にEGFRの遺伝子変異を持つ癌細胞の群(①、②)とEGFRの遺伝子変異を持たない癌細胞の群(③、④)が確認された。IV-17と同様にIV-70、IV-14、IV-110、IV-21の計6症例で癌細胞の多様性が確認された。21症例を解析した結果を表5-2に示す。6症例は混在例であり(グループa)その他の15症例は全ての癌細胞がEGFR遺伝子変異を持っていたが、この15症例のうち8症例はEGFR変異の中央値が4割から6割であり(グループa)他の7症例は領域あたりのEGFR変異7割を超えていた(グループb)。

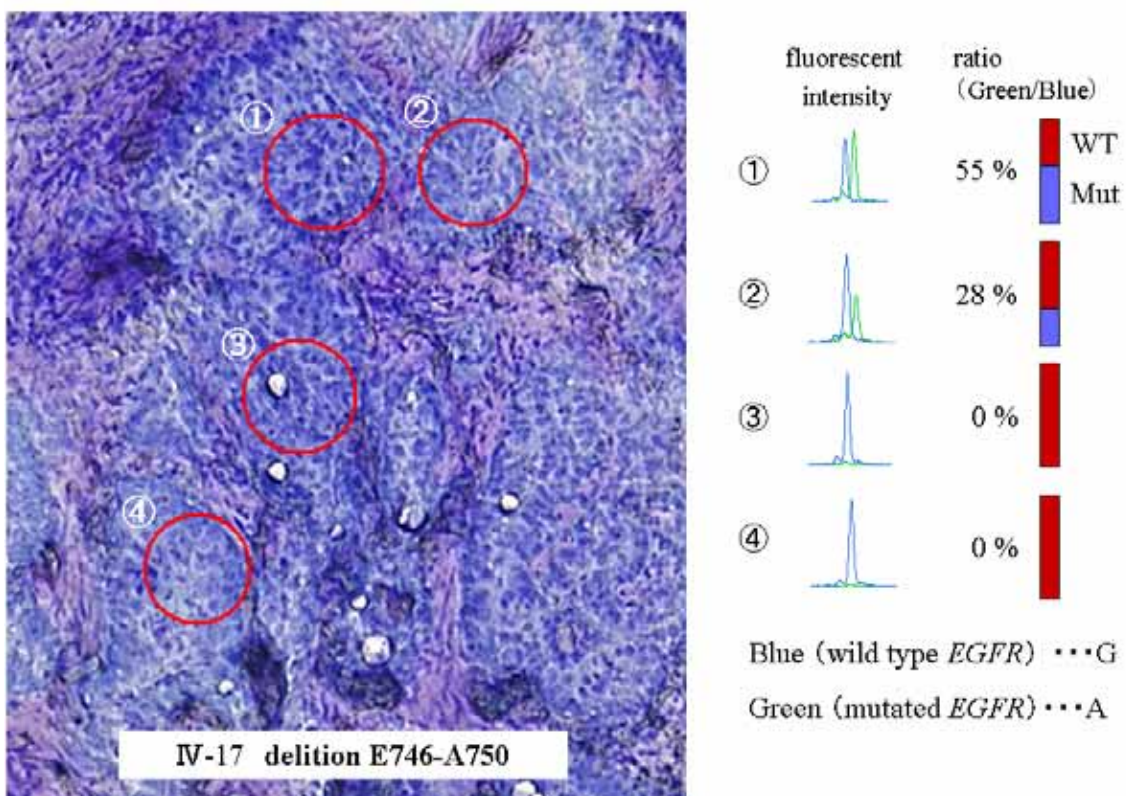


図5-4: EGFR 遺伝子変異に関する混在例の一例

から は、それぞれ赤丸の範囲からゲノムDNAを抽出し、SNaPshotにより解析した。その解析結果を中央に示し、右に棒グラフとパーセンテージに直して示した。

表 5 - 2 : EGFR 遺伝子変異に関する 21 症例の解析結果

	Patient	Sex	Age	Smoking	Stage	Histology	Mutation	median(%)	Group
1	Ⅲ-7	M	64	Former	Ⅲa	Ad	L858R	47.9	I a
2	Ⅲ-5	F	53	Never	Ⅱb	Ad	L858R	48.0	I a
3	Ⅳ-11	M	74	Former	Ⅱa	Ad	L858R	50.4	I a
4	Ⅳ-26	F	67	Never	Ⅲb	Ad	L858R	53.2	I a
5	Ⅳ-15	F	65	Never	Ⅲa	Ad	L858R	53.9	I a
6	Ⅳ-25	F	49	Never	Ⅱb	Ad	delE746-A750	58.7	I a
7	Ⅳ-8	F	53	Never	Ⅲa	Ad	delE746-A750	58.6	I a
8	Ⅳ-18	F	59	Never	I a	Ad	delE746-A750	61.5	I a
9	I-4	M	71	Former	Ⅱb	Ad	L858R	74.2	I b
10	V-1	F	53	Never	I a	Ad	delE746-A750	75.1	I b
11	V-4	F	48	Never	I a	Ad	delE746-A750	81.2	I b
12	Ⅲ-16	F	74	Never	I b	Ad	delE746-A750	82.8	I b
13	V-3	F	70	Never	Ⅲa	Ad	delL747-T751	84.5	I b
14	Ⅳ-20	F	53	Never	I b	Ad	L861Q, G719S	82.9	I b
15	Ⅲ-17	F	38	Never	I b	Ad	delE750-K758	93.9	I b
16	Ⅱ-70	F	69	Never	Ⅱb	Ad	L858R	24.1	Ⅱ
17	V-2	M	54	Never	Ⅲa	Ad	delE746-A750	48.3	Ⅱ
18	Ⅳ-17	F	57	Never	Ⅲa	Ad	delE746-A750	49.8	Ⅱ
19	Ⅳ-14	F	58	Never	Ⅲb	Ad	E746-T751 insIP	54.3	Ⅱ
20	Ⅱ-110	F	70	Former	Ⅱb	Ad	delE746-A750	58.1	Ⅱ
21	Ⅳ-21	M	59	Former	Ⅲb	AdSc	delA859-L883 insV	NA	Ⅱ

Patient : 症例番号

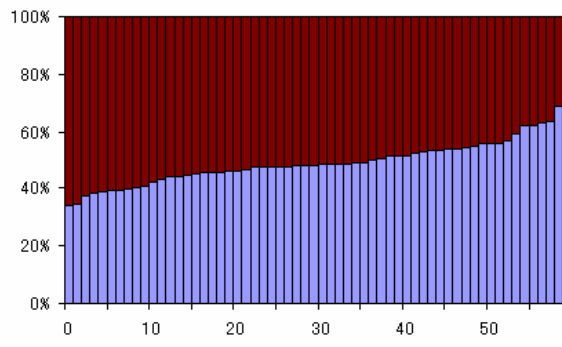
Smoking : 喫煙歴、Forme : 喫煙者、Never : 非喫煙者

Stage : 腫瘍の WHO の分類に基づくステージ

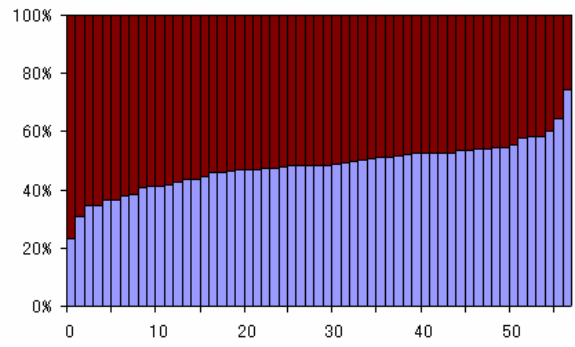
Histology : Ad : 腺癌、AdSc : 腺扁平上皮癌

Median (%) : 変異領域のうちの変異割合の中央値

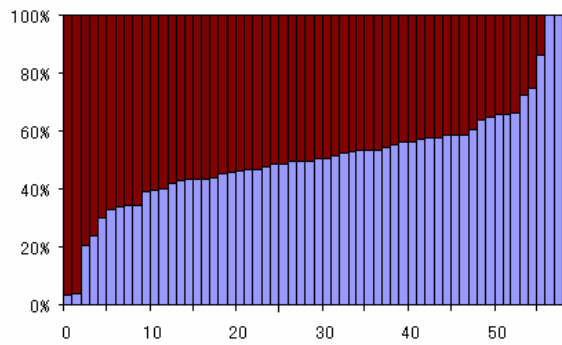
次に、全症例の解析結果を図にして示す。縦の一つの棒グラフは一領域の結果であり、各症例 50 から 60 箇所全解析結果を一つの図にまとめて示した(図 5-5、5-6、5-7)。また、図はグループごとに示した。図 5-21 は変異がどの領域にも見られなかったが、この症例はシーケンスにより変異の存在が確認されている。シーケンスに用いた組織と異なる部位から切片を作成したために、多スポットの解析では変異が同定されなかったものと考えられる。



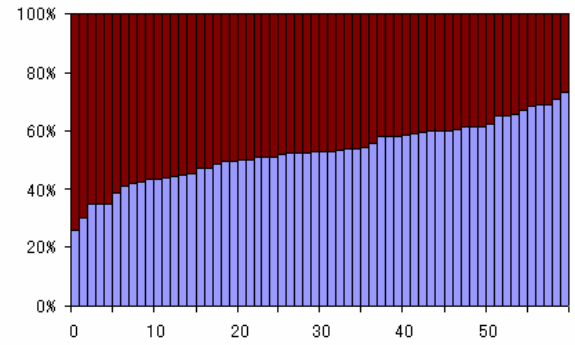
III-7



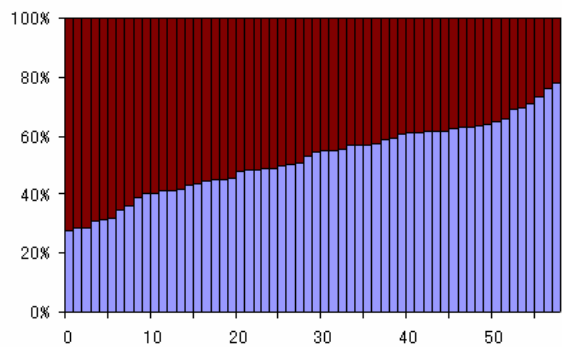
III-5



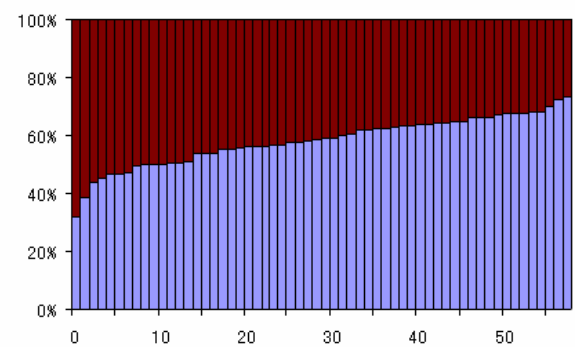
IV-11



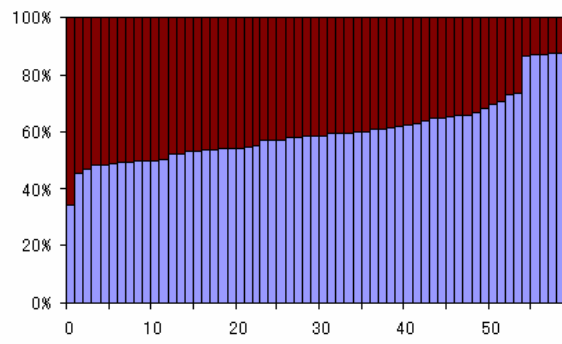
IV-26



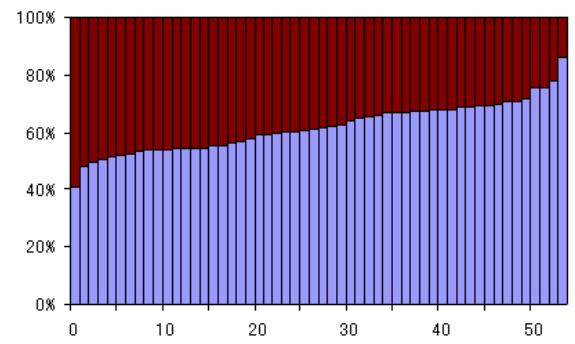
IV-15



IV-25



IV-8

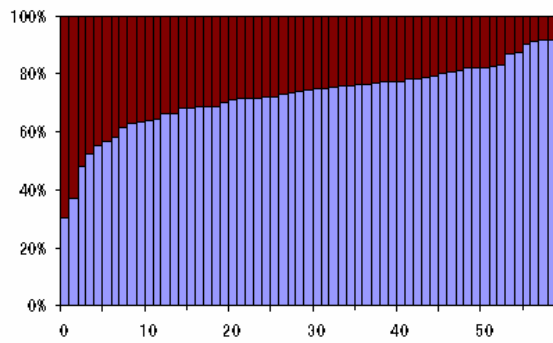


IV-18

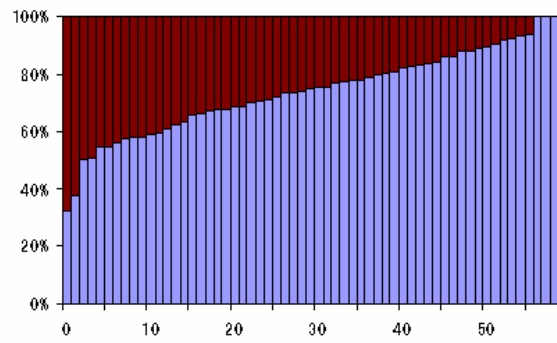
図 5 - 5 : グループ a の解析結果

縦軸 : 正常配列に対する変異配列の蛍光強度比

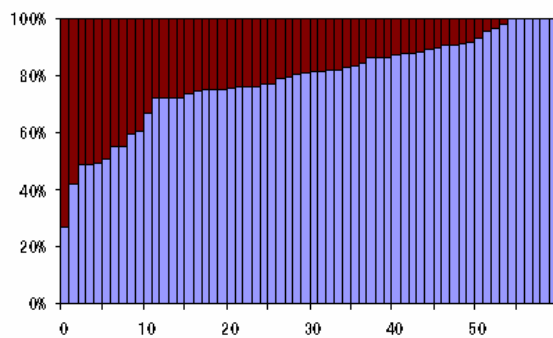
横軸 : 各領域



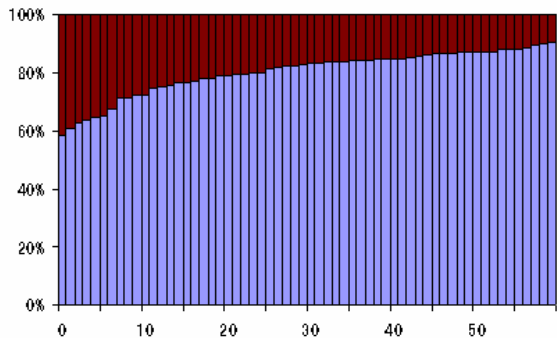
I-4



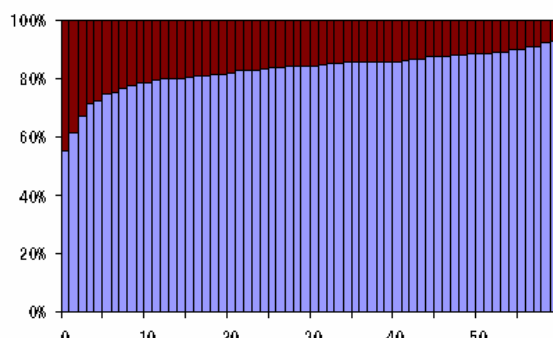
V-1



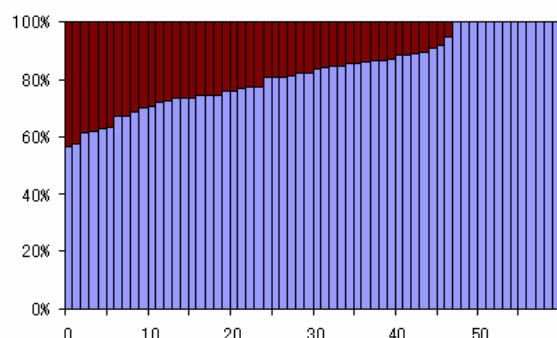
V-4



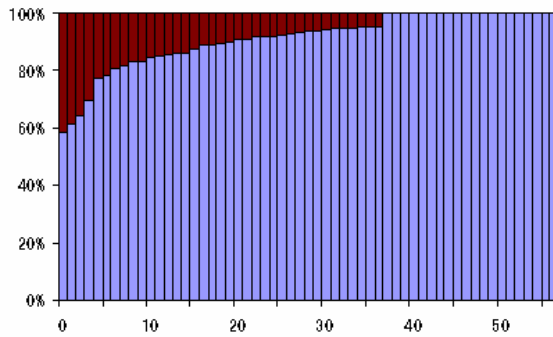
III-16



V-3



IV-20



III-17

図 5 - 6 : グループ b の解析結果

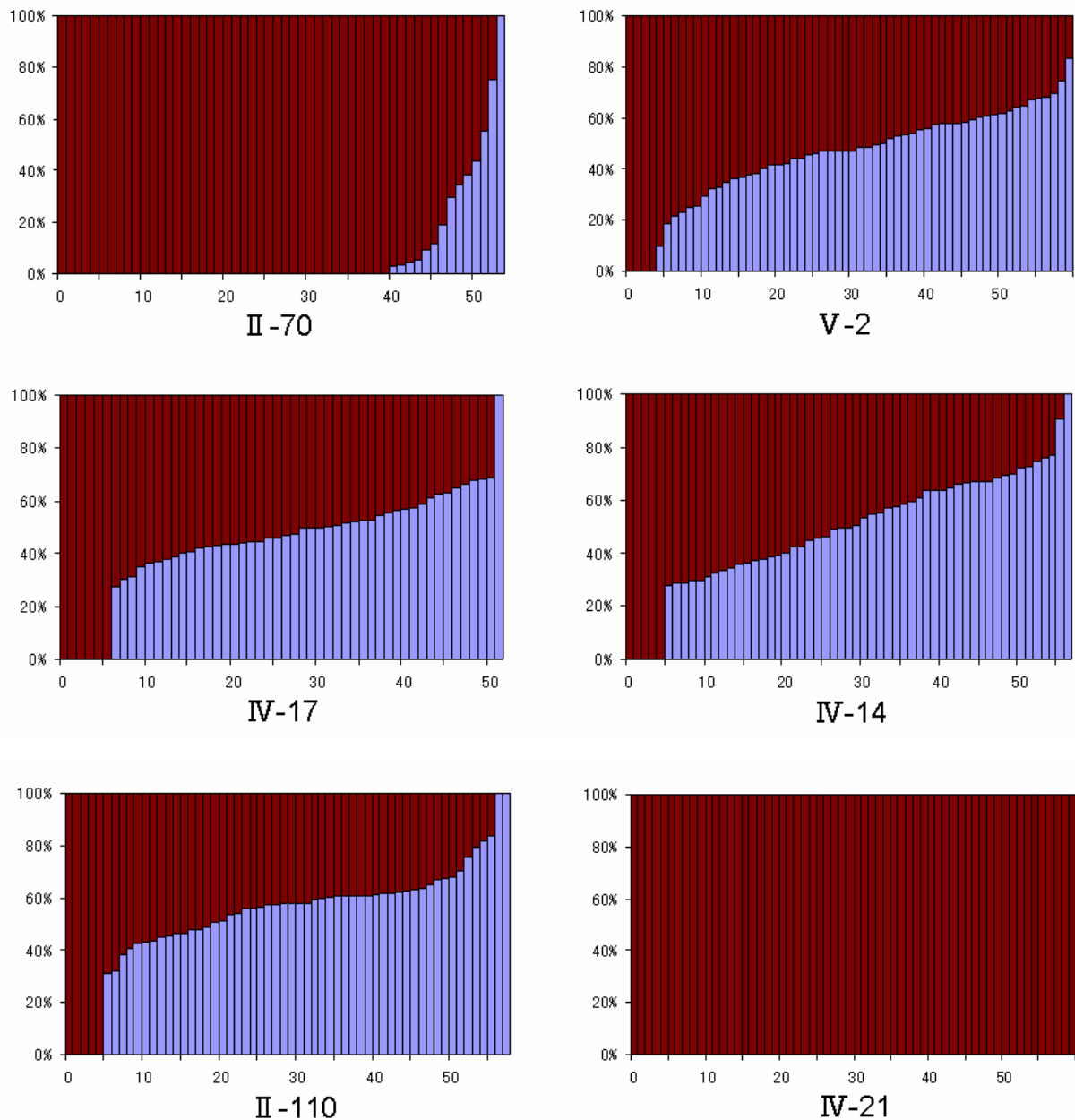


図 5 - 7: グループ の解析結果

次にグループ とグループ のゲフィチニブに対する感受性を解析するために、ゲフィチニブ投与後の全生存期間、及び無増悪期間に関して Kaplan-Meier 法による生存曲線解析、及び Log-rank 検定を行った (図 5 - 8)。グループ の全生存期間の平均値は 38 ヶ月であり、グループ の平均値は 18 ヶ月であった。グループ の全生存期間の中央値は打ち切り症例が多いため計算できなかった。また、両群を Log-rank 法により検定すると有意にグループ の生存期間が長かった ($p=0.003$)。さらに、無増悪期間においてもグループ の中央値が 22 ヶ月であるのに対し、グループ では 3 ヶ月であり、両群間に有意な差が認められた ($p=0.009$)。

次にグループ a とグループ b の全生存期間、及び無増悪期間に関して

Kaplan-Meier 法による生存曲線解析、及び Log-rank 検定を行った (図 5 - 9)。グループ a の全生存期間の平均値は 37 ヶ月であり、グループ b の平均値は 38 ヶ月であった。また、両群を Log-rank 法により検定すると両者に有意な差は無かった ($p=0.585$)。さらに、無増悪期間においてもグループ a の平均値が 17 ヶ月であるのに対し、グループ b では 21 ヶ月であり、両群間に有意な差は認められなかった ($p=0.275$)。

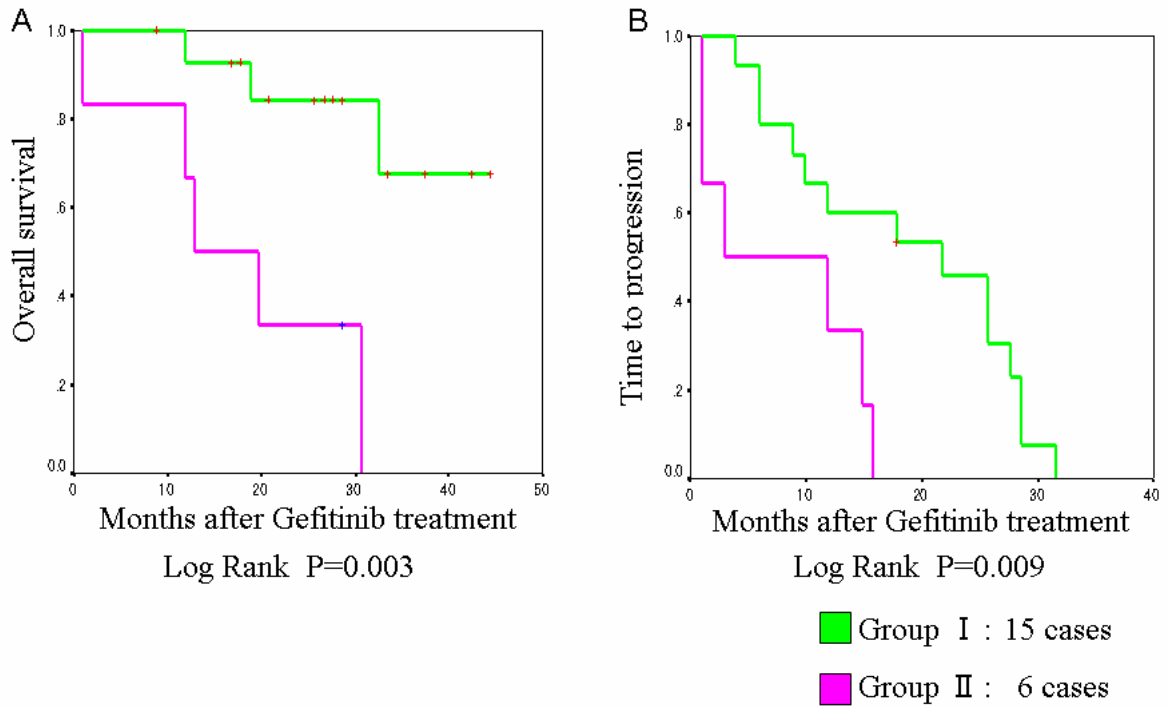


図 5 - 8 : グループ と の生存曲線解析

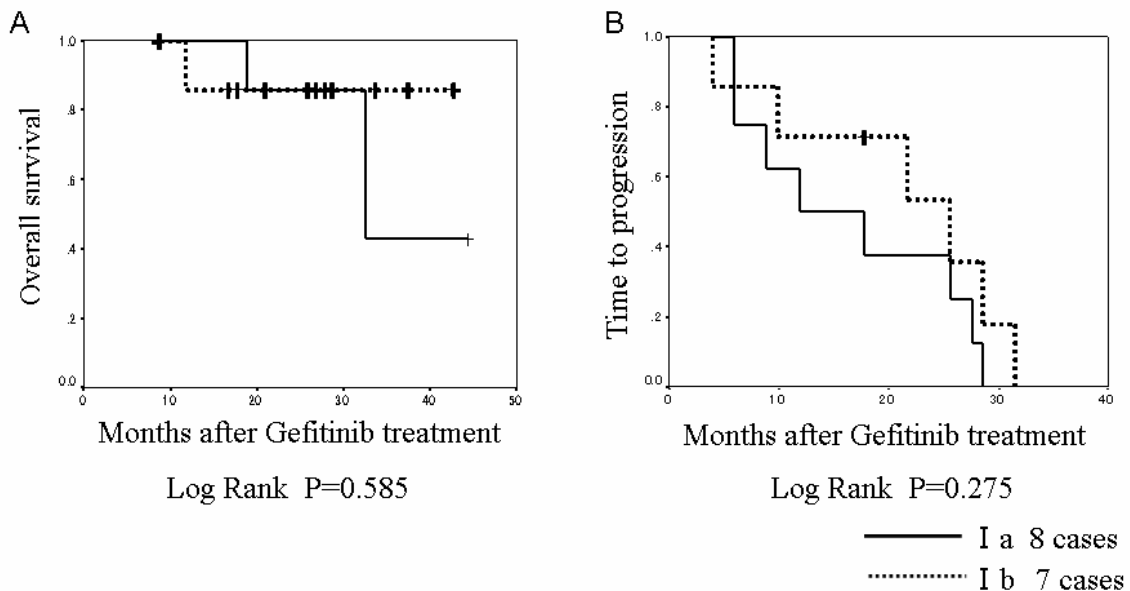


図 5 - 9 : グループ a と b の生存曲線解析

5.4. 考察

Gefinitibをはじめとする EGFR チロシンキナーゼ阻害剤は有用な抗癌剤であるが、これまでのところ一部の患者にしか効果が無い。最大限の治療効果をあげるためには、患者の選別が不可欠である。日本では、EGFR TK ドメインにおける体細胞性の変異が Gefinitib の予測効果を判定する最もよい指標となっている (Mitsudomi T.et.al.2005)。現在までの解析では、組織内の EGFR 遺伝子変異は全ての癌細胞で均一のものであるとして解析がなされてきたが、本研究において EGFR 遺伝子変異を持つ 21 症例中に 6 症例の EGFR 遺伝子変異のない癌細胞を持つ混在例が存在することを明らかになった。さらに、これらの混在例はゲフィチニブに抵抗性を示しており、ゲフィチニブ投与後の全生存期間、及び無増悪期間は全ての細胞で EGFR 遺伝子変異を持つ癌に比べると有意に短かった。EGFR 遺伝子変異のない癌細胞はゲフィチニブに反応せず、腫瘍の増殖の原因となったと考えられる。先行研究では、EGFR 遺伝子変異を持つ 39 症例中に 7 症例のゲフィチニブに抵抗性を示す症例があった (Mitsudomi T.et.al.2005)。このゲフィチニブに抵抗性を示す症例の割合は、今回の解析での混在例の割合とほぼ一致しており、ゲフィチニブへの抵抗性の主な原因が組織内の多様性にあると考えられる。現在、日本において EGFR 遺伝子変異によりゲフィチニブの感受性を予測することが一部の臨床の現場で行われているが、EGFR 遺伝子変異の多様性を調べることはより精度の高い予測を可能にすると思われる。

第三相臨床試験である Iressa survival evaluation in advanced lung cancer (ISEL) 試験は、進行非小細胞肺癌患者に対するゲフィチニブ投与群とプラセボ群の比較を 28 ヶ国、210 施設 (日本と米国は含まれていない) で行った (Thatcher N.et.al.2005)。1692 例が登録されたが、全生存期間では有意な予後の延長は認められなかった。しかし、サブセット解析では東洋人、非喫煙者の群で予後の延長が認められた。また、EGFR の遺伝子変異は人種により大きな偏りがあることが分かっており、日本人を含めた東洋人に多い。人種と EGFR 遺伝子変異の組織内多様性に関しても人種による偏りがある可能性がある。アジア以外の諸国では、ゲフィチニブの感受性と EGFR 遺伝子変異の関連が無いという報告も多いが、これには EGFR 遺伝子変異の組織内多様性が関わっている可能性がある。

ゲフィチニブ抵抗性に関しては、T790M の二次的な変異が関連していることが明らかになっている (Kobayashi S.et.al.2005, Pao W.et.al.2005)。しかし、T790M は原発巣ではほとんど確認されず、ゲフィチニブ投与後の早期の抵抗性には関連が無いと考えられる。早期の抵抗性を示す原因はこれまで分かっていなかったが、組織内多様性がゲフィチニブの早期の抵抗性の原因と考えられた。

本研究は非小細胞肺癌の原発巣組織を用いて解析を行った。ゲフィチニブの効果は再発巣の組織に対するものであり、原発巣と再発巣で遺伝子変異や組織内の多様性を保っているのかが問題となる。EGFR 遺伝子変異に関する原発巣と再発巣の比較データは無いが、Her2 の実験データから遺伝子変異や組織内の多様性を保っていることが予想される。Trastuzumab は乳癌の Her2 高発現症例に有効な Her2 を標的分子とした

分子標的薬である。この Her2 の遺伝子発現は原発巣と再発巣とで 90% から 95% 一致していた (Gancberg D.et.al.2002,Simon R.et.al.2001)。

全ての癌細胞が *EGFR* 遺伝子変異を持つグループ はその変異 *EGFR* と正常 *EGFR* の比から 2 群に分けられた。細胞内には通常、*EGFR* 遺伝子を父親由来のアレルと母親由来のアレルを持つ。一方の遺伝子に変異が入れば変異 *EGFR* と正常 *EGFR* の比は 1 対 1 となり変異 *EGFR* の割合は 5 割となる。グループ a では変異 *EGFR* の割合が 5 割に近く、一方の遺伝子に変異が入った細胞群であると考えられる。一方、グループ b では変異 *EGFR* の割合が 7 割を超えており、変異 *EGFR* 遺伝子を持つ染色体の増幅または遺伝子増幅が起こっていると考えられる。それに対して、正常 *EGFR* の割合が高い症例は無かった。このことから、変異 *EGFR* 遺伝子を持つアレルは増幅しやすい傾向があることが示唆された。*EGFR* 遺伝子の増幅は多くの症例で観察されており、*EGFR* 遺伝子の増幅がゲフィチニブの感受性に関連しているという報告もある (Takano T.et.al.2005)。ゲフィチニブ の感受性と *EGFR* 遺伝子の増幅には関連性が無く、*EGFR* 遺伝子に変異が入った症例に遺伝子増幅がよく起こっているのかもしれない。

本研究では非小細胞肺癌内の *EGFR* 遺伝子変異の多様性を示し、性質が異なる細胞が癌組織内に存在していることを明らかにした。これらの細胞の差異は分子標的薬の感受性に関連しており、組織内の多様性を明らかにすることはゲフィチニブの感受性を予測するうえで重要であると考えられる。また、他の分子標的薬に関しても、抗腫瘍効果を遺伝子変異等で予測する際には、組織内の多様性を考慮することが必要であると考えられた。

6. 結語

日本人の3人に1人が癌で亡くなっており癌研究の重要性は非常に高い。診断技術の発達や検診受診率の向上により早期発見できる割合は高くなっているが、手術による根治例は一般的に限られている。再発例など手術が不可能な場合には、放射線療法や化学療法による治療が行われる。放射線化学療法の治療成績を向上させるためには、癌の個体差を理解し適切な抗癌剤を投与することが必要である。そこで、本研究では第一に癌の個体差を理解するために遺伝子発現プロファイリングの手法を用いて解析を行った。また、現在のところ抗癌剤による癌の根治はほとんど不可能である。これは、耐性癌細胞の出現や癌幹細胞など癌組織内に多様性が存在することが原因の一つであると考えられている。そこで、第二に、癌組織内の癌細胞の多様性を理解するために、実際に一つの癌組織から多スポットの小領域を切り出し、遺伝子変異の多様性を明らかにした。

遺伝子発現プロファイリングはヒトゲノム解析に伴い広く用いられるようになったゲノムワイドな解析手法の一つである。遺伝子発現プロファイリングや遺伝子変異による癌の分類は、いわゆる癌の個別化医療において重要な役割を担っている。日本における癌の個別化医療のためのアプローチとしては SNP による薬剤代謝の違いを明らかにする試みが代表的であるが、各研究室レベルでは遺伝子発現プロファイリングによる癌の分類が盛んに行われている。SNP の解析には多くの症例数が必要であるのに対して遺伝子発現プロファイル解析は数十例から数百例の解析で行われることが多いためである。本研究では、遺伝子発現プロファイリングを用いて良性の腫瘍との術前鑑別が困難な腫瘍の一つである甲状腺濾胞癌の鑑別を試みた。60 遺伝子を用いた Weighted voting アルゴリズムによる分類器を構築し、この分類器でテストサンプルの鑑別が 83.3%の精度で可能であった。先行研究では症例数の関係から十分なテストサンプル数での解析が行われていなかったが、本研究により遺伝子発現プロファイルを用いた甲状腺濾胞癌の補助診断の可能性を示すことができたと考えている。また、遺伝子発現プロファイリング解析には RT-PCR を基本にした ATAC-PCR を用いた。遺伝子発現プロファイリングの手法として一般的にはマイクロアレイが用いられているが RT-PCR に比べると発現量の少ない遺伝子では検出感度が低い (Holland MJ, et al. 2002)。ATAC-PCR は遺伝子数が多いと実験が繁雑で時間がかかるが、遺伝子数を絞った解析を行うことはマイクロアレイを用いるより容易であり、様々な癌の解析が現在も行われている。

遺伝子発現プロファイリングのデータに対する統計解析手法も現在行われている重要な研究分野である。遺伝子発現プロファイリングを行う場合には適切な統計解析手法を用いることにより有用な情報の抽出が可能となる。一般的には2群間の比較を行う研究が多いが、癌の分類問題では様々な組織型の癌を一度に解析することが必要である。本研究では、新規多クラス分類器の開発を行い、遺伝子発現プロファイルによる多クラス分類法の改良に貢献した。分類手法の進展により、組織病理学的に分類困難な組織型等の分類が可能であることを示すことが出来たといえる。また、遺伝子

発現プロファイルのデータは本研究室が構築した CGED データベースに登録し、情報を公開している。遺伝子発現プロファイリング研究の分野では、公開済みの情報をまとめて再解析を行うようなメタ解析が行われるようになってきており、さらに大規模なデータベースの構築や遺伝子のオントロジー等の情報との統合により、正確で有用な知見の獲得と世界中の研究者がその情報にアクセスできるようなシステム開発による研究の発展が期待される。データ取得に関する再現性の問題など解決すべき課題は残されているが、遺伝子発現プロファイルは組織病理学的な診断の補助的手段として利用が進んでいくものと考えられる。

癌の個体差は癌細胞の多様性を示しているが、一つの腫瘍の中にも癌細胞の多様性があることが明らかとなっている。癌は一個の癌細胞から発生すると考えられており、癌が進展する過程で新たな変異を獲得すると共に癌細胞の多様性も獲得する。癌組織中の癌細胞がどの程度の速さで新たな変異を獲得し、その変異がどのように腫瘍の中で広がっていくのかは癌細胞の性質を理解する上で重要な研究課題である。また、現在注目を集めている癌幹細胞に関する研究分野も癌細胞の多様性の一端を研究対象にしている。多様性の程度が癌の悪性度と関連しているとの研究報告もあり (Roth MJ, et al. 2001) 多様性そのものを判定する技術の進展も望まれている。今回、Laser Micro Dissection と SNaPshot を組み合わせた独自の実験手法により非小細胞肺癌内の *EGFR* 遺伝子変異の多様性を明らかにし、これがゲフィチニブの感受性と関連していることを初めて報告した。この結果は、分子標的薬の抗癌剤感受性に組織内多様性の理解が重要であることを示している。非小細胞肺癌に対する分子標的薬はゲフィチニブと同様に *EGFR* を標的分子とするエルロチニブの他に、EKB-559 や Canertinib のように他の ErbB ファミリーも同時に標的としている薬剤、また、ZD6474 のように *EGFR* と vascular endothelial growth factor を標的とする薬剤などが臨床試験中であり、これらの分子標的薬の感受性と分子プロファイルとの関連性の解析結果が待たれる。

遺伝子発現プロファイリングの研究及び、組織内多様性の研究により癌の多様性の理解と分子プロファイルに基づく分類に貢献することができたと考えている。分子プロファイリングによる癌の分類は今後さらに進展し、個別化医療が着実に実現されていくものと思われる。特に、現在多くの薬が開発されている分子標的薬に対する抗腫瘍効果の予測には標的分子の状態把握が必須であり、他の抗癌剤に関しても今後は組織内の多様性を考慮したより精度の高い分類手法の確立が望まれる。さらに、化学療法による癌の根治を目指す場合には組織内の多様性の理解無しには不可能であると考えられ、遺伝子変異の多様性の理解が組織内の癌細胞の多様性を理解する上で重要であることは間違いない。また、遺伝子発現プロファイリングは癌の分類のみならず新規抗癌剤の候補の同定や、抗癌剤の薬物代謝の評価など薬剤開発にも用いられており、新規多クラス分類器の活用方法は幅広い。今後、遺伝子変異、増幅、発現、タンパク発現等の分子プロファイルの解析精度が向上し、選択できる抗癌剤の種類を増加することで癌の治療成績が向上していくことが期待される。

7. 謝辞

論文主査をお引き受け頂きました竹家達夫教授、論文審査をお引き受け頂きました川市正史教授に感謝いたします。また、奈良先端科学技術大学院大学大正製薬ゲノム機能解析学講座、大阪府立成人病センター研究所において、松尾浩子博士、小泉恭子博士、齋藤さかえ博士、川本祥子博士に実験指導及び適切な助言を頂きました。本研究で使用した甲状腺癌の検体は、大阪大学医学部乳腺内分泌外科、野口眞三郎教授に提供していただきました。また、肺癌の検体は、大阪府立成人病センター呼吸器外科、東山聖彦先生、岡見次郎先生に提供して頂き、共同研究者として多くの助言を頂きました。肺癌組織の正常細胞と癌細胞の識別に関しては、病理医の西澤恭子博士に指導して頂きました。また、奈良先端科学技術大学院大学情報科学研究科論理生命学分野の行縄直人博士、大羽成征助手、石井信教授には多クラス分類器の共同研究のみならず、統計解析手法について指導していただきました。この場を借りて、厚く御礼申し上げます。並びに御指導、御協力を頂きました疾患分子遺伝学講座の皆様、大正製薬ゲノム機能解析学講座の先輩方を始め、本学での五年間に及ぶ研究生活を支えて下さった多くの方々、両親に心から感謝いたします。

最後になりましたが、本研究を進めるにあたり、御指導、御鞭撻を賜った大阪府立成人病センター研究所加藤菊也所長に深く感謝申し上げます。

8 . 参考文献

Allwein E.L., Schapire R.E., Singer Y. Reducing multiclass to binary: A unifying approach for margin classifiers. (2000). In Proc 17th International Conf on Machine Learning. 9-16.

Alon U., Barkai N., Notterman D.A., Gish K., Ybarra S., Mack D., and Levine A.J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA. 96,6745-6750.

Arriagada R., Bergman B., Dunant A., Le Chevalier T., Pignon J.P., Vansteenkiste J.; International Adjuvant Lung Cancer Trial Collaborative Group. (2004). Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. N Engl J Med. 350(4),351-360.

Baloch Z.W., Fleisher S., LiVolsi V.A., Gupta P.K. (2002). Diagnosis of `follicular neoplasm` a gray zone in thyroid fine needle aspiration cytology. Diagn Cytopathol. 26,41-44.

Balsara B.R., Testa J.R. (2002). Chromosomal imbalances in human lung cancer. Oncogene. 21(45),6877-6883.

Barden C.B., Shister K.W., Zhu B., Guiter G., Greenblatt D.Y., Zeiger M.A., Fahey T.J. 3rd. (2003). Classification of follicular thyroid tumors by molecular signature: results of gene profiling. Clin Cancer Res. 9,1792-1800.

Beer D.G., Kardia S.L., Huang C.C., Giordano T.J., Levin A.M., Misek D.E., Lin L., Chen G., Gharib T.G., Thomas D.G., Lizyness M.L., Kuick R., Hayasaka S., Taylor J.M., Iannettoni M.D., Orringer M.B., Hanash S. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med. 8,816-824.

Bodner S.M., Minna J.D., Jensen S.M., D'Amico D., Carbone D., Mitsudomi T., Fedorko J., Buchhagen D.L., Nau M.M., Gazdar A.F., et al. (1992). Expression of mutant p53 proteins in lung cancer correlates with the class of p53 gene mutation. Oncogene. 7(4),743-749.

Cerutti J.M., Delcelo R., Amadei M.J., Nakabashi C., Maciel R.M., Peterson B., Shoemaker J., Riggins G.J. (2004). A preoperative diagnostic test that distinguishes benign from malignant thyroid carcinoma based on gene expression. J Clin Invest. 113,1234-1242.

Chomczynski P., Sacchi N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* 162,156-9.

Ciardiello F., De Vita F., Orditura M., Tortora G. (2004). The role of EGFR inhibitors in nonsmall cell lung cancer. *Curr Opin Oncol.* 16,130-135.

Ciardiello F., Tortora G. (2001). A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor. *Clin Cancer Res.* 7,2958-2970.

Cohen S., Carpenter G., King L. Jr. (1980). Epidermal growth factor-receptor-protein kinase interactions. Co-purification of receptor and epidermal growth factor-enhanced phosphorylation of activity. *J Biol Chem.* 255:4834-4842.

Crammer K., Singer Y. (2001). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research.* 2,265-291.

Devesa S.S., Shaw G.L., Blot W.J. (1991). Changing patterns of lung cancer incidence by histological type. *Cancer Epidemiol Biomarkers Prev.* 1,29-34.

Dietterich T.G., Bakiri G. (1995). Solving multiclass learning problem via errorcorrecting output codes. *Journal of Artificial Intelligence Research.* 2,263-286.

Duda R.O., Hart P.E., Stork D.G. (2000). *Pattern Classification.* Wiley Interscience.

Dziedziszko R., Witta S.E., Cappuzzo F., Park S., Tanaka K., Danenberg P.V., Baron A.E., Crino L., Franklin W.A., Bunn P.A. Jr, Varella-Garcia M., Danenberg K.D., Hirsch FR. (2006). Epidermal growth factor receptor messenger RNA expression, gene dosage, and gefitinib sensitivity in non-small cell lung cancer. *Clin Cancer Res.* 12(10),3078-3084.

Friend S.H., Bernards R., Rogelj S., Weinberg R.A., Rapaport J.M., Albert D.M., Dryja T.P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature.* 323(6089):643-646.

Freund Y., Schapire R.E. (1996). Experiments with a new boosting algorithm. *International Conference on Machine Learning.* 148-156.

Fukuoka M., Yano S., Giaccone G., Tamura T., Nakagawa K., Douillard J.Y., Nishiwaki Y., Vansteenkiste J., Kudoh S., Rischin D., Eek R., Horai T., Noda K., Takata I., Smit E., Averbuch S., Macleod A., Feyereislova A., Dong R.P., Baselga J. (2003). Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial). *J Clin Oncol.* 21(12),2237-2246.

Gancberg D., Di Leo A., Cardoso F., Rouas G, Pedrocchi M., Paesmans M., Verhest A., Bernard-Marty C., Piccart M.J., Larsimont D. (2002). Comparison of HER-2 status between primary breast cancer and corresponding distant metastatic sites. *Ann Oncol.* 13(7),1036-1043.

Garber M.E., Troyanskaya O.G, Schluens K., Petersen S., Thaesler Z., Pacyna-Gengelbach M., van de Rijn M., Rosen G.D., Perou C.M., Whyte R.I., Altman R.B., Brown P.O., Botstein D., Petersen I. (2002). Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A.*99,1098.

Gerhold, D.L., Jensen R.V., Gullans S.R. (2002) . Better therapeutics through microarrays. *Nat.Genet.*32 supplement,547-552.

Giaccone G, Herbst R.S., Manegold C., Scagliotti G, Rosell R., Miller V., Natale R.B., Schiller J.H., Von Pawel J., Pluzanska A., Gatzemeier U., Grous J., Ochs J.S., Averbuch S.D., Wolf M.K., Rennie P., Fandi A., Johnson D.H. (2004). gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 1. *J Clin Oncol.* 22(5),777-784.

Gonzalez-Garcia I., Sole R.V., Costa J. (2002). Metapopulation dynamics and spatial heterogeneity in cancer. *Proc Natl Acad Sci U S A.* 99(20),13085-13089.

Greaves T.S., Olvera M., Florentine B.D., Raza A.S., Cobb C.J., Tsao-Wei D.D., Groshen S., Singer P., Lopresti J., Martin S.E. (2000). Follicular lesions of thyroid: a 5-year fine-needle aspiration experience. *Cancer.*90,335-341.

Hajra A., Liu P.P., Wang Q., Kelley C.A., Stacy T., Adelstein R.S., Speck N.A., Collins F.S. (1995). The leukemic core binding factor beta-smooth muscle myosin heavy chain (CBF beta-SMMHC) chimeric protein requires both CBF beta and myosin heavy chain domains for transformation of NIH 3T3 cells. *Proc Natl Acad Sci U S A.* 92,1926-1930.

Hastie T., Tibshirani R. (1998). Classification by pairwise coupling. In *Advances in Neural Information Processing System*. 10,507-513.

Hedenfalk I., Ringner M., Ben-Dor A., Yakhini Z., Chen Y., Chebil G., Ach R., Loman N., Olsson H., Meltzer P., Borg A., Trent J. (2003). Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc Natl Acad Sci U S A*. 100(5),2532-2537.

Herbst R.S., Giaccone G., Schiller J.H., Natale R.B., Miller V., Manegold C., Scagliotti G., Rosell R., Oliff I., Reeves J.A., Wolf M.K., Krebs A.D., Averbuch S.D., Ochs J.S., Grous J., Fandi A., Johnson D.H. (2004). gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 2. *J Clin Oncol*. 22(5),785-94.

Higashiyama M., Kodama K., Yokouchi H., Takami K., Doi O., Kobayashi H., Tanisaka K., Minamigawa K. (1998). Immunohistochemical p53 protein status in nonsmall cell lung cancer is a promising indicator in determining in vitro chemosensitivity to some anticancer drugs. *J Surg Oncol*. 68(1),19-24.

Histological Typing of Lung and Pleural Tumours. (1999) 3rd edition, International Histological Classification of Tumours, World Health Organization.

Holland M.J. (2002). Transcript abundance in yeast varies over six orders of magnitude. *J Biol Chem*. 277,14363-14366.

Huang S.F., Liu H.P., Li L.H., Ku Y.C., Fu Y.N., Tsai H.Y., Chen Y.T., Lin Y.F., Chang W.C., Kuo H.P., Wu Y.C., Chen Y.R., Tsai S.F. (2004). High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell Lung Cancers related to gefitinib responsiveness in Taiwan. *Clin Cancer Res*.10,8195-8203.

Huang Y., Prasad M., Lemon W.J., Hampel H., Wright F.A., Kornacker K., LiVolsi V., Frankel W., Kloos R.T., Eng C., Pellegata N.S., de la Chapelle A. (2001). Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc Natl Acad Sci U S A*.98,15044-15049.

IARC TP53 Mutation Database, International Agency for Research on Cancer, <http://www-p53.iarc.fr/>

Ichinose Y., Yano T., Yokoyama H., Inoue T., Asoh H., Tayama K., Takanashi N. (1994). Postrecurrent survival of patients with non-small-cell lung cancer undergoing a complete resection. *J Thorac Cardiovasc Surg.* 108(1),158-161.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature.* 409,860-921.

Ippolito A., Vella V., La Rosa G.L., Pellegriti G., Vigneri R., Belfiore A. (2001). Immunostaining for Met/HGF receptor may be useful to identify malignancies in thyroid lesions classified suspicious at fine-needle aspiration biopsy. *Thyroid.* 11,783-787.

Iwao K., Matoba R., Ueno N., Ando A., Miyoshi Y., Matsubara K., Noguchi S., Kato K. (2002). Molecular classification of primary breast tumors possessing distinct prognostic properties. *Hum Mol Genet.* 11,199-206.

Janne P.A., Gurubhagavatula S., Yeap B.Y., Lucca J., Ostler P., Skarin A.T., Fidias P., Lynch T.J., Johnson B.E. (2004). Outcomes of patients with advanced non-small cell lung cancer treated with gefitinib (ZD1839, "Iressa") on an expanded access study. *Lung Cancer.* 44,221-230

Kato K. (1997). Adaptor-tagged competitive PCR: a novel method for measuring relative gene expression. *Nucleic Acids Res.* 25,4694-4696.

Khan J., Wei J.S., Ringner M., Saal L.H., Ladanyi M., Westermann F., Berthold F., Schwab M., Antonescu C.R., Peterson C., Meltzer P.S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med.* 7(6),658-659.

Kosaka T., Yatabe Y., Endoh H., Kuwano H., Takahashi T. and Mitsudomi T. (2004). Mutations of the epidermal growth factor receptor gene in Lung Cancer : Biological and Clinical Implications. *Cancer Res.*64,8919-8923.

Kris M.G., Natale R.B., Herbst R.S., Lynch T.J. Jr, Prager D., Belani C.P., Schiller J.H., Kelly K., Spiridonidis H., Sandler A., Albain K.S., Cella D., Wolf M.K., Averbuch S.D., Ochs J.J., Kay A.C.. (2003). Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA.* 290(16),2149-58.

Kurokawa Y., Matoba R., Nagano H., Sakon M., Takemasa I., Nakamori S., Dono K., Umeshita K., Ueno N., Ishii S., Kato K., Monden M. (2004). Molecular prediction of response to 5-fluorouracil and interferon-alpha combination chemotherapy in advanced hepatocellular carcinoma. *Clin Cancer Res.* 10(18),6029-6038.

Kurokawa Y., Matoba R., Takemasa I., Nakamori S., Tsujie M., Nagano H., Dono K., Umeshita K., Sakon M., Ueno N., Kita H., Oba S., Ishii S., Kato K., Monden M. (2003). Molecular features of non-B, non-C hepatocellular carcinoma: a PCR-array gene expression profiling study. *J Hepatol.* 39(6),1004-1012.

Levinson A.D., Oppermann H., Levintow L., Varmus H.E., Bishop J.M. (1978). Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein. *Cell.* 2.561-572.

Li T., Zhang C., Ogihara M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics.* 20(15),2429-2437.

Lynch T.J., Bell D.W., Sordella R., Gurubhagavatula S., Okimoto R.A., Brannigan B.W., Harris P.L., Haserlat S.M., Supko J.G., Haluska F.G. et.al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med*;350:2129-2139.

MacDonald T.J., Brown K.M., LaFleur B., Peterson K., Lawlor C., Chen Y., Packer R.J., Cogen P., Stephan D.A. (2001). Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Genet.* 29,143-152.

Maley C.C., Galipeau P.C., Finley J.C., Wongsurawat V.J., Li X., Sanchez C.A., Paulson T.G., Blount P.L., Risques R.A., Rabinovitch P.S., Reid B.J. (2006). Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet.* 38. (4),468-473.

Mazzaferri E.L. (1993). Management of a solitary thyroid nodule. *N Engl J Med.* 328,553-559.

Motoori M., Takemasa I., Yano M., Saito S., Miyata H., Takiguchi S., Fujiwara Y., Yasuda T., Doki Y., Kurokawa Y., Ueno N., Oba S., Ishii S., Monden M., Kato K. (2005). Prediction of recurrence in advanced gastric cancer patients after curative resection by gene expression profiling. *Int J Cancer.* 114(6),963-968

Muro S., Takemasa I., Oba S., Matoba R., Ueno N., Maruyama C., Yamasita R., Sekimoto M., Yamamoto H., Nakamori S., Monden M., Isii S., and Kato K. (2003). Identification of expressed genes linked to malignancy of human colorectal carcinoma by parametric clustering of quantitative expression data. *Genome Biology*.4,R21,1-10

Murren J., Glatstein E., Pass H.I. (2001). Small cell lung cancer. In: DeVita V.T., Hellman S., Rosenberg S.A., eds. *Cancer Principles and Practice of Oncology*, 6th edn. Philadelphia: Lippincott Williams & Wilkins, 983–1018.

Nielsen H.L., Ronnov-Jessen L., Villadsen R., Petersen O.W. (2002). Identification of EPSTI1, a novel gene induced by epithelial-stromal interaction in human breast cancer. *Genomics*. 79(5),703-710.

Noel M., Delehaye M.C., Segond N., Lasmoles F., Caillou B., Gardet P., Fragu P., Moukhtar M.S. (1991). Study of calcitonin and thyroglobulin gene expression in human mixed follicular and medullary thyroid carcinoma. *Thyroid*. 1,249-256.

Griffith O.L., Melck A., Jones S.J., Wiseman S.M. (2006). Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *J Clin Oncol*. 24(31),5043-5051.

Oba S., Sato M., Ishii S. (2000). Variational Bayes method for Mixture of Principal Component Analyzers. In *Proceeding for 7th International Conference on Neural Information Proceeding*. 2,1416-1421.

Paez J.G., Janne P.A., Lee J.C., Tracy S., Greulich H., Gabriel S., Herman P., Kaye F.J., Lindeman N., Boggon T.J, et.al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*.304(5676),1497-1500.

Parkin D.M., Bray F., Ferlay J., Pisani P. (2005). Global cancer statistics, 2002. *CA Cancer J Clin*. 55(2),74-108.

Pao W., Miller V., Zakowski M., Doherty J., Politi K., Sarkaria I., Singh B., Heelan R., Rusch V., Fulton L., Mardis E., Kupfer D., Wilson R., Kris M., Varmus H. (2004) EGF receptor gene mutations are common in lung cancer from “never smokers” and are associated with sensitivity of gefitinib and erlotinib. *Proc Natl Acad Sci U S A*; 101:13306-13311.

Pomeroy S.L., Tamayo P., Gaasenbeek M., Sturla L.M., Angelo M., McLaughlin M.E., Kim J.Y., Goumnerova L.C., Black P.M., Lau C., Allen J.C., Zagzag D., Olson J.M., Curran T., Wetmore C., Biegel J.A., Poggio T., Mukherjee S., Rifkin R., Califano A., Stolovitzky G., Louis D.N., Mesirov J.P., Lander E.S., Golub T.R. (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*. 415,436-442.

Ramaswamy S., Tamayo P., Rifkin R., Mukherjee S., Yeang C.H., Angelo M., Ladd C., Reich M., Latulippe E., Mesirov J.P., Poggio T., Gerald W., Loda M., Lander E.S., Golub T.R. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A*. 98 (26),15149-15154.

Roth M.J., Hu N., Emmert-Buck M.R., Wang Q.H., Dawsey S.M., Li G., Guo W.J., Zhang Y.Z., Taylor P.R. (2001). Genetic progression and heterogeneity associated with the development of esophageal squamous cell carcinoma. *Cancer Res*. 61(10),4098-4104.

Ruggeri R.M., Villari D., Simone A., Scarfi R., Attard M., Orlandi F., Barresi G., Trimarchi F., Trovato M., Benvenga S. (2002). Co-expression of interleukin-6 (IL-6) and interleukin-6 receptor (IL-6R) in thyroid nodules is associated with co-expression of CD30 ligand/CD30 receptor. *J Endocrinol Invest*. 25,959-966.

Rusch V., Baselga J., Cordon-Cardo C., Orazem J., Zaman M., Hoda S., McIntosh J., Kurie J., Dmitrovsky E. (1993). Differential expression of the epidermal growth factor receptor and its ligands in primary non-small cell lung cancers and adjacent benign lung. *Cancer Res*. 53,2379-2385.

Samowitz W.S., Slattery M.L. (1999). Regional reproducibility of microsatellite instability in sporadic colorectal cancer. *Genes Chromosomes Cancer*. 26(2),106-114.

Scagliotti G.V., Fossati R., Torri V., Crino L., Giaccone G., Silvano G., Martelli M., Clerici M., Cignetti F., Tonato M.; Adjuvant Lung Project Italy/European Organisation for Research Treatment of Cancer-Lung Cancer Cooperative Group Investigators. (2003). Randomized study of adjuvant chemotherapy for completely resected stage I, II, or IIIA non-small-cell Lung cancer. *J Natl Cancer Inst*. 95(19),1453-1461.

Scholkopf B., Burges C., Smola A. (1999). *Advances in Kernel Methods Support Vector Learning*. The MIT Press.

Scholkopf B., Burges C., Vapnik V. (1995). Extracting support data for a given task. In Proceedings of the First International Conference on Knowledge Discovery and Data Mining. 252-257.

Sherman S.I. (2003). Thyroid carcinoma. *Lancet*. 361(9356),501-511.

Simon R., Nocito A., Hubscher T., Bucher C., Torhorst J., Schraml P., Bubendorf L., Mihatsch M.M., Moch H., Wilber K., Schotzau A., Kononen J., Sauter G. (2001). Patterns of her-2/neu amplification and overexpression in primary and metastatic breast cancer. *J Natl Cancer Inst*. 93(15),1141-1146.

Sordella R., Bell D.W., Haber D.A., Settleman J. (2004). Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science*.305,1163-1167.

Statnikov A., Aliferis C.F., Tsamardinos I., Hardin D., Levy S. (2005). A comprehensive evaluation of multiclassification methods for microarray gene expression cancer diagnosis. *Bioinformatics*. 21(5),631-643.

Takano T., Ohe Y., Sakamoto H., Tsuta K., Matsuno Y., Takeishi U., Yamamoto S., Nokihara H., Yamamoto N., Sekine I., Kunitoh H., Shibata T., Sakiyama T., Yoshida T. and Tamura T. (2005). Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. *J Clin Oncol.*; 23(28):6829-6837.

Takano T., Miyauchi A., Matsuzuka F., Yoshida H., Kuma K., Amino N. (2003). Ubiquitous expression of galectin-3 mRNA in benign and malignant thyroid tumors. *Cancer Lett*. 199,69-73.

Takano T., Miyauchi A., Yoshida H., Kuma K., Amino N. (2004). High-throughput differential screening of mRNAs by serial analysis of gene expression: Decreased expression of trefoil factor 3 mRNA in thyroid follicular carcinomas. *Br J Cancer*. 90(8),1600-1605

Tanaka T., Umeki K., Yamamoto I., Sakamoto F., Noguchi S., Ohtaki S. (1995). CD26 (dipeptidyl peptidase IV/DPP IV) as a novel molecular marker for differentiated thyroid carcinoma. *Int J Cancer*. 64,326-331.

Taron M., Ichinose Y., Rosell R., Mok T., Massuti B., Zamora L., Mate J.L., Manegold C., Ono M., Queralt C., Jahan T., Sanchez J.J., Sanchez-Ronco M., Hsue V., Jablons D., Sanchez J.M., Moran T. (2005). Activating mutations in the tyrosine kinase domain of the epidermal growth factor receptor are associated with improved survival in gefitinib-treated chemorefractory lung adenocarcinomas. *Clin Cancer Res.* 11(16),5878-5885.

Tax D., Duin R.P.W., Using two-class classifiers for multi-class classification. (2002). In *Proceedings 16th International Conference on Pattern Recognition.* 2,124-127.

Therasse P., Arbuck S.G., Eisenhauer E.A., Wanders J., Kaplan R.S., Rubinstein L., Verweij J., Van Glabbeke M., van Oosterom A.T., Christian M.C., Gwyther S.G. (2000). New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst.* 92(3),205-216.

Thatcher N., Chang A., Parikh P., Rodrigues Pereira J., Ciuleanu T., von Pawel J., Thongprasert S., Tan E.H., Pemberton K., Archer V., Carroll K. (2005). gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet.* 366(9496),1527-37.

Tibshirani R., Hastie T., Narasimhan B., Chu G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A.* 99(10),6567-6572.

Tokumo M., Toyooka S., Kiura K., Shigematsu H., Tomii K., Aoe M., Ichimura K., Tsuda T., Yano M., Tsukuda K., Tabata M., Ueoka H., Tanimoto M., Date H., Adi F. Gazdar and Shimizu N. (2005). The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell Lung Cancers. *Clin Cancer Res.* 11,1167-1173.

Tonon G., Wong K.K., Maulik G., Brennan C., Feng B., Zhang Y., Khattry D.B., Protopopov A., You M.J., Aguirre A.J., Martin E.S., Yang Z., Ji H., Chin L., Depinho R.A. (2005). High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci U S A.* 102(27),9625-9630.

van 't Veer L.J., Dai H., van de Vijver M.J., He Y.D., Hart A.A., Mao M., Peterse H.L., van der Kooy K., Marton M.J., Witteveen A.T., Schreiber G.J., Kerkhoven R.M., Roberts C., Linsley P.S., Bernards R., Friend S.H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 415,530-536.

Vapnik V. (1998). *Statistical Learning Theory*. Wiley, NY.

Weston J., Watkins C. (1998). Multi-class support vector machine. Technical report, University of London.

Yatabe Y., Konishi H., Mitsudomi T., Nakamura S., Takahashi T. (2000). Topographical distributions of allelic loss in individual non-small-cell lung cancers. *Am J Pathol.* 157(3),985-993.

Yukinawa N. (2006). 統計的手法による遺伝子発現情報からの細胞状態の同定に関する研究. 奈良先端科学技術大学院大学 博士論文 is-DD0361037.

Zhou J., Wang H., Lu A., Hu G., Luo A., Ding F., Zhang J., Wang X., Wu M., Liu Z. (2002). A novel gene, NMES1, downregulated in human esophageal squamous cell carcinoma. *Int J Cancer.* 101,311-316.

Zhu K., Wang J., Zhu J., Jiang J., Shou J., Chen X. (1999). p53 induces TAP1 and enhances the transport of MHC class I peptides. *Oncogene.* 18,7740-7747.

Zou T.T., Selaru F.M., Xu Y., Shustova V., Yin J., Mori Y., Shibata D., Sato F., Wang S., Olaru A., Deacu E., Liu T.C., Abraham J.M., Meltzer S.J. (2002). Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene.* 21,4855-4862.