

An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest

Rahul S. Desikan,^a Florent Ségonne,^c Bruce Fischl,^{b,c} Brian T. Quinn,^b
Bradford C. Dickerson,^h Deborah Blacker,^d Randy L. Buckner,^{b,e,f} Anders M. Dale,^g
R. Paul Maguire,^j Bradley T. Hyman,^h Marilyn S. Albert,ⁱ and Ronald J. Killiany^{a,*}

^aDepartment of Anatomy and Neurobiology, Boston University School of Medicine, 715 Albany Street, W701, Boston, MA 02118, USA

^bAthinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA 02118, USA

^cComputer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA 02139, USA

^dDepartment of Psychiatry, Massachusetts General Hospital, Charlestown, MA 02129, USA

^eDepartment of Psychology, Harvard University, Cambridge, MA 02138, USA

^fHoward Hughes Medical Institute, USA

^gDepartments of Neuroscience, Radiology and Cognitive Science, University of California, San Diego School of Medicine, San Diego, CA 92093, USA

^hDepartment of Neurology, Massachusetts General Hospital, Charlestown, MA 02129, USA

ⁱDepartment of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

^jPfizer Global Research and Development, Groton, CT 06340-8012, USA

Received 1 August 2005; revised 26 October 2005; accepted 12 January 2006

Available online 10 March 2006

In this study, we have assessed the validity and reliability of an automated labeling system that we have developed for subdividing the human cerebral cortex on magnetic resonance images into gyral based regions of interest (ROIs). Using a dataset of 40 MRI scans we manually identified 34 cortical ROIs in each of the individual hemispheres. This information was then encoded in the form of an atlas that was utilized to automatically label ROIs. To examine the validity, as well as the intra- and inter-rater reliability of the automated system, we used both intraclass correlation coefficients (ICC), and a new method known as mean distance maps, to assess the degree of mismatch between the manual and the automated sets of ROIs. When compared with the manual ROIs, the automated ROIs were highly accurate, with an average ICC of 0.835 across all of the ROIs, and a mean distance error of less than 1 mm. Intra- and inter-rater comparisons yielded little to no difference between the sets of ROIs. These findings suggest that the automated method we have developed for subdividing the human cerebral cortex into standard gyral-based neuroanatomical regions is both anatomically valid and reliable. This method may be useful for both morphometric and functional studies of the cerebral cortex as well as for clinical investigations aimed at tracking the evolution of disease-induced changes over time, including clinical trials in which MRI-based measures are used to examine response to treatment.

© 2006 Elsevier Inc. All rights reserved.

Introduction

Structural magnetic resonance imaging (MRI) provides extensive detail about the anatomical structure of the brain. It is becoming increasingly important for characterizing cortical changes associated with the normal aging process (see Raz, 2004 for review) and further differentiating these from the degenerative changes associated with dementing illnesses such as Alzheimer's disease (AD). Furthermore, structural MRI has now become an essential tool for the clinical care of patients with brain disease and is of increasing use in clinical trials to identify response to treatment. For example, MRI is now a secondary endpoint in clinical trials of patients with multiple sclerosis (MS), as MS lesions can now be quantified quickly and reliably (see Bakshi et al., 2005). Research and clinical investigations of patients with AD are beginning to incorporate MRI measurements, but these have been primarily restricted to assessments of whole brain atrophy (Freeborough and Fox, 1998a,b; Fox et al., 1999, 2001, 2005) or manual measures of the hippocampus (Jack et al., 1995, 1997, 2003).

The application of MRI to research and clinical studies has been limited by the ability to quantify the critical dimensions of interest. Methods have been developed to automatically quantify regions of interest (ROI), but these have not as of yet been incorporated into clinical trials. Some of these methods have focused on a single ROI, such as the hippocampus (Hsu et al., 2002; Csernansky et al., 2000, 2005; Wang et al., 2003), and the cingulate gyrus (Miller et al., 2003) or sets of subcortical ROIs (Fischl et al., 2002). Developing semi-automated procedures for quantifying cortical

* Corresponding author.

E-mail address: killiany@bu.edu (R.J. Killiany).

Available online on ScienceDirect (www.sciencedirect.com).

ROIs has been more challenging due to the substantial inter-individual variability of the topographic features of the cortex (Zilles et al., 1988; Ono et al., 1990; Kennedy et al., 1998).

Initial efforts at measuring cortical ROIs on MRI scans required substantial operator involvement (e.g., Damasio and Damasio, 1989; Rademacher et al., 1992). More recent and more automated methods employ a variety of approaches to the problem of labeling cortical features, including template-driven warping approaches, where a local correspondence is established between a manually labeled atlas brain and individual subject's brain images (Thompson et al., 1996; Sandor and Leahy, 1997; Hammers et al., 2003; Buckner et al., 2004; Mega et al., 2005), watershed-based approaches to extract cortical sulci (Lohmann, 1998; Rettmann et al., 2002), and graph-based techniques which represent sulci as vertices on a graph (Mangin et al., 1995; Le Goualher et al., 1999).

We recently reported a probabilistic labeling algorithm (Fischl et al., 2004) that was applied to two different systems for defining cortical regions of interest (Rademacher et al., 1992; Destrieux et al., 1998). The strength of this algorithm is that it is not tied to a specific neuroanatomical template, but instead incorporates not only the probable location of a region of interest, but also the potential inter-subject variance of the location of the region, derived from whatever training set is employed.

In the present study, we have expanded on this work in several ways. First, we have developed the definitions of the regions of interest using curvature based information (i.e., sulcal representations) available on images of the cortex that are 'inflated' (Dale and Sereno, 1993; Dale et al., 1999; Fischl et al., 1999a,b, 2001; Fischl and Dale, 2000); anatomic curvature is visually represented well on inflated images as they provide a view of the brain in which the entire cortical surface is exposed, including the tissue deep in the sulci. Moreover, since the same type of anatomic curvature information is utilized by the probabilistic labeling algorithm, we hypothesized that defining labels on the inflated surface using curvature information would improve the accuracy of the manual definitions of the cortical regions of interest. Second, we employed a training set consisting of 40 MRI scans that included young, middle-aged and elderly controls, as well as patients with Alzheimer's disease (AD). These 40 scans were manually labelled for 34 cortical regions of interest, and an atlas was generated.

With regard to brain atlases, two approaches can be used towards developing atlases. The first approach involves the identification and selection of a highly selected group of individuals (such as cases of AD patients with CDR 1.0) and building an atlas that is optimized for that group alone. Though it is hypothetically possible to construct an atlas from such a selected set of homogenous cases, apply this atlas onto a larger cohort of similar cases and thus achieve high accuracy, it has limited practical value for large morphometric studies aiming to assess anatomic changes across different population types. The second approach, such as the one presented in this manuscript, involves the development of a more generalized atlas that incorporates a wide range of anatomic and atrophic variance. When applied, the result produces an atlas that is likely to be slightly less accurate within a selected group (i.e., when applied only to cases of AD patients with CDR 1.0) but is applicable across several different groups and ultimately, is more accurate across these groups. This more generalized approach to atlas building is important to utilize when the underlying variable is continuous making group distinctions somewhat arbitrary.

We used intraclass correlation coefficients (ICC) to assess validity and reliability. Since we were particularly interested in the anatomical accuracy of the regions of interest, we developed a method that uses the mean distance of "mislabeling" on the cortical surface (known as mean distance maps) to detect the geographical mismatch between manual and automated regions or between sets of regions generated by the same (intra-rater) or different (inter-rater) operators applying the automated process. Finally, since we were interested in assessing the applicability of our automated atlas, we employed a jackknife/leave-one-out technique (a statistical re-sampling method) to test the reliability of our atlas on novel datasets.

Materials and methods

Subjects

The participants in the study were enrolled by the Washington University Alzheimer's Disease Research Center (ADRC) in St. Louis. As such, all were screened for neurological impairment, depression and psychoactive medications use (see Fotenos et al., 2005 for details of this sample). As a part of this assessment, all subjects were screened for the presence of major vascular risk factors (e.g., atrial fibrillation, diabetes). Subjects found to have clinically relevant abnormalities on MRI (e.g., tumors, infarcts) were also excluded. Individuals with white matter hyperintensities and other non-specific findings on MRI were included in the sample. Adults 60 and over were also clinically screened for dementia and classified based on the Clinical Dementia Rating (Morris, 1993).

The MRI scans of 40 subjects from this cohort were used in the present analyses. As noted above, for the development of a cortical atlas, we wanted to include subjects with a range of atrophy. We therefore selected MRI scans from subjects that varied widely in age and clinical status to incorporate the types of variance we find in our typical studies of aging and dementia. The 40 subjects were therefore divided into four groups: Group 1—young adults ($n = 10$; mean age = 21.5, age range 19–24; 6 females, 4 males); Group 2—middle-aged adults ($n = 10$; mean age = 49.8, age range 41–57; 7 females, 3 males); Group 3—elderly adults ($n = 10$; mean age = 74.3, age range 66–87; 8 females, 2 males); and Group 4—patients with AD ($n = 10$; mean age = 78.2, age range 71–86; 5 females, 5 males).

MRI image acquisition

The MRI scans were acquired on a 1.5T Vision system (Siemens, Erlangen Germany). T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) scans were obtained according to the following protocol: two sagittal acquisitions, FOV = 224, Matrix = 256 × 256, Resolution = 1 × 1 × 1.25 mm, TR = 9.7 ms, TE = 4 ms, Flip angle = 10°, TI = 20 ms, TD = 200 ms. Two acquisitions were averaged together to increase the contrast-to-noise ratio. The images were collected as part of the ongoing operations of the Washington University ADRC and have appeared in prior publications (e.g., Buckner et al., 2004; Fotenos et al., 2005). Anonymized comparable data (from Head et al., 2005) can be freely obtained from the fMRI Data Center (<http://www.fmridc.org/fmridc>; accession number 2-2004-1168X).

Cortical surface generation from MRI scans

Each scan was first corrected for motion, averaged, normalized for intensity and resampled to isotropic dimensions of $1 \times 1 \times 1$ mm using previously published algorithms that are distributed in the FreeSurfer software package (<http://www.martinos.org/freesurfer>) (Dale and Sereno, 1993; Dale et al., 1999; Fischl et al., 1999a). Next, the skull was removed from the images using a skull-stripping algorithm (Ségonne et al., 2004) and the images were segmented to identify the dorsal, ventral and lateral extent of the gray/white matter boundary to provide a surface representation of the cortical white matter (Dale and Sereno, 1993; Dale et al., 1999; Fischl et al., 1999a). The quality of the skull stripping and accuracy of the gray/white matter boundary for each subject was reviewed by two anatomically skilled operators (RSD, BTQ). The cortical white matter surfaces generated by the steps above were automatically corrected for topological defects (Fischl et al., 2001; Ségonne et al., 2005), and thereafter utilized in a deformation procedure that locates the pial (gray matter) surface of the brain (Fischl and Dale, 2000).

Manual delineation of cortical regions of interest

The cerebral hemispheres were subdivided into 34 regions each by one operator (RSD), who was blind to participants' age, gender, and group status. A 'sulcal' approach (manual tracing from the depth of one sulcus to another, thus incorporating the gyrus within) was used to define most structures (see details below).

Several sources of information were used to guide the delineation of neuroanatomical ROIs on volumetric MRI images, including: (1) standard neuroanatomical conventions based on brain atlases (Duvernoy, 1991; Ono et al., 1990), (2) modifications to previously published definitions (Killiany et al., 1993, 2000; Wible et al., 1995, 1997; Crespo-Facorro et al., 2000; Van Hoesen et al., 2000; Halliday et al., 2003; Yamasue et al., 2004; Ballmaier et al., 2004; Onitsuka et al., 2004) and (3) consultations with Drs. Thomas Kemper and Douglas Rosene (Kemper and Rosene, personal communication). This information was then used to define the ROIs on the T1 images. These volumetric ROIs were then transposed onto the 'inflated' cortical surface of each reconstructed brain (Dale and Sereno, 1993; Fischl et al., 1999a) and using anatomic information regarding local curvature (e.g., the presence of sulci), the final anatomic labels were made. As noted above, the inflated surface allows for visualization of anatomic information across the entire cortical surface (i.e., both the sulci and gyri) without interference from cortical folding. For example, as noted in Fig. 1, the cortex around the perimeter of the central sulcus is buried between the pre- and post-central gyri and thus not visible on the pial surface (see white asterisk noting location), but the presence of this cortex is clearly identifiable in the inflated image (see yellow asterisks noting location). This type of anatomic curvature information is potentially very useful for delineating manual regions of interest on the inflated images and moreover is utilized by the probabilistic labeling procedure to develop the automated atlas. The anatomic definitions utilized for each region were as follows:

Temporal lobe—medial aspect

Entorhinal cortex. The rostral and caudal boundaries of the entorhinal cortex were the rostral end of the collateral sulcus and the caudal end of the amygdala, respectively. The medial boundary

was the medial aspect of the temporal lobe and the lateral boundary was the collateral sulcus.

Parahippocampal gyrus. The rostral and caudal boundaries of the parahippocampal gyrus were the posterior end of the entorhinal cortex and the caudal portion of the hippocampus (where it could be identified inferomedial to the trigone of the lateral ventricle), respectively. The medial boundary was designated as the medial aspect of the temporal lobe and the lateral boundary was the collateral sulcus.

Temporal pole. The temporal pole lies in the anterior portion of the temporal lobe (rostral boundary) and extends caudally to the entorhinal cortex. The medial and lateral boundaries were the medial aspect of the temporal lobe and the superior or inferior temporal sulci, respectively.

Fusiform gyrus. The rostral boundary of the fusiform gyrus was the rostral extent of the collateral sulcus. The caudal boundary was defined on the inflated surface as the rostral limit of the lateral occipital cortex. The medial and lateral boundaries were the collateral sulcus and the occipitotemporal sulcus, respectively.

Temporal lobe—lateral aspect

Superior temporal gyrus. The rostral boundary of the superior temporal gyrus was the rostral extent of the superior temporal sulcus. The caudal boundary was the caudal portion of the superior temporal gyrus (posterior to becoming continuous with the supramarginal gyrus). The medial boundary was the lateral fissure (and when present, the supramarginal gyrus), and the superior temporal sulcus was utilized as the lateral extent.

Middle temporal gyrus. The rostral boundary of the middle temporal gyrus was the rostral extent of the superior temporal sulcus, and the caudal boundary was the temporo-occipital incisure on the cortical surface. The superior temporal sulcus was the medial boundary, and the inferior temporal sulcus was the lateral boundary.

Inferior temporal gyrus. The rostral boundary of the inferior temporal gyrus was the rostral extent of the inferior temporal sulcus, and the caudal boundary was designated as the lateral occipital cortex on the cortical surface. The occipitotemporal sulcus was the medial boundary, and the inferior temporal sulcus was the lateral boundary.

Transverse temporal cortex. The rostral boundary was the rostral extent of the transverse temporal sulcus, and the caudal boundary was the caudal portion of the insular cortex. The lateral fissure and the superior temporal gyrus were utilized as the medial and lateral boundaries, respectively.

Banks of the superior temporal sulcus. (defined as the posterior aspect of the superior temporal sulcus). The rostral boundary was the superior temporal gyrus and the caudal boundary was the middle temporal gyrus.

Frontal lobe

Superior frontal gyrus. The rostral boundary of the superior frontal gyrus was the rostral extent of the superior frontal sulcus, and the caudal boundary was the paracentral sulcus on the inflated surface.

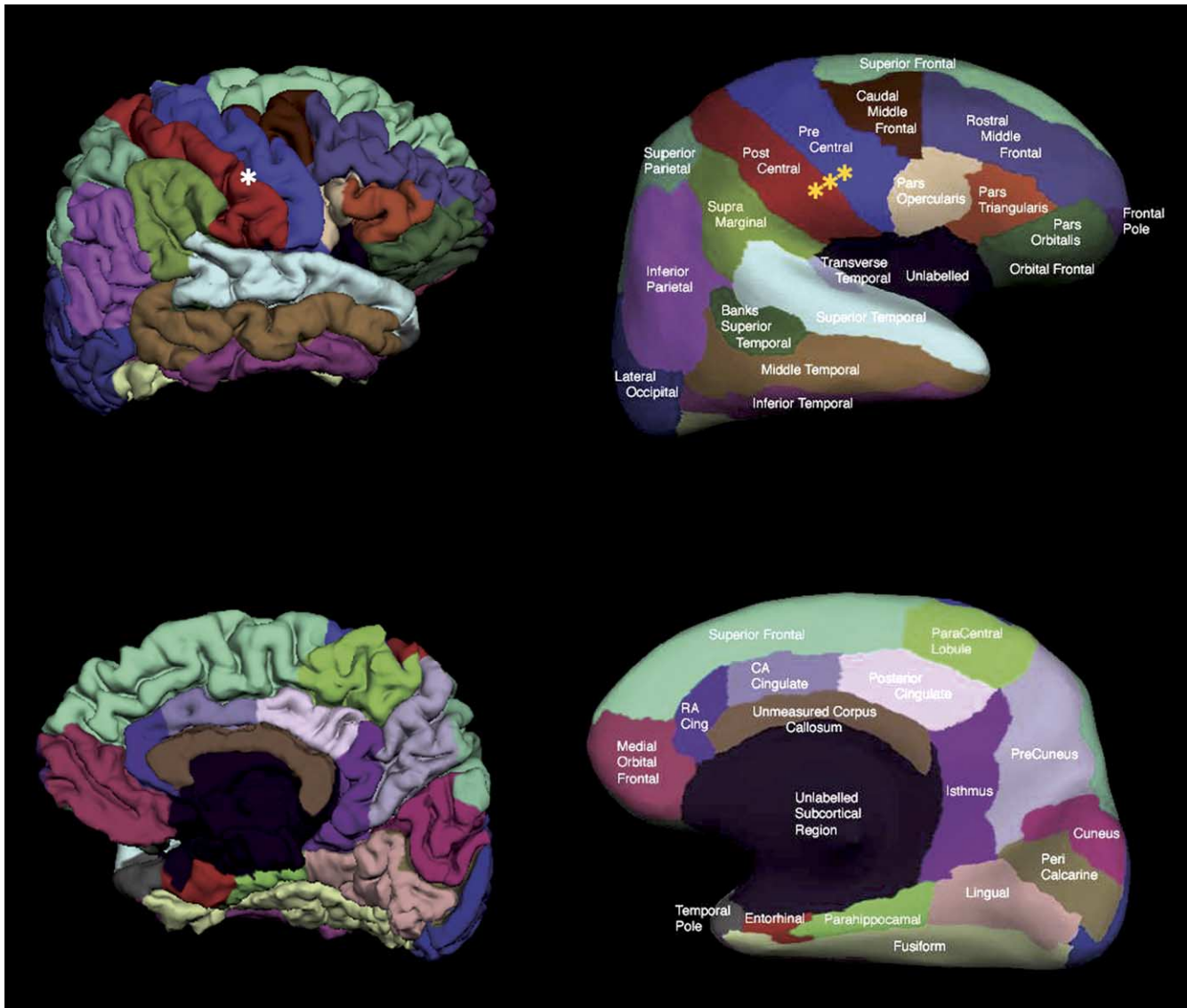


Fig. 1. Pial (left) and inflated (right) cortical representations of the regions of interest in one hemisphere. The top row illustrates the lateral view of the hemisphere while the bottom row shows the medial view of the hemisphere. The white asterisk on the pial surface (left) indicates the cortex around the perimeter of the central sulcus that is buried within the gyri and thus not visible. The yellow asterisks on the inflated surface (right) indicate the cortex around the perimeter of the central sulcus that has been ‘inflated’ and is now visible.

The medial and lateral boundaries were designated as the medial aspect of the frontal lobe and the superior frontal sulcus, respectively.

Middle frontal gyrus. Subdivided into:

(a) *Rostral division.* The rostral boundary was the rostral extent of the superior frontal sulcus, and the caudal boundary was the caudal extent of the middle frontal gyrus. The medial and lateral boundaries were the superior frontal sulcus and the inferior frontal sulcus, respectively.

(b) *Caudal division.* The rostral boundary was the rostral extent of the middle frontal gyrus, and the caudal boundary was the precentral gyrus. The medial and lateral boundaries were designated as the superior frontal sulcus and the inferior frontal sulcus, respectively.

Inferior frontal gyrus. First, the whole of the inferior frontal gyrus was labelled volumetrically, and then was subdivided on

the surface. The rostral boundary was the rostral extent of the inferior frontal sulcus, and the caudal boundary was the precentral gyrus. The medial and lateral boundaries were the lateral bank of the inferior frontal sulcus and the medial bank of the lateral orbital sulcus and/or the circular insular sulcus, respectively.

(a) *Pars opercularis.* The first gyrus from the precentral gyrus.

(b) *Pars triangularis.* The second gyrus from the precentral gyrus.

(c) *Pars orbitalis.* The remainder of the inferior frontal gyrus once the pars opercularis and triangularis have been defined.

Orbitofrontal cortex. Subdivided into:

(a) *Lateral division.* The rostral boundary of the lateral division of the orbitofrontal cortex was the rostral extent of

the lateral orbital gyrus (where it appears with the frontomarginal sulcus), and the caudal boundary was the caudal portion of the lateral orbital gyrus. The medial and lateral boundaries were the midpoint of the olfactory sulcus and the lateral bank of the lateral orbital sulcus and/or the circular insular sulcus, respectively.

(b) *Medial division.* The rostral boundary of the medial division of the orbitofrontal cortex was the rostral extent of the medial orbital gyrus, and the caudal boundary was the caudal portion of the medial orbital gyrus/gyrus rectus. The medial and lateral boundaries were the cingulate cortex on the inflated surface and the medial bank of the superior frontal gyrus (or the cingulate gyrus when visible), respectively.

Frontal pole. The rostral and caudal boundaries of the frontal pole were the superior frontal gyrus and the rostral division of the middle frontal gyrus, respectively. Note that the frontal pole was manually designated using an exclusionary criterion (other frontal lobe regions were first designated and the remaining portion was called frontal pole) and is not actually used as a measure of the frontal pole itself.

Precentral gyrus. The rostral and caudal extents of the central sulcus were the rostral and caudal boundaries of the precentral gyrus, respectively. The medial boundary was specific frontal gyri (superior, middle and inferior), and the lateral boundary was the medial bank of the central sulcus.

Paracentral lobule. The rostral boundary of the paracentral lobule was the posterior extent of the superior frontal gyrus, and the caudal boundary was the rostral extent of the precuneus cortex. The medial and lateral boundaries were the medial aspect of the cingulate cortex and the superior frontal gyrus (or pre- and postcentral gyri when visible), respectively.

Parietal lobe

Postcentral gyrus. The rostral and caudal extents of the central sulcus were the rostral and caudal boundaries of the postcentral gyrus, respectively. The medial and lateral boundaries were the lateral bank of the precentral gyrus and the lateral fissure and/or the medial bank of the superior parietal gyrus, respectively.

Supramarginal gyrus. The caudal extent of the superior temporal gyrus was the rostral boundary, and the rostral extent of the superior parietal gyrus was the caudal boundary. The medial and lateral boundaries were the lateral banks of the intraparietal sulcus and the medial banks of the lateral fissure and/or the superior temporal gyrus, respectively.

Superior parietal cortex. The rostral and caudal boundaries of the superior parietal cortex were the precentral gyrus and lateral occipital cortex, respectively. The medial and lateral boundaries were the precuneus and/or cuneus cortex and the inferior parietal cortex, respectively.

Inferior parietal cortex. The inferior parietal cortex region includes the inferior parietal gyrus and the angular gyrus and lies inferior to the superior parietal gyrus. The rostral and

caudal boundaries were the supramarginal gyrus and the lateral occipital cortex, respectively. The medial and lateral boundaries were the superior parietal gyrus and the middle temporal gyrus, respectively.

Precuneus cortex. The rostral boundary was the posterior extent of the paracentral lobule, and the caudal boundary was the lingual gyrus. The medial and lateral boundaries were the parieto-occipital fissure and the superior parietal gyrus, respectively.

Occipital lobe

Lingual gyrus. The rostral boundary of the lingual gyrus was the posterior extent of the parahippocampal gyrus, and the caudal boundary was the most posterior portion of the occipital cortex. The medial and lateral boundaries were the medial portion of the temporal and occipital cortices and the medial bank of the collateral sulcus, respectively.

Pericalcarine cortex. The rostral boundary of the pericalcarine cortex was the rostral extent of the calcarine sulcus, and the caudal boundary was the most posterior portion of the occipital cortex. The medial and lateral boundaries were the medial portion of the temporal and occipital cortices and the inferomedial end of the calcarine sulcus, respectively.

Cuneus cortex. The rostral and caudal extents of the calcarine sulcus were designated as the rostral and caudal boundaries of the cuneus cortex. The medial boundary was the most medial portion of the occipital and parietal cortices. The supero-lateral boundary was the parieto-occipital fissure, and the inferolateral boundary was the pericalcarine cortex.

Lateral occipital cortex. The rostral and caudal boundaries of the lateral occipital cortex were the superior parietal gyrus and as the last visible portion of occipital cortex, respectively. The medial and lateral boundaries were the cuneus/pericalcarine cortex and the inferior temporal/inferior parietal gyri, respectively.

Cingulate cortex

Rostral anterior division. The rostral boundary was the rostral extent of the cingulate sulcus (inferior to the superior frontal sulcus), and the caudal boundary was the genu of the corpus callosum. The medial boundary was the medial aspect of the cortex. The supero-lateral boundary was the superior frontal gyrus, and the inferolateral boundary was defined as the medial division of the orbitofrontal gyrus.

Caudal anterior division. The rostral boundary was the genu of the corpus callosum, and the caudal boundary was established as the mammillary bodies. The medial and lateral boundaries were the medial aspect of the cortex and the superior frontal gyrus, respectively.

Posterior division. The rostral and caudal extent were the caudal anterior and the isthmus divisions of the cingulate cortex, respectively. The medial and lateral boundaries were the corpus callosum and the superior frontal gyrus and/or paracentral lobule, respectively.

Isthmus division. The rostral and caudal boundaries were the posterior division of the cingulate cortex and the parahippocampal gyrus, respectively. The medial and lateral boundaries were the medial wall (area unknown) and the precuneus, respectively.

Corpus callosum. The rostral and caudal extents of the corpus callosum were the medial division of the orbitofrontal cortex and the isthmus division of the cingulate cortex, respectively. The medial and lateral boundaries were the medial wall (area unknown) and the divisions of the cingulate cortex, respectively. Note this region of interest was included to restrict the boundaries of other regions and is not actually used as a measure of the corpus callosum itself since the procedures described in this manuscript measure the volume between the white matter and the gray matter surfaces.

Construction of cortical atlas

Once the manually drawn ROIs for both hemispheres of all 40 brains were completed, a cortical atlas was generated using a registration procedure that aligns the cortical folding patterns (Fischl et al., 1999b) and probabilistically assigns a neuroanatomical region to every point on the cortical surface (Fischl et al., 2004). The probabilistic algorithm, in an initial step, generates a spherical representation of each brain by minimizing the metric distortion between the cortical and the spherical representations. Next, all spherical surfaces are registered together. An energy functional measuring the alignment of the cortical folding patterns with the average is iteratively minimized by gradient-descent onto the sphere in a multi-scale manner (the full energy functional is described in Fischl et al., 1999b). This establishes a spherical surface-based coordinate system that is adapted to the folding pattern of each individual subject thus allowing for increased precision in registering anatomic features of the human brain across subjects (Fischl et al., 1999b). Then, a spherical statistical atlas is used to label the cortical surfaces into neuroanatomical regions of interest. This procedure models the labeling system as a first order anisotropic non-stationary Markov random field on the curvature of the cortical surface, allowing it to capture the spatial relationships and variance between regions that are present in the training set (see Fischl et al., 2002, 2004 for more details on the probabilistic labeling procedure). Therefore, the automated system incorporates information about sulcal and gyral geometry with spatial information regarding the location of brain structures (derived from the manually drawn regions of interest) and the variance of that information included in the training set to determine the regions of interest.

Data analysis

Intraclass correlations

Validity was assessed by comparing the volumes generated manually to those generated by the automated system. Reliability was assessed by comparing the volumes generated by the automated system on two occasions by the same operator, blind to both the clinical status and the identity of each subject (i.e., intra-rater reliability) or by two operators independently processing all of the cases, blind to both the clinical status and identities of subject (i.e., inter-rater reliability). In both cases, volumes generated were compared using intraclass correlation coefficients. For comparisons between the manual and automated system (assessing validity), we

used an ICC based on a two-way analysis of variance (ANOVA) with fixed effects since we were interested in the consistency between these two methods (Streiner and Norman, 1989). For comparisons between operators (inter-rater reliability) and between occasions (intra-rater reliability), we used an ICC based on a two-way ANOVA with random effects, since we were interested in how these occasions or operators were representative of the larger spectrum of occasions or operators (Streiner and Norman, 1989).

Mean distance maps

In order to assess the anatomical consistency of the regions, within subject comparisons (manual vs. automated, or between occasions or operators processing the automated procedure) were made by overlaying all of the regions on the inflated surface and determining the mismatch error present between the individual regions when the two labeling systems (manual vs. automated) were compared.

First, two distance maps, $D_{i,1}$ and $D_{i,2}$ (one per labeling system) were generated for each subject i in the dataset. For each vertex v with label $l(v)$ in a specific labeling system, the geodesic distance $d(v)$ to the label border was computed on the inflated surface S : $d(v) = \min_{x \in S(l(x) \neq l(v))} g(x, v)$, where $g(x, v)$ represents the geodesic distance between vertex x and vertex v on the inflated surface S . Next, individual mean distance maps D_i were generated as the average of the previous distance maps: $D_i = \frac{D_{i,1} + D_{i,2}}{2}$. In addition to individual distance maps, an average distance map D was computed as the average of all individual maps $D = \frac{1}{n} \sum_i D_i$ where n is the number of subjects in the dataset.

Essentially, mean distance maps provide a point-for-point estimate of the degree of mismatch (the mean distance of error) between labeling systems. Furthermore, unlike intraclass correlation coefficients which assess only the correspondence between the volumetric measurements, mean distance maps allow for the visualization of the boundary areas of mismatch (e.g., the sulcal boundaries) within a region of interest when labeling systems are compared. Such differences can be displayed in color (see Fig. 2, where a discrepancy of 0.5 mm is shown in red and a discrepancy of 1.0 mm is portrayed in yellow). Finally, when comparing labeling systems, mean distance maps may provide a more meaningful estimate of error than intraclass correlation coefficients by showing the actual spatial distribution of error and thus allowing for the assessment of which errors are meaningful. As is presented in Fig. 2, the mismatch is almost entirely at the boundaries of the regions, and the errors are less than 1 mm. It is likely that the boundaries cannot be localized with any greater precision and thus these may not be 'errors' in the true sense of the word. (Fischl et al., 2004).

Jackknifing/Leave-one-out reliability

In order to assess the reliability of the automated atlas when applied to novel datasets, we employed a statistical technique termed jackknifing (Efron, 1982; Efron and Tibshirani, 1993). Using this method, from the original set of 40 subjects, the manual regions of interest from 39 subjects were used as a training set to construct an automated atlas thus leaving out one dataset whose ROIs were not included as part of the training set. The atlas from the 39 subjects was then applied to the left-out dataset and the mismatch error between the automated atlas (built from the training set of 39 subjects) and the manual regions for the left-out dataset was quantified point-for-point across the entire cortical surface. This process was then repeated until all 40 subjects had been left-out once. The average mismatch error resulting from this jackknifing procedure was compared with

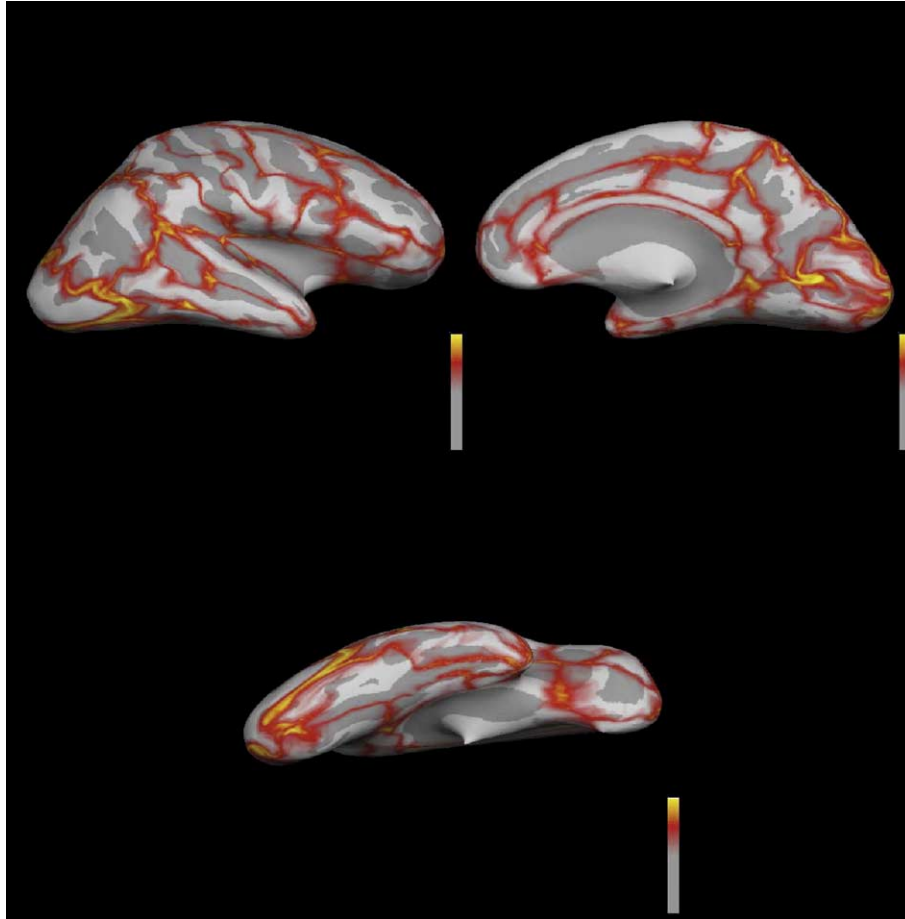


Fig. 2. Inflated cortical representation (lateral, medial and ventral views) displaying mean distance error maps between the manual and the automated regions of interest for all 40 subjects. The dark gray overlay represents sulcal cortical regions and the light gray overlay represents gyral cortical regions. The colorscale represents a maximum value for the mismatch between the two sets of images as follows: 0 mm (gray), 0.5 mm (red), 1 mm (yellow). Note that the mismatch between the image sets appears to lie entirely along the borders of the regions of interest.

the average mismatch error from the original automated versus manual comparison to determine the equivalence of applying the automated atlas to a dataset from which the atlas was constructed and to a novel dataset that did not contribute to the construction of the automated atlas (determined from the jackknife versus manual comparison).

Results

Manual versus automated analysis (validity)

Table 1 lists the intraclass correlation coefficients for 32 of the 34 labels in each hemisphere. One region, the corpus callosum, was excluded a priori, since it is a white matter structure and, as noted above, was only included in order to better define the regions around it. A second region, the frontal pole, was excluded after examination of the preliminary analyses. This region proved to be unreliable (average ICC 0.26), in all likelihood because it was defined as that region in the most anterior portion of the brain that remained when all other regions near it were outlined.

The average ICC for the comparison of the 32 manual and automated regions in each hemisphere was 0.835, with values ranging from a low of 0.623 for the left banks of the superior

temporal sulcus and right pericalcarine cortex to a high of 0.977 for the right superior frontal gyrus (see Table 2). Sixty of the sixty-four labels had intraclass correlation coefficients higher than 0.700. The regions with smaller surface areas (and thus smaller volumes) resulted in smaller ICC values (c.f., left temporal pole and left superior frontal gyrus) when the automated and manual regions were compared. This was not surprising as a label with an area of A on the inflated surface is of approximate size $r = \sqrt{\frac{A}{\pi}}$. Since our labeling system is accurate of the order of the mm, we can roughly estimate the expected labeling error C in function of the size r by: $C(r) = \frac{A(r-1)}{A(r)} = \frac{(r-1)^2}{r^2} = 1 - \frac{2}{r} + \frac{1}{r^2}$ or in function of the area A by: $C(A) = 1 - 2\sqrt{\frac{\pi}{A}} + \frac{\pi}{A}$. For small areas of approximate size $r = 6$ mm, such as the temporal pole, this leads to an estimate of $C(r) = 25/36 \approx 69\%$. This expected labeling error partly explains the smaller ICC values found for smaller labels. For larger areas with $r = 30$ mm, such as the superior frontal gyrus, we find that $C(r) = 841/900 \approx 93\%$. This seems to imply that our method, which uses curvature information alone (through a non-stationary anisotropic Markov random field) in order to locate each region, might not be accurate enough when small regions are targeted. Integration of additional information (for example, incorporating the relative location of subcortical structures to better locate cortical labels) should be able to alleviate this limitation and is of significant interest for us as future work to make the smaller regions more reliable.

Table 1
Intraclass correlations (two-way ANOVA with fixed effects)—manual versus automated comparison for the volumes of each region of interest

Region	Left hem	Right hem
Banks superior temporal sulcus	0.623	0.733
Caudal anterior-cingulate cortex	0.768	0.809
Caudal middle frontal gyrus	0.897	0.907
Cuneus cortex	0.767	0.797
Entorhinal cortex	0.818	0.737
Fusiform gyrus	0.812	0.880
Inferior parietal cortex	0.935	0.960
Inferior temporal gyrus	0.862	0.870
Isthmus–cingulate cortex	0.729	0.729
Lateral occipital cortex	0.873	0.900
Lateral orbital frontal cortex	0.865	0.814
Lingual gyrus	0.854	0.923
Medial orbital frontal cortex	0.834	0.907
Middle temporal gyrus	0.891	0.892
Parahippocampal gyrus	0.857	0.804
Paracentral lobule	0.839	0.861
Pars opercularis	0.817	0.792
Pars orbitalis	0.664	0.729
Pars triangularis	0.745	0.819
Pericalcarine cortex	0.732	0.623
Postcentral gyrus	0.916	0.880
Posterior-cingulate cortex	0.833	0.812
Precentral gyrus	0.967	0.972
Precuneus cortex	0.839	0.945
Rostral anterior cingulate cortex	0.811	0.835
Rostral middle frontal gyrus	0.878	0.908
Superior frontal gyrus	0.965	0.977
Superior parietal cortex	0.912	0.856
Superior temporal gyrus	0.921	0.944
Supramarginal gyrus	0.915	0.894
Temporal pole	0.649	0.729
Transverse temporal cortex	0.712	0.719

Mean distance error maps for the labels on the medial, lateral and ventral surface revealed that the mismatch was almost entirely along the boundaries between the structures and on the magnitude of 1 mm when the manual and automated labeling schemes were compared (see Fig. 2). Furthermore, the mismatch error derived from the jackknifing/leave-one-out procedure was used to assess the applicability of the automated atlas. The average mismatch error from the jackknife procedure revealed no difference with the average mismatch error from the automated scheme (less than 0.01%) thus lending support to the fact that the automated labeling scheme is highly reliable when applied to new brains that are not part of the original training set.

Intra- and inter-rater comparisons (reliability)

Intra-rater reliability was measured by having one operator (RSD) process all of the cases twice, blind to both the clinical status and the identity of each subject. Table 2 lists the intraclass correlation coefficients for the same 32 labels in each hemisphere that were provided in Table 1. The average ICC for intra-rater reliability was 0.998, ranging from a low of 0.993 to a high of 0.999 (see Table 2). Mean distance error maps for the labels on the medial, lateral and ventral surface revealed that the variability seen was almost entirely the boundaries between the structures and on the magnitude of less than 1 mm (see Fig. 3).

Inter-rater reliability was measured by having a second operator (RJK) independently process all of the cases also blind to both the clinical status and identities of subject. Table 2 lists the intraclass correlation coefficients for the same 32 labels in each hemisphere that were provided in Table 1. The average inter-rater reliability was 0.998, with values ranging from a low of 0.993 to a high of 0.999 (see Table 2). Mean distance error maps for the labels on the medial, lateral and ventral surface revealed that the variability was all along the boundaries between the structures and on the magnitude of less than 1 mm (see Fig. 4).

Discussion

Automated systems for labeling cortical structures provide an efficient way to undertake a complex and otherwise labor-intensive process, provided the accuracy of these methods are

Table 2
Intraclass correlations (two-way ANOVA with random effects)—intra- and inter-rater comparisons of the volumes for each region of interest

Region	Intra-rater		Inter-rater	
	Left hem	Right hem	Left hem	Right hem
Banks superior temporal sulcus	0.997	0.998	0.996	0.998
Caudal anterior cingulate cortex	0.998	0.999	0.998	0.999
Caudal middle frontal gyrus	0.999	0.999	0.999	0.999
Cuneus cortex	0.999	0.999	0.999	0.998
Entorhinal cortex	0.993	0.994	0.995	0.989
Fusiform gyrus	0.999	0.999	0.999	0.999
Inferior parietal cortex	0.999	0.999	0.999	0.999
Inferior temporal gyrus	0.999	0.999	0.999	0.999
Isthmus-cingulate cortex	0.997	0.997	0.998	0.998
Lateral occipital cortex	0.999	0.999	0.999	0.999
Lateral orbital frontal cortex	0.999	0.999	0.999	0.998
Lingual gyrus	0.999	0.999	0.999	0.999
Medial orbital frontal cortex	0.999	0.999	0.999	0.999
Middle temporal gyrus	0.999	0.999	0.999	0.999
Parahippocampal gyrus	0.998	0.995	0.999	0.993
Paracentral lobule	0.999	0.999	0.999	0.999
Pars opercularis	0.999	0.998	0.999	0.998
Pars orbitalis	0.997	0.998	0.999	0.998
Pars triangularis	0.998	0.999	0.998	0.999
Pericalcarine cortex	0.999	0.999	0.999	0.999
Postcentral gyrus	0.999	0.999	0.999	0.999
Posterior-cingulate cortex	0.998	0.997	0.998	0.998
Precentral gyrus	0.999	0.999	0.999	0.999
Precuneus cortex	0.999	0.999	0.999	0.999
Rostral anterior cingulate cortex	0.999	0.998	0.999	0.998
Rostral middle frontal gyrus	0.999	0.997	0.999	0.999
Superior frontal gyrus	0.999	0.999	0.999	0.999
Superior parietal cortex	0.999	0.999	0.999	0.999
Superior temporal gyrus	0.999	0.999	0.999	0.999
Supramarginal gyrus	0.999	0.999	0.999	0.999
Temporal pole	0.996	0.998	0.996	0.998
Transverse temporal cortex	0.998	0.997	0.996	0.995

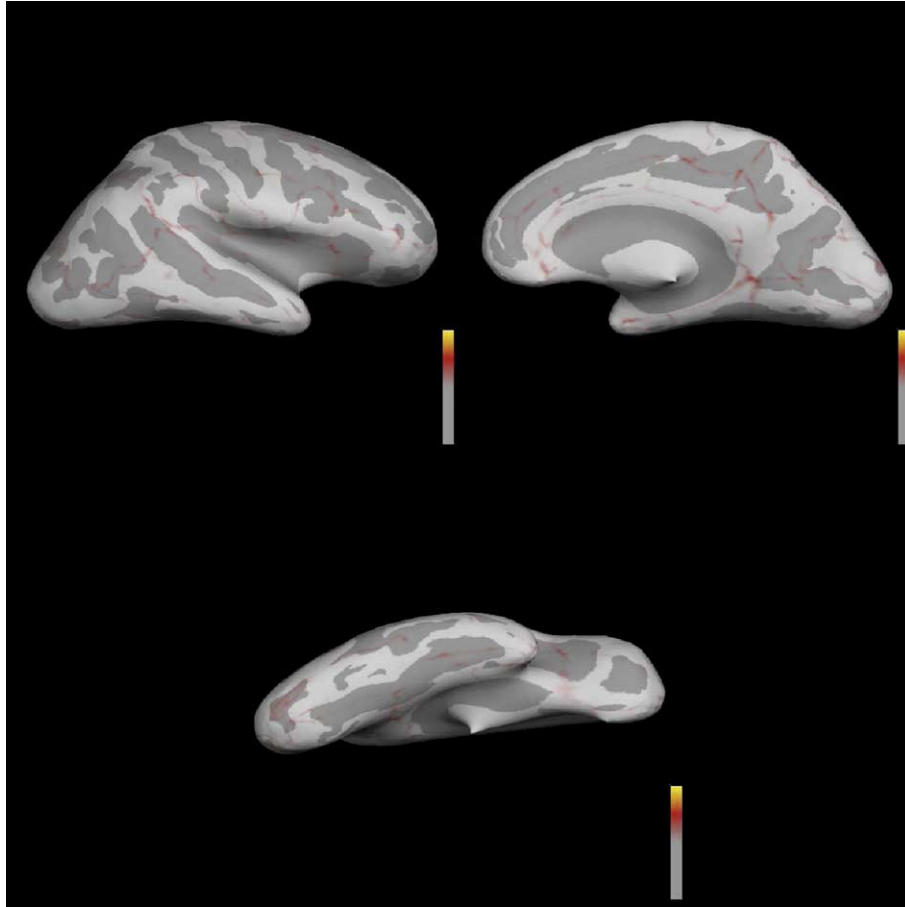


Fig. 3. Inflated cortical representation (lateral, medial and ventral views) displaying mean distance error maps between two sets of automated labels generated by the same operator for all 40 subjects (intra-rater). The dark gray overlay represents sulcal cortical regions and the light gray overlay represents gyral cortical regions. The colorscale represents a maximum value for the mismatch between the two sets of images as follows: 0 mm (gray), 0.5 mm (red), 1 mm (yellow). Note the presence of small mismatch between the image sets; this mismatch appears to lie entirely along the borders of the regions of interest.

sufficient. In the work presented here our motivation was to develop an automated anatomic labeling system from a range of subjects to better account for cortical inter-subject variability that could then be applied to a variety of subjects in studies involving the cerebral cortex. The current findings suggest that the automated system for subdividing the cortex into regions of interest described here is valid when compared to manual procedures and has a very high degree of reliability.

The anatomic labeling scheme presented here is comparable to several other approaches that have been utilized to parcellate the human brain into neuroanatomical regions of interest. Using a volumetric labeling technique, Tzourio-Mazoyer et al. (2002) manually delineated regions of interest for a single subject existing in common stereotaxic space in order to provide a more robust anatomic basis for functional activation studies, not for determining the absolute anatomic localization of brain structures that is required for morphometric studies. Other examples of parcellation techniques involve the spatial transformation to a common stereotaxic space of a fixed set of manual labels in order to compute probabilistic maps of the anatomic regions (which encode information regarding inter-subject anatomic variability) that can then be used to further label new datasets (Hammers et al., 2003; Mega et al., 2005). This work is similar to the one presented here in that the manual labeling was performed on a range of subjects and proposes probabilistic

approaches to potentially label new brains but differs from our work in the usage of whole-brain, anatomic labels and volumetric averaging approaches (Hammers et al., 2003; Mega et al., 2005) instead of surface-based techniques developed specifically for the cerebral cortex. Surface-based approaches similar to the one presented here have also been employed to the problem of labeling cortical features including the modelling of specific cortical regions to examine patterns of gray matter distribution across various subject groups (Thompson et al., 2001), the extraction of cortical sulci using a watershed based approach (Rettmann et al., 2002), and the probabilistic determination of sulcal patterns using manually labelled datasets (Van Essen, 2005). Though these approaches isolate certain cortical features, a detailed neuroanatomical labeling scheme along with its incorporation into an automated algorithm for probabilistically labeling future datasets is not presented.

Prior methods defining anatomic regions and labeling the cortex have typically used Pearson's correlations and/or intraclass correlation coefficients to judge the validity of the techniques, using manually drawn regions of interest as the standard (Mega et al., 2005) and assess reliability between operators (Crespo-Facorro et al., 2000; Goncharova et al., 2001; Buckner et al., 2004). We used these standard procedures to assess validity of the techniques described here and found an average ICC for all structures of 0.835 when the automated and manual labeling schemes were

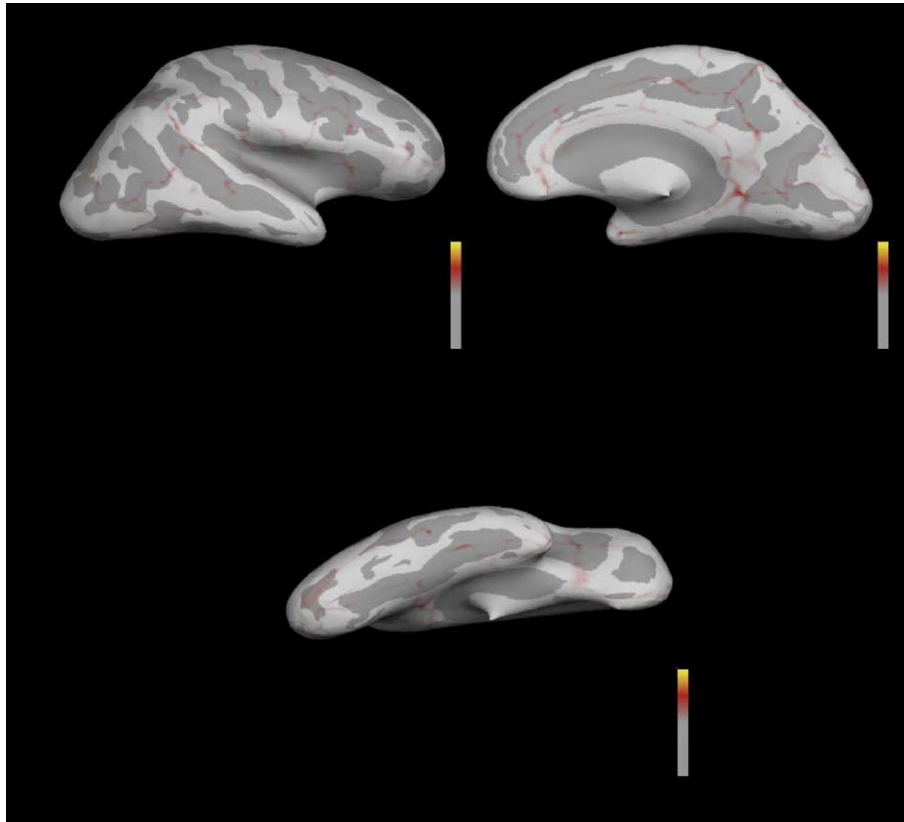


Fig. 4. Inflated cortical representation (lateral, medial and ventral views) displaying mean distance error maps between two sets of automated labels generated by the two separate operators for all 40 subjects (inter-rater). The dark gray overlay represents sulcal cortical regions and the light gray overlay represents gyral cortical regions. The colorscale represents a maximum value for the mismatch between the two sets of images as follows: 0 mm (gray), 0.5 mm (red), 1 mm (yellow). Note the presence of small mismatch between the image sets; this mismatch appears to lie entirely along the borders of the regions of interest.

compared. In general, the regions with smaller surface areas (and thus smaller volumes) yielded lower ICC values when compared to regions with larger surface areas (c.f., left temporal pole and left superior frontal gyrus). This seems to imply that for these smaller structures curvature information alone may not be sufficient for the automated algorithm to accurately locate the regions of interest. Additional information (such as information regarding the relative location of subcortical structures such as the hippocampus and amygdala to cortical structures such as the temporal pole and entorhinal cortex) may need to be incorporated into the automated labeling scheme to better determine the location of these smaller regions and as such is of significant interest to us as future work to aid in making the smaller structures more reliable. Additionally, regions such as the banks of the superior temporal sulcus and the pericalcarine cortex have larger inter-individual variability due to being purely sulcal in nature thus explaining their lower ICC values. Finally, we used intraclass correlation coefficients to assess the reliability of the automated system by intra- and inter-rater measurements and found an average ICC for all structures of 0.998.

The intraclass correlations indicate that there is an overall high level of validity and reliability for this automated cortical labeling system. However, these types of assessments evaluate the similarity of the volumes generated for each region of interest rather than the accuracy of the specific location of the anatomic regions. This opens up the possibility that the volumes could be similar, at the same time that the regions of interest could lack congruence. Thus, we developed mean distance maps (i.e., mean distance, in mm, of the

mismatches), in order to get a better estimate of the degree of anatomical mismatch between data sets.

These mean distance maps demonstrated that the variability between the manual and automated regions of interest was along the boundaries of the structures and of a magnitude of 1 mm or less (which is approximately at the limit of the 1 mm isotropic resolution of the resliced images). As can be seen in Fig. 2, there appeared to be little mismatch within the main body of any of the structures (i.e., the red and yellow regions are all at the boundaries of the ROIs). This strongly suggests that the automated system subdivides the cortex equivalently to a manual operator. To assess the applicability of the automated labeling procedure, we utilized a jackknife/leave-one-out technique (Efron, 1982; Efron and Tibshirani, 1993) to apply the atlas onto a new dataset that was not part of the training set from which the atlas was developed. One of the strengths of this technique is that it provides an estimate of the reliability of the automated procedure when applied to datasets with the same general composition as the one employed here (Efron, 1982; Efron and Tibshirani, 1993). The average mismatch error from this jackknife procedure was equivalent to the average mismatch error from the automated procedure thus lending further support to the reliability of our labeling scheme when applied to novel datasets. As part of our ongoing research studies on aging and AD, we are currently in the process of applying this automated labeling scheme to images acquired from a different scanner using differing parameters than the ones presented here and will be presenting this work in future publications. Finally, as can be seen in Figs. 3 and 4, when applied to

assess intra- and inter-rater reliability, we found minimal variability between the sets of regions. This should not be surprising, since one of the strengths of automated systems is the ability to provide consistent measures.

The consistency of these measures is of particular importance for clinical trials and other longitudinal studies, which depend on comparisons of scans over time. Recently, we have completed a study examining the same set of subjects scanned at multiple intervals both across various scanner types (GE and Siemens) and across multiple field strengths (1.5 and 3.0 T). Our laboratory is in the process of using morphometric tools to quantify the overall measurement variability of MRI in these subjects (Quinn et al., 2005; Han et al., submitted for publication).

In addition to being highly consistent and automated, one of the strengths of the current labeling scheme is its development from a range of subjects including young, middle-aged and elderly controls as well as patients with AD. Our hypothesis was that by using a training set of brains from a range of subjects, we would better account for inter-subject anatomic variability and thus allow the automated algorithm to define the cortical regions of interest with improved accuracy. To test this hypothesis, we: (1) constructed an automated atlas from manually labelled cortical regions of interest from our subset of 10 elderly controls and applied this atlas to our subset of 10 patients with AD, (2) constructed an automated atlas from manually labelled cortical regions of interest from our subset of 10 AD patients and applied this atlas to our subset of 10 elderly controls, and (3) compared the average intraclass correlation coefficient (between the automated and manual regions of interest) derived from both of the above mentioned atlases with the average intraclass correlation coefficient derived from the previously constructed atlas of 40 subjects. The results from this study revealed both the elderly and AD atlases to have smaller average ICC values (elderly atlas average ICC = 0.748, AD atlas average ICC value = 0.726) when compared with the atlas constructed from the 40 subjects (average ICC value = 0.835). This suggests that an atlas constructed with a narrower range of anatomic variability (i.e., all elderly, all males, all females, etc.) would be less accurate than one that incorporated a wider range of anatomic variability. Though it is hypothetically possible to construct an atlas from a homogenous set of cases and apply this atlas onto a larger cohort of similar cases (e.g., build an atlas from young females to apply to only young females), it has limited practical value for large morphometric studies aiming to assess anatomic changes across different population types.

In summary, we have presented an automated system for subdividing the human cerebral cortex into standard gyral-based neuroanatomical regions, and demonstrated this procedure to be both anatomically valid and highly reliable. This type of detailed neuroanatomic information can be used for co-registration of MRI scans with other types of in vivo imaging studies (e.g., fMRI, PET, SPECT) and may provide a valuable tool for research studies involving the cerebral cortex as well as clinical trials in which MRI measures are used to examine response to treatment and/or to track the evolution of disease over time.

Acknowledgments

This work was supported by National Institute of Aging Grant P01-AG04953, the National Center for Research Resources (P41-RR14075, R01 RR 16594-01A1) and the NCCR BIRN Morpho-

metric Project BIRN002, U24 RR021382), the National Institute for Biomedical Imaging and Bioengineering (R01 EB001550), the Mental Illness and Neuroscience Discovery (MIND) Institute, and Pfizer Incorporated. The authors would like to thank Drs. Thomas Kemper and Douglas Rosene for their insightful discussions concerning the anatomical boundaries for the various regions of interest, Dr. Howard Cabral for helpful discussions regarding statistical re-sampling methods, and Dr. David Salat for assistance with initial data transfer and for valuable discussions regarding the manuscript.

References

- Bakshi, R., Minagar, A., Jaisani, Z., Wolinsky, J.S., 2005. Imaging of multiple sclerosis: role in neurotherapeutics. *NeuroRx* 2, 277–303.
- Ballmaier, M., Kumar, A., Thompson, P.M., Narr, K.L., Lavretsky, H., Estanol, L., Deluca, H., Toga, A.W., 2004. Localizing gray matter deficits in late-onset depression using computational cortical pattern matching methods. *Am. J. Psychiatry* 161, 2091–2099.
- Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage* 23, 724–738.
- Crespo-Facorro, B., Kim, J., Andreasen, N.C., Spinks, R., O’Leary, D.S., Bockholt, H.J., Harris, G., Magnotta, V.A., 2000. Cerebral cortex: a topographic segmentation method using magnetic resonance imaging. *Psychiatry Res.* 100, 97–126.
- Csernansky, J.G., Wang, L., Joshi, S., Miller, J.P., Gado, M., Kido, D., McKeel, D., Morris, J.C., Miller, M.I., 2000. Early DAT is distinguished from aging by high-dimensional mapping of the hippocampus. *Dementia of the Alzheimer type. Neurology* 55, 1636–1643.
- Csernansky, J.G., Wang, L., Swank, J., Miller, J.P., Gado, M., McKeel, D., Miller, M.I., Morris, J.C., 2005. Preclinical detection of Alzheimer’s disease: hippocampal shape and volume predict dementia onset in the elderly. *NeuroImage* 25, 783–792.
- Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J. Cogn. Neurosci.* 5, 162–176.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage* 9, 179–194.
- Damasio, H., Damasio, A.R., 1989. *Lesion Analysis in Neuropsychology*. Oxford Univ. Press, Oxford.
- Destrieux, C., Halgren, E., Dale, A.M., Fischl, B., Sereno, M.I., 1998. Variability of the human brain studied on the flattened cortical surface. *Abstr.-Soc. Neurosci.* 24, 1164.
- Duvernoy, H., 1991. *The Human Brain*. Springer-Verlag, Vienna.
- Efron, B., 1982. *The Jackknife, The Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to The Bootstrap*. Chapman and Hall, London.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11050–11055.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9, 195–207.
- Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284.
- Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery:

- constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* 20, 70–80.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.
- Fotinos, A.F., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L., 2005. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology* 64, 1032–1039.
- Fox, N.C., Warrington, E.K., Rossor, M.N., 1999. Serial magnetic resonance imaging of cerebral atrophy in preclinical Alzheimer's disease. *Lancet* 353, 2125.
- Fox, N.C., Crum, W.R., Scahill, R.I., Stevens, J.M., Janssen, J.C., Rossor, M.N., 2001. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet* 358, 201–205.
- Fox, N.C., Black, R.S., Gilman, S., Rossor, M.N., Griffith, S.G., Jenkins, L., Koller, M., AN1792(QS-21)-201 Study, 2005. Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology* 64, 1563–1572.
- Freeborough, P.A., Fox, N.C., 1998a. MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease. *IEEE Trans. Med. Imaging* 17, 475–479.
- Freeborough, P.A., Fox, N.C., 1998b. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *J. Comput. Assist. Tomogr.* 22, 838–843.
- Goncharova, I.I., Dickerson, B.C., Stoub, T.R., deToledo-Morrell, L., 2001. MRI of human entorhinal cortex: a reliable protocol for volumetric measurement. *Neurobiol. Aging* 22, 737–745.
- Halliday, G.M., Double, K.L., Macdonald, V., Kril, J.J., 2003. Identifying severely atrophic cortical subregions in Alzheimer's disease. *Neurobiol. Aging* 24, 797–806.
- Hammers, A., Allom, R., Koeppe, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* 19, 224–247.
- Han, X., Jovicich, J., Salat, D., Van Der Kouwe A., Quinn, B., Czanner, S., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Fischl, B., submitted for publication. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *NeuroImage*.
- Head, D., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L., 2005. Frontal-hippocampal double dissociation between normal aging and Alzheimer's disease. *Cereb. Cortex* 15, 732–739.
- Hsu, Y.Y., Schuff, N., Du, A.T., Mark, K., Zhu, X., Hardin, D., Weiner, M.W., 2002. Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. *J. Magn. Reson. Imaging* 16, 305–310.
- Kennedy, D.N., Lange, N., Makris, N., Bates, J., Meyer, J., Caviness, V.S.J., 1998. Gyri of the human neocortex: an MRI-based analysis of volume and variance. *Cereb. Cortex* 8, 372–384.
- Killiany, R.J., Moss, M.B., Albert, M.S., Sandor, T., Tieman, J., Jolesz, F., 1993. Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. *Arch. Neurol.* 50, 949–954.
- Killiany, R.J., Gomez-Isla, T., Moss, M., Kikinis, R., Sandor, T., Jolesz, F., Tanzi, R., Jones, K., Hyman, B.T., Albert, M.S., 2000. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. *Ann. Neurol.* 47, 430–439.
- Jack, C.R.J., Theodore, W.H., Cook, M., McCarthy, G., 1995. MRI-based hippocampal volumetrics: data acquisition, normal ranges, and optimal protocol. *Magn. Reson. Imaging* 13, 1057–1064.
- Jack, C.R.J., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., Kokmen, E., 1997. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology* 49, 786–794.
- Jack, C.R.J., Slomkowski, M., Gracon, S., Hoover, T.M., Felmlee, J.P., Stewart, K., Xu, Y., Shiung, M., O'Brien, P.C., Cha, R., Knopman, D., Petersen, R.C., 2003. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. *Neurology* 60, 253–260.
- Le Goualher, G., Procyk, E., Collins, D.L., Venugopal, R., Barillot, C., Evans, A.C., 1999. Automated extraction and variability analysis of sulcal neuroanatomy. *IEEE Trans. Med. Imaging* 18, 206–217.
- Lohmann, G., 1998. Extracting line representations of sulcal and gyral patterns in MR images of the human brain. *IEEE Trans. Med. Imaging* 17, 1040–1048.
- Mangin, J.F., Froin, V., Bloch, J., Regis, J., Lopez-Krahe, J., 1995. From 3-D magnetic resonance images to structured representations of the cortex topography using topology preserving deformations. *J. Math. Imaging Vis.* 5, 297–318.
- Mega, M.S., Dinov, I.D., Mazziotta, J.C., Manese, M., Thompson, P.M., Lindshield, C., Moussai, J., Tran, N., Olsen, K., Zoumalan, C.I., Woods, R.P., Toga, A.W., 2005. Automated brain tissue assessment in the elderly and demented population: construction and validation of a sub-volume probabilistic brain atlas. *NeuroImage* 15, 1009–1018.
- Miller, M.I., Hosakere, M., Barker, A.R., Priebe, C.E., Lee, N., Ratnathar, J.T., Wang, L., Gado, M., Morris, J.C., Csernansky, J.G., 2003. Labeled cortical mantle distance maps of the cingulate quantify differences between dementia of the Alzheimer type and healthy aging. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15172–15177.
- Morris, J.C., 1993. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 43, 2412–2414.
- Onitsuka, T., Shenton, M.E., Salisbury, D.F., Dickey, C.C., Kasai, K., Toner, S.K., Frumin, M., Kikinis, R., Jolesz, F.A., McCarley, R.W., 2004. Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an MRI study. *Am. J. Psychiatry* 161, 1603–16011.
- Ono, M., Kubik, S., Abernathy, C.D., 1990. *Atlas of The Cerebral Sulci*. Thieme Medical Publishers, Inc., New York.
- Quinn, B., Fenstermacher, E., Han, X., Pacheco, J., Czanner, S., Van Der Kouwe, A., Maguire, P., Raunig, D., Albert, M., Makris, N., Desikan, R., Killiany, R., Dickerson, B., Fischl, B., 2005. Test–retest reliability assessment for longitudinal MRI studies: a comparison of the effects of different T1-weighted protocols, scanner platforms, and field strengths on semi-automated hippocampal volume measures. 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto, 2005 (poster presentation).
- Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., Caviness, V.S.J., 1992. Human cerebral cortex: localization, parcellation and morphometry with magnetic resonance imaging. *J. Cogn. Neurosci.* 4, 352–374.
- Raz, N., 2004. The aging brain: Structural changes and their implications for cognitive aging. In: Dixon, R., Bäckman, L., Nilsson, L.-G. (Eds.), *New Frontiers in Cognitive Aging*. Oxford Univ. Press, pp. 115–134.
- Rettmann, M.E., Han, X., Xu, C., Prince, J.L., 2002. Automated sulcal segmentation using watersheds on the cortical surface. *NeuroImage* 15, 329–344.
- Sandor, S., Leahy, R., 1997. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Trans. Med. Imaging* 16, 41–54.
- Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. *NeuroImage* 22, 1060–1075.
- Ségonne, F., Grimson, E., Fischl, B., 2005. A genetic algorithm for the topology correction of cortical surfaces. *IPMI-LNCS* 3964, 393–405.
- Streiner, D.L., Norman, G.R., 1989. *Health Measurement Scales A Practical Guide To Their Development and Use*. Oxford Univ. Press, Oxford, UK.
- Thompson, P.M., Schwartz, C., Toga, A.W., 1996. High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain. *NeuroImage* 3, 19–34.
- Thompson, P.M., Mega, M.S., Woods, R.P., Zoumalan, C.I., Lindshield,

- C.J., Blanton, R.E., Moussai, J., Holmes, C.J., Cummings, J.L., Toga, A.W., 2001. Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb. Cortex* 11, 1–16.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289.
- Van Essen, D.C., 2005. A PopulationAverage, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *NeuroImage* 28, 635–662.
- Van Hoesen, G.W., Parvizi, J., Chu, C.C., 2000. Orbitofrontal cortex pathology in Alzheimer's disease. *Cereb. Cortex* 10, 243–251.
- Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernansky, J.G., 2003. Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. *NeuroImage* 20, 667–682.
- Wible, C.G., Shenton, M.E., Hokama, H., Kikinis, R., Jolesz, F.A., Metcalf, D., McCarley, R.W., 1995. Prefrontal cortex and schizophrenia. A quantitative magnetic resonance imaging study. *Arch. Gen. Psychiatry* 52, 279–288.
- Wible, C.G., Shenton, M.E., Fischer, I.A., Allard, J.E., Kikinis, R., Jolesz, F.A., Iosifescu, D.V., McCarley, R.W., 1997. Parcellation of the human prefrontal cortex using MRI. *Psychiatry Res.* 28, 29–40.
- Yamasue, H., Iwanami, A., Hirayasu, Y., Yamada, H., Abe, O., Kuroki, N., Fukuda, R., Tsujii, K., Aoki, S., Ohtomo, K., Kato, N., Kasai, K., 2004. Localized volume reduction in prefrontal, temporo limbic, and paralimbic regions in schizophrenia: an MRI parcellation study. *Psychiatry Res.* 15, 195–207.
- Zilles, K., Armstrong, E., Schleicher, A., Kretschmann, H.J., 1988. The human pattern of gyrification in the cerebral cortex. *Anat. Embryol. (Berl.)* 179, 173–179.