Construction of 4D Infant Cortical Surface Atlases with Sharp Folding Patterns via Spherical Patch-based Group-wise Sparse Representation

Zhengwang Wu[1], Li Wang[1], Inbar Fried[1], Weili Lin[1], John H. Gilmore[1], Gang Li[1*], Dinggang Shen[1,2,*]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599
[2] Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea
*Co-corresponding authors: gang_li@med.unc.edu, dgshen@med.unc.edu

1.1 Parameter selection

There are two categories of parameters involved in the atlas construction methods. One set is the *registration related parameters*, and the other set is the *method related parameters*, such as the patch size, typical patch number, neighbor size, and regularization weights.

For the registration related parameters, we determine them in a straightforward manner. Specifically, we first compute the correlation of cortical attributes (the average convexity and curvature) across the registered cortical surfaces and then select the optimal registration parameter set as the one leading to the highest correlation coefficients. Once the optimal registration parameter set is determined, we fix it in all atlas construction methods for a fair comparison.

For the method related parameters, we determine them based on a grid search in the cross-validation framework. Among the method related parameters, the patch size and the typical patch (or highly-correlated patch) number are involved in all comparison methods, while the neighbor size and the regularization weights are only involved in the sparse representation methods. Of note, we need to consider 2 criteria when selecting the parameters, i.e., the variance of the atlas to each individual subject and the sharpness. Practically, it would be too expensive to explore the entire parameter space and consider the variance and sharpness simultaneously, since there are too many combinations. Our strategy is that we first determine a candidate set for the patch size and typical patch number based on the cortical attributes variance from the registered cortical surfaces, which will lead to the optimal atlas using the top *M* patch-based method. And then, we further choose the optimal neighbor size and regularization weights for the two sparse representation methods. This strategy enables us to obtain an atlas that has the similar variance to the top *M* patch-based method. More importantly, it will effectively reduce the parameter searching space for all comparison methods, since we do not need to tune all parameters at the same time.

For the patch size, a larger one will enable the local patch to capture a larger view of the local cortical attributes. **Fig. S1** shows the 1-ring, 2-ring, and 3-ring patches, respectively, color-coded with the curvature. Given a larger patch, the atlas construction needs to maintain the local cortical attribute pattern consistency across all subjects. With richer context information from the larger patch, the constructed atlas is more robust to the local cortical attribute pattern. However, as the patch size increases, the constructed atlas is also prone to become over-smooth. Therefore, a patch size that balances the robustness and sharpness needs to be determined. In the

experiments, we empirically tested 3 candidate patch sizes, i.e., 1-, 2-, and 3-ring neighbors, since the 3-ring patch has already been large enough for capturing the local cortical attribute pattern.

The necessity of selecting the top *M* highly-correlated patches as the representation target is explained in **Fig. S2**, which shows the corresponding center patches after the two-stage registration. **Fig. S2 (a)** shows the patches with the majority cortical attribute pattern; while **Fig. S2 (b)** shows a sample patch that has less agreement of cortical attribute pattern with the majority patches, which is caused by the inevitable registration error. Note that if we augment the neighboring patches into the dictionary, e.g., the patch centered at the black point, then the sparse representation can potentially select it for the atlas construction; **Fig. S2 (c)** shows the sample patch that has less agreement with the majority cortical attribute pattern (in the current available dataset), due to the substantial inter-subject variation of the cortical attribute. We can regard the patches having less agreement with the majority patches as the noisy patches. By selecting highly-correlated patches (typical patches), we can filter out these noisy patches. This is also consistent with that the atlas patch should be a good representation of the population (majority) patches. Basically, a larger typical patch number enables the atlas to cover a wider population. However, it might also introduce the noisy patches. On the contrary, a smaller typical patch number helps to filter out noisy patches effectively, but it may lead to coverage of less individuals. In the experiment, we have tested the top 60% and 80% highly-correlated patches as the representation target for the atlas construction (note that the top 100% would be similar to the straightforward average).

The neighbor size determines the tolerance of registration errors. In **Fig. S2 (b)**, we can see that if we augment the patch centered at the black point (a 2-ring neighbor), then the dictionary will include the patch consistent with the majority cortical attribute pattern for this subject. Through the sparse representation, it can be selected for atlas construction. A larger neighbor size can potentially incorporate more patches and thus improve the tolerance to the registration errors, but also leads to a more complex representation. In the experiment, we have tried to augment the 1-, 2-, and 3-ring neighbors into the dictionary, respectively.

The two sparse representation based methods involve the regularization weights. The independent sparse representation uses the elastic net to represent each cortical attribute independently, whose regularization weights is determined by cross-validation (Wu et al., 2017). The group-wise sparse representation also uses a grid search for determining the optimal regularization weights ρ_1, ρ_2. For ρ_1, we empirically set its searching range as [0.1,1.5], and then search the optimal value with a step size 0.1; for ρ_2, we set its searching range as [0.1, 1] and the searching step size as 0.1.

Fig. S3 shows the average convexity and curvature variances of the atlases (constructed using the top *M* patch-based method) with different parameters from the infants at 12 months of age. From this figure, it can be seen that with the 2-ring patch and the top 80% highly-correlated patches, we can have the smallest variance (minimal area-under-curve). Then, with the initialized patch size and the typical patches number, we further use the grid search to obtain the optimal neighbor size and regularization weights based on the registration performance when aligning individual subjects onto the atlas. Through the comparison, we found that, with 163,842 vertices on each cortical surface, by setting the patch size as 2-ring neighbor, the typical patches number as 80% of the total patch number, the neighbor size as the 3-ring neighbors, and ρ_1=1, ρ_2=0.5, the atlas

achieves the optimal registration performance. Therefore, we choose this parameter combination for our atlas construction.

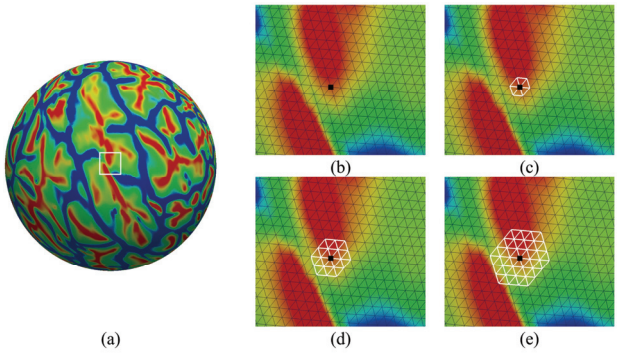

Fig. S1. Different size of patches. (a) Spherical cortical surface color-coded with the curvature. (b) The zoomed region with the black point indicates the current patch center. (c) 1-ring patch, as indicated by the white edge. (d) 2-ring patch. (e) 3-ring patch.

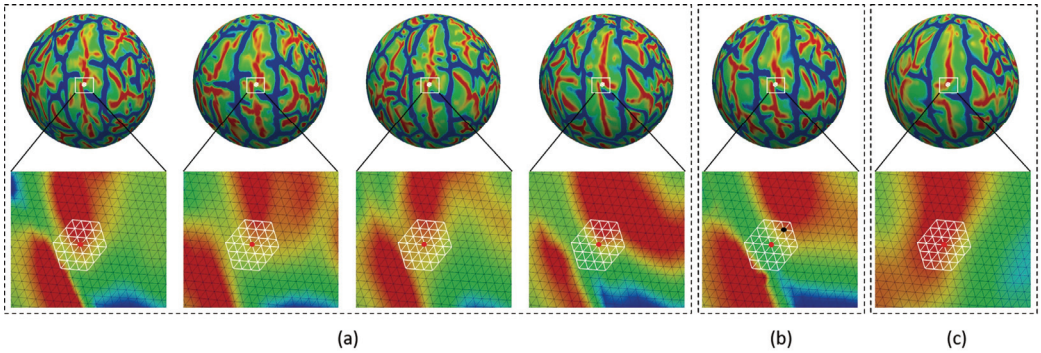

Fig. S2. Corresponding patches after registration. Red points indicate patch center. These patches are from 1-month infants after registration. (a) Patches having the majority cortical attribute pattern. (b) A patch with less agreement with the majority cortical attribute pat-tern, caused by inaccurate registration. Note that the correct corresponding patch is cen-tered at the black point, which belongs to its 2-ring neighbors. (c) A patch with less agreement with the majority cortical attribute pattern, due to the substantial inter-individual variability of cortical attribute.

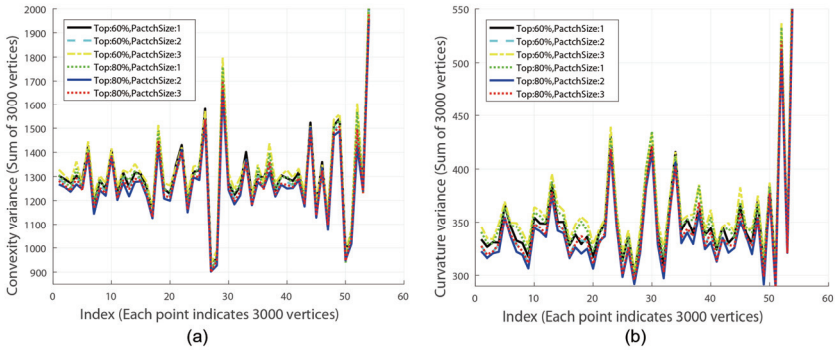

Fig. S3. The cortical attributes variance of the constructed atlases with different parameters with respect to the individual 12-month-old infants. (a) Average convexity variance; (b) Curvature variance.

1.2 Necessity of augmenting neighboring patches

Augmenting the neighboring patches into the dictionary enables our atlas construction be tolerant to potential registration errors. These augmented patches are selected during the sparse representation. Actually, this can be explained both mathematically and experimentally.

Mathematically, let us first revisit the representation model in Eq. (1). if we denote $\boldsymbol{\mu}_j(\boldsymbol{v}_i) = (\frac{1}{M})\sum_{m=1}^{M}\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}(\boldsymbol{v}_i)$, then the optimization of $\boldsymbol{W}$ is equivalent to optimize the following equation (we removed position $\boldsymbol{v}_i$ for simplicity and the proof is presented in supplementary file).

$$\underset{\boldsymbol{W}}{\operatorname{argmin}}\left[M\sum_j\left\|\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}\boldsymbol{\omega}_j-\boldsymbol{\mu}_j\right\|_2^2+\rho_1\|\boldsymbol{P}\|_{\infty,1}+\rho_2\|\boldsymbol{Q}\|_1\right] \tag{3}$$

Without losing any generality, we can denote dictionary matrix as $\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}=[\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{C})};\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{A})}]$, and also the representation vector as $\boldsymbol{\omega}_j=\begin{bmatrix}\boldsymbol{\omega}_j^{(\mathbb{C})}\\ \boldsymbol{\omega}_j^{(\mathbb{A})}\end{bmatrix}$, where $\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{C})}$ is the dictionary contains local patches from different subjects; $\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{A})}$ is the dictionary contains all neighboring patches; $\boldsymbol{\omega}_j^{(\mathbb{C})}$ is the representation weights for $\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{C})}$, and $\boldsymbol{\omega}_j^{(\mathbb{A})}$ is the representation weights for $\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{A})}$.

Therefore, we can re-write Eq. (3) as (Where $Reg(W)$ denotes the regularization term),

$$\underset{\boldsymbol{W}}{\operatorname{argmin}}\left[M\sum_j\left\|[\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{C})};\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{A})}]\begin{bmatrix}\boldsymbol{\omega}_j^{(\mathbb{C})}\\ \boldsymbol{\omega}_j^{(\mathbb{A})}\end{bmatrix}-\boldsymbol{\mu}_j\right\|_2^2+Reg(W)\right] \tag{4}$$

Without augmenting the neighboring patches into the dictionary, the above equation becomes,

$$\underset{\boldsymbol{W}}{\operatorname{argmin}}\left[M\sum_j\left\|[\boldsymbol{\mathcal{D}}_{\mathcal{M}_j}^{(\mathbb{C})};\boldsymbol{0}]\begin{bmatrix}\boldsymbol{\omega}_j^{(\mathbb{C})}\\ \boldsymbol{0}\end{bmatrix}-\boldsymbol{\mu}_j\right\|_2^2+Reg(W)\right] \tag{5}$$

Comparing Eq. (4) to (5), we can easily see that for any optimal solution of Eq. (5), we can always find a solution of Eq. (4), which has lower or at least equal cost. Because the optimal solution of Eq. (5) is included in the solution of Eq. (4). In the worst case, they will have the same cost, which can be regarded as the upper bound of Eq. (4). Therefore, augmenting neighboring patches will at least not decrease the quality of the atlas.

Experimentally, we have also quantitatively compared the spatial normalization performance of the atlas constructed without augmenting neighboring patches. We still choose the top 80% typical patches with the 2-ring patch to maintain the initial lowest variance. Then, we optimize the regularization weights by checking the performance of registering the subjects from the testing sets onto the constructed atlas. The quantitative registration performance is evaluated by the average information entropy of the sulcal and gyral regions after aligning the testing cortical surfaces onto the atlas. The result is reported in **Table S1**, which shows the atlas constructed without neighboring patch augmentation gets consistently inferior performance, even worse than the atlas constructed by independent sparse representation (comparing to Table 2).

Table S1. The average information entropy with and without neighboring patch augmentation.

Testing set 1											Testing set 2		
01	03	06	09	12	18	24	36	48	60	72	01	12	24
0.403	0.392	0.391	0.393	0.382	0.398	0.360	0.327	0.408	0.405	0.356	0.427	0.439	0.446

1.3 Proof of optimization equivalence of Eq. (1) and (3).

Denote $\boldsymbol{\mu}_j = (\frac{1}{M}) \sum_{m=1}^{M} \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}$, then

$$\sum_{m=1}^{M} \left\| \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j - \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \right\|_2^2 = \sum_{m=1}^{M} \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j - \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}, \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j - \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle$$

$$= \sum_{m=1}^{M} \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j \rangle - 2 \sum_{m=1}^{M} \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle + \sum_{m=1}^{M} \langle \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}, \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle$$

$$= \sum_{m=1}^{M} \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j \rangle - 2 \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \sum_{m=1}^{M} \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle + \sum_{m=1}^{M} \langle \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}, \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle$$

$$= M \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j \rangle - 2M \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \boldsymbol{\mu}_j \rangle + \sum_{m=1}^{M} \langle \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}, \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle \tag{6}$$

Note, if we let

$$var\left(\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}\right) = \sum_{m=1}^{M} \left\| \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} - \boldsymbol{\mu}_j \right\|_2^2 = \sum_{m=1}^{M} \langle \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} - \boldsymbol{\mu}_j, \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} - \boldsymbol{\mu}_j \rangle$$

$$= \sum_{m=1}^{M} \langle \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}, \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle - M \langle \boldsymbol{\mu}_j, \boldsymbol{\mu}_j \rangle$$

Obviously, once $\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}$ is selected, $var\left(\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}\right)$ is a constant, and

$$\sum_{m=1}^{M} \langle \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}, \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \rangle = M \langle \boldsymbol{\mu}_j, \boldsymbol{\mu}_j \rangle + var\left(\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}\right) \tag{7}$$

Based on Eq. (6) and (7), we have,

$$\sum_{m=1}^{M} \left\| \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j - \widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)} \right\|_2^2 = M \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j \rangle - 2M \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j, \boldsymbol{\mu}_j \rangle + M \langle \boldsymbol{\mu}_j, \boldsymbol{\mu}_j \rangle + var\left(\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}\right)$$

$$= M \langle \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j - \boldsymbol{\mu}_j, \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j - \boldsymbol{\mu}_j \rangle + var\left(\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}\right) = M \left\| \boldsymbol{\mathcal{D}}_{\mathcal{M}_j} \boldsymbol{\omega}_j - \boldsymbol{\mu}_j \right\|_2^2 + var\left(\widehat{\boldsymbol{p}}_{\mathcal{M}_j}^{(m)}\right)$$

So, the optimization of $\boldsymbol{W}$ on **Eq. (1)** and **(3)** is equivalent.

1.4 HCP parcellation warping

To check the quality of the warped parcellation, we have presented HCP MMP parcellation on both the original HCP cortical surface atlas and our constructed atlas. From **Fig. S4**, it can be seen the ROI regions are quite consistently overlaid on both the original HCP cortical surface atlas and our constructed surfaces atlas.

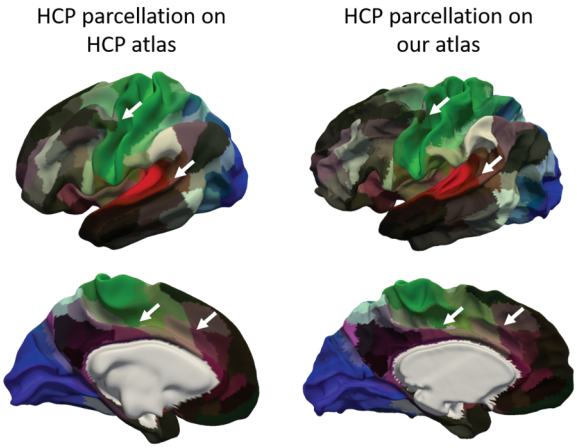

Fig. S4. A comparison figure of the HCP MMP parcellation on the original HCP atlas and our atlas.

1.5 Variance of the constructed atlas

Besides the cortical attribute patterns on the atlas, we also present the variance of the cortical attributes on our atlas in **Fig. S5,** which is a reference for better inspecting the consistency of the cortical attributes across different subjects and different time points.

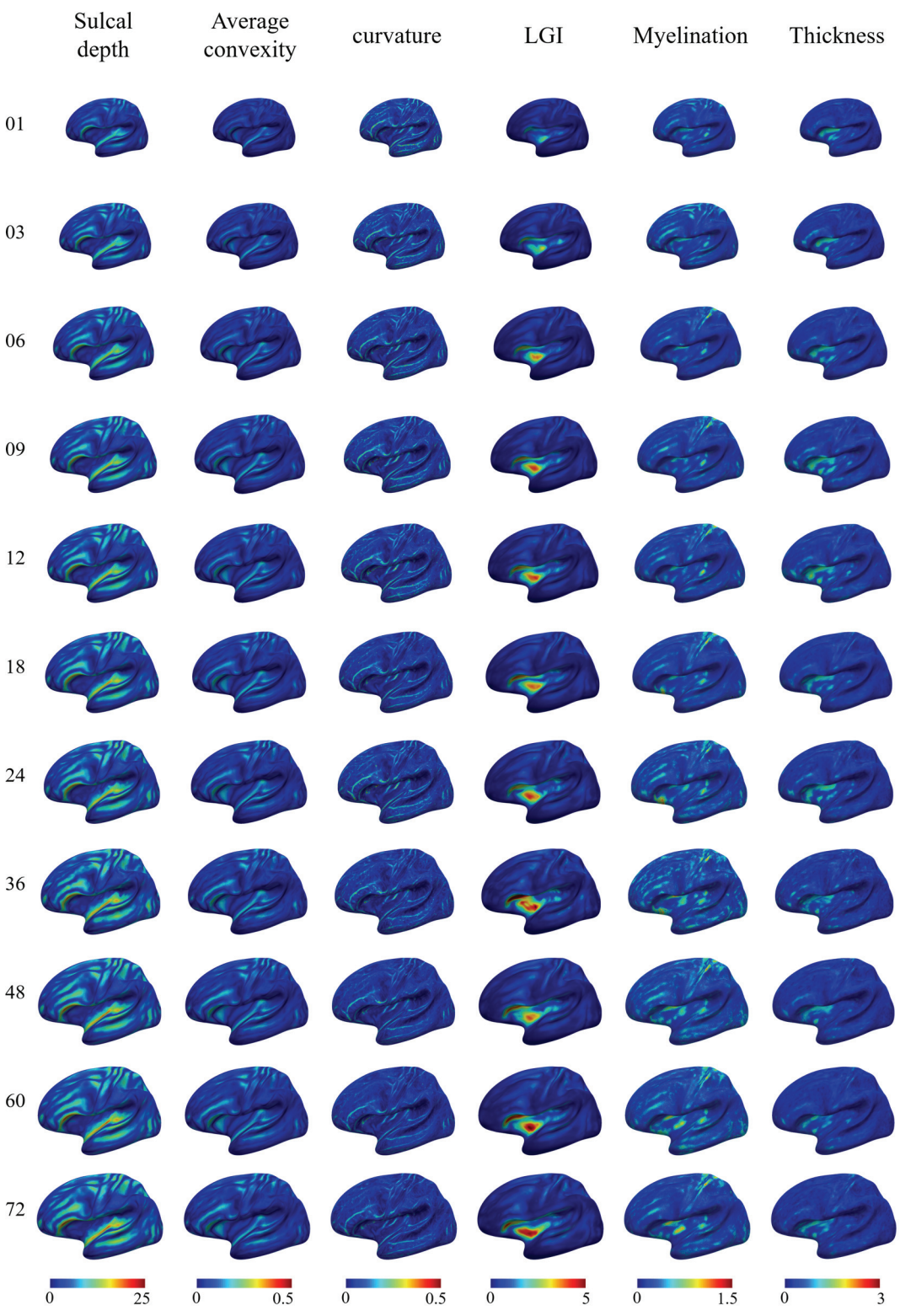

Fig. S5. The cortical attributes variance of the constructed cortical surface atlas.