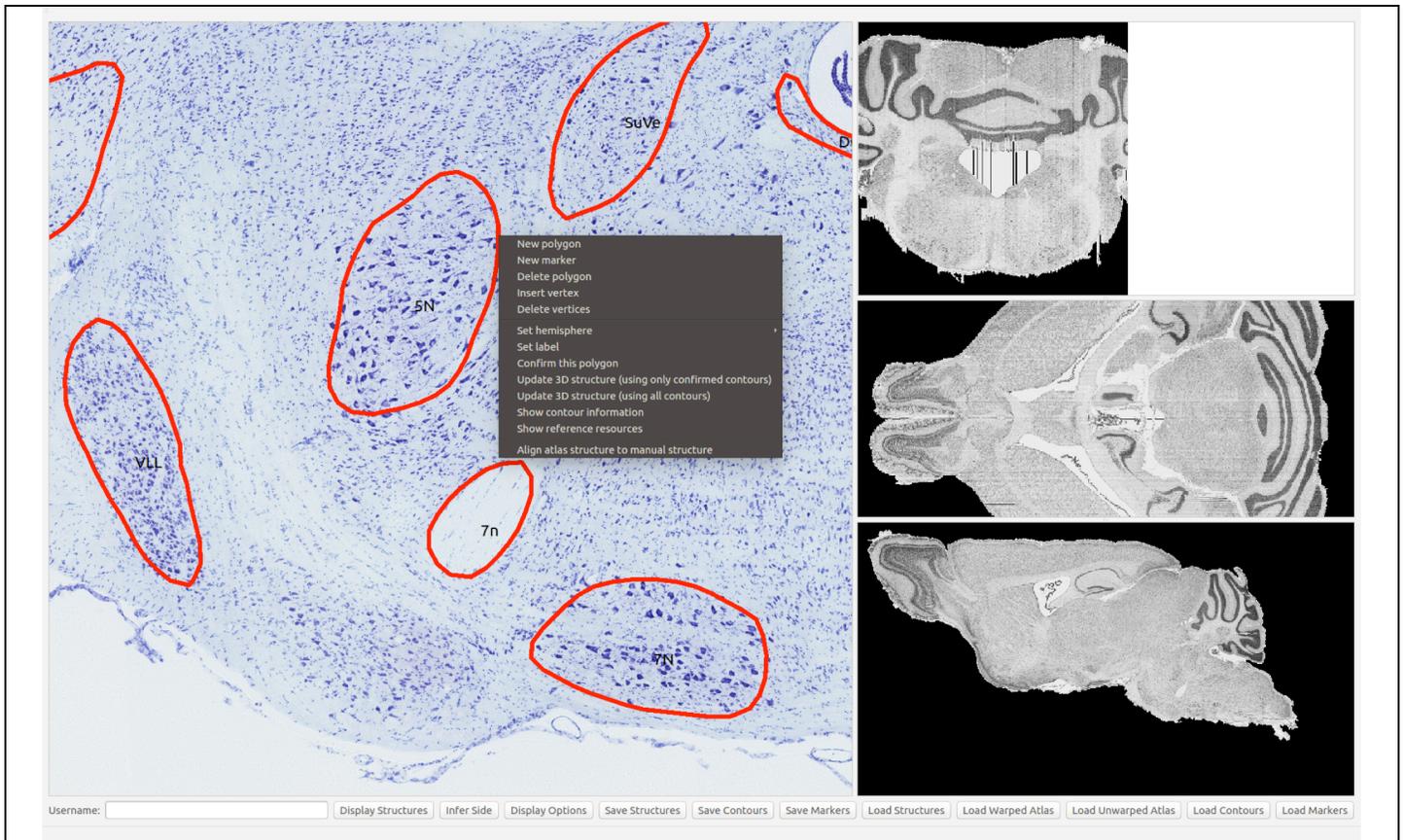


**Supplementary Figure 1**

Reconstruction of one expertly annotated brain out of three used to bootstrap the atlas.

(a) Two-dimensional contours mapped to three-dimensional subject space. The gray level volume is shown for reference.

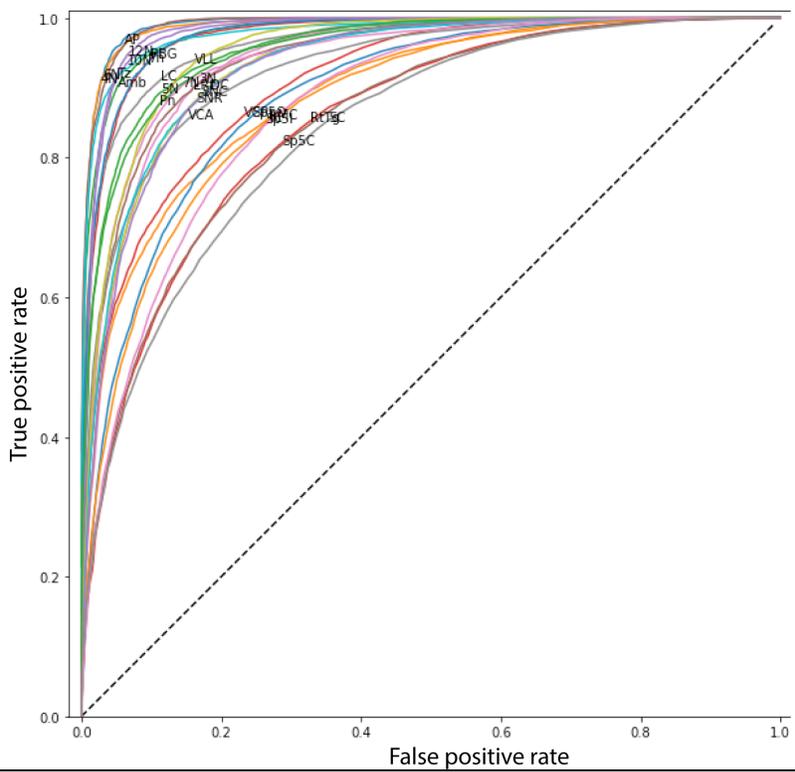
(b) Reconstructed landmark structures.



## Supplementary Figure 2

Graphical interface for annotation of landmarks.

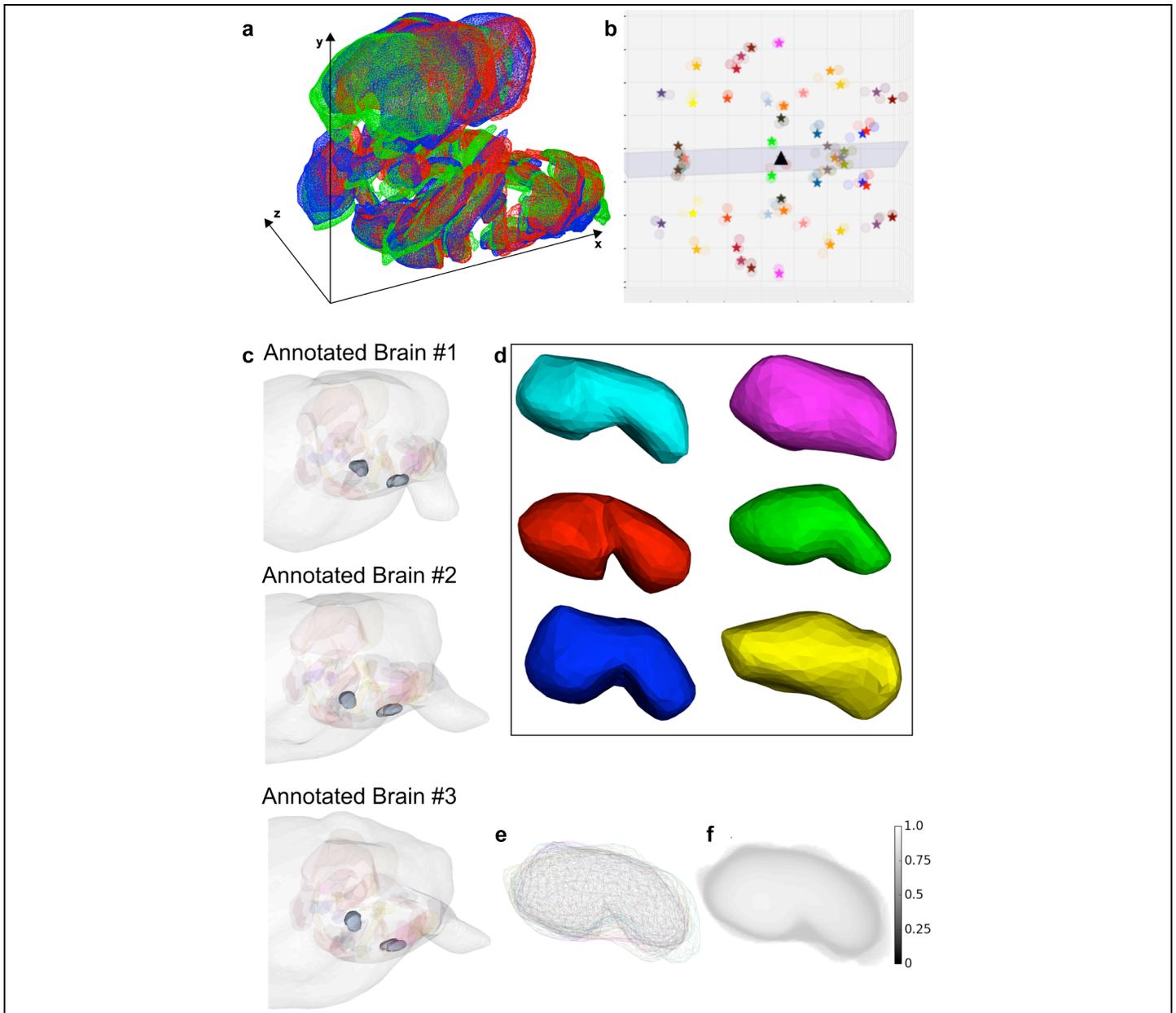
The main panel shows the original full-resolution section image and structure contours. The side panels show virtual sections of the reconstructed gray level volume in three orthogonal planes. All panels are synchronized.



**Supplementary Figure 3**

Classification performance.

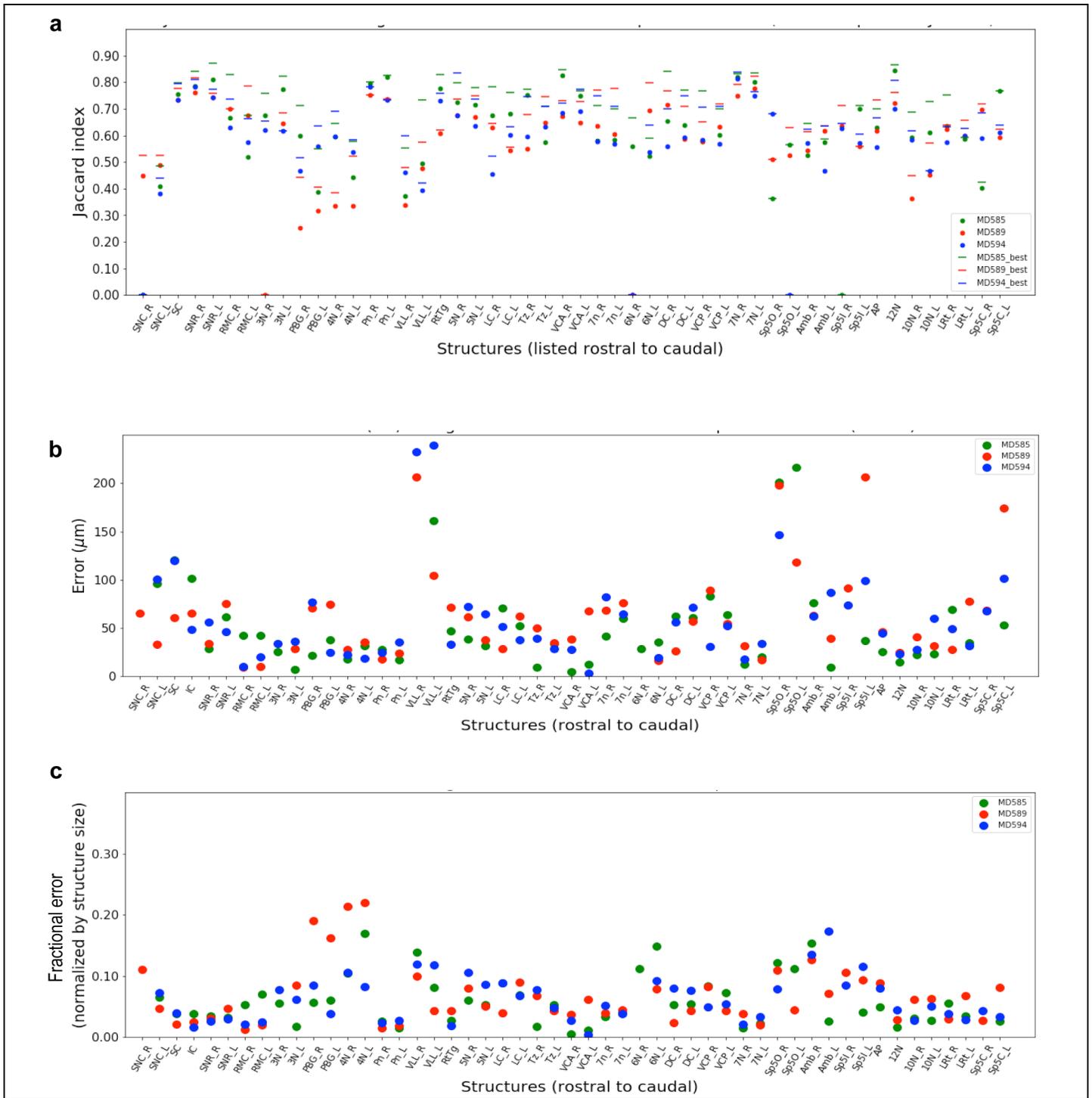
The receiver operator characteristics (ROC) curves for the classifiers of different structures for one particular combination of training brain and one test brain for all 51 landmarks. The ROC measures classifier performance for each landmark.



#### Supplementary Figure 4

Estimation of locations and shapes for the bootstrapped reference atlas.

- (a) Illustration of three annotated brains (blue, green, red) brought in registration in atlas space by global affine transforms.
- (b) Top-down view of the atlas space. Each structure is represented by a color. Circles are instance centroids. Stars are the nominal centroids. Shaded plane is the common mid-sagittal plane. Note the symmetry of the nominal centroids of paired structure with respect to the mid-sagittal plane.
- (c) Reconstructions of three annotated brains with facial motor nucleus (7N) in both sides highlighted.
- (d) An example of the alignment of one landmark using all six instances of the facial motor nucleus; one from each hemisphere.
- (e) All six instances of the facial motor nucleus aligned using rigid transforms.
- (f) Probabilistic average shape of the facial motor nucleus obtained by voxel averaging.

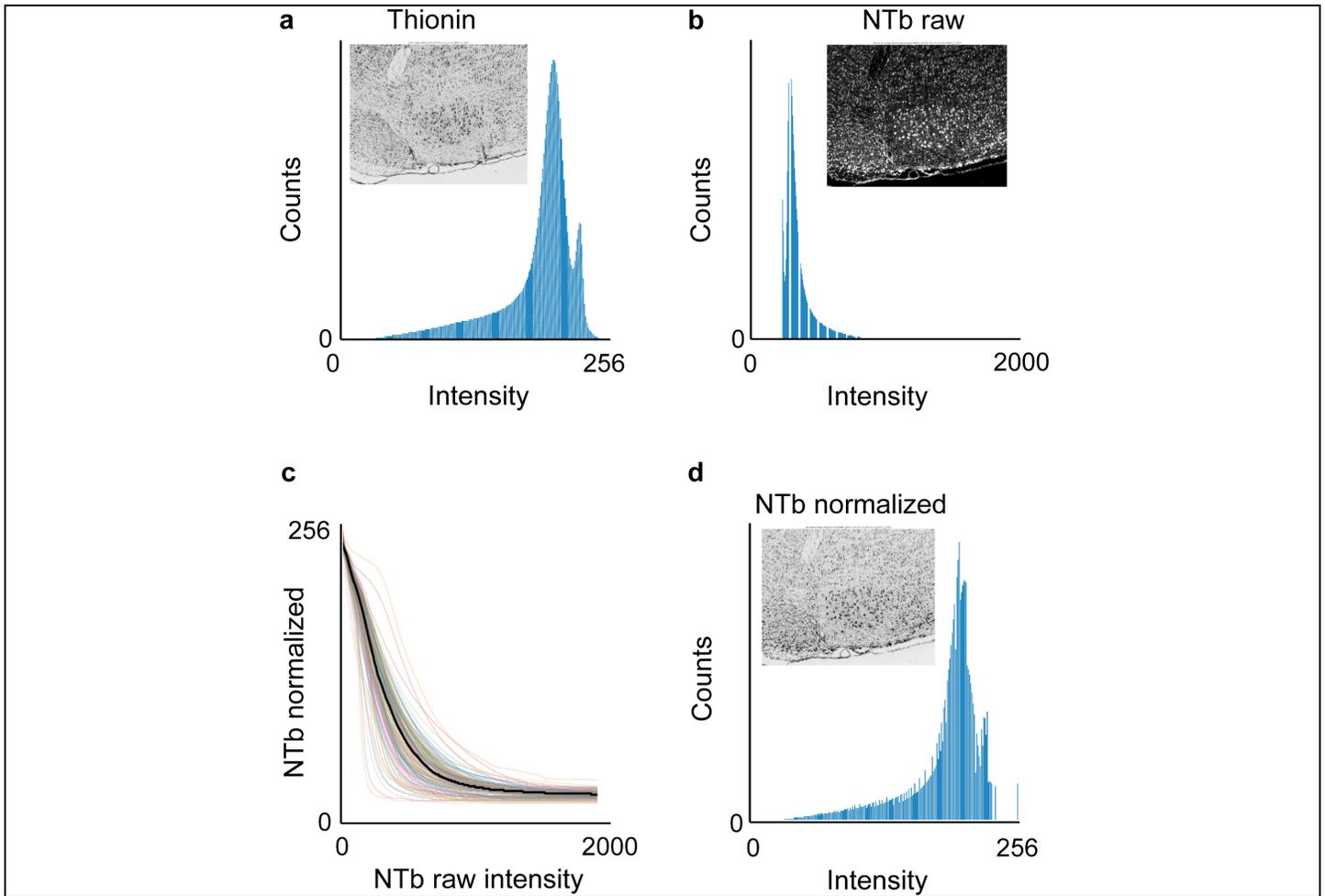


**Supplementary Figure 5**

Consistency between automatic and manual annotation. See Supplementary\_Figure\_5\_Source\_Data for listing of all data points.

(a) Jaccard index, which ranges from 0 for completely disjoint to 1 for an exact overlap.

(b,c) The deviation of the centroid of a registered landmark from the expert annotation in absolute units (panel b) and normalized to the size of the landmark (panel c).



### Supplementary Figure 6

Mapping of fluorescent intensities from Neurotrace blue stained sections to thionin sections.

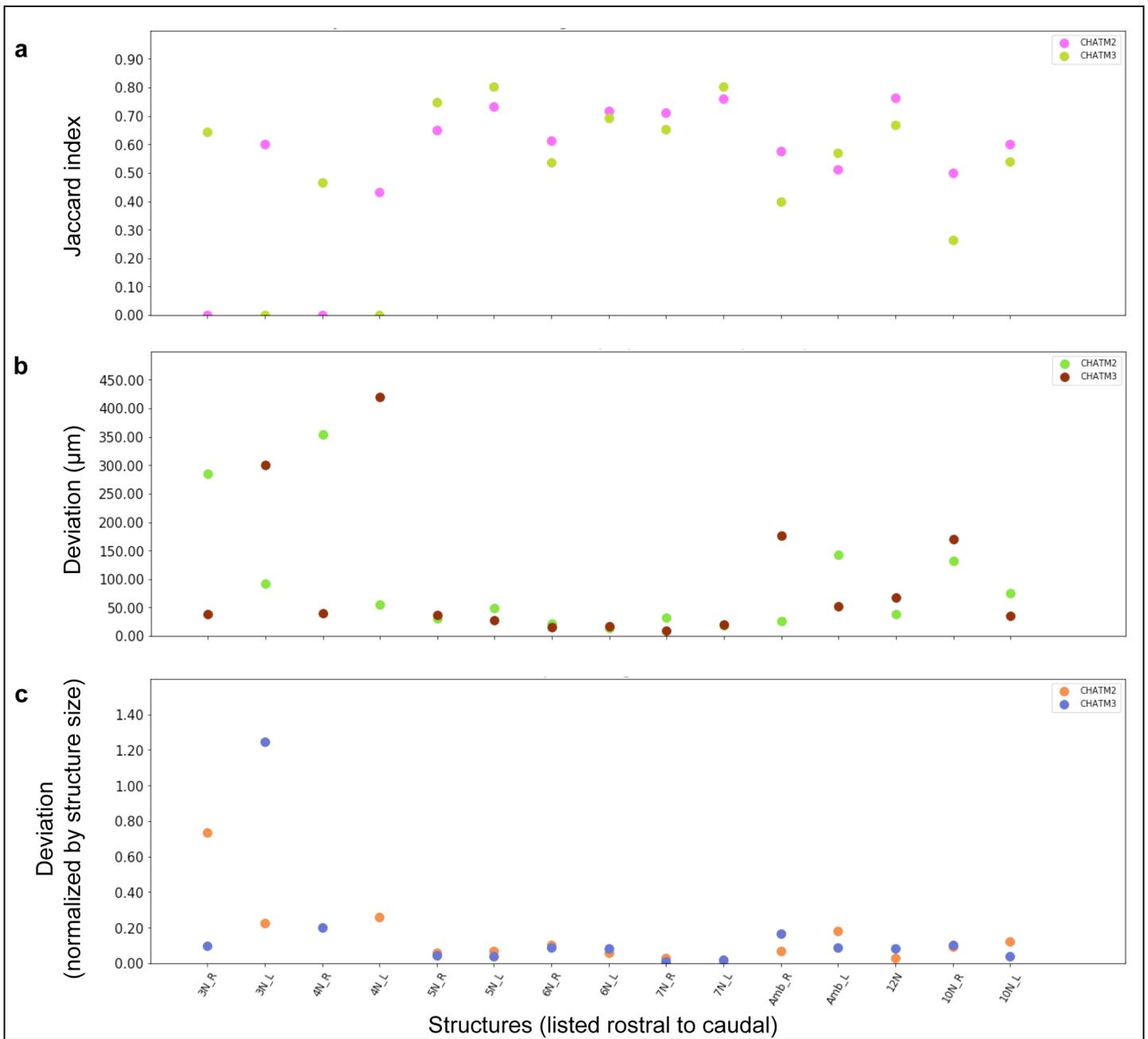
Data was obtained from brains with alternate sections stained with Neurotrace blue versus thionin. The sections are rigidly aligned by *Elastix* with mutual information as criteria. We then randomly sample 10 pairs of regions, each 500  $\mu\text{m}$  by 500  $\mu\text{m}$ , from many pairs of adjacent sections and match the intensity histograms of corresponding regions. We match histograms of moderately-sized regions rather than entire images because the global tissue content is likely to be different even for adjacent sections, while using regions of limited extent reduces this variance.

(a) Histogram of the pixel intensity of a region from a thionin section; histology shown as an inset.

(b) Histogram of the pixel intensity of a region from a Neurotrace blue section at the same level in the brain; histology shown as an inset.

(c) The estimated nonlinear mapping between the intensities of Neurotrace blue to those of thionin. We collected 1000 such curves from five brains. The thick black line is for the section in panels a and b, while the other lines are for other pairs of sections.

(d) The histogram and image of the Neurotrace blue data in panel b after correction.



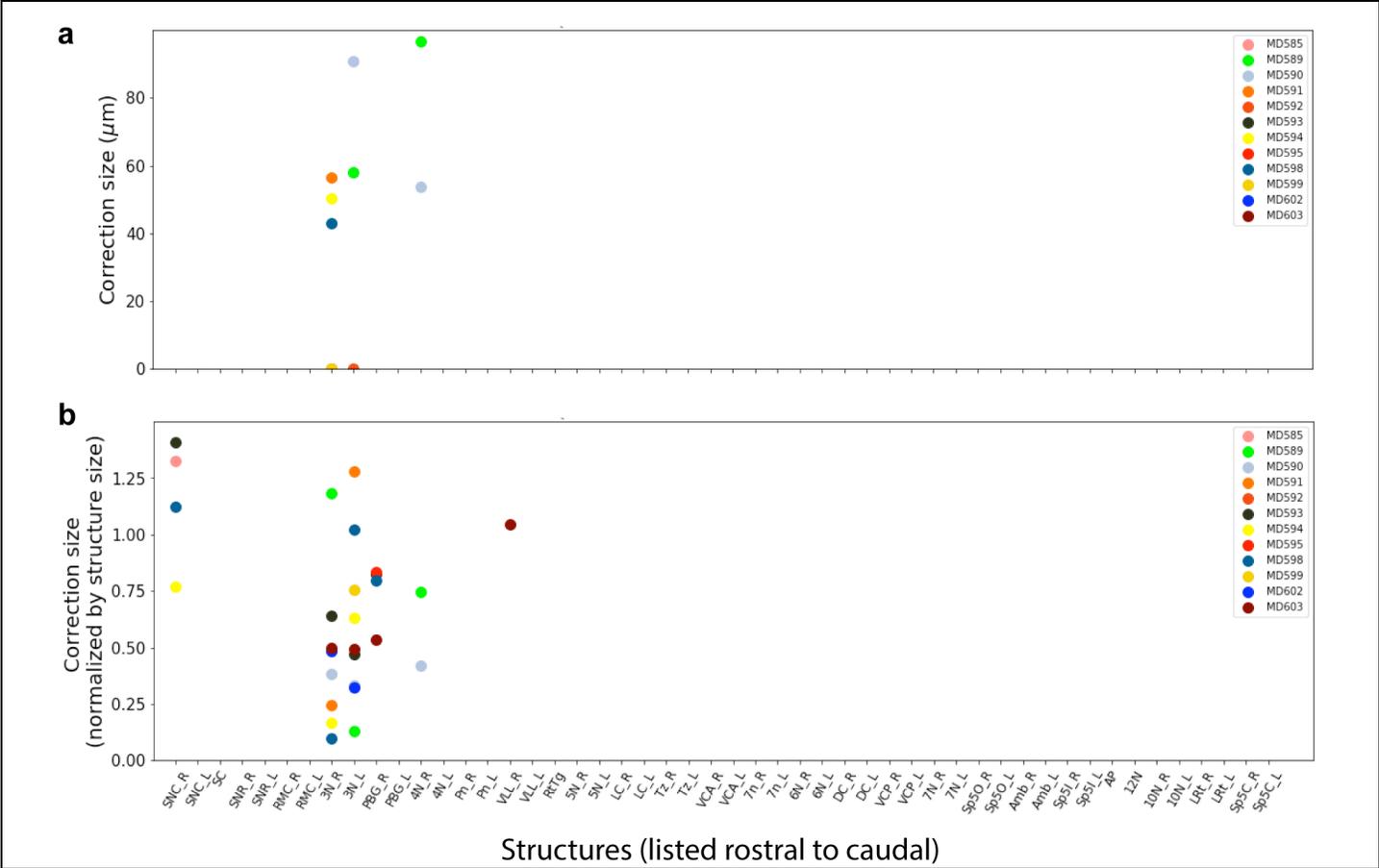
**Supplementary Figure 7**

Consistency between ChAT- and texture-based annotation. See Supplementary\_Figure\_7\_Source\_Data for listing of all data points.

This analysis is over two ChAT brains.

(a) Jaccard index.

(b,c) The deviation of the centroid of a registered landmark from the expert annotation in absolute units (panel b) and normalized to the size of the landmark (c).

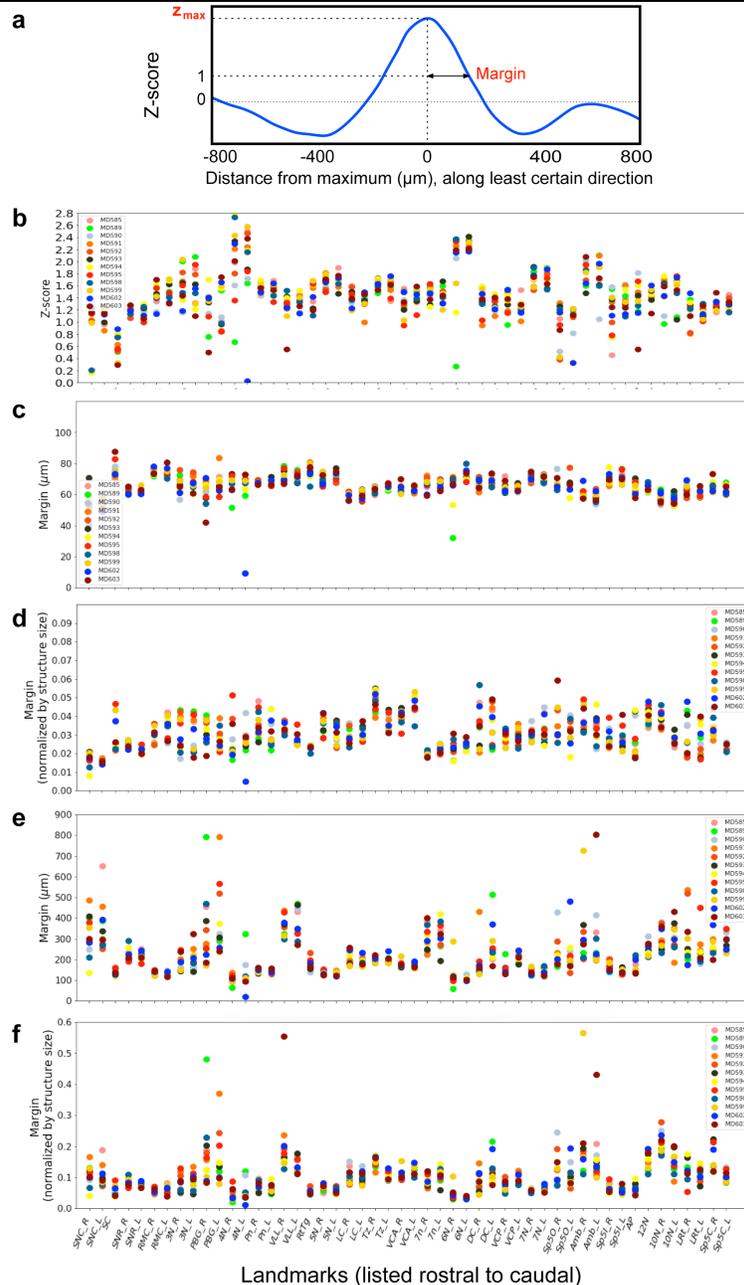


**Supplementary Figure 8**

Quantification of human correction. See Supplementary\_Figure\_8\_Source\_Data for listing of all data points.

This analysis is over thirteen brains. Only selected landmarks required corrections.

- (a) Corrections in absolute units.
- (b) Corrections normalized to the size of the landmark.



### Supplementary Figure 9

Measures of registration confidence. See Supplementary\_Figure\_9\_Source\_Data for listing of data points in panels b-f.

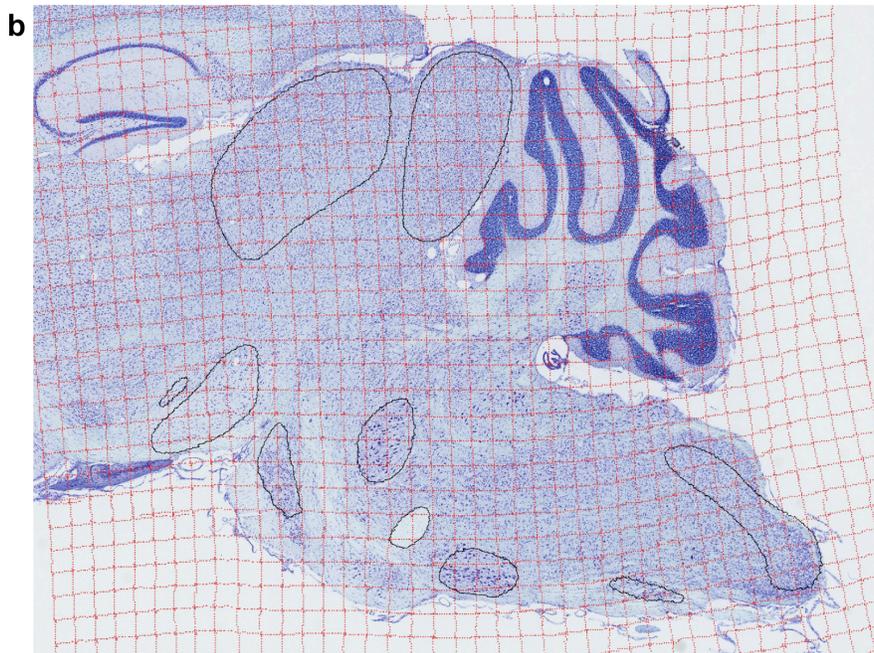
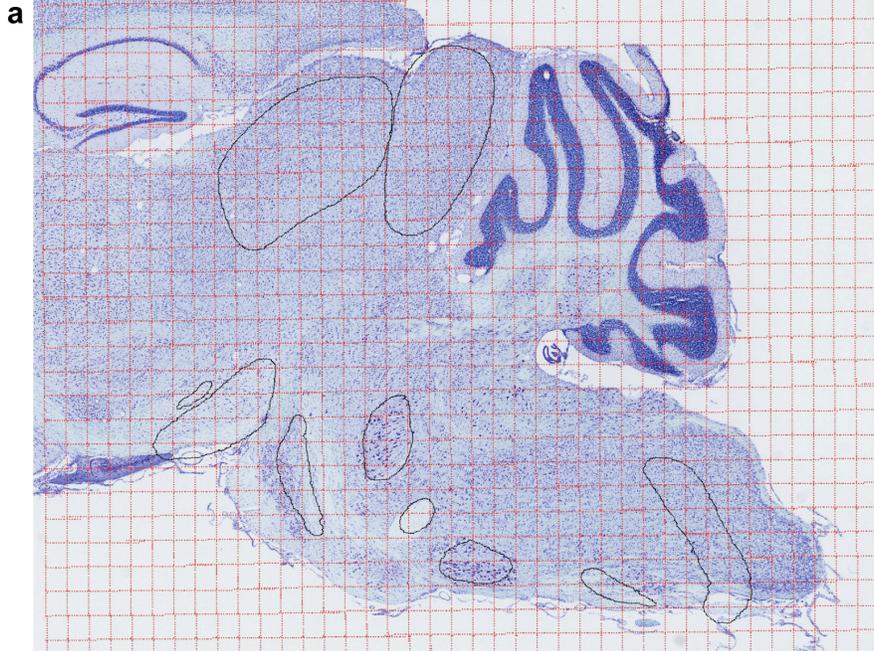
(a) Landscape of the objective function for a particular registration in one mouse. Magnitude is normalized to yield z-scores. Significance metrics are the z-score of the estimated maximum, and the margin, i.e., the distance from the maximum where z-score drops to unity.

(b) The z-scores of all structure-specific registrations across 12 mice, as indicated.

(c,d) Lower bound of all structure-specific registrations in absolute (panel c) and normalized (panel d) coordinates across 12 mice, as indicated.

(e,f) Upper bound of all structure-specific registrations in absolute (panel e) and normalized (panel f) coordinates across 12 mice, as indicated.





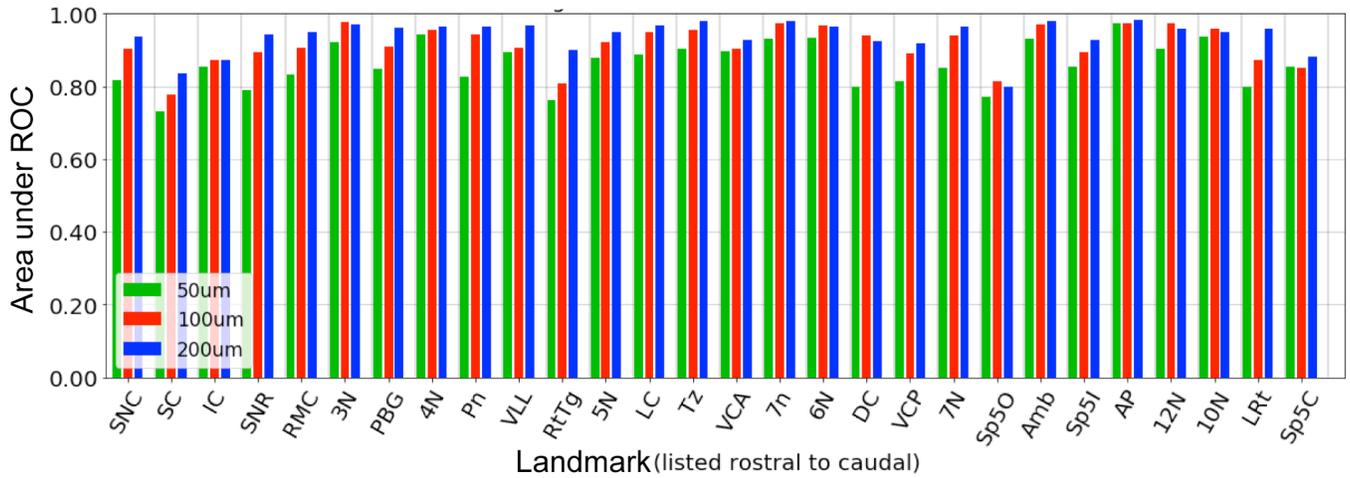
**Supplementary Figure 11**

Deformation fields derived from global registration and landmark specific registration for an example section.

Contours are the cross-sections of 0.5-level iso-surfaces of the aligned atlas structures. Grid lines represent the transformed result of a regular grid defined in atlas space.

(a) Results after global registration. Structures are placed reasonably close to the correct positions, but individual adjustment is still necessary. Grid lines exhibit an affine transformation.

(b) Results after structure-specific registration. Structure pose and locations are improved. Warped grid lines demonstrate the final deformation field.



**Supplementary Figure 12**

Results for the area under the receiver operator curve (ROC) for three different patch sizes. See Supplementary\_Figure\_12\_Source\_Data for listing of all data points.

Classifiers were trained using two brains, MD585 and MD589, and the accuracy was measured against a third brain, MD594.