SUPPLEMENTARY INFORMATION

Supplementary Methods

Animals

Animal procedures were approved by the Cold Spring Harbor Laboratory Animal Care and Use Committee. All animals were housed under constant temperature and light conditions (12 hour cycle lights ON: 0600, lights OFF: 1800) and given food and water *ad libitum*. The c-fos-GFP mice were obtained from the Jackson Laboratory as a double transgenic strain B6;DBA-Tg(Fos-tTA,Fos-EGFP*)1Mmay Tg(tetO-lacZ,tTA*)1Mmay/J, (stock number 008344). The mice were bred with C57BL/6 to remove the tetO-lacZ,tTA* transgene and then continued to backcross to C57BL/6 for >10 generations as Tg(Fos-tTA,Fos-EGFP) line, which, for simplicity, we call c-fos-GFP mice in our study. These mice comprise two transgenes integrated in the same genomic site: Fos-tTA driving the expression of the tetracycline transactivator (tTA) and Fos-EGFP driving c-fos-GFP fusion protein. The c-fos-GFP transgene includes all 4 exons and all introns of the *c-fos* gene, with the EGFP sequence fused in-frame at NcoI site to the exon 4; this design was successfully used in two other independently generated c-fos reporter transgenic mice (Barth et al., 2004; Schilling et al., 1991; Wilson et al., 2002). In our study we used the direct c-fos-GFP signal, whereas several other studies used the tTA protein to drive other reporter molecules (Garner et al., 2012; Liu et al., 2012; Matsuo et al., 2008; Reijmers et al., 2007). The comparison between c-fos-GFP expression visualized by STP tomography and native c-fos expression visualized by anti-c-fos immunohistochemistry was used to validate the social interaction-evoked c-fos-GFP induction in several brain region (Figure S5).

Behavioral tests

Heterozygous c-fos-GFP male mice were group-housed before the test. One week before experiments, adult (8 – 11 week old mice) were transferred to a designated area in the animal room and separated single per cage to lower the variability in baseline c-fos expression, as described (Ferguson et al., 2001; Ferguson et al., 2000). Note that the one-week adult isolation is too brief to evoked chronic behavioral and neuroendocrine stress responses (Lupien et al., 2009). All experiments were done between 11 am and 12:30 pm, and the mice were killed between 2 pm and 3:30 pm. The behavioral stimuli were: transfer of the animal to the experimental arena (handling control) or transfer of the animal plus introduction of an OVX conspecific female (male-female group), conspecific male (male-male group), 50 ml falcon tube (object group), and 50 ml falcon tube with a side-opening in which was cotton ball with isoamyl acetate (1:100 in mineral oil, 40 μl per experiment, freshly made each day) (olfactory group) (Yang and Crawley, 2009). The stimulus was placed in the home cage for 90 seconds and then removed. The OVX female mice and intruder males were 2-4 month old; the c-fos-GFP male mice were 2-3 month old. Behavioral response of c-fos-GFP male mice to the OVX females were also compared to wild-type females (Figure S4). The behavior was video-recorded and was manually scored off-line by Avidemux for the time spent in active social interaction (close following, anogenital sniffing, nose-to-nose touch, and other social sniffing) in the social groups (Figure S4; Movie S2). After the behavioral stimulus was removed, the

mice remained in the home cage for additional 3 hrs and then killed by transcardial perfusion with 0.1 M phosphate-buffered saline (PB) followed by 4% paraformaldehyde (PFA) in 0.1 M PB. Since the handling group and the object were not statistically different, we used only the handling group for comparison to the male-female, male-male and object+odor groups.

Brain preparation and STP tomography.

The brains were prepared as described in our previous study (Ragan et al., 2012). Briefly, after perfusion the brains were postfixed overnight in 4% PFA at 4 ^{o}C, then kept for 48 hrs in 0.1 M glycine / 0.1 M PB, and stored in 0.05 M PB until imaging. The brains were embedded in oxidized 4% agarose in 0.05 M PB using a custom built holder to maintain consistent embedding position. The embedded brains were crosslinked in 0.2% sodium borohydrate solution (in 0.05 M sodium borate buffer, pH 9.0-9.5) and imaged as 280 serial sections, each comprised of a mosaic of 12 x 16 FOVs (X-Y 700 x 700 μm). The raw image tiles (16 bit tif; 70 GB per dataset) were corrected for illumination, stitched in 2D in matlab and aligned in 3D in Fiji (Ragan et al., 2012). Volocity (Perkin-Elmer) was used to visualize the whole brains and cell counts in 3D.

Automated c-fos-GFP+ cell counting

The selection of the CNs for detection of c-fos-GFP+ cells was done by training on ground truth data marked up by an expert biologist. The ground truth data comprised 72 FOVs randomly selected from a whole-brain dataset.

The CN configuration of 3 hidden layers (each layer was 20 units wide, the output of the network was 1 pixel and the filters were 5 x 5 in size) was chosen after training multiple CNs with different sized filters and different number of units in each layer (Figure S1). The ground truth data set was then divided into 6 folds (i.e. 5 training sets, each with 60 FOVs; the remaining 12 FOVs were assigned as a test set). Five CNs were trained with each training set. The threshold for each CN output was varied from 0.90 to 0.99 in steps of 0.01 and the threshold value with the maximum F-score was chosen as the threshold for that network. Of the five CNs, the CN with maximum F-score on its test data was chosen for the analysis in the current study. The CN training was done using Cortical network simulator (CNS) (Mutch et al., 2010). The training was accelerated on a NVIDIA GPU.

The CN performance was scored based on the F-score on a dataset of 10 FOVs marked by three experts, which also served to evaluate human inter-expert variability (F score = the harmonic mean of the precision and recall, where precision is the ratio of correctly predicted cells divided by all predicted cells and recall is ratio of correctly predicted cells divided by ground true positive cells; ~1100 c-fos-GFP+ cells were marked). The commonly marked up cells agreed by all experts were used as a ground truth data to score the CN performance. To calculate inter-experts variability, each expert mark-up was set as a ground truth to score the other two experts and the averaged recall and precision from the three comparisons was used as the final inter-experts F score.

In the CN output images, signal smaller than 40 μm^2 was removed as noise and single c-fos-GFP+ cells were identified as circles of radii 4 to 14 μm. In this study, we did not analyze c-fos-GFP induction in the cerebellum by CN, because of a high a false positive rate due to a cellular autofluorescence specific to this brain region. We have

analyzed cerebellum by visual inspection and detected only a few c-fos-GFP+ cells in either the social or object groups, suggesting a lack of c-fos induction in this brain region in our experiments (data not shown). In addition, the analysis of c-fos-GFP in the olfactory bulb was done using a separate CN trained specifically on OB images. This was because OB granule cells are tightly packed and smaller in diameter, which makes the OB signal not well comparable to the rest of the mouse brain.

For CN data analysis the brightness of the signal of each sample was normalized by the mean and standard deviation of tissue autofluorescence signal from a coronal section at a bregma position of +0.20 mm.

c-fos-GFP density calculation

The following stereological procedure (Williams and Rakic, 1988) was used to generate the 2D-to-3D conversion ratio for calculating the densities of c-fos-GFP+ cells per activated ROIs. First, one brain was imaged at *xyz* resolution 1 x 1 x 2.5 μm (i.e. as a 5600 serial section dataset) and c-fos-GFP+ cells were manually counted in 3D in fourteen "counting boxes" of 300 x 300 x 50 μm (*xyz*) randomly selected from the whole brain. Second, the 2D-to-3D conversion was derived by dividing the manual 3D counts by single 2D counts measured by the CN from the middle section in each box. Third, the obtained conversion factor of 2.5 was used to multiply the 2D ROI counts in order to estimate the total numbers of c-fos-GFP+ cells, and the total counts were divided by the ROI volumes in order to estimate the densities of c-fos-GFP+ cell.

c-fos immunohistochemistry and manual c-fos-GFP+ cell counting

Wild type C57BL/6 mice (8 to 10 week old) underwent the same behaviors as the c-fos-GFP mice of the social (OVX female) and handling groups. The mice were killed and perfused 1 hour later and the brains were fixed O/N (overnight) in 4% PFA, then cut as 50 micron coronal sections and stored in cryoprotectant (30% ethylenglycol and 25% glycerol in 0.05 M PB) at -20 °C. For immunohistochemistry, sections were washed 3 x 10 min in 0.1 M PBS, incubated in 1.5% H_2O_2 in 0.1 M PBS for 15 min, washed 3 x 10 min in 0.1 M PBS, incubated in PBS+ (10% donkey serum and 0.3% triton X-100) for 1 hour, and exposed to rabbit anti-c-fos antibody (1:10000, Santa Cruz SC052) O/N at 4 °C. The following day, the sections were washed 3 x 10 min in 0.1 M PBS, exposed to secondary biotinylated donkey anti-rabbit antibody (1:500) in PBS+ for 1 hour, washed 4 x 10 min in 0.1 M PBS, incubated for 1 hour in elite ABC mixture (6 ml of solution A and B in 1 ml of PBS+, Vector Labs), washed 4 x times 10 min in 0.1 M PBS, exposed to DAB solution (10 mg DAB; Sigma D5905 in 20 ml 0.1 M PBS) for 4 minutes, rinsed with 0.1 M PBS for 1 min, washed 3 x 10 min in 0.1 M PBS, and mounted on slides. Dried section on slides was dehydrated by 2 min each in 50%, 70%, 80%, 90%, 95%, 100% of EtOH, and xylene, and cover-slipped. Sections were imaged by light microscopy (Leica Axiovision, 10x lens). For c-fos IHC cell counting per area, we first used FIJI (imageJ) trakEM2 to segment each anatomical ROI and used Volocity (Perkin-Elmer) for cell counting. For the STP images, we also used FIJI to segment corresponding c-fos IHC analyzed ROI and used CNs detection for cell counting.

Time course of c-fos-GFP induction

Freshly prepared isoamylacetate (1:100 in mineral oil, 40 μl) was added to gauze in 50 ml conical tube with side opening, which was introduced into the mouse home cage for a brief period of 90 sec. Once the tube was removed, the mice were left undisturbed until they were sacrificed by transcardial perfusion at selected time points of 0.5, 1.5, 3 and 5 hours post ISO stimulation. Olfactory bulbs were imaged with STP tomography and c-fos-GFP+ cells were quantified in the MOB granular cell layer to examine the time course of c-fos-GFP induction (Figure S5).

Correlation between c-fos-GFP cell counts and social behavior

Significantly activated brain regions from the male-female and male-male versus the handling group comparisons (with FDR $q < 0.01$ as cutoff) were selected for the correlation analysis. First, Pearson correlation R values were calculated between c-fos-GFP cell counts and time spent in social behaviors, including anogenital sniffing, close following, nose-to-nose sniffing, and other sniffing. Second, the R values were transformed to one-tailed *p*-value, which were corrected by FDR for multiple comparison correction

3D brain registration

3D registration methods were the same as described (Ragan et al., 2012), but with modified parameters. The affine transform was calculated using 4 resolution levels, while the B-spline step used 3 resolution steps. Mattes Mutual information was used as the similarity measure between the moving and fixed images. The image similarity function was estimated and minimized for a set of randomly chosen samples with the images at each resolution in an iterative way. The registration takes 1 hour on 650 × 450 × 300 voxel sized images on 8-core central processing unit (CPU) with 16 Gb RAM. The images involved in the registration have a 20 μm × 20 μm × 50 μm pixel spacing. The entire image warping experiment is set up using Elastix (Klein et al., 2010), an image registration toolbox based on Insight's ITK. The precision of the registration was measured by the displacement of 13 landmark points in 6 different mouse brains after warping each dataset onto the average RSTP brain (Figure S2).

Statistics

Power analysis. A region of interest (ROI) count is defined as the sum of c-fos-GFP cells within its boundary. For every ROI, we used the CN c-fos-GFP measurements of the experiment samples to estimate the maximum likelihood parameters (μ = mean and θ = shape) of a negative binomial distribution fitted to its count (McCullagh and Nelder, 1989; Venables and Ripley, 2002). These served as the starting points for our Monte Carlo simulations. For each set of estimated parameters, we generated two datasets of a sample size *n*, the first from a negative binomial with parameters (μ, θ), and the second with parameters ($e^*\mu$, θ), where *e* is a scaling factor quantifying the effect size being introduced. For every region, we repeated this 30,000 times while modifying the effect size over the range 0.1 to 1.5. The power of a statistical test is defined as the probability of achieving a significant result given that the null hypothesis is false. We estimated it in the following way. For every simulated dataset we applied our statistical test (defined below) to obtain a p-value. If the p-value was below our selected significance level($\alpha <$ 0.05), the test result was deemed significant and assigned a 1, otherwise, it was assigned

a 0. An estimate of the statistical power is simply the average of these test results, or put another way, the proportion of significant test results to the total number of tests run at that parameter setting. To determine an 'optimal' sample size for a given effect size (0.6), we plotted the number of ROIs having over 80% power as the sample size was varied over a reasonable range (5 to 30), and chose the *n* where we observed an 'elbow'. The 'elbow' represents the sample size where the contribution to power gained from adding another sample begins to wane.

Statistical power for the correlation of c-fos-GFP counts and social behaviors follows previously established power curve for the Pearson correlation. With the selected significance level ($\alpha < 0.05$) and the sample size (N = 26 brains), we obtained >80% statistical power to detect significant correlation values >0.47.

Statistical analysis between experimental groups. We ran statistical comparisons between different behavioral groups based on either ROIs or evenly spaced voxels. Voxels were overlapping 3D spheres with 100 μm diameter each and spaced 20 μm apart from each other. The cell count of each voxel was calculated as the number of nuclei found within 100 μm from the center of the voxel in all 3D. We assumed the cell counts at a given location, *Y*, follow a negative binomial distribution whose mean is linearly related to one or more experimental conditions, *X:* $E[Y] = \alpha + \beta X$. For example, when testing a social group versus a control group, our *X* is a single column showing the categorical classification of mouse sample to group id, i.e. 0 for the control group and 1 for the social group (O'Hara and Kotze, 2010; Venables and Ripley, 2002). We found the maximum likelihood coefficients α and β through iterative reweighted least squares, obtaining estimates for sample standard deviations in the process, from which we obtained the significance of the β coefficient. A significant β means the group status is related to the cell count intensity at the specified location. The z-values in our summary tables correspond to this β coefficient normalized by its sample standard deviation, which under the null hypothesis of no group effect, has an asymptotic standard normal distribution. The p-values give us the probability of obtaining a β coefficient as extreme as the one observed by chance assuming this null hypothesis is true. In the current case of three groups, we utilized Tukey's Honest Significance test to adjust the p-values of the group factor coefficients to control for multiple comparisons: group1v2, group1v3 and group2v3. To account for multiple comparisons across all voxel/ROI locations, we thresholded the p-values and reported false discovery rates with the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). In contrast to correcting for type I error rates, this method controls the number of false positives among the tests that have been deemed significant.

To compare voxel activation between the female and male stimuli (Figures 3-6), voxels that passed the FDR cutoff 0.05 were pseudo-colored (red and green) for each dataset and the activation maps were overlaid on the RSTP brain (e.g. Figure 3C-F)

Supplementary Figures

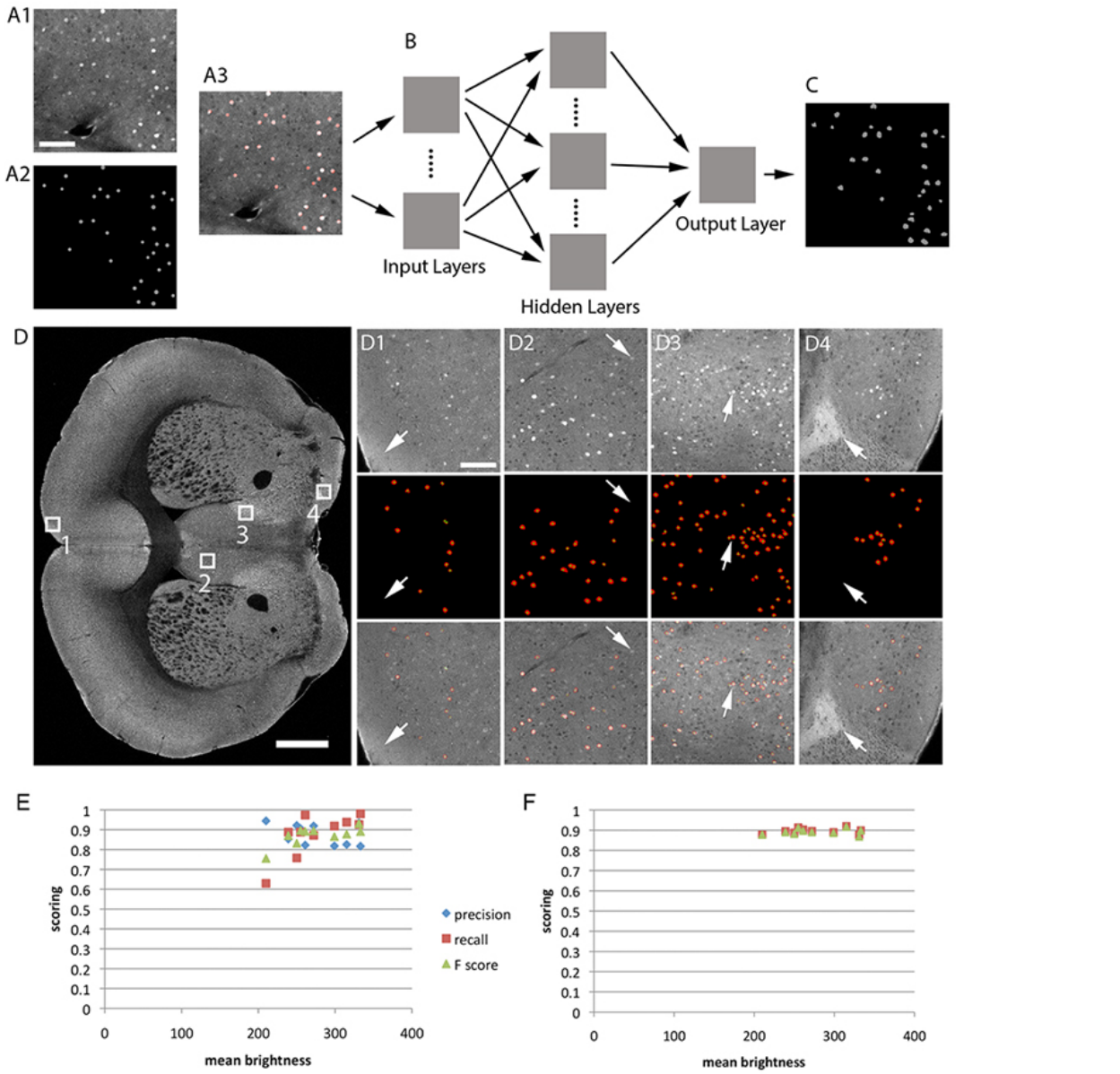

Figure S1. Training of CNs for detection of c-fos-GFP+ cells shown in Figure 1. (A-C) Schematic representation of the CN training procedure. (A) Ground truth c-fos-GFP+ cells were marked up by an expert biologist: A1 = raw image; A2 = human mark-up; A3 = overlay; the scale bar = 100 μm. (B) CNs learn image features of the c-fos-GFP signal through multiple network layers to achieve (C) c-fos-GFP+ cell detection comparable to the human expert. (D) Examples of CN-based detection of c-fos-GFP+ cells. Four (1-4) selected regions are shown in panels d1-d4 as examples of varied background autofluorescence and SNR. Top panels show raw data, middle panels show CN detection, and bottom panels show the overlay. In the middle panels, red color represents c-fos-GFP+ nuclei detected by CNs and the green dots represent centroid position of each c-fos-GFP+ cell after separation of nearby merged cells. Arrows in d1

and D4 point to regions of high autofluorescence, which cause high rates of false positive detection by other tested methods, but not by CNs. Arrows in d2 indicate an example of dim cells that were not detected by CNs. Arrows in d3 show an example of two neighboring cells that were detected by CNs and further separated by the cell separation algorithm. Scale bar in d and d1 = 1 mm and 100 μm, respectively. (E-F) Evaluation of CN and human performance depending on the background autofluorescence. Y axis shows precision, recall, and F score from 10 FOV tiles from the ground truth dataset; X axis shows autofluorescence brightness of the brain regions in the particular tiles. The CN (E) and human (F) performance was overall independent of the background, with the exception of the "darkest tile", which had a lower CN recall of 0.63 (F score 0.76, and precision 0.94). This tile included the caudal olfactory tubercle area, which comprises myelinated fibers passing from the dorsal striatum. This suggests that increased light scattering in areas with myelinated tracks may somewhat lower CN recall performance.

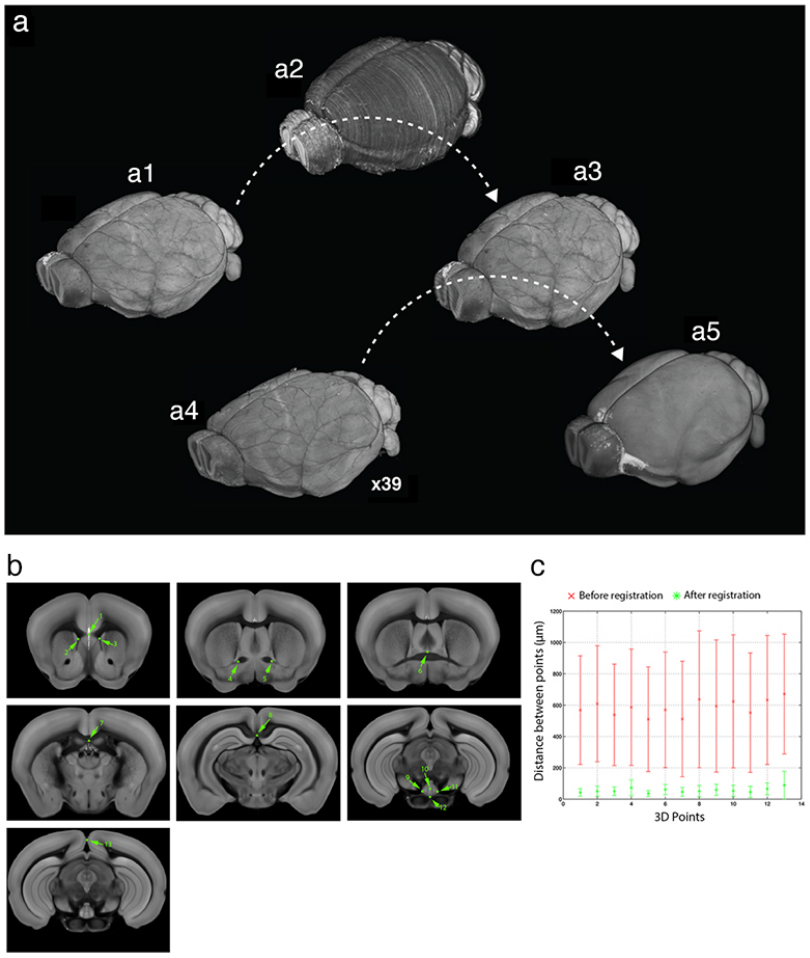

Figure S2. Generation of the Reference STP (RSTP) for image registration shown in Figure 1.

(A) Generation of the RSTP Brain. A single STP brain dataset (A1) was registered by 3D affine and B-spline transformations to the z-stack of Nissl-stained ABA coronal sections (A2). This generated the first transformed STP brain dataset (A3) matched in 3D to the Nissl ABA brain. Next, 39 other STP brains (A4) were registered to the transformed STP dataset (A3), generating the averaged RSTP brain (A5) (see Movie S1). (b-c) Validation of 3D registration accuracy. (B) Thirteen unique points were marked up in the RSTP Brain and in 6 sample STP tomography brains. These points were previously identified as unique 3D landmarks in the Waxholm space (Hawrylycz et al., 2011): (1) frontal middle 1, (2) frontal right 2, (3) frontal left 2, (4) anterior commissure right, (5) anterior commissure left, (6) crossing of the anterior commissure, (7) corpus callosum middle, (8) hippocampus middle, (9) interpeduncular nucleus right, (10) interpeduncular nucleus middle, (11) interpeduncular nucleus right, (12) pontine nucleus middle, (13) cortex middle (http://scalablebrainatlas.incf.org/main/coronal3d.php?template=WHS11&). (C) The distance for each point between the RSTP Brain and the 6 sample brains was (mean $\pm$ SD): 682.6 $\pm$ 327.6 μm and 65.0 $\pm$ 39.9 μm before and after 3D registration by affine and B-spline transformation, respectively.

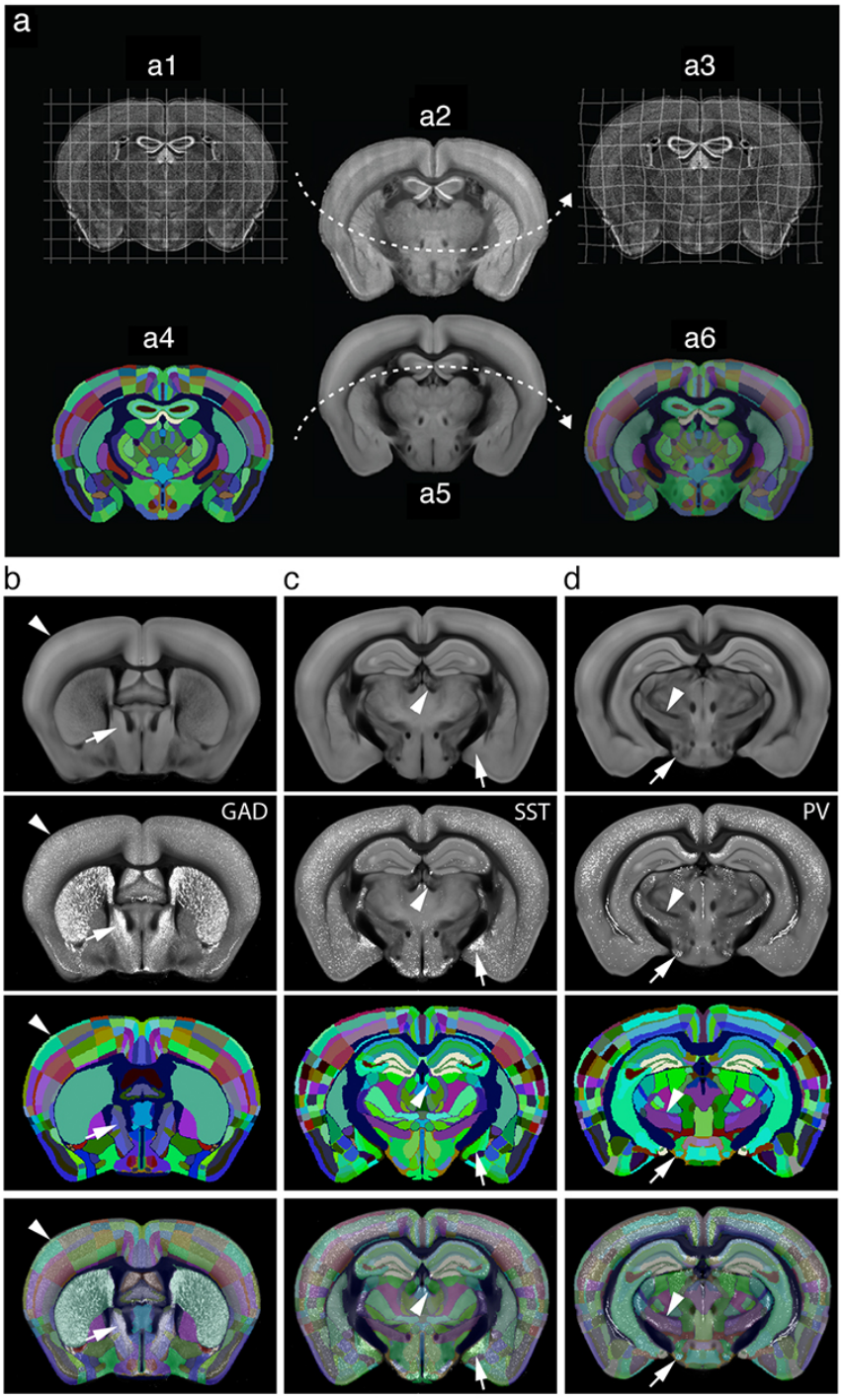

Figure S3. Registration and validation of the ABA anatomical labels in the RSTP brain shown in Figure 1.

(A) Each ABA Nissl-stained section (A1) was registered by 2D transformations to corresponding sections from the reporter CAG-Keima brain (A2) that was previously registered to the RSTP brain. A cellular Nissl-like fluorescent signal of the Keima FP (Kogure et al., 2006) (A2) improved the precision of the alignment between the two

datasets (ABA and RSTP brains) (A3). ABA anatomical labels (A4), as based on the original Nissl dataset (A1), were transformed using the 2D parameters from the Nissl to CAG-Keima registration (A1-3). This further improved the precision of the alignment of the ABA labels to the RSTP brain (A5-6). (B-D) Validation of the alignment of the ABA anatomical labels and anatomical structures in the RSTP Brain. First row panel: RSTP brain autofluorescence signal can be used to align the brain surface contour (arrowhead in B), as well as borders of internal structures, such as habenula (arrowhead in C), and zona incerta (arrowhead in D). Second row panel: fluorescence signal from STP brains of H2B-GFP interneuron cell-type reporter mice (Taniguchi et al., 2011) registered to the RSTP brain can be used to delineate internal structures not clearly visible in the autofluorescence signal. These include the bed nuclei of the stria terminalis, posterior division, principal nucleus (BSTpr) in the glutamic acid decarboxylase (GAD) brain; Medial amygdalar nucleus, posterodorsal part (MEApd) in the somatostatin (SST) brain; and Lateral mammillary nucleus (LM) in the parvalbumin (PV) brain. Third row panel: the ABA anatomical labels registered to the RSTP brain. Fourth row panel: The ABA anatomical labels registered and overlaid to the RSTP brain.

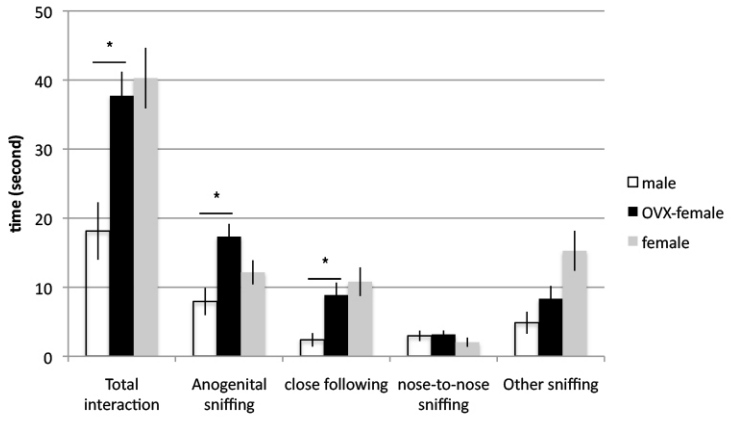

Figure S4. Characterization of experimental behaviors used in Figures 3 – 7.
Characterization of 90-sec interaction between two males (white), male and an OVX female (black), and male and an intact female (gray). There was no significant difference between male interaction with OVX female or intact female (n = 11). The male spent more time interacting with OVX female (n = 13) than another male (n = 13) in the total interaction time (37.7 ± 3.5 and 18.1 ± 4.2; $p = 0.001$), anogenital sniffing (17.3 ± 1.8 and 7.9 ± 2.0; $p = 0.002$) and close following (8.9 ± 1.8 and 2.3 ± 1.0; $p = 0.004$). No significant difference was seen in nose-to-nose sniffing (3.2 ± 0.6 and 3.0 ± 0.8; $p = 0.8$) and other sniffing (8.3 ± 1.9 and 3.0 ± 0.8; $p = 0.16$). All values are in seconds.

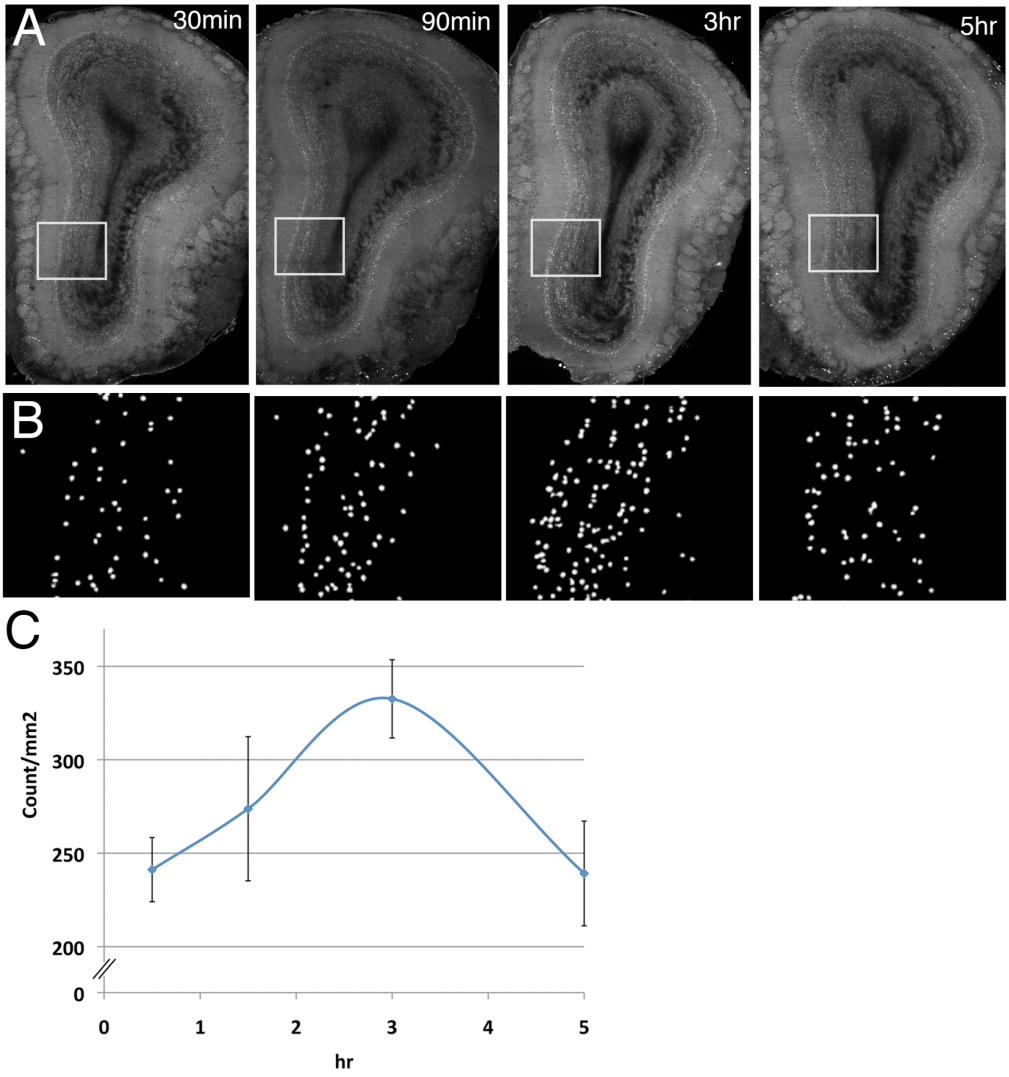
A
30min
90min
3hr
5hr
B
C
Count/mm2
350
300
250
200
0
0
1
2
3
4
5
hr
D
E
F
G
H
cells/mm²
800
700
600
500
400
300
200
100
0
ILA
PL
ORBm
BLAa
COApl
MEA
DG
PIR
n.s.
IHC, Handling
IHC, Female
STP, Handling
STP, Female

Figure S5. c-fos-GFP induction time course and c-fos-GFP+ cell count compared to native c-fos+ expression: validation data for experiments shown in Figures 3-7.
(A) c-fos-GFP induction in the olfactory bulb was examined at 0.5, 1.5, 3 and 5 hours after 90 sec ISO stimulation (Methods). (B) CNs detection of cfos-GFP in the boxed area from (a). Note the highest number of c-fos-GFP+ cells at 3 hours after the stimulation. (C). Quantification (mean ± SD) of the c-fos-GFP+ cell counts in the MOB granular cell layer shows a peak of induction at about 3 hours, which returns to the baseline by 5 hours post ISO stimulation.
(D-I) Comparison of c-fos-GFP and native c-fos induction. (D-F) Representative images of c-fos-GFP labeling in the ORBm cortex after handling (D) and social behavior (E) imaged by STP tomography; the panel (F) shows the location of the zoomed-in views in the corresponding coronal section. (G-H) Representative images of anti-c-fos immunohistochemistry from C57BL/6 mice in the matching ORBm cortex after handling (G) and social stimulation (H) imaged by bright field microscopy. Scale bar = 100 μm. (I) Quantitation of c-fos+ and c-fos-GFP+ cell counts in eight selected regions: ILA (infralimbic area), PL (prelimbic area), ORBm (orbital medial cortex), BLAa (anterior basal lateral amygdala), COApl (cortical amygdala, posterior lateral), MEA (medial amygdalar nucleus), DG (dentate gyrus), and PIR (piriform cortex); all values are mean ± SEM; asterisk = $p < 0.05$; n.s. = not significant. On average, the c-fos-GFP cell counts detected by STP tomography represented ~59% of the c-fos counts detected by immunohistochemistry in which the c-fos protein signal is enhanced by antibody staining. Importantly, both STP tomography and immunohistochemistry detected comparable induction changes between the female and handling groups: ILA = 1.8 and 2.1, PL = 1.9 and 2.8, ORBm = 2.4 and 2.7, BLAa = 2.0 and 1.9, COApl = 2.6 and 2.4, MEA = 1.9 and 2.0; DG = 1.2 and 1.3, and PIR = 2.2 and 1.8.

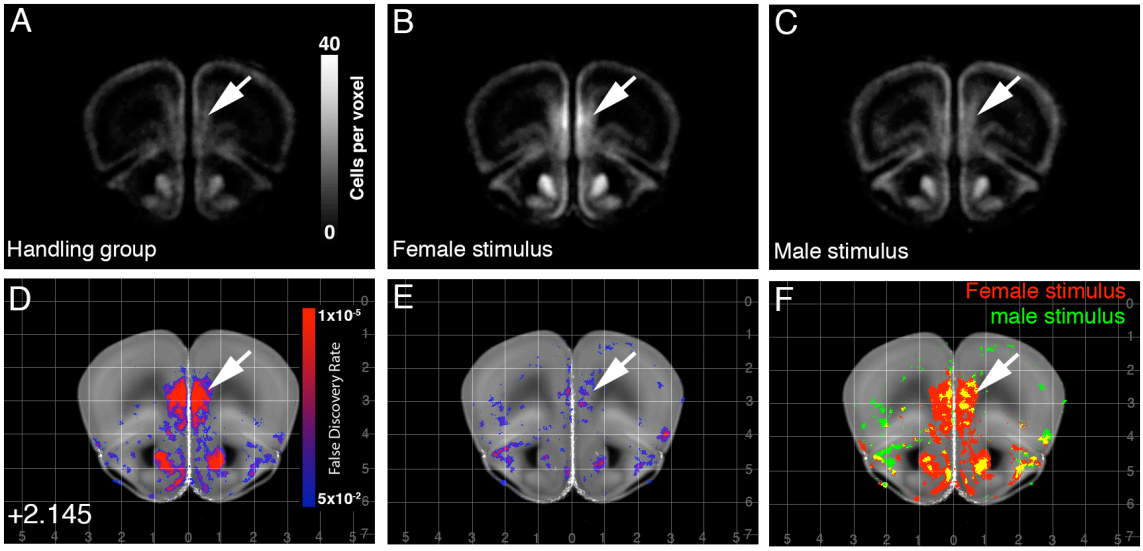

Figure S6. Principles of voxel based statistical analysis used in Figures 3 - 6.
(A-C) Mean voxel-based c-fos-GFP+ cell counts of 13 brains from the handling (A), male-female (B), and male-male (C) groups. The voxels are spheres of 100 μm diameter, with 20 μm spacing; the brightness of the signal is based on the number of cells per voxel, as shown in the heat map index in (A). (D-E) The statistically significant voxels from the handling to the male-female group (D), and from the handling to the male-male group comparison (E) are color-coded according to the level of the statistical significance, as shown in heat map index in (D). (F) Activated voxels (with FDR $q < 0.05$) were binarized, given different color depending on stimulation (red for male-female and green for male-male evoked activation), and overlaid in the RSTP. The arrow points to an example of a large hotspot of activation in the medial orbital cortex. See also Movie S3; the A/P bregma position is indicated at the lower left, the grid for M/L and D/V bregma position is overlaid).

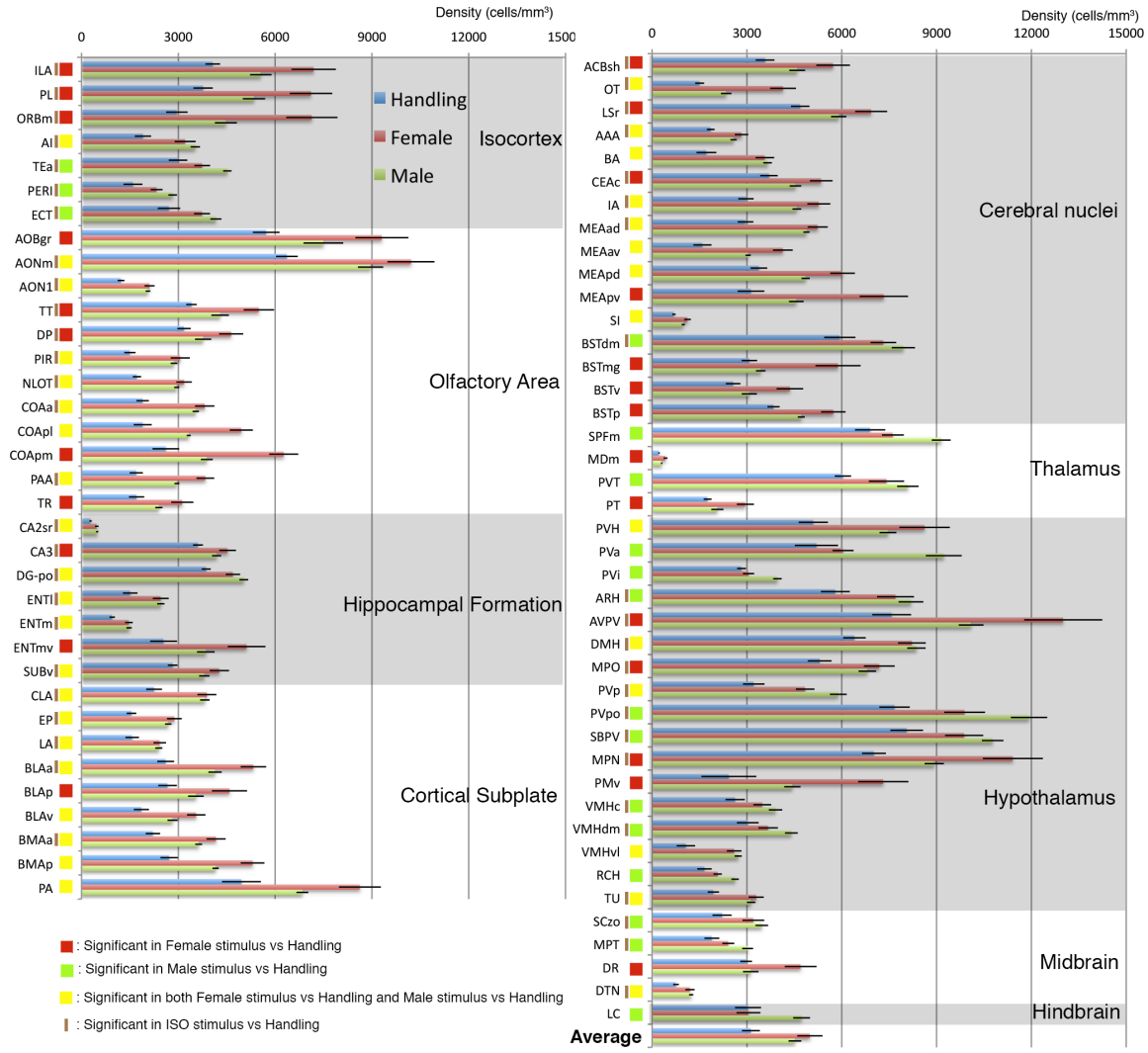

Figure S7. c-fos-GFP cell density in male- and female-stimulus evoked brain regions based on results shown in Figures 3 – 7.

Density of significantly activated brain regions by either male- (green bar) or female-stimulation (red bar) in comparison to handling groups (blue bar) is displayed in the whole brain. Red, green, yellow box, and brown bar next to the ROI indicates that the corresponding ROIs were activated by the female stimulation, male stimulation, both the female and male stimulations, and ISO stimulation with FDR cutoff 0.01, respectively. "Average" in the last bar graph represents the mean values of the all significantly activated ROIs.

TABLES

Table S1. ROI-based analysis of c-fos-GFP induction of the handling, male-female, male-male and ISO groups against the baseline group, related to Figure 3 - 7.
The Table S1a shows the comparison between the handling and baseline groups, the Table S1b shows the male-female and baseline comparison, the Table S1c shows the male-male and baseline comparison, and the Table S1d shows the ISO and baseline comparison. The ROIs with statistically increased c-fos-GFP induction are color-coded: light-blue for 1 x 10^{-2} to 1 x 10^{-3}, green for 1 x 10^{-3} to 1 x 10^{-5}, and red for <1 x 10^{-5}. Column A = the abbreviation names of the ABA ROIs; column B = the full name of the ABA ROIs; column C = the unique numerical ID for each ROI; column D = the hierarchical structure order of each ROI; columns E-H = mean and SD for each ROI in the corresponding experimental groups; columns I-K = the statistical z-scores, uncorrected p values, and corrected FDR q value, respectively; column M-N shows color coding information based on FDR values. The anatomical location of ROIs can be easily viewed in online Allen Brain Atlas (http://atlas.brain-map.org).

Table S2. ROI-based analysis of c-fos-GFP induction of the male-female, male-male and ISO groups against the handling group, related to Figure 3 - 7.
The Table S2a shows the comparison between the male-female and handling groups, the Table S2b shows the male-male vs. handling group comparison, and the Table S2c shows the ISO vs. handling group comparison. The format of the table is the same as the Table S1.

Table S3. ROI-based analysis of c-fos-GFP induction of the male-female and male-male groups against ISO group, related to Figure 3 - 7.
The Table S3a shows the comparison between the male-female and ISO groups with male-female higher ROI highlighted, S3b shows the comparison between the male-female and ISO groups with ISO higher ROI highlighted, S3c shows the comparison between the male-male and ISO groups with male-male higher ROI highlighted, and S3d shows the comparison between the male-male and ISO groups with ISO higher ROI highlighted. The format of the table is the same as the Table S1.

MOVIES

Movie S1. RSTP Mouse Brain, related to Figure 1
Bregma coordinates (in mm) imported from Allen Reference Atlas were overlaid; Numbers in the left bottom corner represent A/P coordinates. Vertical and horizontal grid lines represent M/L and D/V coordinates, respectively.

Movie S2. Video recording of a representative example of a male-female interaction, a male-male interaction, and a handling control group, related to Figure 3 – 7
Annotation in the right bottom corner represents specific social behavior manually scored (see also Figure S4).

Movie S3. Voxel based analysis, related to Figure 3 - 6
The male-female vs. handling group comparison (left), the male-male vs. handling group comparison (middle), and the binarized male and female vs. handling group comparison (right). The visualization of the voxel-based statistical results is overlaid on the RSTP brain. The voxel output is shown with FDR q = 0.05 as a cutoff (red for male-female, green for male-male, yellow for shared activation). The A/P bregma position is indicated at the bottom left and the grid for M/L and D/V bregma position is overlaid.

References

Barth, A.L., Gerkin, R.C., and Dean, K.L. (2004). Alteration of neuronal firing properties after in vivo experience in a FosGFP transgenic mouse. J Neurosci *24*, 6466-6475.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological), 289-300.

Ferguson, J.N., Aldag, J.M., Insel, T.R., and Young, L.J. (2001). Oxytocin in the medial amygdala is essential for social recognition in the mouse. J Neurosci *21*, 8278-8285.

Ferguson, J.N., Young, L.J., Hearn, E.F., Matzuk, M.M., Insel, T.R., and Winslow, J.T. (2000). Social amnesia in mice lacking the oxytocin gene. Nature genetics *25*, 284-288.

Garner, A.R., Rowland, D.C., Hwang, S.Y., Baumgaertel, K., Roth, B.L., Kentros, C., and Mayford, M. (2012). Generation of a synthetic memory trace. Science *335*, 1513-1516.

Hawrylycz, M., Baldock, R.A., Burger, A., Hashikawa, T., Johnson, G.A., Martone, M., Ng, L., Lau, C., Larson, S.D., Nissanov, J.*, et al.* (2011). Digital atlasing and standardization in the mouse brain. PLoS computational biology *7*, e1001065.
Klein, S., Staring, M., Murphy, K., Viergever, M.A., and Pluim, J.P.W. (2010). Elastix: a toolbox for intensity based medical image registration. IEEE Transactions on Medical Imaging *29*, 196 - 205.
Kogure, T., Karasawa, S., Araki, T., Saito, K., Kinjo, M., and Miyawaki, A. (2006). A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. Nature biotechnology *24*, 577-581.
Liu, X., Ramirez, S., Pang, P.T., Puryear, C.B., Govindarajan, A., Deisseroth, K., and Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. Nature *484*, 381-385.
Lupien, S.J., McEwen, B.S., Gunnar, M.R., and Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nature reviews *10*, 434-445.
Matsuo, N., Reijmers, L., and Mayford, M. (2008). Spine-type-specific recruitment of newly synthesized AMPA receptors with learning. Science *319*, 1104-1107.
McCullagh, P., and Nelder, J.A. (1989). Generalized linear model, Vol 37 (Chapman & Hall/CRC).
Mutch, J., Knoblich, U., and Poggio, T. (2010). CNS: a GPU-based framework for simulating cortically-organized networks. Massachusetts Institute of Technology, Cambridge, MA, Tech Rep MIT-CSAIL-TR-2010-013/CBCL-286.
O'Hara, R.B., and Kotze, D.J. (2010). Do not log-transform count data. Methods in Ecology and Evolution *1*, 118-122.
Ragan, T., Kadiri, L.R., Venkataraju, K.U., Bahlmann, K., Sutin, J., Taranda, J., Arganda-Carreras, I., Kim, Y., Seung, H.S., and Osten, P. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. Nature methods *9*, 255–258.
Reijmers, L.G., Perkins, B.L., Matsuo, N., and Mayford, M. (2007). Localization of a stable neural correlate of associative memory. Science *317*, 1230-1233.
Schilling, K., Luk, D., Morgan, J.I., and Curran, T. (1991). Regulation of a fos-lacZ fusion gene: a paradigm for quantitative analysis of stimulus-transcription coupling. Proceedings of the National Academy of Sciences of the United States of America *88*, 5665-5669.
Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., Kvitsani, D., Fu, Y., Lu, J., Lin, Y.*, et al.* (2011). A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex. Neuron *71*, 995–1013.
Venables, W., and Ripley, B. (2002). MASS: modern applied statistics with S (Springer, New York. http://www. stats. ox. ac. uk/pub/MASS4).
Williams, R.W., and Rakic, P. (1988). Three-dimensional counting: an accurate and direct method to estimate numbers of cells in sectioned material. The Journal of comparative neurology *278*, 344-352.
Wilson, Y., Nag, N., Davern, P., Oldfield, B.J., McKinley, M.J., Greferath, U., and Murphy, M. (2002). Visualization of functionally activated circuitry in the brain. Proceedings of the National Academy of Sciences of the United States of America *99*, 3252-3257.

Yang, M., and Crawley, J.N. (2009). Simple behavioral assessment of mouse olfaction. Current protocols in neuroscience / editorial board, Jacqueline N Crawley [et al *Chapter 8*, Unit 8 24.