# Heterospecific and conspecific social cognition in the anterior cingulate cortex

Jun Shinozaki[a], Takashi Hanakawa[a,b] and Hidenao Fukuyama[a]

[a]Human Brain Research Center, Kyoto University Graduate School of Medicine, Kyoto and [b]Department of Cortical Function Disorders, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan

Correspondence to Dr Takashi Hanakawa, MD, PhD, Section Chief, Department of Cortical Function Disorders, National Institute of Neuroscience, National Center of Neurology and Psychiatry, 4-1-1 Ogawahigashi, Kodaira, Tokyo 187-8502, Japan
Tel: +81 42 341 2711 (ext) 5173; fax: +81 42 346 1748; e-mail: hanakawa@ncnp.go.jp

Received 3 March 2007; accepted 21 March 2007

The development of human social cognition has allowed interactions with other species and the formation of a cooperative multispecies society. This feature posited keen attention on the following question: is heterospecific social cognition represented in the same brain areas as conspecific social cognition? Here we investigated brain activity accompanying the facial recognition of familiar humans and of companion dogs, both of whom had real social interactions with participants. The rostroventral anterior cingulate cortex responded to both species whereas the caudal anterior cingulate cortex was sensitive only to familiar humans. Social cognition processes may be dual-layered: the rostroventral anterior cingulate cortex is associated with fundamental and intuitive aspects, whereas the caudal anterior cingulate cortex is concerned with the analysis of complex social interactions. *NeuroReport* 18:993–997 © 2007 Lippincott Williams & Wilkins.

**Keywords**: anterior cingulate cortex, conspecific, facial recognition, functional MRI, heterospecific, social cognition

## Introduction

Animals that form societies, particularly humans, have developed the ability to deal with complex interindividual relationships through social cognition [1]. Although these interactions typically occur between conspecifics, they can also extend to heterospecifics [2]. One example is the social interaction that occurs between humans and companion animals, which are often treated like family members. Many humans receive emotional support from companion species [3] and have sincere concern for their welfare [4]. Interactions with animals might also help children to develop empathetic ability [5]. Thus, it seems to be important to know the neural mechanisms underlying heterospecific social interactions or the roles of cognition and emotion in this process.

Recent imaging studies have begun to explore the neural underpinnings of conspecific social cognition using various tasks, including face recognition [6]. Faces convey critical information for social cognition, as the recognition of individuals and their emotional states constitute the fundamentals of interpersonal relationships [1]. In particular, recognizing familiar faces, such as those of family members, activates the amygdala [7] and the medial prefrontal areas, including the anterior cingulate cortex [8]. Parts of the anterior cingulate cortex are rich in spindle-shaped output neurons that are present in great apes [9]. In addition, abundant connections with the limbic and prefrontal areas render the anterior cingulate cortex adequate for representing society [10], including the family, which is one of its most critical constituents. The anterior cingulate cortex can be divided into the affective rostroventral anterior cingulate cortex and the cognitive caudal anterior cingulate cortex [11]. Several types of activity were expected to occur in the anterior cingulate cortex during the recognition of heterospecific and conspecific family members. We hypothesized that faces of human and canine 'family' members would evoke similar responses in the rostroventral anterior cingulate cortex, which might support the processing of social information closely related to emotion. In addition, we predicted that the recognition of familiar human faces would evoke complex cognitive reflection, which might involve the analytical caudal anterior cingulate cortex regions [11–14].

## Materials and methods

### Participants

Seventeen healthy volunteers (11 men and six women) participated in this study. All participants had one or two companion dogs and none reported critical family problems. Four additional volunteers (one woman) participated in a control experiment for recognition of familiar objects. All individuals gave written informed consent to participate in the study and the protocol was approved by the local ethics committee.

### Stimuli

The stimuli comprised grayscale digital photographs of faces and mosaic patterns. Luminance and contrast were roughly adjusted across the stimuli. In total, six types of face stimulus were prepared specifically for each participant: family, newly learned and unfamiliar human faces (FH, NH

and UH, respectively), and family, newly learned and unfamiliar dog faces (FD, ND and UD, respectively). The human facial expressions ranged from neutral to a faint smile. Two types of mosaic pattern were created: newly learned mosaics (NM) and unfamiliar mosaics (UM). The numbers of face and mosaic images in each category varied from 14 to 20, according to availability. Each stimulus subtended visual angles of ~10° vertically and horizontally. All facial stimuli were superimposed onto a white background of 640 × 480 pixel size, after removing the original backgrounds of the photographs. Similarly, pictures of personally familiar and newly learned objects were prepared for the object-recognition experiment. The objects were personal belongings such as cell phones. The stimuli were presented for 300 ms in a semirandom order and the interstimulus intervals were varied between 5 and 10 s.

**Experimental procedures**

Participants were asked to memorize a predetermined set of stimuli, consisting of NH, ND and NM. To assess the complete memorization of NH, ND and NM, participants saw pictures of FH, FD, NH, ND, NM and UM (18 pictures in total) and judged whether they knew them or not before the imaging experiment. During scanning, the stimulus presentation was controlled by a personal computer synchronized with the MRI scanner. The onset of the stimulus presentation was semirandomly staggered with the acquisition onset of each functional image. In the MRI scanner, the participants judged whether they had previously seen the presented images and reported their answer by pressing one of the two buttons. The accuracy and response times were recorded. The participants maintained visual fixation on a crosshair at baseline. In total, 439 volumes were acquired in each imaging run (18 min 17.5 s). Two imaging runs were administered for each participant. The same stimulus was not presented twice within a single run and stimuli from unfamiliar categories were not repeated in the second run.

After the functional MRI (fMRI), the participants reported on how much support they received everyday from their human family members and companion dogs, using the Scale of Expectancy for Social Support (SESS) [15] and the Animal Companionship and Support Scale (ACSS) [16], respectively. Nine participants also rated subjective valence and arousal to each facial stimulus, using scales ranging from 1 (most negative or least arousal) to 9 (most positive or most arousal).

**Image acquisition**

FMRI experiments were conducted on a 3-T scanner equipped with a volume headcoil (Siemens Trio, Erlangen, Germany). Functional images were obtained in a T2*-weighted gradient-echo echo-planar imaging sequence with prospective motion correction. The image-acquisition parameters were as follows: repetition time (TR) = 2.5 s; echo time (TE) = 30 ms; flip angle (FA) = 90°; field of view (FOV) = 192 mm; matrix = 64 × 64; 40 interleaved axial slices with 3-mm thickness without gaps (3-mm cubic voxels). The first two volumes were not saved to allow for signal stabilization.

For anatomic images, T1-weighted three-dimensional structural images were also obtained using a magnetization-prepared rapid-gradient echo sequence. Each participant lay supine on a scanner bed, with a button-response device held in the right hand. The participants viewed visual stimuli that were back-projected onto a screen through a built-in mirror. Foam pads were used to minimize head motion.

**Image analysis**

The fMRI data were analyzed with SPM2 (Wellcome Department of Imaging Neuroscience, University College London, London, UK) implemented on MATLAB 6.5 (MathWorks, Natick, Massachusetts, USA), using the principles of the general linear model [17]. The functional images were corrected for differences in slice-acquisition timing and were then spatially realigned to the first image of the initial run to adjust for residual head movements after prospective motion correction. The realigned images were spatially normalized to fit to a Montreal Neurological Institute template [18] based on the standard stereotaxic coordinate system [19]. Subsequently, all images were smoothed with an isotropic Gaussian kernel of 12-mm full-width at half-maximum. Each of the eight stimulus conditions was separately modeled as a regressor for the first-level multiregression analysis. This analysis was performed for each participant, to test the correlation between the MRI signals and a train of delta functions (representing event onsets) convolved with the canonical hemodynamic response function and its temporal derivative. Global signal normalization was performed only between runs. Low-frequency noise was removed using a high-pass filter with a cutoff of 128 s and serial correlation was adjusted using an AR(1) model. By applying the appropriate linear contrast to the parameter estimates, mean-effect images reflecting the magnitude of correlation between the signals and the model of interest were computed. These were used for the subsequent second-level random-effect model analysis. Group-level statistical parametric maps were produced using the one-sample *t*-test.

For the correlation analysis, the raw SESS and ACSS scores were used as regressors at the second level. The statistical parametric map (SPM){*t*} was first thresholded at $P<0.001$ and the significance of the activity was defined by the cluster-level $P$ value corrected for multiple comparisons for the entire brain ($P<0.05$). For the caudal anterior cingulate cortex, a small volume correction analysis was applied (false discovery rate corrected, $P<0.05$) for both the interaction and correlation analyses. The center coordinates ($x=-9$, $y=6$, $z=42$) of a spherical region-of-interest with a 7.5 mm radius was determined on the basis of a previous report [13] showing the relationship of the caudal anterior cingulate cortex with higher-order interindividual interactions. The resulting activation maps were displayed on the anatomically normalized mean T1 image derived from all participants, to identify the anatomical correlates of the activity. The results are reported on the basis of the coordinates of the Montreal Neurological Institute template.

## Results

**Behavioral results**

The accuracy was on average over 90% for all stimuli. A Wilcoxon signed-rank test was used to analyze the hit rates. No significant differences were found in the hit rates during fMRI between FH and NH ($P=0.480$), or between FD and ND ($P=0.686$). The hit rates demonstrated similar levels of robustness in the familiarity of recognizable faces between the human and dog conditions. The averaged response

times were 762, 799, 774 and 871 ms for FH, NH, FD and ND, respectively. The response times were analyzed with two-way repeated-measure analysis of variance (ANOVA) followed by Bonferroni's posthoc two-tailed *t*-test. Significant effects of both familiarity and species were noted on response times ($P<0.001$ and $=0.003$, respectively). Specifically, the response to FH and FD tasks was significantly faster than that to NH and ND tasks ($P=0.034$ and 0.001, respectively).

**Species-invariant and human-selective activities**

To explore specific family-related activity, the familial and newly learned faces were contrasted separately for the conspecific (FH minus NH) and heterospecific (FD minus ND) conditions. Conjunction analyses revealed species-invariant activation during familial face recognition only in the rostroventral anterior cingulate cortex (coordinates: $x=-12$, $y=32$, $z=-12$; Z-value = 4.34) at the whole-brain corrected threshold ($P<0.05$; Fig. 1a). Interaction analyses were then conducted to identify familial human-selective activity (conspecific minus heterospecific) or dog-selective activity (heterospecific minus conspecific). The former was identified in the caudal anterior cingulate cortex (coordinates: $x=-8$, $y=12$, $z=44$; Z-value = 3.13) (Fig. 1b) after a small volume correction ($P<0.05$), whereas the latter did not reach statistical significance.

**Psychological scales and anterior cingulate cortex activities**

The imaging data (e.g. FH relative to baseline) were reanalyzed to identify any brain activity that covaried with these scores. Participants with higher perceived ACSS scores showed increased activity in the rostroventral anterior cingulate cortex, right anterior insula and pulvinar during FD recognition (Fig. 2b). The ACSS-correlated rostroventral anterior cingulate cortex region overlapped with that showing species-invariant rostroventral anterior cingulate cortex activity during familial face recognition. The SESS scores were positively correlated with caudal anterior cingulate cortex activity during FH recognition (Fig. 2c). This relationship was statistically significant after a small volume correction ($P<0.05$). This further supported the role of the caudal anterior cingulate cortex in processing only human interpersonal relationships. A confirmation analysis upheld the specificity of these correlations; there was no correlation between the caudal anterior cingulate cortex activity during family dog recognition and the ACSS score ($r=0.387$, $P=0.124$), or between the rostroventral anterior cingulate cortex activity during family human recognition and the SESS score ($r=0.143$, $P=0.583$).

The observed medial frontal activities could not be attributed to task difficulty, as the FH and FD tasks were relatively easier than the NH and ND tasks according to the response-time analysis. To determine whether valence or arousal differences could account for our findings, nine participants rated these parameters for each stimulus after the fMRI. The results showed that familiarity and species had no significant effects on the arousal rating ($P=0.095$ and 0.088, respectively). Significant effects on the valence rating were, however, observed for familiarity ($P=0.004$) and species ($P=0.039$). Valence was rated more positively for the family face condition than the other conditions ($P=0.014$ versus newly learned, $P=0.008$ versus unfamiliar). Notably, the overall effect of species was primarily due to the significantly higher valence ratings for UD than for UH ($P=0.005$). Brain activity was thus compared between UD and UH in the rostroventral and caudal anterior cingulate areas to examine the general effects of valence. Despite the significant valence disparity, the brain activity in the UD and UH conditions did not differ in the rostroventral anterior cingulate cortex ($P=0.351$) or the caudal anterior cingulate cortex ($P=0.130$). Thus, the arousal or simple valence component alone appeared unable to explain the observed rostroventral and caudal anterior cingulate cortex activities.

Finally, to ensure that the anterior cingulate activities did not merely reflect different degrees of familiarity, brain activity was compared between the newly learned and unknown faces. Simple familiarity-related activation was not detected in the medial frontal area according to this

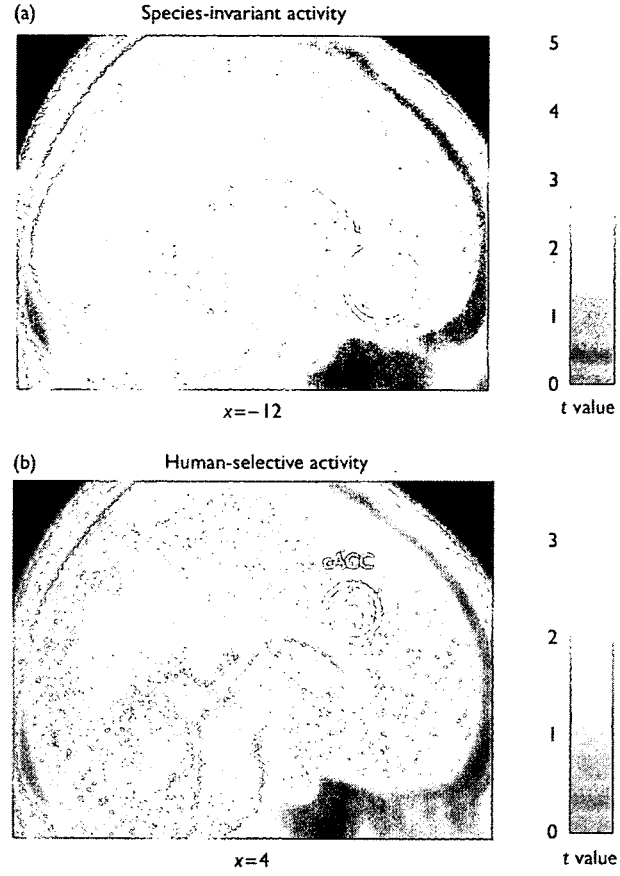


**Fig. 1** Species-invariant and human-selective activities in the anterior cingulate cortex. (a) Species-invariant activity coequally associated with the recognition of familial human and canine faces was detected in the rostroventral anterior cingulate cortex (rvACC; coordinates, $x=-12$, $y=32$, $z=-12$; Z-value = 4.34) at a cluster-level corrected threshold of $P<0.05$. (b) Human-selective activity during the recognition of familial faces was found in the caudal anterior cingulate cortex (cACC; coordinates, $x=-8$, $y=12$, $z=44$; Z-value = 3.13) at a false discovery rate corrected threshold of $P<0.05$ after a small volume correction. When a uniform threshold of uncorrected $P<0.001$ was applied, the numbers of activated voxels were 244 and 51 in the rostroventral anterior cingulate cortex and the caudal anterior cingulate cortex, respectively. Activity is shown at a height threshold of $P<0.001$ (uncorrected) with an extent threshold of 50 voxels.

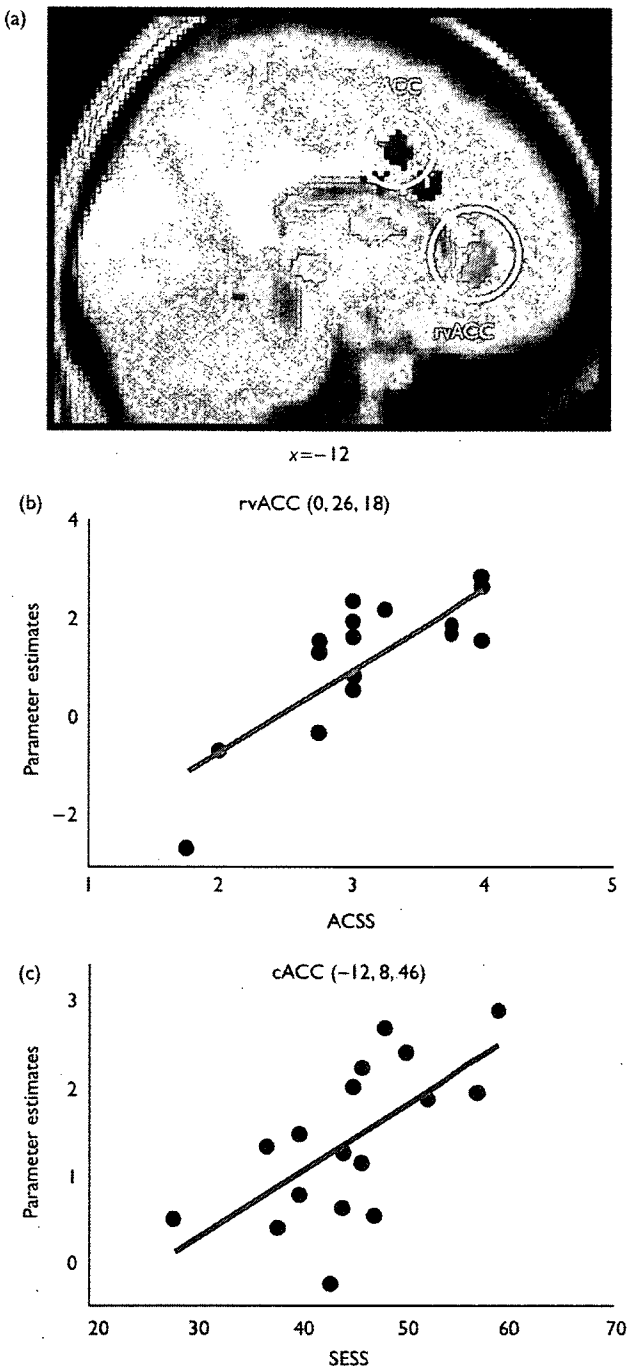


**Fig. 2** Correlation between anterior cingulate cortex activities and support from family members or companion dogs. (a) Brain activity correlated with the Scale of Expectancy for Social Support (SESS) scores for human family members is shown in green. This activity was significant during human familial face recognition in the caudal anterior cingulate cortex (cACC) after a small volume correction ($P<0.05$). Brain activity correlated with the Animal Companionship and Support Scale (ACSS) scores is shown in red. This activity was significant in the rostroventral anterior cingulate cortex (rvACC) at a cluster-level corrected threshold for the whole brain ($P<0.05$). The SESS-correlated activity was thresholded at uncorrected $P<0.025$, and the ACSS-correlated activity was thresholded at uncorrected $P<0.005$ for display purposes. (b) The rostroventral anterior cingulate cortex activity during family dog recognition was significantly correlated with the ACSS score ($r=0.805$, $P<0.001$). (c) The caudal anterior cingulate cortex activity during familial human recognition was positively correlated with the SESS score ($r=0.639$, $P=0.006$).

confirmatory analysis ($P<0.01$, uncorrected). To assess the possible effects of nonsocial, long-term familiarity, brain activity was compared between personally familiar and newly learned objects in four participants. Long-term familiarity to nonliving objects did not significantly activate the medial frontal area in a fixed-effect model group analysis ($P<0.01$, uncorrected). These additional data further supported the hypothesis that the rostroventral and caudal anterior cingulate activities more likely reflected long-term social familiarity than different degrees of familiarity or long-term nonsocial familiarity.

## Discussion

At least two distinct activities were identified in the anterior cingulate cortex during the recognition of human and canine faces associated with long-term social familiarity: species-invariant activity corresponding to the affective rostroventral anterior cingulate cortex and species-selective activity corresponding to the cognitive caudal anterior cingulate cortex. Recently, the medial prefrontal cortex was shown to be active during the analysis of the psychological state of both dogs and humans [20]. Anatomically, the species-invariant rostroventral anterior cingulate cortex activity occurred in the most ventral parts of the medial prefrontal cortex and also partly in the orbitofrontal cortex, both of which have been implicated in social cognition [6]. For instance, previous imaging studies reported enhanced rostroventral anterior cingulate cortex activity when mothers saw their babies' faces [8]. Damage to the medial prefrontal cortex impairs decision-making ability [21,22]. This condition is related to a failure to cooperate with family members and society [22] and is accompanied by emotional flattening [21]. These interactions between emotion and social cognition might be subserved by the reciprocal connections between the amygdala and rostroventral anterior cingulate cortex, which could be engaged in the emotional and autonomic analysis of internal states [23]. As we used an implicit social cognition task in this study, we might be able to approach the intuitive aspects of social cognition. In addition, rostroventral anterior cingulate cortex activity was associated with emotional support from companion dogs, but not with social support from human family members. This suggests that the social cognition represented in the rostroventral anterior cingulate cortex is basic, intuitive and emotional and might underpin long-term interindividual cross-species relationships.

A previous study demonstrated greater caudal anterior cingulate cortex responses when Caucasian participants saw the faces of individuals from a different ethnic group than those from the same group [24]. This activity was significant only when the stimuli were presented for a relatively long duration, suggesting that cognitively controlled analysis was involved. These results suggest that the caudal anterior cingulate cortex is associated with the analysis of facial information according to social norms. In addition, Singer and colleagues [13] reported that caudal anterior cingulate cortex activity persisted in individuals who were subjected to painful stimuli or saw others in pain. Activation of the caudal anterior cingulate cortex is also induced by action-related judgment about humans relative to the one about dogs [14]. Taken together, these findings imply that the caudal anterior cingulate cortex is related to conspecific

cognitive analysis, which may realize complex social interactions and empathy in humans.

## Conclusion

The current study has clarified the representations of species-general and species-specific social cognition. The rostroventral anterior cingulate cortex and caudal anterior cingulate cortex appear to represent different levels of social cognition. Consistent with this, Somerville and colleagues recently showed that the rostroventral anterior cingulate cortex was associated with emotional evaluation, whereas the caudal anterior cingulate cortex was related to cognitive conflicts in social acceptance/rejection judgment based on face recognition [12]. Thus, the rostroventral anterior cingulate cortex is probably associated with fundamental aspects of social cognition, which are closely related to affection. This function appears to apply to interspecies interactions without involving complex cognitive analysis. By contrast, the caudal anterior cingulate cortex might be related to higher-order social cognition, involving analyses of profit-and-loss and social norms. Thus, balancing the judgment of cognitive analysis and emotional intuition seems crucial for social interactions.

## Acknowledgements

The authors thank K. Taneich and S. Tanaka for their helpful comments on the psychological testing, T. Kochiyama for support with the fMRI data analysis and D.H. Duy Thuy, N. Sawamoto and T. Murai for their critical comments on the paper. This study was partly supported by a Grant from the Companion Animal Information and Research Center (CAIRC) and also at a writing stage by a Grant-in-Aid on Fundamental Research (C) (17500210) and on Priority Areas (Mobilligence Project 17022023) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, to T.H., by a Grant-in-Aid for Scientific Research on Priority Areas System study on higher-order brain functions (18020014) from the Japan Society for the Promotion of Science, and by a grant from New Energy and Industrial Technology Development, Japan (51101244-0) to H.F.

## References

1. Adolphs R. Cognitive neuroscience of human social behaviour. *Nat Rev Neurosci* 2003; 4:165–178.
2. Adolphs R. The neurobiology of social cognition. *Curr Opin Neurobiol* 2001; 11:231–239.
3. Gunter R. *Pets & people: the psychology of pet ownership*. London: Whurr Pub Ltd; 1999.
4. Hills AM. The motivational bases of attitudes toward animals. *Soc Anim J Hum-Anim Stud* 1993; 1:111–128.
5. Robinson I. *The Waltham book of human animal interaction: benefits and responsibilities of pet ownership*. New York: Pergamon; 1995.
6. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* 2006; 7:268–277.
7. Leibenluft E, Gobbini MI, Harrison T, Haxby JV. Mothers' neural activation in response to pictures of their children and other children. *Biol Psychiatry* 2004; 56:225–232.
8. Bartels A, Zeki S. The neural correlates of maternal and romantic love. *Neuroimage* 2004; 21:1155–1166.
9. Nimchinsky EA, Vogt BA, Morrison JH, Hof PR. Spindle neurons of the human anterior cingulate cortex. *J Comp Neurol* 1995; 355:27–37.
10. Allman JM, Watson KK, Tetreault NA, Hakeem AY. Intuition and autism: a possible role for von economo neurons. *Trends Cogn Sci* 2005; 9:367–373.
11. Bush G, Luu P, Posner MI. Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci* 2000; 4:215–222.
12. Somerville LH, Heatherton TF, Kelley WM. Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nat Neurosci* 2006; 9:1007–1008.
13. Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD. Empathy for pain involves the affective but not sensory components of pain. *Science* 2004; 303:1157–1162.
14. Mason MF, Banfield JF, Macrae CN. Thinking about actions: the neural substrates of person knowledge. *Cereb Cortex* 2004; 14:209–214.
15. Hisada M, Senda S, Miguchi M. The approaches to create a scale for students: the scale of expectancy for social support. The 30th annual meeting for the Japanese society of social psychology [in Japanese]. Tokyo: The Japanese Society of Social Psychology; 1989. pp. 143–144.
16. Taneichi K. Animal Companionship and support scale development and comparisons of scores in pet owners. The annual meeting for Japanese animal hospital association [in Japanese]. Tokyo: Japanese Animal Hospital Association; 2001. pp. 19–20.
17. Friston KJ. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapping* 1995; 2:189–210.
18. Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM. 3D statistical neuroanatomical models from 305 MRI volumes. IEEE-nuclear science symposium and medical imaging conference. San Francisco, California, USA: IEEE Services Center; 1993. pp. 1813–1817.
19. Talairach J, Tournoux P. *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers; 1988.
20. Mitchell JP, Banaji MR, Macrae CN. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage* 2005; 28:757–762.
21. Damasio AR. *Descartes' error: emotion, reason, and the human brain*. New York: Avon Books; 1995.
22. Anderson SW, Bechara A, Damasio H, Tranel D, Damasio AR. Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat Neurosci* 1999; 2:1032–1037.
23. Devinsky O, Morrell MJ, Vogt BA. Contributions of anterior cingulate cortex to behaviour. *Brain* 1995; 118 (Pt 1):279–306.
24. Cunningham WA, Johnson MK, Raye CL, Chris Gatenby J, Gore JC, Banaji MR. Separable neural components in the processing of black and white faces. *Psychol Sci* 2004; 15:806–813.

NMR IN BIOMEDICINE
*NMR Biomed.* 2007; **20**: 743–748
Published online 8 March 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI:10.1002/nbm.1139



# Fractional anisotropy and mean diffusivity: comparison between 3.0-T and 1.5-T diffusion tensor imaging with parallel imaging using histogram and region of interest analysis

**Yasutaka Fushimi,[1†] Yukio Miki,[1*,†] Tsutomu Okada,[1] Akira Yamamoto,[1] Nobuyuki Mori,[1] Takashi Hanakawa,[2,3] Shin-ichi Urayama,[2] Toshihiko Aso,[2] Hidenao Fukuyama,[2] Ken-ichiro Kikuta[4] and Kaori Togashi[1]**

[1]Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan
[2]Human Brain Research Center, Kyoto University Graduate School of Medicine, Kyoto, 606-8507, Japan
[3]Department of Cortical Function Disorders, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan
[4]Department of Neurosurgery, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan

*Received 15 August 2006; Revised 28 November 2006; Accepted 29 November 2006*

ABSTRACT: We performed a comparison study focusing on differences in fractional anisotropy (FA) and mean diffusivity (MD) between 3-T and 1.5-T diffusion tensor imaging (DTI) with parallel imaging. Thirty healthy volunteers underwent DTI with an eight-channel phased-array coil at both 3 T and 1.5 T. Histogram and region of interest (ROI) analyses were performed. Paired *t* tests were applied for statistical analysis. Signal-to-noise ratios of these regions were also measured. For histogram analysis, peak location of FA was significantly lower at 3 T than at 1.5 T ($P = 0.04$). Mean FA was significantly higher at 3 T than at 1.5 T ($P = 0.002$). Peak location of MD was significantly lower at 3 T than at 1.5 T ($P < 0.001$). Mean MD was significantly lower at 3 T than at 1.5 T ($P < 0.001$). In ROI analysis, FA was significantly larger at 3 T than at 1.5 T in the centrum semiovale ($P < 0.001$), middle cerebellar peduncle ($P < 0.001$), cerebral peduncle ($P = 0.006$), posterior limb of the internal capsule ($P = 0.007$), genu ($P < 0.001$) and splenium ($P < 0.001$). FA was significantly lower at 3 T than at 1.5 T in the globus pallidus ($P < 0.001$). MD was significantly smaller at 3 T than at 1.5 T in the globus pallidus ($P = 0.007$), thalamus ($P < 0.001$), centrum semiovale ($P < 0.001$), middle cerebellar peduncle ($P < 0.001$), cerebral peduncle ($P = 0.01$), posterior limb of the internal capsule ($P < 0.001$), genu ($P = 0.01$) and splenium ($P < 0.001$). Significant differences in FA and MD exist between 3 T and 1.5 T for whole-brain histogram analysis and ROI analysis. Copyright © 2007 John Wiley & Sons, Ltd.

KEYWORDS: diffusion tensor imaging; fractional anisotropy; mean diffusivity; histogram analysis; high-field imaging; parallel imaging

## INTRODUCTION

Diffusion tensor imaging (DTI) is an MRI technique sensitive to the orientation of mobility in intravoxel water molecules (1,2), and it has been widely used for studying the dependency of water proton diffusion in the brain. DTI has proven indispensable for evaluating brain structures, particularly white matter. White matter shows high anisotropy, because water diffusion is faster in the direction of fibers (3,4). *In vivo*, diffusion anisotropy in a voxel is determined by macrostructural features of the tissue, such as intravoxel fiber–tract coherence, and microstructural features of fiber density and fiber diameter (4). Degree of myelination also affects diffusion anisotropy (5).

In DTI, fractional anisotropy (FA) and mean diffusivity (MD) are two of the most commonly used scalars (6). FA decreases in isotropic water diffusion when water molecules move in all directions, and increases in anisotropic water diffusion when movement of water molecules is restricted to a specific direction. MD is equal to one-third of the diffusion tensor trace and represents a parameter of average molecular motion. FA and MD can be applied to several pathological situations in which microstructural properties of the brain are altered, such as multiple sclerosis (7–9), Alzheimer's disease (10,11) and amyotrophic lateral sclerosis (12,13). Histogram analyses offer a useful approach to whole-brain DTI analysis, as histograms can deal with whole-brain images and can be

**Correspondence to:* Y. Miki, Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, 54 Shogoin-Kawaharacho, Sakyo-ku, Kyoto, 606-8507, Japan.
E-mail: mikiy@kuhp.kyoto-u.ac.jp
†These authors contributed equally to the study.
*Contract/grant sponsor:* Health and Labour Sciences Research of Japan; *contract/grant number:* H15-003.

**Abbreviations used:** DTI, diffusion tensor imaging; FA, fractional anisotropy; MD, mean diffusivity; ROI, region of interest; SNR, signal-to-noise ratio.

compared among many subjects with low observer variability; in contrast, region of interest (ROI) analyses are observer dependent (14–16).

High-field MRI has become popular in clinical situations and research. Numerous reports have featured the differences in MRI at 3 T and 1.5 T, but few investigations have focused on differences in DTI (17,18). Parallel imaging using multi-channel head coils has now been introduced, and its advantages for DTI at higher fields have been reported (19). However, to the best of our knowledge, no studies have yet focused on differences in DTI with parallel imaging between 3 T and 1.5 T.

The purpose of this study was to elucidate differences in FA and MD for whole-brain images derived from DTI between 3 T and 1.5 T, using parallel imaging techniques.

## MATERIALS AND METHODS

### Subjects

The subjects were 30 healthy volunteers (15 men, 15 women; mean age, 28 years; range 21–46 years) with no history of neurological injury or psychiatric disease. Either a neurologist or a neurosurgeon examined each subject. None displayed any abnormal neurological signs or symptoms. Institutional review board approval was obtained for all study protocols, and all subjects provided written informed consent.

### Data acquisition

All subjects underwent both 3 T and 1.5 T DTI consecutively in random order on the same day, with a whole-body 3-T MR scanner (Magnetom Trio; Siemens, Erlangen, Germany) and a 1.5 T MR scanner (Magnetom Symphony; Siemens). All underwent both 3 T and 1.5 T MRI with an interval of less than 1 h. DTI was performed using an integrated parallel acquisition technique and a receiver-only eight-channel phased-array head coil for both MR units. DTI sequences for both 3 T and 1.5 T used the same single-shot spin echo echo planar sequences. The generalized autocalibrating partially parallel acquisitions algorithm was applied with a reduction factor of 2 to shorten the *TE*: *TR*, 5200 ms; *TE*, 79 ms; field of view, 220 mm; matrix, 128 × 128; 3 mm thickness without interslice gap (matrix size, 1.7 mm × 1.7 mm × 3 mm); and four times repetition. Motion-probing gradients were applied along 12 different non-colinear directions with a $b$ factor of 700 s/mm$^2$. A total of 40 slices covered the entire brain including brainstem and cerebellum. Scan time for both 3 T and 1.5 T DTI was 7 minutes 40 seconds.

### DTI post-processing

MR data from DTI were transferred to a Windows PC workstation from the MR scanners. Tensor calculations were made using DtiStudio version 2.02 software (H. Jiang, S. Mori; Department of Radiology, Johns Hopkins University, Baltimore, MD, USA) (20,21). Binary brain mask images were made for diffusion-unweighted ($b = 0$) images by using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK) and Matlab version 6.5 (The MathWorks, Natick, MA, USA). As FA and MD images were created from both diffusion-weighted and diffusion-unweighted images, they shared the same matrices. Skull and cerebrospinal fluid space were removed from data using the brain mask, and FA and MD images of the brain were created.

### Histogram analysis

Histograms were created for each FA and MD image with a binwidth of 1% of maximum using ImageJ (National Institutes of Health, Bethesda, MD, USA), and were divided by each entire brain voxel for normalization. Peak locations, peak heights, and mean FA and MD were derived and evaluated.

### ROI analysis

ROI-based analysis was also performed. The cerebral peduncle and middle cerebellar peduncle were selected as infratentorial structures, and the thalamus, globus pallidus, centrum semiovale, posterior limb of the internal capsule, and genu and splenium of the corpus callosum as supratentorial structures (Fig. 1). ROIs were placed manually on diffusion-unweighted images. Signal-to-noise ratios (SNRs) of the ROIs (middle cerebellar peduncle, globus pallidus, centrum semiovale, splenium) was calculated from the means of the ROI and standard

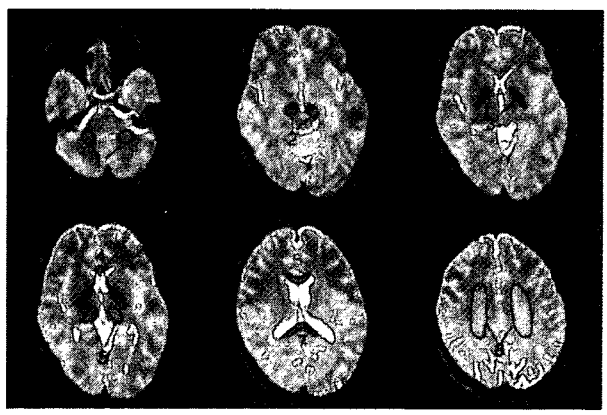

**Figure 1.** Diffusion-unweighted imaging ($b = 0$). The following ROIs were manually selected: 1, middle cerebellar peduncle; 2, cerebral peduncle; 3, globus pallidus; 4, posterior limb of the internal capsule; 5, thalamus; 6, genu and 7, splenium of the corpus callosum; 8, centrum semiovale

deviation of the background according to previous reports (17,22). These ROIs were selected for the following reasons: middle cerebellar peduncle for white matter in the posterior fossa; globus pallidus for magnetic susceptibility induced by ferritin; and centrum semiovale and splenium for white matter in the supratentorial region. SNR index was determined as SNR at 3 T divided by SNR at 1.5 T.

### Statistical analysis

JMP 5.1 (SAS Institute Inc., Cary, NC, USA) was used for statistical analysis. A paired $t$ test was performed on the following histogram data at both 3 T and 1.5 T: peak location; peak height; mean FA; mean MD; normalized voxels for each bin of FA and MD. For the MD histogram, any bin containing <0.5% of normalized voxels was considered as noise and excluded from statistical analysis. A paired $t$ test was also applied for each value of ROI for FA and MD at both 3 T and 1.5 T. $P < 0.05$ was considered statistically significant.

## RESULTS

### Histogram analysis

Representative FA and MD images at both 3 T and 1.5 T are shown in Fig. 2. The results of FA histogram analysis are shown in Table 1. Peak location of FA was significantly lower at 3 T than at 1.5 T ($P = 0.04$). Mean FA was significantly higher at 3 T than at 1.5 T ($P = 0.002$). The results of MD histogram analysis are shown in Table 2. Peak location of MD was significantly lower at 3 T than at 1.5 T imaging ($P < 0.001$). Mean MD was significantly lower at 3 T than at 1.5 T ($P < 0.001$). The mean FA and MD histograms on which all the data were averaged by each bin are shown in Figs 3 and 4, respectively. These were different from individual histograms, and the information on peak locations and peak heights were compensated; however, they represent gross characteristics of FA and MD at 3 T and 1.5 T.

### ROI analysis

FA was significantly larger at 3 T than at 1.5 T in the centrum semiovale ($P < 0.001$), middle cerebellar peduncle ($P < 0.001$), cerebral peduncle ($P = 0.006$), posterior limb of the internal capsule ($P = 0.007$), genu ($P < 0.001$) and splenium ($P < 0.001$) (Table 3). It was significantly smaller at 3 T than at 1.5 T in the globus pallidus ($P < 0.001$). MD was significantly smaller at 3 T than at 1.5 T in the globus pallidus ($P = 0.007$), thalamus ($P < 0.001$), centrum semiovale ($P < 0.001$). middle cerebellar peduncle ($P < 0.001$), cerebral peduncle ($P = 0.01$), posterior limb of the internal capsule ($P < 0.001$), genu ($P = 0.01$) and splenium ($P < 0.001$) (Table 4).

Table 5 shows SNR index values. All were larger than 1, including that for the middle cerebellar peduncle for the posterior fossa.

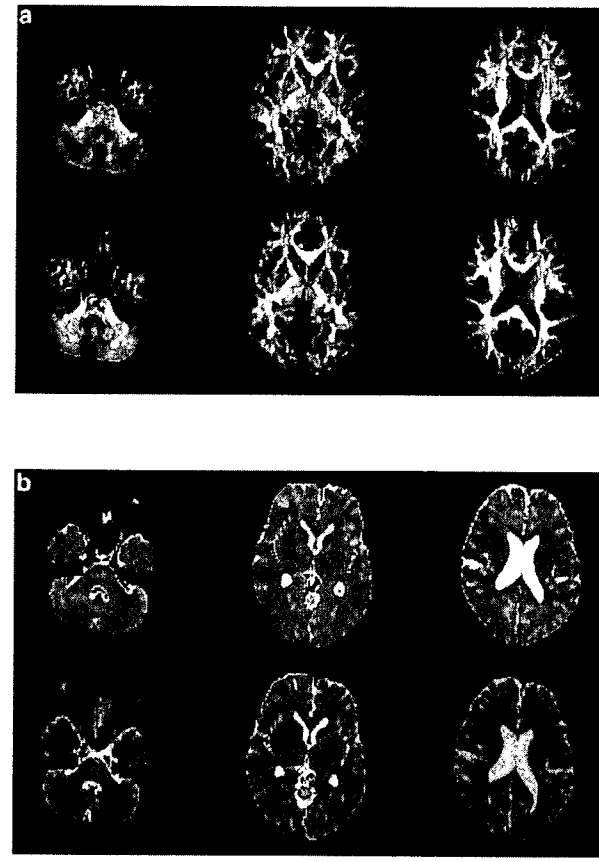


**Figure 2.** Transverse single-shot spin-echo echo-planar diffusion-tensor MRI at 3 T (*TR* 5200 ms, *TE* 79 ms) and 1.5 T (*TR* 5200 ms, *TE* 79 ms). (a) Top row shows images of FA obtained at 1.5 T. Bottom row shows images of FA obtained at 3 T. (b) Top row shows images of MD obtained at 1.5 T. Bottom row shows images of MD obtained at 3 T

**Table 1. Histogram analysis for FA. Values are mean ± SD[Q3]**

| | 1.5 T | 3 T | *P* value |
|---|---|---|---|
| Peak location | 0.111 ± 0.010 | 0.106 ± 0.013 | 0.04 |
| Peak height (%)* | 3.660 ± 0.129 | 3.681 ± 0.205 | 0.49 |
| Mean FA | 0.248 ± 0.007 | 0.253 ± 0.010 | 0.002 |

*Peak height for each volunteer was divided by the number of whole-brain voxels for normalization, therefore the data are percentages.

## DISCUSSION

Several comparison studies of DTI at 1.5 T have been reported (23,24), but few have focused on the differences

**Table 2. Histogram analysis for MD. Values are mean ± SD**

| | 1.5 T | 3 T | *P* value |
|---|---|---|---|
| Peak location ($10^{-3} \times mm^2/s$) | 0.862 ± 0.027 | 0.833 ± 0.025 | <0.001 |
| Peak height (%)* | 8.094 ± 0.773 | 7.883 ± 1.176 | 0.102 |
| Mean MD ($10^{-3} \times mm^2/s$) | 1.019 ± 0.040 | 0.974 ± 0.041 | <0.001 |

*Peak height for each volunteer was divided by the number of whole-brain voxels for normalization, therefore the data are percentages.

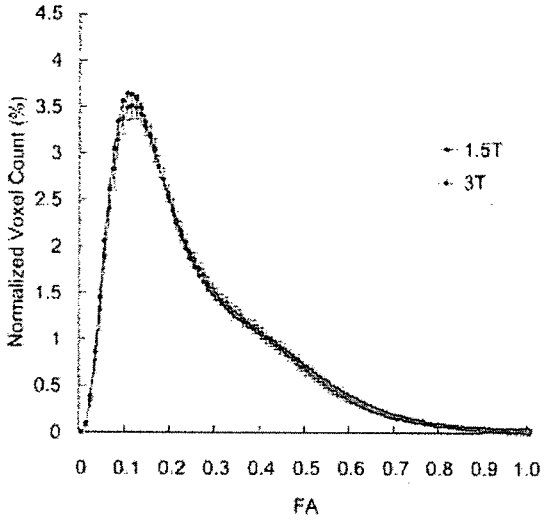


**Figure 3.** Mean histograms of FA at 3 T and 1.5 T. Error bars indicate standard deviations. Histograms were created for each FA image with a binwidth of 1% of maximum, and were divided by whole-brain voxels of that subject for normalization. The 3 T histogram is higher at both smaller value and larger values of FA than the 1.5 T histogram

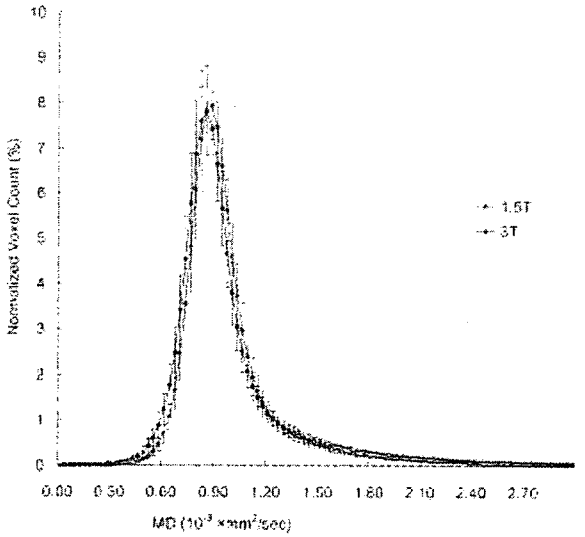


**Figure 4.** Mean histograms of MD at 3 T and 1.5 T. Error bars indicate standard deviations. Histograms were created for each MD image with a binwidth of 1% of maximum, and were divided by whole-brain voxels of that subject for normalization. The 3 T histogram shows a shift to the left

**Table 3. ROI analysis of FA at 3 T and 1.5 T. Values are mean ± SD**

| | 1.5 T | 3 T | *P* value |
|---|---|---|---|
| GP | 0.235 ± 0.039 | 0.197 ± 0.038 | <0.001 |
| Th | 0.295 ± 0.032 | 0.286 ± 0.021 | 0.17 |
| CS | 0.477 ± 0.041 | 0.551 ± 0.040 | <0.001 |
| mCP | 0.566 ± 0.058 | 0.615 ± 0.041 | <0.001 |
| CP | 0.644 ± 0.067 | 0.682 ± 0.049 | 0.006 |
| pIC | 0.682 ± 0.040 | 0.701 ± 0.037 | 0.007 |
| genu | 0.766 ± 0.052 | 0.821 ± 0.041 | <0.001 |
| splenium | 0.799 ± 0.037 | 0.847 ± 0.033 | <0.001 |

GP, Globus pallidus; Th, thalamus; CS, centrum semiovale; mCP, middle cerebellar peduncle; CP, cerebral peduncle; pIC, posterior limb of the internal capsule.

**Table 4. ROI analysis of MD at 3 T and 1.5 T. Values are mean ± SD**

| MD ($10^{-3} \times mm^2/s$) | 1.5 T | 3 T | *P* value |
|---|---|---|---|
| GP | 0.863 ± 0.045 | 0.819 ± 0.063 | 0.007 |
| Th | 0.866 ± 0.027 | 0.800 ± 0.024 | <0.001 |
| CS | 0.846 ± 0.026 | 0.776 ± 0.026 | <0.001 |
| mCP | 0.845 ± 0.031 | 0.778 ± 0.027 | <0.001 |
| CP | 0.905 ± 0.118 | 0.808 ± 0.139 | 0.01 |
| pIC | 0.826 ± 0.027 | 0.785 ± 0.023 | <0.001 |
| Genu | 0.863 ± 0.090 | 0.810 ± 0.065 | 0.01 |
| Splenium | 0.795 ± 0.049 | 0.717 ± 0.058 | <0.001 |

GP, Globus pallidus; Th, thalamus; CS, centrum semiovale; mCP, middle cerebellar peduncle; CP, cerebral peduncle; pIC, posterior limb of the internal capsule.

**Table 5. SNR index of DTI. Values are mean ± SD**

| | SNR index |
|---|---|
| GP | 1.40 ± 0.31 |
| mCP | 1.48 ± 0.29 |
| CS | 1.71 ± 0.38 |
| genu | 1.67 ± 0.48 |

SNR index was calculated as SNR at 3 T divided by SNR at 1.5 T. GP, Globus pallidus; CS, centrum semiovale; mCP, middle cerebellar peduncle.

between 3 T and 1.5 T (17,18). Hunsche *et al.* (17) reported differences in DTI among seven healthy volunteers at 3 T and 1.5 T using ROI analysis; however, the data were obtained with the conventional technique without using parallel imaging. There are no articles

featuring the differences in recent DTI using parallel imaging between 3 T and 1.5 T. In our study, the differences in FA and MD at both magnets were evaluated by using histogram and ROI analyses.

Histogram analyses showed that mean FA at 3 T was significantly higher than at 1.5 T, and the peak location of FA was significantly lower at 3 T than at 1.5 T. In histograms, the peak of FA represents a smaller value of FA, i.e. less anisotropic tissue such as gray matter. Higher mean FA and lower peak location at 3 T may represent better contrast between the larger and smaller value of FA. ROI analysis of FA at 3 T showed a significantly larger value than at 1.5 T for the centrum semiovale, middle cerebellar peduncle, cerebral peduncle, posterior limb of the internal capsule, genu, and splenium. These ROIs indicate relatively anisotropic tissues. Our results may support better visualization of DT tractography at 3 T than at 1.5 T (17,18).

The MD histogram at 3 T showed significantly lower peak location and mean MD than at 1.5 T. ROI analyses also revealed that all the ROIs of MD at 3 T were smaller than at 1.5 T. Our results are consistent with a previous study (17). FA of the globus pallidus was significantly lower at 3 T than at 1.5 T, and this might be attributable to increased susceptibility at 3 T of ferritin in the globus pallidus.

The SNR for DTI at 3 T showed a $1.40 \pm 0.31$ to $1.70 \pm 0.38$ gain compared with that at 1.5 T. Hunsche *et al.* (17) reported a $1.4 \pm 0.2$ gain in white matter at 3 T using a standard head coil; however, they excluded infratentorial regions from the analysis. Gonen *et al.* (25) reported a 1.23–1.46 gain at 3 T using a standard head coil. In the present study, parallel imaging was applied for both 3 T and 1.5 T DTI in order to reduce geometric distortion and acquisition time. Parallel imaging is essential to modern DTI (26,27) and may be attributable to higher SNR gain, including the posterior fossa and basal ganglia, compared with previous studies. In experimental studies, FA is sensitive to low SNR, because the value that is sensitive to isotropic diffusion, such as in cerebrospinal fluid, erroneously shows relatively high anisotropy in low SNR images (28,29). Conversely, MD is less sensitive to low SNR, as MD always underestimates true diffusion anisotropy (28).

Parallel imaging is an up-to-date technique and essential in DTI, especially at high-field magnetization, as it can reduce geometric distortion and susceptibility effects by shortening the echo train of the echo planar sequence (26,27). Furthermore, parallel imaging enables DTI with higher spatial resolution by reducing the echo train. We applied the same algorithm of parallel imaging to both 3 T and 1.5 T DTI and compared whole-brain histogram analyses at both magnets among many subjects, which has not been previously reported.

Determining differences in FA and MD under different magnetic fields is necessary, as DTI is performed at many institutes using various settings. The results of histogram and ROI analyses in this study add to our understanding of different magnets. Particularly when studying subtle changes in FA and MD, the use of MR machines of identical magnetic field strength is important.

We focused on FA and MD in this study as these two scalars are commonly used in clinical studies. They are rotationally invariant and appropriate for comparison among individuals because FA represents anisotropy of the diffusion tensor ellipsoid and MD is an orientationally averaged value of the diffusion tensor (30). On the other hand, the first eigenvector with the largest eigenvalue is assumed to represent the local fiber direction and is used for fiber tractography (20,21). However, there might be large differences in the first eigenvector among individuals, as it represents only one direction of the ellipsoid and it is not truly rotationally invariant.

## CONCLUSION

In conclusion, this study shows differences in FA and MD at 3 T and 1.5 T with whole-brain histogram and ROI analysis. DTI with parallel imaging at 3 T showed higher FA values than at 1.5 T in ROI analysis. Mean FA at 3 T was higher than at 1.5 T, and peak location at 3 T was lower than at 1.5 T in histogram analysis, which means anisotropic tissue such as white matter shows higher FA values at 3 T than at 1.5 T. Significant differences in FA and MD were shown at 3 T and 1.5 T in this study, which was different from the previous study (17). The SNR increase resulting from the use of parallel imaging and reduction in geometric distortion especially at 3 T may be the reasons. Furthermore, our data are supported by whole-brain analysis of a relatively large population, which was also different from the previous study (17). The SNR of DTI at 3 T increased with parallel imaging. Jaermann *et al.* (27) reported improved resolution at 3 T by using parallel imaging, which is due to the higher SNR and fewer susceptibility artifacts. However, there is no comparison study of DTI with parallel imaging at both 3 T and 1.5 T. We have revealed higher SNR in DTI using parallel imaging at 3 T than at 1.5 T.

## Acknowledgement

This study was supported in part by a Japanese Health and Labour Sciences Research Grant (H15-003).

## REFERENCES

1. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J.* 1994; **66**: 259–267.
2. Beaulieu C. The basis of anisotropic water diffusion in the nervous system: a technical review. *NMR Biomed.* 2002; **15**: 435–455.

3. Chenevert TL, Brunberg JA, Pipe JG. Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo. *Radiology* 1990; **177**: 401–405.
4. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. *Radiology* 1996; **201**: 637–648.
5. Sakuma H, Nomura Y, Takeda K, Tagami T, Nakagawa T, Tamagawa Y, Ishii Y, Tsukamoto T. Adult and neonatal human brain: diffusional anisotropy and myelination with diffusion-weighted MR imaging. *Radiology* 1991; **180**: 229–233.
6. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B* 1996; **111**: 209–219.
7. Filippi M, Cercignani M, Inglese M, Horsfield MA, Comi G. Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology* 2001; **56**: 304–311.
8. Bammer R, Augustin M, Strasser-Fuchs S, Seifert T, Kapeller P, Stollberger R, Ebner F, Hartung HP, Fazekas F. Magnetic resonance diffusion tensor imaging for characterizing diffuse and focal white matter abnormalities in multiple sclerosis. *Magn. Reson. Med.* 2000; **44**: 583–591.
9. Werring DJ, Clark CA, Barker GJ, Thompson AJ, Miller DH. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology* 1999; **52**: 1626–1632.
10. Bozzali M, Franceschi M, Falini A, Pontesilli S, Cercignani M, Magnani G, Scotti G, Comi G, Filippi M. Quantification of tissue damage in AD using diffusion tensor and magnetization transfer MRI. *Neurology* 2001; **57**: 1135–1137.
11. Muller MJ, Greverus D, Dellani PR, Weibrich C, Wille PR, Scheurich A, Stoeter P, Fellgiebel A. Functional implications of hippocampal volume and diffusivity in mild cognitive impairment. *Neuroimage* 2005; **28**: 1033–1042.
12. Ellis CM, Simmons A, Jones DK, Bland J, Dawson JM, Horsfield MA, Williams SC, Leigh PN. Diffusion tensor MRI assesses corticospinal tract damage in ALS. *Neurology* 1999; **53**: 1051–1058.
13. Sach M, Winkler G, Glauche V, Liepert J, Heimbach B, Koch MA, Buchel C, Weiller C. Diffusion tensor MRI of early upper motor neuron involvement in amyotrophic lateral sclerosis. *Brain* 2004; **127**: 340–350.
14. Nusbaum AO, Tang CY, Wei T, Buchsbaum MS, Atlas SW. Whole-brain diffusion MR histograms differ between MS subtypes. *Neurology* 2000; **54**: 1421–1427.
15. Cercignani M, Iannucci G, Rocca MA, Comi G, Horsfield MA, Filippi M. Pathologic damage in MS assessed by diffusion-weighted and magnetization transfer MRI. *Neurology* 2000; **54**: 1139–1144.
16. Cercignani M, Inglese M, Pagani E, Comi G, Filippi M. Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis. *AJNR Am. J. Neuroradiol.* 2001; **22**: 952–958.
17. Hunsche S, Moseley ME, Stoeter P, Hedehus M. Diffusion-tensor MR imaging at 1.5 and 3.0 T: initial observations. *Radiology* 2001; **221**: 550–556.
18. Okada T, Miki Y, Fushimi Y, Hanakawa T, Kanagaki M, Yamamoto A, Urayama S, Fukuyama H, Hiraoka M, Togashi K. Diffusion-tensor fiber tractography: intraindividual comparison of 3.0-T and 1.5-T MR imaging. *Radiology* 2006; **238**: 668–678.
19. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn. Reson. Med.* 1997; **38**: 591–603.
20. Mori S, van Zijl PC. Fiber tracking: principles and strategies: a technical review. *NMR Biomed.* 2002; **15**: 468–480.
21. Mori S, Frederiksen K, van Zijl PC, Stieltjes B, Kraut MA, Solaiyappan M, Pomper MG. Brain white matter anatomy of tumor patients evaluated with diffusion tensor imaging. *Ann. Neurol.* 2002; **51**: 377–380.
22. Edelstein WA, Bottomley PA, Pfeifer LM. A signal-to-noise calibration procedure for NMR imaging systems. *Med. Phys.* 1984; **11**: 180–185.
23. Cercignani M, Bammer R, Sormani MP, Fazekas F, Filippi M. Inter-sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. *AJNR Am. J. Neuroradiol.* 2003; **24**: 638–643.
24. Papanikolaou N, Karampekios S, Papadaki E, Malamas M, Maris T, Gourtsoyiannis N. Fractional anisotropy and mean diffusivity measurements on normal human brain: comparison between low- and high-resolution diffusion tensor imaging sequences. *Eur. Radiol.* 2006; **16**: 187–192.
25. Gonen O, Gruber S, Li BS, Mlynarik V, Moser E. Multivoxel 3D proton spectroscopy in the brain at 1.5 versus 3.0 T: signal-to-noise ratio and resolution comparison. *AJNR Am. J. Neuroradiol.* 2001; **22**: 1727–1731.
26. van den Brink JS, Watanabe Y, Kuhl CK, Chung T, Muthupillai R, Van Cauteren M, Yamada K, Dymarkowski S, Bogaert J, Maki JH, Matos C, Casselman JW, Hoogeveen RM. Implications of SENSE MR in routine clinical practice. *Eur. J. Radiol.* 2003; **46**: 3–27.
27. Jaermann T, Crelier G, Pruessmann KP, Golay X, Netsch T, van Muiswinkel AM, Mori S, van Zijl PC, Valavanis A, Kollias S, Boesiger P. SENSE-DTI at 3 T. *Magn. Reson. Med.* 2004; **51**: 230–236.
28. Bastin ME, Armitage PA, Marshall I. A theoretical study of the effect of experimental noise on the measurement of anisotropy in diffusion imaging. *Magn. Reson. Imaging* 1998; **16**: 773–785.
29. Armitage PA, Bastin ME. Selecting an appropriate anisotropy index for displaying diffusion tensor imaging data with improved contrast and sensitivity. *Magn. Reson. Med.* 2000; **44**: 117–121.
30. Basser PJ, Jones DK. Diffusion-tensor MRI: theory, experimental design and data analysis: a technical review. *NMR Biomed.* 2002; **15**: 456–467.

*Jpn J Clin Oncol 2007;37(8)615–622*
*doi:10.1093/jjco/hym064*

# Inter-observer Variations in FDG-PET Interpretation for Cancer Screening

**Akiko Suzuki[1], Yuji Nakamoto[2], Takashi Terauchi[3], Masami Kawamoto[4], Yoshihiro Okumura[5], Yutaka Suzuki[6], Toshihiko Sato[7], Nobukazu Takahashi[8], Jin Lee[8], Michio Senda[9], Kimiichi Uno[10] and Tomio Inoue[8]**

[1]Department of Radiology, School of Medicine, Yokohama City University, Yokohama, [2]Department of Diagnostic Radiology, Kyoto Graduate University School of Medicine, Kyoto, [3]Research Center for Cancer Prevention and Screening, Cancer Screening Division, National Cancer Center, Tokyo, [4]Diagnostic Imaging Center, Radiology, Yuai Clinic, Yokohama, [5]Department of Radiology, PET/RI center, Okayama Kyokuto Hospital, Okayama, [6]HIMEDIC Imaging Center at Lake Yamanaka, Yamanashi, [7]Utsunomiya Central Clinic PET Center, Utsunomiya, [8]Department of Radiology, School of Medicine, Yokohama City University, Yokohama, [9]Institute of Biomedical Research and Innovation, Kobe and [10]Nishidai Clinic Diagnostic Imaging Center, Tokyo, Japan

Received October 16, 2006; accepted March 20, 2007; published online August 18, 2007

**Background:** Diagnostic guidelines for the use of 2-(fluorine 18) fluoro-2 deoxy-D-glucose (FDG)-positron emission tomography (PET) in cancer screening have yet to be established. We assessed inter-observer variability in screening FDG-PET.
**Methods:** Subjects comprised 40 individuals who underwent FDG-PET and computed tomography (CT) for cancer screening. To assess various patterns of FDG uptakes, three subsets of the cases were selected: 'Cancer', 15 cases with cancer; 'Not malignant', 15 cases with suspected cancer by FDG-PET who were confirmed as cancer-free; and 'Normal', 10 cases without remarkable FDG uptake who were confirmed as cancer-free. A total of 68 lesions made up of malignancy ($n = 18$), benign ($n = 21$), and physiological FDG uptake ($n = 29$) were interpreted by six physicians. Each observer reviewed each case three times. Step 1 involved interpretation of PET images alone, Step 2 involved side-by-side reading of PET and CT images, and Step 3 involved re-evaluation of findings with the results of other screening tests. We assessed inter-observer agreement for each step.
**Results:** Inter-observer agreement for all lesions at each step was moderate, compared to fair agreement for 'Normal' subjects. Inter-observer agreement of 'Cancer' and 'Not malignant' subjects in Step 1 were better than those in Step 2 and 3; however, the differences were not statistically significant.
**Conclusion:** The interpretation of FDG-PET is adequately reproducible, while that of 'Normal' subjects is less reproducible. Improvement of inter-observer variability in assessing physiological FDG uptakes requires universal reporting criteria in FDG-PET. Correlative interpretation of PET, CT and other information may require standardization in subjects with suspected cancer by FDG-PET.

*Key words: radiology – PET – radiology – CT/MRI – cancer screening*

## INTRODUCTION

2-(fluorine 18) fluoro-2 deoxy-D-glucose (FDG)-positron emission tomography (PET) plays an important role in the detection of malignant tumors, although the effectiveness of whole-body FDG-PET imaging in cancer screening remains uncertain (1,2). FDG-PET scans have been performed for cancer screening in Japan on many asymptomatic individuals who had previously never been diagnosed with cancer. FDG-PET is considered to be useful for whole-body survey because it can detect cancers of various organs that any single conventional organ-specific screening test cannot cover. High detection rates for a wide variety of cancers in cancer-screening FDG-PET has been reported by Yasuda (3)

For reprints and all correspondence: Akiko Suzuki, Department of Radiology, School of Medicine, Yokohama City University, 3–9 Fukuura, Kanazawa-ku, Yokohama, Kanagawa, 236-0004, Japan. E-mail: akiko225@yokohama-cu.ac.jp

and Chen (4); however, FDG-PET cannot be an alternative to other conventional screening tests such as physical examination, laboratory studies, mammography and thoracic computed tomography (CT), because FDG-PET analysis has obvious limitations in detecting urological cancers, cancers of low cell density, small cancers and hypometabolic or FDG-negative cancers (1,3,5). Diagnostic guidelines for the use of whole-body FDG-PET imaging in cancer screening have yet to be established. Furthermore, FDG-PET may be better interpreted with reference to CT images in cancer screening, as the determination of the precise location of FDG-avid lesions using PET alone can be challenging (6). The role of whole-body FDG-PET and CT in cancer screening is yet to be evaluated.

We surveyed a large number of cancer-screening centers in Japan in January 2005 to investigate the actual situation of cancer screening by FDG-PET. Thirty cancer-screening centers answered the questionnaire and the results were reported (data not published). The recall rate (i.e. the rate recommending diagnostic work-up due to positive findings suggesting possible cancer) varied widely from 1 to 44% betweem the centers. We hypothesized that inter-observer variation in FDG-PET interpretation for cancer screening affected clinical decisions to recommend either close examination or follow-up and caused variability in the recall rate. Inter-observer variation in FDG-PET for cancer screening is of particular interest because of the challenge involved in detecting suspected lesions, of which the incidence is very low, out of numerous cases of equivocal FDG uptake; some radiologists tend to over-diagnose FDG uptake to avoid potential false negative outcomes. Numerous investigations have shown that considerable variability exists among radiologists in the interpretation of screening tests such as mammography and thoracic CT without training or computer-aided diagnosis (7–9). This variability affects the diagnostic accuracy of screening studies and clinical decisions to recommend either close examination or follow-up. Herder et al. reported that inter-observer agreement of FDG-PET between clinical and final stage was good in patients with suspected lung cancer (10). Inter-observer agreement of interpretation in relevant focal pulmonary abnormality of FDG-PET was also reported to be good by Joshi et al. (11). To the best of our knowledge, variability in radiologists' interpretations of whole-body FDG-PET for cancer screening has yet to be examined. The purpose of the present study is to assess inter-observer variations in screening FDG-PET.

## MATERIALS AND METHODS

### Case Materials

FDG-PET and CT data of 40 subjects (21 male, 19 female, median age 57 years) were collected from seven cancer-screening centers in Japan and were used in this study. The scanning took place between April 2004 and March 2005; all subjects were symptom-free and underwent FDG-PET and CT on the same day for cancer screening together with other physical and laboratory tests. The 40 subjects consisted of Group 1 'Cancer', 15 true positive cases with suspected cancer by FDG-PET who were confirmed to have cancer by biopsy; Group 2 'Not malignant', 15 false positive cases with suspected cancer by FDG-PET and recommended for close examination who were confirmed as cancer-free on follow-up or biopsy; and Group 3 'Normal', 10 true negative cases who had no suspected lesion detected in FDG-PET and were confirmed to be cancer-free at 1-year follow-up. It should be noted that the number of cases in each group does not reflect the fractional occurrence of each outcome in cancer screening using FDG-PET. This study did not deal with false negative cases on PET because this is peripheral to the primary aim of assessing inter-observer variation in the interpretation of malignant lesions without remarkable FDG uptakes. Among the 15 'Cancer' subjects, primary disease involved the lung ($n = 3$), thyroid ($n = 3$), colon ($n = 2$), breast ($n = 2$), stomach ($n = 2$), pancreas ($n = 1$) and malignant lymphoma ($n = 2$). These cancers are commonly detected in FDG-PET and CT during cancer screening (1,4).

On reviewing the 40 subjects together with all the reference data, a total of 103 lesions were identified, presenting varying intensities of FDG uptake. Thirty-five of 103 lesions that were considered true negative lesions in 'Cancer' and 'Not malignant' cases were excluded because of the absence of confirmed reference data. A final total of 68 lesions were diagnosed as malignant ($n = 18$), benign ($n = 21$), or physiological FDG uptake ($n = 29$). Among the 15 'Cancer' subjects, 18 malignant lesions were detected and enrolled as true positive lesions. In the 15 'Not malignant' subjects, 13 benign lesions and five physiological uptakes were enrolled as false positive lesions. In the 10 'Normal' subjects, eight benign lesions and 24 physiological uptakes were enrolled as true negative lesions. All malignant lesions and seven benign lesions were confirmed on biopsy. At 1 year after FDG-PET, 14 benign lesions and 29 physiological uptakes were confirmed as stable or diminished in uptake. In the present study, we examine the interpretations of these 68 lesions made by each physician.

### Scanning of Whole-Body PET and CT Study

FDG-PET and CT of the 40 cases were performed in seven different institutions between April 2004 and March 2005. All PET images were obtained using a standardized protocol in accordance with the FDG-PET guidelines for cancer screening issued in 2004 by the Japanese Society of Nuclear Medicine. Patients fasted for at least 5 h prior to scanning. We obtained a whole-body PET image from the head to the thigh using a PET or PET/CT scanner at 50–60 min following the injection of 300–450 MBq of FDG. Transmission images were obtained to correct for photon attenuation using a germanium-68 line source. For PET/CT, PET attenuation

correction factors were calculated from the CT images. We reconstructed image datasets using the ordered-subsets expectation maximization algorithm. We acquired a whole-body CT image from the head to the pelvis without intravenous contrast agent using a CT scanner or PET/CT scanner. The CT scanners and technical parameters were as follows: (i) Robusto (Hitachi Medico, Tokyo, Japan) multi-detector four row CT, 120 kVp, 100–160 mAs, beam pitch 1.75 and 10 mm thickness; (ii) Light Speed Ultra (GE Medical Systems, Tokyo, Japan) multi-detector eight row CT, 120 kVp, 175 mAs maximum with automatic exposure control system, 1.35 pitch and 5 mm thickness; (iii) Brilliance 16 (Philips Electronics Japan, Tokyo, Japan) multi-detector 16 row CT, 120 kVp, 150 mAs maximum with automatic exposure control system, 0.9 pitch and 5 mm thickness; (iv) Biograph LSO (Siemens-Asahi Medical Technologies, Tokyo, Japan) PET/CT with multi-detector two row CT, 130 kVp, 80 mAs maximum with automatic exposure control system, 1.15 pitch and 4 mm thickness; (v) Biograph LSO (Siemens-Asahi Medical Technologies), PET/CT with multi-detector two row CT, 130 kVp, 60 mAs, 1.5 pitch and 3 mm thickness; (vi) Eminence-SOPHIA (Shimadzu Corporation, Kyoto, Japan) helical CT, 120 kVp, 187.5 mAs, 1.4 pitch and 7 mm thickness; (vii) CT-Turbo (Hitachi Medico) helical CT, 120 kVp, 100–120 mAs, and 5–10 mm thickness. The mAs settings were selected to optimize spatial and contrast resolution. Tube current modulation was used to minimize the radiation dose to the individuals.

The PET/CT images were divided into PET and CT images, which were interpreted by side-by-side reading.

### Observing Radiologists

FDG-PET and CT data were interpreted by six physicians with experience in both FDG-PET and CT, but who had not previously seen the study cases. The six observers were based at the six different cancer-screening centers with various recall rates from 1 to 44% [1, 10, 12, 23, 35, 44] that were mentioned in the introduction; each observer had between 4 and 10 years experience in reading screening FDG-PET.

### Image Interpretation

Each observer reviewed each case in three steps. Step 1 involved interpretation of PET images alone, Step 2 involved side-by-side reading of PET and CT, and Step 3 involved re-evaluation of findings with reference to past history, smoking and drinking habits, and the results of other screening tests performed at the same time such as blood tests, fecal occult blood inspection and other imaging modalities that included magnetic resonance (MR) imaging for lower abdomen assessment and ultrasonography (US) for upper abdomen and thyroid gland assessment. In Step 2, each observer interpreted PET and CT by side-by-side reading without using fusion images. Findings of FDG uptake in each step were recorded as site and score depending on the likelihood of malignancy (1–5 points: 1, definitely not malignant; 2, probably not malignant; 3, equivocal; 4, probably malignant; 5, definitely malignant).

Each observer gave a score rating for every lesion that he/she considered to represent remarkable FDG uptake in each case. We analyzed each observer's interpretation of the 68 lesions that had been identified in advance as having confirmed reference data. Any of the 68 lesions that were not recognized by the observer as remarkable FDG uptake were given a score of 1 (definitely not malignant). Any lesions that were detected other than the 68 previously mentioned were excluded from analysis because of the absence of confirmed reference data.

All images were viewed with the same software using Synapse, medical imaging and information management network system, housed at the Fujifilm's demonstration showroom in Ginza, Tokyo, Japan.

### Data Analyses

We assessed observer accuracy and variation using the scores given by each observer based on the relevant lesion. Sensitivity, specificity and positive predictive value (PPV) data were calculated from the interpretations of the six observers for each step. Receiver-operating characteristic (ROC) analysis was also performed as the standard method for evaluating observer accuracy, as sensitivity and specificity offer an incomplete description of accuracy and depend on the decision threshold selected by the observer to define positive diagnoses. The area under the ROC curve (Az) was used as a summary index of accuracy. Sensitivity for malignancy was calculated as the proportion of malignancies given a rating of 3–5. Specificity was defined as the fraction of benign lesions or physiological FDG uptakes for which a rating of 1–2 was reported. Inter-observer agreement in 68 lesions on the likelihood of malignancy (1–5 points) for each step was also assessed using the $\kappa$ statistic.

### Statistical Analysis

The Wilcoxon matched pairs signed rank sum test was applied to the sensitivity, specificity, and PPV means for each step to test for significant differences. Values of $P < 0.05$ were considered indicative of statistically significant differences. We calculated weighted $\kappa$ values to describe concordance in reporting as 'slight' (0.00–0.20), 'fair' (0.21–0.40), 'moderate' (0.41–0.60), 'substantial' (0.61–0.80), or 'almost perfect' (0.81–1.00) (12,13). We conducted all analyses using MedCalc for Windows, version 7.6.0.0 (MedCalc Software, Mariakerke, Belgium), except for ROC analysis, which was performed using the software ROCKIT (C. Metz, University of Chicago, Chicago, IL, USA). ROC software was used to fit a binormal ROC curve

to the data from each observer and to compare Az for each pair using a univariate z-score test (14).

## RESULTS

### Diagnostic Accuracy of Each Observer

A summary of ROC curves obtained from interpretation in Step 1 using PET alone is shown in Fig. 1. Az values did not differ significantly between the six observers for each step. The means and dispersion of sensitivity, specificity, PPV and Az are shown in Table 1. Although Az values did not differ significantly, sensitivity, specificity and PPV varied widely between the observers.

### Effect of Reference to CT on Diagnostic Accuracy

The mean specificity increased significantly when observers referred to CT ($P < 0.05$), although sensitivity, PPV and Az did not change significantly (Table 1).

### Variability in Interpretation

The mean $\kappa$ value and the strength of inter-observer agreement for each step are shown in Table 2. Inter-observer agreement for all lesions in each step was moderate ($\kappa = 0.58$ for Step 1; $\kappa = 0.55$ for Step 2; and $\kappa = 0.53$ for Step 3).

Inter-observer agreement was higher for 'Cancer' and 'Not malignant' lesions than for 'Normal' lesions for each step (moderate versus fair).

Although the $\kappa$ values for each group in Step 1 were higher than those in Steps 2 and 3, the differences were not statistically significant. The $\kappa$ values of 'Cancer' and 'Not malignant' lesions in Step 2 were higher than those in Step 3; however, the differences were not statistically significant.

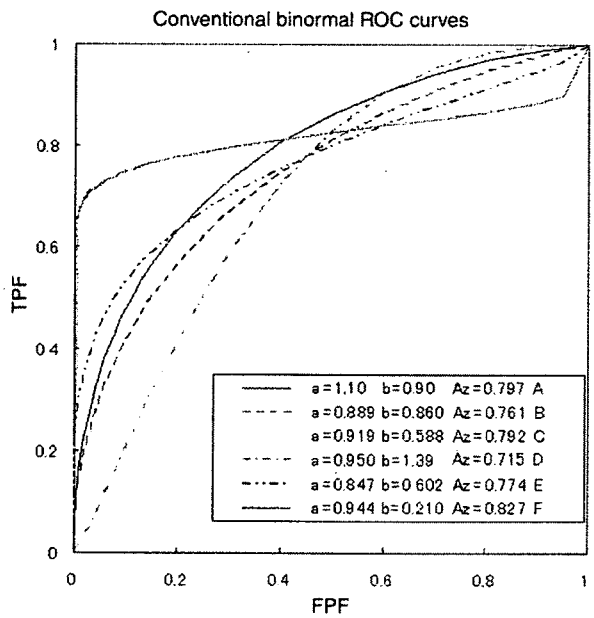


**Figure 1.** Summary receiver-operating characteristic curves obtained from interpretation in Step 1 of evaluations with positron emission tomography alone (please note that a colour version of this figure is available as supplementary data at http://www.jjco.oxfordjournals.org). TPF, true positive fraction; FPF, false positive fraction.

### Patterns of FDG Uptake with Poor Agreement

In assessing inter-observer agreement for each organ, we determined the organs that were interpreted with difficultly by PET alone. The numbers of sites presenting poor or good agreement in each organ for the 68 lesions in Step 1 using PET alone are shown in Tables 3–5. Lesions for which fewer than five observers agreed in diagnosis were considered as poor agreement, while lesions for which five or six observers agreed in diagnosis were considered as good agreement.

FDG uptakes in the 'Normal' 32 true negative lesions that presented poor agreement included nine physiological uptakes in the larynx, mediastinum, intestine and ovary, and four benign lesions in the thyroid, neck, lung and uterus (Table 3). The case with physiological FDG uptake in the ascending colon is shown in Fig. 2.

FDG uptakes in the 18 'Cancer' true positive lesions that presented poor agreement included 10 malignant lesions in the thyroid, hilum, breast, colon and stomach (Table 4).

FDG uptakes in the 18 'Not malignant' false negative lesions that presented poor agreement included eight benign lesions in the thyroid, lung, colon and joint, and four physiological uptakes in the hilum, intestine and ovary (Table 5).

## DISCUSSION

### Diagnostic Accuracy of Each Observer

We assessed observer accuracy using the scores representing a rating on the likelihood of malignancy for 68 lesions by each observer. No significant differences were identified for any step in Az of the six observers. Wide variation in sensitivity, specificity and PPV detected between observers was caused by differing decision thresholds during interpretation. Selection of different thresholds does not cause Az to vary, as an ROC curve depicts all of the tradeoffs available as the threshold is varied. Therefore, variability of the decision threshold between observers exists where no significant differences were identified in diagnostic accuracy as quantified with Az. That is, the scores given for each lesion could vary between the six observers even though there were no significant differences in diagnostic accuracy indicated by Az.

**Table 1.** Mean (range of six observers) sensitivity, specificity, PPV and Az of FDG-PET cancer screening based on lesions

| | Step 1 | Step 2 | Step 3 |
|---|---|---|---|
| | PET alone | PET + CT | PET + CT + Other information |
| Sensitivity | 78.7 (72.2–83.3) | 75.9 (66.7–83.3) | 72.2 (61.1–83.3) |
| Specificity | 64.7* (58.0–74.0) | 71.3* (64.0–78.0) | 74.0* (66.0–82.0) |
| PPV | 44.9 (41.7–51.9) | 49.2 (41.9–57.7) | 50.6 (39.3–60.9) |
| Az** | 0.778 (0.715–0.827) | 0.788 (0.718–0.851) | 0.794 (0.704–0.848) |

PET, positron emission tomography; CT, computed tomography; PPV, positive predictive value.
*Mean specificity of PET + CT and PET + CT + Other information was higher than that of PET alone (Wilcoxon matched pairs signed rank sum test, $P < 0.05$).
**Az, area under curve in ROC.

**Table 2.** Mean $\kappa$ value and strength of inter-observer agreement for the likelihood of malignancy of 68 FDG-avid lesions in 40 cases

| | | Step 1 | Step 2 | Step 3 |
|---|---|---|---|---|
| | | PET alone | PET + CT | PET + CT + Other information |
| Total lesions ($n = 68$) | $\kappa$ | 0.58 | 0.55 | 0.53 |
| | Agreement | Moderate | Moderate | Moderate |
| 'Normal' ($n = 32$) | $\kappa$ | 0.30 | 0.28 | 0.28 |
| | Agreement | Fair | Fair | Fair |
| 'Cancer' ($n = 18$) | $\kappa$ | 0.57 | 0.44 | 0.42 |
| | Agreement | Moderate | Moderate | Moderate |
| 'Not malignant' ($n = 18$) | $\kappa$ | 0.51 | 0.46 | 0.41 |
| | Agreement | Moderate | Moderate | Moderate |

FDG, 2-(fluorine 18) fluoro-2 deoxy-D-glucose.
$\kappa$ value <0, poor; 0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect.
'Normal': 32 true negative lesions that were not suspected malignant in FDG-PET cancer screening and were confirmed as physiological uptakes or benign lesions after 1 year.
'Cancer': 18 true positive lesions that were suspected malignant in FDG-PET screening and were later diagnosed as malignant lesions.
'Not malignant': 18 false positive lesions that were suspected malignant in FDG-PET screening and were later diagnosed as benign lesions or physiological uptakes.

### Effect of Reference to CT on Diagnostic Accuracy

We defined sensitivity as the proportion of malignancies given a rating of 3–5 and specificity as the fraction of benign lesions or physiological FDG uptakes given a rating of 1–2. This classification has critical meaning, therefore, Az has less value when comparing test performance between Steps 1 to 3. The mean specificity increased significantly in this study when observers referred to CT. Chen et al. also reported that additional CT for localization and lesion characterization showed an increased specificity of PET for cancer screening in asymptomatic individuals (4). In the present study, however, mean sensitivity and PPV did not change significantly when observers referred to CT. Several investigators report that the combination of FDG-PET and CT significantly improves diagnostic accuracy in the diagnosis of malignancy (15–17). The fact that the present results demonstrate no improvement in sensitivity and PPV may be due to selection bias: the present study did not include false negative lesions of PET that are recognized in CT. Sensitivity on general FDG-PET screening may be improved when observers refer to CT, given the inclusion of CT positive lesions that are without remarkable FDG uptakes such as bronchioloalveolar lung carcinoma (18).

### Variability in Interpretation

We assessed inter-observer agreement on the scores for likelihood of malignancy (1–5 points) in 68 lesions. The 68 lesions that presented varying intensities of FDG uptake were founded in whole-body FDG-PET performed for cancer screening of 40 asymptomatic individuals.

Inter-observer agreement for all lesions in each step was moderate. Berg et al. reported that inter-observer agreement

**Table 3.** Agreement in each site of 'Normal' 32 true negative lesions in Step 1 using PET alone

| Sites | Diagnosis | No. of sites | No. of each agreement | |
|---|---|---|---|---|
| | | | Poor** | Good*** |
| Larynx | Physiological* | 2 | 2 | 0 |
| Oral cavity | Physiological | 4 | 0 | 4 |
| Mediustinum | Physiological | 2 | 1 | 1 |
| Epigastrium | Physiological | 5 | 0 | 5 |
| Intestine | Physiological | 6 | 4 | 2 |
| Ovary | Physiological | 1 | 1 | 0 |
| Uterus | Physiological | 1 | 0 | 1 |
| Prostate | Physiological | 2 | 0 | 2 |
| Ureter | Physiological | 1 | 0 | 1 |
| Thyroid | Thyroadenitis | 1 | 1 | 0 |
| Neck | Lymphadenopathy | 1 | 1 | 0 |
| Lung | Pneumonia | 2 | 1 | 1 |
| Rib | Fracture | 1 | 0 | 1 |
| Shoulder | Arthritis | 1 | 0 | 1 |
| Liver | Cyst | 1 | 0 | 1 |
| Uterus | Myoma | 1 | 1 | 0 |
| Total | | 32 | 12 | 20 |

*Physiological: physiological FDG uptake.
**Poor: fewer than five observers agreed in diagnosis.
***Good: five or six observers agreed in diagnosis.

**Table 4.** Agreement in each site of the 18 'Cancer' true positive lesions in Step 1 using PET alone

| Sites | Diagnosis | No. of sites | No. of each agreement | |
|---|---|---|---|---|
| | | | Poor | Good |
| Thyroid | Thyroid Cancer | 3 | 3 | 0 |
| Lung | Lung Cancer | 2 | 0 | 2 |
| Hilum | LN metastasis or lung cancer | 2 | 2 | 0 |
| Breast | Breast cancer | 2 | 1 | 1 |
| Colon | Colon cancer | 2 | 2 | 0 |
| Stomach | Gastric cancer | 2 | 2 | 0 |
| Pancreas | Pancreas cancer | 1 | 0 | 1 |
| Subclavicular | Lymphoma | 1 | 0 | 1 |
| Mediustinum | Lymphoma | 1 | 0 | 1 |
| Paraaorta | Lymphoma | 1 | 0 | 1 |
| Mesenterium | Lymphoma | 1 | 0 | 1 |
| Total | | 18 | 10 | 8 |

LN metastasis, lymph node metastasis.

among radiologists on mammogram screening after training in Breast Imaging Reporting and Data System was moderate (9). Our results suggest that interpretation of FDG-PET in cancer screening is adequately reproducible as a whole.

Inter-observer agreement for all lesions at each step was moderate, compared to fair agreement for 'Normal' subjects. The higher prevalence of malignant lesions (18/68) means that the set used in this study was not strictly representative of FDG-PET within the general screening population. Inter-observer agreement on general FDG-PET screening might normally be lower, given the inclusion of a larger number of normal healthy subjects. Low inter-observer agreement may cause the marked variability in recall rate among the institutions that perform screening FDG-PET.

Inter-observer agreement was lower for 'Normal' lesions than for 'Cancer' and 'Not malignant' lesions for each step (fair versus moderate). Since sensitivity are calculated based on the data for 'Cancer' and specificity are based on those for 'Not malignant' and 'Normal', inter-observer variation observed for 'Cancer' and 'Not malignant and Normal' are corresponding to variability in sensitivity and specificity between observers (58.0–74.0, 72.2–83.3 in Step 1, respectively).

Inter-observer agreement decreased when observers referred to CT, however, the differences were not statistically

**Table 5.** Agreement in each site of the 18 'Not malignant' false positive lesions in Step 1 using PET alone

| Sites | Diagnosis | No. of sites | No. of each agreement | |
|---|---|---|---|---|
| | | | Poor | Good |
| Thyroid | Goiter or thyroadenitis | 3 | 2 | 1 |
| Parotid gland | Warthin tumor | 1 | 0 | 1 |
| Lung | Pneumonia or pleural tumor | 4 | 2 | 2 |
| Sternoclavicular joint | Arthritis | 1 | 1 | 0 |
| Colon | Adenoma or polyp | 4 | 3 | 1 |
| Hilum | Physiological | 1 | 1 | 0 |
| Intestine | Physiological | 1 | 1 | 0 |
| Stomach | Physiological | 1 | 0 | 1 |
| Ovary | Physiological | 2 | 2 | 0 |
| Total | | 18 | 12 | 6 |

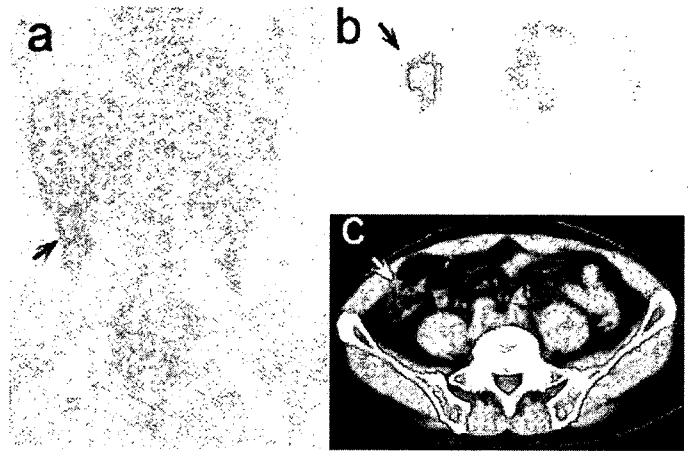


**Figure 2.** Physiological 2-(fluorine 18) fluoro-2 deoxy-D-glucose uptake in the intestine of 'Normal' subjects that presented poor agreement: a 54-year-old male. (a) Coronal and (b) transaxial FDG-PET image revealed focal uptake in the ascending colon (arrow). (c) The corresponding lesion was not demonstrated on CT. CT, computed tomography.

significant. Although side-by-side reading of PET and CT improves lesion localization and supports lesion characterization, correlative interpretation of PET and CT differed between observers. Metser et al. concluded that in-line PET/CT offers better lesion localization relative to visual fusion of PET and CT, especially for small lymph nodes, lesions adjacent to mobile organs and lesions adjacent to the chest or abdominal wall (19). Syed et al. reported that PET/CT increases inter-observer agreement and confidence in disease localization of FDG-avid lesions in patients with head and neck cancers (20). By precisely localizing FDG uptakes, interpretation of image fusion by integrated PET/CT might offer higher inter-observer agreement in comparison to interpretation of PET images alone or side-by-side interpretation of PET and CT images.

Although inter-observer agreement of 'Cancer' and 'Not malignant' lesions decreased when observers referred to other information, the differences were not statistically significant. In cancer screening, positive lesions are eventually recommended for diagnostic work-up or observation with close follow-up. The clinical recommendation is determined by evaluation of the PET and CT findings with reference to past history, smoking and drinking habits, and results of other screening tests. Correlative interpretation of PET, CT and other information may need to be standardized to achieve greater agreement in subjects with suspected cancer by FDG-PET.

### Patterns of FDG Uptake with Poor Agreement

We evaluated the organs that were most difficult to interpret by PET alone. Some observers tended to over-diagnose FDG uptake to avoid potential false negative outcome, while others did not pick up suspected lesions in 'Normal' subjects. Physiological FDG uptake is recognized at various sites in various degrees. A focal intense uptake in intestine mimics FDG uptake of colon tumor as shown in Fig. 1. Reporting criteria for various patterns of FDG uptakes in intestine differed between observers. For higher agreement on results for the 'Normal' 32 true negative lesions in FDG-PET, interpretation of physiological FDG uptake for the larynx, mediastinum, intestine and ovary should be standardized.

Various FDG uptakes in goiter, pneumonia, colon adenoma and arthritis confound image interpretation of FDG-PET and act to reduce inter-observer agreement. Interpretation of FDG-avid lesions in the thyroid, lung, hilum, breast, colon, stomach and ovary may require standardization for higher agreement in 'Cancer' and 'Not malignant' subjects.

## CONCLUSIONS

Our results suggest that interpretation of FDG-PET in cancer screening is adequately reproducible, whereas interpretation of physiological FDG uptake in 'Normal' subjects is less reproducible. Improvement of inter-observer variability in assessing physiological FDG uptakes requires universal reporting criteria in FDG-PET. Furthermore, correlative interpretation of PET, CT and other information may require standardization in subjects with suspected cancer by FDG-PET.

### Acknowledgments

This research was supported in part by a Grant from Foundation for Promotion of Cancer Research and by a Health and Labor Sciences Research Grant for the project titled 'Third Term Comprehensive Control Research for Cancer' from the Ministry of Health, Labor, and Welfare, Japan.

**Conflict of interest statement**
None declared.

## References

1. Yasuda S, Ide M. PET and cancer screening. *Ann Nucl Med* 2005;19:167–77.
2. Kubota K. From tumor biology to clinical PET: a review of positron emission tomography (PET) in oncology. *Ann Nucl Med* 2001;15:471–86.
3. Yasuda S, Ide M, Fujii H, Nakahara T, Mochizuki Y, Takahashi W, et al. Application of positron emission tomography imaging to cancer screening. *Br J Cancer* 2000;83:1607–11.
4. Chen YK, Ding HJ, Su CT, Shen YY, Chen LK, Liao AC, et al. Application of PET and PET/CT imaging for cancer screening. *Anticancer Res* 2004;24:4103–8.
5. Nakai K, Yasuda S, Ide M, Kawada S, Shohtsu A. Cancer screening with 18F FDG-PET. In: Oeher P, Biersack HJ, Coleman RE editors. *PET and PET-CT in Oncology*. Heidelberg: Springer 2004; 309–20.
6. Ak I, Stokkel MP, Pauwels EK. Positron emission tomography with 2-(18F) fluoro-2-deoxy-D-glucose in oncology. Part II. The clinical value in detecting and staging primary tumours. *J Cancer Res Clin Oncol* 2000;126:560–74.
7. Jiang Y, Nishikawa RM, Schmidt RA, Toledano AY, Doi K. Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. *Radiology* 2001;220:787–94.
8. Rubin GD, Lyo JK, Paik DS, Sherbondy AJ, Chow LC, Leung AN, et al. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. *Radiology* 2005;234:274–83.
9. Berg WA, D'Orsi CJ, Jackson VP, Bassett LW, Beam CA, Lewis RS, et al. Does training in the Breast Imaging Reporting and Data System (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? *Radiology* 2002;224:871–80.
10. Herder GJ, Kramer H, Hoekstra OS, Smit EF, Pruim J, van Tinteren H, et al. Traditional versus up-front (18F) fluorodeoxyglucose-positron emission tomography staging of non-small-cell lung cancer: a Dutch cooperative randomized study. *J Clin Oncol* 2006;24:1809–6.
11. Joshi U, Raijmakers PG, van Lingen A, Comans EF, Pijpers R, Teule GJ, et al. Evaluation of pulmonary nodules: comparison of a prototype dual crystal (LSO/NAI) dual head coincidence camera and full ring positron emission tomography (PET). *Eur J Radiol* 2005;55:250–4.
12. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003;228:303–8.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
14. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 1989;24:234–45.
15. Eubank WB, Mankoff DA, Schmiedl UP, Winter TC, III, Fisher ER, Olshen AB, et al. Imaging of oncologic patients: benefit of combined CT and FDG PET in the diagnosis of malignancy. *Am J Roentgenol* 1998;171:1103–10.
16. Aquino SL, Asmuth JC, Alpert NM, Halpern EF, Fischman AJ. Improved radiologic staging of lung cancer with 2-(18F)-fluoro-2-deoxy-D-glucose-positron emission tomography and computed tomography registration. *J Comput Assist Tomogr* 2003;27:479–84.
17. Yoshida Y, Kurokawa T, Kawahara K, Tsuchida T, Okazawa H, Fujibayashi Y, et al. Incremental benefits of FDG positron emission tomography over CT alone for the preoperative staging of ovarian cancer. *Am J Roentgenol* 2004;182:227–33.
18. Higashi K, Ueda Y, Seki H, Yuasa K, Oguchi M, Noguchi T, et al. Fluorine-18-FDG PET imaging is negative in bronchioloalveolar lung carcinoma. *J Nucl Med* 1998;39:1016–20.
19. Metser U, Golan O, Levine CD, Even-Sapir E. Tumor lesion detection: when is integrated positron emission tomography/computed tomography more accurate than side-by-side interpretation of positron emission tomography and computed tomography? *J Comput Assist Tomogr* 2005;29:554–9.
20. Syed R, Bomanji JB, Nagabhushan N, Hughes S, Kayani I, Groves A, et al. Impact of combined (18)F-FDG PET/CT in head and neck tumours. *Br J Cancer* 2005;92:1046–50.

ELSEVIER

Journal of the Neurological Sciences xx (2006) xxx–xxx

Journal of the Neurological Sciences

www.elsevier.com/locate/jns

Short communication

# Dopaminergic neuronal dysfunction associated with parkinsonism in both a Gaucher disease patient and a carrier

Satoshi Kono [a,*], Kentaro Shirakawa [a], Yasuomi Ouchi [b], Masanobu Sakamoto [b], Hiroyuki Ida [c], Takeshi Sugiura [a], Hiroyuki Tomiyama [d], Hitoshi Suzuki [a], Yoshitomo Takahashi [a], Hiroaki Miyajima [a], Nobutaka Hattori [d], Yoshikuni Mizuno [d]

[a] *First Department of Medicine, Hamamatsu University School of Medicine, 1-20-1 Handayama, Hamamatsu 431-3192, Japan*
[b] *Department of Neurology, Positron Medical Center, Hamamatsu Medical Center, Hamamatsu, Japan*
[c] *Department of Pediatrics, Jikei University School of Medicine, Tokyo, Japan*
[d] *Department of Neurology, Juntendo University School of Medicine, Tokyo, Japan*

Received 8 September 2006; received in revised form 23 October 2006; accepted 30 October 2006

**Abstract**

A clinical association between Gaucher disease and parkinsonism has been demonstrated. We herein report a Japanese patient with type 3 Gaucher disease who was compound heterozygous for F213I and L444P mutations in the glucocerebrosidase gene while his father was heterozygous for the L444P mutation. They both presented with parkinsonism characterized by a predominance of akinetic-rigid signs and a favorable response to anti-Parkinson therapies. We investigated the dopaminergic neuronal function using positron emission tomography (PET) with radioligands, [$^{11}$C] CFT and [$^{11}$C] raclopride. PET studies of both patients demonstrated the [$^{11}$C] CFT uptake to be severely decreased in the putamen and the caudate nucleus, however, the [$^{11}$C] raclopride uptake was normal in the basal ganglia. Although the majority of Gaucher disease patients with parkinsonism tend to be refractory to anti-Parkinson therapies. The clinical features and the findings of the PET studies suggest that patients with parkinsonism associated with the mutation in the glucocerebrosidase gene, even in heterozygosis, may be related to the presynaptic dopaminergic neuronal dysfunction reported in Parkinson's disease. A PET study to evaluate the dopaminergic neuronal function in Gaucher disease would provide both a better understanding of the effects of anti-Parkinson therapies and a help to improve our ability to make an early diagnosis of parkinsonism associated with Gaucher disease.


*Keywords:* Gaucher disease; Parkinsonism; Glucocerebrosidase; PET

## 1. Introduction

Gaucher disease is an autosomal recessive lysosomal disorder resulting from a deficiency of the lysosomal enzyme glucocerebrosidase which lead to the systemic storage of glycosphingolipids [1]. This disease is caused by mutations in the glucocerebrosidase gene located on1q21. Recent studies have revealed an association between Gaucher disease and Parkinson's disease due to a concurrence of type 1 Gaucher disease and parkinsonism and the identification of glucocerebrosidase mutations in patients with sporadic Parkinson's disease [2–7]. Initial studies of the patients affected from Gaucher disease with the parkinsonism showed that the parkinsonism was characterized by an early onset and it tended to be refractory to levodopa therapy [2,3,5], however, there is an increasing number of reports which showed parkinsonism to demonstrate the following signs of typical Parkinson disease: namely, the asymmetric onset of rigidity, resting tremor, bradykinesia, and a favorable response to Parkinson therapies [4,8]. Treatment-refractory parkinsonism suggests that mutations in the glucocerebrosidase gene may affect either postsynaptic dopaminergic neurons or both post- and presynaptic dopaminergic neurons.

* Corresponding author. Tel.: +81 53 435 2261; fax: +81 53 434 9447.
*E-mail address:* satokono@hama-med.ac.jp (S. Kono).

0022-510X/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jns.2006.10.019

Please cite this article as: Kono S et al. Dopaminergic neuronal dysfunction associated with parkinsonism in both a Gaucher disease patient and a carrier. J. Neurol. Sci. (2006), doi:10.1016/j.jns.2006.10.019