# Versatile P(acman) BAC Libraries for Transgenesis Studies in *Drosophila melanogaster*

Koen J. T. Venken[1], Joseph W. Carlson[2,8], Karen L. Schulze[3,8], Hongling Pan[3], Yuchun He[3], Rebecca Spokony[4], Kenneth H. Wan[2], Maxim Koriabine[5], Pieter J. de Jong[5], Kevin P. White[4], Hugo J. Bellen[1, 3, 6, 7, *] and Roger A. Hoskins[2]

[1] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[2] Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 97420, USA

[3] Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA

[4] Institute for Genomics and Systems Biology and Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

[5] Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA

[6] Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA

[7] Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA

[8] These authors contributed equally to this work.

[*] Correspondence (H.J.B.): E-mail, hbellen@bcm.tmc.edu; Tel, 713-798-5272; Fax, 713-798-3694.

**ABSTRACT**

We constructed *Drosophila melanogaster* BAC libraries with 21-kb and 83-kb inserts in the P(acman) system. Clones representing 12-fold coverage and encompassing more than 95% of annotated genes were mapped onto the reference genome. These clones can be integrated into predetermined *attP* sites in the genome using ΦC31 integrase to rescue mutations. They can be modified through recombineering, for example to incorporate protein tags and assess expression patterns.

**MAIN TEXT**

Genetic model systems such as *Drosophila melanogaster* are powerful tools for investigating developmental and cell biological processes, properties of inheritance, the molecular underpinnings of behavior, and the molecular bases of disease [1]. The approaches used in model systems rely on the identification of mutations in genes and the characterization of the gene products, often aided by transgenesis techniques [2].

We recently developed a new transgenesis platform for *D. melanogaster*, the P(acman) (P/ΦC31 artificial chromosome for manipulation) system, that allows modification of cloned fragments by recombineering and germ-line transformation of genomic DNA fragments up to 133 kb in length [3]. P(acman) combines a conditionally amplifiable BAC [4], the ability to use recombineering in *E. coli* for retrieval and manipulation of DNA inserts [5], and bacteriophage ΦC31 integrase-mediated germ-line transformation into the *D. melanogaster* genome [6,7]. Clones are maintained at low-copy number to improve plasmid stability and facilitate recombineering, but can be induced to high-copy number for plasmid isolation to facilitate microinjection of embryos. Recombineering can be used to insert protein tags for *in vivo* protein localization or acute protein inactivation [8], and to create deletions [9] and point mutations [5] for structure/function analysis. ΦC31-mediated transgenesis integrates DNA constructs at specific pre-determined *attP* sites dispersed throughout the genome [3,6,7,10], eliminating the need to map integration events and reducing variability in expression due to position effects [10]. The technique allows rescue of mutations in large genes [3] and facilitates comparative expression analysis of engineered DNA constructs [7,10-12]. Previously, genomic regions of interest were cloned into P(acman) by gap-repair from available mapped BAC clones [3]. Here, we describe a more efficient approach: we

constructed two genomic BAC libraries in the P(acman) system and mapped the cloned inserts by alignment of paired end sequences to the reference genome sequence.

We engineered a novel P(acman) BAC vector for construction of genomic libraries **(Fig. 1a**). In addition to the published features [3], we included a polylinker embedded within a mutant *α-lacZ* fragment. It became apparent that in the low-copy-number condition necessary to ensure stability of large genomic inserts, standard *α-lacZ* fragments are expressed at insufficient levels to permit reliable blue-white colony screening. We isolated a mutant with significantly enhanced β-galactosidase activity resulting from a premature stop codon in the *α-lacZ* fragment (**Supplementary Fig. 1**) that permits blue-white selection for cloned inserts at low-copy number using an automated colony picking device.

To create a resource for manipulation and analysis of *D. melanogaster* genes, we constructed two P(acman) libraries (**Supplementary Fig. 2**). For analysis of most genes, we constructed a library with an insert size of 20 kb. Ninety percent of protein-coding gene annotations in *D. melanogaster* are less than 12.1 kb in length, and a 20 kb insert size should provide sufficient flanking genomic sequence to contain most genes, including regulatory sequences required for normal expression. For analysis of large genes and gene complexes, we constructed a library with an insert size of 80 kb. High molecular weight genomic DNA was prepared from the *D. melanogaster* strain used to produce the reference genome sequence. The DNA was fragmented by partial restriction digestion, and size fractions in the 20 kb and 80 kb ranges were recovered and cloned separately to produce two genomic BAC libraries. The libraries produced from the 20 kb and 80 kb fractions were designated CHORI-322 and CHORI-321, respectively. We stocked 73,728 CHORI-322 clones and 36,864 CHORI-321 clones.

## P(acman) BAC Libraries

To map P(acman) BACs on the genome, paired end sequences were determined and aligned to the reference genome sequence. We mapped consistent paired ends of 33,314 CHORI-322 clones representing 4.3-fold coverage of the X chromosome and 5.9-fold coverage of the autosomes, and 12,328 CHORI-321 clones representing 8.2-fold coverage of the X chromosome and 9.3-fold coverage of the autosomes. The mapped paired end sequences show that the average insert sizes of the CHORI-322 and CHORI-321 libraries are 21.0 kb (+/- 4.0 kb) and 83.3 kb (+/- 21.5 kb), respectively. An additional 18,767 CHORI-322 clones and 11,571 CHORI-321 clones were partially mapped to the genome sequence by alignment of one end sequence only. The two libraries together represent deep coverage of the genome and span most annotated genes (**Supplementary Table 1**). The mapped CHORI-322 and CHORI-321 clones span 88.9% and 99.3% of annotated genes, respectively. P(acman) clones containing genes and genomic regions of interest can be identified through a web-accessible genome browser (http://pacmanfly.org/) (**Fig. 1b**) and are available for distribution from the BACPAC Resources Center (http://bacpac.chori.org/).

We tested the P(acman) library resource for transformation efficiency using clones encompassing several genes. For each gene, we identified a clone containing substantial flanking sequences biased toward the 5' end of the gene annotation. These clones are likely to include the regulatory sequences necessary for normal expression of the gene. For small genes (≤12 kb), a CHORI-322 clone was preferred over a CHORI-321 clone, as smaller clones tend to have higher transformation efficiencies [3]. When a mapped CHORI-322 clone was not available for a small gene (e.g. *hh*, *vas* and *shi*) or sufficient 5' regulatory sequence did not appear to be present in a mapped CHORI-322 clone (e.g. *jar*, *lt* and *cta*), we chose a CHORI-321 clone instead. In total, we selected 38 clones from the CHORI-322 library (**Table 1**) and 24 clones from the CHORI-

321 library (**Table 2**). The largest clone, encompassing *Hnf4*, has an insert size of 105 kb. Each clone was isolated and tested for integration into a genomic *attP* docking site, either VK37 on chromosome arm 2L or VK33 on chromosome arm 3L [3], using ΦC31 integrase [6,7]. The transformation efficiency of each clone was defined as the percentage of G0 fertile crosses that yielded at least one transgenic animal. We were successful in obtaining at least one transformant for all CHORI-322 clones (**Table 1**) and 13 of the 24 CHORI-321 clones (**Table 2**). In addition, 16 of 17 CHORI-322 clones used for recombineering-mediated tagging (see below) were successfully integrated (**Supplementary Table 2**). Moreover, 53 of 72 CHORI-321 clones have been integrated successfully in an independent experiment to generate a set of duplication lines, each carrying a clone from a tiling path of overlapping CHORI-321 clones spanning the entire X chromosome (Ellen Popodi and Thom Kaufman, personal communication). These data show that more than 98% (54/55) of CHORI-322 clones and at least 68% (66/96) of CHORI-321 clones can be successfully integrated. For all transformants, the presence of the expected DNA fragment sizes at the integration junctions - indicative of site-specific integration at the respective docking site - was confirmed by multiplex PCR that tests simultaneously for the presence of *attP*, *attB*, *attR* and *attL* sites (**Supplementary Fig. 3**).

The range of integration efficiencies observed is surprisingly broad. Efficiencies ranged from 0% to 28.1 % for CHORI-322 clones and from 0% to 11.6 % for CHORI-321 clones. The insert sizes of CHORI-322 clones are very similar to each other, so the observed range suggests that some fragments are less efficiently transformed than others due to sequence content or specific interference between certain fragments and docking sites (e.g. *Csp* and *wg*). Notably, the high efficiency observed for some CHORI-321 clones (e.g. CH321-16H04, CH321-64G01 and

CH321-79N05) suggests that further optimization of the integration efficiency of large clones is possible.

We tested transgenic insertions of ten CHORI-322 and six CHORI-321 clones for their ability to complement lethal mutations in genes. All CHORI-322 clones tested, encompassing the genes *CG6017*, *chc*, *dap160*, *drp1*, *endo*, *Eps15*, *n-syb*, *sqh*, *synj* and *vha100-1*, rescue lethal mutations in the corresponding genes. To our knowledge, rescue of mutations in *endo*, *n-syb* and *vha100-1* using genomic fragments has not been reported previously. Similarly, CHORI-321 clones encompassing the genes *cac*, *Dscam*, *lt* and *shakB* complement lethal mutations in the corresponding genes. Rescue of *cac*, *lt* and *shakB* using genomic fragments has also not been reported previously. Rescue of a lethal mutation in *lt* with a 92 kb genomic fragment inserted in euchromatin is surprising, because full expression of *lt* and several other heterochromatic genes has been shown to be dependent on their heterochromatic context [13]. Only one of three clones tested complemented *lt* lethality, suggesting that essential regulatory elements or sufficient genomic context were absent in the other two clones.

To test the utility of recombineering in P(acman) BACs, we introduced EGFP reporter tags into 17 genes encoding transcription factors with well-documented embryonic expression patterns. We inserted the coding region of EGFP in-frame at the 3' end of the open reading frame, replacing the stop codon and creating C-terminal protein fusions [14] (**Supplementary Fig. 4**). Both the untagged and tagged constructs were tested for integration using ΦC31 integrase (**Supplementary Table 2**). Eleven tagged constructs were tested for expression of the fusion protein. Since this EGFP does not fold efficiently in embryos prior to stage 15, we performed immunohistochemistry on embryos with an anti-GFP antibody (**Fig. 2** and **Supplementary Fig. 5a,b**). EGFP fluorescence could be used to visualize fusion protein expression in live embryos

only in the late stages of embryonic development (**Supplementary Fig. 5c**). The expression patterns of *eve*, *D*, *cad*, *Dfd*, *tll*, *slp2* , and *exd* are reproduced by the transgenic fusion constructs (**Supplementary Discussion**). The *en* and *h* gene expression patterns appeared to be exceptions (**Fig. 2k-l**). For *h*, only two stripes (1 and 5) of expression in the embryo were observed, instead of eight [15]. Interestingly, enhancers for stripes 1 and 5 are located in the 7 kb region proximal to the transcription start site, whereas the regulatory elements for the other stripes are located more distally [16]. The latter regulatory elements are lacking in CH322-135D17 used to tag *h*. Hence, the tagged construct is expressed in the expected pattern. Similarly, *en* expression was only observed in 13 stripes and not the head region [17]. This may be due to the absence of regulatory regions in the *en* clone CH322-92I14 (Judith Kassis, personal communication). These experiments show that recombineering-mediated deletion of genomic sequences in P(acman) constructs can be used to dissect the control of transcription by *cis*-regulatory elements.

In conclusion, we have described a versatile P(acman) BAC library resource for functional analysis of transgenes in *D. melanogaster*. This new resource can be used to integrate small and large P(acman) transgenes at defined docking sites in the fruit fly genome using ΦC31 integrase. This allows the rescue of mutations in genes and the engineering of fusion proteins using recombineering that recapitulate endogenous expression patterns. These and many other potential applications are made possible by the novel integration of library design and P(acman) technology (**Supplementary Discussion**). We conservatively estimate that the new resource enables *in vivo* analysis of more than 95% of *D. melanogaster* genes including large genes, gene complexes and heterochromatic genes (**Supplementary Fig. 6**). Moreover, protein tagging should prove a valuable alternative to antibody production, particularly when proteins are poorly immunogenic. Finally, the flexibility of recombineering [5] permits the integration of a variety of

protein tags for numerous applications [18]. The few genes and gene complexes that are too large to be contained within clones in the P(acman) libraries or are otherwise not represented in them can be obtained using the previously described gap-repair procedure [3] and previously mapped and end-sequenced BAC libraries constructed from the same isogenized strain [19,20].

**ACKNOWLEDGMENTS**

We thank the Washington University Genome Sequencing Center for their excellent BAC end sequencing services. We thank the Bloomington Drosophila Stock Center, NCI-Frederick, N. Copeland (NCI Frederick), A. Hyman (Max Planck Institute), R. Karess (CNRS), R. Ordway (Penn State University), J. Reinitz (Stony Brook University), D. Schmucker (Harvard Medical School), T. Schwarz (Children's Hospital, Boston), B. Wakimoto (University of Washington), S. Warming (NCI Frederick) and L. Zipursky (UCLA) for reagents. We are especially thankful to J. Bischof, K. Basler (University of Zurich) and F. Karch (University of Geneva) for providing germ-line ΦC31 sources and information about their use. We thank J. Cohen for help with recombineering, N. Giagtzoglou and A. Rajan for help with microscopy, C. Amemiya and D. Frisch for helpful communications and discussions. We are grateful to B. Wakimoto for critical reading of the manuscript. Confocal microscopy was supported by the BCM Intellectual and Developmental Disabilities Research Center. This work was supported by a grant from the Howard Hughes Medical Institute to H.J.B. and the NIH modENCODE project in collaboration with K.P.W. Work at LBNL was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and LBNL award number LB07003227 to R.A. H.H.J.B. is an Investigator of the Howard Hughes Medical Institute.

## ACCESSION NUMBERS

The sequence of the *attB*-P(acman)-$Cm^R$-BW vector and the P(acman) BAC end sequences have been deposited in GenBank under accession numbers FJ931533 and FI329972 to FI494724, respectively.

## METHODS

Methods and associated references are available as supplementary online material at http://www.nature.com/naturemethods/.

## NOTE

Supplementary information is available on the Nature Methods website.

**FIGURES**

# Figure 1

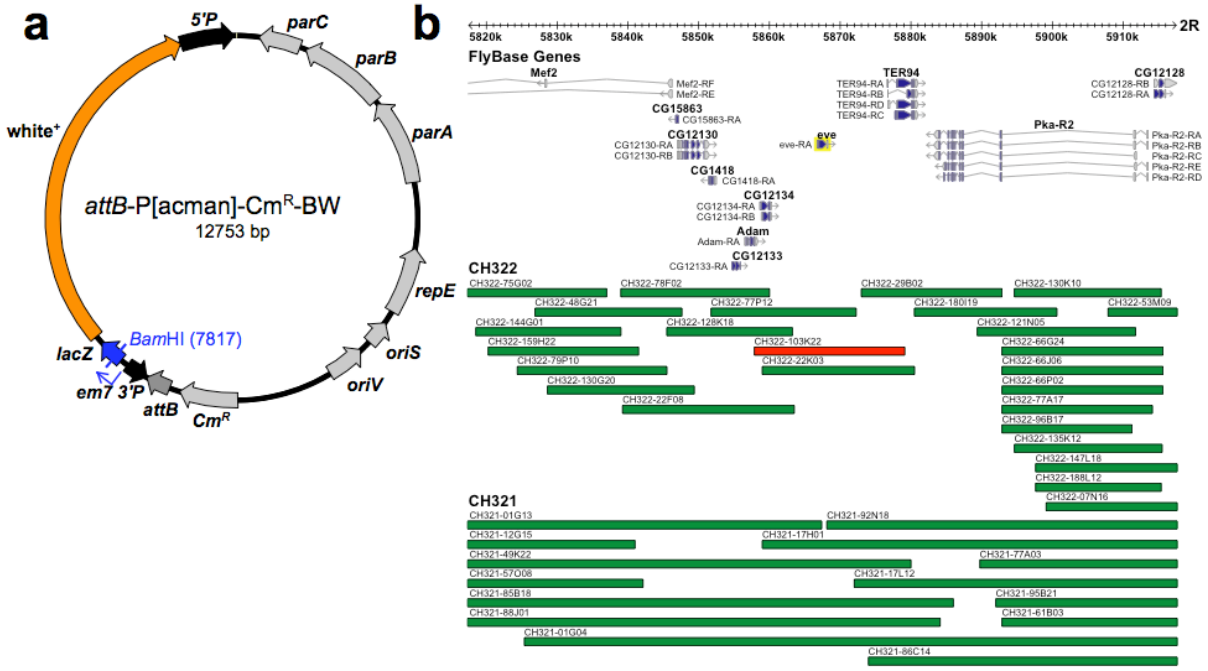



**Figure 1. The P(acman) BAC Vector and Mapped Clones in the *eve* Region.** (**a**) Map of *attB*-P(acman)-Cm$^R$-BW with partitioning functions (*parA*, *parB* and *parC*), low-copy replication functions (*repE* and *oriS*), selectable marker (chloramphenicol acetyl transferase, *Cm$^R$*), conditionally inducible origin of replication (*oriV*) and dominant eye color marker *white$^+$*. Genomic DNA was cloned into the *Bam*HI site (**Supplementary Fig. 2**). (**b**) A 100-kb region surrounding the *eve* gene (yellow) on chromosome arm 2R. Mapped CHORI-321 and CHORI-322 clones are indicated below the FlyBase R5.9 gene annotation. CH322-103K22, selected for transformation (**Supplementary Table 2**) and protein tagging (**Fig. 2**), is indicated in red.

## Figure 2

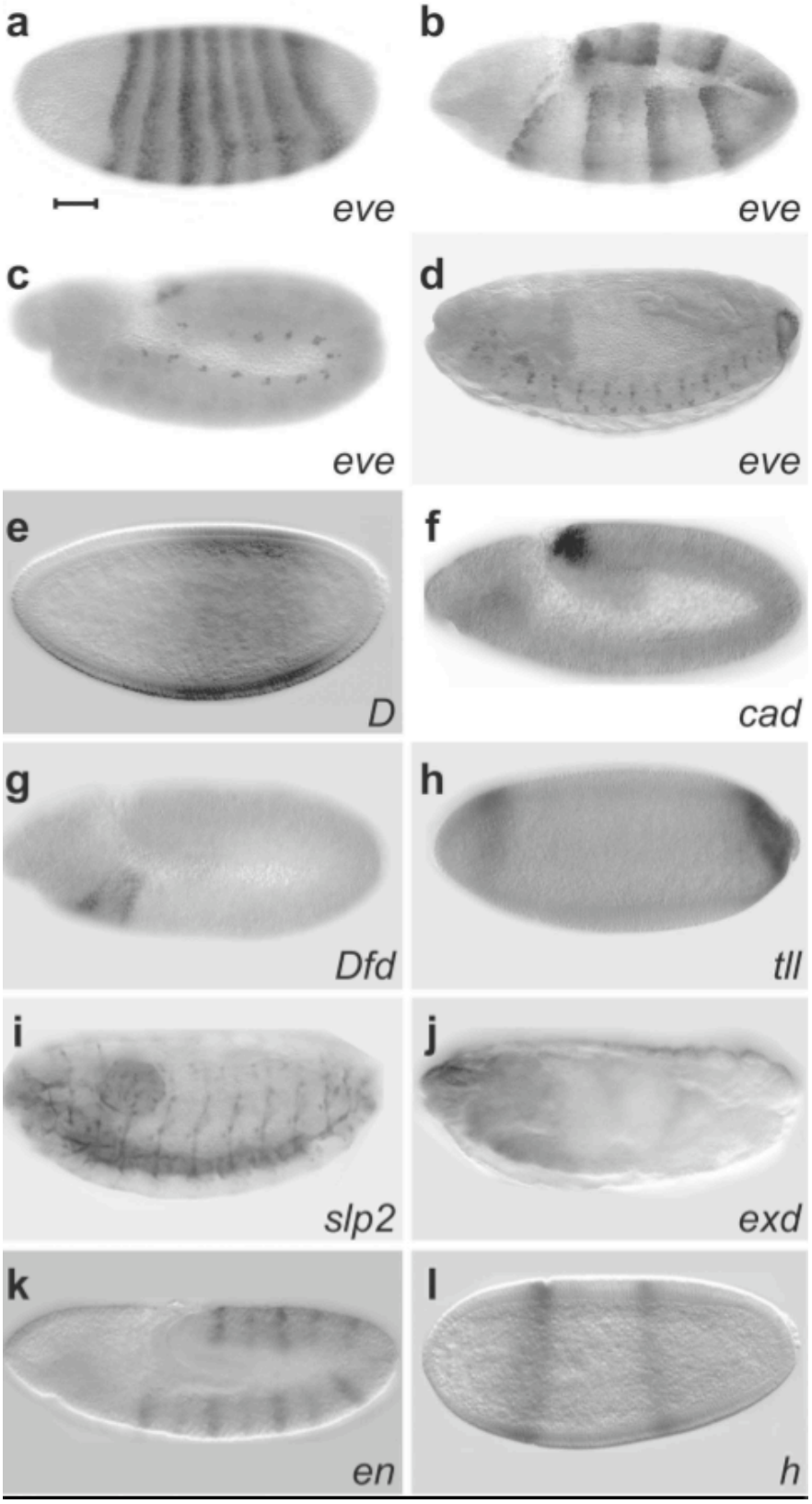


**Figure 2. Expression of EGFP Fusion Proteins in Transgenic Embryos.** Fusion proteins were detected using an anti-GFP antibody and peroxidase staining. (**a-d**) Expression of an *even skipped* fusion construct recapitulates the native pattern of Eve expression. (**a**) embryonic stage 5, (**b**) embryonic stage 9, (**c**) embryonic stage 11, (**d**) embryonic stage 15. (**e**) *Dichaete*, embryonic stage 5. (**f**) *caudal*, embryonic stage 9. (**g**) *Deformed*, embryonic stage 11. (**h**) *tailless*, embryonic stage 5. (**i**) *sloppy paired 2*, embryonic stage 17. (**j**) *extradenticle*, embryonic stage 15. (**k**) *engrailed*, embryonic stage 9. (**l**) *hairy*, embryonic stage 6. Scale bar (**a**) indicates 50 μm.

**TABLES**

| Gene(s) | Clone | Insert | VK# | G0 | Tr | % |
|---|---|---|---|---|---|---|
| ***Act42A*** | *12M14* | 19,321 | 33 | 84 | 2 | 2.4% |
| ***Act87E*** | *158M20* | 19,317 | 37 | 65 | 5 | 7.7% |
| ***alphaTub84B*** | *158D05* | 20,091 | 37 | 51 | 1 | 2.0% |
| ***bcd*** | *100D18* | 18,452 | 37 | 59 | 4 | 6.8% |
| ***CG14438*** | *191E24* | 20,115 | 33 | 103 | 4 | 3.9% |
| ***CG6017*** | *55J22* | 19,815 | 37 | 24 | 3 | 12.5% |
| ***Chc*** | *123J21* | 17,647 | 33 | 57 | 8 | 14.0% |
| ***Clc*** | *92D22* | 20,882 | 37 | 32 | 2 | 6.3% |
| ***Csp*** | *06D09* | 23,790 | 16 | 44 | 5 | 11.4% |
| ***Csp*** | *"* | " | 22 | 11 | 2 | 18.2% |
| ***Csp*** | *"* | " | 37 | 115 | 0 | 0.0% |
| ***Dap160*** | *154I22* | 22,027 | 33 | 73 | 1 | 1.4% |
| ***DIP1*** | *146O15* | 20,342 | 33 | 69 | 2 | 2.9% |
| ***Drp1*** | *83H15* | 19,373 | 33 | 115 | 2 | 1.7% |
| ***endoA*** | *19L12* | 20,365 | 37 | 25 | 1 | 4.0% |
| ***Eps-15*** | *150F15* | 21,222 | 33 | 61 | 1 | 1.6% |
| ***ERR*** | *54A09* | 20,644 | 37 | 70 | 1 | 1.4% |
| ***His2Av*** | *97I07* | 20,846 | 37 | 51 | 1 | 2.0% |
| ***Hr4*** | *137G06* | 22,025 | 33 | 56 | 3 | 5.4% |
| ***Hr96*** | *155C21* | 22,752 | 37 | 38 | 2 | 5.3% |
| ***Khc*** | *162G07* | 21,662 | 33 | 63 | 1 | 1.6% |
| ***ncd*** | *118A01* | 22,196 | 37 | 104 | 4 | 3.8% |
| ***Nmnat*** | *175G15* | 24,376 | 37 | 87 | 2 | 2.3% |
| ***Nrx-IV*** | *154P15* | 21,688 | 37 | 45 | 3 | 6.7% |
| ***n-syb*** | *83G13* | 21,536 | 37 | 48 | 1 | 2.1% |
| ***ogre*** | *155A19* | 22,329 | 33 | 109 | 1 | 0.9% |
| ***Pak*** | *05M18* | 21,251 | 37 | 76 | 4 | 5.3% |
| ***pen*** | *08M17* | 22,049 | 33 | 71 | 1 | 1.4% |
| ***piwi*** | *103C03* | 24,728 | 33 | 43 | 2 | 4.7% |
| ***polo*** | *104P12* | 23,954 | 37 | 28 | 1 | 3.6% |
| ***Rab5*** | *97N16* | 20,416 | 33 | 59 | 4 | 6.8% |
| ***sec15/Rab11*** | *152E24* | 21,979 | 37 | 58 | 2 | 3.4% |
| ***sens*** | *01N16* | 19,175 | 37 | 49 | 9 | 18.4% |
| ***spn-E*** | *93D04* | 23,226 | 37 | 39 | 2 | 5.1% |
| ***sqh*** | *130G10* | 20,074 | 33 | 65 | 8 | 12.3% |
| ***stau*** | *15P05* | 20,978 | 33 | 57 | 3 | 5.3% |
| ***synj*** | *188H18* | 18,311 | 33 | 36 | 1 | 2.8% |

| | | | | | | |
|---|---|---|---|---|---|---|
| ***Syx1A*** | *142K16* | 18,384 | 37 | 45 | 1 | 2.2% |
| ***Vha100-1*** | *119J05* | 19,299 | 37 | 53 | 3 | 5.7% |
| ***wg*** | *192I14* | 22,791 | 13 | 21 | 0 | 0.0% |
| ***wg*** | *"* | " | 31 | 35 | 5 | 14.3% |
| ***wg*** | *"* | " | 33 | 78 | 0 | 0.0% |

**Table 1. Characterization of CHORI-322 Clones.** Genes contained in 38 CHORI-322 clones are indicated. For each clone, the deduced genomic insert length in bp (insert), *attP* VK docking site used (VK#), number of fertile G0 crosses (G0), number of vials resulting in at least one transgenic animal (Tr) and integration efficiency (%) are indicated. Note that only the gene of interest is shown; most clones contain more than one gene.

| Gene(s) | Clone | Insert | VK# | G0 | Tr | % |
|---|---|---|---|---|---|---|
| ***cac*** | *60D21* | 77,150 | 33 | 62 | 2 | 3.2% |
| ***cta*** | *03L03* | 83,080 | 33 | 65 | 1 | 1.5% |
| ***cta*** | *04I17* | 91,022 | 33 | 81 | 1 | 1.2% |
| ***dpp*** | *23O18* | 86,898 | 33 | 41 | 1 | 2.4% |
| ***Dscam*** | *22M14* | 87,314 | 33 | 99 | 2 | 2.0% |
| ***eag*** | *77E01* | 97,072 | 33 | 56 | 1 | 1.8% |
| ***ftz-f1*** | *47I12* | 92,406 | 37 | 51 | 0 | 0.0% |
| ***gfa*** | *57O14* | 77,006 | 33 | 76 | 0 | 0.0% |
| ***hh*** | *61H05* | 101,201 | 37 | 61 | 0 | 0.0% |
| ***Hnf4*** | *12P12* | 104,925 | 33 | 63 | 1 | 1.6% |
| ***Hr38*** | *25N09* | 85,274 | 33 | 107 | 0 | 0.0% |
| ***Hr46*** | *23L02* | 86,475 | 33 | 62 | 1 | 1.6% |
| ***jar*** | *76B03* | 92,518 | 37 | 46 | 0 | 0.0% |
| ***lt*** | *16H04* | 92,084 | 33 | 66 | 6 | 9.1% |
| ***lt*** | *64G01* | 92,368 | 33 | 60 | 4 | 6.7% |
| ***lt/cta*** | *05E14* | 78,102 | 33 | 64 | 2 | 3.1% |
| ***para*** | *18K02* | 98,254 | 33 | 68 | 0 | 0.0% |
| ***rl*** | *36L01* | 53,704 | 33 | 63 | 0 | 0.0% |
| ***rl*** | *81D16* | 102,491 | 33 | 79 | 0 | 0.0% |
| ***shakB*** | *27E22* | 88,712 | 33 | 76 | 2 | 2.6% |
| ***shi*** | *71G22* | 70,514 | 33 | 53 | 0 | 0.0% |
| ***syt*** | *08F02* | 82,244 | 33 | 59 | 0 | 0.0% |
| ***tweek*** | *79N05* | 76,842 | 33 | 69 | 8 | 11.6% |
| ***vas*** | *69O09* | 80,238 | 33 | 62 | 0 | 0.0% |

**Table 2. Characterization of CHORI-321 Clones.** Genes contained in 24 CHORI-321 clones are indicated. Colum headings are identical to those in **Table 1**.

## ONLINE METHODS

### P(acman) BAC Vector Construction

The P(acman) vector *attB*-P(acman)-Cm$^R$-BW was constructed using previously described methods [3] and the DY380 strain (gift from NCI-Frederick) [21]. The multiple cloning site of *attB*-P(acman)-Cm$^R$ was replaced with rpsL-Neo (Genebridges) amplified by PCR using primers BW-rpsL-Neo-F and -R (**Supplementary Table 3**). Next, rpsL-Neo was replaced with an engineered *em7-α-lacZ* obtained by chimeric PCR [22]. The *em7* promoter was amplified from pGalK (gift from S. Warming and N. Copeland) [23] using primers BW-em7-LacZ-F and em7-LacZ-R (**Supplementary Table 3**), and the *α-lacZ* fragment was amplified from pBeloBAC11 [24] using primers em7-LacZ-F and BW-em7-LacZ-R (**Supplementary Table 3**). The two amplicons were fused together using BW-em7-LacZ-F and -R. The resulting plasmid, *attB*-P(acman)-Cm$^R$-BW, was tested for *P* transposase-[25] and ΦC31 integrase-mediated transgenesis[6,7].

### BAC Library Construction

BAC libraries were constructed as described [26] with modifications. High molecular weight genomic DNA from adult flies of the *y$^1$; cn$^1$ bw$^1$ sp$^1$* strain [27] used to produce the reference *D. melanogaster* genome sequence was prepared in agarose blocks, as described [28]. For the CHORI-321 library, agarose-embedded DNA was digested with *Mbo*I (5 U/ml) for 4 min at 37°C. After two rounds of size-fractionation on CHEF gels, the 80-100 kb fraction was ligated into the *Bam*HI site of *attB*-P(acman)-Cm$^R$-BW. For the CHORI-322 library, digestion was performed similarly, but samples were removed at 4, 20 and 36 min. In a first size-fractionation on a CHEF gel, 20-40 kb fractions were isolated. To ensure a narrow insert size range, a second

size fractionation was performed using standard gel electrophoresis. The 20-25 kb fractions were ligated into the *Bam*HI site of *attB*-P(acman)-Cm$^{R}$-BW. Following transformation of the two libraries into EPI300$^{TM}$-T1R cells (Epicentre), clones with inserts (white colonies) were selected on LB plates (12.5 μg/ml Chl, X-Gal and IPTG). The CHORI-321 and CHORI-322 libraries were arrayed into 96 and 192 384-well microtiter plates, respectively.

**BAC End Sequencing, Genomic Alignment and Genome Browser**

P(acman) BAC cultures were grown in LB (12.5 μg/ml Chl), and plasmid copy number was induced using CopyControl$^{TM}$ solution (Epicentre). Paired BAC end sequences were determined at the Washington University Genome Sequencing Center using the sequencing primers Pac-BW-F and -R ( **Supplementary Table 3** and **Fig. 1b**), BigDye v3.1 terminator chemistry and the direct sequencing from culture procedure (Applied Biosystems).

Sequence reads were basecalled using KB (v1.2) (Applied Biosystems). Reads were trimmed to remove vector sequences and low quality data. Trimmed reads were aligned to the *D. melanogaster* Release 5 reference genome sequence (http://www.fruitfly.org/sequence/release5genomic.shtml) using BLASTN [29] with parameters appropriate for the low polymorphism rate in the isogenized strain (Q = 35, R = 10, gapL = 1.37, gapK = 0.71, gapH = 1.31, -wordmask dust). High-scoring alignments of at least 95% identity over 200 bp were identified. Clones with a unique pair of high-scoring alignments with proper orientation and a deduced insert size of 15-30 kb for CHORI-322 or 40-120 kb for CHORI-321 were defined as “mapped”. Clones with just one end sequence with a unique high-scoring alignment were defined as “partially mapped”. Rare clones with inconsistent paired end sequence alignments were rejected.

Limited curation was performed to refine the mapping results. First, five artifactual alignments associated with two gaps in the genome sequence were identified and deselected. Four clones appeared to span the large (> 500 kb) tandem array of histone genes on chromosome arm 2L in polytene division 39C, and one clone appeared to span the unsized gap on 2L in 40B. These clones were curated to deselect artifactual alignments to repeated sequences. Second, to reduce or span gaps in genome coverage by CHORI-321 clones, 68 partially mapped clones were curated to select consistent alignments of the paired end sequence that did not pass the conservative criteria used in automated mapping. Forty-three of these clones map to the X chromosome where library coverage is lower. In addition to the expected 17 gaps at the tip and base of each chromosome arm and at five gaps in the genome physical map [28], there are eight gaps in mapped P(acman) clone coverage of the Release 5 chromosome arm sequences, five of which are on the X chromosome. Coordinates of mapped and partially mapped clone ends on the Release 5 genome sequence are available on the P(acman) Resources home page (http://pacmanfly.org/).

An implementation of Gbrowse [30] was created to provide public online access to data on mapped and partially mapped P(acman) BAC clones. The browser is available via a link on the P(acman) Resources home page (http://pacmanfly.org/) (**Fig. 1b** and **Supplementary Fig. 6**). The graphical user interface supports searches by gene symbol, clone name, polytene map location, or genome sequence coordinates. Searches return genome browser displays of clone locations, the current FlyBase gene annotation, and any other user-defined annotation tracks. Individual clone report pages present genomic coordinates of clone ends and deduced sequences of cloned inserts.

**ΦC31 Integrase-Mediated Transformation of *D. melanogaster***

P(acman) clones were integrated in two docking sites in the genome: VK37 (*PBac{y*$^{+}$*-attP-3B}VK00037*) located on chromosome 2 and VK33 (*PBac{y*$^{+}$*-attP-3B}VK00033*) on chromosome 3 [3]. The docking sites were chosen based on high hatching rate of injected embryos, high rate of fertile crosses, high integration efficiencies of *attB*-P(acman)-Ap$^{R}$ [3] (~50%) and the 18 kb insert containing *sens-L* clone [3] (~30%), and receptivity for the large insert clones *Dscam-1* [3], *E(Spl)-C* [3] and/or *Dscam-2* [3] (data not shown). Chromosomes bearing each docking site were brought together with the X-linked ΦC31 integrase transgene *M{vas-int.B}ZH-2A* transgene (gift of J. Bischof, K. Basler and F. Karch) [7]. Fragments containing X or second chromosome genes were integrated into VK33. Fragments containing third chromosome genes were integrated into VK37. Fragments that did not integrate into VK33 (*wg*) or VK37 (*Csp*), were integrated into VK13 and VK31 (*wg*), or VK16 and VK22 (*Csp*), using similarly designed stocks.

On average, 124 embryos were injected for each construct using standard procedures, resulting in 89 hatching larvae (72%). Single G0 males were crossed to five *y w* virgins, or one G0 virgin female was crossed to three *y w* males in vials, resulting in 50 fertile crosses (56%) on average for each construct. The transformation efficiency is defined as the percentage of G0 fertile crosses that produce one or more transformant. Transgene inserts were balanced using *y*$^{1}$ *w*$^{67c23}$*; In(2LR)Gla, wg*$^{Gla-1}$*/SM6a* (Bloomington Drosophila Stock Center, BDSC #6600) for VK37 integrants and *y w; D/TM6B, Tb, Hu* for VK33 integrants. The ΦC31 integrase-containing chromosome was removed from each transgenic line. For nomenclature purposes, transgenic insertions were named "Docker::Insert" (e.g. VK37::*Bcd*/*CH322-100D18*).

**Molecular Characterization of Integration Events**

For PCR verification, DNA was extracted from 10 flies using the Gentra Puregene Tissue Kit (Qiagen). DNAs isolated from *y w*, VK33 and VK37 were used as controls. A multiplex PCR procedure was developed that verifies correct integration into VK [3], *P{CaryP}* [6,10] and *M{3xP3-RFP.attP}* [7] docking sites. PCR was performed with four primers (attP-F and -R, P(acman)-F and -R) (**Supplementary Table 3**), and HotStart Taq polymerase (Qiagen) in the presence of 5% DMSO. PCR products were separated on 2.5% agarose gels. Correct integration events were identified by loss of the *attP* (134 bp) PCR product, specific to the original docking site, and the appearance of both *attL* (227 bp) and *attR* (454 bp) PCR products, specific for the integration event, as illustrated for six different VK docking sites **(Supplementary Fig. 3)**.

**Recombineering-Mediated Protein Tagging**

Recombineering was performed as described [14] using the SW102 strain [23] (gift from NCI Frederick) with modifications (**Supplementary Fig. 4**). Tags containing EGFP followed by a kanamycin resistance marker [14] were PCR amplified from the C-terminal LAP-tag containing plasmid (gift from A. Hyman) using the recombineering primers, GOI-F and GOI-R (**Supplementary Table 4**). PCR was performed using Phusion Taq Polymerase (Finnzymes). PCR products were purified with the Qiaquick PCR purification kit (Qiagen) and verified on agarose gels. P(acman) clones for each gene (**Supplementary Table 2**) were identified and verified with PCR using primers GOI_F and GOI_R, located 100 bp upstream and downstream of the stop codon respectively (**Supplementary Table 5**). PCR reactions were performed using Analytical PCR (Promega). From verified clones, plasmid DNA was prepared using the Qiaprep Spin Miniprep Kit (Qiagen) after CopyControl (Epicentre) induction, electroporated into SW102 cells, and selected on LB plates (12.5 µg/ml Chl, 10 µg/ml Tet) at 32°C. Recombineering was

performed using the PCR amplified tags. Recombinants were isolated on LB plates (12.5 μg/ml Chl, 10 μg/ml Tet, 25 μg/ml Kan) at 32°C. Single colonies were screened for presence of the tag using PCR, as described above. Modified P(acman) clones were transferred into EPI300 cells (Epicentre) for plasmid copy induction. Correct tag recombination was verified by sequencing.

## REFERENCES


1. Bier, E. Drosophila, the golden bug, emerges as a tool for human genetics. *Nat. Rev. Genet.* **6**, 9-23 (2005).
2. Venken, K. J. & Bellen, H. J. Transgenesis upgrades for Drosophila melanogaster. *Development* **134**, 3571-3584 (2007).
3. Venken, K. J., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science* **314**, 1747-1751 (2006).
4. Wild, J., Hradecna, Z. & Szybalski, W. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res.* **12**, 1434-1444 (2002).
5. Sawitzke, J. A. *et al.* Recombineering: in vivo genetic engineering in E. coli, S. enterica, and beyond. *Methods Enzymol.* **421**, 171-199 (2007).
6. Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic Drosophila by using the site-specific integrase from phage ϕC31. *Genetics* **166**, 1775-1782 (2004).
7. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for Drosophila using germ-line-specific ϕC31 integrases. *Proc. Natl. Acad. Sci. U. S. A* **104**, 3312-3317 (2007).
8. Venken, K. J. *et al.* Recombineering-mediated tagging of Drosophila genomic constructs for in vivo localization and acute protein inactivation. *Nucleic Acids Res.* **36**, e114 (2008).
9. Pepple, K. L. *et al.* Two-step selection of a single R8 photoreceptor: a bistable loop between senseless and rough locks in R8 fate. *Development* **135**, 4071-4079 (2008).
10. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E. & Perrimon, N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat. Genet.* **40**, 476-483 (2008).
11. Ni, J. Q. *et al.* Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster. *Nat. Methods* **5**, 49-51 (2008).
12. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in Drosophila. *Proc. Natl. Acad. Sci. U. S. A* **105**, 9715-9720 (2008).
13. Yasuhara, J. C. & Wakimoto, B. T. Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet.* **22**, 330-338 (2006).
14. Poser, I. *et al.* BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5**, 409-415 (2008).
15. Hooper, K. L., Parkhurst, S. M. & Ish-Horowicz, D. Spatial control of hairy protein expression during embryogenesis. *Development* **107**, 489-504 (1989).
16. Howard, K. R. & Struhl, G. Decoding positional information: regulation of the pair-rule gene hairy. *Development* **110**, 1223-1231 (1990).
17. DiNardo, S., Kuner, J. M., Theis, J. & O'Farrell, P. H. Development of embryonic pattern in D. melanogaster as revealed by accumulation of the nuclear engrailed protein. *Cell* **43**, 59-69 (1985).
18. Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-224 (2006).
19. Adams, M. D. *et al.* The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-2195 (2000).

20. Hoskins, R. A. *et al.* Sequence finishing and mapping of Drosophila melanogaster heterochromatin. *Science* **316**, 1625-1628 (2007).
21. Lee, E. C. *et al.* A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65 (2001).
22. Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques* **32**, 728-730 (2002).
23. Warming, S., Costantino, N., Court DL, Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).
24. Kim, U. J. *et al.* Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**, 213-218 (1996).
25. Rubin, G. M. & Spradling, A. C. Genetic transformation of Drosophila with transposable element vectors. *Science* **218**, 348-353 (1982).
26. Osoegawa, K. *et al.* An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**, 1-8 (1998).
27. Brizuela, B. J., Elfring, L., Ballard, J., Tamkun, J. W. & Kennison, J. A. Genetic analysis of the brahma gene of Drosophila melanogaster and polytene chromosome subdivisions 72AB. *Genetics* **137**, 803-813 (1994).
28. Hoskins, R. A. *et al.* A BAC-based physical map of the major autosomes of Drosophila melanogaster. *Science* **287**, 2271-2274 (2000).
29. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).
30. Stein, L. D. *et al.* The generic genome browser: a building block for a model organism system database. *Genome Res.* **12**, 1599-1610 (2002).

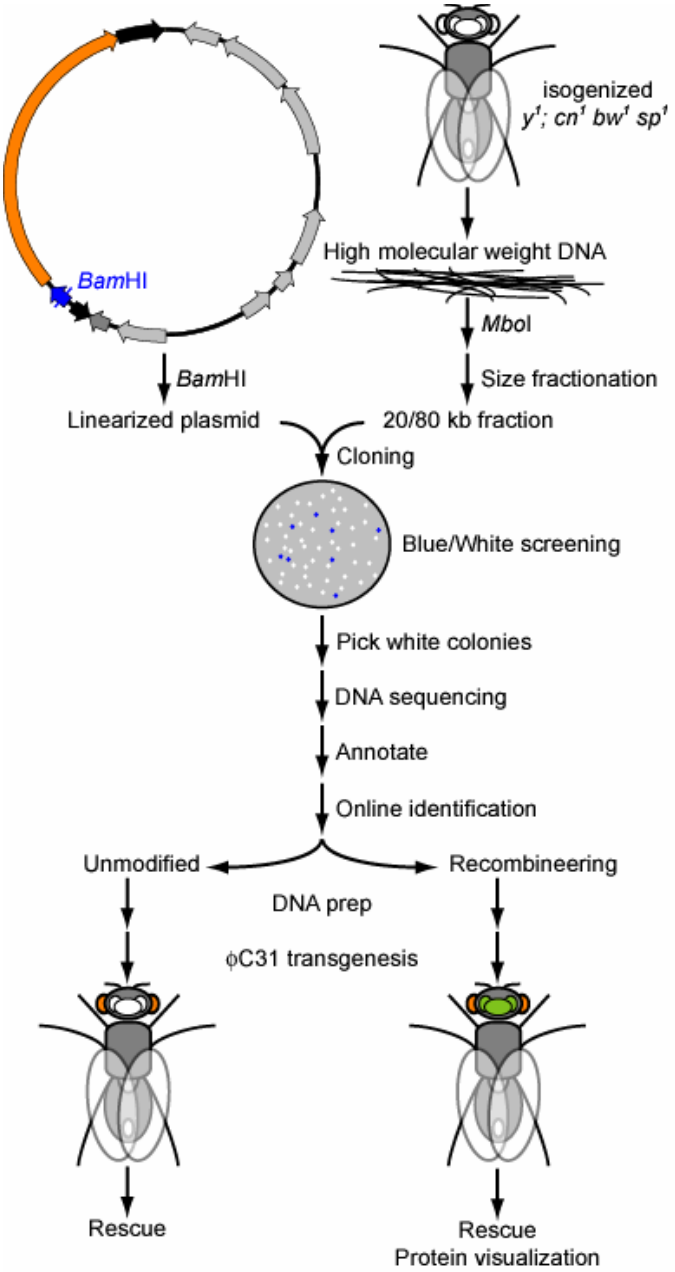
## LIST OF SUPPLEMENTARY ITEMS

| Supplementary File | Title |
|---|---|
| **Supplementary Figure 1** | The em7-lacZ Marker Gene. |
| **Supplementary Figure 2** | Overview of the P(acman) BAC System. |
| **Supplementary Figure 3** | Verification of ΦC31-Mediated Site-Specific Integration by Multiplex PCR. |
| **Supplementary Figure 4** | Overview of the Recombineering Strategy. |
| **Supplementary Figure 5** | Embryonic Expression of EGFP Fusion Proteins in Transgenic Flies. |
| **Supplementary Figure 6** | P(acman) BAC Coverage of a Gene Complex and a Heterochromatic Region. |
| **Supplementary Table 1** | P(acman) BAC Coverage of Genes and the Genome. |
| **Supplementary Table 2** | Tagging and Transformation of CHORI-322 Clones. |
| **Supplementary Table 3** | Cloning primers used in Online Methods. |
| **Supplementary Table 4** | Primers for Recombineering. |
| **Supplementary Table 5** | Primers for PCR and Sequencing. |
| **Supplementary Discussion** | |
| **Supplementary Methods** | |
| **Supplementary References** | |

**Supplementary Figure 1**


**Supplementary Figure 1. The em7-lacZ Marker Gene.** Sequence of the *attB*-P(acman)-$Cm^R$-BW vector in the region of the cloning site and the *em7-lacZ* marker gene, indicated in blue text. The *em7* promoter is indicated in lower case text, and the truncated *lacZ* coding sequence is indicated in upper case text. The premature stop codon in the *lacZ* allele, the sequences of primers Pac-BW-F and Pac-BW-R and the *Bam*HI cloning site are underlined.

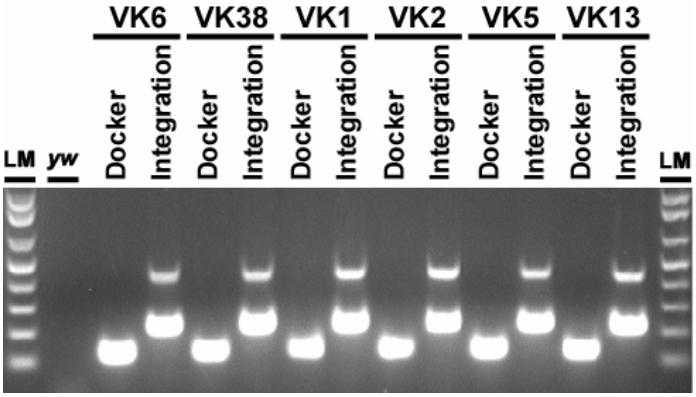
**Supplementary Figure 2**




**Supplementary Figure 2. Overview of the P(acman) BAC System.** High molecular weight DNA, isolated from the isogenized *y*$^{1}$*; cn*$^{1}$ *bw*$^{1}$ *sp*$^{1}$ strain, was partially digested with *Mbo*I and ligated into the *Bam*HI site of the *attB*-P(acman)-Cm$^{R}$-BW plasmid. White recombinant clones

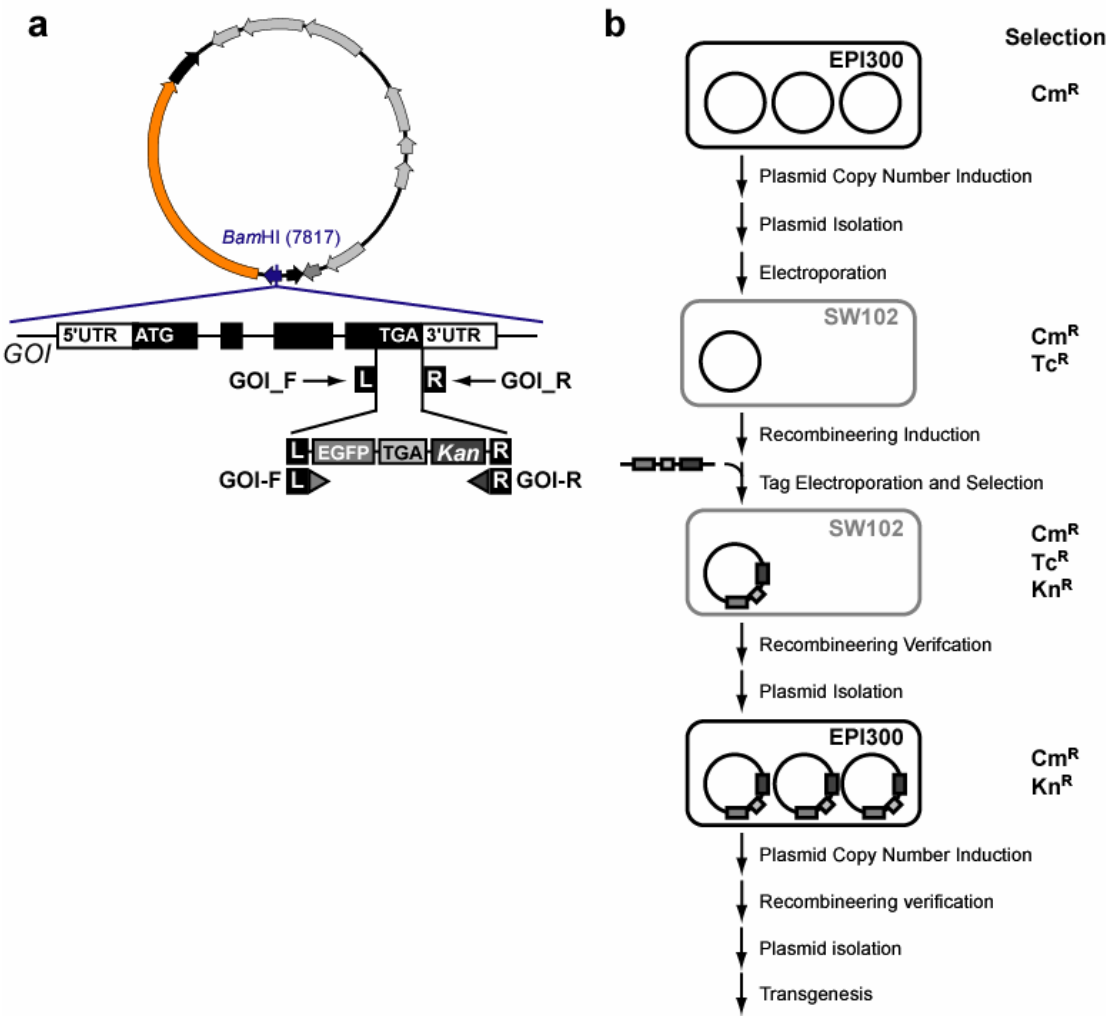
were identified after blue/white screening on LB plates using X-Gal. Clones were end-sequenced with primers Pac-BW-F and -R (A), and the sequences were mapped to the reference genome. Clones containing genes of interest are identified in an online genome browser www.pacmanfly.org and obtained from the BACPAC Resources Center. Clones can be modified through recombineering (**Supplementary Fig. 3**), for example by incorporating protein tags such as EGFP. Both unmodified and modified constructs are integrated into the *D. melanogaster* germ line using ΦC31 integrase, and transgenic flies are identified as *white*$^{+}$ (orange eyes) flies. Transgenic flies, containing untagged or tagged constructs, can be tested in mutant rescue experiments (**Tables 1**, **2** and **Supplementary Table 2**). In addition, tagged clones can be used for *in vivo* protein localization using EGFP (green) (**Fig. 2**) or other tags, and in other experiments requiring engineered protein or peptide tags.

**Supplementary Figure 3**




**Supplementary Figure 3. Verification of ΦC31-Mediated Site-Specific Integration by Multiplex PCR.** Integration events were identified by loss of the *attP* (134 bp) PCR product specific for the docking site (Docker), and the appearance of both *attL* (227 bp) and *attR* (454 bp) PCR products specific for the integration event (Integration), as illustrated for six different VK docking sites: VK6 and VK38 on the X chromosome, VK1 and VK2 on chromosome 2, and VK5 and VK13 on chromosome 3. A fly strain containing no docking sites (*y w*) is the negative control. LM, molecular weight marker.

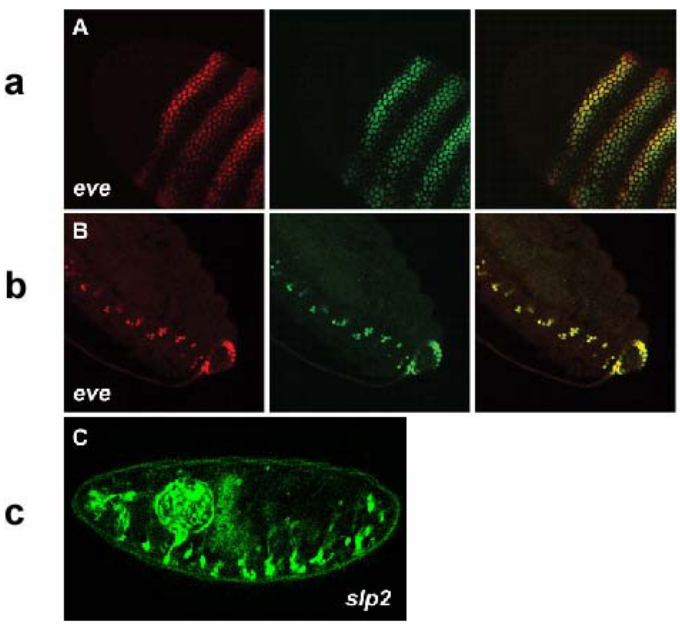
**Supplementary Figure 4**




**Supplementary Figure 4. Overview of the Recombineering Strategy.** (**a**) Schematic of a P(acman) BAC clone containing a gene of interest (GOI) and the recombineering strategy. The 5' untranslated region (UTR), start codon (ATG), stop codon (TGA) and 3'UTR are indicated. Exons are illustrated as boxes: UTRs and coding exons are indicated in white and black, respectively. The tag, consisting of EGFP followed by a stop codon and a kanamycin resistance marker (*Kan*), is amplified by PCR using recombineering primers GOI-F and GOI-R

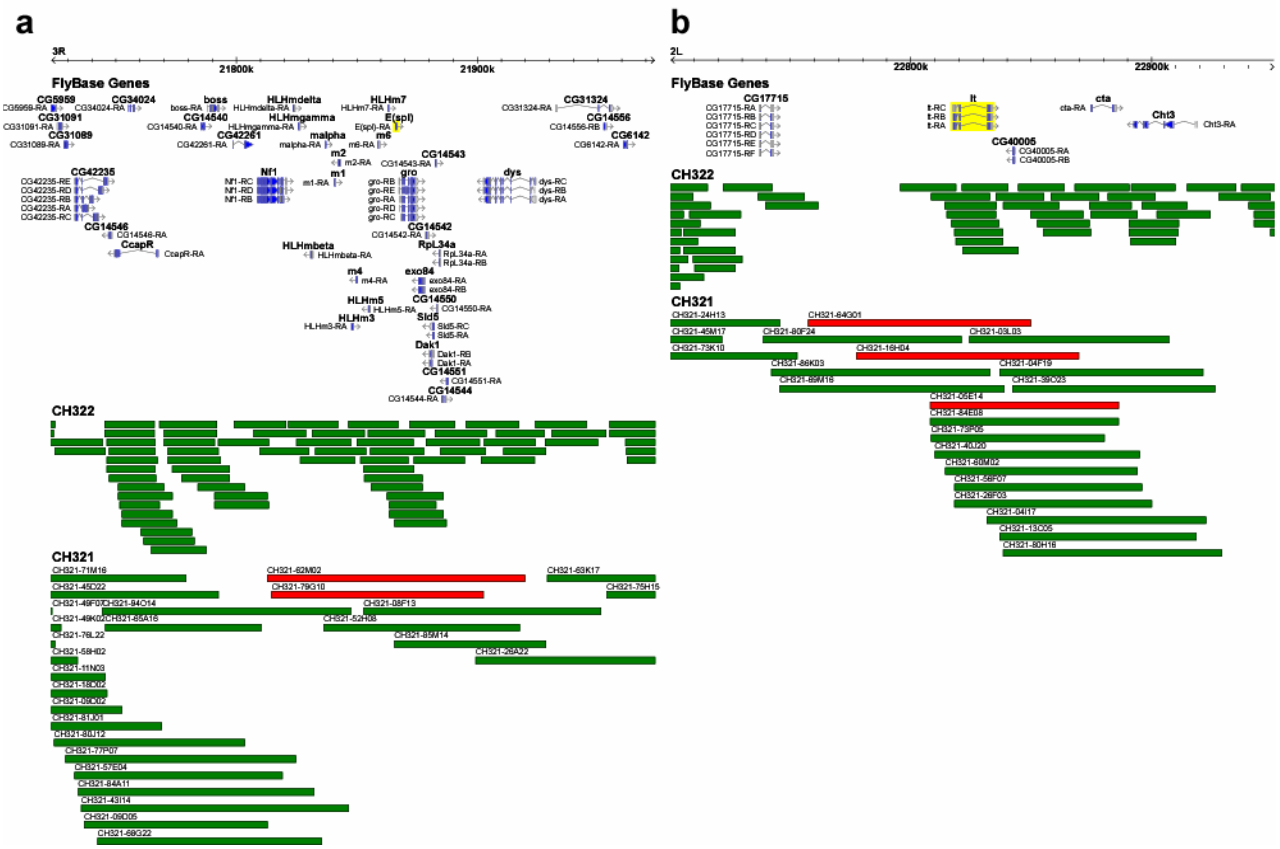
(**Supplementary Table 5**). Recombineering homology regions are indicated on the left (L) and right (R). Recombineering replaces the endogenous stop codon; a novel stop codon is introduced downstream of the EGFP tag. Verification of correct recombineering is performed using primers GOI_F and GOI_R (**Supplementary Table 6**). (**b**) Flow diagram of the procedure. The plasmid copy number of the P(acman) BAC ($Cm^R$) is induced in EPI300 cells. Plasmid is isolated and electroporated into the tetracycline resistant ($Tc^R$) SW102 strain. Recombineering functions are induced, and the strain is electroporated with PCR product encompassing EGFP, kanamycin resistance ($Kn^R$) and homology arms (A). Potential recombinants are selected ($Cm^R$, $Tc^R$, $Kn^R$) and verified. The recombinant construct is isolated, transformed into EPI300 and selected ($Cm^R$, $Kn^R$). The plasmid copy number is induced, the recombination event is verified, and plasmid is isolated for transformation into the fly genome.

**Supplementary Figure 5**




**Supplementary Figure 5. Embryonic Expression of EGFP Fusion Proteins in Transgenic Flies.** (**a, b**) Double labeling with anti-Eve (red) and anti-GFP (green) antibodies confirms proper expression of the transgene at embryonic stage 5 (**a**) and embryonic stage 15 (**b**). (**c**) EGFP fluorescence in a live stage 17 embryo bearing the *sloppy paired 2* fusion construct. Compare to **Fig. 2i**.

**Supplementary Figure 6**




**Supplementary Figure 6. P(acman) BAC Coverage of a Gene Complex and a Heterochromatic Region.** Mapped CHORI-321 and CHORI-322 P(acman) BAC clones (green) are indicated below the FlyBase R5.9 gene annotation. Note that the clone names for the CHORI-322 library are not indicated due to the large size of both genomic regions. (**a**) A 250-kb region surrounding the *E(Spl)* complex on chromosome 3R at polytene division 96F10. Two clones, CH321-62M02 and CH321-79G10, that encompass the entire gene complex are indicated in red. (**b**) A 250-kb region surrounding the *lt* gene (yellow) on chromosome arm 2L at polytene division 40B in heterochromatin. CH321-64G01, CH321-16H04 and CH321-05E14, selected for

transformation rescue experiments, are indicated in red. CH321-64G01 complemented the lethality of a *lt* mutant; CH321-16H04 and CH321-05E14 did not.

**Supplementary Table 1**

| Library | Genes Spanned | | | Genome Coverage | |
|---|---|---|---|---|---|
| | +/- 0 kb | +/- 2 kb | +/- 5 kb | X | A |
| CHORI-321 | 14,410 (99.3%) | 14,362 (98.9%) | 14,270 (98.3%) | 8.2X | 9.3X |
| CHORI-322 | 12,898 (88.9%) | 12,121 (83.5%) | 9,750 (67.2%) | 4.3X | 5.9X |

**Supplementary Table 1. P(acman) BAC Coverage of Genes and the Genome.** The number and percentage of FlyBase R5.9 gene models spanned by P(acman) clones in the two BAC libraries, including 0 kb, 2 kb and 5 kb of additional sequence at the 5' and 3' ends of each model to account for unannotated transcribed and regulatory sequences, is shown. The redundancy of coverage of the genomic sequences of the X chromosome (X) and the autosomes (A) is shown.

**Supplementary Table 2**

| Inj | Gene | Clone | Insert | VK# | G0 | Tr | % |
|---|---|---|---|---|---|---|---|
| **1-U** | ***bcd*** | *100D18* | 18,452 | 37 | 47 | 8 | 17.0% |
| **1-T** | | | | | 22 | 2 | 9.1% |
| **2-U** | ***hb*** | *05M03* | 24,093 | 37 | 56 | 0 | 0.0% |
| **2-T** | | | | | 76 | 0 | 0.0% |
| **3-U** | ***en*** | *92I14* | 21,093 | 33 | 34 | 9 | 26.5% |
| **3-T** | | | | | 100 | 16 | 16.0% |
| **4-U** | ***gt*** | *05H16* | 24,464 | 33 | 31 | 1 | 3.2% |
| **5-U** | ***eve*** | *103K22* | 21,174 | 33 | 118 | 2 | 1.7% |
| **5-T** | | | | | 35 | 1 | 2.9% |
| **6-U** | ***run*** | *115B21* | 19,432 | 33 | 80 | 9 | 11.3% |
| **6-T** | | | | | 21 | 0 | 0.0% |
| **7-U** | ***cad*** | *115J21* | 21,274 | 33 | 59 | 4 | 6.8% |
| **7-T** | | | | | 39 | 6 | 15.4% |
| **8-U** | ***tll*** | *83J21* | 23,996 | 37 | 31 | 1 | 3.2% |
| **8-T** | | | | | 49 | 2 | 4.1% |
| **9-U** | ***slp2*** | *127L06* | 21,598 | 33 | 57 | 16 | 28.1% |
| **9-T** | | | | | 54 | 3 | 5.6% |
| **10-U** | ***Dfd*** | *75M21* | 20,767 | 37 | 56 | 2 | 3.6% |
| **10-T** | | | | | 54 | 1 | 1.9% |
| **11-U** | ***Chi*** | *91I05* | 18,227 | 33 | 34 | 1 | 2.9% |
| **11-T** | | | | | 114 | 1 | 0.9% |
| **12-U** | ***Bro*** | *16P24* | 26,146 | 37 | 37 | 9 | 24.3% |
| **12-T** | | | | | 30 | 1 | 3.3% |
| **13-U** | ***exd*** | *35A16* | 22,572 | 33 | 108 | 2 | 1.9% |
| **13-T** | | | | | 90 | 1 | 1.1% |
| **14-U** | ***D*** | *169N20* | 22,404 | 37 | 43 | 1 | 2.3% |
| **14-T** | | | | | 52 | 1 | 1.9% |
| **15-U** | ***h*** | *135D17* | 18,449 | 37 | 19 | 2 | 10.5% |
| **15-T** | | | | | 20 | 2 | 10.0% |
| **16-U** | ***ems*** | *143M02* | 19,969 | 37 | 46 | 2 | 4.3% |
| **16-T** | | | | | 61 | 5 | 8.2% |
| **17-U** | ***kni*** | *25A12* | 23,149 | 37 | 31 | 4 | 12.9% |
| **17-T** | | | | | 71 | 0 | 0.0% |

**Supplementary Table 2. Tagging and Transformation of CHORI-322 Clones.** Genes spanned by 17 CHORI-322 clones were tagged by recombineering; untagged (U) and tagged (T) constructs are indicated. For each clone, the genomic insert length in bp (Insert), *attP* VK

docking site used (VK#), number of fertile G0 crosses (G0), number of vials resulting in at least one transgenic animal (Tr) and integration efficiency (%) are indicated. Note that only the gene of interest is indicated; most clones contain more than one gene. Construct 4-T resulted in an aberrant supercoiled plasmid migration pattern after gel electrophoreses and was not injected.

**Supplementary Table 3**

| Primer name | Sequence |
|---|---|
| **BW-rpsL-Neo-F** | AGTTTTATTTTTAATAATTTGCGAGTACGCAACCGGTGGCCTGGTGATGATGGCGGGATC |
| **BW-rpsL-Neo-R** | AAAAATGGGTTTTATTAACTTACATACATACTAGAATTCTCAGAAGAACTCGTCAAGAAG |
| **BW-em7-LacZ-F** | TAAAGTTTTATTTTTAATAATTTGCGAGTACGCAACCGGTCCTGTTGACAATTAATCATC |
| **em7-LacZ-R** | CTTGGCGTAATCATGGTCATGATCTGCCTCCTGGGTTTAG |
| **em7-LacZ-F** | CTAAACCCAGGAGGCAGATCATGACCATGATTACGCCAAG |
| **BW-em7-LacZ-R** | GCAAAAATGGGTTTTATTAACTTACATACATACTAGAATTCCTATGCGGCATCAGAGCAG |
| **Pac-BW-F** | ATCGGCATAGTATATCGGCATAG |
| **Pac-BW-R** | GATGTGCTGCAAGGCGATTAAGT |
| **attP-F** | AGGTCAGAAGCGGTTTTCGGGAGTAGTG |
| **attP-R** | GGTCGTAAGCACCCGCGTACGTGTCCAC |
| **P(acman)-F** | ACGCCTGGTTGCTACGCCTGAATAAGTG |
| **P(acman)-R** | CCCACGGACATGCTAAGGGTTAATCAAC |

**Supplementary Table 3. Primers for Cloning and Sequencing (see Online Methods).**

**Supplementary Table 4**

| Primer name | Sequence |
|---|---|
| **bicoid-F** | ATCCCCATCGGAACGCCGCGGGCAACTCGCAGTTTGCCTACTGCTTCAATGATTATGATATTCCAACTACTG |
| **bicoid-R** | AGTGGTTAACCTAAAGCTAATGAAACTCTCTAACACGCCTCTCATCCAGGTCAGAAGAACTCGTCAAGAAG |
| **hunchback-F** | GCGACGGACCCGTCGGCCTCTTCGTTCACATGGCCAGGAATGCTCACTCCGATTATGATATTCCAACTACTG |
| **hunchback-R** | ATATGATAATAGTGATAAATAATAATAACAAGGTGATGGTGATGGGGAACTCAGAAGAACTCGTCAAGAAG |
| **engrailed-F** | TGACCAAGGAGGAGGAGGAGCTCGAGATGCGCATGAACGGGCAGATCCCCGATTATGATATTCCAACTACTG |
| **engrailed-R** | ACGCCCCCGTCAGAAGGAGTAACCCCTTATGGGTAACCATTGGTCAGCGCTCAGAAGAACTCGTCAAGAAG |
| **giant-F** | CCCTCAAGGTCCAGCTGGCCGCCTTCACCTCCGCCAAAGTAACCACCGCCGATTATGATATTCCAACTACTG |
| **giant-R** | ACATACGATTCGGATCCTCGCGTTCAACGCATCAAGAGAGGAGTGGACCTTCAGAAGAACTCGTCAAGAAG |
| **eve-F** | TGATTGCGGAGCCCAAGCCGAAGCTCTTCAAGCCCTACAAGACTGAGGCGGATTATGATATTCCAACTACTG |
| **eve-R** | CTTTTGGGGGAGCATGGGGGGGGGGGAGAGAGTGTGTGTGGATCGCGGGCTCAGAAGAACTCGTCAAGAAG |
| **runt-F** | CCAAGATCAAGAGCGCCGCCGTGCAGCAGAAGACCGTGTGGCGGCCCTACGATTATGATATTCCAACTACTG |
| **runt-R** | TATATCACTTTGTTTTCTTCATTCCTCCAGATTTTTGGGGATCAGATGCCTCAGAAGAACTCGTCAAGAAG |
| **caudal-F** | AGCACAGTGCGCAGATGTCCGCTGCGGCGGCAGTGGGCACGCTCTCGATGGATTATGATATTCCAACTACTG |
| **caudal-R** | TCCATGTAGTTGTTACTGTCGCCGCTCGCCGCATAACAGGAATGGTCGTGTCAGAAGAACTCGTCAAGAAG |
| **tailless-F** | ACATCACCATTGTGCGCCTCATCTCCGACATGTACAGTCAGCGCAAGATCGATTATGATATTCCAACTACTG |
| **tailless-R** | ACTTGGAAGGCACTTCGAGTGCGGCGATTAGTCTAGGCTCTACATACTTTTCAGAAGAACTCGTCAAGAAG |
| **slp2-F** | CCTCGCCACAGCCTCTTCACAAACCCGTCACCGTAGTCTCCCGCAATAGCGATTATGATATTCCAACTACTG |
| **slp2-R** | TATGCTGCTTCTCAAGCCAAAGAATCCTCCGATTCTTCTCGCCAGAATCTTCAGAAGAACTCGTCAAGAAG |
| **Deformed-F** | TGACCAATCTTCAGCTACACATCAAGCAGGACTACGATCTGACGGCCCTGGATTATGATATTCCAACTACTG |
| **Deformed-R** | GGTATTTAATTTACAATGCTGAAGTCATCTTGAAGATATCCCTGCTGATTTCAGAAGAACTCGTCAAGAAG |
| **Chip-F** | ACAACGCGAATATATCTGACATAGATAAAAAGAGCCCCATTGTATCGCAAGATTATGATATTCCAACTACTG |
| **Chip-R** | GTATGAAAATACATTTGACAATATAGCGAAATATTATGTTTTATTAAGTTTCAGAAGAACTCGTCAAGAAG |
| **Brother-F** | ATCACCATCACAGAGGTGGGCCTGGTCTGCCTAGAGGACCAATGGGATGGGATTATGATATTCCAACTACTG |
| **Brother-R** | TTGTTGATTTACTTAGATGTTACTTAAAGCTATGAGGGTAATATATAGCATCAGAAGAACTCGTCAAGAAG |
| **extradenticle-F** | ACAACAGCATGGGCGGCTACGACCCAAATCTCCATCAGGATCTAAGCCCCGATTATGATATTCCAACTACTG |
| **extradenticle-R** | CAACTGTATGAGGGATTCTCCGGACTGGGAGTGGTTCACAGCCCTGATCCTCAGAAGAACTCGTCAAGAAG |
| **Dichaete-F** | TGGATGGTTCCATGGACAGTGCCCTGAGGCGACCGGTTCCGGTGCTCTATGATTATGATATTCCAACTACTG |
| **Dichaete-R** | AATGCGTCAACAATTGTACTCTACTCTGTACTCTAACCTAAAACTCGACTTCAGAAGAACTCGTCAAGAAG |
| **hairy-F** | TGGTGATCAAGAAGCAGATCAAGGAGGAGGAGCAGCCCTGGCGGCCCTGGGATTATGATATTCCAACTACTG |
| **hairy-R** | AGATTCGATAGGGGTGGCTATGCTATATGATATGCATATGCAGACACCCTTCAGAAGAACTCGTCAAGAAG |
| **ems-F** | AGATGGACGAGTGTCCCAGCGATGAGGAGCACGAGCTGGACGCCAGCCACGATTATGATATTCCAACTACTG |
| **ems-R** | AGGATAAGGAACTCACGGACCGCCGCAGGACTATCTCCTGGCCGCTTCTCTCAGAAGAACTCGTCAAGAAG |
| **knirps-F** | CGGTGGCCCACAATGCGGCTAGTGCCATGAGGGGAATATTCGTGTGTGTCGATTATGATATTCCAACTACTG |
| **knirps-R** | AACGAGGGTTTTTGGGGCGACTCCTCCCACTTGGTTTTTTCGCCGTGTACTCAGAAGAACTCGTCAAGAAG |

**Supplementary Table 4. Primers for Recombineering (see Online Methods).**

**Supplementary Table 5**

| Primer name | Sequence |
|---|---|
| **bicoid_F** | TCCCTGGGAACCATTTACAC |
| **bicoid_R** | GGCGAATCGAACCAATATCA |
| **hunchback_F** | AGCGGCTTAATTGGCTTATG |
| **hunchback_R** | CCGTGCTCTACACCATTCAC |
| **engrailed_F** | TCGCTACGGATGGGTCTTAC |
| **engrailed_R** | CGGGCCAAGATCAAGAAGT |
| **giant_F** | AAGAAAGTTGAGCCCCCTAAA |
| **giant_R** | CGCATCAAGGAGGATGAGAT |
| **even-skipped_F** | CTCCACTGACCACCACCAG |
| **even-skipped_R** | CAATCTGTACAATCTTCGGGAAT |
| **runt_F** | CTACCACGAGAGCGGTCCT |
| **runt_R** | CGCTGCCGTTTATCCTTCTA |
| **caudal_F** | ATGGCGGCGATGAACATT |
| **caudal_R** | GGGAGCACTGAAGGACACAT |
| **tailless_F** | TGATGCACAAGGTCTCAAGC |
| **tailless_R** | GGCTCGACTCCTGGATATGA |
| **sloppy-paired-2_F** | ACCCATCACCACCCACAC |
| **sloppy-paired-2_R** | AAGCGTCTAGCGCCTCAGT |
| **Deformed_F** | CAATCAAATGGGTCACACGA |
| **Deformed_R** | TGCCTGAAGTTCAAGGTCATT |
| **Chip_F** | AAGAACAGAAGTATCCTACGGTTTAAT |
| **Chip_R** | CTCGAATCCGTGGAGCAT |
| **Brother_F** | TGGCTGAGTCCAGGAGAATC |
| **Brother_R** | GGAAACTCCTTGTGAAATGAGA |
| **extradenticle_F** | CACAGGATTCCATGGGCTAT |
| **extradenticle_R** | TCTTGGCGAAGAAGCAAATA |
| **Deformed_F** | ATCTGAAGTGCGCGAAGATT |
| **Deformed_R** | ACTCTACCCATCCTCGTCCA |
| **hairy_F** | ACATCAAGCCATCGGTCATC |
| **hairy_R** | GGCGAATGATGTGAAATGTCT |
| **empty-spiracles_F** | ATGCAGCAGGAGGATGAGAA |
| **empty-spiracles_R** | CCTTGGGATCGCTCTAAGAA |

**Supplementary Table 5. Primers for PCR and Sequencing.** Primers used to verify the corresponding P(acman) BAC by PCR and to sequence the tag junctions following recombineering.

## SUPPLEMENTARY DISCUSSION

The primary BAC libraries used in the *D. melanogaster* genome sequencing projects were produced from genomic DNA fragmented by partial digestion with restriction enzymes that cleave at six base-pair recognition sequences (*Eco*RI, *Hin*dIII) [1,2]. More recently, *D. melanogaster* BAC libraries have been constructed by physical shearing of genomic DNA [3]. Our goal was to produce P(acman) BAC libraries with high densities of fragment endpoints and high depth of mapped clone coverage. We therefore chose partial digestion with a restriction enzyme that cleaves at a four base-pair recognition sequence, *Mbo*I, as a compromise between the lower density of endpoints resulting from restriction digestion and the lower cloning efficiency inherent to the physical shearing approach. In addition, to produce the shorter insert CHORI-322 library, we performed partial digestion at three reaction conditions to further reduce bias in the distribution of fragment end points. A P1 library [4] and a BAC library [1] had previously been made using isoschizomers of *Mbo*I (*Sau*3AI and *Nde*II, respectively). However, no thorough analysis of the distribution of endpoints of cloned fragments in large-insert genomic libraries across the *D. melanogaster* genome has been reported. As we have shown, our strategy produced libraries containing mapped clones that span nearly all annotated *D. melanogaster* genes including their transcriptional regulatory sequences. Where no P(acman) BAC clone with mapped paired ends is available for a region of interest, the partially mapped clones for which only one end is aligned to the genome provide additional options. The location of the unmapped end of such clones can be determined by conventional techniques such as end sequencing and restriction mapping.

Previously available *D. melanogaster* BAC and P1 libraries were constructed in conventional low-copy cloning vectors. Clones in these libraries are not transformation-ready,

and plasmid copy number is not inducible, hampering the efficient isolation of DNA for embryo microinjection. Transformation-ready cosmid libraries were previously constructed for *D. melanogaster* [5,6]. However, cosmids have a limited insert size. The available cosmid libraries are not mapped onto the reference genome and do not support ΦC31 integrase-mediated germ-line transformation. Moreover, the medium-copy number backbone of cosmids interferes with some recombineering paradigms. The new P(acman) libraries do not suffer from these disadvantages. The libraries are mapped onto the reference genome. The vector supports both *P* element transposase-mediated and bacteriophage ΦC31 integrase-mediated germ-line transformation in *D. melanogaster*. The clones can be maintained at low-copy number to ensure plasmid stability and to facilitate recombineering, and can be induced to high-copy number to improve plasmid yield. Hence, the development of mapped BAC libraries in the P(acman) transgenesis system provides a versatile resource for efficient manipulation of individual genes and for genome-wide applications.

The expression patterns illustrated by the transgenic fusion constructs (**Fig. 2**) are similar to those described for the normal expression of the *eve* [7], *D* [8,9], *cad* [10], *Dfd* [11], *tll* [12], *slp2* [13], and *exd* [14] genes and proteins.

## SUPPLEMENTARY REFERENCES


1. Benos, P. V. *et al.* From first base: the sequence of the tip of the X chromosome of Drosophila melanogaster, a comparison of two sequencing strategies. *Genome Res.* **11**, 710-730 (2001).
2. Hoskins, R. A. *et al.* A BAC-based physical map of the major autosomes of Drosophila melanogaster. *Science* **287**, 2271-2274 (2000).
3. Osoegawa, K. *et al.* BAC clones generated from sheared DNA. *Genomics* **89**, 291-299 (2007).
4. Smoller, D. A., Petrov, D. & Hartl, D. L. Characterization of bacteriophage P1 library containing inserts of Drosophila DNA of 75-100 kilobase pairs. *Chromosoma* **100**, 487-494 (1991).
5. Haenlin, M., Steller, H., Pirrotta, V. & Mohier, E. A 43 kilobase cosmid P transposon rescues the fs(1)K10 morphogenetic locus and three adjacent Drosophila developmental mutants. *Cell* **40**, 827-837 (1985).
6. Tamkun, J. W. *et al.* brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**, 561-572 (1992).
7. Macdonald, P. M., Ingham, P. & Struhl, G. Isolation, structure, and expression of even-skipped: a second pair-rule gene of Drosophila containing a homeo box. *Cell* **47**, 721-734 (1986).
8. Nambu, P. A. & Nambu, J. R. The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. *Development* **122**, 3467-3475 (1996).
9. Russell, S. R., Sanchez-Soriano, N., Wright, C. R. & Ashburner, M. The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. *Development* **122**, 3669-3676 (1996).
10. Mlodzik, M. & Gehring, W. J. Expression of the caudal gene in the germ line of Drosophila: formation of an RNA and protein gradient during early embryogenesis. *Cell* **48**, 465-478 (1987).
11. Martinez-Arias, A., Ingham, P. W., Scott, M. P. & Akam, M. E. The spatial and temporal deployment of Dfd and Scr transcripts throughout development of Drosophila. *Development* **100**, 673-683 (1987).
12. Pignoni, F. *et al.* The Drosophila gene tailless is expressed at the embryonic termini and is a member of the steroid receptor superfamily. *Cell* **62**, 151-163 (1990).
13. Grossniklaus, U., Pearson, R. K. & Gehring, W. J. The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev.* **6**, 1030-1051 (1992).
14. Rauskolb, C., Peifer, M. & Wieschaus, E. extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1. *Cell* **74**, 1101-1112 (1993).