



# Automated Spatial Brain Normalization and Hindbrain White Matter Reference Tissue Give Improved [<sup>18</sup>F]-Florbetaben PET Quantitation in Alzheimer's Model Mice

Felix Overhoff<sup>1†</sup>, Matthias Brendel<sup>1†</sup>, Anna Jaworska<sup>2,3</sup>, Viktoria Korzhova<sup>2</sup>, Andreas Delker<sup>1</sup>, Federico Probst<sup>1</sup>, Carola Focke<sup>1</sup>, Franz-Josef Gildehaus<sup>1</sup>, Janette Carlsen<sup>1</sup>, Karlheinz Baumann<sup>4</sup>, Christian Haass<sup>2,5,6</sup>, Peter Bartenstein<sup>1,5</sup>, Jochen Herms<sup>2,5</sup> and Axel Rominger<sup>1,5\*</sup>

## OPEN ACCESS

### Edited by:

John Ashburner,  
UCL Institute of Neurology, UK

### Reviewed by:

Hidenao Fukuyama,  
Kyoto University, Japan  
Stephen John Sawiak,  
University of Cambridge, UK

### \*Correspondence:

Axel Rominger  
axel.rominger@med.uni-muenchen.de

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 10 November 2015

**Accepted:** 01 February 2016

**Published:** 29 February 2016

### Citation:

Overhoff F, Brendel M, Jaworska A, Korzhova V, Delker A, Probst F, Focke C, Gildehaus F-J, Carlsen J, Baumann K, Haass C, Bartenstein P, Herms J and Rominger A (2016) Automated Spatial Brain Normalization and Hindbrain White Matter Reference Tissue Give Improved [<sup>18</sup>F]-Florbetaben PET Quantitation in Alzheimer's Model Mice. *Front. Neurosci.* 10:45. doi: 10.3389/fnins.2016.00045

<sup>1</sup> Department of Nuclear Medicine, Ludwig-Maximilians-University of Munich, Munich, Germany, <sup>2</sup> DZNE-German Center for Neurodegenerative Diseases, Munich, Germany, <sup>3</sup> Laboratory of Neurodegeneration, International Institute of Molecular and Cell Biology, Warsaw, Poland, <sup>4</sup> Roche Pharma Research and Early Development, Neuroscience Discovery, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland, <sup>5</sup> Munich Cluster for Systems Neurology (SyNergy), Munich, Germany, <sup>6</sup> Biomedical Center, Ludwig-Maximilians-University of Munich, Munich, Germany

Preclinical PET studies of  $\beta$ -amyloid (A $\beta$ ) accumulation are of growing importance, but comparisons between research sites require standardized and optimized methods for quantitation. Therefore, we aimed to evaluate systematically the (1) impact of an automated algorithm for spatial brain normalization, and (2) intensity scaling methods of different reference regions for A $\beta$ -PET in a large dataset of transgenic mice. PS2APP mice in a 6 week longitudinal setting ( $N = 37$ ) and another set of PS2APP mice at a histologically assessed narrow range of A $\beta$  burden ( $N = 40$ ) were investigated by [<sup>18</sup>F]-florbetaben PET. Manual spatial normalization by three readers at different training levels was performed prior to application of an automated brain spatial normalization and inter-reader agreement was assessed by Fleiss Kappa ( $\kappa$ ). For this method the impact of templates at different pathology stages was investigated. Four different reference regions on brain uptake normalization were used to calculate frontal cortical standardized uptake value ratios ( $SUV_{CTX/REF}$ ), relative to raw  $SUV_{CTX}$ . Results were compared on the basis of longitudinal stability (Cohen's  $d$ ), and in reference to gold standard histopathological quantitation (Pearson's  $R$ ). Application of an automated brain spatial normalization resulted in nearly perfect agreement (all  $\kappa \geq 0.99$ ) between different readers, with constant or improved correlation with histology. Templates based on inappropriate pathology stage resulted in up to 2.9% systematic bias for  $SUV_{CTX/REF}$ . All  $SUV_{CTX/REF}$  methods performed better than  $SUV_{CTX}$  both with regard to longitudinal stability ( $d \geq 1.21$  vs.  $d = 0.23$ ) and histological gold standard agreement ( $R \geq 0.66$  vs.  $R \geq 0.31$ ). Voxel-wise analysis suggested a physiologically implausible longitudinal decrease by global mean scaling. The hindbrain white matter reference ( $R_{mean} = 0.75$ )

was slightly superior to the brainstem ( $R_{\text{mean}} = 0.74$ ) and the cerebellum ( $R_{\text{mean}} = 0.73$ ). Automated brain normalization with reference region templates presents an excellent method to avoid the inter-reader variability in preclinical A $\beta$ -PET scans. Intracerebral reference regions lacking A $\beta$  pathology serve for precise longitudinal *in vivo* quantification of [ $^{18}\text{F}$ ]-florbetaben PET. Hindbrain white matter reference performed best when considering the composite of quality criteria.

**Keywords:** Alzheimer's disease,  $\beta$ -amyloid, [ $^{18}\text{F}$ ]-florbetaben, small animal PET, reference region, brain normalization

## INTRODUCTION

The steadily growing number of patients suffering from Alzheimer's disease (AD) will place a great burden on healthcare systems in the coming decades, barring development of an effective intervention therapy (Schneider, 2013). Molecular imaging of  $\beta$ -amyloid (A $\beta$ ) with positron emission tomography (PET) has given new insight into the progression of AD pathology and has entered clinical diagnostic use (Johnson et al., 2013). Furthermore, PET imaging is increasingly used for detecting cerebral amyloidosis in transgenic mouse models of AD (Manook et al., 2012; Rominger et al., 2013). Small animal PET studies of longitudinal design afford monitoring of the rate of  $\beta$ -amyloid accumulation, and present the possibility of testing interventions for attenuating plaque formation.

Preclinical brain PET imaging frequently suffer from shortcomings such as underpowered study groups, reader dependence of endpoints, varying approaches to data analysis, and inadequate blinding of investigators to treatment groups (Jucker, 2010). Automated template-based normalization of rodent brain to standard coordinates, in analogy to standard methods for analysis of human PET data, has the potential to minimize biases from reader dependence and imperfect blinding, although the reliability of such approaches has yet to be systematically investigated for preclinical A $\beta$ -PET. In particular, the choice of the optimal template for automated spatial normalization of A $\beta$ -PET images may be influenced by pathological features of the particular AD mouse model (Rohlfing et al., 2009). Since this issue has not hitherto been raised in preclinical A $\beta$ -PET imaging, we planned to validate an automated, user-independent approach for spatial normalization by comparing binding results from readers at different training levels before and after normalization, with histopathological examination of fibrillar A $\beta$  as the gold standard. Furthermore, we objected to study the influence of template selection at different pathology stages on the automated spatial normalization of mouse brain A $\beta$ -PET images.

An additional point of contention in preclinical A $\beta$ -PET concerns the method for image intensity scaling, with normalization either to the injected dose, or to tracer uptake in intracerebral reference regions. Just as in human PET studies with A $\beta$ -PET tracer, the choice of reference regions is crucial for preclinical models, which have characteristic patterns of A $\beta$  deposition. In human A $\beta$ -PET imaging, the whole cerebellum and the cerebellar gray matter have been the preferred reference regions for most large scale quantitative investigations

(Vandenberghe et al., 2010; Barthel and Sabri, 2011; Clark et al., 2011). However, recent human studies with fluorinated amyloid tracers have revealed longitudinal stabilization of standardized uptake value ratios (SUVR), with use of white matter or brain stem reference regions, this despite the ongoing accumulation of A $\beta$  plaques as confirmed by histology (Landau et al., 2014, 2015; Brendel et al., 2015a; Chen et al., 2015). There has so far been no systematic comparison of scaling methods for mouse A $\beta$ -PET imaging. Consequently, we aimed to compare different scaling methods and reference regions in a large series of A $\beta$  studies with [ $^{18}\text{F}$ ]-florbetaben (FBB) in an established AD mouse model, with regard to accuracy of the PET method in capturing longitudinal changes in A $\beta$ -deposition, with terminal histological plaque quantitation serving as the gold standard. The overall aims of this study were to investigate if automated brain spatial normalization is beneficial for quantitation of preclinical A $\beta$ -PET, and to identify an optimally performing method for intensity scaling.

## MATERIALS AND METHODS

### Animal Model and Study Design

All experiments were performed in compliance with the National Guidelines for Animal Protection, Germany, with approval of the local animal care committee of the Government of Oberbayern (Regierung Oberbayern), and overseen by a veterinarian. Anesthesia was performed with isoflurane 1.5%. Mice were killed by cervical dislocation in a state of deep narcosis.

The transgenic B6.PS2APP (line B6.152H) is homozygous for both the human presenilin (PS) 2, N141I mutation and the human amyloid precursor protein (APP) K670N, M671L mutation. The APP and PS2 transgenes are driven by mouse Thy-1 and mouse prion promoters, respectively. This line had been created by co-injection of both transgenes into C57Bl/6 zygotes (Richards et al., 2003). Homozygous B6.PS2APP mice show first appearance of plaques in the cortex and hippocampus at 5–6 months of age (Ozmen et al., 2009).

A $\beta$ -PET scans ( $N = 114$ ) were used from two studies wherein drug-naive longitudinal FBB-PET recordings were obtained in PS2APP mice ( $N = 37$ ) at a short interval (8 and 9.5 months of age). Furthermore, we also used terminal A $\beta$ -PET recordings ( $N = 40$ ; range: 13–16 months;  $N = 24$  from the longitudinal set and  $N = 16$  from another historical investigation), in which histologically quantified terminal plaque load was measured. We defined the different ages for A $\beta$ -PET imaging as “BL” for the

baseline at 8 months ( $N = 37$  PS2APP mice), “FU” for the follow-up of the same group at 9.5 months ( $N = 37$  PS2APP mice), and “TER” for the terminal A $\beta$ -PET recordings obtained between 13 and 16 months of age ( $N = 40$  PS2APP mice).

## Radiochemistry

The [ $^{18}\text{F}$ ]-florbetaben precursor (Piramal Imaging, Berlin) was radiolabeled by the method of Zhang et al. (2005), with slight modifications. This procedure is described elsewhere (Rominger et al., 2013), and yields a radiochemical purity exceeding 98% and specific activity of 50–90 GBq/ $\mu\text{mol}$  at the end of synthesis.

## Amyloid Pet Acquisition and Reconstruction

Image acquisition and reconstruction followed a standardized protocol (Brendel et al., 2015b). Mice were anesthetized with isoflurane (1.5%, delivered via a mask at 3.5 L/min in oxygen) and received bolus injection of  $10.1 \pm 2.3$  MBq of [ $^{18}\text{F}$ ]-florbetaben in 150  $\mu\text{L}$  of saline to a tail vein. Following placement in the tomograph (Siemens Inveon DPET), a single frame emission recording for the interval 30–60 min p.i. followed by a 15 min transmission scan was obtained using a rotating [ $^{57}\text{Co}$ ] point source. The image reconstruction procedure consisted of an three-dimensional ordered subset expectation maximization (OSEM) with four iterations and 12 subsets followed by a maximum *a posteriori* (MAP) algorithm with 32 iterations. Scatter and attenuation correction were performed and a decay correction for [ $^{18}\text{F}$ ] was applied. With a zoom factor of 1.0 and a  $128 \times 128 \times 159$  matrix, a final voxel dimension of  $0.78 \times 0.78 \times 0.80$  mm was obtained.

## Image Analysis

### Spatial Normalization

A $\beta$ -PET images were first blinded to the reader by coding of the A $\beta$ -PET files. Images were coregistered to a 3T magnetic resonance imaging (MRI) mouse brain template (Dorr et al., 2007) by a rigid manual fusion, using the PMOD FUSION tool (v. 3.4 PMOD Technologies, Zürich). The spatial normalization was independently performed by an expert (>1000 fusions), an experienced reader (~150 fusions) as well as a novice reader who had been trained for 4 h over the course of 2 days. The expert fusion was repeated to assess test-retest variability (%) for operators with high training level.

All spatial normalized images of the expert reader were intensity scaled to the injected dose and multiplied by the individual body weight to give standardized uptake value (SUV) images. Then age-dependent groups ( $N = 37$  or  $N = 40$  images each) from PS2APP mice were averaged for the generation of three different templates defined at increasing stages of amyloid pathology (Figure 1A).

Subsequently, fused images from all readers were spatially normalized on all different templates by applying an automated nonlinear transformation, which removed the main differences in brain size and shape using the normalization algorithm in SPM5 (Wellcome Department of Cognitive Neurology, London, UK) implemented within the PMOD FUSION tool (V. 3.4 PMOD Technologies, Zürich). This algorithm was set

to mouse brain dimensions using standardized settings (equal modality; nonlinear warping; 16 iterations; frequency cutoff 3; regularization 1.0; no thresholding). The procedure was tested with and without transient input smoothing by applying a 0.6 mm Gaussian to the image before initiation of warping operations. The calculated transformation matrix was saved and finally applied to the “raw” input image to avoid loss of resolution due to repeated image resampling.

### Inter-Reader Analyses

For inter-reader analyses, Fleiss Kappa ( $\kappa$ ) was calculated for TER results (SUV<sub>CTX</sub> and SUVR<sub>CTX/REF</sub> as described below) from expert vs. experienced, experienced vs. novice, expert vs. novice, both before and after automated brain normalization. Furthermore, inter-reader variability (%) was calculated, and compared to the expert test-retest variability (%). For statistical testing of absolute differences for SUV<sub>CTX</sub> and SUVR<sub>CTX/REF</sub> between different readers before and after brain normalization a permutation test was performed as described in Section Statistics below and was also used to test for significant differences between inter-reader variability (%) and expert test-retest variability (%).

### Analysis of Template Influence on Automated Brain Intensity Normalization

To assess the impact of stage specific templates, TER results of the expert before brain normalization were compared with TER results after brain normalization to TER, FU, and BL templates by calculation of error-(%), with findings of the expert considered as the standard of truth. A paired *t*-test was used for statistical testing of significant alterations in SUV<sub>CTX</sub> and SUVR<sub>CTX/REF</sub> (as described below) resulting from different templates.

### Scaling

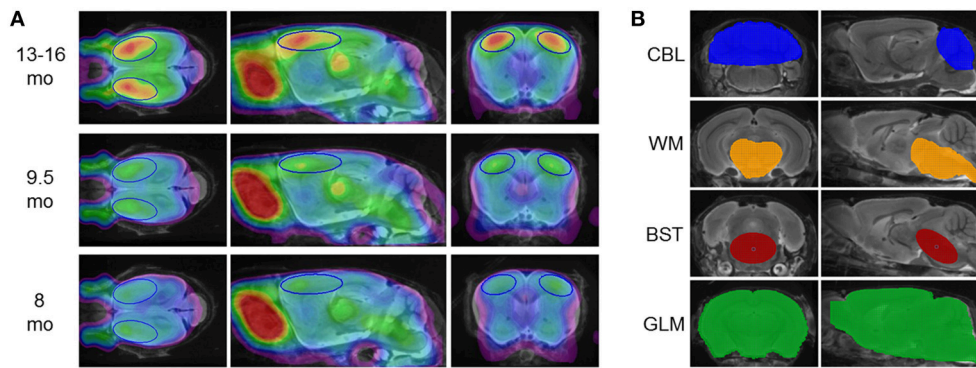
A total of five intensity scaling methods were performed, all using VOIs predefined in the final image space, which was identical for MRI and [ $^{18}\text{F}$ ]-florbetaben templates. Beside calculation of frontal cortical SUV<sub>CTX</sub>, we tested intensity scaling by four different reference regions (Figure 1B):

1. Whole cerebellum (CBL) with a volume of 65 mm<sup>3</sup>.
2. Hindbrain white matter (WM), including high unspecific tracer retention areas (threshold-based) such as pons, midbrain, cerebellar peduncle, cranial hypothalamus and caudal thalamus, with a total volume of 67 mm<sup>3</sup>.
3. Brainstem (BST), an oval shaped region extending from pons to midbrain, with a volume of 24 mm<sup>3</sup>.
4. Whole brain as the global mean (GLM), with a volume of 525 mm<sup>3</sup>.

As target regions two bilateral frontal cortex VOIs comprising 12 mm<sup>3</sup> each (Figure 1A), were employed for calculation of [ $^{18}\text{F}$ ]-florbetaben cortex-to-reference SUVR<sub>CTX/REF</sub>.

### Longitudinal stability

Scaling methods were tested against each other by evaluating the longitudinal stability of the endpoint (SUV<sub>CTX</sub> or SUVR<sub>CTX/REF</sub>) over the brief 6 week interval from BL to FU in  $N = 37$  animals. Pearson's coefficients of correlation (R) were calculated



**FIGURE 1 | (A)** [ $^{18}\text{F}$ ]-florbetaben PET templates at different pathology stages, deriving from mean findings in groups of PS2APP mice aged 8, 9.5, and 13–16 months. PET images are superimposed on an MRI-based mouse brain atlas (Dorr et al., 2007) for anatomical reference. The frontal cortical target VOI is depicted in blue. **(B)** Reference region VOIs are illustrated on the same MRI mouse atlas; from top to bottom: cerebellum (blue; CBL), hindbrain white matter (orange; WM), brainstem (red; BST), whole brain (green; GLM).

for  $\text{SUV}_{\text{CTX}}$  and  $\text{SUVR}_{\text{CTX}/\text{REF}}$  values of the four different reference regions, given the assumption that the mouse model is characterized by a nearly linear progression of amyloidosis over time, as supported by findings from our previous study (Brendel et al., 2015b). The variance of BL and FU groups, expressed by  $\text{SD}(\%)$ , was calculated as an indicator of intra-group stability. Effect sizes (Cohen's  $d$ ) for the resulting differences between the two sequential  $\text{A}\beta$ -PET scans were calculated as an additional quality criterion in the longitudinal design.

#### Longitudinal regional analyses

To test the impact of different scaling methods for  $\text{A}\beta$ -PET on the detected differences in longitudinal data independently from the cortical target VOI, we assessed alterations in FBB-binding between BL and FU voxel-wise by statistical parametric mapping (SPM). We used SPM5 routines implemented in MATLAB (version 7.1), adapted from Sawiak et al. (2009) for mouse data. For  $\text{SUV}_{\text{CTX}}$  and each reference region approach, we performed a paired  $t$ -test for  $\text{A}\beta$ -PET images (FU vs. BL) of PS2APP ( $N = 37$ ) mice, and thus assessed increases or decreases over 6 weeks of follow up.

#### Histochemical Analyses

Histochemical analyses were performed in a matching frontal cortex region of interest as the gold standard of amyloid burden, for evaluating reliability of frontal cortical  $\text{SUV}_{\text{CTX}}$  and four different  $\text{SUVR}_{\text{CTX}/\text{REF}}$  results. The procedure followed a standardized protocol wherein cortical plaque load (%) was calculated for each animal (Brendel et al., 2015b). For correlation analyses of the terminal  $\text{A}\beta$ -PET estimates ( $N = 40$ ) with plaque load (%), Pearson's coefficients of correlation ( $R$ ) were calculated with and without brain normalization and for all different intensity normalization methods. Significant differences between correlation coefficients before and after spatial brain normalization, between different readers, and between different intensity normalization methods were assessed by an extended Fisher's transformation approach as described in Section Statistics.

#### Statistics

A permutation test was used to test for significance of not normally distributed differences between two readers before and after normalization and for the comparison of inter-reader variability (%) with test-retest variability (%) of the expert. Absolute values were used for these comparisons. For permutation testing, the results were pooled and a loop rearranging the pooled results into two groups (with 1 million repeats) was coded within Matlab 7.12.0. The originally observed results of difference between manually acquired and normalized data or of difference between test-retest expert variability (%) and inter-reader variability (%) were defined as target values. For every resampled pair the mean result was calculated and each mean result equal to or higher than the target value was counted automatically. Finally, the total count was set in relation to the number of repeats to obtain the  $p$ -value.

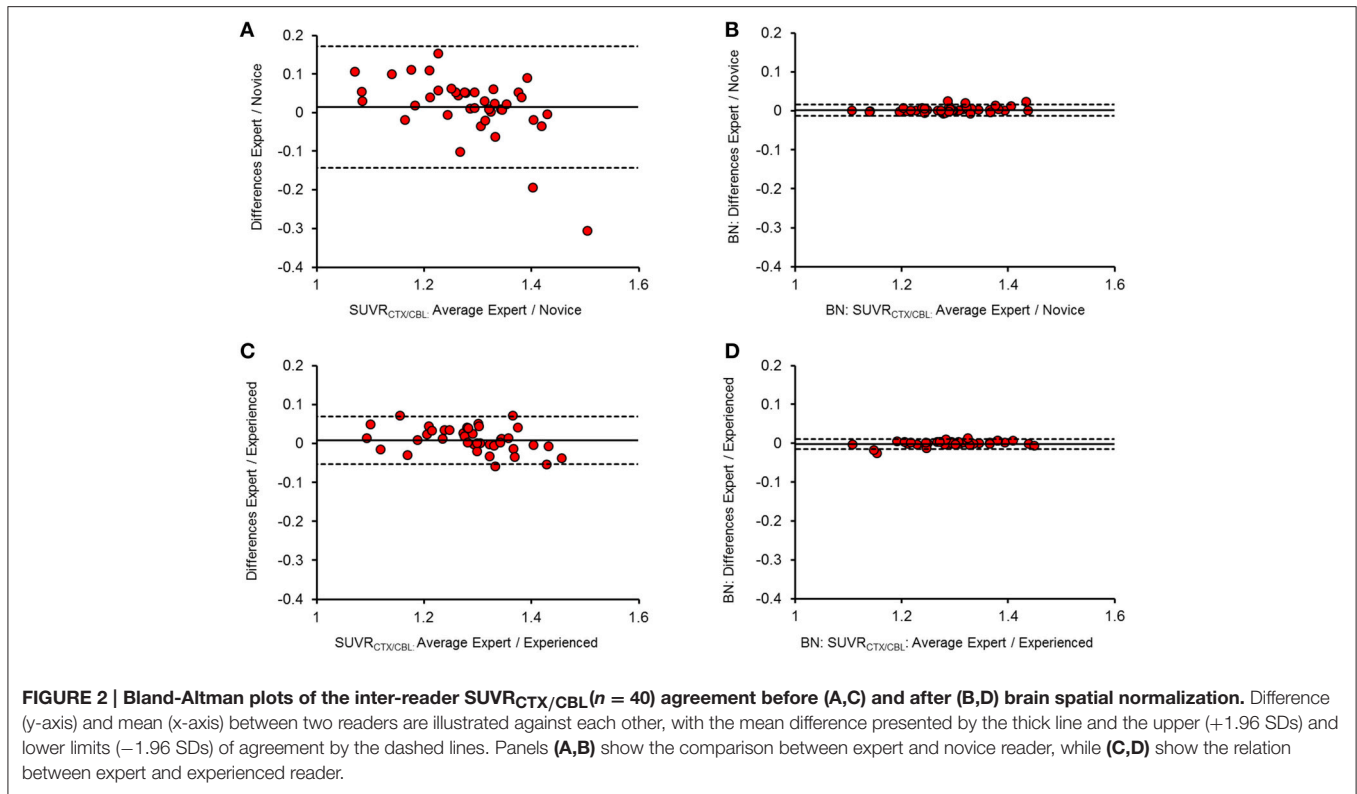
Significant differences between two dependent correlations with one variable in common (plaque load %) were assessed by an extended Fisher's transformation approach (Lee and Preacher, 2013). First, each correlation coefficient was converted into a  $z$ -score using Fisher's  $r$ -to- $z$  transformation. Then, asymptotic covariance of the estimates was computed. Finally, these quantities were used in an asymptotic  $z$ -test. A threshold of  $p < 0.05$  was considered to be significant for rejection of the null hypothesis in all statistical tests.

#### RESULTS

##### Spatial Normalization

##### Automated Brain Normalization Significantly Reduces Inter-Reader Variability

The intra-reader test-retest variability of  $\text{SUVR}_{\text{CTX}/\text{REF}}$  for the expert reader was  $1.4 \pm 1.0\%$  (range: 0.9–2.9% for different reference regions). Inter-reader agreement for  $\text{SUVR}_{\text{CTX}/\text{REF}}$  was very high between the expert and the experienced reader without brain pre-normalization ( $\kappa = 0.97 \pm 0.02$ ; inter-reader variability  $1.4 \pm 0.8\%$ ), indicating a very high reproducibility



when both readers are at a high training level without significant differences when compared to the expert test-retest variability (all  $p = n.s.$ ). However, lower inter-reader agreement including some strong outliers occurred between novice reader and expert ( $\kappa = 0.89 \pm 0.08$ ; inter-reader variability  $3.0 \pm 1.8\%$ ; all  $p < 0.01$ ), or novice and experienced reader ( $\kappa = 0.88 \pm 0.06$ ; inter-reader variability  $2.9 \pm 1.5\%$ ; all  $p < 0.01$ ). Especially the agreement for  $SUVR_{CTX/CBL}$  was lower in the latter two contrasts ( $\kappa = 0.77/0.79$ ; inter-reader variability =  $5.6\%/5.2\%$ ) when only the manual PET-MRI fusions were compared. After automated spatial brain normalization, all  $\kappa$  values were  $\geq 0.99$  and the maximal inter-reader variability was  $0.6\%$ . All differences between readers were significantly lowered by the automated brain normalization procedure ( $p < 0.001$  for all reference regions). Brain normalization without transient input smoothing ( $0.6\text{ mm}$ ) during the automated coregistration process led to some instances of strong image distortions, and was therefore dismissed. Representative approximations of  $SUVR_{CTX/CBL}$  between readers after brain spatial normalization are illustrated in Bland-Altman Plots (Figure 2) and all  $\kappa$ -values/%-variabilities are reported in Table 1.

### Template Characteristics Influence Brain Normalization Results

FBB uptake in the TER scans was underestimated when spatially normalized to templates from BL and FU disease stages, under the assumption that the expert manual fusion serves as the standard of truth. This effect was most obvious with the cerebellum serving as the reference region (BL-template:  $-0.6 \pm 2.1\%$ ,  $p < 0.05$ /FU-template:  $-2.9 \pm 2.4\%$ ,

$p < 0.001$ ; Figure 3A). With the stage specific TER template, low  $SUVR_{CTX/REF}$  values tended to be overestimated, whereas high results were underestimated, but the mean error-(%) was only  $\pm 0.2\%$ . Based on these findings, we attempted to quantify the bias deriving from the deviation of individual  $SUVR_{CTX/REF}$  from the template  $SUVR_{CTX/REF}$  by constructing a function defining the relationship between error-(%) and the deviation-(%) =  $(\text{individual} - SUVR_{CTX/REF} - \text{template} - SUVR_{CTX/REF}) / \text{template} - SUVR_{CTX/REF} * 100$ . Using this function we estimated a bias of  $\pm 13.8\%$  (linear regression:  $R = -0.44$ ;  $p < 0.01$ ) for  $SUVR_{CTX/CBL}$  (Figure 3B), meaning that an individual mouse with a 10% higher  $SUVR_{CTX/CBL}$  compared to the template results gives on average a 1.38% lower result after brain intensity normalization. Respective bias was similarly high for  $SUVR_{CTX/GLM}$  ( $\pm 10.1\%$ ; linear regression:  $R = -0.34$ ;  $p < 0.05$ ), but distinctly lower for  $SUVR_{CTX/WM}$  ( $\pm 1.9\%$ ; linear regression:  $R = -0.13$ ;  $p = n.s.$ ; Figure 3C) and  $SUVR_{CTX/BST}$  ( $\pm 1.7\%$ ; linear regression:  $R = -0.16$ ;  $p = n.s.$ ). Absolute biases in  $SUVR_{CTX/CBL}$  and  $SUVR_{CTX/GLM}$  were significantly higher when compared to  $SUVR_{CTX/WM}$  or  $SUVR_{CTX/BST}$  (all  $p < 0.001$ ).

### Longitudinal Analysis

#### Intracerebral Reference Regions give Lower Variance and Higher Longitudinal Effect Sizes When Compared to SUV

Implausibly low longitudinal increases of FBB binding were obtained through generation of ordinary  $SUVR_{CTX}$  estimates ( $R = -0.12$ ,  $p = n.s.$ ), which is explicable by the rather high relative standard deviations in the BL (15.9%) and FU (13.2%)

TABLE 1 | Inter-reader agreement.

Reference	$\kappa$ /inter-reader variability (%)			Test-retest variability
	Expert-experienced ( $\kappa$ /%)	Experienced-novice ( $\kappa$ /%)	Expert-novice ( $\kappa$ /%)	Expert (%)
<b>RAW FUSION</b>				
CBL	0.95/2.6	0.79/5.2*	0.77/5.6*	2.9
WM	0.98/1.0	0.91/2.4**	0.93/2.1**	0.9
BST	0.99/0.9	0.94/2.0**	0.96/1.8**	0.9
GLM	0.95/1.2	0.88/2.1**	0.89/2.3**	0.9
MEAN $\pm$ SD	0.97 $\pm$ 0.02/1.4 $\pm$ 0.8	0.88 $\pm$ 0.06/2.9 $\pm$ 1.5**	0.89 $\pm$ 0.08/3.0 $\pm$ 1.8**	1.4 $\pm$ 1.0
<b>BRAIN NORMALIZATION</b>				
CBL	1.00/0.5	0.99/0.6	1.00/0.5	
WM	1.00/0.2	0.99/0.4	0.99/0.4	
BST	1.00/0.2	0.99/0.4	0.99/0.5	
GLM	1.00/0.2	0.99/0.5	0.99/0.4	
MEAN $\pm$ SD	1.00 $\pm$ 0.00/0.2 $\pm$ 0.1	0.99 $\pm$ 0.00/0.5 $\pm$ 0.1	0.99 $\pm$ 0.00/0.5 $\pm$ 0.1	

Inter-reader agreement expressed by Fleiss  $\kappa$  and inter-reader variability (%) before and after brain normalization (BN) for all different reference regions. For the raw fusion the test-retest variability of the expert is given. \* $p < 0.01$ ; \*\* $p < 0.001$  for interreader-variability (%) vs. expert test-retest variability (%), permutation test.

groups. Each of the four intracerebral reference regions stabilized the A $\beta$ -PET SUVR<sub>CTX/REF</sub> estimates, and resulted in distinctly lower variance (2.3–4.6%), with highest agreement between BL and FU for SUVR<sub>CTX/WM</sub> ( $R = 0.64$ ,  $p < 0.001$ ). Highest effect sizes for the age-dependent increases were found with SUVR<sub>CTX/WM</sub> and SUVR<sub>CTX/BST</sub> ( $d = 1.64$ ), which exceeded that for SUVR<sub>CTX/CBL</sub> ( $d = 1.23$ ) and SUVR<sub>CTX/GLM</sub> ( $d = 1.21$ ). Details of this analysis are provided in Table 2.

### Global Mean Scaling Gives Physiologically Implausible Longitudinal Intensity Decreases in the Hindbrain

Voxel-wise analyses revealed a longitudinal progression of amyloidosis from BL to FU in the forebrain for all reference regions in 648,691 voxels (CBL), 662,374 voxels (WM), 612,517 voxels (BST), and 47,247 voxel (GLM) (all FDR-corrected;  $p < 0.05$ ; Figure 4A). Together with the distinctly lower number voxels with temporally increasing SUVR to GLM scaling, 156,066 voxels in the hindbrain (pons, midbrain, and cerebellum) indicated an implausible decrease for SUVR from BL to FU (FDR-corrected;  $p < 0.05$ ; Figure 4B). No amyloid or vascular pathology is known in these regions in PS2APP mice, nor were there any plausible physiological or pathophysiological explanations for a marked apparent decrease in the relative decrease in the hindbrain A $\beta$ -PET signal in a 6 weeks' follow-up. No significantly changing voxels were found when performing SPM with raw SUV images.

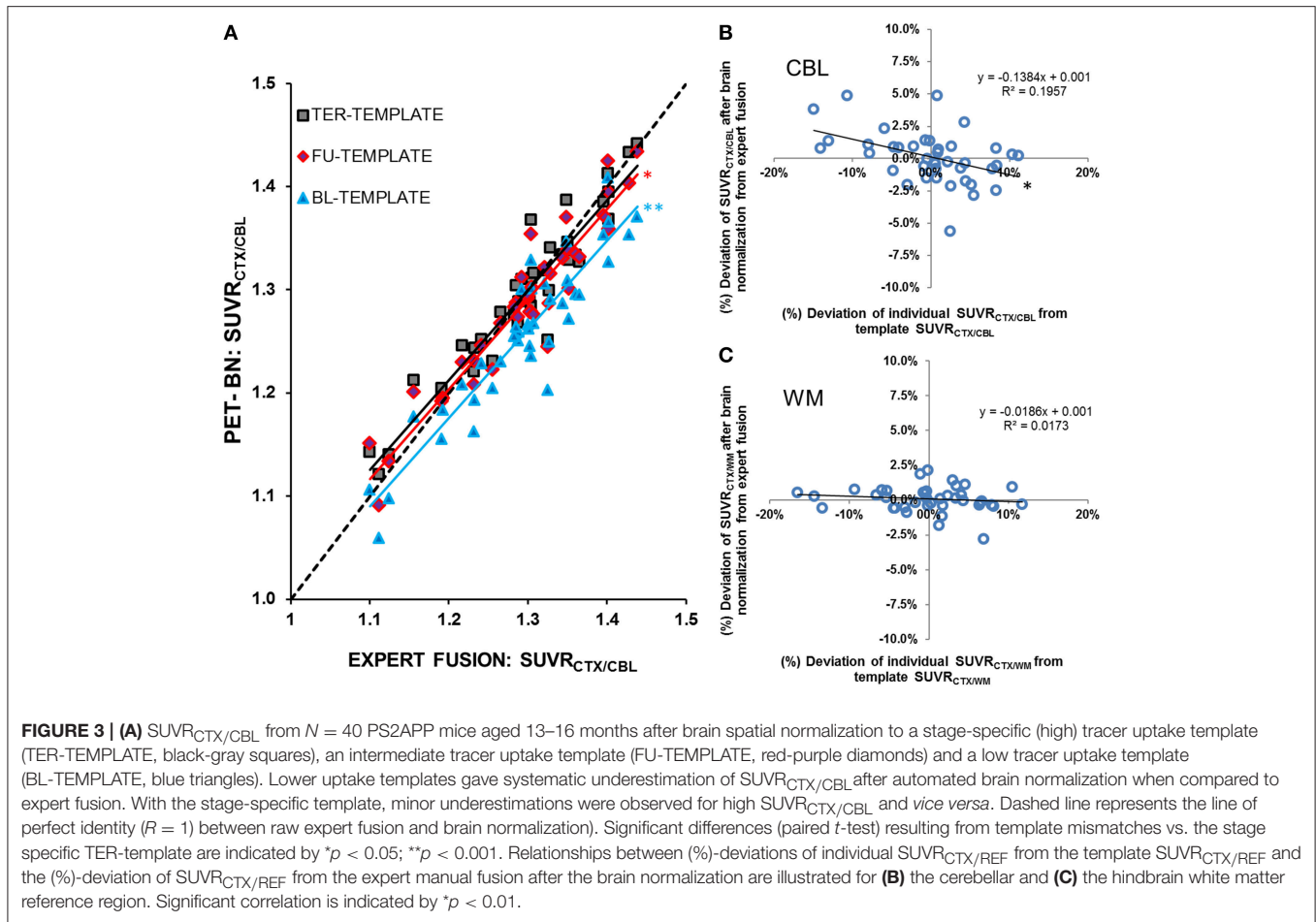
### Intensity Scaling by Intracerebral Reference Regions Gives Superior Agreement with Histology Compared to SUV, and is Further Improved by Automated Brain Spatial Normalization

The cortical plaque load (%) in  $N = 40$  PS2APP aged 13–16 months mice was  $10.8 \pm 1.2\%$ . The correlation between A $\beta$ -PET

and plaque load (%) was poor ( $R \leq 0.34$ ,  $p < 0.05$ ) for the case of plain SUVR<sub>CTX</sub> (Figure 5A), and consistent significantly lower when compared to any intensity scaling to intracerebral reference regions (all  $p < 0.001$ ). SUVR<sub>CTX/REF</sub> estimates (Figure 5B) revealed a high correlation ( $R \geq 0.66$ ,  $p < 0.001$ ) with plaque load (%) in this homogeneous sample of PS2APP mice regardless of the operator training status, or application of automated brain spatial normalization. Without automated brain normalization, the correlation to histology of the manual coregistration by the novice reader was significantly lower ( $R_{\max} = 0.70$ ) when compared to the expert ( $R_{\max} = 0.78$ ;  $p < 0.01$ ) or experienced reader ( $R_{\max} = 0.76$ ;  $p < 0.05$ ). Application of spatial brain normalization harmonized this correlation between different readers ( $R_{\max} = 0.75$ – $0.76$ ; no significant differences between readers). In this regard, manual coregistration by the novice reader was significantly improved by automated brain normalization (all reference regions:  $p < 0.05$ ), whereas the changes for the expert and experienced reader did not show any significant alterations (all reference regions:  $p = \text{n.s.}$ ). Scaling to WM or GLM ( $R_{\text{mean}} = 0.75$ ) was slightly superior ( $p = \text{n.s.}$ ) to BST ( $R_{\text{mean}} = 0.74$ ) or CBL ( $R_{\text{mean}} = 0.73$ ). All correlations are provided in Table 3.

### Best Global Performance is Archived by Hindbrain White Matter Intensity Scaling

Based on the different analyses (brain normalization, group analysis, longitudinal assessment, and correlation with histology as the gold standard) we defined five categories, which need to be addressed when considering the optimal intensity scaling for [<sup>18</sup>F]-florbetaben A $\beta$ -PET imaging in mice (Table 4). (1) The impact of single animal results analyzed by templates used for automated brain normalization should be minimal, but this was not the case for cerebellar and global mean intensity scaling. (2) Variance in groups of mice should not be artifactually raised by the method, which was the case for raw SUVR<sub>CTX</sub>. (3) The effect sizes regarding longitudinal progression of amyloidosis



**TABLE 2 | Longitudinal 6-week follow-up.**

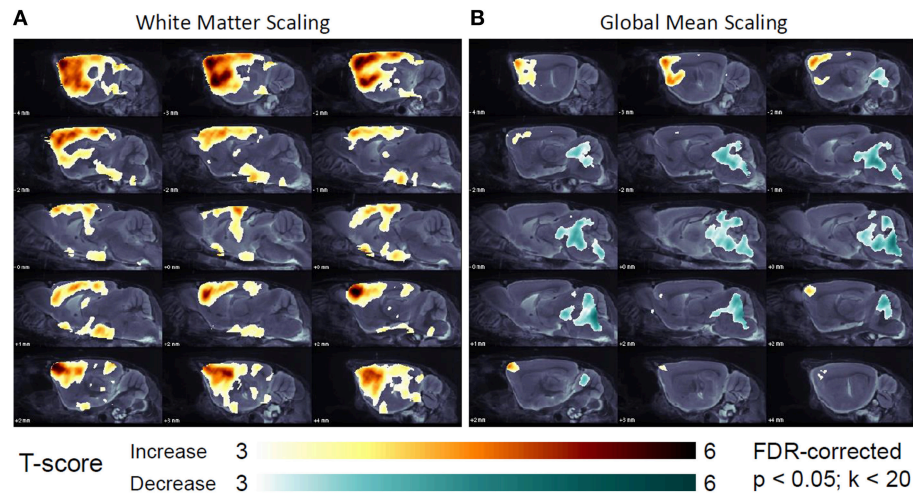
	BL MEAN	BL SD (%)	FU MEAN	FU SD (%)	d	R
SUV <sub>CTX</sub>	0.48	0.08 (15.9%)	0.50	0.07 (13.2%)	0.23	-0.12
SUVR <sub>CTX/CBL</sub>	1.10	0.04 (3.6%)	1.16	0.05 (4.4%)	1.23	0.51
SUVR <sub>CTX/WM</sub>	0.92	0.03 (3.1%)	0.98	0.04 (4.1%)	1.64	0.64
SUVR <sub>CTX/BST</sub>	0.90	0.03 (3.6%)	0.96	0.04 (4.6%)	1.64	0.60
SUVR <sub>CTX/GLM</sub>	1.06	0.02 (2.3%)	1.10	0.04 (3.4%)	1.21	0.63

Results from the longitudinal 6 week follow-up in (N = 37) PS2APP mice. Mean cortical FBB PET values and their SD are given for BL and FU scans for the different scaling methods together with the Cohen's effect size (d) and BL-FU correlation (R).

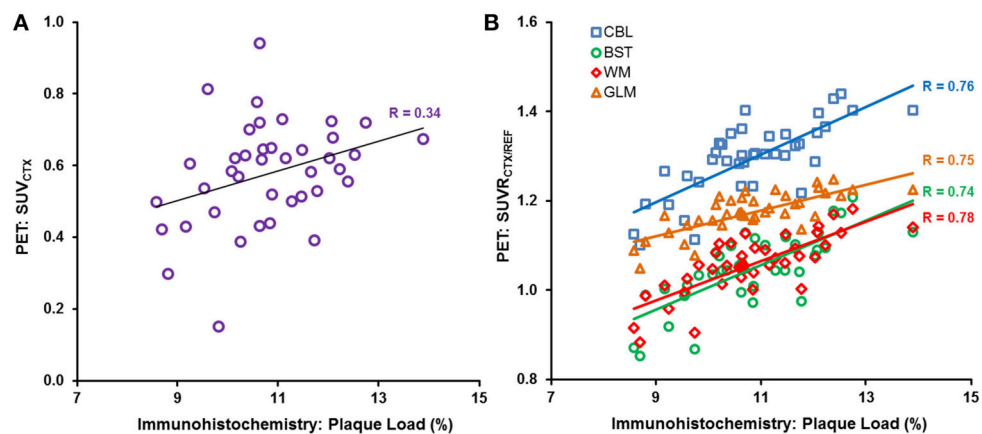
in the frontal cortical target region should be captured as best possible, which was the case for hindbrain white matter and brainstem intensity scalings. (4) Longitudinal results should not be affected by presence of amyloid pathology in the reference region, which was the case for the global mean intensity scaling. (5) Aβ-PET signal in single animals should correlate as best possible with the histological gold standard, which was the case for all intracerebral reference regions. There was trend toward better agreement for hindbrain white matter and global mean intensity scaling, compared to cerebellum or brainstem scaling. Thus, in summary the hindbrain white matter scaling performed best when considering the composite of factors.

## DISCUSSION

We present a large-scale evaluation of automated spatial brain normalization and a systematic comparison of different intensity scaling methods for preclinical Aβ-PET imaging with FBB. Our established Aβ-PET methodologies were challenged by investigation of a homogeneous PS2APP data set, including a short-term longitudinal follow-up, and histologically validated gold standard results, all characterized by a rather narrow range of plaque burden; this bodes well for sensitive detection of treatment effects. Automated spatial brain normalization reliably harmonized the results of different



**FIGURE 4 |** Longitudinal changes in regional [ $^{18}\text{F}$ ]-florbetaben uptake in the short-term 6 week follow-up, study of ( $N = 37$ ) PS2APP mice after intensity scaling to the hindbrain white matter (**A**) and the global mean uptake (**B**) as assessed by SPM (FDR-corrected;  $p < 0.05$ ;  $k < 20$ ). T-contrasts expressing significant longitudinal increases (yellow-red) and decreases (turquoise-green) are projected upon sagittal slices of the MRI mouse brain atlas.



**FIGURE 5 |** Individual (**A**)  $\text{SUV}_{\text{CTX}}$  and (**B**)  $\text{SUVR}_{\text{CTX}/\text{REF}}$  PET correlations (expert manual fusion) with gold standard plaque load (%), as assessed by histochemistry, in ( $N = 40$ ) PS2APP mice aged 13–16 months, are illustrated for  $\text{SUV}_{\text{CTX}}$  (purple circles) and the different reference regions, i.e., cerebellum (CBL, blue squares), hindbrain white matter (WM, red diamonds), brain stem (BST, green circles), and global mean (GLM, orange triangle).

readers even if poorly trained, while the correspondence with histological gold standard assessments remained stable or even improved. Scaling by intracerebral reference regions ( $\text{SUVR}_{\text{CTX}/\text{REF}}$ ) was distinctly superior to plain  $\text{SUV}_{\text{CTX}}$ , both with regard to longitudinal stabilization and agreement with histology. Global mean intensity scaling revealed a comparable correlation with histology when contrasted to reference regions demonstrably devoid of  $\text{A}\beta$  pathology, but this procedure impaired detection of longitudinal increases in  $\text{A}\beta$  in cortex, and returned physiologically implausible decreases in brain regions lacking  $\text{A}\beta$  burden. A hindbrain white matter reference was slightly superior to the cerebellar reference due, we suppose, to lesser bias from bone uptake and template differences.

## Automated Brain Normalization

While small animal  $\text{A}\beta$ -PET studies experienced growing interest in the recent years the methodological tools used for their interpretation were rather heterogeneous. Spatial normalization has hitherto mostly been done on MRI templates (Poinsel et al., 2012; von Reutern et al., 2013), or with a hybrid PET-CT apparatus (Snellman et al., 2013). One previous investigation of this type used SPM-based automated brain normalization (Rojas et al., 2013), and in most studies, the tracer uptake was scaled to a cerebellum reference region (Maeda et al., 2007; Kuntner et al., 2009; Manook et al., 2012; Poinsel et al., 2012; Snellman et al., 2013), although some have employed plain SUV for their analyses (Waldron et al., 2015). However, with respect to the small effect sizes to brief follow-up in AD mice, and the susceptibility of



**TABLE 3 | Histology agreement.**

Reader	Expert		Experienced		Novice	
	RAW	BN	RAW	BN	RAW	BN
SUV <sub>CTX</sub>	0.34**	0.34**	0.34**	0.34**	0.32**	0.31**
SUVR <sub>CTX/CBL</sub>	0.76	0.75	0.76	0.74	0.66*	0.74
SUVR <sub>CTX/WM</sub>	0.78	0.76	0.76	0.76	0.69*	0.75
SUVR <sub>CTX/BST</sub>	0.74	0.74	0.74	0.73	0.70*	0.73
SUVR <sub>CTX/GLM</sub>	0.75	0.76	0.76	0.76	0.70*	0.75

Correlations (*R*) with plaque load (%) as assessed by histochemistry for expert, experienced and novice reader for the different scaling methods before (RAW) and after brain normalization (BN). Significant differences between correlation coefficients are indicated for SUV<sub>CTX</sub> vs. any SUVR<sub>CTX/REF</sub> and for raw fusions vs. brain normalizations of the same reader \**p* < 0.05; \*\**p* < 0.001.

**TABLE 4 | Intensity scaling summary.**

	Template affection	Variance	Longitudinal effect size	Pathology affection	Histology
SUV <sub>CTX</sub>	o	–	–	o	–
SUVR <sub>CTX/CBL</sub>	–	+	+	o	+
SUVR <sub>CTX/WM</sub>	o	+	++	o	+(+)
SUVR <sub>CTX/BST</sub>	o	+	++	o	+
SUVR <sub>CTX/GLM</sub>	–	+	+	–	+(+)

Significant advantages (moderate: +, light green; large: ++, dark green)/disadvantages (–, light red) of different SUV<sub>CTX</sub> and SUVR calculations are illustrated in five categories, which are deemed relevant for analysis of [<sup>18</sup>F]-florbetaben measurements in the mouse brain. "o" indicate neutral condition; bracketed "+" indicate a trend to advantage. "Template affection" refers to results Section Template Characteristics Influence Brain Normalization Results. "Variance" refers to result Section Intracerebral Reference Regions Give Lower Variance and Higher Longitudinal Effect Sizes When Compared to SUV. "Longitudinal effect size" refers to result Section Intracerebral Reference Regions Give Lower Variance and Higher Longitudinal Effect Sizes When Compared to SUV. "Pathology affection" refers to result Section Global Mean Scaling Gives Physiologically Implausible Longitudinal Intensity Decreases in the Hindbrain. "Histology" refers to result Section Intensity Scaling by Intracerebral Reference Regions Gives Superior Agreement with Histology Compared to SUV, and is Further Improved by Automated Brain Spatial Normalization.

our SUV<sub>CTX</sub> results to methodological inaccuracies, it is crucial that Aβ-PET data be analyzed with robust, reliable and reader-independent strategies, so as to ensure accuracy of the study outcomes.

To our knowledge there have not been any studies systematically investigating the inter-reader variability as a limitation of small animal Aβ-PET studies, or indeed for any other classes of radioligand. Our investigations were able to show that manual coregistration of Aβ-PET images contributes substantially to variation of results between readers, especially when they have less experience. A further disadvantage of a manual approach arises from the blinding which is necessary for preclinical drug trials (Jucker, 2010). Thus, an experienced reader is usually able to visually identify animals with different extent of the Aβ pathology. So even if differing groups are blinded to the reader, the visual impressions of the images potentially influence the reader. The automated brain normalization diminishes the impact of these influences toward zero as SUVR<sub>CTX/REF</sub> values

show nearly perfect inter-reader agreement after processing as inter-reader differences were significantly lowered by the method (Figure 2). Thus, the brain normalization brings about a pseudo-blinding for small animal Aβ-PET analysis, and guarantees minimization of inter-reader variance.

While aiming to improve the inter-reader agreement, the automated brain normalization method should not reduce the ultimate correlation between the PET endpoint with the histological gold standard of amyloid burden. Our results clearly indicate that correlation coefficients of PET results with plaque burden remain equal after brain spatial normalization for well-trained readers, whereas the agreement between PET and histology is significantly improved for less trained operators (Table 3): this is an important precondition for stable results in longitudinal or multicentre studies.

We also observed an effect of applying Aβ-PET templates at different pathology stages. Although this effect was never more than 2.9% of the group mean, this avoidable systematic error could potentially bias Aβ-PET outcomes in treatment studies with small effect sizes (Balducci et al., 2014; Brendel et al., 2015c). This bias of inappropriate templates is unsurprising, as automated brain normalization orientates mainly at edges and high uptake regions of the image (Gispert et al., 2003). High plaque burden in the forebrain results in a cortical hot-spot and sharper cortex-to-extracortical contrasts for Aβ-PET mouse brain images, which influences the quality of spatial normalization when this hot region is absent from the utilized template or *vice versa*. In practice, this mismatch tends to shrink the scaled PET image, leading to underestimation in mice with higher plaque burden than in the template and an overestimation in mice with lower plaque burden. Interestingly SUVR<sub>CTX/WM</sub> and SUVR<sub>CTX/BST</sub> suffered a lower bias (<2% relative to the deviation of the individual Aβ-PET image to its template) when compared to SUVR<sub>CTX/CBL</sub> and SUVR<sub>CTX/GLM</sub> (>10% relative to the deviation of the individual Aβ-PET to its template), supporting the preferred use of the WM and BST reference regions (Figure 3), with matching of stage/age. Furthermore, transient smoothing of the input image with a 0.6 mm filter proved indispensable to avoid strong distortions in the automated spatial normalization.

## Intensity Scaling Methods

Variance of plain SUV<sub>CTX</sub> was rather high in BL and FU groups of this investigation (15.9/13.2%), thus indicating inadequate intensity scaling of Aβ-PET images by this method. Well-known sources of error in the SUV<sub>CTX</sub> calculation include paravenous leakage of the injectate, variability in brain perfusion, and imprecise measurement of radiotracer dosage. We typically calculate injected dose by well-counting syringes before and after tracer application, whereas measurement of the injected dose in a whole body VOI can be more precise (Rominger et al., 2013). However, our scanning configuration does not always capture the entire body of the mice. While SUV measurements are improved with arterial blood sampling, this is inherently linked with a high logistic and economic effort, scarcely feasible in a longitudinal setting. Our SUV<sub>CTX</sub> findings were distinctly inferior to the several SUVR<sub>CTX/REF</sub> results, with respect both to longitudinal

and histological validation analyses. Thus, SUV methods are not to be recommended for preclinical A $\beta$ -PET studies.

Having ascertained the need to make SUVR calculations, we need to identify the optimal reference region for intensity scaling. Since a whole brain VOI is easily obtained, scaling to the global mean brain uptake is convenient. Indeed we find a high correlation between SUVR<sub>CTX/GLM</sub> and the histological gold standard of A $\beta$  accumulation. However, when large parts of the brain are affected by the amyloid pathology, the denominator for scaling is of course raised by the high tracer retention. This resulted in the spurious detection of declining relative FBB uptake in the hindbrain to follow-up at 6 weeks, an artifact that could mimic a real clearance of fibrillar A $\beta$  in longitudinal interventional designs. Furthermore, global mean scaling attenuated the apparent cluster size and effect size for increases in A $\beta$  in the longitudinal part of the study, relative to findings with scaling to reference regions expected to be devoid of A $\beta$  pathology. Our findings are supported by earlier [<sup>18</sup>F]-FDG-PET investigations of metabolic rate in human brain, in relation to artifactual findings arising as a consequence of global normalization (Borghammer et al., 2009). Scaling to [<sup>18</sup>F]-FDG uptake in a reference cluster unaffected by pathology clearly improved the detection of zones of true hypometabolism in patients with AD or fronto-temporal dementia (Yakushev et al., 2009; Dukart et al., 2013). Thus, we are confident that scaling by reference regions devoid of plaque pathology should perform best for detection of A $\beta$  in the present mouse study.

In most transgenic mouse models, the hindbrain remains relatively unaffected by A $\beta$  plaque accumulation (Teipel et al., 2011), but due to the characteristically high retention of [<sup>18</sup>F]-labeled A $\beta$  tracers in white matter, the binding in hindbrain regions is quite heterogeneous. With regard to FBB-PET, we see low non-specific binding in the cerebellum (Rominger et al., 2013), which led to our adoption of cerebellum as a reference region in small animal A $\beta$ -PET analysis for this and related ligands (Manook et al., 2012). While the cerebellum can serve as an accurate reference region, it remains vulnerable to spill-in from the overlying cranium and adjacent vascular structures (Mille et al., 2012). In addition, cerebellum VOIs are at risk for contamination from non-cerebral voxels, if the caudal border of the brain is not captured precisely. Together, these factors probably account for instances of false high and false low SUVR<sub>CTX/CBL</sub>; resultant higher variance especially in longitudinal analyses of SUVR<sub>CTX/CBL</sub> are responsible for the lower correlation between BL and FU in the present longitudinal arm, as compared to the more precise findings with SUVR<sub>CTX/WM</sub> or SUVR<sub>CTX/BST</sub>. While spill-over and imperfect capture of the caudal brain edges are less of an issue for human A $\beta$ -PET, WM reference scaling seems to give more stable results to follow-up than does cerebellum scaling (Chen et al., 2015; Brendel et al., 2015a). Similarly, the highest effect size ( $d = 1.64$ ) in the present 6-week longitudinal setting was observed for SUVR<sub>CTX/WM</sub> and SUVR<sub>CTX/BST</sub>.

The whole point of A $\beta$ -PET is to depict accurately the plaque load. We designed the present study to test the limits of sensitivity of the method, by examining reference tissue methods in a cohort of PS2APP mice with rather narrow inter-individual range

in plaque load (8.6–13.9%). This is in comparison to the ten-fold range used in previous mouse studies of the correlation between A $\beta$ -PET and histological plaque load (1.0–9.2%, 0.3–13.3%; Manook et al., 2012; Rominger et al., 2013). Results with all intracerebral reference regions correlated highly with the histological gold standard, although WM normalization gave a slightly higher correlation. The slight superiority of SUVR<sub>CTX/WM</sub> over SUVR<sub>CTX/BST</sub> can be explained by the larger volume, giving less statistical noise in the results. The slight superiority of SUVR<sub>CTX/WM</sub> over SUVR<sub>CTX/CBL</sub>, attributed above to imperfect delineation of the caudal limit of the brain, is also demonstrated by the lower correlation of SUVR<sub>CTX/CBL</sub> by the novice reader with histology ( $R = 0.66$ ); visual interpretation of caudal limit of the fusion images is a matter of skill-learning.

In summary, SUVR<sub>CTX/WM</sub> performed best for the detection of A $\beta$  by PET in transgenic mice, which should prove advantageous in preclinical imaging studies of longitudinal design (Table 4). Thus, quantitation of [<sup>18</sup>F]-florbetaben uptake in interventional trials, where accurate monitoring in single animals is mandatory, can be improved implementing hindbrain white matter intensity scaling, in conjunction with automated spatial normalization.

## Limitations

As mentioned above, we did not use the best possible means to measure the injected FBB dose; we suppose that whole mouse VOIs would have propagated to better correlations of SUVR<sub>CTX</sub> with the histological gold standard. While our PS2APP mice are characterized by almost complete absence of A $\beta$  pathology in the hindbrain, present findings are not translatable to mouse models, which may express relevant A $\beta$  plaque burden in the reference regions tested in this paper. Thus, prior knowledge about the regional deposition of A $\beta$  is indispensable for appropriate selection of the reference in preclinical A $\beta$ -PET studies with FBB or related tracers.

Excessive reliance on automated methods for spatial normalization of mouse brain should be avoided. Although we did not detect any failures of the method, they might still occur, were artificial hot spots present in the image. Thus, we saw distinctly more cases of image distortions, especially in brains with more heterogeneous FBB uptake, in the absence of transient input smoothing. As in all PET studies, it is important to visually control post-processed images for potentially failed spatial normalizations. In consideration of this issue, we first attempted automated brain normalization with larger templates (data not shown) encompassing adjacent extracerebral regions, so as to support a correction of effects of partial volume (Brendel et al., 2014). This approach yielded some excessive distortions of the brain, when the VOI extended to include sources in the neck or spine. We note that templates fulfilling the differing requirements for partial volume effect correction and automated brain normalization remain to be validated.

## CONCLUSION

Automated spatial normalization of mouse brain can be applied to A $\beta$ -PET studies with FBB, thus ensuring pseudo-blinding,

and by eliminating inter-reader variability due to differing skill learning. This improves the reproducibility of the endpoint, even if the reader is poorly trained. Intracerebral reference regions lacking A $\beta$  pathology are necessary for accurate *in vivo* quantification of A $\beta$ -PET with FBB; use of a hindbrain WM reference tissue gave the best performance predominately due to advantages in longitudinal designs. SUV<sub>CTX</sub> generation and global mean scaling were distinctly inferior in performance to the intensity scaling by pathology-free intracerebral reference regions.

## AUTHOR CONTRIBUTIONS

FO carried out the PET experiments, performed the data analysis, and drafted the manuscript; MB participated in the design of the study, contributed to global data analysis, performed the statistical analysis, and drafted the manuscript; AJ carried out the histological experiments, performed the histological data analysis, and drafted the manuscript; VK carried out the histological experiments, and performed the histological data analysis; AD performed statistical programming and analysis; FP carried out the PET experiments, and performed the PET data analysis; CF carried out the PET experiments, and performed the PET data analysis; FG carried out the radiochemistry; JC participated in the PET experiments and helped to draft the manuscript; KB participated in the design of the study and helped

to draft the manuscript; CH participated in the design of the study and increased the intellectual content; PB participated in the design of the study and helped to draft the manuscript; JH conceived of the study, and participated in its design and coordination and drafted the manuscript; AR conceived of the study, and participated in its design and coordination, contributed to interpretation of the data and drafted the manuscript. All authors read and approved the final manuscript.

## FUNDING

The study was financially supported by the SyNergy Cluster (Core 2 project). AJ was supported by the Foundation for Polish Science within the International PhD Project “Studies of nucleic acids and proteins—from basic to applied research,” co-financed by European Union—Regional Development Fund (MPD/2009-3/2).

## ACKNOWLEDGMENTS

A part of this paper originated from the doctoral thesis of FO. We thank Karin Bormann-Giglmair and Rosel Oos for excellent technical assistance. Florbetaben precursor was kindly provided by Piramal Imaging. The authors acknowledge Inglewood Biomedical Editing for professional editing of the manuscript.

## REFERENCES

- Balducci, C., Mancini, S., Minniti, S., La Vitola, P., Zotti, M., Sancini, G., et al. (2014). Multifunctional liposomes reduce brain beta-amyloid burden and ameliorate memory impairment in Alzheimer's disease mouse models. *J. Neurosci.* 34, 14022–14031. doi: 10.1523/JNEUROSCI.0284-14.2014
- Barthel, H., and Sabri, O. (2011). Florbetaben to trace amyloid-beta in the Alzheimer brain by means of PET. *J. Alzheimers Dis.* 26, 117–121. doi: 10.3233/JAD-2011-0068
- Borghammer, P., Cumming, P., Aanerud, J., Förster, S., and Gjedde, A. (2009). Subcortical elevation of metabolism in Parkinson's disease—a critical reappraisal in the context of global mean normalization. *Neuroimage* 47, 1514–1521. doi: 10.1016/j.neuroimage.2009.05.040
- Brendel, M., Delker, A., Rotzer, C., Boning, G., Carlsen, J., Cyran, C., et al. (2014). Impact of partial volume effect correction on cerebral beta-amyloid imaging in APP-Swe mice using [(18)F]-florbetaben PET. *Neuroimage* 84, 843–853. doi: 10.1016/j.neuroimage.2013.09.017
- Brendel, M., Högenauer, M., Delker, A., Sauerbeck, J., Bartenstein, P., Seibyl, J., et al. (2015a). Improved longitudinal [(18)F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction. *Neuroimage* 108, 450–459. doi: 10.1016/j.neuroimage.2014.11.055
- Brendel, M., Jaworska, A., Griessinger, E., Rotzer, C., Burgold, S., Gildehaus, F. J., et al. (2015b). Cross-sectional comparison of small animal [18F]-florbetaben amyloid-PET between transgenic AD mouse models. *PLoS ONE* 10:e0116678. doi: 10.1371/journal.pone.0116678
- Brendel, M., Jaworska, A., Herms, J., Trambauer, J., Rötzer, C., Gildehaus, F. J., et al. (2015c). Amyloid-PET predicts inhibition of *de novo* plaque formation upon chronic gamma-secretase modulator treatment. *Mol. Psychiatry* 20, 1179–1187. doi: 10.1038/mp.2015.74
- Chen, K., Roontiva, A., Thiyyagura, P., Lee, W., Liu, X., Ayutyanont, N., et al. (2015). Improved power for characterizing longitudinal amyloid-beta PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. *J. Nucl. Med.* 56, 560–566. doi: 10.2967/jnumed.114.149732
- Clark, C. M., Schneider, J. A., Bedell, B. J., Beach, T. G., Bilker, W. B., Mintun, M. A., et al. (2011). Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA* 305, 275–283. doi: 10.1001/jama.2010.2008
- Dorr, A., Sled, J. G., and Kabani, N. (2007). Three-dimensional cerebral vasculature of the CBA mouse brain: a magnetic resonance imaging and micro computed tomography study. *Neuroimage* 35, 1409–1423. doi: 10.1016/j.neuroimage.2006.12.040
- Dukart, J., Perneczky, R., Förster, S., Barthel, H., Diehl-Schmid, J., Draganski, B., et al. (2013). Reference cluster normalization improves detection of frontotemporal lobar degeneration by means of FDG-PET. *PLoS ONE* 8:e55415. doi: 10.1371/journal.pone.0055415
- Gispert, J. D., Pascau, J., Reig, S., Martínez-Lázaro, R., Molina, V., García-Barreno, P., et al. (2003). Influence of the normalization template on the outcome of statistical parametric mapping of PET scans. *Neuroimage* 19, 601–612. doi: 10.1016/S1053-8119(03)00072-7
- Johnson, K. A., Minoshima, S., Bohnen, N. I., Donohoe, K. J., Foster, N. L., Herscovitch, P., et al. (2013). Appropriate use criteria for amyloid PET: a report of the amyloid imaging task force, the society of nuclear medicine and molecular imaging, and the Alzheimer's association. *J. Nucl. Med.* 54, 476–490. doi: 10.2967/jnumed.113.120618
- Jucker, M. (2010). The benefits and limitations of animal models for translational research in neurodegenerative diseases. *Nat. Med.* 16, 1210–1214. doi: 10.1038/nm.2224
- Kuntner, C., Kesner, A. L., Bauer, M., Kreamlehner, R., Wanek, T., Mandler, M., et al. (2009). Limitations of small animal PET imaging with [18F]FDDNP and FDG for quantitative studies in a transgenic mouse model of Alzheimer's disease. *Mol. Imaging Biol.* 11, 236–240. doi: 10.1007/s11307-009-0198-z
- Landau, S. M., Fero, A., Baker, S. L., Koeppe, R., Mintun, M., Chen, K., et al. (2015). Measurement of longitudinal beta-amyloid change with 18F-florbetapir

- PET and standardized uptake value ratios. *J. Nucl. Med.* 56, 567–574. doi: 10.2967/jnumed.114.148981
- Landau, S. M., Thomas, B. A., Thurfjell, L., Schmidt, M., Margolin, R., Mintun, M., et al. (2014). Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. *Eur. J. Nucl. Med. Mol. Imaging* 41, 1398–1407. doi: 10.1007/s00259-014-2753-3
- Lee, I. A., and Preacher, K. J. (2013). *Calculation for the Test of the Difference between two Dependent Correlations with One Variable in Common*. Available online at: <http://quantpsy.org>
- Maeda, J., Ji, B., Irie, T., Tomiyama, T., Maruyama, M., Okouchi, T., et al. (2007). Longitudinal, quantitative assessment of amyloid, neuroinflammation, and anti-amyloid treatment in a living mouse model of Alzheimer's disease enabled by positron emission tomography. *J. Neurosci.* 27, 10957–10968. doi: 10.1523/JNEUROSCI.0673-07.2007
- Manook, A., Yousefi, B. H., Willuweit, A., Platzer, S., Reder, S., Voss, A., et al. (2012). Small-animal PET imaging of amyloid-beta plaques with [<sup>11</sup>C]PiB and its multi-modal validation in an APP/PS1 mouse model of Alzheimer's disease. *PLoS ONE* 7:e31310. doi: 10.1371/journal.pone.0031310
- Mille, E., Cumming, P., Rominger, A., La Fougère, C., Tatsch, K., Wangler, B., et al. (2012). Compensation for cranial spill-in into the cerebellum improves quantitation of striatal dopamine D(2)/(3) receptors in rats with prolonged [(1)(8)F]-DMFP infusions. *Synapse* 66, 705–713. doi: 10.1002/syn.21558
- Ozmen, L., Albientz, A., Czech, C., and Jacobsen, H. (2009). Expression of transgenic APP mRNA is the key determinant for beta-amyloid deposition in PS2APP transgenic mice. *Neurodegener. Dis.* 6, 29–36. doi: 10.1159/000170884
- Poisnel, G., Dhilly, M., Moustie, O., Delamare, J., Abbas, A., Guilloteau, D., et al. (2012). PET imaging with [<sup>18</sup>F]AV-45 in an APP/PS1-21 murine model of amyloid plaque deposition. *Neurobiol. Aging* 33, 2561–2571. doi: 10.1016/j.neurobiolaging.2011.12.024
- Richards, J. G., Higgins, G. A., Ouagazzal, A. M., Ozmen, L., Kew, J. N., Bohrmann, B., et al. (2003). PS2APP transgenic mice, coexpressing hPS2mut and hAPPswe, show age-related cognitive deficits associated with discrete brain amyloid deposition and inflammation. *J. Neurosci.* 23, 8989–9003.
- Rohlfing, T., Sullivan, E. V., and Pfefferbaum, A. (2009). Subject-matched templates for spatial normalization. *Med. Image Comput. Comput. Assist. Interv.* 12(Pt 2), 224–231. doi: 10.1007/978-3-642-04271-3\_28
- Rojas, S., Herance, J. R., Gispert, J. D., Abad, S., Torrent, E., Jimenez, X., et al. (2013). *In vivo* evaluation of amyloid deposition and brain glucose metabolism of 5XFAD mice using positron emission tomography. *Neurobiol. Aging* 34, 1790–1798. doi: 10.1016/j.neurobiolaging.2012.12.027
- Rominger, A., Brendel, M., Burgold, S., Keppler, K., Baumann, K., Xiong, G., et al. (2013). Longitudinal assessment of cerebral  $\beta$ -amyloid deposition in mice overexpressing Swedish mutant  $\beta$ -amyloid precursor protein using 18F-florbetaben PET. *J. Nucl. Med.* 54, 1127–1134. doi: 10.2967/jnumed.112.114660
- Sawiak, S. J., Wood, N. I., Williams, G. B., Morton, A. J., and Carpenter, T. A. (2009). Voxel-based morphometry in the R6/2 transgenic mouse reveals differences between genotypes not seen with manual 2D morphometry. *Neurobiol. Dis.* 33, 20–27. doi: 10.1016/j.nbd.2008.09.016
- Schneider, L. S. (2013). Alzheimer disease pharmacologic treatment and treatment research. *Continuum (Minneapolis)* 19, 339–357.
- Snellman, A., Lopez-Picon, F. R., Rokka, J., Salmona, M., Forloni, G., Scheinin, M., et al. (2013). Longitudinal amyloid imaging in mouse brain with 11C-PIB: comparison of APP23, Tg2576, and APPswe-PS1dE9 mouse models of Alzheimer disease. *J. Nucl. Med.* 54, 1434–1441. doi: 10.2967/jnumed.112.110163
- Teipel, S. J., Buchert, R., Thome, J., Hampel, H., and Pahnke, J. (2011). Development of Alzheimer-disease neuroimaging-biomarkers using mouse models with amyloid-precursor protein-transgene expression. *Prog. Neurobiol.* 95, 547–556. doi: 10.1016/j.pneurobio.2011.05.004
- Vandenbergh, R., Van Laere, K., Ivanoiu, A., Salmon, E., Bastin, C., Triau, E., et al. (2010). 18F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. *Ann. Neurol.* 68, 319–329. doi: 10.1002/ana.22068
- von Reutern, B., Grünecker, B., Yousefi, B. H., Henriksen, G., Czisch, M., and Drzezga, A. (2013). Voxel-based analysis of amyloid-burden measured with [(11)C]PiB PET in a double transgenic mouse model of Alzheimer's disease. *Mol. Imaging Biol.* 15, 576–584. doi: 10.1007/s11307-013-0625-z
- Waldron, A. M., Wyffels, L., Verhaeghe, J., Bottelbergs, A., Richardson, J., Kelley, J., et al. (2015). Quantitative  $\mu$ PET imaging of cerebral glucose metabolism and amyloidosis in the TASTPM double transgenic mouse model of Alzheimer's disease. *Curr. Alzheimer Res.* 12, 694–703. doi: 10.2174/1567205012666150710104713
- Yakushev, I., Hammers, A., Fellgiebel, A., Schmidtman, I., Scheurich, A., Buchholz, H. G., et al. (2009). SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnesic mild cognitive impairment from healthy aging. *Neuroimage* 44, 43–50. doi: 10.1016/j.neuroimage.2008.07.015
- Zhang, W., Oya, S., Kung, M. P., Hou, C., Maier, D. L., and Kung, H. F. (2005). F-18 stilbenes as PET imaging agents for detecting beta-amyloid plaques in the brain. *J. Med. Chem.* 48, 5980–5988. doi: 10.1021/jm050166g

**Conflict of Interest Statement:** KB is an employee of F. Hoffmann-La Roche; PB received consultant fees from GE and Piramal Imaging, and honoraria from Siemens; AR received consultant fees from Piramal Imaging and GE. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Overhoff, Brendel, Jaworska, Korzhova, Delker, Probst, Focke, Gildehaus, Carlsen, Baumann, Haass, Bartenstein, Herms and Rominger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.