あって、受診されている方が、70～80%。日本の乳癌検診受診率は、10%ちょっと。70%と10%の差ってすごいです。全然桁違い。ですからこの受診率を上げていくということ、これがとっても大切なことになります。検診の施設は検診精度を高める努力をする。市町村では女性の乳癌死を減らすという目的のために、いくらお金が出せるかという話が出てくると思うのですけど、やはり意識を高めて、是非受けたいという声が出てくるように指導をして頂いて、せめて30、40、50%というふうに受診率が上がるような努力、これが必要と思います。

乳癌死を減らすためには、医療関係者は精度の高い検診、あるいは病院勤務者は精密検査の充実、治療法をきちんとするということ、こういう努力が必要ですし、また保健師さんとか、そうでない方も含めて国民的な課題というふうに考えて、高い検診受診率にするために、「受診行った？」「マンモグラフィ撮った？」というのが合い言葉になるような、そういう空気を作っていく運動、これも必要だろうと思います。

乳癌罹患はまだまだ増えます。増えても乳癌死亡が減ったというところが、この乳癌検診の目標であります。これを保健師さんはじめ市町村健康管理担当者の目標として、是非一緒に持って頂きたいということをお願いして私のお話を終わらせて頂きます。ご静聴ありがとうございました。

青木先生：

いたれりつくせりのお話、ありがとうございました。私なりにまとめさせていただきますと、1つは乳癌患者がどんどん増えていると、つまり若い年齢で増えているということ、30代以上で増えている。それから死亡を減らすには検診の効率が良いということでした。ただ検診には適切な機器を使って、しっかりした技術を持った人が良い写真を撮って、正確に読影をすること、そしてその結果、精密検診が必要な人はすぐ受診すること、又マンモグラフィの専門医のおられる施設に行っていただくことが非常に大事である。ということでした。マンモグラフィの読影は容易ではありません。、私は素人ですので先日から一生懸命見ておりますけれども、なかなか正確に判断できないのです。やはり専門家に任せるよりしょうがないと思います。そういうことで、今後皆様方は、対象の女性に対して受診を勧めるということ、受診を勧めるならばしっかりした所で、撮ってもらって、しっかりしたところで診断してもらうこと。健康管理をやってらっしゃる方は、後のことも考え、1年後2年後3年後どうなったかについて気をくばってやっていただきますとたいへん役に立つと思います。

先生本当にどうもありがとうございました。

質問がありましたらお受けします。

質問者：

先生どうもありがとうございました。私は岐阜県の神戸町という所で働いています保健師ですが、30代の女性の方で授乳をされている方からお電話がありまして、乳癌検診マンモグラフィを受けたいのですが、まだ授乳中ということでその赤ちゃんに対しての影響がないかということが心配というお電話がありまして、先ほどのお話でほとんど影響がないということは分かったのですが、そのリスクから考えて、その30代ということで、乳腺の問題とかでその検診を受けた方がいいのか、授乳中でも受けてもらった方が良いとお勧めした方が良いのか、その辺がリスクと効果というところでどうなのかということをお聞きしたいのでお願いします。

遠藤先生：

分かりました。30歳代で、特に授乳していると言いますと、どういうおっぱいの状態になっているかと言いますと、写真では真っ白な状態です。普段そうでもない方がこうなってしまう。どうしてこうなるかと言うと、乳腺が大きくなる。中にミルクがある。つまり水分がいっぱいあるのです。先ほどほとんど脂肪ばかりというのから順番に見ていただきました。1番最後の真っ白のがありました。あれよりもっと真っ白になるのです。だから中にもし病気があっても、発見できる率と発見できない率を考えると、発見できないことがいっぱいあるのです。だからまず効果があまりない。効果がないのに少ない量であってもX線を浴びせて良いのかということになると、効果がないのに浴びせるというのは、要するに害しかないわけです。全く分からないかと言うと、何かよほど大きなものがあった時には見えることがあるのですけど、逆のこともあるのです。もし中にあるのだけれども見えなかった。大丈夫ですよと言ってしまって、逆に手遅れになってしまう。そういうこともあるので、むしろやらない方が良いというのが結論です。私、今40歳代というふうなことでマ

ンモグラフィ検診が有効ですという話をしました。30歳代には全く触れませんでした。今の女性のライフスタイルを見ていきますと、30を過ぎてからぼちぼち結婚を考えて、35歳になりそうだから結婚する。40になりそうだから出産するというのが、けっこうな頻度だと思うのです。もちろん早い方は20代になる前からでも出産される方はいらっしゃるのですけど、頻度的に言うと平均の結婚年齢って、相当上がってます。そうなりますと30歳代というのはもうまさに妊娠出産授乳の真っ最中ということで、あまり撮った方が良いというふうにお勧め出来る方が多くない。それとそういうような状況に無くても、まだ30歳代というのは、乳腺が非常に充実している方が多い。一方、この前あったことなのですけど、28歳で、大きなおっぱいをしてみえて、マンモグラフィを撮って、そこにちょっと影があって、超音波をやってよく分からないけど、これだと思って刺したら癌が出たという人もいるので、全員が全員無効というわけではないのですが、大勢を対象にしてお勧めするかどうかというお話になると、お勧めしないです。では20歳代、20歳後半から30歳代、40歳くらいまでどうするのかということなんですけど、１番有効なのは、自分で自分のおっぱいの状態を分かっていただく自己触診です、自己検診をいかに上手にしてもらうかっていう指導、これが１番だと思います。この頃この辺少し硬い気がするって言われた方、ほんとうにこの前あったんですが、硬い気がするだけ。でも若い人はけっこう硬い乳腺が普通というのがあるんです。だから医者が触診をして、けっこう乳腺をたくさん触っている私が触診をして、少し硬いけど乳腺症でも十二分に説明が付くと思う。超音波をやりました。乳腺症でも良いけど、乳腺症の所見と普通読みます。でも本人がそう言っていらっしゃるし、ちょっと念のために針を刺して細胞を見ましょう。超音波で見ながら針を刺してみました。癌が疑われますという結果が来ました。そういうことがあるのです。そこで何が癌の発見に有効だったかというと自分で触ってた。自分の体ですから自分のちょっとした変化というのはすごく敏感に感じ取られるんです。ですから正しい触診の仕方というのを保健師さんには修得していただいて広めてもらうということが一番良いだろうと思います。25歳を過ぎたら自分のおっぱいに対して自分で触診をする。こういう習慣を付けてもらう。これは是非保健師さんに働いて頂きたいと思ってます。そして異常があったら受診してもらって、それも、乳腺をしっかり勉強した医者がいるところに行って頂くということをお勧めして欲しいと思います。

質問者：

愛知県の海部郡の蟹江町で保健師をしております。一番聞きたい部分なんですけれども、テストに合格された方々が所属していらっしゃる機関が公表されることはございますでしょうか。

遠藤先生：

マンモグラフィ検診精度管理中央委員会、精中委です。ここがホームページを持っております。聖マリアンナ医科大学の乳腺内分泌外科ですが、その聖マリアンナ医科大学というところから入って頂いてもすぐに入れますし、それから最近はＮＨＫからでもリンクして、入れるようになっていますけど、その精中委のホームページの中にリストがあります。読影医の試験、それから技師の試験にうかってホームページに名前と所属を載せても良いという本人の意思表示があった方だけ、つまり人によっては、今勤めているところの関係でそういうようなことをしてはいけないと言われている方もあって、（意思表示が無いと載せないので）全員100％ではないのですけども、少なくてもそういうことがＯＫという人の名前と所属は載ってます。それから施設画像評価についても、100施設を越えたものですから載っております。それから講習の予定も載ってます。これからある講習の予定も載ってる。そういうホームページがありますのでご覧下さい。

質問者：

先生貴重なご講演ありがとうございました。三重県健康管理事業センターの吉田でございます。先ほどの保健師さんのご質問にも関連があるんですが、実際にうちのセンターで検診をしまして、精密検査で引っかかりまして、行かれる方が多数出てきているんですが、まずどこの施設に行くかということを聞かれます。それは私は自分が外科医ですので、何とかクリアをしまして、ご紹介をするんですが、運悪く乳腺を専門とされていない先生に当たりまして、マンモで引っかかって行ったのに、視触診とエコーだけで異常なしで帰されてしまい、その後どうしたら良いのですかという問い合わせがかかってくるんです。たまたま私の場合は自分がお願い出来るんですけれど、実際保健師さんは非常に困ると思うのですが、そういところは愛知県ではどういうふうにク

リアをしているのでしょうか。

遠藤先生：

とっても難しい問題で、そういうふうに問い合わせをして下さる受診者の方は、まだ自分で自分の活路を見いだせるんですけど、そうでなくって、あっ大丈夫なんだなっていうふうに検診よりも精密検査、医者の目の前にこう座ってやってくれた人の方が信頼できると思ってしまって、あー大丈夫だ良かったというふうにして済んでしまうことがあるんです。そうしますと検診のマンモグラムを読んだ先生方が、「どうなったかな、えー異常なし、カテゴリー５と付けたのに異常なしということで、もう１回どこか受診してもらいたい。」と、逆に愛知県の場合は、そういうようなストレスがあって、でもそこをどうやってクリアするかというと、今のところは、医師のネットワークで裏からと言うのか、どこへ行ったかということを、返ってきたときの施設名から拾って、そこの施設に存じ上げている方がいらっしゃれば、そこに電話を入れて、こういう所見なので、そのつもりでいないと危ない、見逃してしまうよという話を提供させて頂いているのですけれども、やはり精密検査をして良い施設という所をやはりきちんと色分けしてしてしまわないといけないのではないかというふうに思います、最終的にはです。そういう精密検査施設となりたかったら、やはり努力をしなければいけない。

今検診施設さんは一所懸命努力して、技師も受講させている。それから施設も管理もです。画像評価も受けている。読影もしっかりとＡ、Ｂを取っている先生にお願いしている。で、チェックされたのに癌を異常なしというふうに返してくる精密検査というのはもう否定されるものと思うのです。県の方から指定ということは精密検査を医療と捉えるとできないのです。でも検診と捉えれば、指定できるのです。だから精密検査イコール二次検診というふうに色づけして、つまりそこは治療までしなくてもしっかりと診断が出来ればよい施設というふうに考えれば、二次検診だから検診の基準をクリアしていなければいけないというふうな形で推進するしかないかなというふうに思います。

質問者：

平和町で保健師をしております岡本と申します。ありがとうございました。たまたまこの間、精密検査という結果が出た方にお電話をして追跡をしてたんですが、毎年検診を受けてるんだけれど、毎年引っかかって、毎年それで精密検査で病院の方に行くと、問題ないというふうに返ってくるのですと言われて、今年はだからもう受けないつもりなんですけどというふうにお話があったのですが、そのような場合、こちら側としてどのようにその方に対して指導していったら良いのかということを迷いましたので、今日お伺いしたいと思います。

遠藤先生：

それは、検診の記録の問題だと思います。つまり私が外来をしてましても、あなたは検診を受けてはいけません、もう医療機関に通っているのだから、検診を受けてはいけませんという人がいるのです。そういう人が現にいるわけ。それでそういう人に対しては、毎年医療機関で、あるいは半年に１回医療機関で診療を受けている。診療を受けてるという人たちは、検診は受けるべきではないのです。もう１つの場合として、検診を受ける、今までの検診の仕方だったら、チェックされてしまう。けれどもそれは検診の引っかける基準が間違っている、そういう場合があるのです。マンモグラフィの場合だったら、その基準がきちっと決まってます。決まってるんです。なので、前の年にやって今年もやったんだったらフイルムを前回と比べて変わっていなければ、カテゴリーのランクは下がるんです。最初の年、前の年の情報が無ければⅢでチェックしたものが、次の年はⅡになる、同じだったら。という形で落とせるのです。だからそういうところでも、検診の質の問題が問われるのです。今の受診者の反応というのは、もう検診は当てにしないけど、検診は受けてしまうという変な矛盾したことになってるんですけれども、そこでもしそれがマンモグラムで起きてることであれば、マンモグラムの読み方に関して、それでよろしいのでしょうかという質問を読影者の方に投げかけて下さい。そういうようなフィードバックがあれば読影も質が良くなっていくはずです。よろしくお願いします。ありがとうございました。

どうもありがとうございました。まだいろいろご質問がありましょうが、時間がまいりましたので、ではどうも先生ありがとうございました。

Int. J. Cancer: **106,** 45–51 (2003)

Publication of the International Union Against Cancer

RASSF1A GENE INACTIVATION IN NON-SMALL CELL LUNG CANCER AND ITS CLINICAL IMPLICATION

Hideki ENDOH[1,3], Yasushi YATABE[2], Shigeki SHIMIZU[1], Kohei TAJIMA[1,3], Hiroyuki KUWANO[3], Takashi TAKAHASHI[4] and Tetsuya MITSUDOMI[1]*

[1]*Department of Thoracic Surgery, Aichi Cancer Center Hospital, Nagoya, Japan*
[2]*Department of Pathology and Molecular Diagnosis, Aichi Cancer Center Hospital, Nagoya, Japan*
[3]*Department of Surgery I, Gunma University School of Medicine, Maebashi, Japan*
[4]*Division of Molecular Oncology, Aichi Cancer Center Research Institute, Nagoya, Japan*

Methylation-associated inactivation of *RASSF1*, a putative tumor suppressor identified at 3p21.3, is reported in several cancers. We examined *RASSF1* in non-small lung cancer (NSCLC) to search for clinical implications. RT-PCR analysis showed no expression of *RASSF1A* in 12 of 20 lung cancer cell lines. Loss of expression correlated well with promoter methylation status of these lines. Sequence analysis revealed 2 polymorphisms (codons 21 and 133) in *RASSF1A* transcripts, but not in *RASSF1C* transcripts. No somatic mutations were found. Of 7 cell lines with *K-ras* mutations at codon 12 or 61, 2 lost expression of *RASSF1A*, whereas in 13 cell lines with wild-type *K-ras* gene, 10 lost *RASSF1A* gene expression ($p = 0.0521$). We investigated methylation status of this putative tumor suppressor gene in 100 primary NSCLCs to determine whether there is a clinical significance. Forty-two of primary NSCLCs demonstrated methylated allele. There is no correlation between promoter methylation of *RASSF1A* and clinicopathological findings, including histological type or grade, tumor staging, *p53* and *K-ras* mutational status, or patients' survival. In the cases of Stage I and II disease, however, *RASSF1A* methylation was associated with earlier recurrence ($p = 0.0247$). Epigenetic silencing of *RASSF1A* is a frequent event in non-small lung cancer and will provide novel opportunities to develop diagnosis and therapy of NSCLC.

Key words: *RASSF1*; NSCLC; 3p21; methylation; TSG

Lung cancer is the leading cause of cancer-related deaths in Japan, claiming 50,000 lives annually and the situation is still worsening.[1] Less than 50% of NSCLC patients are candidates for potentially curative resection and <50% of these patients survived for more than 5 years.[2]

Allelic loss at chromosome 3p21 is the one of the most frequent genetic changes found in various types of human cancers, including those in the lung.[3] Despite an intensive search for putative tumor suppressor genes at this locus, however, no such gene has been identified that fits Knudson's prototype of tumor suppressor genes (*i.e.*, 1 allele deleted with the other mutated). The *RASSF1* gene at 3p21.3 that encodes more than 7 isoforms derived from alternative mRNA splicing and promoter usage has been cloned recently.[4] Among them, *RASSF1A* and *RASSF1C* are major transcripts. Loss of expression and methylation of the promoter region of *RASSF1A* but not of *RASSF1C* is found frequently in various types of human cancers including examples in the lung,[4–8] breast,[5,6,9] ovary,[5,10] kidney,[10–12] bladder,[13] nasopharynx,[14] stomach[15] and colon.[16]

RASSF1A is predicted to encode a 39-kd peptide that contains an N-terminal diacylglycerol (DAG)-binding domain and a Ras-association domain.[4] *RASSF1C* is predicted to encode a 32-kd peptide that lacks the DAG-binding but contains the Ras-association domain.[17,18] This latter is >50% identical and >70% similar to the carboxyl terminal 225 residues of mouse *Nore1*,[19] identified as a potential ras effector. In addition, ectopic expression of *RASSF1A* can suppress growth of tumor cell lines in both *in vivo* and *in vitro*.[4,6]

Promoter methylation is an important means for transcriptional repression of a number of cancer-associated genes including $p16^{INK4a}$, $p14^{ARF}$, *MGMT*, *APC*, *DAPK*, *GSTP1*, *RARβ2*, *CDH1* and *RASSF1A*.[20,21] Although there is evidence for epigenetic silencing of TSGs and its contribution to carcinogenesis, the underlying mechanisms for *de novo* methylation in cancer cells remains unclear. Two studies have addressed recently the issue on the association between expression level or polymorphism of methyl-group metabolism genes and methylation status of these genes in human tumors including lung cancer. Some polymorphism might explain methylation status[22] but expression level of these genes does not seem to have direct association with methylation status.[23]

We examined lung cancer cell lines for expression, methylation status and for somatic mutations of the *RASSF1A* gene. We investigated 100 consecutive, resected NSCLC specimens for promoter methylation with reference to clinical parameters in an attempt to determine the significance of alterations of the *RASSF1* gene.

MATERIAL AND METHODS

Cell lines

Five SCLC and 15 NSCLC cell lines were used in our study. Details of their derivation and culture conditions have been described previously.[24,25] Cells were cultured in RPMI1640 medium supplemented with 5% FCS, containing 100 U/ml penicillin and 0.1 mg/ml streptomycin.

Patients

Primary tumor samples were obtained from 100 consecutive patients with NSCLCs who underwent potentially curative resection at the Aichi Cancer Center Hospital from April 1996–March 1998. We obtained appropriate institutional review board approval and patients' written informed consent. Patients included 68 males and 32 females with an age at diagnosis ranging from 32–77 years (median = 62). Fifty-seven patients had Stage I disease, 11 had Stage II and 32 had Stage III. There were 67 adenocarcinomas, 26 squamous cell carcinomas, 4 large cell carcinomas and 3 adenosquamous carcinomas. Sixty-two patients were smokers, with a median smoking index (*i.e.* number of cigarettes/day × years) of 480, the remaining 38 being never-smokers. Tumor samples were

Abbreviations: COBRA, combined bisulfite-restriction analysis; *GAPDH*, glyceraldehyde-3-phosphate dehydrogenase; MSP, methylation specific PCR; NSCLC, non-small cell lung carcinoma; RT-PCR, reverse transcription-polymerase chain reaction; SCLC, small cell lung carcinoma; TSG, tumor suppressor gene

*Correspondence to: Department of Thoracic Surgery, Aichi Cancer Center Hospital, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan. Fax: +81-52-764-2963. E-mail: mitsudom@aichi-cc.jp

Received 6 September 2002; Revised 2 January 2003; Accepted 23 January 2003

DOI 10.1002/ijc.11184
Published online 29 April 2003 in Wiley InterScience (www.interscience.wiley.com).

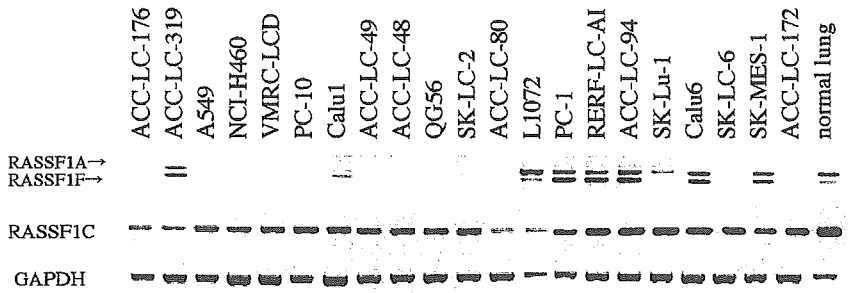


FIGURE 1 – *RASSF1A, RASSF1F* and *RASSF1C* mRNAs detected by isoform-specific RT-PCR in lung cancer cell lines. *Top*: Primers were S2 and AS3 and bands of 372 base-pairs (bp) (*RASSF1A*) and 262 bp (*RASSF1F*) were obtained. *Middle*: 182 bp bands of *RASSF1C* were found in all cell lines using primers 2γ and AS4. *Bottom*: RT-PCR for *GAPDH* as a control. L1072 was a resected adenocarcinoma sample and the corresponding normal lung was used as a positive control. SCLC cell lines were ACC-LC-49, ACC-LC-48, SK-LC-2, ACC-LC-80, ACC-LC-172, SK-LC-6. The others were NSCLC cell lines (derived from 5 adenocarcinomas, 7 squamous cell carcinomas, and 3 large cell carcinomas).

obtained at the time of surgery, rapidly frozen in liquid nitrogen and stored at −80°C until use. Total RNA was isolated using the acid guanidium thiocyanate/cesium chloride procedure and genomic DNA was extracted by the sodium dodecylsulfate-proteinase K and phenol-chloroform extraction method.[26]

Expression of RASSF1A *and* RASSF1C *assessment by RT-PCR*

Expression of *RASSF1A* and *RASSF1C* was evaluated by RT-PCR. Primers for *RASSF1A* were S2 (5′-ACCCCACACGGCAGCTGGTCC-3′) and AS2 (5′-CTCGTCCACGTTCGTGTCCCGC-3′), or S2 and AS3 (5′-ATGAAGCCTGTGTAAGAACCGTCCTTG-3′) , and for RASSF1C were 2γ (5′-CGGAGGCGCCTTCTTTCGAAATGACCT-3′) and AS4 (5′-GTGACAGCACATGTAGGTGCTTGACAG-3′). PCR conditions were 95°C for 1 min, 64°C for 1 min and 72°C for 1 min for 40 cycles. RT-PCR for *GAPDH* transcripts was carried out with the forward primer GAPDH-S (5′-CCCCTTCATTGACCTCAACTAC-3′) and the reverse primer GAPDH-AS (5′-TACTTCTCATGGTTCACACCCA-3′). PCR products were separated by 2% agarose gel electrophoresis and stained with ethidium bromide.

Mutation analysis of the RASSF1 *and* K-ras *genes*

Isoform-specific regions of *RASSF1A* and *RASSF1C* (1α, 2αβ and 2γ[4]) and exons 3-6 of *RASSF1* were amplified with 4 sets of primers : 1α; S1 (5′-AAAGCCAGCGAAGCACGGGC-3′)-AS1 (5′-ATGAAGTCGCCACAGAGGTCGCAC-3′), 2αβ; S2-AS2, 2γ and exon3; 2γ-AS4, exon3-6; S4 (5′-GAGTACAATGCCCAGATCAACAGCAAC-3′)- AS6 (5′-CCTTCCACCTGGGGGTACAAG-3′). PCR conditions were as described above (annealing temperatures 60–65°C) and products were directly sequenced using an automatic sequencer (ABI PRISM 3700, Applied Biosystems, Foster City, CA) without further purification.

The mutational status of the *K-ras* gene exons I and II was also examined by direct sequencing of the RT-PCR products. The primers used were K1F (5′-GGCCTGCTGAAAATGACTGA-3′) and K2R (5′-CACAAAGAAAGCCCTCCCCA-3′).

Methylation analysis of the promoter regions of the RASSF1A *gene*

We investigated the methylation status of the RASSF1A promoter regions by 3 methods after bisulfite modification of genomic DNA; *i.e.*, sequencing, detection of restriction length polymorphisms (COBRA) and allele specific PCR (MSP).

Bisulfite modification of genomic DNA was carried out as described elsewhere.[27,28] Briefly, 200 ng of genomic DNA was denatured by NaOH (0.3 M) in the presence of 4% low melting agarose (NuSieve GTG Agarose, BioWhittaker Molecular Applications, Rockland, MA) in a 20 μl volume to prevent renaturation (agarose bead method).[29] Five hundred microliters of a bisulfite solution (125 mM hydroquinone and 2.5M sodium bisulfate, pH 5.0) were added followed by incubation at 50°C for 15 hr. DNA samples were then washed with 0.3 M NaOH and Tris-EDTA. For experiments with cell lines, the methylation status of the *RASSF1A* gene was evaluated by sequencing. We analyzed 16 CpGs in a 204-bp fragment containing 3 Sp1 consensus binding sites and the putative transcription and translation initiation sites.[4] Primers were MU379 (5′-GTTTTGGTAGTTTAATGAGTTTAGGTTTTTT-3′) and ML730 (5′-ACCCTCTTCCTCTAACACAATAAAAC-TAACC-3′), and semi-nested PCR primers were MU379 and ML561 (5′-CCCCACAATCCCTACACCCAAAT-3′). PCR conditions were 95°C for 1 min, 60°C for 1 min, 72°C for 1 min, for 40 cycles, and semi-nested PCR conditions were 95°C for 1 min, 65°C for 1 min, 72°C for 1 min, for 40 cycles. The PCR products were then cloned into pCR2.1 vectors using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA) and 5 clones of each specimen were sequenced using an ABI 3700. For COBRA,[30] 15 μl of the 204 base-pair (bp) semi-nested PCR product of bisulfite modified genomic DNA was incubated with 20 U of *Taq*I (New England BioLabs, Beverly, MA) for 2 hr. The expected sizes of the *Taq*I restriction enzyme digestion products are 173, 112, 92, 81 and 31 bp.[5] The restriction enzyme digest was visualized by electrophoresis in a 3.5% agarose gel stained with ethidium bromide. Furthermore, we evaluated *RASSF1A* gene methylation by MSP.[31] The primers were MS (5′-GGGTTTTGCGAGAGCGCG-3′) and MA (5′-GCTAACAAACGCGAACCG), and the unmethylation specific PCR primers were US (5′-GGTTTTGTGAGAGTGTGTTTAG-3′) and UA (5′-CACTAACAAACACAAACCAAAC-3′), using 50 ng of bisulfite-modified genomic DNA as template for 45 cycles at 95°C for 50 sec, 64 or 59°C for 50 sec and 72°C for 50 sec.[6] The PCR products were resolved on a 2.5% agarose gel.

Statistical analysis

For comparisons of proportions, the χ^2 test or Fisher's exact test were used. The Kaplan-Meier method was employed to estimate the probability of survival as a function of time and survival differences were analyzed by the log-rank test. The 2-sided significance level was set at $p < 0.05$. To identify which independent factors jointly had a significant influence on the overall survival, the Cox proportional hazards modeling technique was applied . All analyses were carried out using StatView (version 5, SAS institute Inc., Cary, NC) software on a Macintosh computer.

RESULTS

Expression, mutation and methylation status of RASSF1 *A and C gene in lung cancer cell lines and relationship with* K-ras *and* p53 *mutations*

We investigated whether *RASSF1A* and *RASSF1C* genes were expressed in lung cancer cell lines by RT-PCR. Although *RASSF1C* was expressed in all cell lines, *RASSF1A* was not expressed in 12 of 20 (60%) (Fig. 1). *RASSF1F*, which yielded a

PCR product 110 bp smaller than *RASSF1A*, was expressed in cases positive for *RASSF1A*, in accordance with a previous report.[6] We then evaluated sequence alterations for the *RASSF1A* gene by direct sequencing of PCR products and detected the 2 base changes shown in Table I, AAG to CAG (codon 21) resulting in Lys to Gln and GCT to TCT (codon 133) resulting in Ala to Ser. These changes have been described previously,[5,6] and we confirmed them to be polymorphisms by sequencing the gene in several tumors and corresponding normal tissues.

We next determined the CpG methylation status in the 5′ region of the *RASSF1A* gene by sequencing bisulfite-modified DNA (Fig. 2, Table I). At least 5 independent clones were assessed for each cell line examined. Of 20 cell lines, 5 cell lines did not have any methylation in any CpG sites in the promoter region of the *RASSF1A* gene. It was noted that these all expressed *RASSF1A* (Fig. 2). In contrast, in 12 cell lines that did not express *RASSF1A*, most of the CpG sites were methylated. In the other 3 cell lines, in which methylation was partial (Fig. 2), one of these expressing *RASSF1A* normally, and the others showing reduction. We could not find any CpG sites that specifically correlated with expression status in these experiments.

K-ras mutations occurring at codons 12, 13 or 61 were detected in 7 of 20 lung cancer cell lines (Table I). Of these mutations, 2 had lost expression of the *RASSF1A* gene (29%), whereas in the 13 cell lines that had wild-type *K-ras* gene, this was the case for 10 (77%). Although this difference was intriguing, the level of significance was border-line ($p = 0.0521$, Fisher's exact test). With regard to *p53* mutations, 16 of 20 cell lines had proved positive, loss of expression of *RASSF1A* being apparent in 9 of the 16 with mutant *p53* (56%), and in 3 of 4 cell lines with wild-type *p53* (75%). This difference was not significant ($p = 0.6186$)(Table I).

Methylation status of the RASSF1A *gene in resected, primary NSCLCs and its clinical implications*

In the experiments with lung cancer cell lines, we found that *RASSF1A* methylation status corresponded well with expression status and that there was no bias in terms of CpG sites with regard to expression status. Therefore, we evaluated *RASSF1A* changes in resected NSCLCs by examining several CpG sites of methylation as a surrogate for gene silencing.

To determine the overall frequency of *RASSF1A* methylation, we carried out both MSP analysis (examining methylation of CpG positions 9–12)[4] and COBRA (*Taq*I digestion assay, examining methylation positions 6 and 16)[4] (Fig. 3). Of 100 primary NSCLCs, 42 were found to be methylated by both assays, 1 only by COBRA and 1 only by MSP. These latter 2 cases were considered to retain expression of the *RASSF1A* from the experiment using cell lines and treated as cases with active *RASSF1A* (see above).

Table II summarizes data for associations between *RASSF1A* promoter methylation and clinical-pathological and demographic characteristics in NSCLC patients. There were no significant associations with such factors as age, gender, smoking index, histologic types, stage, or *p53* mutational status ($p = 0.3165$, Fisher's exact test) (Shimizu and Mitsudomi, manuscript in preparation). We also examined the specimens for *K-ras* mutations and found 3 mutations at codon 12 and 1 at codon 61 in 4 adenocarcinomas. Interestingly, of 4 tumors with *K-ras* mutation, 3 had their *RASSF1A* gene unmethylated showing the similar tendency as data obtained in cell lines. This difference was not significant statistically, however, probably due to low incidence of *K-ras* mutation in our cohort ($p = 0.6370$).

The allele frequency for the Lys→Gln polymorphism at codon 21 was 46% in the patients, but there was no significant correlation between this polymorphism and the methylation status (Lys/Lys: 13/27 (48%), Lys/Gln + Gln/Gln: 29/73 (40%), $p = 0.4901$).

We analyzed our data for any prognostic significance of *RASSF1A* gene methylation. With all patients or patients with

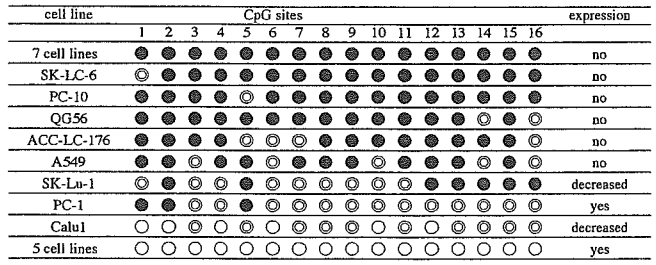


FIGURE 2 – Methylation status of each CpG in the *RASSF1A* promoter and *RASSF1A* expression in the lung cancer cell lines. The numbers for the CpG sites correspond to those listed in Dammann *et al.*[4] ●: 80–100% methylated; ◎ 20–60% methylated; ○: 0% (unmethylated). Group of 7 cell lines comprised of: ACC-LC-49, ACC-LC-48, NCI-H460, SK-LC-2, ACC-LC-80, ACC-LC-172, and VMRC-LCD; 5 cell lines group comprised of: ACC-LC-94, ACC-LC-319, RERF-LC-AI, Calu6 and SK-MES-1.

TABLE I – METHYLATON STATUS OF THE *RASSF1A* GENE AND MUTATIONS OF *K-RAS* AND *P53* GENES IN THE CELL LINES TESTED

Cell line	Histologic type	*RASSF1A*			*K-ras* mutation	*p53* mutation[1]
		Expression	Methylation	Polymorphism		
ACC-LC-94	Ad	+	No	Lys21Gln	Gly12Asp	Asp281Gly
ACC-LC-319	Ad	+	No	Lys21Gln	Gln61Leu	1-bp deletion (codon 71)
A549	Ad	−	Yes	Ala133Ser	Gly12Ser	—
VMRC-LCD	Ad	−	Yes	—	—	Arg175His
SK-Lu-1	Ad	+/−	Partial	—	Gly12Asp	His193Arg
ACC-LC-176	Sq	−	Yes	Ala133Ser	—	—
PERF-LC-AI	Sq	+	No	Lys21Gln	—	Gln104stop
PC-1	Sq	+	Partial	Lys21Gln	—	—
PC-10	Sq	−	Yes	—	—	Gly245Cys
Calu 1	Sq	+/−	Partial	—	Gly12Cys	Homozygous del
QG56	Sq	−	Yes	—	—	Pro249Ser
SK-MES-1	Sq	+	No	Lys21Gln	—	Glu298stop
NCI-H460	La	−	Yes	—	Gln61His	—
Calu 6	La	+	No	Ala133Ser	Gln61Lys	Arg196stop
SK-LC-6	La	−	Yes	Ala133Ser	—	Arg158Leu
ACC-LC-48	SCLC	−	Yes	—	—	Arg249Ser
ACC-LC-49	SCLC	−	Yes	Lys21Gln	—	1-bp deletion (codon 110)
ACC-LC-80	SCLC	−	Yes	—	—	1-bp insertion (codon 152)
ACC-LC-172	SCLC	−	Yes	—	—	Tyr126Cys
SK-LC-2	SCLC	−	Yes	—	—	1-bp deletion (codon 85)

[1]Data from Shimizu and Mitsudomi, manuscript in preparation.

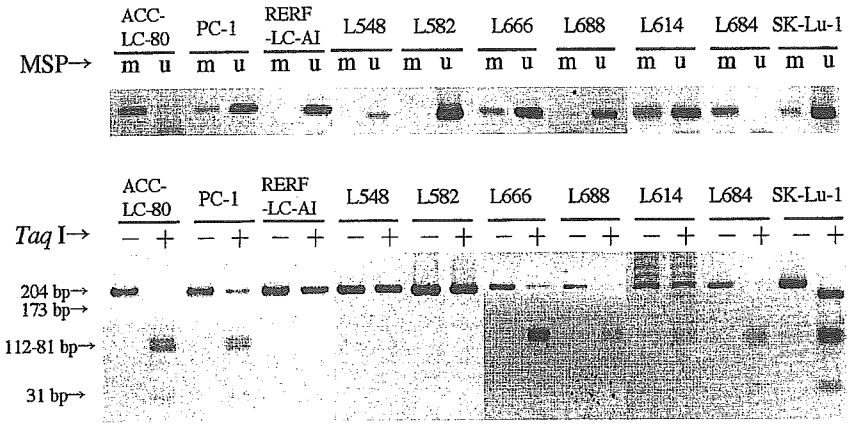


FIGURE 3 – Examples of the results of methylation analysis of *RASSF1A* in NSCLC. *Top*: MSP m: use of primers specific for methylated, MS and MA primer pair u: use of primers specific for unmethylated, US and UA primer pair Each primer set generated a 169 bp product. *Bottom*: COBRA On the left (*Taq*I−), the PCR product after semi-nested PCR using the MU379 and ML561 primer pair was 204 bp in size. Products of digestion by *Taq*I and the 204 bp undigested product are indicated (*Taq*I+). RERF-LC-AI, L548 and L582 samples are unmethylated.

TABLE II – RELATIONSHIP BETWEEN THE INCIDENCE OF *RASSF1A* METHYLATION AND VARIOUS CLINICAL-PATHOLOGIC FEATURES OF THE NSCLCS

Factor	Category	Total cases (n = 100)	Methylated cases (42%)	*p*
Gender	Male	68	29 (43)	0.9999
	Female	32	13 (41)	
Age	63≦	48	17 (35)	0.2280
	<63	52	25 (48)	
Smoking index	400<	53	22 (42)	0.9999
	<400	47	20 (43)	
Histologic type	Squamous	26	8 (31)	0.2484
	Non-squamous	74	34 (46)	
Stage	I	57	25 (44)	0.6876
	II–III	43	17 (40)	
K-ras mutation	Present	4	1 (25)	0.6370
	Absent	96	41 (43)	
p53 mutation[1]	Present	44	21 (48)	0.3165
	Absent	56	21 (38)	

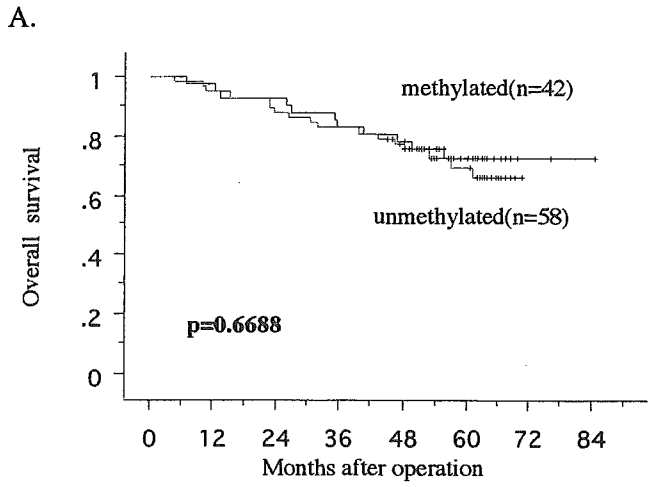
[1]Data from Shimizu and Mitsudomi, manuscript in preparation.

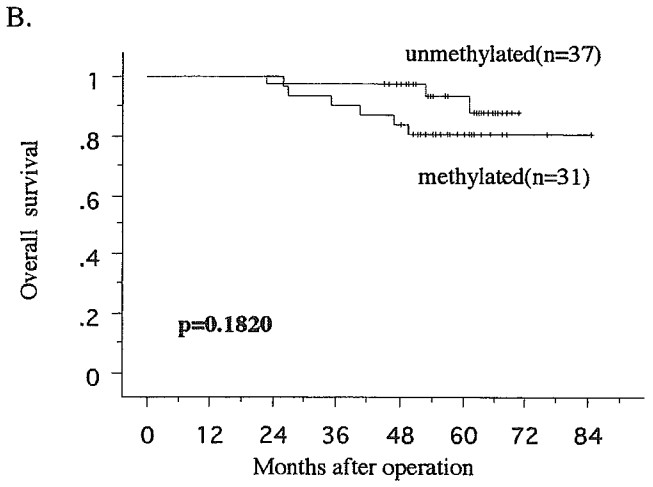
Stage III disease *RASSF1A* methylation was not a significant prognostic factor (Figs. 4*a,c*, 5*a,c*). For patients with Stages I and II disease, *RASSF1A* methylation was predictive for earlier recurrence (Fig. 5*b*), but not for overall survival (Fig. 4*b*). The interrelationships of possible prognostic factors and survival were analyzed by Cox proportional hazard model in the entire cohort, using gender, age, histological type, disease stage (including only in the entire cohort), *p53*, *RASSF1A* methylation status and *RASSF1A* polymorphism. Significant independent prognostic factors for the entire cohort were female gender (hazard ratio [HR] = 0.17, p = 0.006), a smoking index <400 (HR = 0.26, p = 0.01), and disease Stage I (HR = 0.21, p = 0.0003), whereas *RASSF1* gene methylation was not significant (HR = 0.98, p = 0.96) (Table III).

DISCUSSION

In our present study, 5/5 (100%) SCLC and 7/15 (47%) NSCLC lines were found to have lost expression of *RASSF1A* but not *RASSF1C*, in good agreement with previous reports.[4–6] For example, Burbee *et al.*, reported that 100% of SCLC and 65% of NSCLC lines lost *RASSF1A* expression. Because the *RASSF1A* expression status corresponded well with the promoter methylation status, we examined primary, resected tumor samples for methylation of *RASSF1A* as a surrogate marker for gene silencing. We found 42 of 100 cases (42%) to exhibit *RASSF1A* promoter methylation, again, in line with previous reports.[4–6] Dammann *et al.*,[4] and Burbee *et al.*,[6] found 22/58 (38%) and 32/107 (30%) NSCLCs respectively to have their *RASSF1A* promoter methylated. Tomizawa *et al.*[8] found 35/110 (32%) Stage I adenocarcinomas to have their *RASSF1A* promoter methylated. In contrast, *RASSF1C* expression was ubiquitous in line with the finding of a unmethylated status for CpG sites in the presumed *RASSF1C* promoter region in cancer cell lines.[4–6] All our cell lines also showed *RASSF1C* expression well.

It is known that activated ras is associated with senescence and terminal differentiation or apoptosis as well as growth promotion and transformation.[32,33] Vos[34] reported that *RASSF1C* binds ras in a GTP-dependent manner, both *in vitro* and *in vivo*, and that RASSF1 can function as a novel ras effector which may be used by ras to mediate apoptosis. Because isoforms A and C encode the identical ras association domain, it is likely that RASSF1A binds to ras in a fashion similar to *RASSF1C*. It has been shown alternatively that RASSF1A is able to heterodimerize with Nore, a ras effector identified recently, and thereby associate with ras-like GTPases.[35] Loss of expression of *RASSF1A* may shift the balance of ras activities toward a growth promoting effect without the necessity of ras mutations. In our present study, cell lines with ras mutations tend to retain *RASSF1A* expression than those without ras mutations, with border-line significance, although some cell lines such as A549 and NCI-H460 (present study) and T24, TSU-Pr1 and HT1197[13] exhibit both ras mutations and epigenetic silencing of *RASSF1*. It has been reported recently that colorectal cancers with wild-type *K-ras* have significantly higher incidence of *RASSF1* methylation than those with *K-ras* mutations (34/135 *vs.* 11/87, p = 0.023).[16] We found that resected NSCLC with wild-type *K-ras* tends to have a higher incidence of *RASSF1A* methylation than those with mutant *K-ras* (41/96 wild-type *K-ras vs.* 1/4 mutant *K-ras*). The data did not reach statistical significance (p = 0.6370). Tomizawa *et al.*[8] found a similar tendency in resected Stage I lung adenocarcinomas (33/97 wild-type *K-ras vs.* 2/13 mutant *K-ras*, p = 0.2193). Furthermore, it might be of





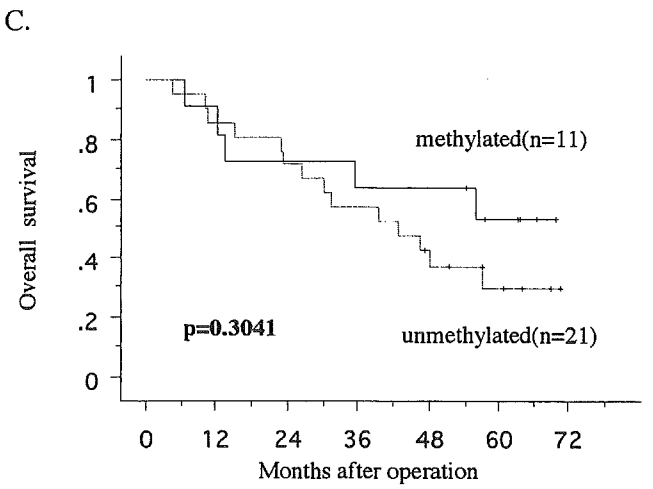
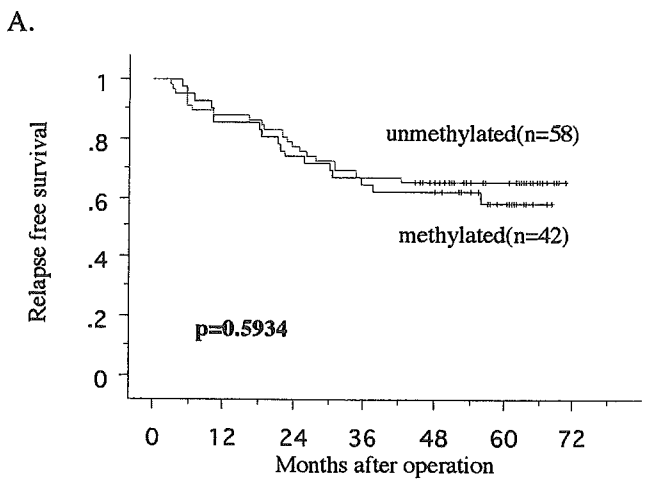
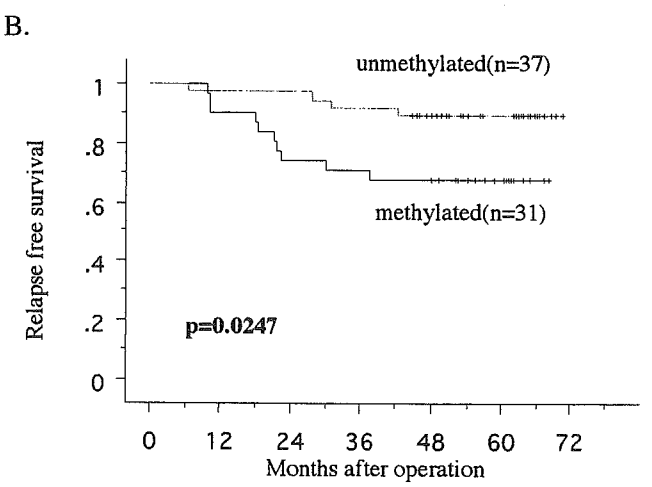


FIGURE 4 – Survival curves with reference to *RASSF1A* promoter methylation. (*a*) Overall survival curves for all 100 NSCLC cases. (*b*) Survival curves for patients with early disease (Stages I and II). (*c*) Survival curves for patients with advanced disease (Stage III).





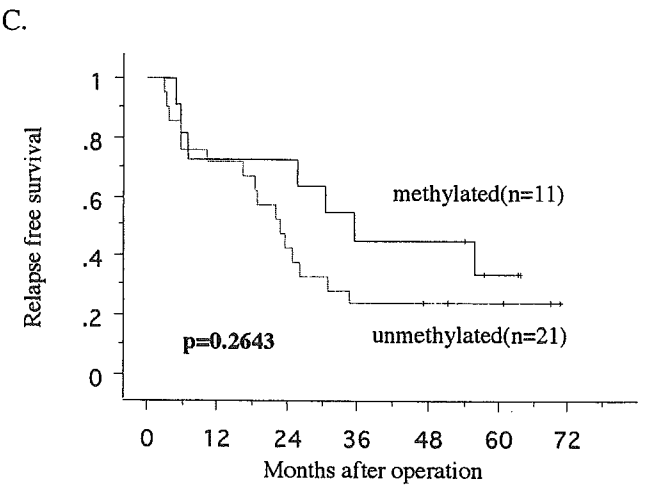


FIGURE 5 – Relapse-free survival curves with reference to *RASSF1A* promoter methylation. (*a*) Relapse free curves for all 100 NSCLC cases. (*b*) Relapse free curves for patients with early disease (Stages I and II). (*c*) Relapse free curves for patients with advanced disease (Stage III).

TABLE III – MULTIVARIATE ANALYSIS OF OVERALL SURVIVAL USING THE COX PROPORTIONAL HAZARDS MODEL

Variable	Category	HR	95% CI	*p*
Gender	Female/male	0.173	0.049–0.611	0.0064
Age	63≦/<63	1.947	0.852–4.450	0.1143
Smoking index	400</<400	0.261	0.086–0.789	0.0174
Histologic type	Non-squamous/Squamous	0.540	0.198–1.474	0.2292
Pathological stage	I/II–III	0.208	0.088–0.492	0.0003
p53 mutation	Mutant/wild-type	0.841	0.365–1.937	0.6843
RASSF1A	Methylated/unmethylated	0.979	0.432–2.221	0.9599
Codon 21 polymorphism	(Lys/Lys)/(Lys/Gln+ Gln/Gln)	0.405	0.147–1.121	0.0819

interest that SCLCs almost never to have *K-ras* mutation[36] had a high frequency of *RASSF1A* methylation in our study and several reports.[4–7] Taken together, these data suggest potential interaction between *K-ras* activation (mutation) and *RASSF1A* inactivation (methylation) in lung cancer *in vivo*.

The present examination of *RASSF1A* methylation status in 100 resected NSCLCs demonstrated 42% to be positive, but without any association significant statistically to several clinical factors. This confirms the previous findings.[6] It was reported a strong prognostic influence of *RASSF1A* methylation in NSCLCs,[6,8] but we detected significance only in the early stage subset. The reason for this discrepancy is not clear. Differences in the ethnicity of the patients and the relatively short observation period of our cohort may be relevant.

In conclusion, we found that a significant proportion of NSCLCs to have their *RASSF1A* gene inactivated through promoter methylation. This (*RASSF1A* methylation) may have prognostic importance for the patients with early stage disease, but further studies are needed to confirm this. The finding of frequent *RASSF1A* inactivation may provide novel opportunities to develop diagnostic approaches or therapeutic interventions targeted at reversing *RASSF1A* silencing[6] or the downstream consequences of *RASSF1A* inactivation.

ACKNOWLEDGEMENT

H. Endoh, S. Shimizu and T. Mitsudomi were supported in part by the Aichi Cancer Research Foundation.

REFERENCES

1. Tominaga S, Oshima A, Cancer mortality and morbidity statistics. Gann Monogram Cancer Research, 1999. No. 47.
2. Vaporciyan AA, Nesbitt JC, Lee JS, Stevens C, Komaki R, Roth JA. Cancer of the lung. In: Bast RC, Kufe DW, Pollock RE, Weichselbaum RR, Holland JF, Frei E, eds. Cancer medicine, 5th ed. Hamilton: B.C. Decker Inc., 2000. 1227–92.
3. Kok K, Naylor SL, Buys CH. Deletions of the short arm of chromosome 3 in solid tumors and the search for suppressor genes. Adv Cancer Res 1997;71:27–92.
4. Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat Genet 2000;25:315–9.
5. Agathanggelou A, Honorio S, Macartney DP, Martinez A, Dallol A, Rader J, Fullwood P, Chauhan A, Walker R, Shaw JA, Hosoe S, Lerman MI, et al. Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. Oncogene 2001;20:1509–18.
6. Burbee DG, Forgacs E, Zochbauer-Muller S, Shivakumar L, Fong K, Gao B, Randle D, Kondo M, Virmani A, Bader S, Sekido Y, Latif F, et al. Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. J Natl Cancer Inst 2001;93:691–9.
7. Dammann R, Takahashi T, Pfeifer GP. The CpG island of the novel tumor suppressor gene RASSF1A is intensely methylated in primary small cell lung carcinomas. Oncogene 2001;20:3563–7.
8. Tomizawa Y, Kohno T, Kondo H, Otsuka A, Nishioka M, Niki T, Yamada T, Maeshima A, Yoshimura K, Saito R, Minna JD, Yokota J. Clinicopathological significance of epigenetic inactivation of RASSF1A at 3p21.3 in stage I lung adenocarcinoma. Clin Cancer Res 2002;8:2362–8.
9. Dammann R, Yang G, Pfeifer GP. Hypermethylation of the CpG island of Ras association domain family 1A (RASSF1A), a putative tumor suppressor gene from the 3p21.3 locus, occurs in a large percentage of human breast cancers. Cancer Res 2001;61:3105–9.
10. Yoon JH, Dammann R, Pfeifer GP. Hypermethylation of the CpG island of the RASSF1A gene in ovarian and renal cell carcinomas. Int J Cancer 2001;94:212–7.
11. Morrissey C, Martinez A, Zatyka M, Agathanggelou A, Honorio S, Astuti D, Morgan NV, Moch H, Richards FM, Kishida T, Yao M, Schraml P, et al. Epigenetic inactivation of the RASSF1A 3p21.3 tumor suppressor gene in both clear cell and papillary renal cell carcinoma. Cancer Res 2001;61:7277–81.
12. Dreijerink K, Braga E, Kuzmin I, Geil L, Duh FM, Angeloni D, Zbar B, Lerman MI, Stanbridge EJ, Minna JD, Protopopov A, Li J, et al. The candidate tumor suppressor gene, RASSF1A, from human chromosome 3p21.3 is involved in kidney tumorigenesis. Proc Natl Acad Sci USA 2001;98:7504–9.
13. Lee MG, Kim HY, Byun DS, Lee SJ, Lee CH, Kim JI, Chang SG, Chi SG. Frequent epigenetic inactivation of RASSF1A in human bladder carcinoma. Cancer Res 2001;61:6688–92.
14. Lo KW, Kwong J, Hui AB, Chan SY, To KF, Chan AS, Chow LS, Teo PM, Johnson PJ, Huang DP. High frequency of promoter hypermethylation of RASSF1A in nasopharyngeal carcinoma. Cancer Res 2001;61:3877–81.
15. Byun DS, Lee MG, Chae KS, Ryu BG, Chi SG. Frequent epigenetic inactivation of RASSF1A by aberrant promoter hypermethylation in human gastric adenocarcinoma. Cancer Res 2001;61:7034–8.
16. van Engeland M, Roemen GMJM, Brink M, Pachen MMM, Weijenberg MP, de Bruine AP, Arends JW, van den Brandt PA, de Goeij AFPM, Herman JG. K-ras mutations and RASSF1A promoter methylation in colorectal cancer. Oncogene 2002;21:3792–5.
17. Lerman MI, Minna JD. The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumor suppressor genes. The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium. Cancer Res 2000;60:6116–33.
18. Hurley JH, Newton AC, Parker PJ, Blumberg PM, Nishizuka Y. Taxonomy and function of C1 protein kinase C homology domains. Protein Sci 1997;6:477–80.
19. Vavvas D, Li X, Avruch J, Zhang XF. Identification of Nore1 as a potential Ras effector. J Biol Chem 1998;273:5439–42.
20. Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. Cancer Res 2001;61:3225–9.
21. Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. Oncogene 2002;21:5450–61.
22. Paz MF, Avila S, Fraga MF, Pollan M, Capella G, Peinado MA, Sanchez-Cespedes M, Hermann JG, Esteller M. Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors. Cancer Res 2002;62:4519–24.
23. Sato M, Horio Y, Sekido Y, Minna JD, Shimokata K, Hasegawa Y. The expression of DNA methyltransferases and methyl-CpG-binding proteins is no associated with the methylation status of p14ARF, p16 INK4a and RASSF1A in human lung cancer cell lines. Oncogene 2002;21:4822–9.
24. Takahashi T, Ueda R, Song X, Nishida K, Shinzato M, Namikawa R, Ariyoshi Y, Ota K, Kato K, Nagatsu T. Two novel cell surface antigens on small cell lung carcinoma defined by mouse monoclonal antibodies NE-25 and PE-35. Cancer Res 1986;46:4770–5.
25. Hida T, Ariyoshi Y, Kuwabara M, Sugiura T, Takahashi T, Hosoda K, Niitsu Y, Ueda R. Glutathione s-transferase π levels in a panel of lung cancer cell lines and its relation to chemo-radiosensitivity. Jpn J Clin Oncol 1993;23:14–9.
26. Sambrook J, Fritsch EF, Maniatis T. Isolation of total RNA by extraction with strong denaturants. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press, 1989. 7.18–25.

27. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994;22:2990–7.
28. Yatabe Y, Tavare S, Shibata D. Investigating stem cells in human colon by using methylation patterns. Proc Natl Acad Sci USA 2001; 98:10839–44.
29. Olek A, Oswald J, Walter J. A modified and improved method for bisulphite based cytosine methylation analysis. Nucleic Acids Res 1996;24:5064–6.
30. Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res 1997;25:2532–4.
31. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci USA 1996;93:9821–6.
32. Bar-Sagi D, Feramisco JR. Microinjection of the ras oncogene protein into PC12 cells induces morphological differentiation. Cell 1985;42: 841–8.
33. Mayo MW, Wang CY, Cogswell PC, Rogers-Graham KS, Lowe SW, Der CJ, Baldwin AS Jr. Requirement of NF-κB activation to suppress p53-independent apoptosis induced by oncogenic Ras. Science 1997; 278:1812–5.
34. Vos MD, Ellis CA, Bell A, Birrer MJ, Clark GJ. Ras uses the novel tumor suppressor RASSF1 as an effector to mediate apoptosis. J Biol Chem 2000;275:35669–72.
35. Ortiz-Vega S, Khokhlatchev A, Nedwidek M, Zhang XF, Dammann R, Pfeifer GP, Avruch J. The putative tumor suppressor *RASSF1A* homodimerizes and heterodimerizes with the Ras-GTP binding protein Nore1. Oncogene 2002;21:1381–90.
36. Mitsudomi T, Viallet J, Linnoila RI, Mulshine JL, Minna JD, Gazdar AF. Mutations of ras genes distinguish a subset of non-small cell lung cancer cell lines from small cell lung cancer cell lines. Oncogene 1991;6:1353–62.

Prognostic models in patients with non-small-cell lung cancer using artificial neural networks in comparison with logistic regression

Taizo Hanai,[1,5,6] Yasushi Yatabe,[2,6] Yusuke Nakayama,[1] Takashi Takahashi,[4] Hiroyuki Honda,[1] Tetsuya Mitsudomi[3] and Takeshi Kobayashi[1]

[1]Department of Biotechnology, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 469-8603, [2]Department of Pathology and Molecular Diagnostics, [3]Department of Thoracic Surgery, Aichi Cancer Center Hospital and [4]Division of Molecular Oncology, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681

(Received January 16, 2003/Revised March 17, 2003/Accepted March 17, 2003)

It is difficult to precisely predict the outcome of each individual patient with non-small-cell lung cancer (NSCLC) by using conventional statistical methods and ordinary clinico-pathological variables. We applied artificial neural networks (ANN) for this purpose. We constructed a prognostic model for 125 NSCLC patients with 17 potential input variables, including 12 clinico-pathological variables (age, sex, smoking index, tumor size, p factor, pT, pN, stage, histology) and 5 immunohistochemical variables (p27 percentage, p27 intensity, p53, cyclin D1, retinoblastoma (RB)), by using the parameter-increasing method (PIM). Using the resultant ANN model, prediction was possible in 104 of 125 patients (83%, judgment ratio (*JR*)) and accuracy for prediction of survival at 5 years was 87%. On the other hand, *JR* and survival prediction accuracy in the logistic regression (LR) model were 37% and 78%, respectively. In addition, ANN outperformed LR for prediction of survival at 1 or 3 years. In these cases, PIM selected p27 intensity and cyclin D1 for the 3-year survival model and p53 for the 1-year survival model in addition to clinico-pathological variables. Finally, even in an independent validation data set of 48 patients, who underwent surgery 10 years later, the present ANN model could predict outcome of patients at 5 years with the *JR* and accuracy of 81% and 77%, respectively. This study demonstrates that ANN is a potentially more useful tool than conventional statistical methods for predicting survival of patients with NSCLC and that inclusion of relevant molecular markers as input variables enhances its predictive ability. (Cancer Sci 2003; 94: 473–477)

Lung cancer is the leading cause of cancer death in Japan[1] as well as in Western countries.[2] Lung cancer is divided into two major morphological types; small-cell lung cancers (SCLCs) and non-small-cell lung cancers (NSCLCs). Although about 30% of NSCLC patients are candidates for potentially curative resection, their long-term survival rate remains unsatisfactory.[3–6] Using existing prognostic tools, such as the commonly used TNM classification,[7,8] however, it is often difficult to accurately predict the outcome of each individual NSCLC patient.[3,9] Tumors that show similar morphology under the microscope may have different sets of genetic alterations and, thus, may exhibit different biological behavior in patients.

We have been evaluating various genetic or epigenetic changes of cancer-related genes in a search for clinically relevant prognosticators. In our previous studies, alterations of cancer-related genes, including cyclin D1,[10] retinoblastoma (RB),[10] p53[11] and p27,[12] have been shown to be of prognostic importance. However, no single variable was sufficiently predictive to precisely foresee a patient's outcome. To overcome this problem, statistical methods of regression have been developed to analyze multiple variables simultaneously. For studies with a binary endpoint, logistic regression (LR) has been used, while for studies with time-to-event data, Cox proportional hazards regression has been utilized as a standard method. However, these methods still have limitations and are not ideal tools for prediction of individual patient's outcome.

Artificial neural networks (ANN) have been developed as an alternative statistical technique in the last 40 years[13] and have been applied in the biochemical and medical fields.[14–17] ANN is a computational methodology that performs multifactorial analyses. In analogy with networks of brain neurons, ANN contains layers of simple computing nodes that operate as nonlinear summing devices.[13] These nodes are interconnected by weighted connection lines, enabling tasks such as predicting outcome values, classifying an object, approximating a function and recognizing a pattern in multifactorial data (Fig. 1).[13] In addition, an ANN model has an output for each set of input variables. Therefore, using ANN, it is possible to predict outcome on an individualized basis.

In the present study, we attempted to predict the prognosis of individual patients using ANN, aiming at achieving better clinical management of the individuals based on the expected risk. The cohort used is a non-biased consecutive series, and the variables of the cohort included not only conventional clinico-pathological factors but also biological markers; i.e., p53, p27, cyclin D1 and RB. The accuracy and efficacy of ANN were compared with those of LR.

Methods

Patients. A series of 125 consecutive patients with NSCLC (32–82 years old (median, 63.3), 88 females and 37 males who underwent potentially curative resection at the Department of Thoracic Surgery, Aichi Cancer Center, Nagoya, from 1986 through 1988) were used for the present study. There were 62 adenocarcinomas, 50 squamous cell carcinomas, five adenosquamous carcinomas and eight large-cell carcinomas. The variables listed in Table 1 were used for the analysis in the present study, which included age, sex, smoking index (number of cigarettes×year), tumor size, p factor (pleural involvement with 4 categories: i.e., p0, no invasion; p1, invasion that does not reach the surface; p2, invasion beyond visceral pleura; p3, invasion to chest wall, diaphragm, mediastinal pleura, parietal pericardium), pT, pN, pathological stage according to the 4th edition of the TNM classification, and histologic type according to the new World Health Organization (WHO) classification (1999). In addition, 48 patients who underwent surgery in 1996

[5]To whom correspondence should be addressed at Laboratory for Bioinformatics, Graduate School of Systems Life Sciences, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.
E-mail: taizo@brs.kyushu-u.ac.jp
[6]These authors have an equal contribution to this article.

were used for validation of the models created with the 1986–1988 cohort (test cases). This group comprised 31 men and 17 women with ages ranging 43 to 76 years (median 64) and included 32 adenocarcinomas, eight squamous cell carcinomas, five adenosquamous carcinomas and three large-cell carcinomas. All cases were coded prior to the initiation of this study in order to keep patients' information anonymous.

Immunohistochemical analysis of biological markers. In the present study, we incorporated several biological variables to improve the efficacy of our models (Table 1). To this end, we used data sets comprising the expression status of p27, p53, cyclin D1 and RB, which have been reported as prognostic factors in our previous studies.[10–12] The procedures of preparation and analysis of tissue samples have also been described in detail.[10–12] Briefly, the standard avidin-biotin-peroxidase complex method was used for immunohistochemical examination of paraffin sections using monoclonal antibodies against $p27^{KIP1}$ (Transduction Laboratory, Lexington, KY), p53 (DO-7, DAKO, Copenhagen, Denmark), cyclin D1 (NCL-cyclin D1, Novocastra Laboratories, Newcastle, UK), and RB (3H9, Medical and Biological Laboratories, Nagoya). In the case of evaluation for p27 expression, both percentage and intensity were included (according to the criteria as reported).

Data preprocessing. In order to use these parameters as input variables for ANN, we standardized all these data into numerical data ranging from 0.05 to 0.95. In this study, three independent models to estimate patient survivals at 1, 3, and 5 years after surgery were constructed. For each estimation model, the output values of survival and death were set to 0.05 and 0.95, respectively. Multicolinearity, which may interfere with the construction of a proper model, would be expected to be present among some of the variables. For example, pT is determined by p factor and tumor size, and pathologic stage is based on pN and pT in the present series. However, calculation of the correlation coefficients for every combination of all variables in Table 1 revealed no disturbingly high correlation, and every correlation coefficient was found to be less than 0.85. Therefore, all 17 variables were used as potential inputs for the current analysis.

ANN. In the present study, a three-layered ANN composed of input, hidden and output layers, was designed as shown in Fig. 1. The output layer had only one unit, which represents the survival status of the patients. Initially, connection weights were randomly assigned values between 0 and 1, and subsequently they were automatically altered by the back propagation method[18] to identify the optimal relationships between input and output. This process is called "learning."

In cases where the number of connection weight parameters is much larger than that of the learning data set, the resultant model may have less generalizability and flexibility.[19] Therefore, in this case it was necessary to decrease the number of input and hidden units so that the number of connection weight parameters was decreased for optimization of the model. To this end, we used the parameter-increasing method (PIM).[20] Briefly, the initial step of PIM was to choose the input variable that was most crucial for accurate prediction. In the next step, the second most crucial variable was selected. By repeating this operation, the best combination of input units was selected in the prediction model. Similarly, the number of units in the hidden layer was decreased one by one from 10, and this procedure was continued until the accuracy of the model had dropped sharply. Practically, the numbers of input units and hidden layer units were different every time because of the random assignment of initial weight connections. Therefore, five optimizing procedures were independently done for both input and hidden layer units, and the input/hidden layer neurons that were most frequently selected were considered as the optimized input/hidden layer neurons.

Subsequently, in order to examine the flexibility of an ANN, cross-validation was performed as follows. At first, the data sets were divided into 5 groups. Then, groups 2, 3, 4 and 5 were used for learning data, and the remaining group 1 was used for evaluation of the trained ANN model. In the next step, groups 1, 3, 4 and 5 were used for learning data, and the remaining group 2 was used for evaluation. All groups of data

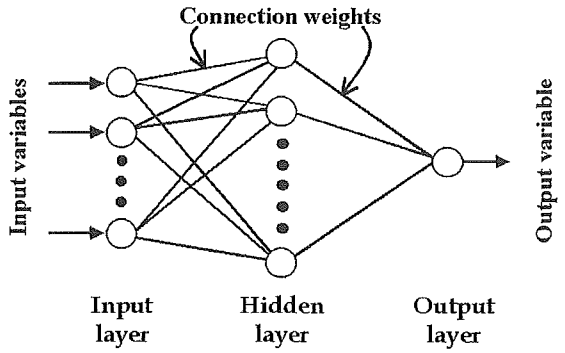


Fig. 1. Artificial neural network.

Table 1. Potential input variables for prognostic models

Input variables	Data type	Value
Age	Continuous	36–82
Sex	Categorical	1: male, 2: female
Smoking index	Continuous	0–2350 cigarettes×year
Tumor size	Continuous	36–110 mm
p factor	Categorical	1: p0, 2: p1, 3: p2, 4: p3
T factor	Categorical	1: T1, 2: T2, 3: T3, 4: T4
N factor	Categorical	1: N0, 2: N1, 3: N2
Pathological stage	Categorical	1: I, 2: II, 3: IIIa, 4: IIIb
Histological types adeno	Categorical	1: yes, 2: no
Histological types squamous	Categorical	1: yes, 2: no
Histological types adenosquamous	Categorical	1: yes, 2: no
Histological types large	Categorical	1: yes, 2: no
p27 percentage	Categorical	1: <5%, 2: 5–30%, 3: 31–60%, 4: 61%<
p27 intensity	Categorical	1: negative, 2: decreased, 3: normal, 4: increased
p53	Categorical	1: negative, 2: positive
Cyclin D1	Categorical	1: negative, 2: positive
RB	Categorical	1: negative, 2: positive

were used as evaluation data in the same way, and the average of all procedures was considered as the estimation ability of ANN. In addition, a completely independent data set was also used to validate the ANN model constructed as described above.

We evaluated the efficacy of our models in terms of two values; i.e., judgment ratio (*JR*) and accuracy. If the patient of interest was dead at a given time point, and the output was larger than the high threshold, the prediction was considered true positive (TP). Conversely, if the output was smaller than the low threshold, the prediction was considered false negative (FN). False positive (FP) and true negative (TN) predictions can be similarly determined. With these numbers, *JR* and accuracy are given in the following equations. The *JR* indicates the proportion of patients on which judgment can be achieved, while accuracy indicates the fraction of *JR* on which the correct judgment was achieved.

$$JR = \frac{N_{TP}+N_{TN}+N_{FP}+N_{FN}}{N_{all}} \times 100 \qquad (1)$$

$$Accuracy = \frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \times 100 \qquad (2)$$

where N_{TP}, N_{TN}, N_{FP}, N_{FN} and N_{all} are the number of TP, TN, FP, FN and all collected data, respectively.

LR modeling. As a control of the modeling method, a conventional statistical prediction model with logistic regression was also constructed. SPSS for Windows (SPSS Regression Models 10.0, SPSS, Inc., Chicago, IL) was used for LR modeling. The input variables for the LR model were optimized by PIM based on the likelihood ratio.

Results

Selected input variables for 1-, 3- and 5-year survival models. Table 2 shows the selected variables for 1-, 3- and 5-year survival models using ANN and LR in the order selected by PIM. Variables selected in the earlier steps are more essential for prognosis prediction. It was noted that several biological variables, especially p53 and p27, were repeatedly selected in the models.

Although a disturbingly high correlation was not observed, we also confirmed that multicolinearity among the input variables does not affect the results by eliminating one of the parameters, such as pathological stage, size, p factor, or p27 expression, from the constructed ANN models. In every attempt, the *JR* and accuracy of the models were lower than those of the models using all input variables (data not shown). The number of the units in hidden layers for the 1-, 3- and 5-year survival estimation models was optimized in a similar manner, resulting in 7, 9 and 10 units, respectively.

Table 2. Selected input variables by the parameter-increasing method

Order of selection	1-year survival		3-year survival		5-year survival	
	ANN	LR	ANN	LR	ANN	LR
1	Stage	Stage	pN	pN	Stage	p
2	SI	SI	p27 intens.	p	Size	pN
3	Age	—	Age	Sex	SI	p27 intens.
4	p	—	pT	SI	p	—
5	p53	—	SI	p27 intens.	p53	—
6	Size	—	Size		Age	—
7	pT	—	CCND1		—	—

Stage, pathological stage; p27 intens., p27 intensity; SI, smoking index; CCND1, cyclin D1.

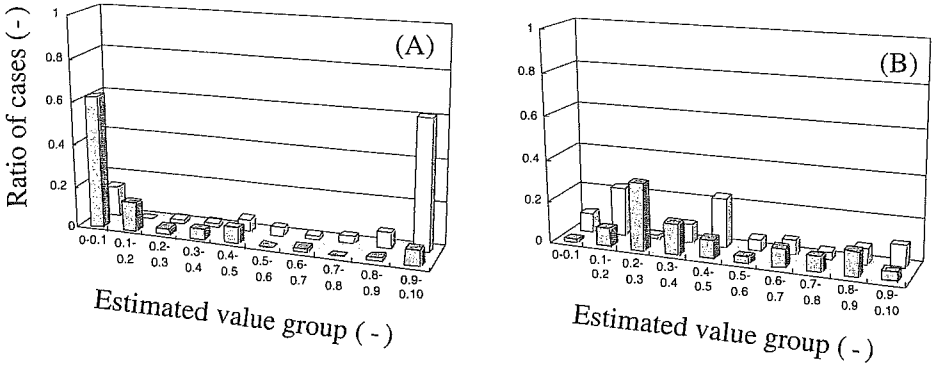


Fig. 2. Results of 5-year survival estimation model using ANN (A) and LR (B). Shaded boxes, "alive" patients; open boxes, "dead" patients.

Table 3. Comparison of predictive models for 1-, 3- and 5-year survival using ANN and LR

	1-year survival		3-year survival		5-year survival	
	ANN	LR	ANN	LR	ANN	LR
JR (%)	77.6	73.6	72.8	50.4	83.2	36.8
No. of patients	97/125	92/125	91/125	63/125	104/125	46/125
Accuracy (%)	93.8	95.7	91.3	87.3	86.8	78.3
No. of patients	91/97	88/92	83/91	55/63	90/104	36/46

Threshold determination. Fig. 2 shows the result of the estimation of 5-year survival using ANN (A) or LR (B). The ANN model discriminated the survival status well, whereas the LR model did not. This clear distinction was also evident in the models estimating 1- and 3-year survival. Since the outputs of the models are given with contiguous values, we configured thresholds for the predictive judgment such that the accuracy of the model was maximized. An increasing accuracy is, to some extent, in conflict with a high *JR*; therefore, we gave accuracy a higher priority than *JR*. By fluctuating the threshold values, the maximal accuracy (86.8%) was achieved with values of 0.2 and 0.8 for the low and high thresholds, respectively. These thresholds were also suitable for the estimation models of 1- and 3-year survival and were therefore used for all estimation models using ANN and LR.

***JR* and accuracy of the estimation models.** Table 3 shows the *JR* and accuracy of the 1-, 3- and 5-year survival estimation models using ANN and LR. Both *JR* and accuracy were superior in the models using ANN to those using LR in most of the cases. For the ANN estimation model of 5-year survival, the *JR* was 83.2% and the accuracy was 86.8%. In contrast, with a 5-year survival model using LR, *JR* was 36.8% and accuracy was 78.3%.

To assess the importance of biological variables, we also constructed models eliminating those variables. *JR* and accuracy of these models were lower than those of the models using all input variables (Tables 4 and 5), suggesting that these biological markers are important for optimal performance of the models. In addition, the impact of p53 and p27 on the prediction was examined by eliminating either of these two biological variables in a similar manner. This also resulted in lower *JR* and accuracy than if all input variables were used in the models. Taken together, these results suggest that biological variables (i.e., p53 and p27) were as important as other clinicopathological variables (Tables 4 and 5).

Validation with an additional independent data set (Table 6). In order to examine the generality of the constructed ANN model, additional data of an independent cohort of 48 patients, who underwent surgery in 1996, were used for validation. *JR* was 77.8% and 81.3% for 3- and 5-year survival, respectively. The *JR* differences between the cohorts with test data (1998 cohort) and learning data (1986–88 cohort) were +5.0% and −1.9%, respectively, for 3-year and 5-year survival. Accuracies with the validation data set (1998 cohort) were 85.7% for 3-year survival and 76.9 % for 5-year survival, which were only 5.6% and 9.9% lower than those with the learning data set (1986–88 cohort).

Table 4. Selected input variables by the parameter-increasing method

Order of selection	5-year survival	5-year survival without variables for p53 and p27	5-year survival without biological variables
1	Stage	Stage	Stage
2	Size	Size	Size
3	SI	SI	SI
4	p	p	p
5	p53	CCND1	Age
6	Age	Age	Sex
7	—	—	—

Stage, pathological stage; SI, smoking index; CCND1, cyclin D1.

Table 5. Comparison of predictive models for 5-year survival

	5-year survival	5-year survival without variables for p53 and p27	5-year survival without biological variables
JR (%)	83.2	59.2	55.2
No. of patients	104/125	74/125	69/125
Accuracy (%)	86.8	83.8	81.2
No. of patients	90/104	62/74	56/69

Table 6. Estimation of 3- and 5-year survival for unlearned data set from 1996 cohort

	3-year survival		5-year survival	
	Learning data	Validation data	Learning data	Validation data
JR (%)	72.8	77.8	83.2	81.3
No. of patients	91/125	14/18	104/125	39/48
Accuracy (%)	91.3	85.7	86.8	76.9
No. of patients	83/91	12/14	90/104	30/39

Learning data were from 1986–1988 cohort and validation data were from 1996 cohort.

Discussion

We have created models for prediction of outcome of each individual NSCLC patient using ANN with immunohistochemical data of biological markers, as well as with conventional clinicopathological variables. *JR* and accuracy of 1-, 3- and 5-year survival estimation models using ANN were superior to those of LR in most cases, except for the accuracy of prediction of 1-year survival by LR, for which only two variables of pathological stage and smoking index were found to be useful (see Table 2). This suggests that the relationship between the input variables and survival status may be correlated rather simply for the prediction of short-term outcome, and that 3- or 5-year survival status may be configured by more complex factors.

We found several biological markers, especially p53 and p27, that were selected as important input variables. Addition of these biological markers to conventional clinico-pathological variables potentiated the predictive ability of the model.[8] Although it is well recognized that the extent of invasion and metastasis is associated with a patient's outcome, additional parameters have been sought to improve accuracy of prediction. However, no single marker known to date has definitive prognostic significance in NSCLC patients. Even for p53, which is one of the molecules that play a central role in cancer development, the prognostic significance is controversial. Although many reports have shown a prognostic significance of p53 status, which was confirmed by meta-analysis,[21] the significance was marginal or at best rather modest in most reports. In the present study, biological markers, including p53, p27 and cyclin D1, were often selected by PIM as more significant input variables for prognostic prediction by ANN, suggesting that the ANN model could resolve their entangled nonlinear relationships. Jefferson *et al.* created predictive models for the outcome of 620 NSCLC patients at 12, 18 and 24 months using a genetic algorithm neural network (GANN, a type [kind] of ANN).[22] The accuracy of 1-year survival estimation using these authors' model was 10% lower than that using our model. This difference may be in part attributable to the fact that these investigators only used clinico-pathological variables (stage, sex, T, N, histologic type and differentiation).

Bellotti *et al.* also constructed an estimation model for prognosis of NSCLC patients using a three-layered ANN.[23] They used 12 variables, including four biological parameters (i.e., S phase fraction, proliferating nuclear antigen, MIB-1 staining

and p53), for the analysis of 67 patients' data.[23] In contrast to our attempt to use PIM to reduce the number of input variables and hidden layers, these investigators used all 12 input variables, 12 units in hidden layers, 1 output variable and 156 connection parameters, which may well be too many for the 67 learning data. Unfortunately, they used all 67 data as learning data and did not evaluate the validity of the model by using an independent data set, which we believe is of crucial importance for the evaluation of generality of the constructed model. In our study, the accuracy observed with the additional independent data set for validation was 9.9% lower than that for the training data set. Taking into account the fact that the validation data set was obtained with patients who underwent surgery about 10 years later, there may have been small but nevertheless important differences in patient selection and management. In addition, there were considerable changes in the distribution of histologies of lung cancers during the decade. Therefore, a decrease in the prediction accuracy by less than 10% may be considered to be reasonably small.

The superiority of ANN to regression models, however, should be interpreted with caution. In a collective review of 28 studies comparing ANN with logistic or Cox regression models,[24] ANN outperformed regression models in 10 of the 28 studies, but was outperformed by regression in 4 studies, and the 2 methods had similar performance in the remaining 14 studies. It is noteworthy that in the 8 largest studies (sample size >5000), regression and ANN tied in 7 cases.

In conclusion, survival estimation models at 1, 3 and 5 years after potentially curative surgery were created using ANN in patients with NSCLC. After optimization of ANN, *JR* and accuracy of ANN models were generally higher than those of LR models. The selected input variables of ANN by PIM contained not only conventional clinico-pathological variables, but also several biological markers. Recent progress in biotechnology now enables us to perform multiple genetic or immunohistological analyses with a short turn-around. cDNA microarray[25,26] and tissue microarray technologies[27] are representatives of these techniques. If they were combined with ANN, this would give the best opportunity to create a highly accurate prognostic model. In this connection, Khan *et al.* have recently shown the potential of combinatorial use of expression profiling and ANN for classification of small round blue-cell tumors into four specific diagnostic categories, which are sometimes difficult to distinguish by light microscopy.[28] In a future study, a larger number of patients should be collected to increase the reliability of the prognostic model. It should become possible to correctly identify patients who are at greater risk of poor clinical outcome using ANN. Such patients are candidates for investigational therapeutic approaches. In clinical trials, this kind of model may also be useful to stratify patients into different prognostic groups.

1. Vital statistics of Japan. Vol. 3. Tokyo: Statistics and Information Department, Ministry of Health and Welfare; 1998. p. 384–411.
2. Wingo PA, Ries LA, Giovino GA, Miller DS, Rosenberg HM, Shopland DR, Thun MJ, Edwards BK. Annual report to the nation on the status of cancer, 1973–1996, with a special section on lung cancer and tobacco smoking. *J Natl Cancer Inst* 1999; **91**: 675–90.
3. Moore DFJ, Lee JS. Staging and prognostic factors: non-small cell lung cancer. In: Pass JBM, Johnson DH, Turrisi AT, editors. Lung cancer: principles and practice. Philadelphia: Lippincott-Raven; 1996. p. 481–94.
4. Lau CL, D'Amico TA, Harpole DH. Clinical and molecular prognostic factors and models for non-small cell lung cancer. In: Pass HI, Mitchell JB, Johnson DH, Turrisi AT, Minna JD, editors. Lung cancer: principles and practice. Philadelphia: Lippincott Williams & Wilkins; 2000. p.602–611.
5. Mountain CF. International staining system for lung cancer. In: Pass HI, Mitchell JB, Johnson DH, Turrisi AT, Minna, JD, editors. Lung cancer: principles and practice. Philadelphia: Lippincott Williams & Wilkins; 2000. p. 591–601.
6. Binsberg RJ, Vokes, EE, Rosenzweig, K. Non-small cell lung cacner. In: dDevita T, Hellman S, Rosenberg, SA, editors. Cancer: principles & practice of oncology. Philadelphia: Lippincott Williams & Wilkins; 2000. p. 925–83.
7. Sobin LH, Wittekind, C. TNM classification of malignant tumours. 5th ed. New York: Wiley-Liss Inc; 1997.
8. Sobin LH. TNM: principles, history, and relation to other prognostic factors. *Cancer* 2001; **91**: 1589–95.
9. Montie JE, Wei JT. Artificial neural networks for prostate carcinoma risk assessment. An overview. *Cancer* 2001; **91**: 1947–52.
10. Nishio M, Koshikawa T, Yatabe Y, Kuroishi T, Suyama M, Nagatake M, Sugiura T, Ariyoshi Y, Mitsudomi T, Takahashi, T. Prognostic significance of cyclin D1 and retinoblastoma expression in combination with p53 abnormalities in primary, resected non-small cell lung cancers. *Clin Cancer Res* 1997; **3**: 1051–8.
11. Nishio M, Koshikawa T, Kuroishi T, Suyama M, Uchida K, Takagi Y, Washimi O, Sugiura T, Ariyoshi Y, Takahashi T, Ueda R. Prognostic significance of abnormal p53 accumulation in primary, resected non-small-cell lung cancers. *J Clin Oncol* 1996; **14**: 497–502.
12. Yatabe Y, Masuda A, Koshikawa T, Nakamura S, Kuroishi T, Osada H, Takahashi T, Mitsudomi T. $p27^{KIP1}$ in human lung cancers: differential changes in small cell and non-small cell carcinomas. *Cancer Res* 1998; **58**: 1042–7.
13. Dayhoff JE, DeLeo JM. Artificial neural networks: opening the black box. *Cancer* 2001; **91**: 1615–35.
14. Baxt WG, Skora J. Prospective validation of artificial neural network trained to identify acute myocardial infarction. *Lancet* 1996; **347**: 12–5.
15. Ball G, Mian S, Holding F, Allibone RO, Lowe J, Ali S, Li G, McCardle S, Ellis IO, Creaser C, Rees RC. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumors and rapid identification of potential biomarkers. *Bioinformatics* 2002; **18**: 395–404.
16. Milik M, Sauer D, Brunmark AP, Yuan L, Vitiello A, Jackson MR, Peterson PA, Skolnick J, Glass CA. Application of an artificial neural network to predict specific class I MHC binding peptide sequences. *Nat Biotechnol* 1998; **16**: 753–6.
17. Lubbert A, Simuits R. Using measuremant data in bioprocess modelling and control. *TIBTECH* 1994; **12**: 304–11.
18. Rumelhart DE, Hinton GE, Williams RL. Learning representations by back-propagating errors. *Nature* 1986; **323**: 533–6.
19. Wasserman PD. Advanced methods in neural computing. New York: Van Nostrand Reinhold, a Division of Wadsworth Inc; 1993.
20. Hanai T, Katayama A, Honda H, Kobayashi T. Automatic fuzzy modeling for Ginjo sake brewing process using fuzzy neural networks. *J Chem Eng Jpn* 1997; **30**: 94-100.
21. Mitsudomi T, Hamajima N, Ogawa M, Takahashi T. Prognostic significance of p53 alterations in patients with non-small cell lung cancer: a meta-analysis. *Clin Cancer Res* 2000; **6**: 4055–63.
22. Jefferson MF, Pendleton N, Lucas SB, Horan MA. Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma. *Cancer* 1997; **79**: 1338–42.
23. Bellotti M, Elsner B, Paez De Lima A, Esteva H, Marchevsky AM. Neural networks as a prognostic tool for patients with non-small cell carcinoma of the lung. *Mod Pathol* 1997; **10**: 1221–7.
24. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer* 2001; **91**: 1636–42.
25. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; **207**: 467–70.
26. Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H. Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm Genome* 1992; **3**: 609–19.
27. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998; **4**: 844–7.
28. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthoud F, Schwab M, Antonescu CR, Peterson C, Meltzer PS. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001; **7**: 673–9.

Draft Diagnostic Guidelines for Non-Mass Image-Forming Lesions by the Japan Association of Breast and Thyroid Sonology (JABTS) and the Japan Society of Ultrasonics in Medicine

Tokiko Endo[1,2], Mitsuhiro Kubota[1], Yutaka Konishi[1], Kazuhiro Shimamoto[1], Kumiko Tanaka[1], Hiroko Tsunoda-Shimizu[1], Hideyuki Hashimoto[1], Norikazu Masuda[1], Mitsuhiro Mizutani[1], Isamu Morishima[1], Hidemitsu Yasuda[1], Takanori Watanabe[1], and Ei Ueno[1]

Background

Recently, ultrasonic instruments have remarkably improved, and smaller or earlier breast cancers have been found. Also, mammographic screening for breast cancer for women of 50 years and older has heightened people's desire to find and diagnose smaller or earlier lesions. The lesions that do not form mass images have been recognized, and a lexicon for reporting these is desired.

We have been discussing the diagnostic guidelines for breast cancer for the past 3 years. Non-mass image-forming lesions are contained as the objects of diagnosis. We present the tentative plan of the guidelines for non-mass image-forming lesions here.

Definition of the Non-Mass Image-Forming Lesions

Non-mass image-forming lesions are those lesions that are difficult to recognize as a "mass image." They may associate with "mass image-forming lesions." The ultrasonic images of breast disease consist of mass image-forming lesions and non-mass image-forming lesions.

Normal Breast Sonograms and Variants

Normal breast sonograms and their variants are the essential knowledge for understanding non-mass image-forming lesions. These factors may have an effect on ultrasonic breast images:

[1] Japan Association of Breast and Thyroid Sonology, Japan
[2] Department of Radiology, National Nagoya Hospital, 4-1-1 Sannomaru, Naka-ku, Nagoya 460-0001, Japan

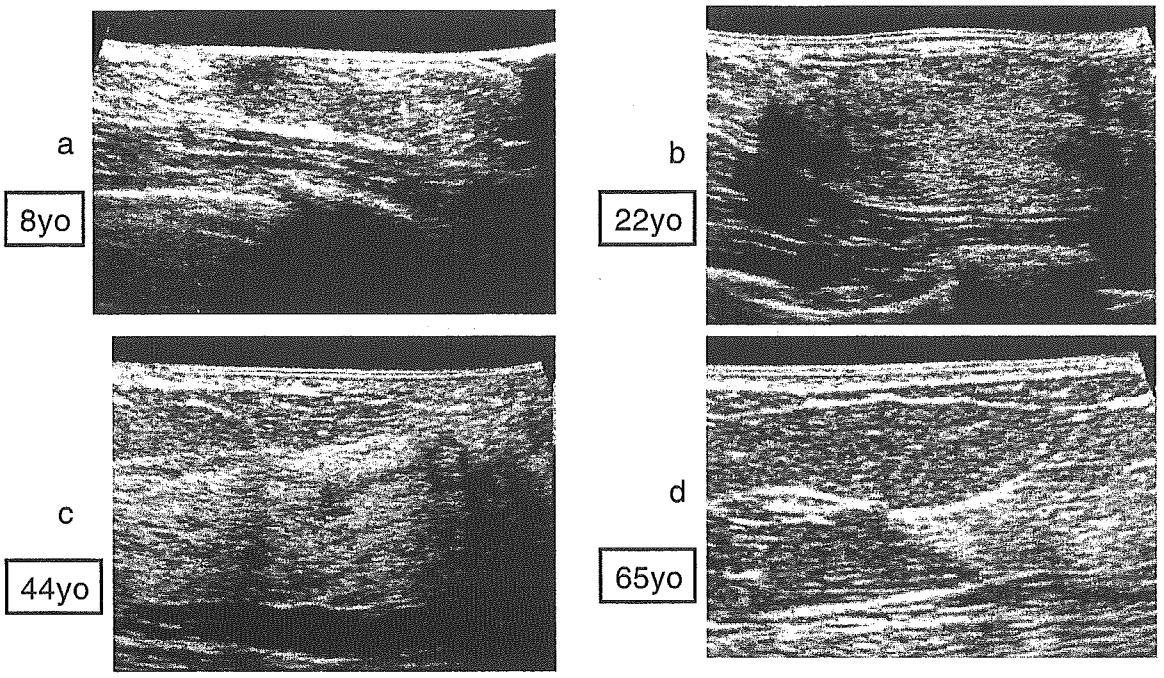


FIG. 1. Ultrasonic breast images and patient age

Age (Fig. 1)
Pregnancy
Breast-feeding
Hormone replacement therapy

Lesions That May Be Observed by Ultrasonic Examination as Non-Mass Image-Forming Lesions

- Duct dilatation
- Duct ectasia (contains plasma cell mastitis)
- Intraductal papilloma, multiple intraductal papilloma
- Mastopathy
- Epithelial hyperplasia
  - Adenosis
  - Multiple cyst
  - Fibroadenomatoid hyperplasia
  - Fibrosis
- Mastitis
  - Lymphocytic mastitis
  - Acute mastitis.
- Radial scar, complex sclerosing lesion
- Noninvasive ductal carcinoma
- Invasive ductal carcinoma with a predominant intraductal component
- Invasive carcinoma

Lexicon for Non-Mass Image-Forming Lesions

- Dilation of the duct
  Dilated ducts with or without internal echoes that may be in any area
- Wall thickening of the duct
  The wall of the duct is increased in thickness more than usual
- Irregularity of the caliber of the duct
  Irregularity of the anechoic area in the duct
- Internal echoes in the duct or tiny cysts
  Echoes in the duct or tiny cysts as follows:
    Solid echoes
    Floating echoes
    Linear high echoes
    High echo spots
    Fine high echo spots (smaller than 1 mm in diameter)
- Multi-vesicular pattern
  Multiple tiny or small cysts in the breast tissue
- Low echo area in the breast tissue
  Low echo area whose character is different from surrounding gland or same area in the ipsilateral breast (Fig. 2)
  Spotted or mottled low echo area
    Relatively small low echo areas form the spotted (or mottled) pattern
- Geographical low echo area
  Low echo area looks like geography as if spotted low echo areas fused into one
- Low echo area with indistinct margin
  Low echo area whose margins are not clearly defined
- Architectural distortion
  Distorted structure without mass image

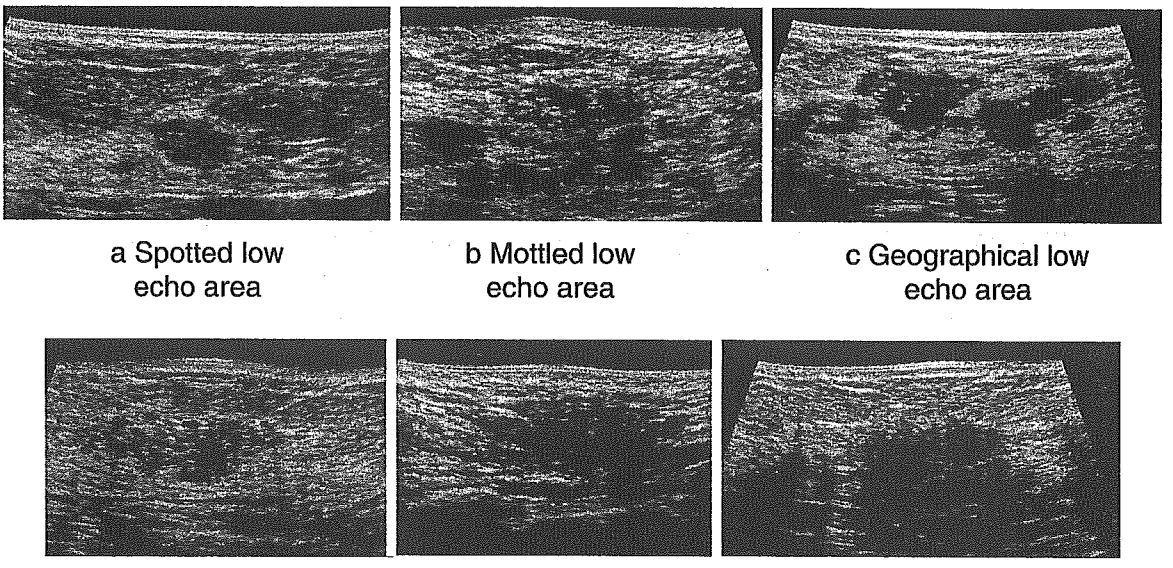


FIG. 2. Low echo area in the breast tissue

Assessment and Categories

Assessment categories are decided as follows:

Category 0: Assessment is incomplete
Category 1: Negative
Category 2: Benign
Category 3: Benign, but malignancy cannot be ruled out
Category 4: Suspicious abnormality
Category 5: Highly suggestive of malignancy

Duct Dilatation (a): Duct Dilatation Without Internal Echoes (Fig. 3)

Dilated ducts with no internal echoes can be seen in the peripheral area outside the areola. They may be complicated with wall thickening by inflammation.

Bilaterally and multiple: category 2
  Dilated ducts
Solitary: category 3
  Dilated duct
  Duct ectasia

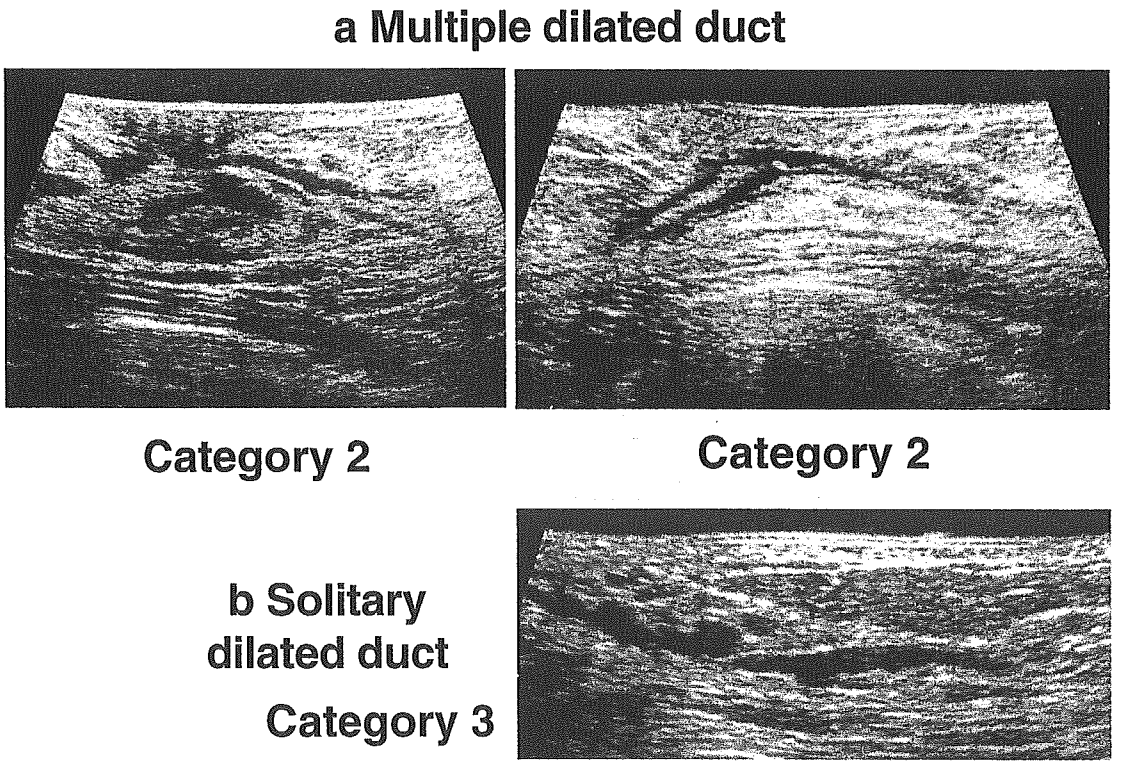


FIG. 3. Duct dilatation without internal echoes

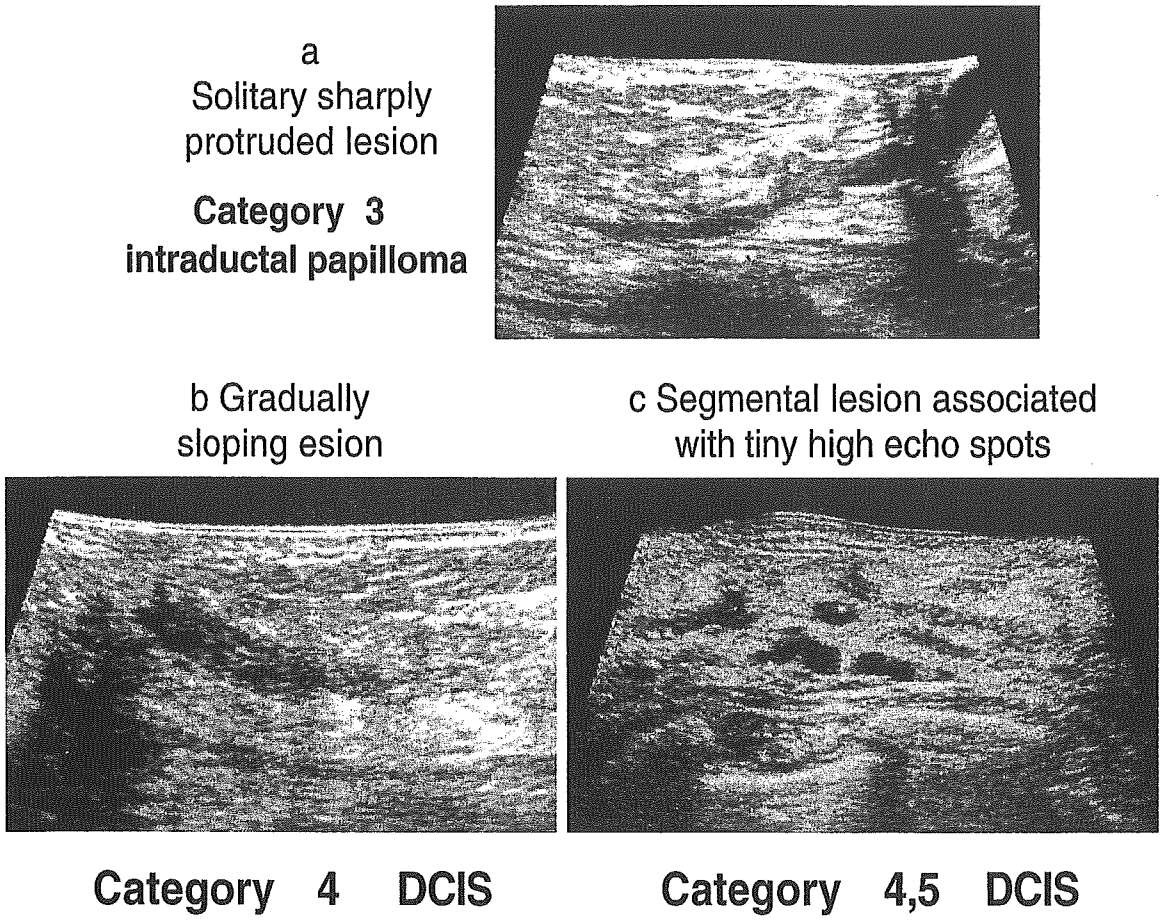


FIG. 4. Dilated ducts with internal solid echoes. *DCIS*, ductal careinoma in situ

Epithelial hyperplasia
Intraductal papillomas, noninvasive ductal carcinoma

*Secondary duct dilatation may be developed from intraductal proliferative lesions.
*When it is difficult to judge whether the internal echoes are there or not, it should be regarded as (b) (below).

Duct Dilation (b): Duct Dilatation with Internal Echoes (Fig. 4)

Intraductal echoes consist of solid echoes, floating echoes, linear high echoes, high echo spots, and fine high echo spots. Solid echoes often result from proliferative lesions; careful observation of the wall is needed. Internal echoes are produced by the floating components in the fluid. Condensed milk or blood is common.

Assessment of Duct Dilatation with Internal Echoes

Shape of the solid echoes

- Sharply protruded: Category 3
  Intraductal papilloma