Supplemental text

PCR and enhancer trap staining quality controls

As quality controls for RT-PCR, different primers for the same gene (Dataset S1) showed consistent RT-PCR signal 93% (26/28) of the time, and positive and negative controls worked as expected (Dataset S1).

Different enhancer trap lines related to the same gene showed consistent expression patterns, sometimes with slight variations (Tables S2 and S3).

Expression Profiling quality controls

We mapped the reads to the annotated *D. melanogaster* genome (genome plus the exon-exon conjuction sequences according to transcript annotation) with stringent criteria, filtering out all reads that mapped to two or more places in the transcriptome. For quality assessment, we verified that biological replicates have a high correlation (0.96) (Figure S3 A). We additionally confirmed the tissue-specificity of expression profiles (Figure S3 C-D, extended experimental procedures).

As an overall count of the number of detectable expressed genes, 84.8% (9981/11764) of all genes, and 49.9% (442/886) of young genes, were expressed in the brain. Old genes are defined as genes that are shared by all 12 sequenced

Drosophilid species, since their age predated the last common ancestor of those species. There are roughly ~20% of the genes in the genome whose ages are not determined.

Of the young genes, if they were expressed in the brain, they were very likely to be expressed in MBs (369/442=83.5%). A gene is considered MB-enriched if the expression level is significantly higher in MB compared to brain (q-value < 0.001, corrected for multiple testing), and MB-depleted if significantly lower (q < 0.001). These data are consistent with the results from both RT-PCR (161/330=49% of young genes detected in the brain) and enhancer trap expression (14/17=82.4% of young brain genes detected in MBs). Because young genes are mostly duplicates, many of their transcript reads will map to several genomic locations. We used stringent unique mapping to avoid signals from paralogs, so many reads that mapped to both a young gene and its paralog(s) were removed. This results in an underestimation of the number of young genes that are expressed in a particular sample, since all reads mapping to more than one genomic location are excluded from the "expressed" dataset. Our use of unique mapping rules out the possibility that the observed signal of young genes comes from their parental genes.

We also analyzed existing genes implicated in several diverse biological functions by retrieving gene sets from literature (Lemaitre and Hoffmann, 2007; Reuter et al., 2003) (Mueller et al., 2005; Vibranovski et al., 2009a). We found

that genes previously implicated in MB development are significantly more abundant in our brain and MB RNA-seq datasets than genes implicated in immunity, seminal fluid production and spermatogenesis. By contrast, genes involved in spermatogenesis are significantly enriched over those involved in MB development in a testis RNA-seq dataset (Figure S3 C-D, Wilcoxon test, $p < 0.001$ for all comparisons).

Around 80% of young genes are detectable in the testis RNA-seq dataset, consistent with the previously proposed "out-of-testis hypothesis" (Kaessmann 2009, 2010), where a new gene is first expressed in the testis, then "jumps" out of the testis and acquires expression in other tissues, such as the brain/MB, as shown here.

Figure S3B shows the confirmation of observation from Figure 3B with different read mapping methods (Extended Experimental Procedure).

Foraging behavior assay setup and quality control

We designed an assay system to measure the foraging ability of *Drosophila* (Figure 5A). In this system, flies must move from an initial empty compartment to a second compartment to find food, through a narrow linker tunnel just large enough for a single fly to pass at a time. We first tested the foraging behavior in two wildtype *D. melanogaster* lines *(Canton-S and Oregon-R)*. We observed that these flies gradually moved out of the original compartment into the second

compartment (Figure 5C). This phenomenon is associated with two possibilities: 1) foraging where flies are attracted to and consume food or 2) olfaction where flies are only attracted to the odor without proceeding to consuming food. Continuous observation by video monitoring reveals that flies are foraging in our system: immediately upon reaching the food source, the starved flies start to taste and feed on the food (Supplemental Video 1). We also observed that the consumption of food in the second compartment was essential for the survival of the flies in our assay. Starved flies died within 2~3 days if no food consumption occurred due to either lack of food in the second compartment (control assay), or failure in foraging (those flies that remained in the first compartment); while those flies that successfully found food can live at least 14 days.

As a negative control, when no food was placed in the second tube, only a small proportion of flies moved out of the original tube by chance in 3 hours, significantly fewer than in the experiment group (KS test, $p < 1x10^{-11}$, Figure 5C). We also quantified the performance of foraging behavior in this system by estimating an index of foraging speed (FSI) for each experiment, and found that wildtype flies showed a 10-fold difference in FSI in the presence versus absence of food (Figure 5D). The wildtype performance and controls demonstrate that our behavior system works efficiently and consistently in detecting foraging ability in wildtype strains of *D. melanogaster*.

Inference of ancestral foraging speed index by MCMC

We inferred the ancestral foraging phenotype using a previously established Markov Chain Monte Carlo method implemented in BayesianTraits software (Pagel and Meade, 2006; Pagel et al., 2004). We inferred the FSI values of the ancestral nodes dated prior to *Xcbp1* origination (nodes C and D, Figure S7A) and of the nodes dated after *Xcbp1* origination (nodes A and B, Figure S7A). We found that FSIs of nodes C and D were lower than those of A and B (KS test, $p < 0.01$, Figure S7A, Figure 7F).

Interestingly, while the FSIs of *D. yakuba* and *D. erecta* are similar to the slow foraging of *Xcbp1* RNAi knocked-down *D.melanogaster*, *D. pseudoobscura* has an even lower FSI value (Figure 7A). We observe an apparent stepwise increase in FSIs that correlates with the evolutionary divergence of different *Drosophilid* clades (Figure 7A). Specifically, the FSI values increased at (I) the node E to node D transition (*pseudoobscura – D. melanogaster* split to *D. yakuba - D. melanogaster* split), and (II) the node D to node B transition (*D. yakuba - D. melanogaster* split to *D. melanogaster* clade split). This implies that multiple ancestral events possibly occurred at these particular divergence points that influenced foraging behavioral evolution.

Extended Experimental Procedures

Additional sequence and molecular analysis

One or more primer sets specific to each gene were designed, avoiding cross-annealing with paralogs, with an optimum annealing temperature (Tm) of ~60℃ and an optimum product size of ~700bp (Batchprimer3, Dataset S1). RNA samples were collected from dissected adult brains. RT-PCR was performed with oligo-dT and gene specific primers using standard protocols. Primer sequences for the 330 young genes and control genes are shown in Table S6 and in Dataset S1.

Genome and protein sequences were retrieved from Flybase (fly genome assembly 2006, Dm3). Evolutionary analyses including age distribution, origination mechanisms, chromosomal distribution and movements were analyzed as described in (Vibranovski et al., 2009b; Zhang et al., 2010). Primary polymorphism data were collected from the Drosophila Population Genomics Project (DPGP). Coding sequences of each gene from the 37 *D. melanogaster* lines and that from the *D. yakuba* ortholog were aligned (ClustalW) to estimate the non-synonymous divergence, synonymous divergence, and polymorphic frequency spectra using the Polymorphorama program (Haddrill et al., 2008). Estimation of the prevalence of natural selection, or the proportion of non-synonymous substitutions driven by positive selection (α values), was carried out with the polymorphism data using the DoFE package (Eyre-Walker and Keightley, 2009), with *D. yakuba* as an outgroup. Results were similar when using *D. simulans* as an outgroup (data not shown). Comparison of α values

between groups was carried out using the Wilcoxon Rank Sum test. Results using the likelihood ratio test were similar (data not shown).

Specifically for *Xcbp1,* the polymorphism data were not covered in the DPGP; therefore, we sequenced the *Xcbp1* locus and acquired polymorphism data, using the 12 African and cosmopolitan *D.melanogaster* lines from (Emerson et al., 2008). Results from sequencing showed that compared to *Cnx99a*, *Xcbp1* has an AG-rich insertion which encodes an additional polyE peptide fragment. Analysis of the *Xcbp1-Gal4* (*pG156*) insertion site and direction was performed by genomic PCR and sequencing using various primer combinations from the genome and the pGawB transposable element construct. Results showed that the pG156 pGawB element inserted 1kb 5' of *Xcbp1* in the same (minus) strand, at chromosomal coordinate X:16420255; the breakpoint sequence is below:

GGAGAGCGCAATGGGCTGGCGAGAGAGCGAGGACTCTCACTCTCCTCCTG
CCTGCGCCGAGAGAGCCAGTGCCCTCTCTCTTGCGCTCTCACTTCTCTGTG
TGATGCCGCCAACAAACACGGACTACAACAATACATTTTTGGGTTCCCTTCT
CTCTTAACGAACCTCCTTTCAATGTGTGCGTGAGCAAAGGAATTCCGTTAGA
AA

Xcbp1, Desr and their homolog protein sequences were retrieved from Flybase and Ensemble, then alignments were generated by ClustalX and Jalview. Secondary structures were predicted by Jpred (Cole et al., 2008; Waterhouse et

al., 2009). 3D structure was predicted with Swiss-Model (Arnold et al., 2006) with template homolog structure PDB 1JHN (Schrag et al., 2001).

RNA-seq and dataset analyses

Sequencing reads were mapped back to the *D. melanogaster* genome (genome sequence plus the exon-exon conjunction sequence according to transcript annotation as described in (Mortazavi et al., 2008) using Bowtie (Langmead et al., 2009), allowing a maximum of 2 mismatches. For those datasets with relatively lower uniquely mapped reads, the unmapped reads (the majority of them have multiple Ns at one or both ends) were trimmed by 2nt at both ends and re-mapped using the same criteria, to increase the number of total uniquely mapped reads. With unique reads, we obtained 3 - 11 fold coverage of the transcriptome for these samples.

We calculated the relative expression level for each gene in the genome, using the number of reads per kilobase per million total reads (RPKM) as described previously (Mortazavi et al., 2008). The expression levels for young genes tend to be underestimated, because most young genes are duplicates from old genes and thus some reads that mapped to the common region of the gene family were filtered out.

Likelihood-ratio tests were performed in a generalized linear model framework as described in (Marioni et al., 2008) with our own R script. Briefly, the numbers of

reads for each gene were modeled as follow a Poisson distribution in each sample. The null hypothesis is that the two distributions share a same rate parameter lambda ($\lambda_1 = \lambda_2$); the alternative hypothesis is that the two distributions have different lambdas ($\lambda_1 \neq \lambda_2$). The p-value and the $\log_2$ ratio of each gene between different tissue type samples were calculated. The FDR method was used for multiple testing corrections. Values for multiple hypothesis testing were corrected using an fdr method as described in (Storey, 2002; Storey and Tibshirani, 2003).

Foraging behavior assay

The foraging behavior assay we developed is comprised of two compartments linked by a tube 20 mm in length and 2 mm in diameter, which is just large enough for one fly to crawl through at a time. The first 7 x 2.5 x 2.5 cm compartment contains no food, while the second 25 x 1.5 x 1.5 cm compartment contains a yogurt-based food mixed with fruit odors such as strawberry and banana (main ingredients in extended experimental procedures), because such odors have been reported to evoke a neuronal response in flies (Hallem and Carlson, 2004a, b, 2006). A cotton plug was inserted into the end of the 2nd compartment to allow air exchange (Figure 6 A).

For foraging assays, all flies were raised at 22-25 ℃ and 40-60% humidity on a 12:12 hour day:night cycle with standard corn food unless otherwise mentioned. Flies of the desired species and genotypes aged 3-7 d were collected and kept in

the aforementioned conditions for 1-2 additional days to allow recovery, thus avoiding any residual effects of CO_2 anesthesia on foraging behavior. Flies were starved but not desiccated for 20-22h by transferring to an empty tube containing dH_2O but no food. In each experiment, 20 flies were gently transferred without CO_2 into the 1st compartment of the assay system. The proportion of flies that crawled into the 2nd compartment containing the food source was observed for 3 hours. For neuronal inactivation experiments, flies were shifted to a 31℃ incubator for 2h prior to transfer into the 1st compartment of the assay system.

Behavioral data were recorded manually with a timer and analyzed with SigmaPlot and R. Foraging Speed Indices (FSIs) were defined as the percentage of flies that entered the 2nd chamber of the assay system (i.e. that found food) per minute. The estimation of FSIs were calculated using a linear model with single independent variable regression with the slope *b* given according the following equation:

$$y = a + bx$$

where:

$$a = \frac{\sum y - b\sum x}{n}$$

$$b = \frac{n\sum (xy) - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

The food sources used in the foraging experiments were either generated by extracting ripe natural fruits including strawberrys, bananas and apples, or purchased from Dannon ®, where in the latter the main ingredients include water, nonfat milk, food starch, fructose, gelatin, vitamin a palmitae, vitamin D3, strawberry extract, banana extract, food flavors, aspatame, potassium sorbate, sodium citrate, annatto extract (coloring), malic acid, acesulfame potassium, sucralose and red 40.

Inference of ancestral FSI

For ancestral state inference of FSI values, a Markov Chain Monte Carlo method was used in BayesianTraits (Pagel and Meade, 2006; Pagel et al., 2004) with branch lengths of estimated *Drosophila* species divergence times (Powell, 1997). We used a standard constant-variance random walk model (Model A), which has a single parameter of the (instantaneous) variance of evolution. For all other parameters, defaults were used unless mentioned. Internal notes for estimation were defined as follows: Node A: the MRCA (most recent common ancestor) of *D.simulans, D.sechellia* and *D.mauritiana*; Node B: the MRCA of *D.melanogaster* and Node A; Node C: the MRCA of *D.yakuba* and *D.erecta*; Node D: the MRCA of *D.pseudoobscura* and Node C.

Supplemental Figure Legends

Figure S1.

Gene movement and sex-biased expression pattern of young brain genes

(A) Location shift of young brain genes with respect to parental genes; Tandem, Tandem duplication; intra-chr interspersed, intra-chromosomal interspersed duplication; inter-chr movement, between chromosome movement;
(B) Inter-chromosomal and intra-chromosomal movements of young brain genes with respect to parental genes; X-A, X chromosome to autosome movement; A-X, autosome to X chromosome movement; A-A, movement between different autosome arms.
(C) Sex-biased gene expression of young brain genes; male biased, higher expression in male adult flies or testis; female biased, higher expression in female adult flies or ovary; unbiased, no differential expression between male and female tissues.
(See also: Figure 1)

Figure S2.

Expression of young brain genes in the mushroom bodies
(A) A simplified schematic representation of the *D. melanogaster* mushroom body lobes. The five major lobes, α, β (green), α', β' (yellow) and γ (blue), are labeled.
(B) Enhancer-trap based expression patterns of the representative young brain genes in *D. melanogaster* mushroom body lobes. The gene where each enhancer-trap is inserted is noted in green text. The α and α' lobes for each image are labeled and outlined in white. The β and β' lobes are labeled and outlined in yellow. The γ lobe is labeled and outlined in blue. Gene-specific

enhancer-trap *Gal4* driven GFP is in green; background staining with the pan-neuropil marker nc82 is in red. Scalebars = 50 μm.

(See also: Figure 2)

Figure S3

The MB RNA-seq experiment.

(A) Statistical summary of RNA-seq experiments; Top, scatterplot of relative expression levels in RNA-seq between two biological replicates of MBs, with a Spearman correlation = 0.96; Bottom, Boxplot of mean RPKM values of brain, MB and testis samples;

(B) Statistical summary of MB-enriched and MB-depleted genes in differentially expressed gene sets; 60-Myr, genes that are younger than 60 million years (n=369, DE=59); 35-Myr, genes that are younger than 35 million years and with additional quality filtering (n=206, DE=25); Old, genes that are older than 60 Myr (n=9197, DE=3057); Random, randomly sampled genes from the genome (n=11289, DE=3684); Compared to Figure 3B, this barplot uses RNAseq data without reads trimming and re-mapping, and shows a similar conclusion;

(C-D) Relative expression levels of four gene sets implemented in four distinct biological processes: (1) MB development (Reuter et al 2003), (2) immunity (Lemaitre et al 2008), (3) accessory gland proteins (ACPs) (Mueller et al 2005) and (4) spermatogenesis (Vibranovski et al 2009), in brain and MB (C) and testis (D) RNA-seq datasets; MB development genes showed highest expression level

in both brain and MB, spermatogenesis genes showed highest expression level in testis (Wilcoxon test, $p < 0.001$ for all comparisons);

(E) Scatter plot of relative expression levels in RNA-seq between young genes and their respective parental genes in average MB RPKM, showing low correlation (0.17);

(F) Estimation of Alpha revealed positive selection on young brain genes. Alpha, proportion of nonsynonymous substitution subjected to positive selection; young brain genes (young br.) have significantly higher alpha compared to old brain genes (old br.);

Stars denote significance in statistical comparison, if applicable.

(See also: Figure 3)

Figure S4.

Expression pattern of *Xcbp1-Gal4* in antennal lobe revealed several major Xcbp1+ glomeruli

(A) Schematic representation of the positions of Xcbp1+ antennal lobe glomeruli;

(B) Scoring summary of *Xcbp1-Gal4* in antennal lobe glomeruli. Columns represent individual samples and lines represent individual glomeruli.

Antennal lobes were split into anterior, middle and posterior views in (A); in both (A) and (B), colors are indicators for relative Xcbp1-Gal4 signal intensity and/or innervation: magenta glomeruli, high signal and strong innervation; light blue, intermediate signal or intermediate innervation; darker blue, little/absent signal or weak/absent innervation.

(See also: Figure 4)

Figure S5

Inhibition of synaptic transmission in *Xcbp1*+ neurons leads to a foraging phenotype.

(A) FSIs of animals tested at permissive temperature (24 ° C). Genotypes are as labeled. n.s., not significant (ANOVA, p > 0.01)

(B) FSI of animals tested at restrictive temperature (31 ° C). Genotypes are shown at the bottom of each column; data are represented as mean +/− SEM. Asterisk denotes statistical significance (ANOVA, p < 0.01) in all pairwise comparisons; n.s., not significant (ANOVA, p > 0.01) in all pairwise comparisons.

(See also: Figure 5)

Figure S6.

Protein evolution of *Xcbp1* and the calnexin family showing rapid divergence

(A) Regional protein sequence alignment between Xcbp1, its parental protein Cnx99a, and representative calnexin protein family members, showing highly conserved regions (left panels), regions with rapid Xcbp1-specific amino acid substitutions (middle panels), and regions with Xcbp1-unique poly-E insertions (right panels). Only partial representative sequences were shown. Secondary structure predicted with Jpred 3;

(B) Homology-based protein structure of Xcbp1 protein based on (Schrag et al, 2001).

(See also: Figure 6)

Figure S7

Evolution of foraging behavior and identification of *Desr*

(A) Inference of ancestral stage of foraging speed.

Inference of FSI values by BayesianTraits at ancestral state (nodes A, B, C and D) during *Drosophila* evolution. Nodes after the origination of *Xcbp1* (Node A and B) show higher FSIs. Node E cannot be inferred because it is the most ancestral node without a further outgroup.

(B) FSI measurements for additional young MB genes identified *Desr*.

The bar chart shows FSI measurements for RNAi knockdown of young MB genes with the MB driver *OK107*, as well as for control animals. Genotypes are shown at the bottom of each column. A single gene, *CG31875 (Desr),* shows a significant foraging phenotype. Data are represented as mean +/− SEM. The asterisk denotes statistical significance ($p < 0.01$); n.s., not significant ($p > 0.01$).

(See also: Figure 7)

Notes on Supplemental Video

This video shows that wildtype *D. melanogaster* flies actually forage in our behavior paradigm. Wildtype flies not only followed food odors to the second compartment of the behavior apparatus, but also display continuous tasting of and feeding on the food.

Supplemental Tables

Table S1

List of young genes and their origin

This table is a list of 566 young genes in the *D. melanogaster* genome, with their parental genes or closest paralogs (if applicable), origination mechanisms, chromosomal arm positions and ages (primary source from (Chen et al., 2010)).

Table S2

Cellular expression patterns of young brain genes

The cellular expression patterns of young brain genes in major brain centers based on enhancer-trap staining, as well as their genomic locations and putative protein domains. Symbols are defined as follows: (+), clearly expressed in this particular region; (-), clearly not expressed in this region; NA, not available or insufficient resolution; weak, weak expression.

Table S3

Enhancer trap lines and cellular expression patterns of young brain genes

Supplemental Datasets in Excel format

Dataset S1. RT-PCR detection of gene expression of young genes in the *D. melanogaster* brain, and the primers used. Brain expression of young genes also assayed for enhancer-trap expression was confirmed with three independent primer pairs per gene.

Dataset S2. Protein domains encoded by young brain genes

Dataset S3. Polymorphism data and estimation of natural selection

Dataset S4. Genome-wide likelihood ratio test of brain vs MB RNA-seq expression profiling, including young genes and old genes

Dataset S5. Genome-wide likelihood ratio test of testis vs MB RNA-seq expression profiling

Dataset S6. Genome-wide likelihood ratio test of testis vs brain RNA-seq expression profiling

Supplemental Material References

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195-201.

Chen, S., Zhang, Y.E., and Long, M. (2010). New genes in Drosophila quickly become essential. Science *330*, 1682-1685.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res *36*, W197-201.

Emerson, J.J., Cardoso-Moreira, M., Borevitz, J.O., and Long, M. (2008). Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster. Science *320*, 1629-1631.

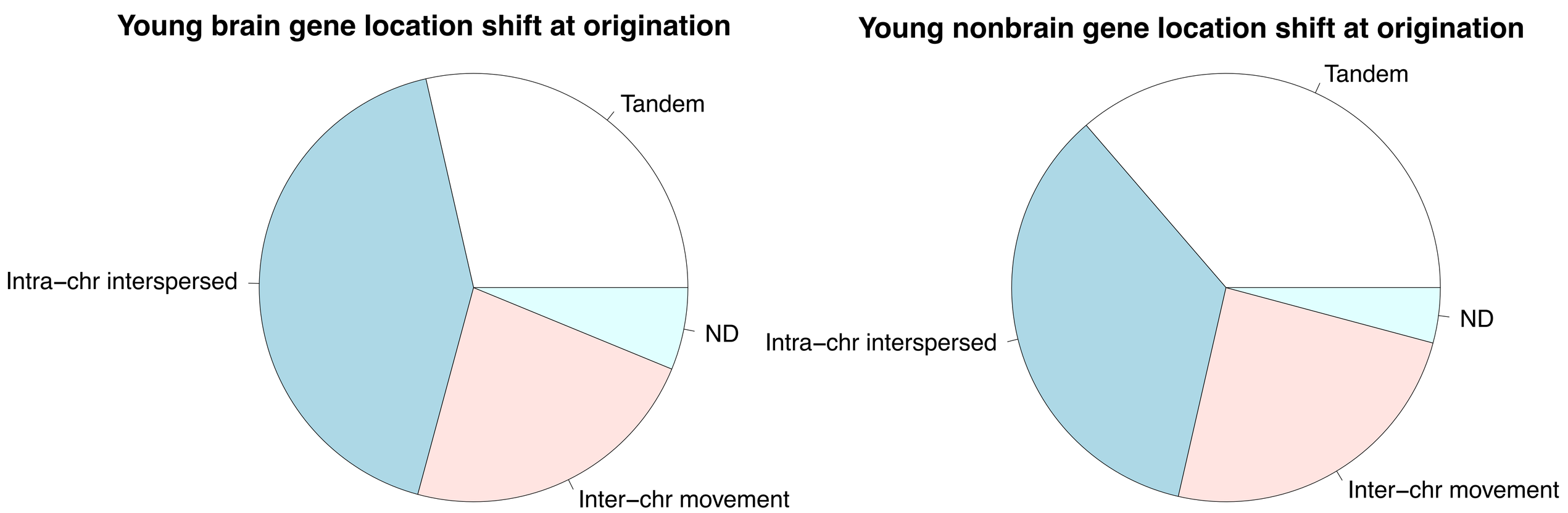
Eyre-Walker, A., and Keightley, P.D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol *26*, 2097-2108.

Haddrill, P.R., Bachtrog, D., and Andolfatto, P. (2008). Positive and negative selection on noncoding DNA in Drosophila simulans. Mol Biol Evol *25*, 1825-1834.
Hallem, E.A., and Carlson, J.R. (2004a). The odor coding system of Drosophila. Trends Genet *20*, 453-459.
Hallem, E.A., and Carlson, J.R. (2004b). The spatial code for odors is changed by conditioning. Neuron *42*, 359-361.
Hallem, E.A., and Carlson, J.R. (2006). Coding of odors by a receptor repertoire. Cell *125*, 143-160.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
Lemaitre, B., and Hoffmann, J. (2007). The host defense of Drosophila melanogaster. Annu Rev Immunol *25*, 697-743.
Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res *18*, 1509-1517.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.
Mueller, J.L., Ravi Ram, K., McGraw, L.A., Bloch Qazi, M.C., Siggia, E.D., Clark, A.G., Aquadro, C.F., and Wolfner, M.F. (2005). Cross-species comparison of Drosophila male accessory gland protein genes. Genetics *171*, 131-143.
Pagel, M., and Meade, A. (2006). Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. Am Nat *167*.
Pagel, M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. Syst Biol *53*, 673-684.
Powell, J.R. (1997). Progress and Prospects in Evolutionary Biology - The Drosophila Model (Oxford University Press).
Reuter, J.E., Nardine, T.M., Penton, A., Billuart, P., Scott, E.K., Usui, T., Uemura, T., and Luo, L. (2003). A mosaic genetic screen for genes necessary for Drosophila mushroom body neuronal morphogenesis. Development *130*, 1203-1213.
Schrag, J.D., Bergeron, J.J., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y., and Cygler, M. (2001). The Structure of calnexin, an ER chaperone involved in quality control of protein folding. Mol Cell *8*, 633-644.
Storey, J.D. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B-Statistical Methodology *64*, 479-498.
Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America *100*, 9440-9445.
Vibranovski, M.D., Lopes, H.F., Karr, T.L., and Long, M. (2009a). Stage-specific expression profiling of Drosophila spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. PLoS Genet *5*, e1000731.
Vibranovski, M.D., Zhang, Y., and Long, M. (2009b). General gene movement off the X chromosome in the Drosophila genus. Genome Res *19*, 897-903.

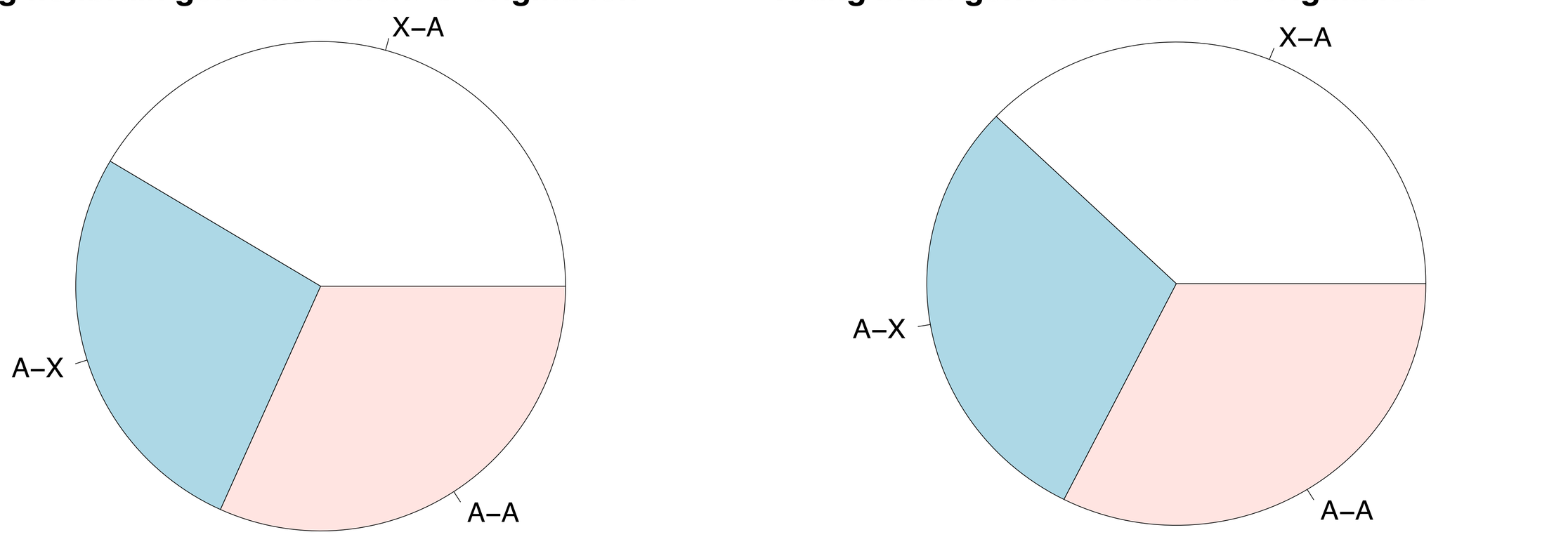
Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.
Zhang, Y., Vibranovski, M.D., and Long, M. (2010). Age-dependent chromosomal distribution of male-biased genes in Drosophil. Genome Research *In press*.

Figure S1

A



B



C

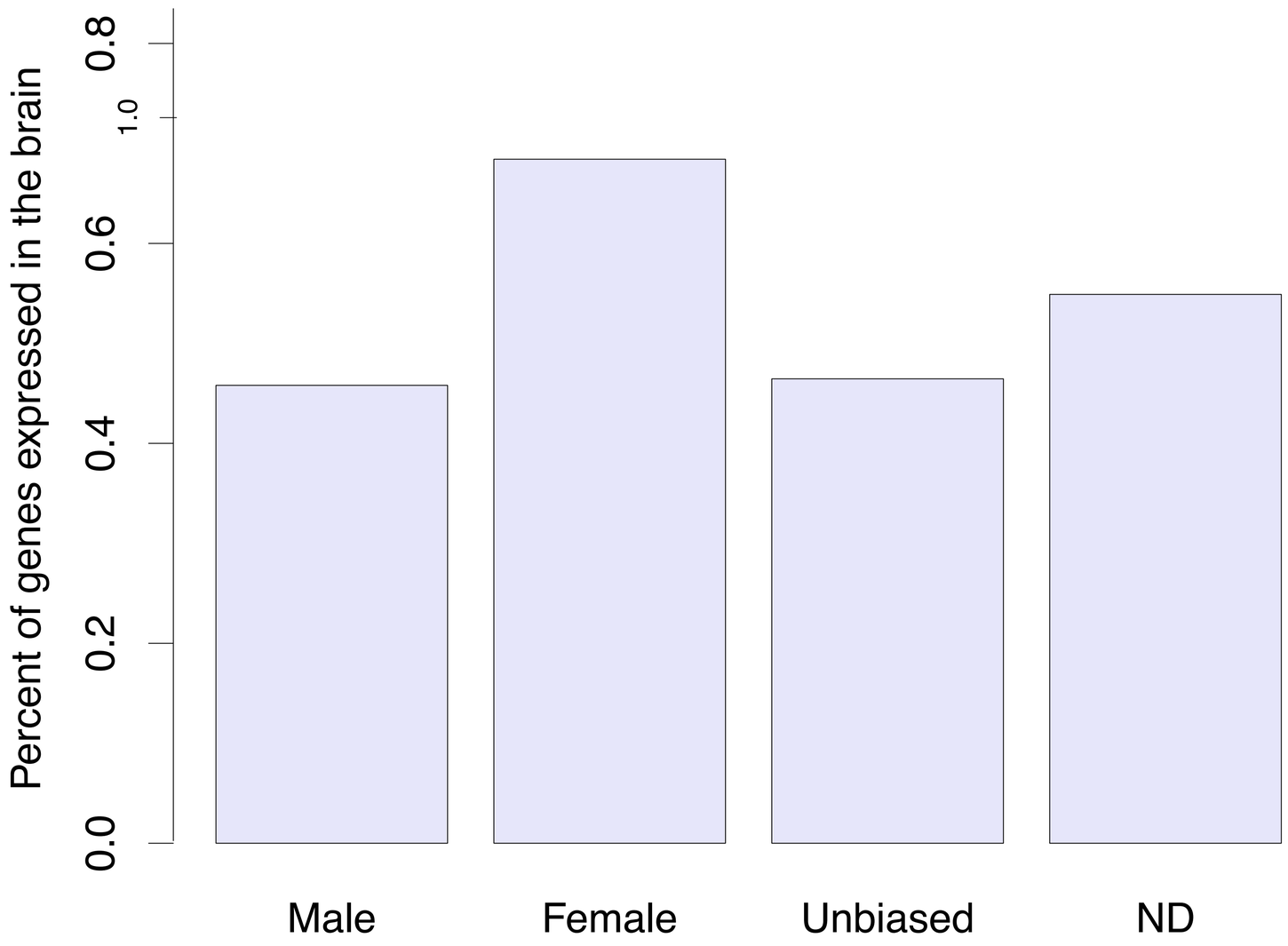
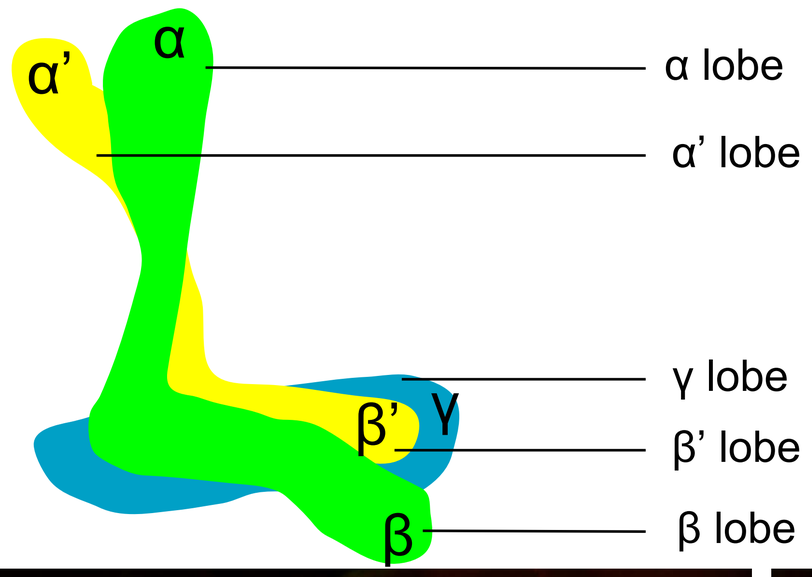
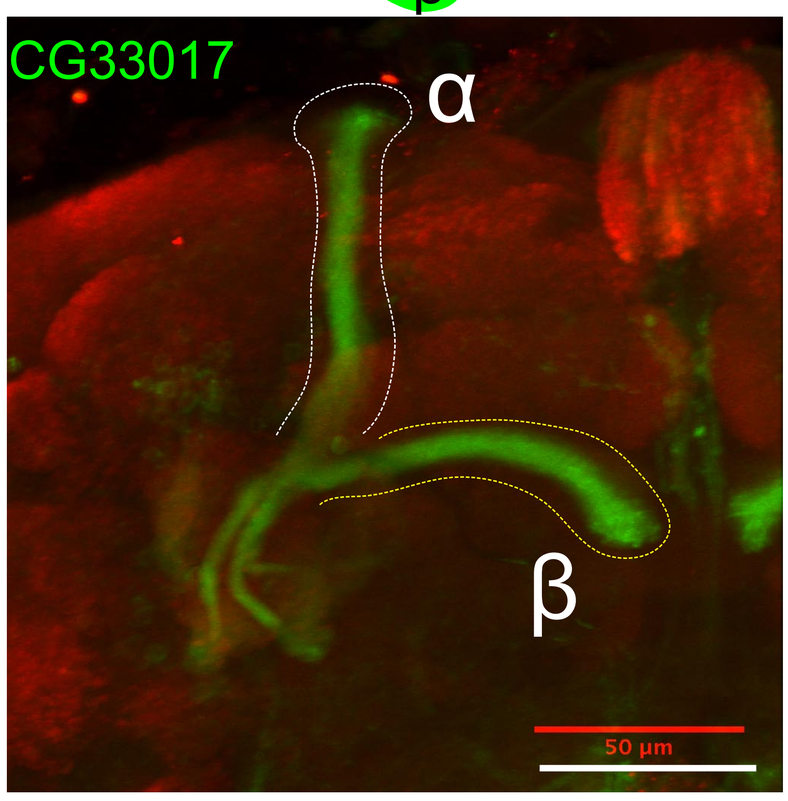


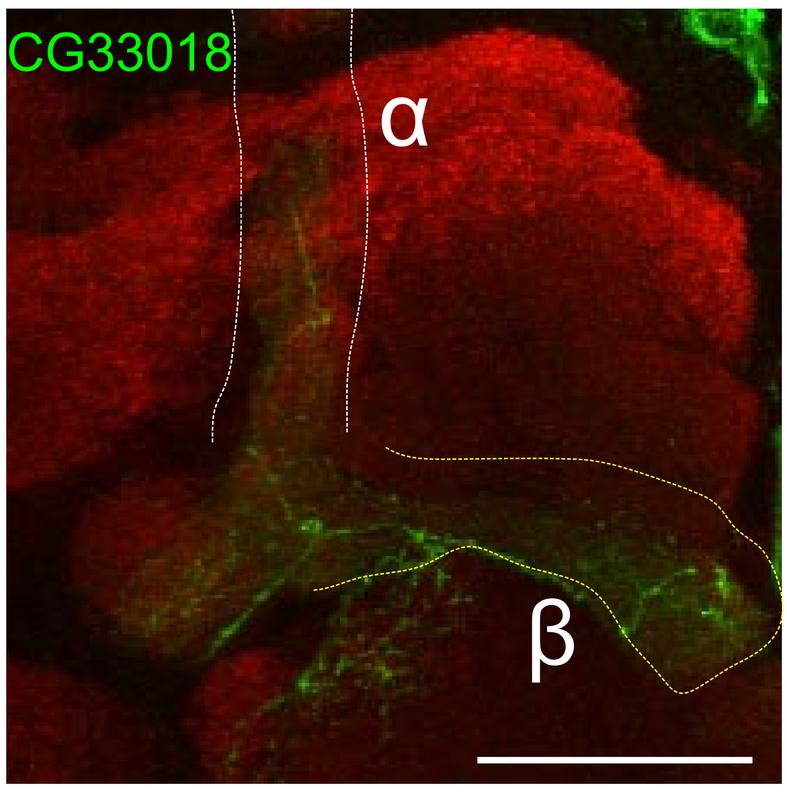
Figure S2

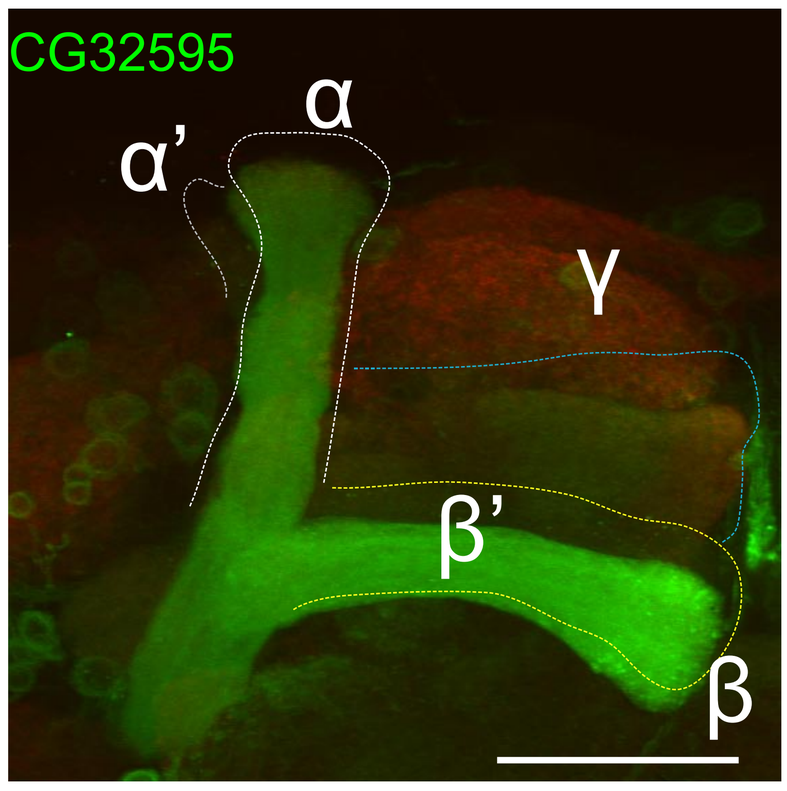
A

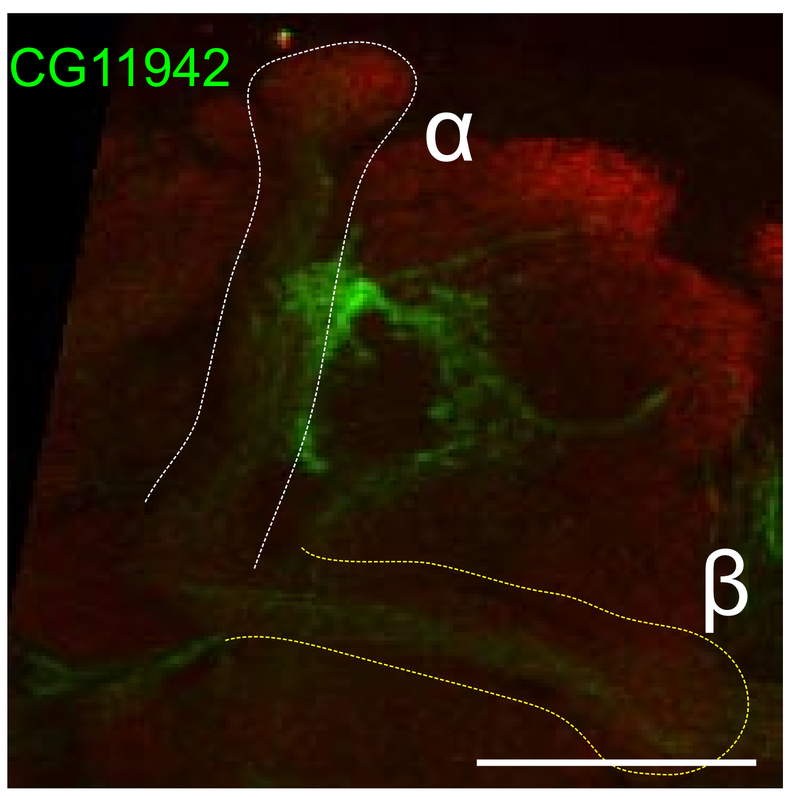

α
α’
α lobe
α’ lobe
γ
β’
γ lobe
β’ lobe
β
β lobe

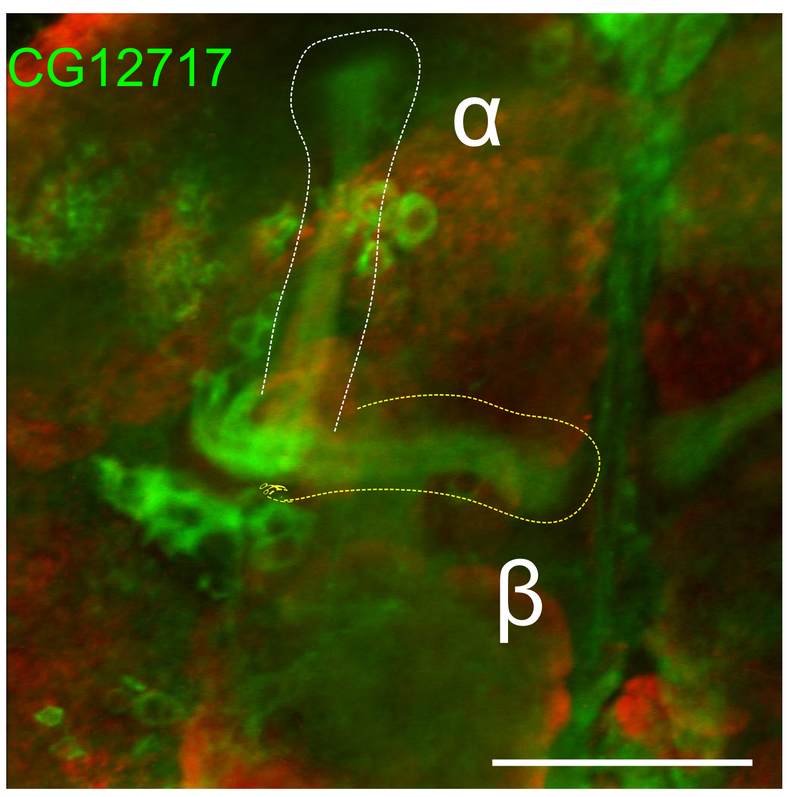
B

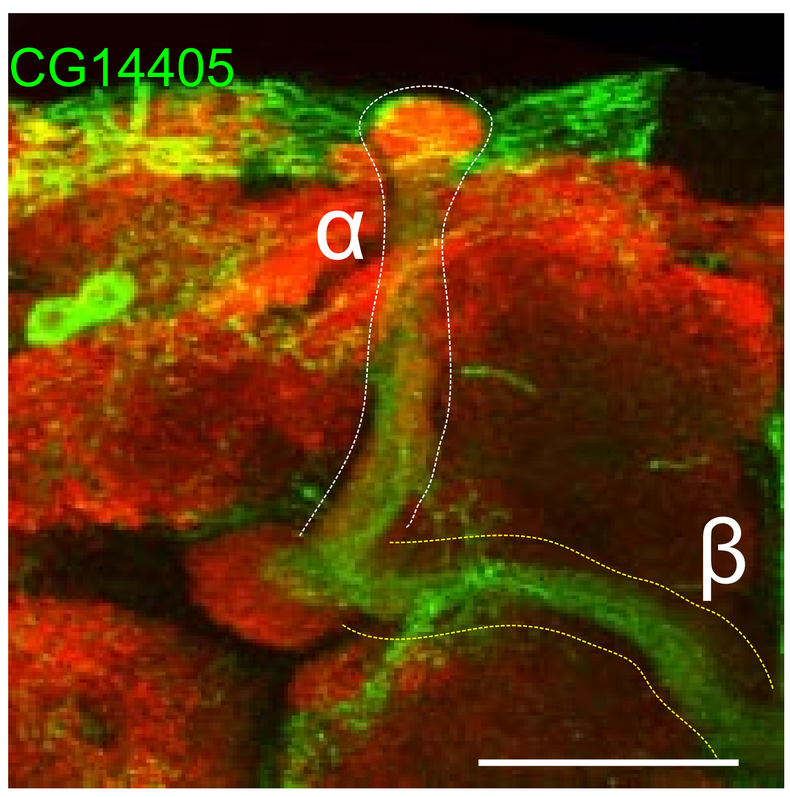

CG33017
α
β
50 µm

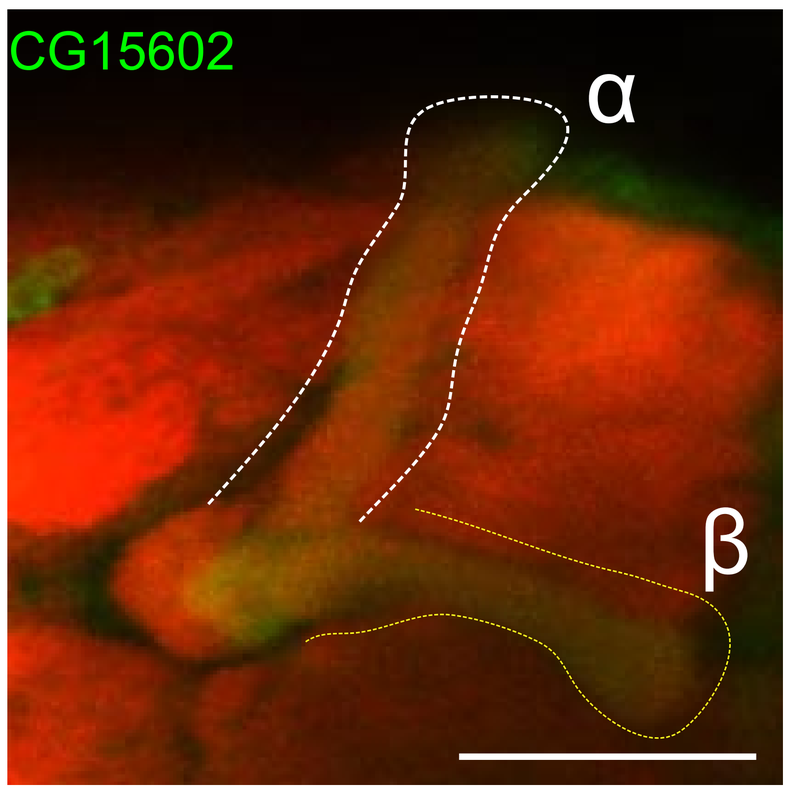

CG33018
α
β

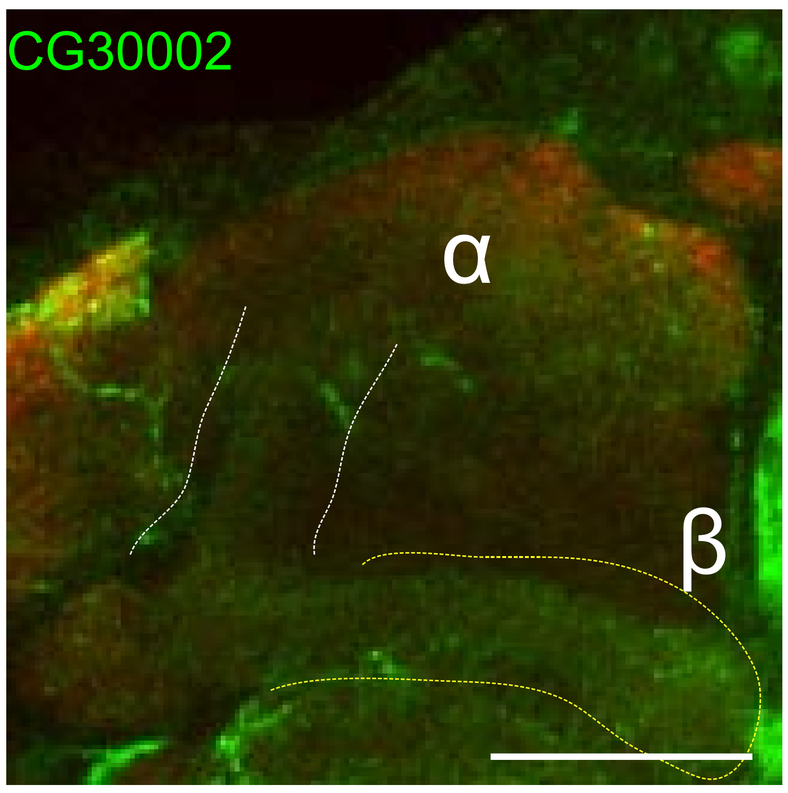

CG32595
α
α’
γ
β’
β

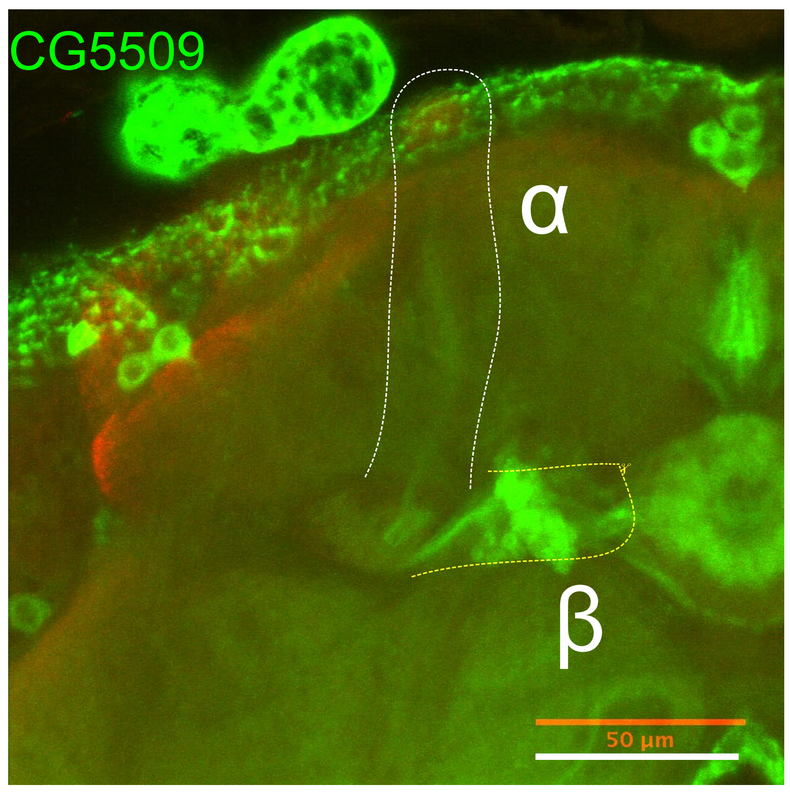

CG11942
α
β

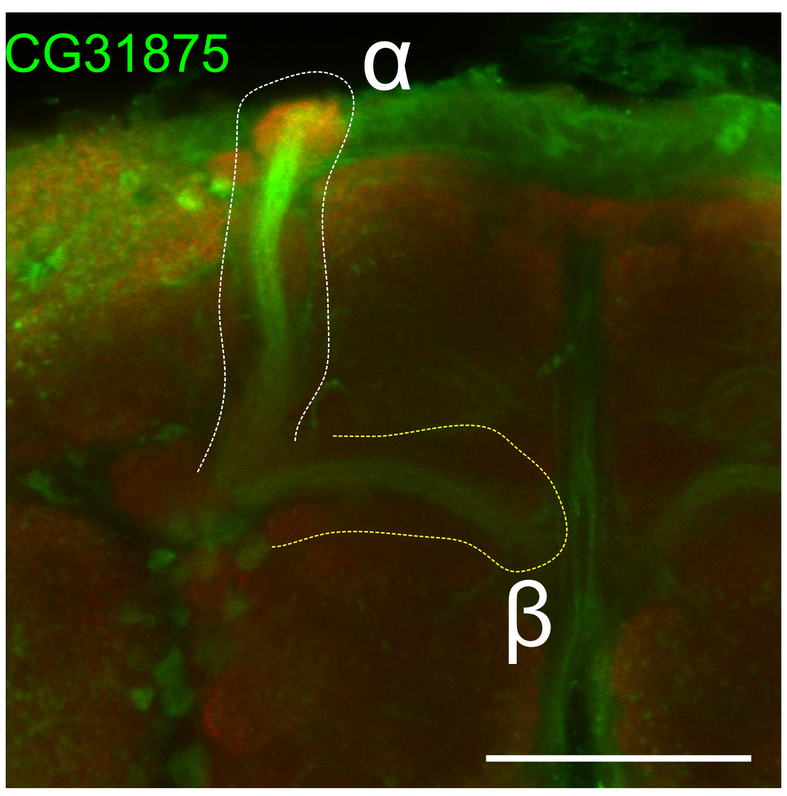

CG12717
α
β

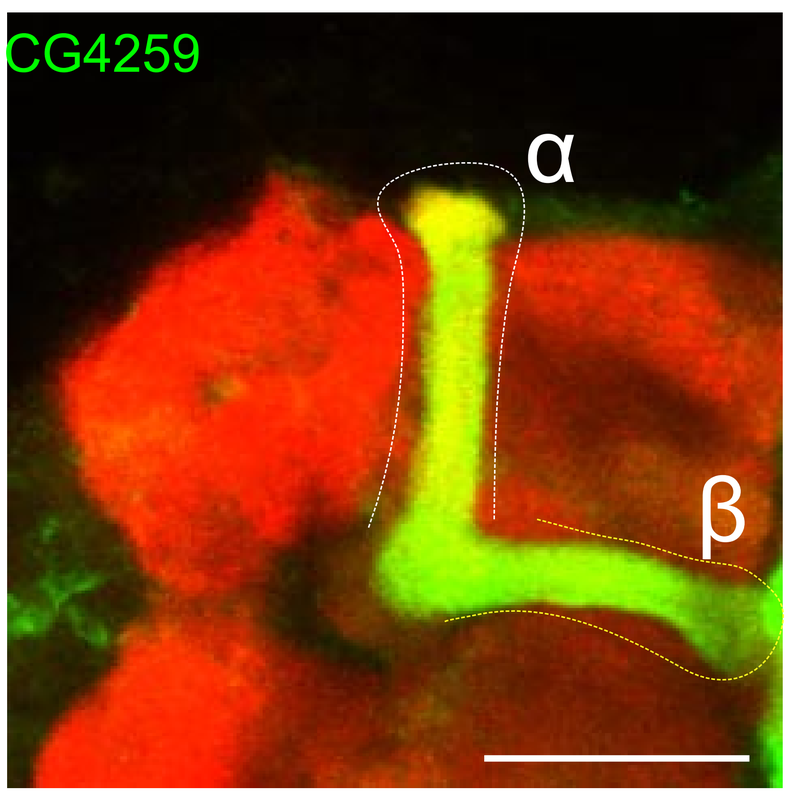

CG14405
α
β


CG15602
α
β


CG30002
α
β


CG5509
α
β
50 µm


CG31875
α
β


CG4259
α
β

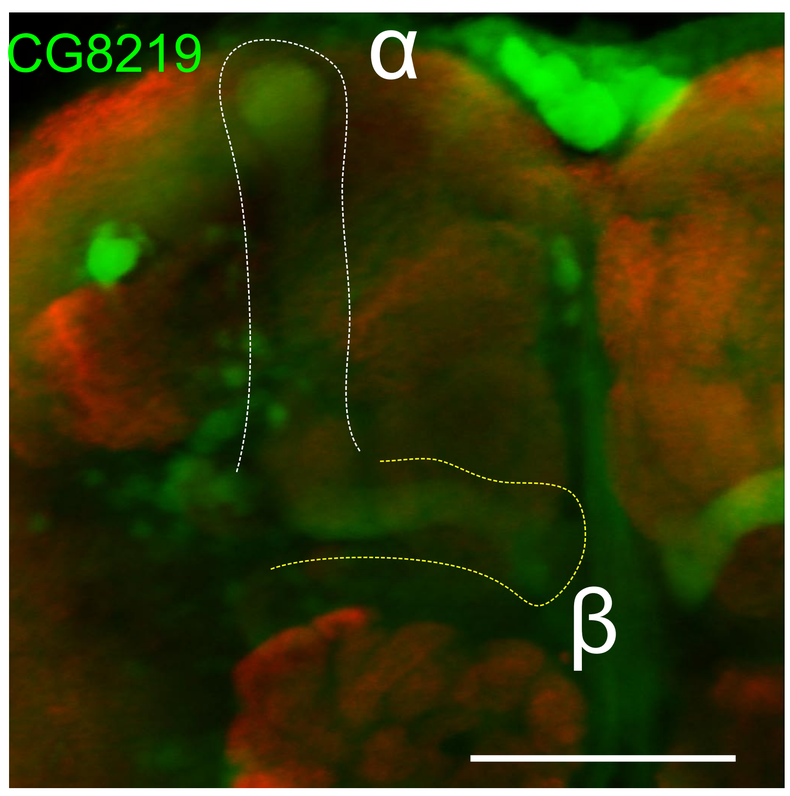
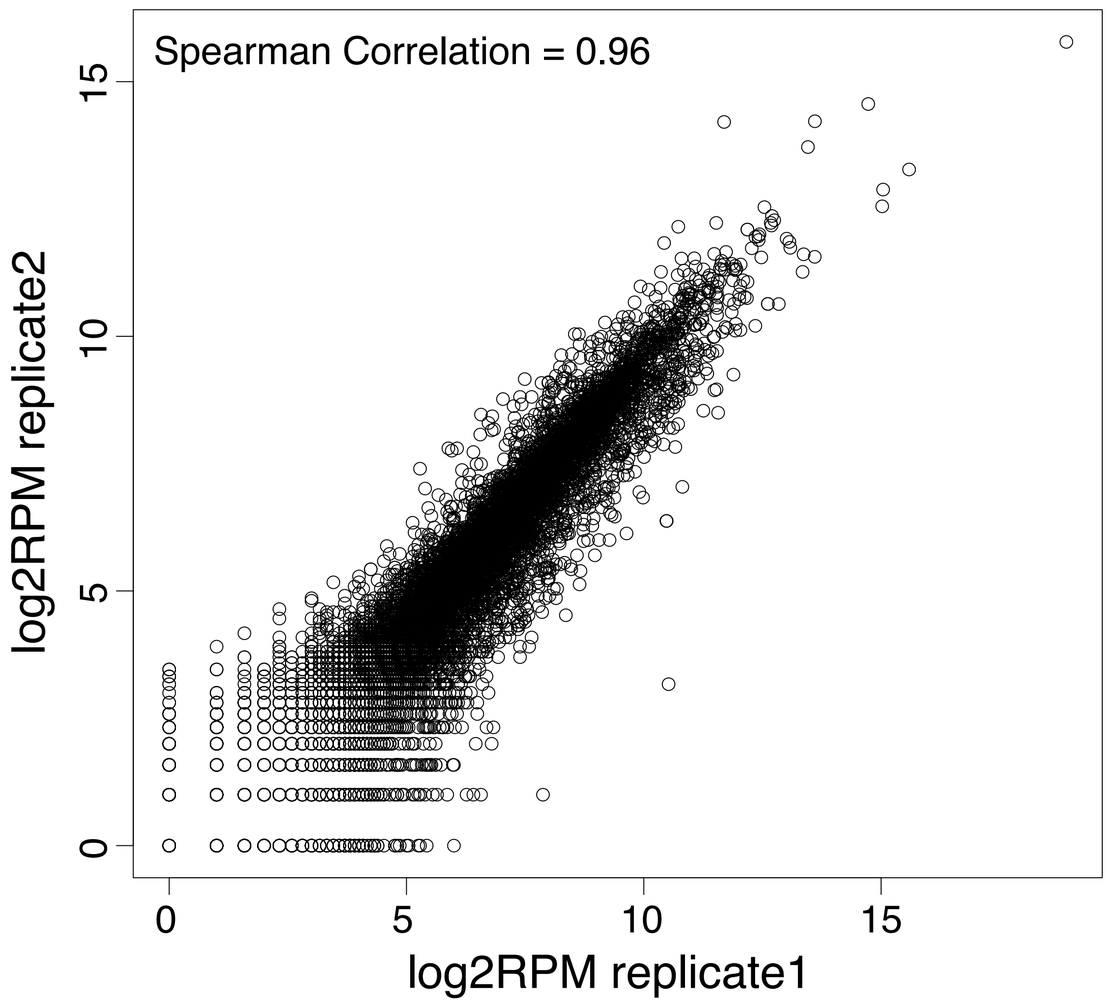
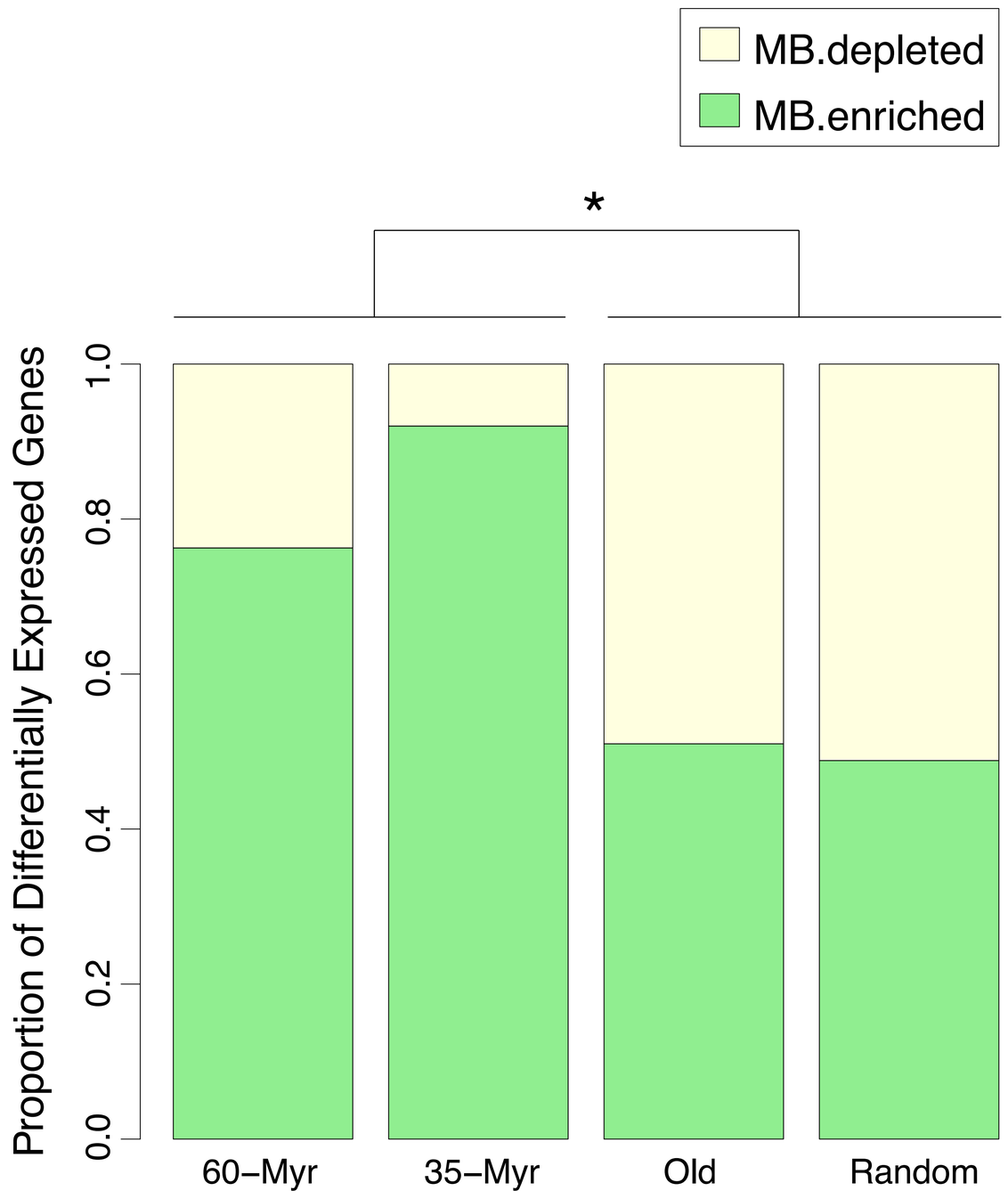

CG8219
α
β

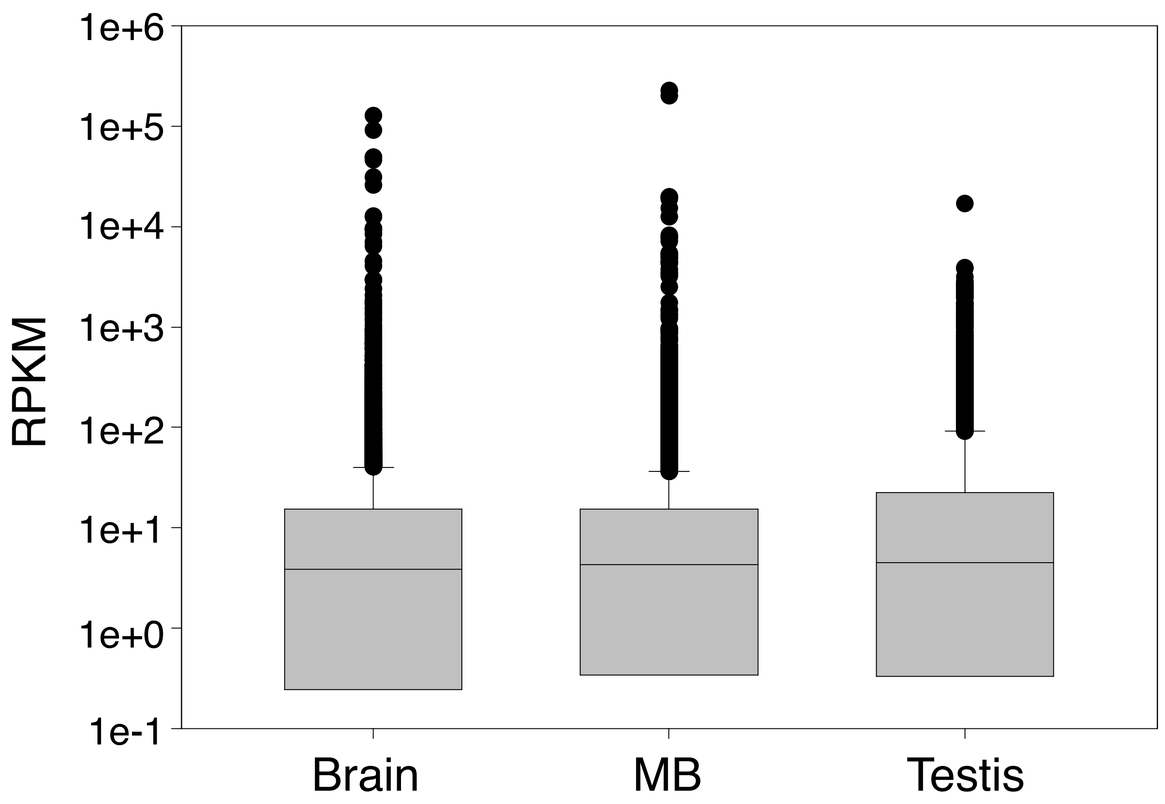
Figure S3

A

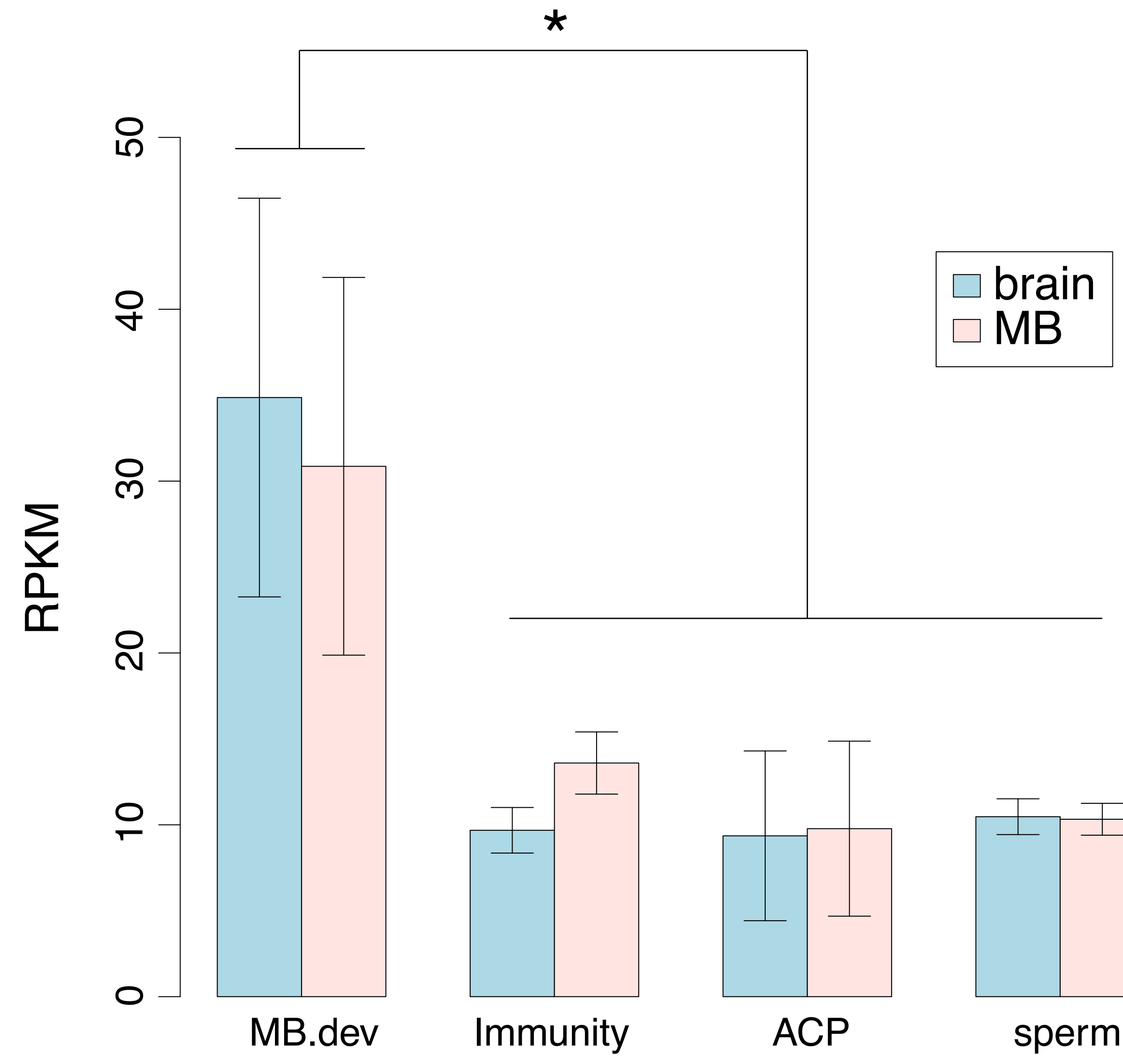


B

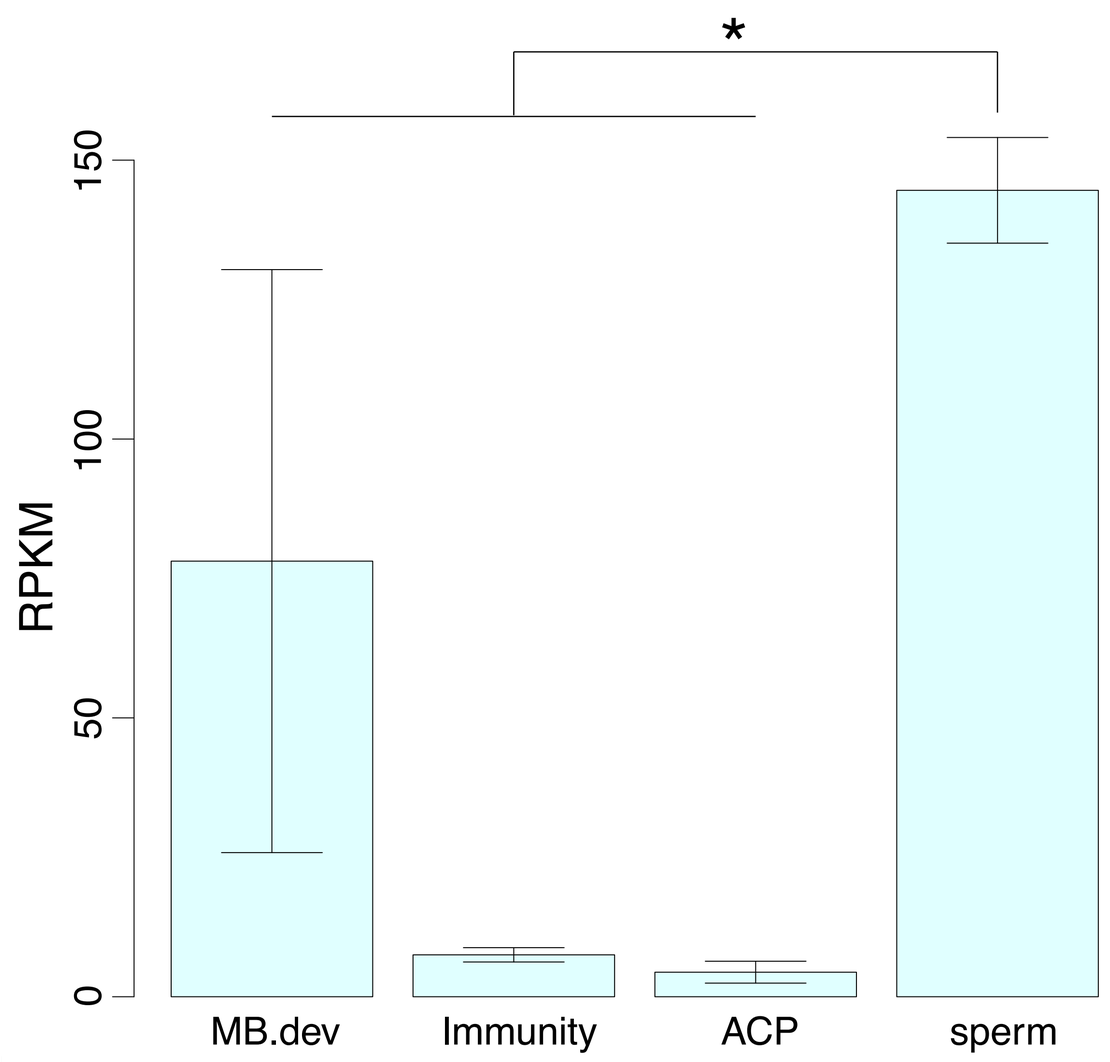




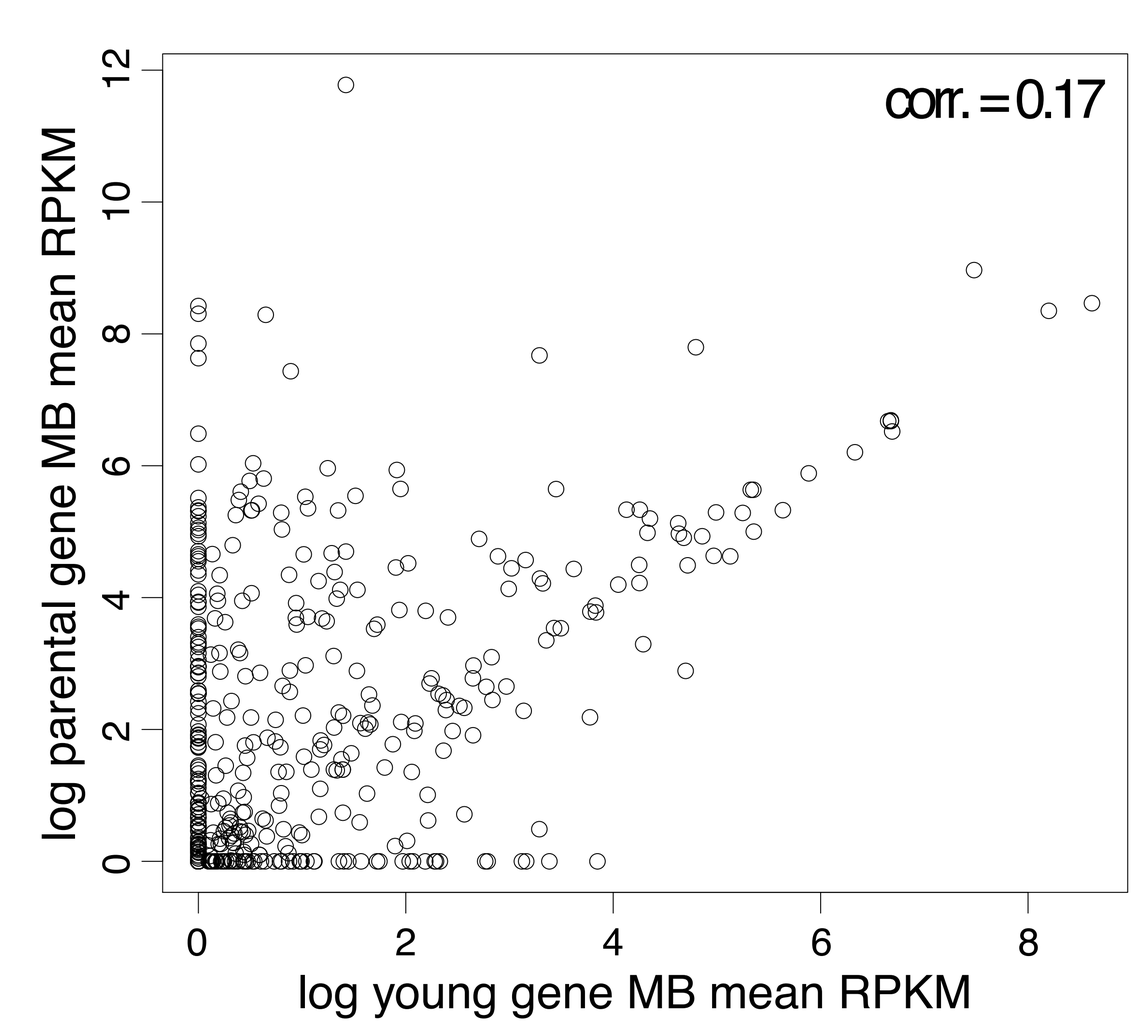
C



D



E



F

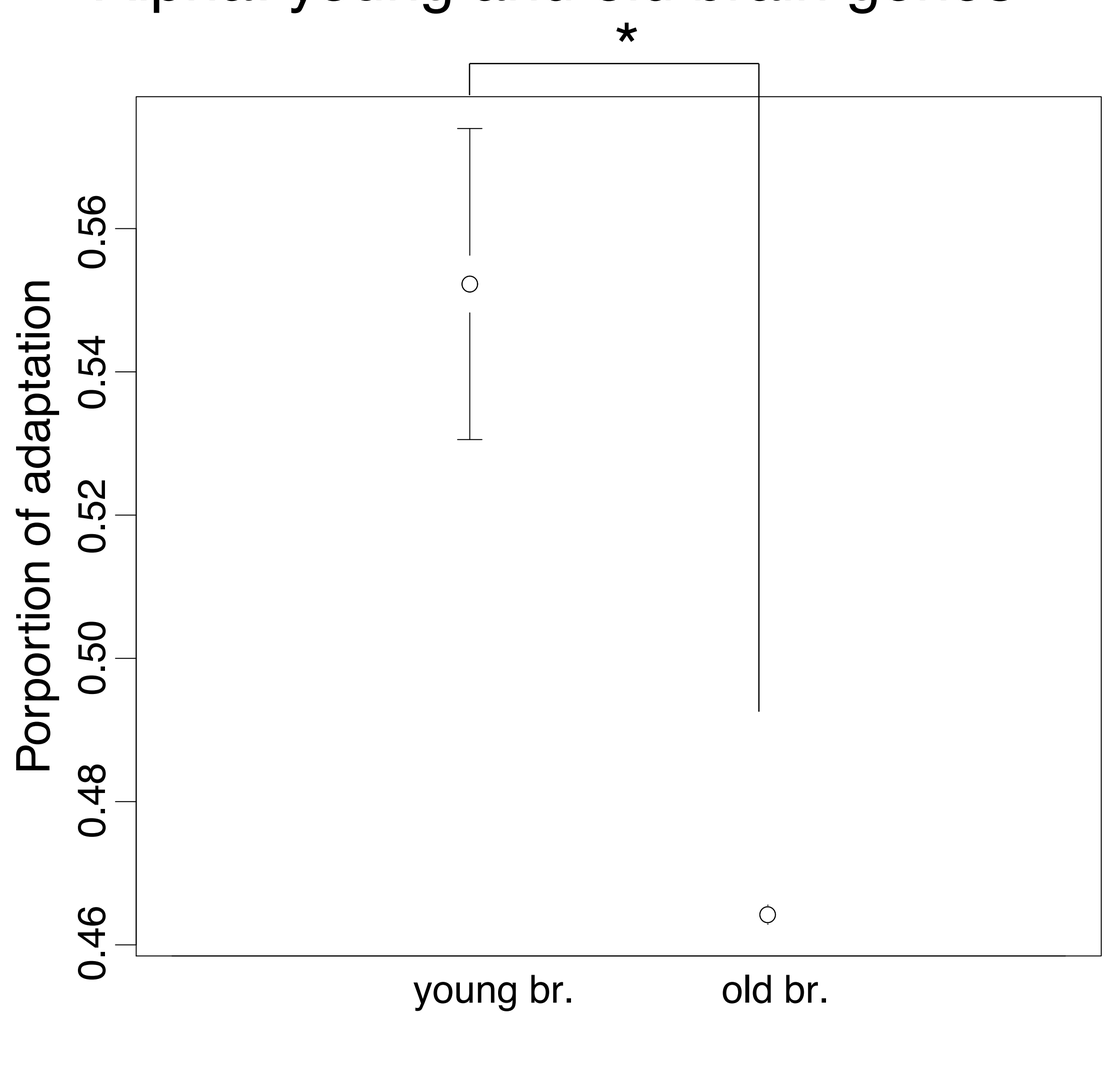
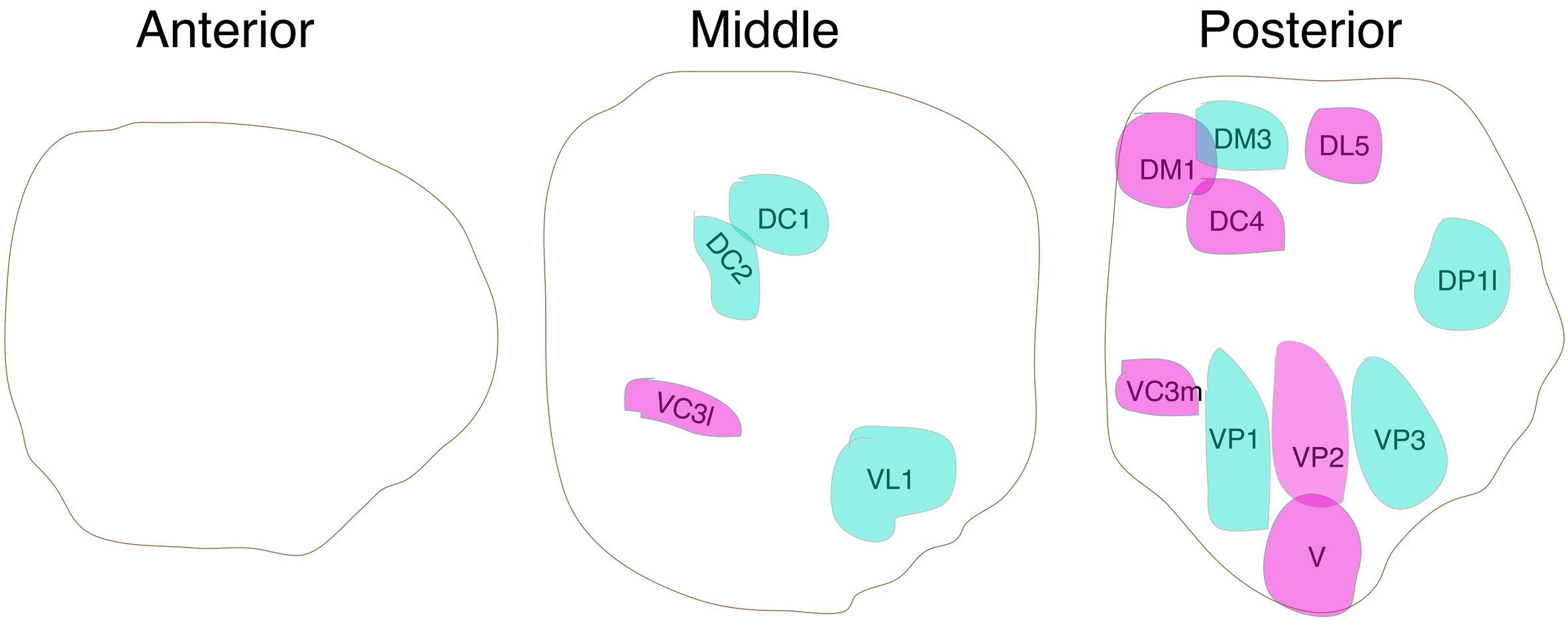


Figure S4

A


Anterior
Middle
Posterior
DC1
DC2
VC3l
VL1
DM1
DM3
DL5
DC4
DP1l
VC3m
VP1
VP2
VP3
V

B

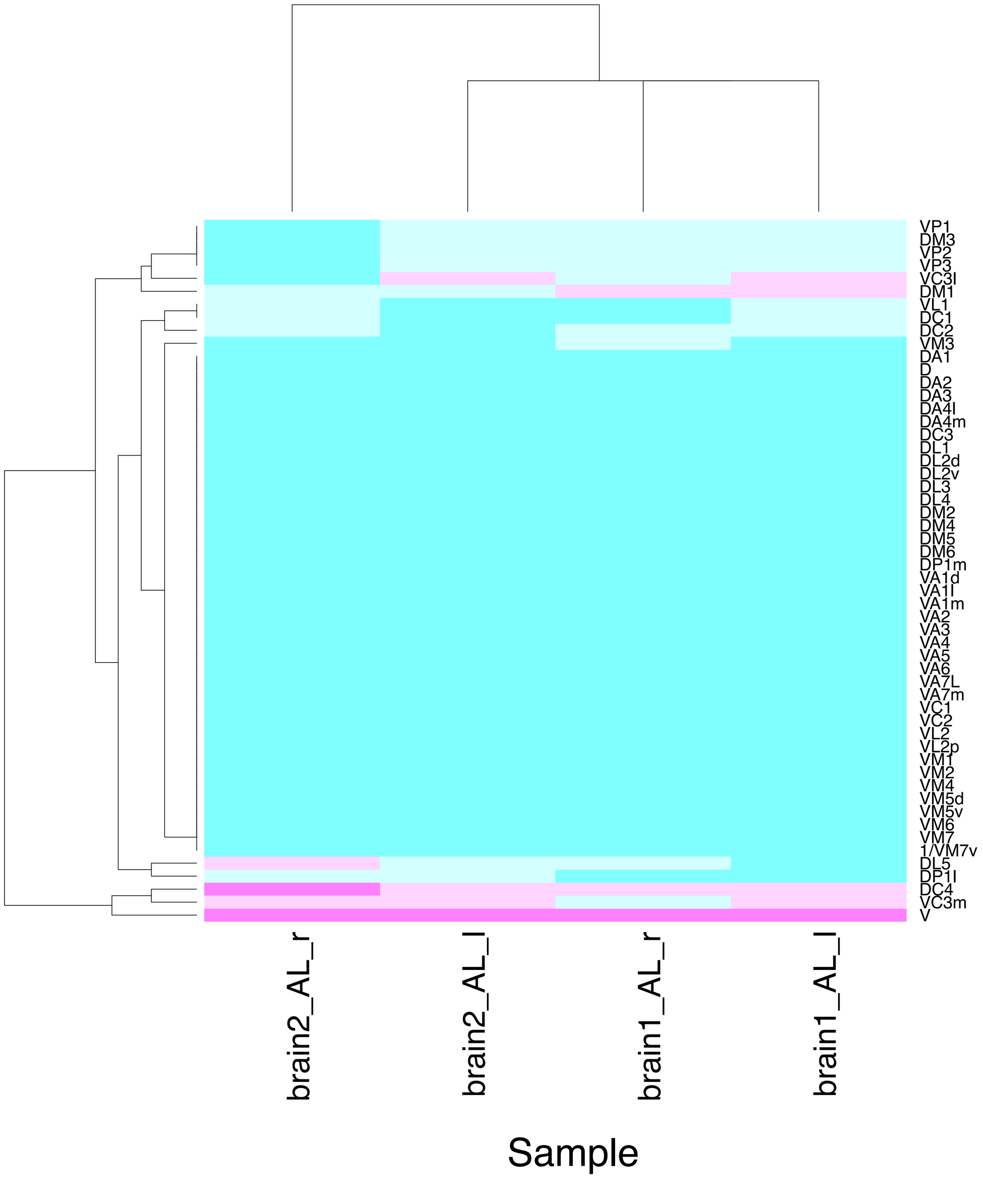
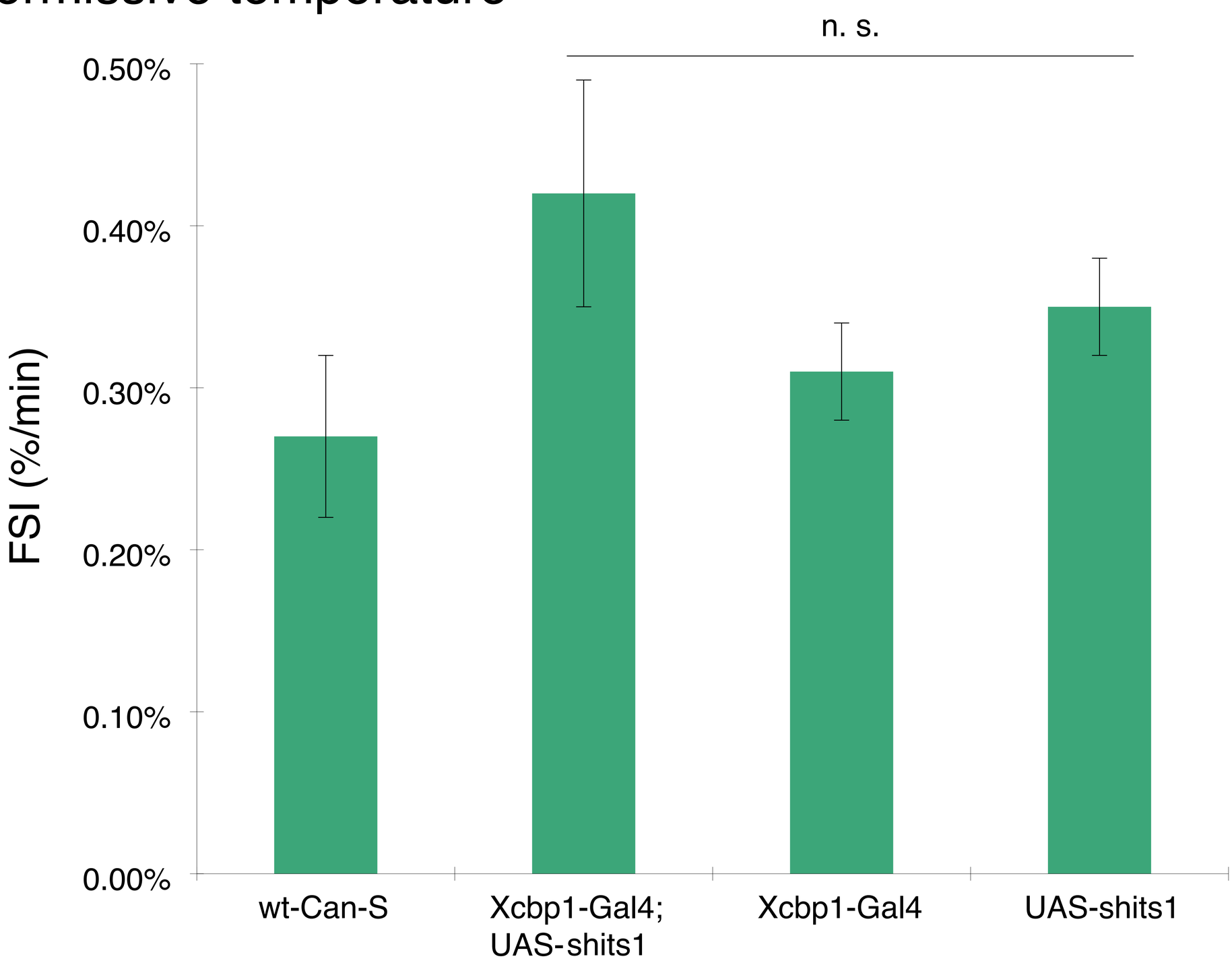

VP1
DM3
VP2
VP3
VC3l
DM1
VL1
DC1
DC2
VM3
DA1
D
DA2
DA3
DA4l
DA4m
DC3
DL1
DL2d
DL2v
DL3
DL4
DM2
DM4
DM5
DM6
DP1m
VA1d
VA1l
VA1m
VA2
VA3
VA4
VA5
VA6
VA7L
VA7m
VC1
VC2
VL2
VL2p
VM1
VM2
VM4
VM5d
VM5v
VM6
VM7
1/VM7v
DL5
DP1l
DC4
VC3m
V
Glomeruli
brain2_AL_r
brain2_AL_l
brain1_AL_r
brain1_AL_l
Sample

Figure S5

A Permissive temperature



B Restrictive temperature

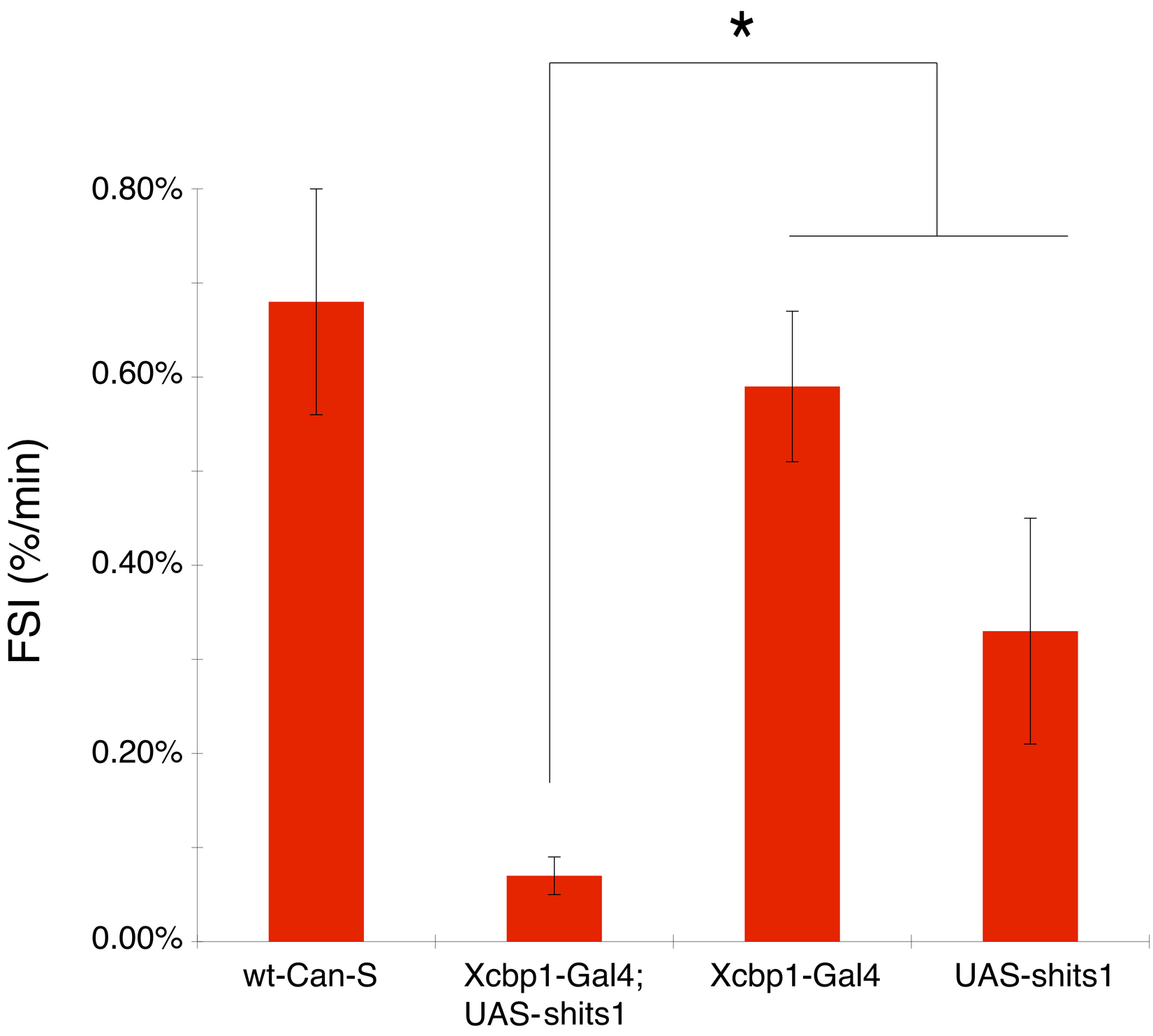
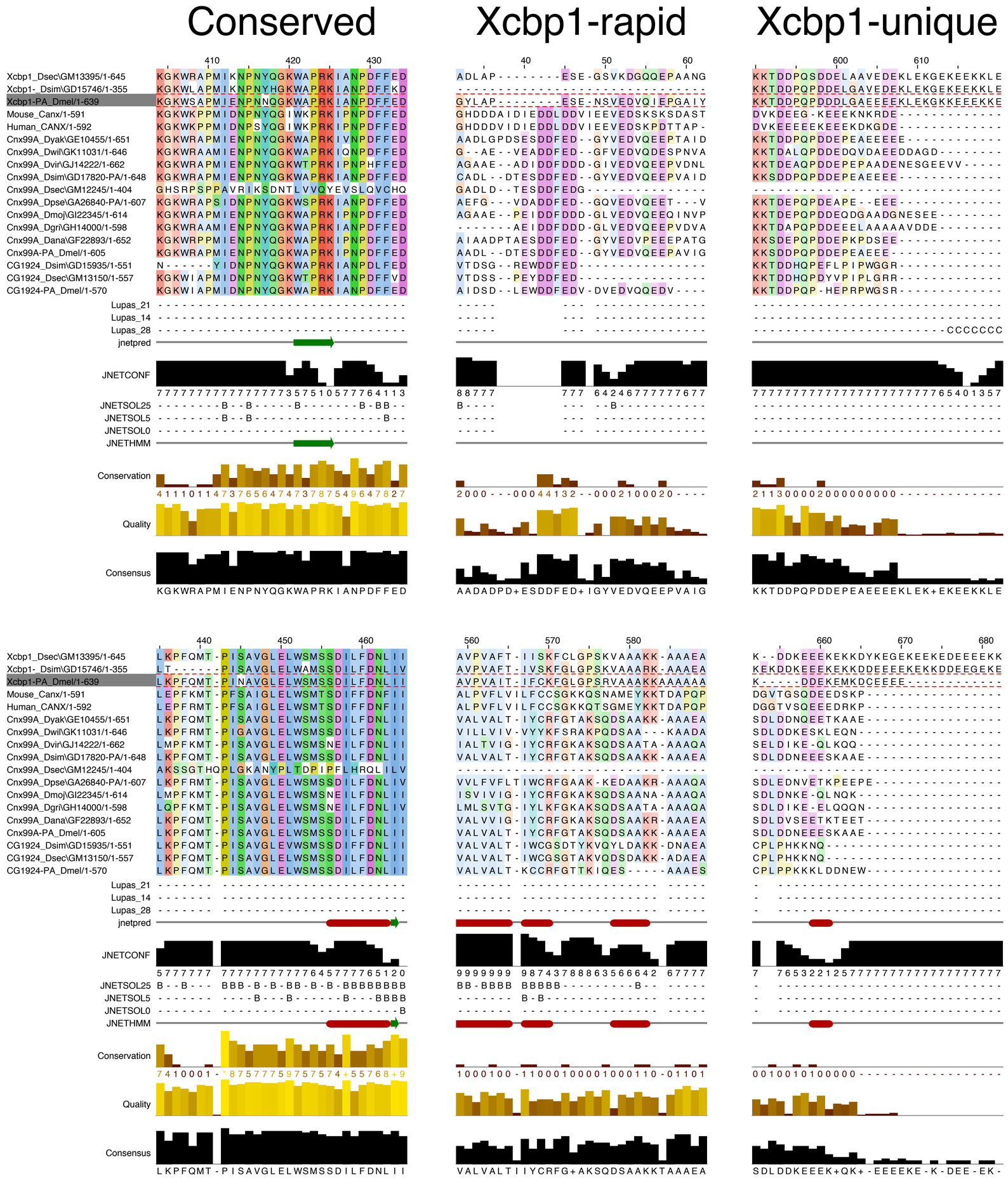
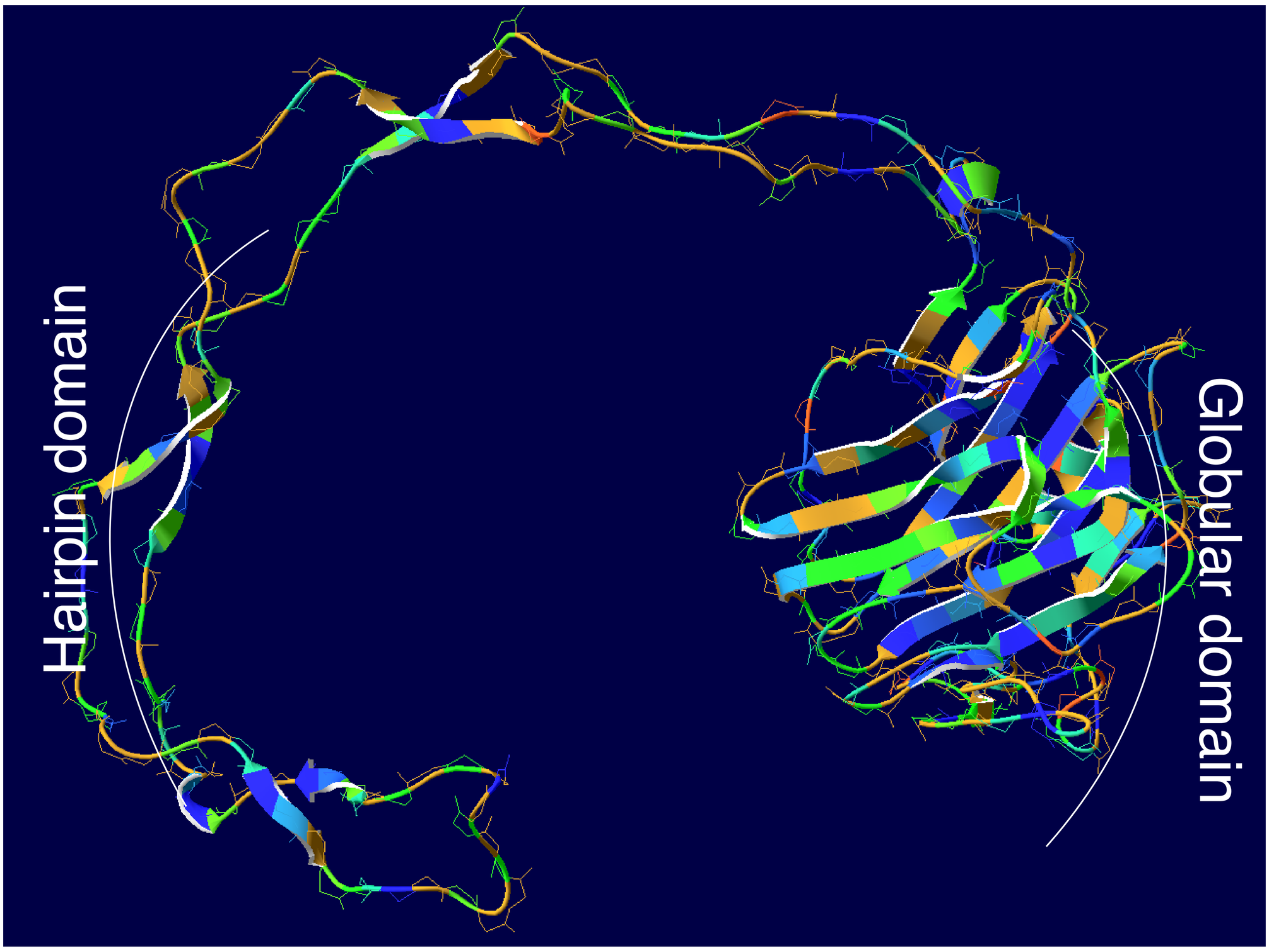


Figure S6

A


Conserved
Xcbp1-rapid
Xcbp1-unique
Xcbp1_Dsec\GM13395/1-645
Xcbp1-_Dsim\GD15746/1-355
Xcbp1-PA_Dmel/1-639
Mouse_Canx/1-591
Human_CANX/1-592
Cnx99A_Dyak\GE10455/1-651
Cnx99A_Dwil\GK11031/1-646
Cnx99A_Dvir\GJ14222/1-662
Cnx99A_Dsim\GD17820-PA/1-648
Cnx99A_Dsec\GM12245/1-404
Cnx99A_Dpse\GA26840-PA/1-607
Cnx99A_Dmoj\GI22345/1-614
Cnx99A_Dgri\GH14000/1-598
Cnx99A_Dana\GF22893/1-652
Cnx99A-PA_Dmel/1-605
CG1924_Dsim\GD15935/1-551
CG1924_Dsec\GM13150/1-557
CG1924-PA_Dmel/1-570
Lupas_21
Lupas_14
Lupas_28
jnetpred
JNETCONF
JNETSOL25
JNETSOL5
JNETSOL0
JNETHMM
Conservation
Quality
Consensus

B


Hairpin domain
Globular domain