

《原 著》

## 201TI SPECT・MRI による脳疾患診断の医師間変動の評価

町田喜久雄\*1 本田 憲業\*1 松本 徹\*2 油井 信春\*3  
 戸川 貴史\*3 久保 敦司\*4 塚谷 泰司\*4 小山田日吉丸\*4  
 石井 勝己\*5 西巻 博\*5 森 豊\*6 宇野 公一\*7  
 村田 啓\*8 丸野 広大\*8

**要旨** 54例の脳腫瘍患者のTI SPECT像を9人の医師が読影する実験を2回に分けて行った。医師は、第1回目はMRIを参照せず、第2回目はMRIを参照して、201TIの異常集積の存在診断と部位の同定および良性・悪性鑑別を行った。これよりMRI参照の有無時における診断結果の医師間変動を比較した。

TI-SPECTの異常所見の検出率はMRIを参照しない時医師全体で84%、参照時94%となり、医師間変動はMRIを参照することで有意に減少した。良性と悪性の鑑別能を表す医師全体のROC曲線の第1動作点はMRIを参照しない時TPR:53%、FPR:28%であり、MRI参照時TPR:74%、FPR:45%となったが、ROC曲線下面積Azおよびその医師間変動はMRI参照の有無と関係を生じなかった。201TI異常所見部位の同定では、前頭葉と側頭葉および脳幹の判断基準について医師間で合意を図る必要性が示唆された。

(核医学 33: 943-952, 1996)

### I. はじめに

われわれは従来より、各種シンチグラム検査の臨床の有効性を測るため、複数の施設が協力して、まず確定診断付きの画像データベースを作成

し、次にこれを多数の医師が読影する実験を行い、その結果をROC解析などを用いて統計解析し、所見検出能や質的診断能およびそれらの医師間変動を定量的、客観的に評価する研究を行ってきた<sup>1-8)</sup>。

今回は、最近、脳腫瘍の質的診断、とくに悪性脳腫瘍の診断に有用<sup>9-14)</sup>といわれて注目されている201TI SPECT像の有効性について検討を加え、多数医師によるその読影結果の変動を評価し、若干の知見を得たので報告する。

### II. 材料および方法

#### 1) 画像データベースの作成

本研究で実験に供された画像データベースは千葉県がんセンターにおいて脳疾患患者54例に対して、三検出器SPECT装置(東芝製GCA 9300A)により撮像された201TI SPECT像とシーメンス社

\*1 埼玉医科大学総合医療センター

\*2 放射線医学総合研究所

\*3 千葉県がんセンター

\*4 慶應義塾大学医学部放射線科

\*5 北里大学医学部放射線科

\*6 東京慈恵会医科大学放射線科

\*7 千葉大学医学部

\*8 虎の門病院核医学科

受付: 8年2月23日

最終稿受付: 8年6月10日

別刷請求先: 川越市鴨田辻道町1981 (☎350)

埼玉医科大学総合医療センター

放射線科学教室

町田 喜久雄

製 Signa 1.5 T により撮像された MRI から成る。

SPECT 像は  $^{201}\text{Tl}$  74 MBq (2 mCi) を静注後約 20 分の early image である。MRI は Gd-DTPA 造影前後の T1 強調像である。

画像表示条件は、 $^{201}\text{Tl}$  SPECT 像は 1 患者ごとに 1 枚の 4 ヶ切フィルムに冠状断、矢状断、横断像を、MRI は 1 患者当たり 2-3 枚の半切フィルムに横断像を一定の形式で出力したものである。

これらの画像のほか、各患者ごとに臨床所見データ (1. 症例番号, 2. 患者氏名: 性別, 年齢, 3. 診断名, 4. 組織診断名, 5. 神経学的所見, 6. 診断根拠) を一定の形式で記録したシートを作成した。

## 2) 読影講習会

$^{201}\text{Tl}$  SPECT 像読影実験の 3 週間前、実験参加医師は  $^{201}\text{Tl}$  SPECT 像および MRI による腫瘍診断の読影基準や T1 所見の存在区域の決め方について医師間で合意を図るため、 $^{201}\text{Tl}$  SPECT 像の診断に習熟した 1 人の医師から講習を受けた。

$^{201}\text{Tl}$  SPECT の読影における悪性、良性の判定に関しては、集積の強さ、病巣数、病巣の大きさ、集積の広がりなどの所見に基づき、各読影担当医が判定することとした。また、正常組織でも集積する部位として脳下垂体、頭皮等があることも考慮することとした。

## 3) 読影医

7 施設に所属する医師 9 人が読影医となった。5-30 年の経験を有する核医学専門医であるが、頭部  $^{201}\text{Tl}$  SPECT 像の読影は、画像データベース提出施設の 1 人を除き、これまでほとんど経験したことがなかった。

## 4) 読影実験

読影実験は以下の手順により 1 日のうちに 2 回行った。

① 最初の実験で、医師は脳疾患患者 54 例の  $^{201}\text{Tl}$  SPECT 像を読影した。その際、前述の臨床所見データのうち、患者の性別、年齢および神経学的所見 (意識状態が正常か異常かのほか、運動麻痺、感覚障害、言語障害: 構語障害および失語、視力障害、頭痛、小脳症状、付随意運動など

の有無) のみが参考にされた。さらに、その他のデータとして脳の横断解剖図が所見の存在区域を判断する際の補助資料として適宜利用された。

読影項目は、「画質評価」、「異常集積の存在確信度 (脳全体)」、(確信度は、存在する、多分存在する、五分五分、多分存在しない、存在しないの 5 段階)、「所見の区域」、「異常集積の程度」、「所見の存在確信度」、および「診断」(正常、良性、多分良性、多分悪性、悪性の 5 段階) である。

$^{201}\text{Tl}$  SPECT 像単独による所見区域の同定は困難と考えられたので、 $^{201}\text{Tl}$  SPECT 正常像の典型例のコピー用紙を用意し、近似的に該当すると思われる横断面上に所見位置をスケッチすることとした。

$^{201}\text{Tl}$  の集積度の「強、中、弱」の判定基準は、頭皮への集積度と比較して頭皮と同程度は「中」、それ以上は「強」、それ以下は「弱」とした。

② 次に  $^{201}\text{Tl}$  SPECT 像に MRI を加えて①と同様に読影した。

最初に「 $^{201}\text{Tl}$  診断を変えない」か「変える」かを判定した。「変える」場合は必要に応じて訂正、変更、追加した。

なお、MRI を参照したこの実験では、医師は  $^{201}\text{Tl}$  SPECT 像で検出した所見の解剖学的区域を判定した。

解剖学的区域は前頭葉、頭頂葉、側頭葉、後頭葉、小脳、基底核、視床、脳幹の 8 区域とその左右からなる 16 区域および脳以外 (頭皮、骨) の合わせて 18 区域に対して、1 つの所見ごとにその解剖学的な広がりを区域をチェックすることで記録した。

## 5) 最終診断

各症例ごとに  $^{201}\text{Tl}$  SPECT 像以外の診断根拠として使われた神経学的所見、剖検、手術、アンギオ、臨床診断、CT、生検、MRI、 $^{67}\text{Ga}$  など、すべての情報を MRI 診断専門医と核医学専門医各 1 名が総合判断し、ゴールドスタンダードとしての最終診断レポートを決定した。なお、上記診断根拠として 1 例当たり平均 4-5 の方法が使われた。MRI は 54 例中 52 例で診断根拠の一つとし

て挙げられているが、MRI のみの例はなかった。診断根拠が MRI・アンギオ・臨診・CT・TI の組み合わせによるものが 5 例、MRI・CT・TI が 3 例、MRI・CT・Ga・TI が 1 例あった。それ以外の症例では手術が診断根拠 41 例、臨床診断 28 例、アンギオ 11 例、生検 11 例、剖検 4 例であり、54 例中 45 例は手術・生検・剖検のいずれかまたはその組み合わせにより診断が確定した症例であった。

脳疾患患者の内訳は転移性脳腫瘍が 36 例(肺癌 23 例、乳癌 5 例、胃癌 2 例、悪性リンパ腫、甲状腺癌、腎癌、直腸癌、悪性線維性組織球腫、骨肉腫、各 1 例)、原発性脳腫瘍 16 例(膠芽腫 8 例、星細胞腫 4 例、松果体細胞腫、髄膜腫、聴神経鞘腫、頭蓋咽頭腫各 1 例)および脳腫瘍 1 例、多発性硬化症 1 例であった。症例の性別・年齢分布は、男性 35 例(内訳：10 歳代 2、20 歳代 1、30 歳代 2、40 歳代 6、50 歳代 9、60 歳代 10、70 歳代 5)、女性 18 例(内訳：10 歳代 0、20 歳代 2、30 歳代 1、40 歳代 7、50 歳代 3、60 歳代 2、70 歳代 3)であった。また、全例とも、<sup>201</sup>Tl SPECT 像上で異常所見を少なくとも 1 つ以上有しており、悪性度分布は正常 0 例、良性 6 例、悪性 48 例で構成された。ただし、読影実験に際して性別、年齢、神経学的所見は読影医に示されたが、各症例がなんらかの異常所見を有していたことや良性・悪性分布に関する情報は示されなかった。

6) 統計解析

各症例ごとにゴールドスタンダードと各医師の読影結果とを比較し、<sup>201</sup>Tl SPECT 像単独時および MRI を追加して <sup>201</sup>Tl SPECT 像を再読影した時(以下 <sup>201</sup>Tl SPECT+MRI という)の <sup>201</sup>Tl SPECT 像による異常所見の検出能、所見位置検出能および悪性鑑別能を医師ごとおよび医師全体で定量し、それらの医師間変動を評価した。

ROC 曲線下面積 (Az) の有意差検定は Hanley らの方法<sup>15)</sup> によった。Az、異常所見の検出数、所見の存在区域の判定数および悪性度診断の判定数に対する 2 群間の平均の差は対応のある t-検定で、2 群間の分散(医師間変動)の相違は F 検定

で行った<sup>16)</sup>。

III. 結 果

1) 画質評価

Fig. 1 に示すごとく本実験で読影に供された <sup>201</sup>Tl SPECT 像の画質は全体の約 98% が読影医により「優、良、可」と評価されたものであった。

2) 所見検出能

Fig. 2 (a, b) に「脳全体」としての異常所見の存在確信度分布を示す。ただし、1 人の医師が 1 症例より所見を複数指摘し、複数の存在確信度を答えた時は最大スコアをその症例に対する「脳全体」の存在確信度とした。

<sup>201</sup>Tl SPECT 像単独ではスコア 5:「存在する」の医師 1 人当たりの判定数の平均は 45.4±4.53(検出率:84.1%)であり、スコア 2:「多分存在しない」およびスコア 4:「多分存在する」の判定が目立つ。一方、<sup>201</sup>Tl SPECT+MRI では、曖昧な判定が減り、スコア 5 の回答数の平均は 50.9±1.691(検出率:94.4%)となり、統計的に有意に増加した(p<0.007)。

また、<sup>201</sup>Tl SPECT+MRI 時のスコア 5「存在す

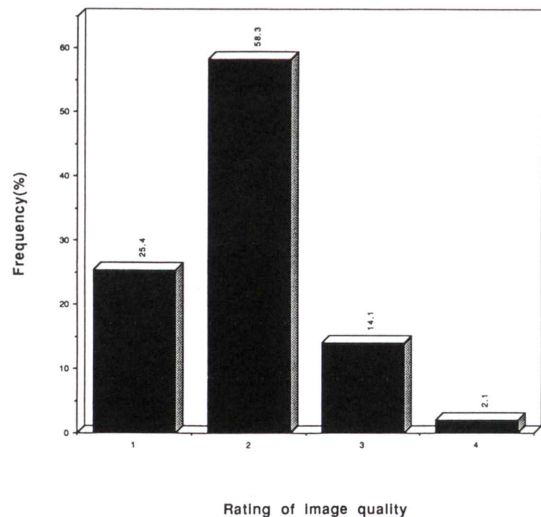
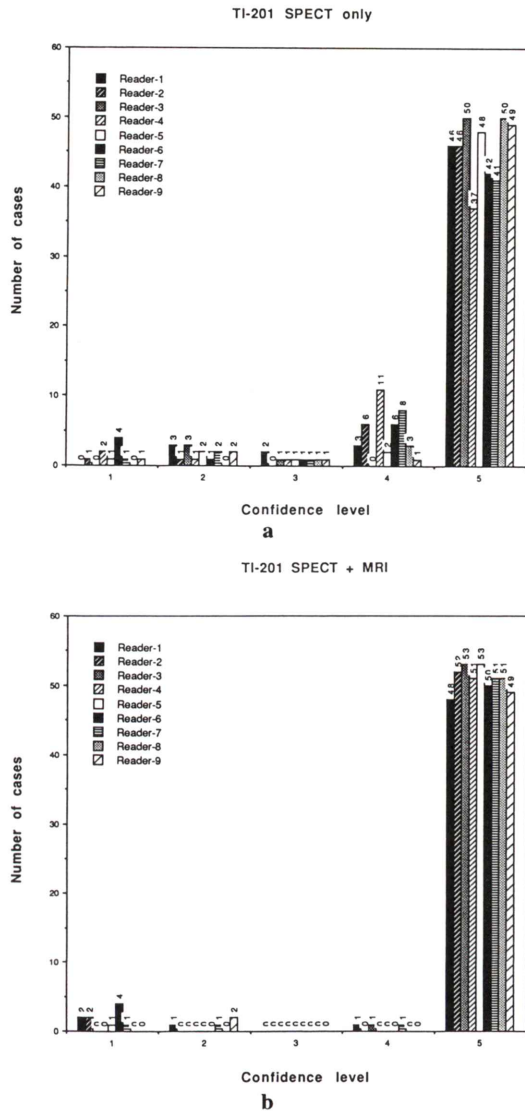


Fig. 1 Evaluation of image quality for <sup>201</sup>Tl SPECT by 9 observers. Score 1: excellent, 2: good, 3: fair, 4: poor.



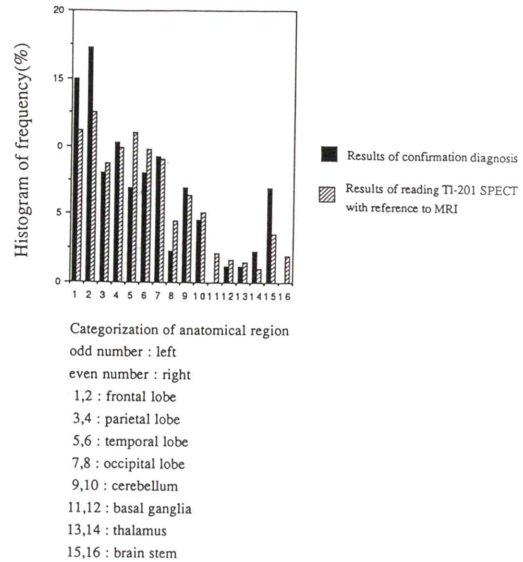


**Fig. 2** Histogram of confidence level of diagnosis of existence of the abnormal accumulation of radionuclide by <sup>201</sup>Tl SPECT without; (a) and with; (b) MRI for the individual observer. Score 1: absent, 2: probably absent, 3: indeterminate, 4: probably present, 5: present.

る」判定数の分散 (医師間変動) は <sup>201</sup>Tl SPECT 像単独時より有意に減少した (p<0.01).

**3) 所見位置判定能**

Fig. 3 に読影医 9 人により判定された異常所見



**Fig. 3** Comparison of the frequency distribution of 16 anatomical regions categorized by reading of <sup>201</sup>Tl SPECT with referring MRI with that of the confirmation diagnosis.

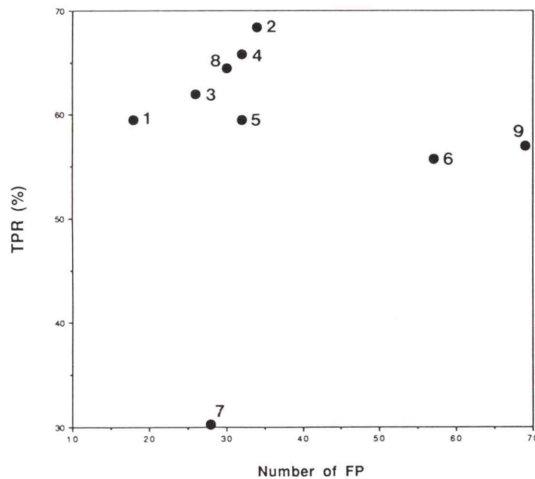
の存在区域の頻度分布と最終診断から得られた区域の頻度分布を比較した。ただし、それぞれの分布は全体が 100% になるよう正規化してある。

読影結果は、最終診断に比べて前頭葉および脳幹の判定が少なく側頭葉の判定が多くなされた。最終診断と読影診断の頻度分布の差は統計的に有意でなかった。

また、読影医 9 人の任意の 2 者間 (全 36 通り) で 16 区域の所見存在判定数の差を同様に検定した結果、8 通りの組み合わせ (22%) で有意の差 (p<0.05) を生じた。

次に、各症例の各区域について、該当区域内の所見の有無 (数不問) の区域判定的中、非的中を下記により求めた。

すなわち、最終診断で各症例ごとに、所見の有無および数に関係なく 16 の解剖学的区域を、<sup>201</sup>Tl SPECT 像の異常所見が存在する区域と存在しない区域に二分した。一方、各医師の読影結果に対しても症例ごとに、検出された所見数に関係なく、異常所見が存在するとしてチェックされた



**Fig. 4** Performance of observers for diagnosis of anatomical region corresponding with abnormal findings of <sup>201</sup>Tl SPECT by the individual observer. Marked interobserver variation is evident.

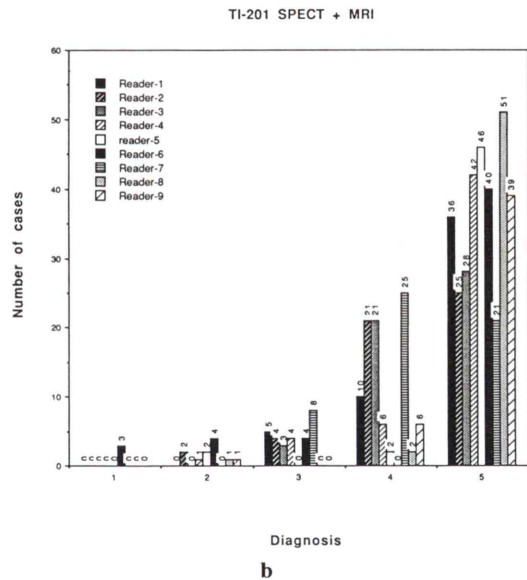
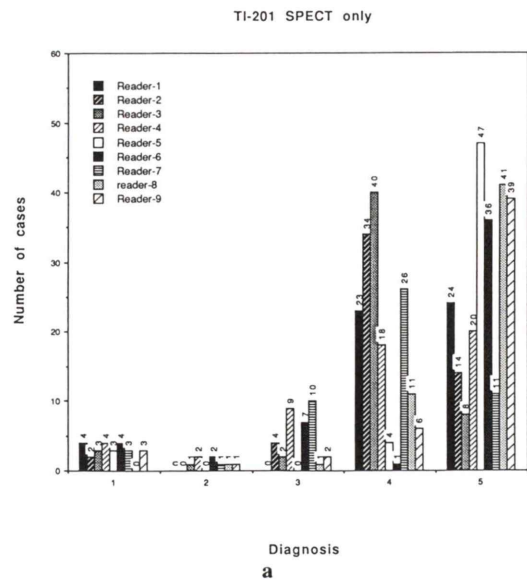
区域とされなかった区域に分類した。上記最終診断と読影診断の対応する左右 16 区域を照合し、最終診断で所見が存在する区域を読影により指摘した：TP，最終診断で所見が存在しない区域を読影により指摘しなかった：TN，最終診断で所見が存在しない区域を読影により指摘した：FP，最終診断で所見が存在する区域を読影により指摘しなかった：FN，それぞれの数をカウントした。

Fig. 4 に FP 数を横軸に、TPR を縦軸にプロットした結果を示す。TP 数は医師間で最小 24 区域から最大 54 区域の範囲に変動した。最終診断では症例全体で 79 区域に所見が存在したので、TPR は 30.3–68.4% 変動したことになる。一方、FP 数は最小 18 から最大 69 の範囲に変動した。

4) 悪性鑑別能

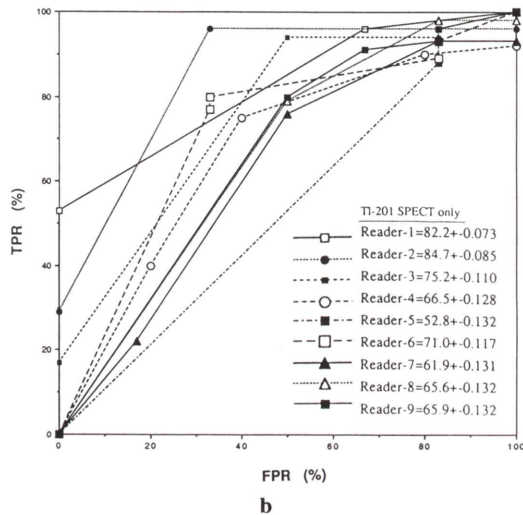
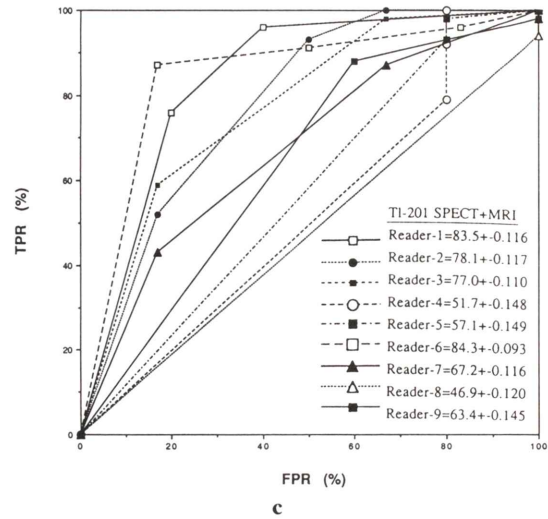
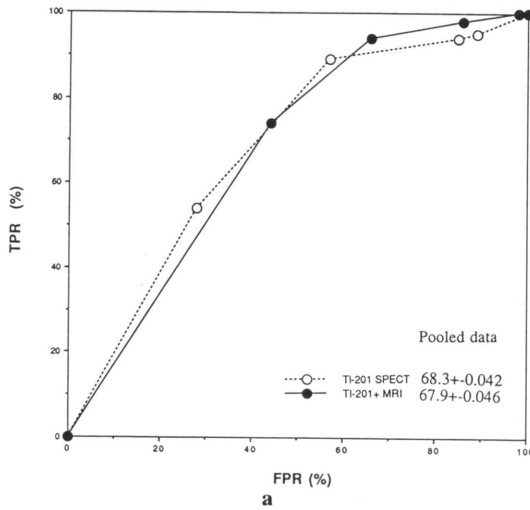
① 最終診断と照合しない場合

Fig. 5 (a, b) に医師 9 人が行った悪性度判定の結果を示す。<sup>201</sup>Tl SPECT 像単独では「悪性」判定数が平均  $26.7 \pm 14.4$  (症例全体の 49%) であったのが、MRI を追加することにより  $36.4 \pm 9.9$  (症例全体の 67.4%) と有意 ( $p < 0.006$ ) に増加した。



**Fig. 5** Histogram of diagnosis of quality of the disease obtained by reading <sup>201</sup>Tl SPECT without; (a) and with; (b) MRI for the individual observer. Score 1: normal, 2: benign, 3: probably benign, 4: probably malignant, 5: malignant.

<sup>201</sup>Tl SPECT 像単独時および <sup>201</sup>Tl SPECT+MRI 時における「悪性」判定数の分散 (医師間変動) の相違は統計的に有意でなかった。



**Fig. 6** ROC curves of diagnosis of quality of the disease by reading  $^{201}\text{Tl}$  SPECT without and with MRI. (a) pooled data by 9 observers, (b) data by reading  $^{201}\text{Tl}$  SPECT alone, (c) data by reading  $^{201}\text{Tl}$  SPECT with reference to MRI. In figure (b), difference of Az (area under the ROC curve) between reader-1 and -5 was statistically significant ( $p < 0.011$ ). In figure (c), the difference of Az between reader-1 and -4, reader-3 and -4, and reader-5 and -6, was statistically significant ( $p < 0.040$ ,  $p < 0.045$ ,  $p < 0.027$ , respectively). The remaining combination between any two readers was not statistically significant. Marked inter-observer variation in the diagnostic performance is evident in figure (b) and (c).

## ② 最終診断と照合した場合

悪性 48 例, 良性 6 例に対する医師 9 人の読影結果について ROC 解析した。

Fig. 6 (a) に pooled data の結果を示す。横軸 (FPR) は最終診断「良性」を「悪性」または「多分悪性」, 「多分良性」, 「良性」と判定した結果の累積%, 縦軸 (TPR) は最終診断「悪性」に対する同様の累積%である。 $^{201}\text{Tl}$  SPECT 像単独時の第 1 動作点の TPR は 53%, FPR は 28% であり,  $^{201}\text{Tl}$  SPECT+MRI 時の TPR は 74%, FPR は 45% であった。ROC 曲線下面積 (Az) で悪性鑑別能を定

量し有意差検定した結果,  $^{201}\text{Tl}$  SPECT 像単独と  $^{201}\text{Tl}$  SPECT+MRI とで差は生じなかった。

Fig. 6 (b, c) に個人ごとの ROC 解析の結果を示す。読影医全体としては,  $^{201}\text{Tl}$  SPECT 像単独と  $^{201}\text{Tl}$  SPECT+MRI とで Az に有意の差は生じなかった。読影医 9 人の任意の 2 者間で Az の差を検定した結果, 有意 ( $p < 0.05$ ) であった組み合わせは  $^{201}\text{Tl}$  SPECT 像単独で 1 組 (3%),  $^{201}\text{Tl}$  SPECT+MRI では 3 組 (8.3%) あった。

$^{201}\text{Tl}$  SPECT 像単独時と  $^{201}\text{Tl}$  SPECT+MRI 時の Az の分散 (医師間変動) の相違は統計的に有意でなかった。



#### IV. 考 察

画像診断の有効度は技術的成熟度と検査法の成熟度の両面から検討する<sup>17)</sup>必要がある。われわれは以前より各種核医学画像の臨床の有効度を評価するため、上記両面からグループ研究を行ってきた<sup>1-8)</sup>。

<sup>201</sup>Tl SPECTによる腫瘍診断に対する有用性の報告がいくつかなされている<sup>18,19)</sup>が、今回は、最近注目を集めている脳の<sup>201</sup>Tl SPECT像を対象とした。<sup>201</sup>Tl SPECT像は脳腫瘍の悪性度評価、放射線治療効果の判定、放射線治療後の再発か放射線壊死かの鑑別に有用であるとされている<sup>9-15)</sup>。現段階では、SPECT像上の異常所見の存在や良性・悪性の判定に関して標準的な基準は定まっていないのが実状と思われる。そこで本研究では、<sup>201</sup>Tl SPECT像の読影基準の統一を図るための第一歩として、7施設に所属する9人の医師が1施設で収集された<sup>201</sup>Tl SPECT像を読影する実験を行い、脳腫瘍の存在診断や質的診断が医師間でどの程度変動するか評価した。

なお、実験に参加した医師9人中1名は画像データベースを提出した施設に所属する医師であったが、他施設の8名と同様本実験で対象とした<sup>201</sup>Tl 脳 SPECT像を本格的に読影するのは今回がはじめてであった。

これまでの研究は自施設の症例について自施設の医師が読影評価したものであり、このような読影実験の報告は少ないのが実状である。また、複数施設の医師が参加した読影実験により、客観的に診断の医師間変動を評価した報告もほとんど知られていない。

画像診断の有効度は医師という人間系をも含めて評価しなくてはならないが、本研究では以前の経験を踏まえて<sup>8)</sup>、1施設の1つの装置で撮像した<sup>201</sup>Tl SPECT像を用いることによって、技術的因子による読影結果の変動を抑える方法を採用した。

本実験のため1施設より読影に供された画像の画質は9人の医師により評価され、そのうちの8

人は他施設の医師であったが、Fig. 1に示すごとく「不可」と評価された<sup>201</sup>Tl SPECT像は僅か2.1%に過ぎず、画質が劣悪であることによる診断能の過小評価はなかったと推測される。また、われわれが以前に行った複数施設から収集した撮像条件・表示条件の異なる<sup>123</sup>I-IMP SPECT像を対象とした読影実験<sup>8)</sup>に比べて、撮像法や画像表示法は一定であり、画質的にも問題は少なかったため、技術的因子が読影結果の変動に寄与した程度は小さかったと考えられる。

医師全体の診断精度の観点から次のようなことが言える。

異常所見の存在診断の的中率はFig. 2に示す通り、<sup>201</sup>Tl SPECT像のみでも「存在する」は84.1% (読影医9人の平均)とかなり高率であった。このように高い値を示すということは、この検査法が有用なことを示す一つの証拠であろう。一方、腫瘍性病変の診断に大きな役割を果たしているMRIを加えて読影すると<sup>201</sup>Tl SPECT像単独で「五分五分」および「多分存在する」の判定が0またはほとんど0%になり「存在する」の割合がその分だけ増加し94.4% (平均)になった。このことは存在診断におけるMRIの有用性を再確認したものである。

Fig. 5の病変が悪性か良性かの質的診断についてみると、MRIを加えることによって「多分悪性」が大幅に減少し、「悪性」判定が有意に増加した。しかし、最終診断と照合した結果得られたROC曲線では、pooled data (Fig. 6 (a))のAzおよび読影医9人のAz (Fig. 6 (a, b))から検定した医師全体としての質的診断能は<sup>201</sup>Tl SPECT像単独時とMRIを追加した時とで有意の差を生じなかった。その原因はMRIを参照したことで最終診断：「悪性」を「悪性」と読影診断する率 (TPR)が増える一方で、最終診断：「良性」を「多分悪性」または「悪性」と読影診断する率 (FPR)も増えたためである。MRIの情報を医師が過大評価する傾向が示されたものと思われるが、本実験では「良性」群が6例と少なく、症例の偏りに依存した結果である可能性も否定できないため、この事

実を一般化するにはさらに今後の検討を要する。

今回の実験では<sup>201</sup>Tl SPECT像の読影が先行したが、日常診療では、MRI検査が先行し、<sup>201</sup>Tl SPECT検査はその後に行われることが多いようである。これは<sup>201</sup>Tl SPECT像がより質的診断に迫れる可能性があると言われていたり<sup>201</sup>Tl SPECTが現在、MRIほど医療施設に普及していない点も理由であろう。そこで、順序を逆にしてMRIを先に読影し、その後で<sup>201</sup>Tl SPECT像を追加しMRIを再読影した場合の診断の変化から、<sup>201</sup>Tl SPECT像の有効性を検討する必要もある。本実験終了後、MRI診断の専門家による、そのような読影実験も行ったが、結果については別に報告する。

存在診断と質的診断の医師個人ごとの読影結果とその医師間変動について次のようなことが言える。

Fig. 2およびFig. 5に示す通り、<sup>201</sup>Tl SPECT像単独に比し、MRIを加えることにより、存在診断および質的診断ともに、曖昧な判定を減少させ、確信度の高い判定を増加させた医師が有意に多くなった。しかし、その増加率は医師により0-2倍近く変動した。変動要因の1つとして、脳疾患MRIに対する核医学専門医の読影経験が関与したことが考えられる。

また、図には示さなかったが、<sup>201</sup>Tl SPECT像の所見検出数は、全体的にTI単独時よりMRIを参照したことで減少した。これは医師がMRIで所見が観察され、<sup>201</sup>Tl SPECT像でTIの集積が「弱」か、「中」以上であっても、MRIで所見が描出されない場合に、これを生理的または技術的アーチファクトないしは脳組織外の集積と判断し、異常所見と採らないことがあったためと思われる。以上の結果、確信度が最も高い(スコア5)所見の検出数(本実験ではすべての症例に少なくとも1つ以上の所見があったため、所見検出能に相当)の医師間変動は有意に減少した。

一方、ROC曲線として求めた悪性鑑別能は、MRIを参照(Fig. 6(c))しても、MRIの過大評価傾向のため全体的にAzは上がらず、またAzの

分散(医師間変動)も<sup>201</sup>Tl SPECT像単独時(Fig. 6(b))と統計的には変化しなかった。しかし、定性的には、2人の医師間で差が無視できない場合もあるように見える。そこで、任意の2者間でAzに差があるかどうか検定した結果、有意( $p < 0.05$ )であったのはFig. 6(b)で1組、Fig. 6(c)で3組のみであった。ここでも、統計的な差が出にくかった原因の1つとして、ROC曲線の横軸のデータが少なかったこと(「良性」6例)が考えられる。

異常所見の位置判定能に関して次のようなことが言える。

<sup>201</sup>Tl SPECT像単独で所見の解剖学的区域まで判定することは所見描出のされ方からして困難が予想されたので、本報では、MRIを参照した時に限り区域判定を行い、その結果を最終診断と比較した。Fig. 3に示すように最終診断と<sup>201</sup>Tl SPECT+MRI読影において、側頭葉と前頭葉の区域判定の頻度分布が逆転した。これは最終診断の前頭葉を読影では側頭葉に、最終診断の側頭葉を読影では前頭葉と判定したことによるとと思われる。側頭葉はやや読みすぎの傾向があり、前頭葉のものをより多く含んでいる可能性が大きい。実験に参加した医師の多くに両区域識別判定の基準に関して、なんらかの系統的なバイアスがかかったものと想像される。

任意の2者間で上記区域判定の頻度分布に差があるかどうか検定した結果、読影医9人から構成される36通りの2人の組み合わせのうち8通りで有意差が観察された。これらの知見を区域判定の基準統一のためどのように活用するかは今後の検討課題である。

Fig. 4に示すごとく、区域的中(TP)および非的中(FP)の医師間変動は非常に大きかった。その原因は、9人中1人がFPは少なかったがTPが他の医師の1/2以下であったこと、2人の医師のFPが他の7人の2倍以上あったためであり、このデータからも区域判定の基準の統一を図る必要性が示唆された。



## V. 結 論

(1) <sup>201</sup>Tl SPECT 像単独による #1 異常所見検出率は 84% (読影医 9 人の平均値) であり, #1 悪性鑑別能 (読影医 9 人の pooled data による ROC 曲線の第 1 動作点) は, TPR: 53%, FPR: 28% であった。

一方, <sup>201</sup>Tl SPECT+MRI では, #2 異常所見検出率は 94.4%, #2 悪性鑑別能は, TPR: 74%, FPR: 45% であった。

#1 対 #2 は統計的に有意差あり, ¥1 対 ¥2 の ROC 曲線下面積 Az は有意差なしであった。

(2) <sup>201</sup>Tl SPECT 像単独による異常所見検出率の医師間変動は <sup>201</sup>Tl SPECT+MRI より有意に大きかった。<sup>201</sup>Tl SPECT 像単独と <sup>201</sup>Tl SPECT+MRI の悪性鑑別能の医師間変動の相違は統計的に有意でなかった。

(3) 本実験では (1), (2) の結果より MRI は異常所見の検出に有効であったが, 悪性鑑別についての有効性は証明できなかった。症例の分布に偏りがあったことが原因の 1 つとして考えられた。

(4) <sup>201</sup>Tl SPECT+MRI による異常所見区域判定率 (TPR) は 30.3–68.4% の範囲に分布した。前頭葉と側頭葉および脳幹の区域判定基準について医師間で合意を図る必要が示された。

謝辞: 本研究で使用した画像データベースの作成に関し技術的援助をいただいた千葉県がんセンター主任診療放射線技師: 木下富士美氏に感謝する。

本研究の一部は文部省科学研究費一般 C (No. 06670933 および基盤研究 (C) (2) 08671048) の援助を受けた。

## 文 献

- 1) 松本 徹, 飯沼 武, 館野之男, 他: 肝シンチグラムの臨床的有効度の定量的評価. 核医学 **19**: 51–65, 1982
- 2) 宇野公一, 内山 暁, 松本 徹, 他: 肝シンチグラム読影における脾腫の判定. Radioisotopes **32**: 469–472, 1983
- 3) 秋山芳久, 油井信春, 松本 徹, 他: Prospective study による骨シンチグラフィの臨床的有効度評価——方法論を中心に. Radioisotopes **33**: 145–154, 1984
- 4) 松本 徹, 飯沼 武, 石川達雄, 他: 肝 SPECT イメージの SOL 検出能——SOL 検出モデルによる解析. Radioisotopes **34**: 414–420, 1985
- 5) 松本 徹, 飯沼 武, 石川達雄, 他: 肝 SPECT イメージの SOL 検出能——ROC 曲線の構造解析. Radioisotopes **34**: 486–492, 1985
- 6) 小山田日吉丸, 町田喜久雄, 松本 徹: 肝 SPECT の臨床的有効度に関する共同研究について. Radioisotopes **36**: 340–346, 1987
- 7) 森 豊, 川上憲司, 町田喜久雄, 他: 肝 SPECT の臨床的有効度に関する研究——とくに肝区域および慢性肝疾患の合併に関する検討. Radioisotopes **36**: 457–464, 1987
- 8) 町田喜久雄, 松本 徹, 本田憲業, 他: 頭部専用リング型断層装置の <sup>123</sup>I-IMP SPECT 像による脳血管障害の診断精度に関するグループ研究. 核医学 **28**: 1301–1312, 1991
- 9) Black KL, Hawkins RA, Kim KT, Becker DP, Lerner C, Marciano D: Use of thallium-201 SPECT to quantitate malignancy grade of gliomas. J Neurosurg **71**: 342–346, 1989
- 10) Kim KT, Black KL, Marciano D, et al: Thallium-201 SPECT imaging of brain tumors: Method and results. J Nucl Med **31**: 965–969, 1990
- 11) Kosuda S, Fujii H, Aoki S, et al: Reassessment of quantitative thallium-201 brain SPECT for miscellaneous brain tumors. Ann Nucl Med **7**: 257–263, 1991
- 12) 織内 昇, 田村 勝, 柴崎 尚, 他: Glioma による <sup>201</sup>Tl 塩化タリウム SPECT の有用性. 核医学 **28**: 1263–1271, 1991
- 13) Carvalho PA, Schwartz RB, Alexander III A, Garada BM, Zimmerman RB, Loeffler JS, et al: Detection of recurrent gliomas with quantitative thallium-201/Tc-99m HMPAO SPECT. J Neurosurg **77**: 565–570, 1992
- 14) 戸川貴史, 油井信春, 木下富士美, 柳沢正道, 難波宏樹: 肺癌脳転移における <sup>201</sup>Tl 脳 SPECT の検討. 核医学 **32**: 217–225, 1995
- 15) Hanley JA, McNeil BJ: A method of comparing the area under receiver operating characteristic curves derived from the same cases. Radiology **148**: 839–843, 1983
- 16) 丹後俊朗: 臨床検査の統計学, 朝倉書店, 東京, 1986
- 17) 飯沼 武: 臨床的有効度の評価基準. 臨床放射線 **28**: 533–536, 1981
- 18) Tonami N, Hisada K: Tl-201 scintigraphy in post-operative detection of thyroid cancer: A comparative study with I-131. Radiology **136**: 461–464, 1980
- 19) Tonami N, Shuke N, Yokoyama K, et al: Tl-201 SPECT in the evaluation of suspected lung cancer. J Nucl Med **30**: 997–1004, 1985

## Summary

### Diagnostic Accuracy of $^{201}\text{Tl}$ SPECT · MRI in Brain Diseases and Inter-reader Variance of Diagnostic Performance

Kikuo MACHIDA\*<sup>1</sup>, Norinari HONDA\*<sup>1</sup>, Toru MATSUMOTO\*<sup>2</sup>, Nobuharu YUI\*<sup>3</sup>, Takashi TOGAWA\*<sup>3</sup>, Atsushi KUBO\*<sup>4</sup>, Yasushi TSUKATANI\*<sup>4</sup>, Hiyoshimaru OYAMADA\*<sup>4</sup>, Katsumi ISHII\*<sup>5</sup>, Hiroshi NISHIMAKI\*<sup>5</sup>, Yutaka MORI\*<sup>6</sup>, Kimiichi UNO\*<sup>7</sup>, Hajime MURATA\*<sup>8</sup> and Hirotaka MARUNO\*<sup>8</sup>

\*<sup>1</sup>Saitama Medical Center, Saitama Medical School

\*<sup>2</sup>National Institute of Radiological Sciences

\*<sup>3</sup>Chiba Cancer Center Hospital

\*<sup>4</sup>School of Medicine, Keio University

\*<sup>5</sup>School of Medicine, Kitazato University

\*<sup>6</sup>Jikei University School of Medicine

\*<sup>7</sup>School of Medicine, Chiba University

\*<sup>8</sup>Toranomon Hospital

Fifteen-four studies of  $^{201}\text{Tl}$  brain tumor SPECT were independently interpreted by 9 nuclear medicine physicians with and without reference magnetic resonance images in 2 separate sessions to define an effect of referring images, and inter-observer variations. The physicians were requested to detect foci of abnormal deposits, and to discriminate whether they were malignant or not according to 5-grade scaling of subjective diagnostic confidence. Receiver-operating characteristics (ROC) analysis was performed. Mean sensitivity for presence of lesions (SFP), and sensitivity and specificity for malignancy of  $^{201}\text{Tl}$  SPECT were 84, and 53 and 55%, which were changed to 94, and 74 and 55% after referring to the MR images. The SFP was significantly improved ( $p < 0.05$ ), but sensitivity

and specificity for malignancy, assessed by areas under the ROC curves, were not significantly improved. Inter-observer variation of the SFP was significantly reduced with addition of the referring MR images. Cerebral lobar localization of lesions by SPECT showed great inter-observer variations (true localization ranged from 30 to 68.4%). It is concluded that  $^{201}\text{Tl}$  brain tumor SPECT has moderate sensitivity and specificity for malignancy, which is not improved by addition of anatomical reference images, that additional MR images reduce inter-observer variation of confidence on lesion presence, and that SPECT localization of lesions has great inter-observer variations.

**Key words:**  $^{201}\text{Tl}$  SPECT, MRI, Diagnostic efficacy, Interobserver variation, ROC analysis.