Tools for neuroanatomy and neurogenetics in *Drosophila*

Barret D. Pfeiffer*, Arnim Jenett*, Ann S. Hammonds[†‡], Teri-T B. Ngo*, Sima Misra[‡], Christine Murphy*, Audra Scully[§], Joseph W. Carlson[†], Kenneth H. Wan[†], Todd R. Laverty*, Chris Mungall[§], Rob Svirskas*, James T. Kadonaga[¶], Chris Q. Doe[‖], Michael B. Eisen[‡], Susan E. Celniker[†], and Gerald M. Rubin*[§††]**

*Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn VA 20147; [†]Department of Genome and Computational Biology, Life Sciences Division, and **Department of Genome Sciences, Genomics Division, Lawrence Berkeley National Laboratory, Berkeley CA 94720; [‡]Department of Molecular and Cellular Biology and [§]Howard Hughes Medical Institute, University of California, Berkeley CA 94720; [¶]Section of Molecular Biology, University of California, San Diego, La Jolla, CA 92093; and [‖]Institutes of Neuroscience and Molecular Biology, Howard Hughes Medical Institute, University of Oregon, Eugene OR 97403

Contributed by Gerald M. Rubin, April 17, 2008 (sent for review March 13, 2008)

We demonstrate the feasibility of generating thousands of transgenic *Drosophila melanogaster* lines in which the expression of an exogenous gene is reproducibly directed to distinct small subsets of cells in the adult brain. We expect the expression patterns produced by the collection of 5,000 lines that we are currently generating to encompass all neurons in the brain in a variety of intersecting patterns. Overlapping 3-kb DNA fragments from the flanking noncoding and intronic regions of genes thought to have patterned expression in the adult brain were inserted into a defined genomic location by site-specific recombination. These fragments were then assayed for their ability to function as transcriptional enhancers in conjunction with a synthetic core promoter designed to work with a wide variety of enhancer types. An analysis of 44 fragments from four genes found that >80% drive expression patterns in the brain; the observed patterns were, on average, comprised of <100 cells. Our results suggest that the *D. melanogaster* genome contains >50,000 enhancers and that multiple enhancers drive distinct subsets of expression of a gene in each tissue and developmental stage. We expect that these lines will be valuable tools for neuroanatomy as well as for the elucidation of neuronal circuits and information flow in the fly brain.

enhancer | gene expression | promoter | transcription | transgenic

The functional elements of the nervous system and the neuronal circuits that process information are not genes but cells. Consequently, the classic genetic methods that have been so powerful in elucidating embryonic development and other processes in *Drosophila melanogaster* are not adequate to probe the function of the nervous system (1). Instead, we will need to be able to assay and manipulate the function of individual neurons with the same facility as we can now assay and manipulate the function of individual genes.

A variety of genetically encoded probes have been developed that allow researchers to visualize individual neurons to study anatomy, as well as to monitor and modulate the activity of neurons to study physiology and behavior. The utility of these probes is highly dependent on the precision with which their expression can be directed to small subsets of neurons in reproducible, controllable, and convenient ways. The primary objective of the work described in this report was to expand the tools available to accomplish such precise, controlled expression in the nervous system of *D. melanogaster*.

Researchers have known for more than 20 years how to identify and, to some extent, manipulate the promoters and enhancers that control the temporal and spatial expression of individual genes in *Drosophila* (2). This work, and similar studies in other animals, has revealed that the complex spatial and temporal expression pattern of a gene usually results from the combined action of a set of individual enhancer elements that act, in a largely autonomous manner, to dictate aspects of the expression of that gene (3, 4). The number of enhancers per gene varies widely but is generally thought to be in the range of 2 to 10 in *Drosophila* (5).

Because individual enhancers appear to represent the fundamental *cis*-acting modules through which gene expression patterns are generated, our objective was to identify a large set of enhancers that could each reproducibly drive expression of a reporter gene in a distinct, small subset of cells in the adult CNS. Ideally, the number of defined expression patterns should be large enough that, in sum, they would cover the entire brain several times over in a variety of overlapping patterns.

The feasibility of this approach depends on a number of factors. First, enhancers from a wide range of genes whose core promoters contain different sequence motifs must each function robustly when placed in a defined genomic location with a common core promoter. Second, the expression pattern driven by a given enhancer must be highly reproducible from animal to animal. Third, the expression patterns driven by individual enhancers should contain an appropriately small fraction of the cells in the brain to make them useful tools for neuroanatomy and behavioral genetics. Finally, the methods for transgenesis and for identifying suitable enhancers must be efficient enough to permit the generation of the required thousands of transgenic lines. Here, we report the development of a strategy that we believe meets all four of these criteria.

Results and Discussion

Overview of Experimental Strategy. We selected 925 genes for which available expression data or predicted function implied expression in a subset of cells in the adult brain, for example, genes encoding transcription factors, neuropeptides, ion channels, transporters, and receptors [supporting information (SI) Dataset S1]. We spanned the flanking upstream and downstream intergenic regions of these genes, as well as any of their introns larger than 300 bp, with fragments of DNA that averaged 3 kb in length and overlapped (in regions that could not be covered by a single fragment) by ≈1 kb. The fragments were generated by PCR from genomic DNA using primers designed to lie in areas of low evolutionary conservation to minimize disruption of individual enhancers. Because the average size of an enhancer element is only a few hundred base pairs (5), we expected that

Author contributions: B.D.P., S.E.C. and G.M.R. designed research; B.D.P., A.J., A.S.H., T.-T.B.N., C. Murphy, A.S., K.H.W., T.R.L., and C.Q.D. performed research; and G.M.R. wrote the paper. A.J. analyzed the brain images; S.M., J.W.C., C. Mungall, R.S., and M.B.E. contributed informatics methods and analyses; and J.T.K. designed the core promoter.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[††]To whom correspondence should be addressed at: Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147. E-mail: rubing@janelia.hhmi.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803697105/DCSupplemental.

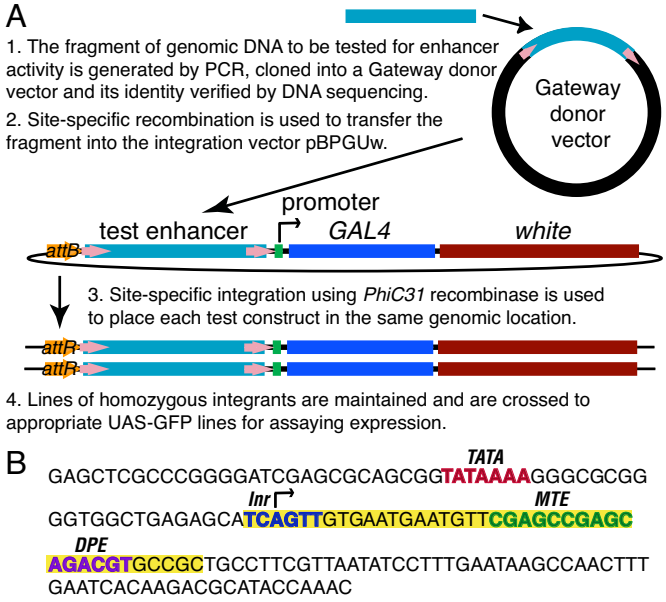

Fig. 1. Strategy for constructing transgenic lines to test DNA segments for enhancer activity. (*A*) Diagram of the vectors and sequential cloning steps. (*B*) Sequence of the *Drosophila* synthetic core promoter (DSCP). Sequences highlighted in yellow were added to the promoter of the *eve* gene. The positions of known promoter motifs are indicated.

nearly all enhancers would be intact in at least one fragment. This process generated 5,200 fragments that were cloned, sequence-verified, and inserted upstream of a core promoter (Fig. 1*A*). In ≈200 cases in which the upstream intergenic region was small, we generated PCR fragments that also contained the start site of transcription and used them to create transcriptional fusion constructs.

Enhancer activity could be tested by imaging the expression patterns that these fragments produce in transgenic animals. In the experiments described here, each enhancer drives the expression of the yeast transcription factor GAL4 (6, 7). We detected the expression of GAL4 either directly by whole mount *in situ* hybridization to its mRNA or by the ability of GAL4 protein to drive the expression of a UAS–GFP fusion gene whose products were then detected by immunocytochemistry and confocal microscopy of whole mount tissue (8).

Drosophila core promoters are ≈80 bp and contain the start site of transcription. An enhancer requires the presence of specific sequence motifs in the core promoter to function properly and the core promoters of different genes vary in their content of these motifs; thus, not all enhancers function efficiently with all core promoters (9). It has been shown that potent core promoters can be created by the incorporation of multiple core promoter motifs into a single promoter (10); more importantly, such promoters would be expected to respond to a wider range of enhancers than naturally occurring core promoters. To assay the enhancer elements from many different genes by using a single core promoter, we constructed a *Drosophila* synthetic core promoter (DSCP) that contains the TATA, Inr, MTE, and DPE sequence motifs (Fig. 1*B*).

We used the *phiC31* site-specific integration system (11) to insert our constructs in the same orientation at the same genomic location. We selected an integration site, *attP2* (Fig. S1) (11), which allows high levels of expression but does not appear to strongly influence the observed pattern of expression of the inserted construct; when inserted at this site, constructs that carry the DSCP but no enhancer lack detectable adult CNS expression (data not shown), many DNA fragments fail to drive any CNS expression, and there are no common pattern elements shared across large numbers of lines that do show CNS expression. Because of the consistent nature of the integration site, we could reliably compare the patterns of expression generated by different enhancer sequences and, once the expression pattern was determined, have confidence that we could drive the expression of other reporter genes in that pattern. Finally, having all constructs inserted at the same genomic location greatly simplifies subsequent genetic manipulations.

Evaluation of a *Drosophila* Synthetic Core Promoter. We compared the expression patterns driven by 40 fragments derived from the dachshund (*dac*), earmuff (*CG31670*), and twin of eyeless (*toy*) genes when the fragments were paired either with their cognate promoter or with the DSCP (Tables S1–S3); the genomic extents of these fragments are shown as blue bars in Figs. 2*A*, 3*A*, and 4*A*. The *dac*, earmuff, and *toy* genes encode evolutionarily conserved transcription factors, with their vertebrate homologs being *Dach*, *Fezl*, and *Pax6*, respectively. An important feature of these genes for our purposes was that they were annotated as having unique transcription start sites, allowing us to select a single endogenous promoter for each.

To compare expression patterns, we established a controlled vocabulary for annotating patterns of axonal and dendritic projections. We first divided the brain into 45 identifiable brain structures, for example, antennal lobe, ellipsoid body, or great commissure. We separately scored each of these regions by using a zero to five scale for three parameters: intensity, distribution, and shape. (See *SI Methods* for a complete description of the controlled vocabulary.) Based on this scoring, we found that the variation between patterns generated with the two promoters was only slightly higher than that seen when comparing the same construct in multiple animals (see below); for 85% of the fragments, the patterns they drove when paired with the DSCP or their cognate promoter were identical in all three parameters in each brain structure. The patterns observed were larger or more pronounced in 10% of the cases with the DSCP and in 5% of the cases with the cognate promoter (see for example Fig. 2 *B–E* and *G–J*); however, even these differences were very subtle. In the embryo, we found that the DSCP routinely drove stronger expression (Fig. 3*B*). These results indicate that, for our purposes, the DSCP serves as an adequate surrogate for the core promoters of individual genes.

Relationship of the Expression Patterns Driven by Individual Fragments to the Expression Pattern of the Gene. Our expectation from previous work was that individual fragments would drive subsets of the endogenous expression pattern of a gene (2–5). However, it was also likely that individual fragments, when taken out of context and freed from negatively acting elements, as well as from the necessity to compete with other enhancers for access to the core promoter, would drive expression in cells where the endogenous gene was not expressed. To address this possibility, we compared the embryonic expression patterns of earmuff (Fig. 3*B*) and *toy* (Fig. 4 *B–D*) with the patterns driven by individual fragments of these genes when combined with the DSCP. As expected, we found that individual fragments generally drove subsets of the wild-type expression pattern and, in sum, appeared to be able to reproduce all of the components of that pattern. However, with some of the fragments, we also saw reproducible expression in cells that do not express the endogenous gene.

The specificity and reproducibility of the patterns driven by individual fragments were illustrated by the expression patterns driven by six different fragments derived from the *toy* gene within the embryonic CNS (Fig. 4 *C* and *D* and Fig. S2). Endogenous *toy* protein was expressed in a highly stereotyped

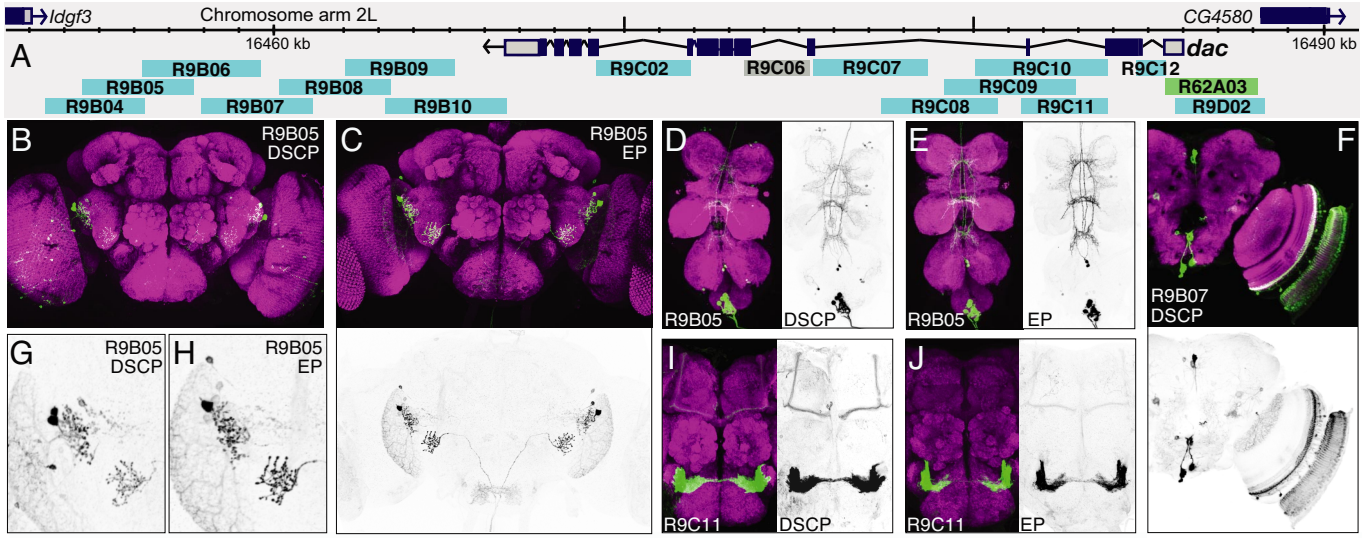

Fig. 2. Testing fragments from the *dac* genetic region for enhancer activity with both the DSCP and the endogenous *dac* promoter. (*A*) Diagram of the chromosomal region surrounding the gene showing the structures of the transcription unit and those of adjacent genes (data taken from *D. melanogaster* genome sequence Release 4.1). The extent and position of the DNA segments that were tested for enhancer activity are shown as light blue bars below the map; R62A03 (green) is a promoter fusion, and we did not obtain data for R9C06 (gray). Total (*B–E* and *G–J*) or partial projections (*F*) of confocal images of the brain (*B*, *C*, *F–J*) or ventral nerve cord (*D* and *E*) of 2- to 5-day old adults of the indicated transgenic line. Fragments were tested with either the DSCP (*B*, *D*, *F*, *G*, and *I*) or the *dac* endogenous promoter (EP) (*C*, *E*, *H*, and *J*). In the color images, gene expression driven by the enhancer fragment is shown in green and the neuropil is counterstained in magenta (see *SI Methods* for details). Embryonic expression patterns are shown in Fig. S3. The gray scale images show only the enhancer-driven expression.

pattern of neurons within the stage 16 CNS: a pair of *toy* medial (TM) neurons, a superficial neuron cluster, a *toy* intermediate (TI) neuron cluster, and three *toy* deep lateral (TDL) neurons. Each fragment drove reproducible expression in a subset of the endogenous *toy*-positive neurons; for example, R9H03 was the only line expressed in the pair of TM neurons, whereas R9G10 was the only line expressed in all three TDL neurons. Each line also showed unique but reproducible expression in a subset of neurons that do not express *toy* protein; for example, R9G09 and R9G10 were the only *toy*-derived lines expressed in the RP2 motor neuron, whereas R9G09 and R9H01 were the only lines expressed in the U motor neurons (data not shown). We conclude that each of these fragments contains sequences that drive expression in a different, reproducible subset of the native *toy* pattern; in addition, when taken out of context, they also drive expression within distinct, reproducible subsets of neurons that do not normally express *toy*.

Enhancers Are Numerous with Each Controlling a Limited Subset of the Total Expression Pattern. We tested 44 fragments derived from genes encoding the transcription factors Earmuff, TOY, DAC, and the G protein-coupled receptor octopamine receptor 2 (Fig. 5 and Table S4) for enhancer activity with the DSCP. Nearly 80% of these fragments generated expression patterns comprising 3 to 1,000 cells in the adult central brain; the central brain corresponds to the brain minus the optic lobes. The mean number of cells showing detectable expression was 95 in these lines; the median number of expressing cells was only 19. These cell numbers were much smaller than observed in a random sample of 27 enhancer trap lines where the observed mean and median were 370 and 180, respectively (Fig. 6); this sample was consistent with the expression patterns generally seen with enhancer trap lines (8). In enhancer trap lines, a transposon carrying a core promoter and a reporter gene is inserted randomly in the genome; the broader expression observed is likely a consequence of individual enhancer trap lines reporting the influence of multiple enhancers.

The patterns driven by a particular fragment are highly dynamic during development. For example, compare fragment R9D11 in the late larva (Fig. 3*F*) and the adult (Fig. 3*J*). The larva showed strong expression in ≈5% of the secondary lineages that produce the cells of the adult central brain, but in the adult central brain, expression is limited to approximately a dozen cells.

Further subdivision of the fragments will be required to determine the extent to which distinct enhancer activities within each fragment can be separated; in the ideal set of lines, each line would represent the expression pattern of a single enhancer. Overlapping fragments often showed overlapping patterns, suggesting that further subdivision would be possible. For example, compare the patterns driven by the fragments R9G08, R9G09, and R9G10, which drove expression in the TI cluster of embryonic TOY-expressing neurons; R9G08 and R9G10 drove expression in distinct subsets, whereas R9G09 drove expression in most or all of the TI neurons (see Fig. 4*D* and Fig. S2).

The Patterns Generated by the Same Enhancer in the Adult Brains of Different Animals Are Highly Reproducible. If the GAL4-expressing lines we created were to have maximum utility, the patterns they produced would have to be highly reproducible from animal to animal. Variability might result from stochastic variation in gene expression activation or in anatomical variation. The degree of variability of adult brain anatomy between individual adult flies of the same genetic makeup has not been well documented. We compared the patterns generated by individual fragments in multiple individuals to address this question. By using the scoring scheme described above and blind study conditions, 95% of the isogenic brains from different animals were scored with identical annotations; even in the 5% that were not scored identically, the differences were subtle. This suggests that variation among animals at the granularity that we were scoring was minimal; however, we did observe that the positions of cell bodies vary much more than arborization patterns. This is illustrated in Fig.

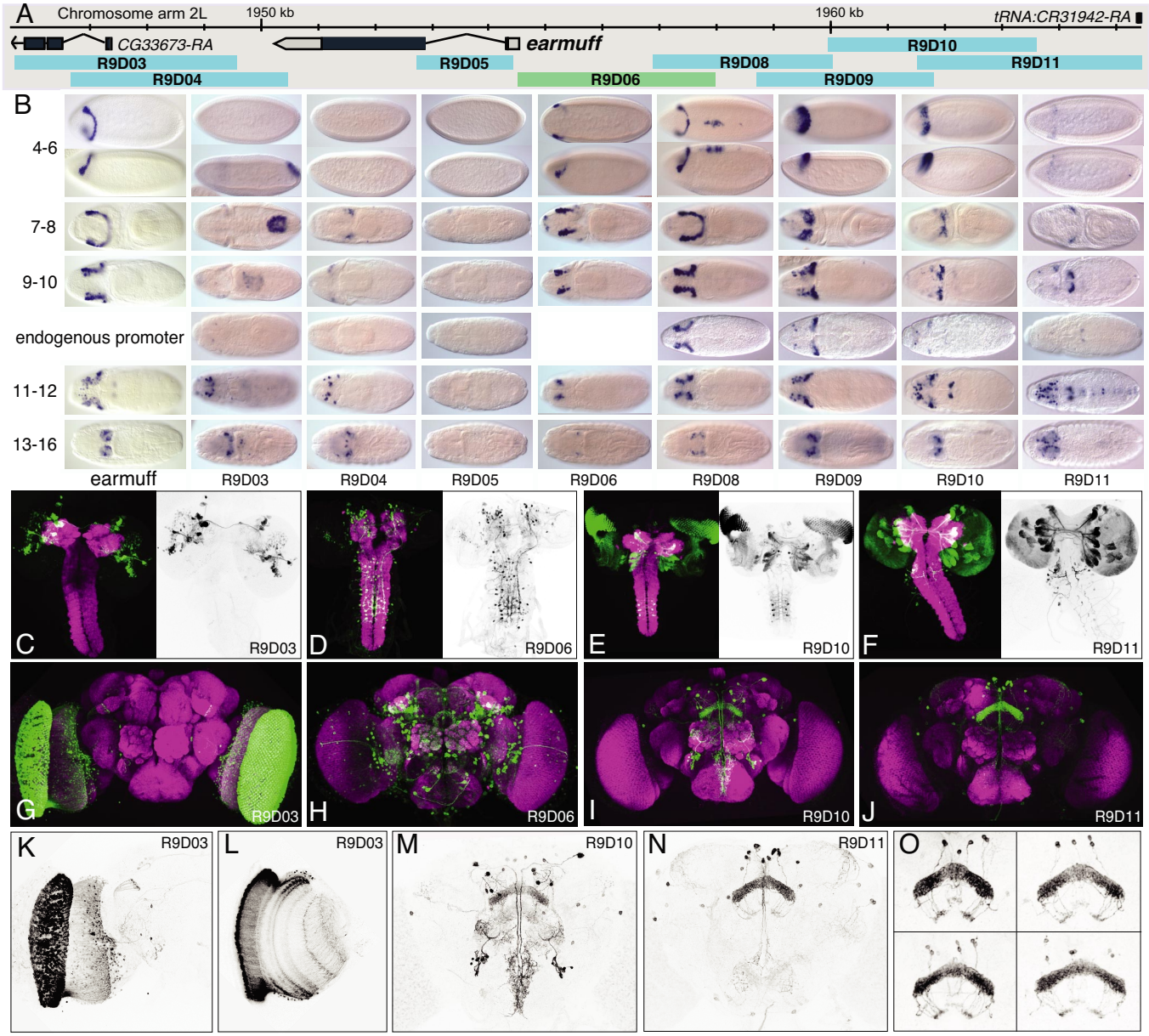

Fig. 3. Patterns generated by fragments of the earmuff gene in embryos, larvae, and adults. (*A*) The genomic map of the earmuff locus. (*B*) Expression of the earmuff gene in embryonic stages 4–6, 7–8, 9–10, 11–12, and 13–16 visualized by whole mount *in situ* hybridization with a probe to earmuff mRNA shown adjacent to the expression produced by the fragments R9D03, R9D04, R9D05, R9D08, R9D09, R9D10, and R9D11 when placed in the enhancer test vector and fragment R9D06 as a promoter fusion. Transgene expression is visualized by whole mount *in situ* hybridization with a probe to *GAL4* mRNA. The enhancer constructs shown use the DSCP; for stages 9–10, we also show data obtained with the endogenous earmuff promoter. Dorsal views are shown except for stages 4–6, where a lateral view is also shown below the dorsal view; anterior is at *Left*. (*C*, *D*, *E*, and *F*) Expression driven by the indicated fragment in late third instar larvae. The clusters of labeled cells seen in *F* represent distinct lineages of secondary neurons; this labeling is not maintained in the adult (*J*). (*G*–*O*) Expression in the adult brain of the indicated lines. A total projection (*K*) and single optical section (*L*) of the optic lobes of the brain shown in *G*. (*O*) Expression in the fan-shaped body of line R9D11 in four different brains.

3*O*, which shows the expression pattern of line R9D11 in the fan-shaped body of four animals.

Concluding Remarks. Our results indicate that it should be possible to establish a collection of transgenic *Drosophila* lines, each directing expression to a small subset of cells in the adult brain, and that, in sum, would cover all cells in the brain. More than 80% of the 44 fragments we tested from four genes gave expression in the adult brain, suggesting that we would generate >4,000 lines with patterns from the analysis of our initial 5,200 putative enhancers. More than half of the fragments we tested gave expression in 10 to 200 cells: 0.03–0.7% of the 30,000 cells estimated to comprise the central brain (W. Pereanu, personal communication). We believe this fraction of cells is a useful number for anatomical, physiological, and behavioral studies. We recognize that the preferred cell number for behavioral studies is not known and will certainly depend on the particular behavior and assay; however, we expect that our lines, singly or in combination, will provide the versatility needed to generate expression patterns of the desired sizes. It is possible that the 2,000 lines we expect to generate having expression patterns within the 10 to 200 range of cell numbers will be adequate to

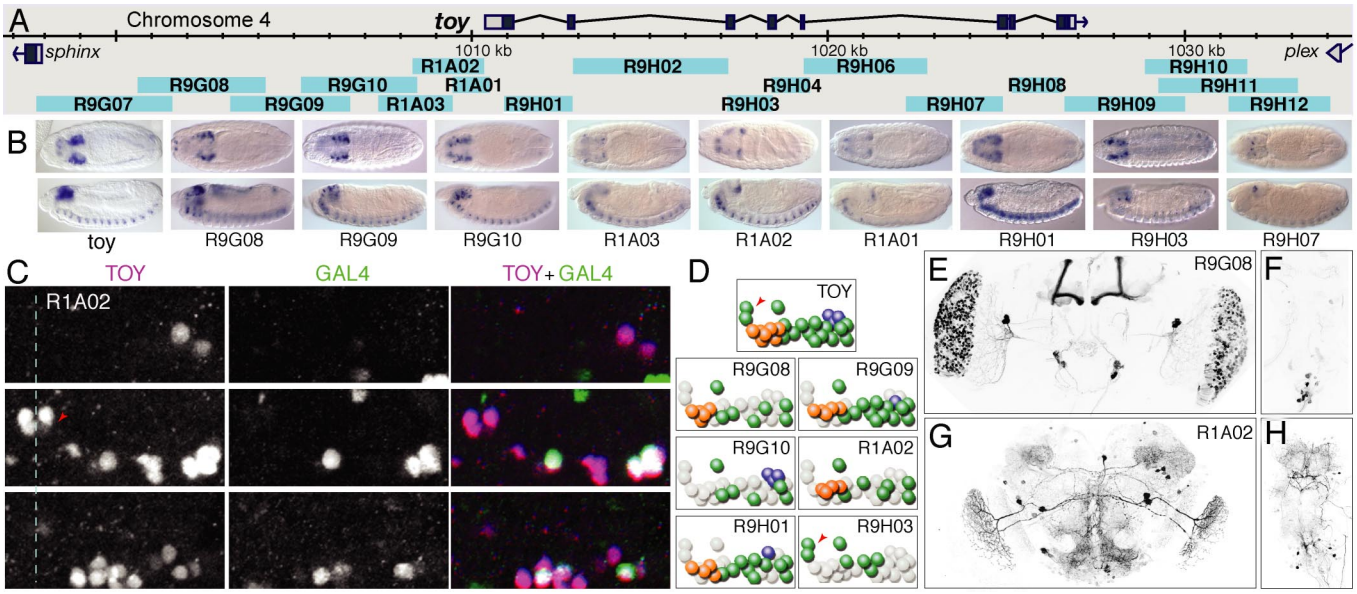

Fig. 4. Patterns generated by fragments of the *toy* gene and a comparison with the expression pattern of the endogenous *toy* gene in the embryo. (*A*) Genomic map of the *toy* locus and the positions of the tested fragments. (*B*) Expression of the endogenous *toy* mRNA and the expression of *GAL4* mRNA driven by the indicated nine fragments shown in stage 13–16 embryos; the other nine fragments shown in *A* did not drive detectable expression at this stage. Dorsal (*Upper*) and lateral (*Lower*) views are shown; anterior is at *Left*. (*C*) Endogenous *toy* protein (magenta in the merged image) and nuclear localized GFP (green in the merged image) expression driven by the R1A02 fragment in an abdominal CNS hemisegment of a stage 16 embryo (anterior, up; midline, dashed line). Three focal planes are shown: deep (*Top*), intermediate (*Middle*), and superficial (*Bottom*); see Fig. S2 for similar data on other fragments. (*D*) Diagrams of endogenous *toy*-positive neurons and the subset of *toy*-positive neurons in which each indicated fragment drives expression (deep neurons, blue; intermediate neurons, green; superficial neurons, orange); each fragment also drives expression in a reproducible set of *toy*-negative neurons that are not shown in these diagrams. The TM neurons are indicated by the red arrowheads. (*E–H*) Total projections of confocal images of the adult brain showing enhancer fragment driven expression in the brain (*E* and *G*) or ventral nerve cord (*F* and *H*) of lines R9G08 (*E* and *F*) and R1A02 (*G* and *H*).

represent all cells in the central brain in a variety of overlapping patterns. If not, it is straightforward to generate additional lines.

Our results reveal several features of genome organization. The number of distinct patterns we observed implies that there are likely to be >50,000 transcriptional enhancers in the *Drosophila* genome; >80% of the fragments we tested showed enhancer activity, it would take ≈50,000 such fragments to cover the entire genome, and many of these fragments are likely to carry multiple enhancers. A gene that is expressed in many tissues and developmental stages might have a single enhancer that controls all aspects of expression in a given stage or tissue; for example, an "adult brain enhancer". Alternatively, the expression pattern at each stage and tissue might be generated by the sum of the actions of multiple enhancers, each controlling a subset of the pattern. Our data strongly favor the latter possibility. Furthermore, our results suggest that enhancers are reused throughout development, although the resolution of our current experiments was not sufficient to distinguish two closely linked enhancers from a single enhancer. It is also clear that enhancers, taken out of context, in addition to driving a subset of the expression pattern of the endogenous gene, often show highly stereotyped expression that is not displayed by the endogenous gene; this expression may reflect either the absence of competition between enhancers or the separation from repressive elements.

We believe that the lines we have generated, where a molecularly defined DNA fragment drives expression, have several

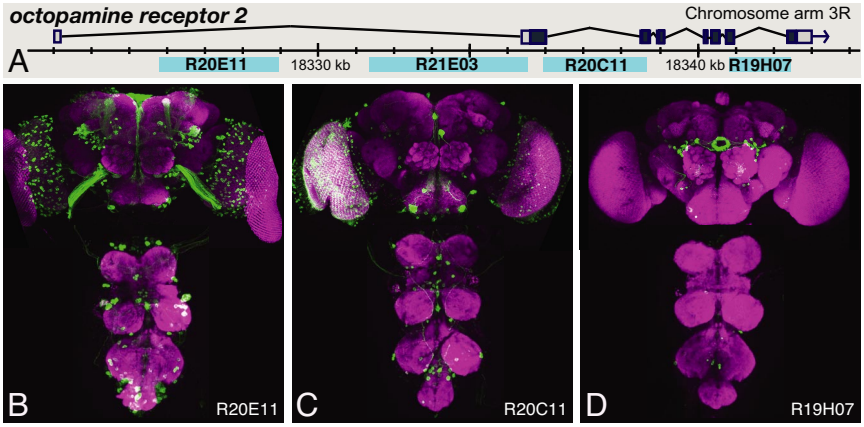

Fig. 5. Distinct expression patterns generated by fragments of the octopamine receptor 2 gene. (*A*) Diagram of the genomic locus. (*B–D*) Expression driven by the indicated fragments in the adult brain and ventral nerve cord.

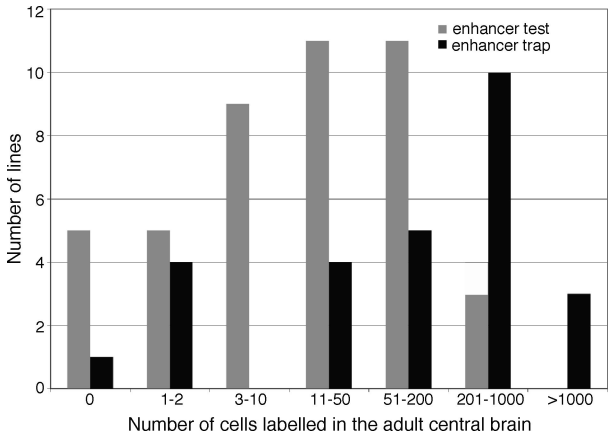

Fig. 6. The patterns driven by individual fragments generally contain fewer cells than those found in enhancer trap lines. The gray bars in the histogram show the number of cells found in the patterns within the central brain of the adult generated by the 44 enhancer fragments shown in blue in Figs. 2*A* (*dac*), 3*A* (earmuff), 4*A* (*toy*), and 5*A* (octopamine receptor 2) genes. The black bars show the number of cells found in 27 enhancer trap lines that were chosen randomly from the unpublished collection of J. Simpson and B. Ganetsky and are typical of the type of patterns seen in other enhancer trap collections (8).

advantages over existing tools for neuroanatomy and neurogenetics. The patterns we observed were less broad than those observed in enhancer trap lines, and our constructs were all at the same genomic location, facilitating subsequent genetic manipulations. It is also straightforward to attempt to produce smaller patterns by subdividing the fragments; these fragments were each large enough to carry several distinct enhancers. Most importantly, because our constructs were all inserted at precisely the same genomic location, the effects of local genomic environment on expression could be held constant. Thus, we should be able to change the gene whose expression is driven by a particular enhancer and have confidence that the expression pattern would remain the same. In this way, we could readily produce lines that drive, instead of GAL4, the expression of other transcription factors such as LexA (12), inhibitors of transcription factor function such as GAL80 (13), or recombinases such as Flp (7, 14) in the same patterns that we have determined for GAL4.

The small number of cells labeled in our lines can facilitate anatomical analysis, and the imaging of the expression patterns of several thousand such lines will provide a good overall view of the morphological range of neurons present in the fly brain and their projection patterns. Such studies can be aided by expression of markers directed to axons, dendrites, or synapses (15, 16). Stochastic labeling of individual cells within these patterns can be accomplished by using recombinases to facilitate detailed anatomy of individual cells (13, 14). It will be informative to examine the extent to which the cells that express a given enhancer share developmental, physiological, or functional properties and how shared activation of a given enhancer is related to the concept of cell type.

Finally, these lines will provide the ability to express genetically encoded indicators of function (17) or modifiers of neuronal activity (18) in well defined small subsets of neurons. We are optimistic that learning how many different behaviors are modified when the function of each of these small cell populations is altered will provide useful insights into the organization of neuronal circuits and information flow within the fly brain.

Methods

Standard molecular and histochemical methods were used; details of the constructs and protocols are given in *SI Methods*. Vectors are available from Addgene.

ACKNOWLEDGMENTS. Garson Tsang, Gina Dailey, Martha Evans-Holmes, and Soo Park provided technical assistance in the molecular biological aspects of this work; Richard Weiszmann and Amy Beaton assisted with *in situ* hybridization and imaging of embryos; Don Hall, Karen Hibbard, Amanda Cavallaro, Megan Hong, and Monti Mercer assisted with genetics and stock maintenance; and Rodney Simmons (Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA) provided media and general laboratory support. Eric Trautman (Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA) provided additional informatics tools. Phuong Chung and Julie Simpson taught us brain dissection and histochemical methods; Adina Bailey and Suzanna Lewis provided advice during the early stages of this work; and Susan Zusman and Michael Tworoger of Genetic Services, Inc., generated most of the transgenic lines. We thank Michael Layden for help in generating the data shown in Fig. 4*C* and Fig. S2. This work was supported by the Howard Hughes Medical Institute and by National Institutes of Health Grants GM076655–01A1 (to S.E.C.) and GM041249 (to J.T.K.).

1. Baker BS, Taylor BJ, Hall JC (2001) Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell* 105:13–24.
2. Hiromi Y, Kuroiwa A, Gehring WJ (1985) Control elements of the Drosophila segmentation gene fushi tarazu. *Cell* 43:603–613.
3. Harding K, Hoey T, Warrior R, Levine M (1989) Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *EMBO J* 8:1205–1212.
4. Goto T, Macdonald P, Maniatis T (1989) Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* 57:413–422.
5. Levine M, Tjian R (2003) Transcription regulation and animal diversity. *Nature* 424:147–151.
6. Brand AH, Dormand EL (1995) The GAL4 system as a tool for unraveling the mysteries of the Drosophila nervous system. *Curr Opin Neurobiol* 5:572–578.
7. Duffy JB (2002) GAL4 system in Drosophila: A fly geneticist's Swiss army knife. *Genesis* 34:1–15.
8. Ito K, Okada R, Tanaka NK, Awasaki T (2003) Cautionary observations on preparing and interpreting brain images using molecular biology-based staining techniques. *Microsc Res Tech* 62:170–186.
9. Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* 72:449–479.
10. Juven-Gershon T, Cheng S, Kadonaga JT (2006) Rational design of a super core promoter that enhances gene expression. *Nat Methods* 3:917–922.
11. Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* 166:1775–1782.
12. Lai SL, Lee T (2006) Genetic mosaic with dual binary transcriptional systems in Drosophila. *Nat Neurosci* 9:703–709.
13. Lee T, Luo L (1999) Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* 22:451–461.
14. Struhl G, Basler K (1993) Organizing activity of wingless protein in Drosophila. *Cell* 72:527–540.
15. Rolls MM, *et al*. (2007) Polarity and intracellular compartmentalization of Drosophila neurons. *Neural Develop* 2:7.
16. Wang J, *et al*. (2004) Transmembrane/juxtamembrane domain-dependent Dscam distribution and function during mushroom body neuronal morphogenesis. *Neuron* 43:663–672.
17. Reiff DF, *et al.* (2005) In vivo performance of genetically encoded indicators of neural activity in flies. *J Neurosci* 25:4766–4778.
18. Kitamoto T (2001) Conditional modification of behavior in Drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons. *J Neurobiol* 47:81–92.

Supporting Information

Pfeiffer *et al.* 10.1073/pnas.0803697105

SI Methods

Construction of the Vectors pBDP, pBPGw, and pBPGUw. Standard methods were used in a multistep cloning process to generate the vectors used in this study, which are part of a larger series of vectors (unpublished data). In brief, the *ori* and *amp* regions from pUC19 (Invitrogen) were combined with the mini-white gene from CaSpeR–AUG–βgal (1), the *PhiC31 attB* sequence from pUASTB (2), and a multiple cloning site to make pBDP. The *GAL4* coding sequence and *hsp70* 3′ UTR sequences from pGaTB (3) were cloned 5′-KpnI to 3′-NotI into pBDP, followed by blunt end ligation of the Gateway *RfA* cassette (Invitrogen) to produce pBPGw. Lastly, the synthetic core promoter, DSCP, was synthesized (DNA2.0) and cloned 5′-FseI to 3′-KpnI to make pBPGUw. Vectors, maps, and sequences are available from Addgene.

Design of Primers for PCR of Genomic DNA. A list of PCR primers was generated by evaluating every 25- to 28-bp segment, keeping only primers with predicted melting temperatures (T_m) between 67°C and 73°C ($T_m = 16.6 \times \log10(0.2) + 41 \times (G+C)/\text{length of primer} - 600/\text{length of primer}$), which were unlikely to form hairpins (evaluated by aligning the primer sequence against its reverse complement and eliminating primers with significant hits) and containing no repeated substring of length five nucleotides or greater. In addition, potential primers that overlapped annotated repeats, which ended with a 12-nt sequence that was present 14 or more times in the genome or a 15-nt sequence present two or more times in the genome, which had four or more G or C nucleotides in the first eight bases or began with CACC (which might interfere with subsequent cloning steps into the Gateway pENTR/D-TOPO vector; Invitrogen) were flagged as "nonconforming" and heavily penalized in subsequent steps. The remaining primers were labeled "conforming".

Primers were designed for all noncoding regions of the genome. Slightly different strategies were used for introns and intergenic regions, as detailed individually below.

Introns. Primer pairs were designed for all annotated introns. Introns containing annotated transposable elements (TEs) were divided, with the different non-TE-containing regions treated as separate introns for the purposes of primer design. Primer pairs were not designed for introns shorter than 300 bp. For introns of longer than 300 bp, but shorter than 4,000 bp, a single primer pair was designed consisting of the closest conforming primers outside of the annotated 5′ and 3′ ends of the intron, unless one end corresponded to a TE; in which case, the first conforming primer inside the relevant end was used to avoid selecting primers within TEs.

Introns longer than 4,000 bp were tiled by selecting multiple primer pairs that would amplify regions covering the intron with overlaps between the different fragments. The optimal tile was considered to have fragments of 3,000 bp overlapping by 1,000 bp. The number of fragments (NF) for each tiled region was set to the closest integer to $1 + (\text{region size} - 3{,}000)/(3{,}000 - 1{,}000)$. Therefore, for example, a 10,000-bp region would be tiled with five fragments. A possible set of primer pairs would produce NF amplified regions of minimum size 1,000 bp and maximum size 5,000 bp, with minimum overlap of 500 bp and maximum overlap of 1,500 bp. The first primer of the first primer pair and last primer of the last primer pair were fixed as the first conforming primers outside (or, in the case of ends determined by TEs inside) of the 5′ and 3′ end of the targeted genomic region. Each possible set of primer pairs was scored by using a dynamic programming algorithm and the optimal (lowest scoring) set selected.

The score for each set of primer pairs was the sum of scores assigned to each primer plus a score based on the sizes of the fragments plus a score based on the extent of fragment overlap. Note that both conforming and nonconforming primers were considered. Each primer was given a score equal to the absolute value of the difference between its predicted T_m and the optimal T_m (70°C). Nonconforming primers were further penalized by 10,000 if they overlapped an annotated repetitive region of the genome, 20,000 if they overlapped a conserved fragment, 40,000 if they contained more than 4 G or C nucleotides in their first eight, 50,000 if they began with CACC, and 50,000 if they began with a 12-nt sequence present 14 or more times in the genome or a 15-nt sequence present two or more times in the genome. Conserved fragments were defined by the "Most Conserved" track downloaded from the University of California, Santa Cruz genome test browser on February 25, 2005. The track was based on phastCons analysis of multiZ alignments of *D. melanogaster* (University of California, Santa Cruz genome droMel2), *D. simulans* (droSim1), *D. yakuba* (droYak1), *D. ananassae* (droAna1), *D. pseudoobscura* (dp3), *D. virilis* (droVir1), *D. mojavensis* (droMoj1), *A. gambiae* (anoGam1), and *A. mellifera* (apiMel1). These penalties were sufficiently large to ensure that nonconforming primers would only be used if no acceptable path containing conforming primers was available. (For example, if there were a conserved fragment in the targeted region larger than the maximum fragment size.) Each fragment received a score equal to the difference between its size and the optimal fragment size (3,000 bp) multiplied by a scaling factor of 0.005. Each overlap between adjacent fragments received a score equal to the difference between the overlap size and the optimal overlap size (1,000 bp) multiplied by a scaling factor of 0.005. The algorithm thus placed a premium on finding paths consisting of pairs of conforming primers, thereby avoiding repetitive and highly conserved regions.

Intergenic regions. For intergenic regions shorter than 4,000 bp, only promoter fusion constructs were designed (see below). For intergenic regions longer than 4,000 bp, a set of overlapping fragments was designed to cover the entire intergenic region, with a 300-bp buffer subtracted from either end if the end is a transcription start site. (This was done to avoid including promoters in constructs designed to detect enhancers.) The regions were then tiled as described for introns above, except that the first primer of the first fragment and last primer of the last fragment were fixed as the first conforming primers inside of the 5′ and 3′ ends of the region, respectively.

Promoter fusions. Promoter fusion primer pairs were designed as follows. The optimal promoter fusion fragment was defined as the region spanning 3,000 bp upstream of the transcription start site (or the end of the closest gene upstream if the gene is within 4,000 bp of the targeted transcription start site) to a point 200 bp into the 5′ UTR of the gene (unless the translation start site was within 300 bp of the transcription start site; in which case, the optimal UTR primer was halfway between the translation and transcription start sites).

The first (upstream) primer was selected as the first conforming primer inside the upstream end of the optimal promoter fusion fragment. The second (UTR) primer was selected as the closest conforming primer to the optimal UTR primer position.

The primers used to generate the fragments used to study the

dac, earmuff, *toy*, and octopamine receptor 2 genes are given in Tables S1, S2, S3, and S4, respectively.

PCR of Fragments from Genomic DNA and Cloning into the Gateway Vector. *PfuUltra* High-Fidelity DNA Polymerase (Stratagene) was used to amplify selected fragments (see above) by using DNA from *y*; *cn bw sp* (4) as a template. The PCR products were confirmed by agarose gel analysis, purified by using the QIAquick PCR Purification Kit (Qiagen), and 9.5 μl was then used in a 12 μl TOPO cloning reaction with pENTR/D-TOPO (Invitrogen) for the fragments derived from the *dac*, *toy*, and earmuff genes. Subsequent PCR fragment cloning was performed by adding 3′ A overhangs to the PCR products with the addition of dATP and *Taq* polymerase in a 10-min incubation at 72°C before Qiagen purification. The products were then used in a TOPO cloning reaction with pCR8/GW/TOPO as described by the manufacturer (Invitrogen). Cloning reactions were allowed to proceed for 30 min at room temperature, and then 2 μl of each reaction was used to transform Mach1 cells (Invitrogen). For each cloning reaction, two isolates were picked, purified, and confirmed by sequence verification.

Sequence Verification of Clones. Two Gateway clones were picked for each enhancer fragment and up to four clones for the promoter fusions, for a total of 11,384 processed clones. A reverse-sequencing primer (Invitrogen; 5′ CAGGAAACAGCTATGACC 3′) was used to generate a single sequence read for comparison with the targeted genomic region by using Sim4 (5). One clone was identified and selected for future studies; enhancer clones in their native genomic orientation were selected when verified clones in both orientations were available. For promoter fusions, only clones in the proper orientation were selected. A rework loop was initiated for the 8% of targets that initially failed at any step of the process. An additional 2,630 clones were selected for rework sequencing. One sequence-verified clone for each target was rearrayed to generate the final Gateway clone set (5,309 clones).

Transfer of Gateway Clones into Integration Vectors. Fifty nanograms of the destination vector, either pBPGw or pBPGUw, were combined with 50 ng of DNA carrying a PCR fragment cloned in the Gateway vector in a one-fifth LR reaction (Invitrogen) and incubated overnight at room temperature. MultiShot Top10 cells (Invitrogen) were transformed with 2 μl of the LR reaction and plated. A single isolate from each reaction was picked into a 96-well Beckman Deepwell block, allowed to grow overnight at 37°C and DNA was prepared by using the QIAprep 96 Turbo Miniprep Kit (Qiagen). The constructs were verified by analysis of restriction enzyme digests. A second isolate was picked in cases where there was a discrepancy between the observed and expected results. DNA for injection was prepared from 5 ml of overnight culture and sent to Genetic Services, Inc. for production of transgenic flies in the *attP2* landing site (2).

***Drosophila* Genetics.** Transformant stocks of each construct were generated as diagramed in Fig. S1. DNA constructs in the pBPGUw or pBPGw vectors (100–200 ng/μl) were microinjected into embryos derived from parents homozygous for both the *attP2* integration site (2) and a fusion gene encoding the *PhiC31* integrase under the control of the *nanos* promoter (*nos–integ*), which provides a maternal source of integrase (6). Single males derived from these embryos were crossed to *y w* females, and males carrying the inserted construct (identified by their w^+ eye color) were selected; note that the source of integrase is also removed in this step. These males were crossed to *y w*; *Dr*, *e*/*TM3*, *Sb* females to establish balanced, homozygous stocks.

Verification of Insertion Site and Fragment Identity by Genomic PCR. To verify the identity of transformant flies and to confirm that all integration events occurred at the *attP2* site, we performed genomic PCR on DNA isolated from homozygous transformant flies. Twenty flies were homogenized and genomic DNA isolated by using the ZR Genomic DNA II Kit (Zymo Research). To assay proper integration in the *attP2* landing site, PCR was performed by using a primer from the *y* gene marker in the *attP2* genomic docking site (TCATGACTTTGTTGCCTTAGA) and a reverse primer from the *w* gene (CGAAAGAGACGGCGATATT) carried in the constructs. Only proper integration events would yield a product of 1839 bp, because *y* and *w* lie more than 2 Mb away in the *Drosophila* genome. Fragment identity was confirmed by using a vector-specific primer (ACAAGTTTGTACAAAAAAGCAGGCT) and a reverse primer specific to the cloned fragment being tested for enhancer activity; the position of the fragment-specific primer was chosen so as to yield a PCR product of 350 to 400 bp.

Embryo Whole Mount *in Situ* Hybridizations. Embryos were collected directly from the homozygous stock. Embryonic whole mount *in situ* RNA hybridizations were performed as described previously (7). RNA probes were generated by amplifying *GAL4* from pBPGUw by using the forward primer TGCGATATTTGCCGACTTA and a reverse primer that contained the promoter sequence for T7 RNA polymerase (TGTAATACGACTCACTATAGGGAACATCCCTGTAGTGATTCCA).

Embryo Whole Mount Histochemistry. Each GAL4-expressing line was crossed to *UAS-GFP:NLS* (which produces nuclear localized GFP), and stage 16 embryos were fixed and stained by using standard methods (8). Briefly, primary antibodies were guinea pig anti-TOY (1:500; gift of Uwe Walldorf, University of Saarland, Homburg, Germany), rabbit anti-GFP (1:1000; Invitrogen), mouse anti-EVE (1:10; see ref. 8). Detection was with species-specific secondary antibodies conjugated to Alexa Fluor 488, 568, or 635 (Molecular Probes–Invitrogen). Embryos were oriented ventral up, and a *z* series was taken at 1-μm intervals throughout an abdominal CNS hemisegment by using a Bio-Rad Radiance confocal. Five-micrometer maximum intensity projections were made at three levels within the CNS to show most of the TOY-positive neurons: one at a "deep" level (contains the TDL neurons and the EVE^+ RP2 motor neuron), one at a "middle" level (contains the TI neurons and the EVE-expressing U motor neurons), and one at a "superficial" level (contains the TS cluster and the EVE-expressing EL interneurons). There is no overlap between endogenous TOY and EVE.

Brain and Ventral Nerve Cord Dissection and Histochemistry. To generate flies for imaging expression in larvae and adults, homozygous males were crossed to homozygous mCD8–GFP females (9). Brains and ventral nerve cords of third instar larvae and adult female flies were dissected in cold PBS and fixed in 0.25% (larval) or 0.8% (adult) paraformaldehyde (Electron Microscopy Science) in PBS at 4°C overnight. After four 30-min washes in 0.5% BSA and 0.1% Triton X-100 in PBS (larval PAT) or in 0.5% BSA and 0.3% Triton X-100 in PBS (adult PAT), samples were permeabilized with 3% normal goat serum (NGS) in larval PAT or adult PAT for 2 h at room temperature. Samples were then incubated in a primary antibody solution containing mouse anti-nc82 (1:50; Developmental Studies Hybridoma Bank) and rabbit anti-GFP IgG (1:1000; no. A11122, Invitrogen) in 3% NGS in larval or adult PAT at 4°C overnight. After four 30-min washes in the appropriate wash buffer, they were incubated at 4°C overnight in a secondary antibody solution containing Alexa Fluor 568 goat anti-mouse IgG (1:1000, Invitrogen) and Alexa Fluor 488 goat anti-rabbit IgG (1:1000, Invitrogen) in the corresponding blocking buffer. After at least

four 30-min washes, samples were rinsed in PBS before mounting in Vectashield mounting medium (Vector Labs) within two conjoining reinforcement ring labels (Avery Consumer Products) or silicone adhesive spacers (no. S24737; Invitrogen). To prevent dehydration, samples were covered with a no. 1.5 glass coverslip before imaging.

Immunolabeled larval and adult brains and ventral nerve cords were imaged with a 510 laser-scanning confocal microscope (Zeiss) under ×20 magnification. Eight-bit z stack images were scanned at 1-μm section intervals with a resolution of 1024 × 1024 pixels.

Annotation of Expression Patterns in the Brain. ***Cell counts.*** We developed an ImageJ macro that allows for machine-assisted counting of a large number of cells. The user marks individual cells with a mouse click in a number of optical sections while the macro counts the cells in the entire brain, as well as in a selected area. To avoid double counting of cells, the macro marks every counted cell three-dimensionally in the data set. The results of the count are stored for each specimen separately as a file containing tab-delimited values.

Controlled vocabulary for annotation of expression patterns. For each of 45 distinct brain structures, the GAL4-driven GFP expression patterns (mCD8–GFP) were described by three parameters: intensity, distribution, and shape. The "intensity" parameter describes the brightness of the expression pattern without regard to the form of the arborization pattern in the neuropil area; it reflects the GFP-expression level as far as it is possible in this qualitative approach. We developed the "distribution" measure of arborization patterns because the expression patterns of most of the evaluated lines arborize only in subregions of the defined neuropil areas; this parameter describes how much of a given neuropil area is interlaced with arborizations. The "shape" parameter describes the form of the arborization pattern in a given neuropil.

These parameters are coded numerically but also have (verbal) descriptive correlates, given below, to facilitate the annotation work and discussion/communication of the evaluated patterns.

Intensity.

0 blank: Even after contrast enhancement, no signal is detectable.
1 faint: The signal is nearly invisible even without the reference staining visualized. The signal can just be detected after enhancing contrast and brightness of the visualization.
2 very weak: In visualization simultaneous with the reference staining, the signal is just recognizable if its position is already known. Without visualizing the reference staining the signal is clearly detectable. Together, the scales "faint" and "very weak" cover the lowest quarter of signal intensity.
3 weak: Signal is dim but still clearly detectable when displayed together with the reference staining pattern. This scale covers the third quartile of signal intensities.
4 strong: Signals are easily detectable in the presence of the reference staining. This scale covers the second quartile of signal intensities.
5 very strong: Very-strong-expressing neurons or strong-expressing neurons with large diameters are detectable. Because of microscope settings, which are set to allow the detection of weak signals, these very strong signals tend to saturate 8-bit signal coding. This scale covers the top quartile of signal intensity.

Distribution.

0 passing tract: The expression pattern is confined to tracts passing the neuropil area. Arborizations of these tracts are not detectable within the given neuropil area.
1 focused: The arborization pattern is aggregated in a very confined space, such as a single glomerulus in the antennal lobe.
2 local, regional: The arborization pattern is confined to one or several distinct, but large, subregions of the given neuropil.
3 widespread: The arborization pattern fills a significant fraction, but less than half of the given neuropil region.
4 prevalent: The arborization pattern fills more than half of the given neuropil region.
5 ubiquitous: The arborization pattern fills the complete neuropil region.

Shape.

0 empty: No fluorescent structures above background noise level are detectable.
1 sparse: Very few fluorescent fibers interlace the area of the arborization pattern. These can be traced individually.
2 diffuse: The arborization pattern is loose enough to allow the recognition of single fibers but is too dense to enable tracing of these fibers.
3 dense: The area of the arborization pattern is interlaced with a close network of fluorescent fibers that are optically not separable from each other.
4 tight: The area of the arborization pattern is densely filled with fluorescent signal, which merges in most of the region to a common signal but also shows small gaps in between these regions.
5 jammed: The area of the arborization pattern is completely filled with fluorescent signal; fibers and/or substructures blur out because of their close proximity, which gives the impression of a nearly homogenous pattern.

1. Thummel CS, Boulet AM, Lipshitz HD (1988) Vectors for *Drosophila* P-element-mediated transformation and tissue culture transfection. *Gene* 74:445–456.
2. Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* 166:1775–1782.
3. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118:401–415.
4. Adams MD *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195.
5. Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* 8:967–974.
6. Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci USA* 104:3312–3317.
7. Tomancak P, *et al.* (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol* 3:research0088.1-0088.14.
8. Doe CQ (1992) Molecular markers for identified neuroblasts and ganglion mother cells in the *Drosophila* central nervous system. *Development* 116:855–863.
9. Lee T, Luo L (1999) Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* 22:451–461.

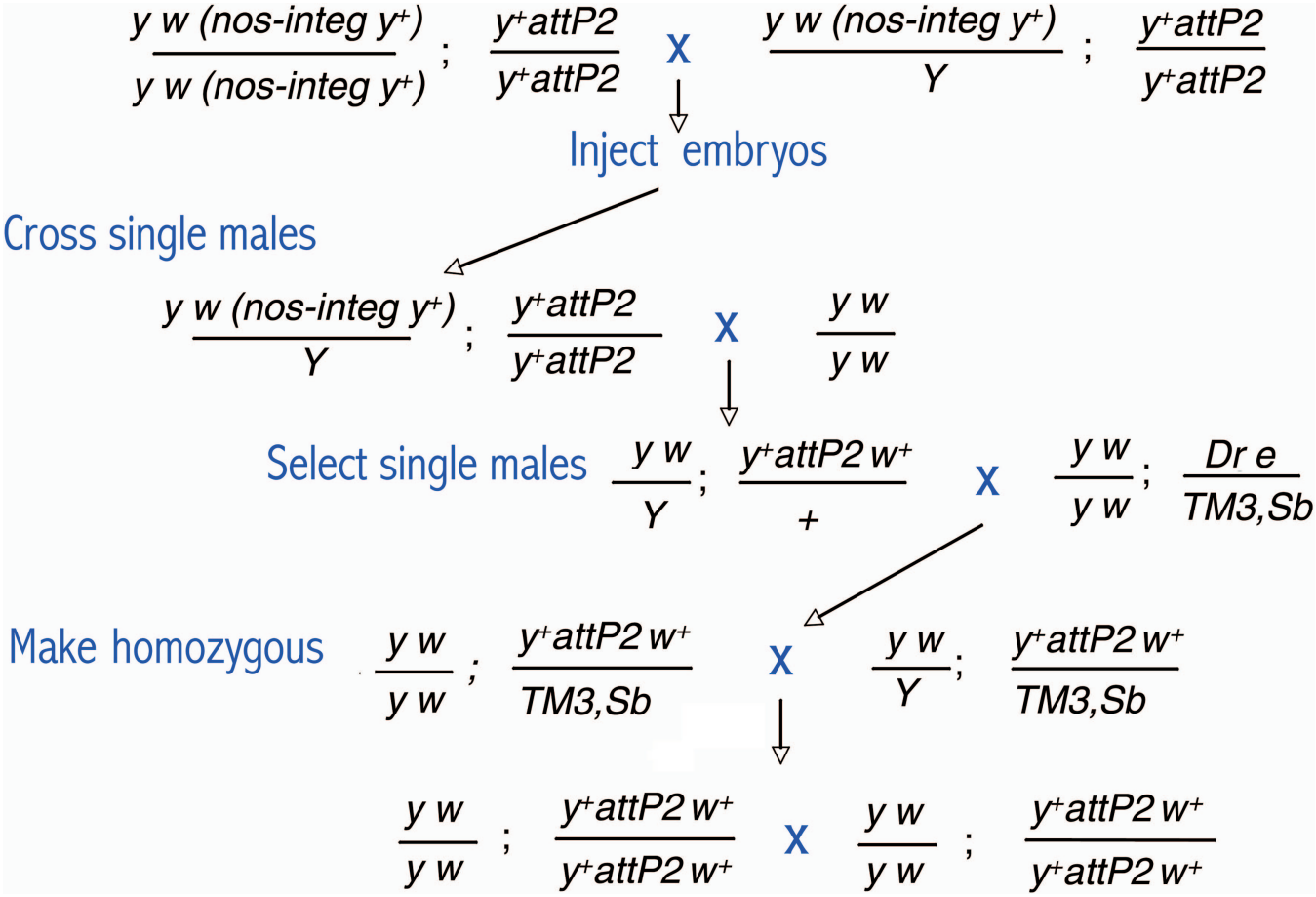

Fig. S1. Genetic crosses used to generate transformant lines at the *attP2* genomic site. See *SI Methods* for details.

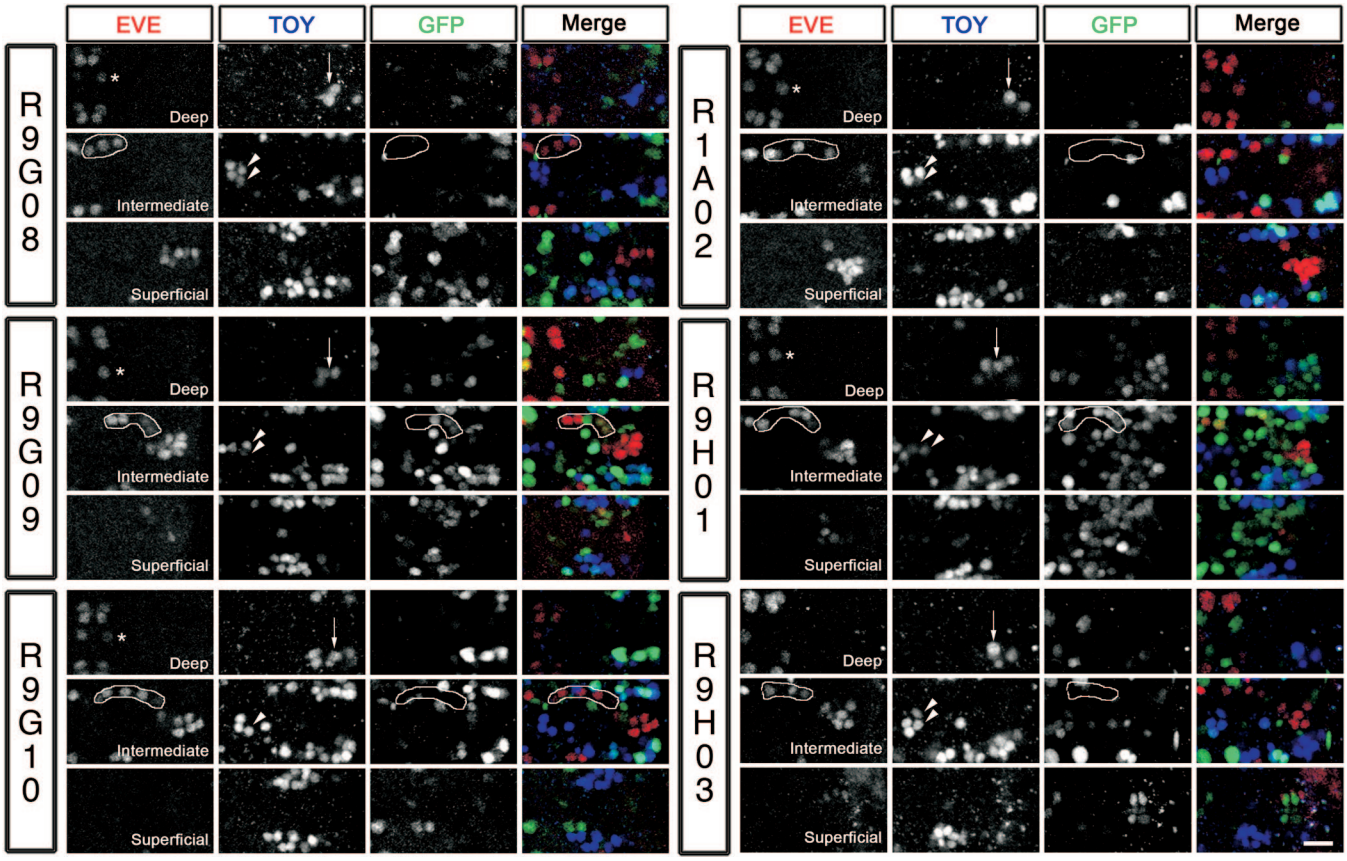

Fig. S2. Patterns generated by fragments of the *toy* gene and a comparison to the expression pattern of the endogenous *toy* gene in the embryonic stage 16 CNS. Endogenous *toy* protein (blue in the merged image), Even-skipped (EVE) (red in the merged image), and nuclear-localized GFP (green in the merged image) expression driven by each of the indicated *toy* fragments. A single abdominal CNS hemisegment is shown, with a small portion of the contralateral hemisegment. Three focal planes are shown: deep (top), intermediate (middle), and superficial (bottom). Arrows, three *toy* deep lateral (TDL) neurons; arrowheads, two *toy* medial (TM) neurons; asterisks, EVE-expressing RP2 motor neuron; outlines, one or more of the EVE-expressing U1–U5 motor neurons. Anterior, up. (Scale bar, 10 μm.)

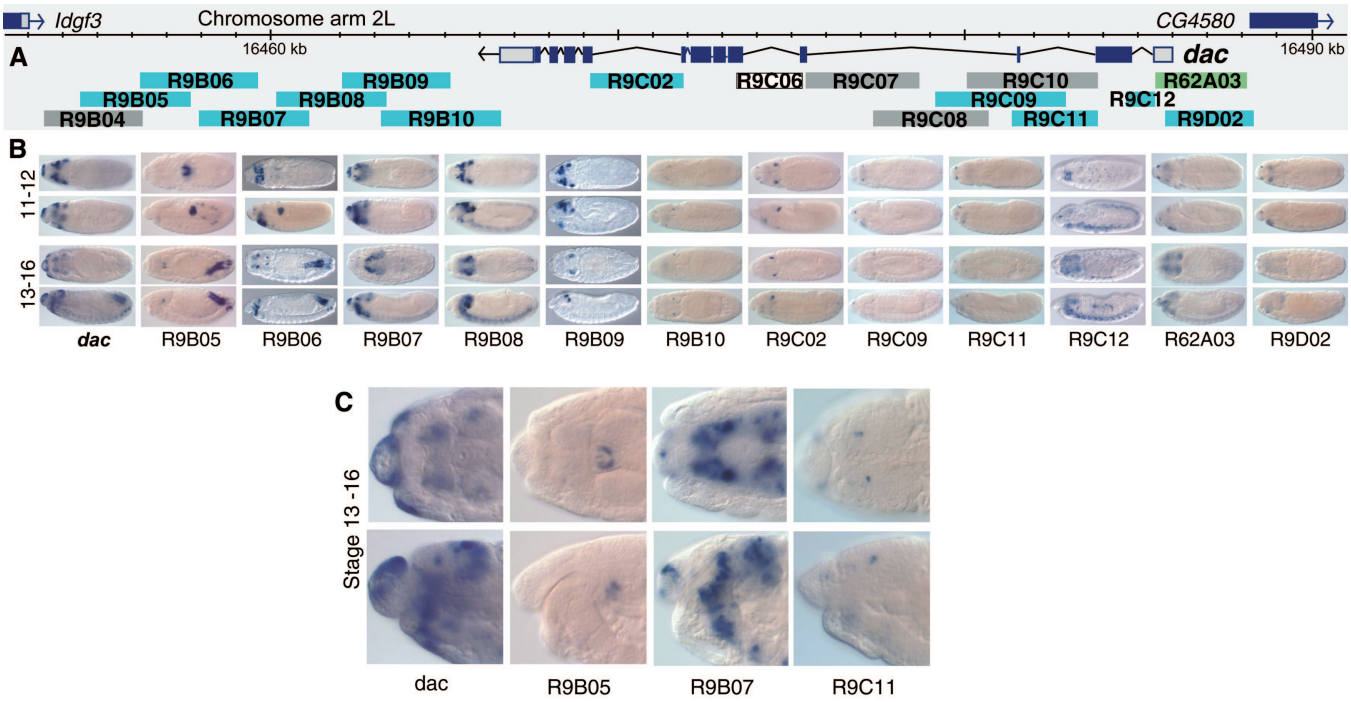

Fig. S3. Patterns generated by fragments of dachshund (*dac*) when paired with the DSCP in embryos. (*A*) The genomic map of the *dac* locus. DNA fragments in gray gave no expression in embryos. Fragment R9C06 was not assayed. (*B*) Expression of the *dac* gene in embryonic stages 11–12 and 13–16; dorsal views are shown above lateral views for each stage, visualized by whole mount *in situ* hybridization by using a probe to *dac* mRNA. Shown adjacent to the wild-type expression patterns are those produced by the fragments R9B05, R9B06, R9B07, R9B08, R9B09, R9B10, R9C02, R9C09, R9C11, R9C12, and R9D02 in the enhancer test vector and fragment R62A03 as a promoter fusion. Transgene expression is visualized by whole mount *in situ* hybridization with a probe to *GAL4* mRNA. (*C*) For the three DNA fragments R9B05, R9B07, and R9C11, we show enlargements (6-fold) of the images shown in *B* focused on the embryonic brain, where expression in distinct subsets of cells can be seen.

Table S1. Specifications of *dac* fragments and endogenous promoter sequences

Fragment	Forward primer	Reverse primer	Length, bp
Endogenous promoter	GAGGCAGCGACGCCACCGTCGCCGT	ATCATGGCCGACACTTTTGTTTATT	973
R62A03 promoter fusion	CGAGCGCACTGAACTTTTGCCATCC	CCGATTGCCCGCTTGTGTTAATCGA	2,639
R9B04	GAATTCACGATCTGGGCG	TCACCCTTTCGGTGCCTA	3,286
R9B05	CAATTCTCCCCAAAAGCG	CGATGGTGACAGGGAAGG	3,211
R9B06	CGGACTGCCAAATGGCATTATTGTG	GCCCAGATCGTGAATTCGTCTGACC	3,360
R9B07	CCAAGATTTGCATCCGCT	CCCCATCATTTCCCATCA	3,204
R9B08	GGCTCTGCCAGAAGTGCT	ACCAACGGGCCTTACCTC	3,221
R9B09	TTGGCTGGGCGTCTTTAC	CCCTCGTTTCTTGCCTGA	3,148
R9B10	ATTAAAATAATACAGCAGCCGACAA	TCGGATATGTTAAAATGGTTCTGAT	3,450
R9C02	GAACGCCAGCTGGTTGATGAACGTA	CTGCTGATAGCGGATGCGGATCTTT	2,690
R9C06	TGGACAACGGCGATTATGCCTATGA	GCCTTGAGCATCTGACTTTCGGGAC	1,884
R9C07	TCTTTGCCAATGATGCCA	TCCAGGCGATGCTTCTTC	3,614
R9C08	TTGATTTTGATCACAGTTGACAGTT	GCAAAATCGAATAAGCTCAAAAATA	3,258
R9C09	GCCGATTTGTTCAAATGAAGGCGAC	TTCAGCCAGCTACTTGCGATTTCCC	3,800
R9C10	CATCCTCTACAGGGACTGCACCACG	TAAATGGCTTTTTGGGGTTTTCCCG	3,796
R9C11	CATCCTCTACAGGGACTGCACCACG	TCGCCCAATTTCCGTTTCGAGTTTA	2,507
R9C12	AGTGCGCGAGTGTCTTGTGTCAGTG	AACATGGCGGTGGTGTTCATATTGC	876
R9D02	CTCACAGGGATTGGCTTCCTTGTCC	CTGGTTCCTCTTCCCCCTCTCGTTC	2,573

The DNA sequences of primers used to generate promoter or enhancer fragments by PCR are shown, along with the expected length in base pairs of the resultant fragments.

Table S2. Specifications of earmuff (CG31670) fragments and endogenous promoter sequences

Fragment	Forward primer	Reverse primer	Length, bp
Endogenous promoter	TGTGTCCTTTTCATATTTCTTAAGG	GGCACTGCCTATGTGGATATCCAGT	272
R9D03	GCGATGCATGTGATGTCC	CCACGTCGAAATTGGGAA	3,888
R9D04	TGGGGTCTGTCAACCGAT	TTCCCGAAAAACAAGGCA	3,808
R9D05	TGTGTGCCAACTGTTGTGTTTTGCC	ATCGAAAACTTTAGGGGACACGCCC	1,702
R9D06 promoter fusion	TGCATACCATTGCAGCGT	AATGCAAATCGAGGGACG	3,465
R9D08	CGAGGAAGCATCGAAGGA	TCAGCGGGAAAGAGATGC	3,166
R9D09	ATTGGCTTTCTTTCGGGG	AGCCATGGGAATGGGAAT	3,114
R9D10	CAGGCCAGTAGCGGAAAA	ACACACCCCCAAAAGCAA	3,664
R9D11	AGAAGTCCAACGCGCTATCCAATGC	TGGTTAGCAGCTGCGGTTTGCTATG	3,955

The DNA sequences of primers used to generate promoter or enhancer fragments by PCR are shown, along with the expected length in base pairs of the resultant fragments.

Table S3. Specifications of *toy* fragments and endogenous promoter sequences

Fragment	Forward primer	Reverse primer	Length, bp
Endogenous promoter	GCCGATGTGTTAGCCGTTGCTCTTG	ATTATTTTTAATGCGGCGTTGGAGA	568
R9G07	ACAAGACGTGCCAGAGGC	CAAAAACTTCAAGCGGCTG	3,807
R9G08	TGCAAAGTGCGTAATGGC	CTTCAGCTAGAGGTGCGCTT	3,560
R9G09	TCGATGGCTAAAGCTAAGAGGCAAA	CCGGAGATTGCTAATATTCGAGCCG	3,394
R9G10	TTATGAAATTTCGCGGTCG	AATAGAGGGCTTGGCGCT	3,244
R1A03	GGAATTTCGTTTATTCAATG	TGAAGTCACACGCATCAACGA	2,021
R1A02	GACAGGTAAAAATACCGAG	CGATAGCTAACAACGTTGCTG	1,930
R1A01	TCGTTGATGCGTGTGACTTC	CGATAGCTAACAACGTTGCTG	1,001
R9H01	AGTGTCCAACGGTTGCGT	CTGGGACATTCCCGTTTG	1,710
R9H02	CGGGAATGTCCCAGCATA	TGTTGCTGAGCTTGCTGC	4,414
R9H03	GCAGCAAGCTCAGCAACA	GAGTCTTCGTCCCCGGAT	1,242
R9H04	ATCCGGGGACGAAGACTC	GTATGTGCCGCCCAAAAC	1,074
R9H06	GCGCGAGAAAGGCTTGCTGATAAAA	ACCGCCCATTCCGCTGCTTATACTT	3,495
R9H07	TGTATTTCGGAACGACCGACCCATT	GGATTATTTGCCGTGCTGGTTCGAC	2,719
R9H08	CCGCGGCTACCCTTAAAT	GGGGTTGCAAGCTGTGTT	1,401
R9H09	ATCTTAGGTGTCATTTCTGCGGGCG	TTTTAAAGCGCAAAACCGATGCTCC	3,322
R9H10	ATGAGCAAAGCGAGGCAC	TAACTCCCACAAACCGCC	2,853
R9H11	CTTTAGCTCGCAACTCGCTTCGCTC	CAATTGCATTCGTGGCTTTGACGAC	4,316
R9H12	AAGGCTCGCGGAAAAACT	CGTCACAAAATGTGAAAGCG	2,843

The DNA sequences of primers used to generate promoter or enhancer fragments by PCR are shown, along with the expected length in base pairs of the resultant fragments.

Table S4. Specifications of octopamine receptor 2 fragments

Fragment	Forward primer	Reverse primer	Length, bp
R19H07	ACTTCAACTCGGCCCTGAACCCCAT	CACGTCGACAGAAGTAGAAAGCGTA	1,565
R20C11	GTCGGTCATGCGGCACCGGAAATTG	CTTGCCCGAGATCATGACGGAAGCA	2,663
R20E11	AGCCCGGCTATCGGGGTCAACTAAA	GGAACCCACCACTCGCATGACCTTA	3,115
R21E03	GTGGGGCTCTACTGCTGACATACAT	CCACGGAATAAGGACGACATGCTGT	4,149

The DNA sequences of primers used to generate promoter or enhancer fragments by PCR are shown, along with the expected length in base pairs of the resultant fragments.

Other Supporting Information File

Dataset S1 (XLS)