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] & Roger A Hoskins[2]

We constructed *Drosophila melanogaster* bacterial artificial chromosome libraries with 21-kilobase and 83-kilobase inserts in the P[acman] system. We mapped clones representing 12-fold coverage and encompassing more than 95% of annotated genes 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.

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 had 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 germline transformation of genomic DNA fragments up to 133 kilobases (kb)[3]. P[acman] combines a conditionally amplifiable bacterial artificial chromosome (BAC)[4], the ability to use recombineering in *E. coli* for retrieval and manipulation of DNA inserts[5] and bacteriophage ΦC31 integrase–mediated germline 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 predetermined *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 new 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 enhanced β-galactosidase activity resulting from a premature stop codon in the *α-lacZ* fragment (**Supplementary Fig. 1** online) 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** online). 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. We prepared high-molecular-weight genomic DNA from the *D. melanogaster* strain used to produce the reference genome sequence. We fragmented the DNA by partial restriction digestion, recovered 20 kb- and 80 kb size range DNA fractions and cloned them separately to produce two genomic BAC libraries. We named the libraries produced from the 20 kb and 80 kb fractions CHORI-322 and CHORI-321, respectively. We stocked 73,728 CHORI-322 clones and 36,864 CHORI-321 clones.

To map P[acman] BACs on the genome, we determined paired-end sequences and aligned them 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

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA. [2]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. [3]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas, USA. [4]Institute for Genomics and Systems Biology and Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. [5]Children's Hospital Oakland Research Institute, Oakland, California, USA. [6]Department of Neuroscience and [7]Program in Developmental Biology, Baylor College of Medicine, Houston, Texas, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to H.J.B. (hbellen@bcm.tmc.edu).

RECEIVED 19 FEBRUARY; ACCEPTED 20 APRIL; PUBLISHED ONLINE 24 MAY 2009; DOI:10.1038/NMETH.1331

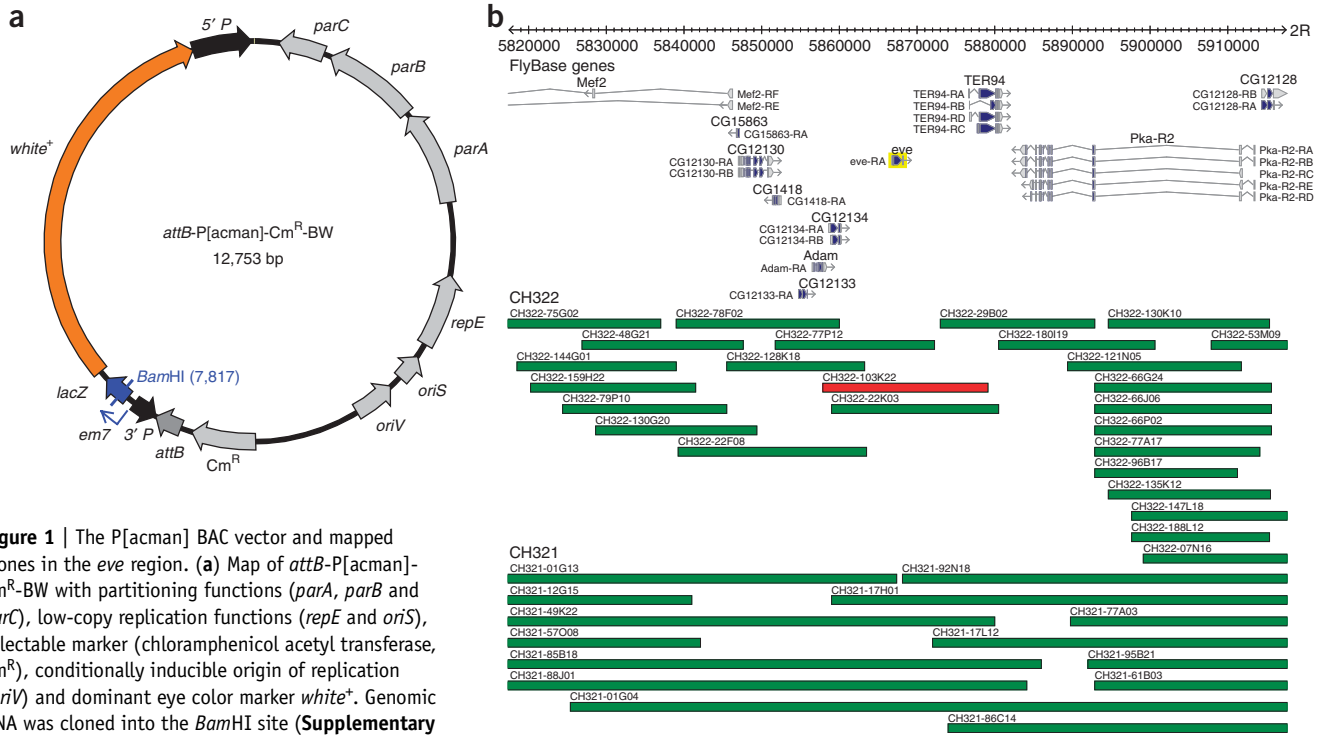

Figure 1 | The P[acman] BAC vector and mapped clones in the *eve* region. (**a**) Map of *attB*-P[acman]-Cm^R-BW with partitioning functions (*parA*, *parB* and *parC*), low-copy replication functions (*repE* and *oriS*), selectable marker (chloramphenicol acetyl transferase, Cm^R), conditionally inducible 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, is indicated in red.

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 showed that the average insert sizes of the CHORI-322 and CHORI-321 libraries were 21.0 kb (± 4.0 kb) and 83.3 kb (± 21.5 kb), respectively. We partially mapped an additional 18,767 CHORI-322 clones and 11,571 CHORI-321 clones to the genome sequence by aligning one end sequence only. The two libraries together represent deep coverage of the genome and span most annotated genes (**Supplementary Table 1** online). The mapped CHORI-322 and CHORI-321 clones spanned 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 (for example, *hh*, *vas* and *shi*) or sufficient 5′ regulatory sequence did not appear to be present in a mapped CHORI-322 clone (for example, *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*, had 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 (ref. 3), using ΦC31 integrase[6,7]. We define the transformation efficiency of each clone as the percentage of G0 fertile crosses that yielded at least one transgenic fly. We obtained 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** online). Moreover, we integrated 53 of 72 CHORI-321 clones in an independent experiment to generate 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 integrated. For all transformants, we confirmed the presence of the expected DNA fragment sizes at the integration junctions, indicative of site-specific integration at the respective docking site, by multiplex PCR that tested simultaneously for the presence of *attP*, *attB*, *attR* and *attL* sites (**Supplementary Fig. 3** online).

Integration 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 were very similar to each other, so the observed range suggested that some fragments are less efficiently transformed than others owing to sequence content or specific interference between certain fragments and docking sites (for example, *Csp* and *wg*). Notably, the high efficiency observed for some CHORI-321 clones (for example, CH321-16H04, CH321-64G01 and CH321-79N05) suggests that additional optimization of the integration efficiency of large clones is possible.

We tested transgenic insertions of 10 CHORI-322 and 6 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*, rescued lethal mutations in the corresponding genes (**Supplementary Note** online). 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* complemented lethal mutations in the corresponding genes. Rescue of *cac*, *lt* and *shakB* using genomic fragments has also not been reported previously to our knowledge. Rescue of a lethal mutation in *lt* with a 92 kb genomic fragment inserted in euchromatin was notable, because full expression of *lt* and several other heterochromatic genes has been shown to be dependent on their heterochromatic context[13]. Only 1 of 3 clones tested complemented *lt* lethality, suggesting that essential regulatory elements or sufficient genomic context were absent in the other two clones.

Table 1 | Characterization of CHORI-322 clones

Gene(s)	Clone	Insert size (bp)	*attP* VK docking site	Fertile G0 crosses	Vials resulting in transgenic fly	Integration efficiency (%)
Act42A	*12M14*	19,321	33	84	2	2.4
Act87E	*158M20*	19,317	37	65	5	7.7
αTub84B	*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
			22	11	2	18.2
			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
			31	35	5	14.3
			33	78	0	0

Genes contained in 38 CHORI-322 clones are indicated. For each clone, the deduced genomic insert length, *attP* VK docking site used, number of fertile G0 crosses, number of vials resulting in at least one transgenic fly and integration efficiency are indicated. Note that only the gene of interest is shown; most clones contain more than one gene. Clone 152E24 contains two genes of interest.

Table 2 | Characterization of CHORI-321 clones

Gene(s)	Clone	Insert size (bp)	*attP* VK docking site	Fertile G0 crosses	Vials resulting in transgenic fly	Integration efficiency (%)
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
gfa	*57O14*	77,006	33	76	0	0
hh	*61H05*	101,201	37	61	0	0
Hnf4	*12P12*	104,925	33	63	1	1.6
Hr38	*25N09*	85,274	33	107	0	0
Hr46	*23L02*	86,475	33	62	1	1.6
jar	*76B03*	92,518	37	46	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
rl	*36L01*	53,704	33	63	0	0
rl	*81D16*	102,491	33	79	0	0
shakB	*27E22*	88,712	33	76	2	2.6
shi	*71G22*	70,514	33	53	0	0
syt	*08F02*	82,244	33	59	0	0
tweek	*79N05*	76,842	33	69	8	11.6
vas	*69O09*	80,238	33	62	0	0

Genes contained in 24 CHORI-321 clones are indicated. Clone 05E14 contains two genes of interest.

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 genes encoding C-terminal protein fusions[14] (**Supplementary Fig. 4** online). We tested both the untagged and tagged constructs for integration using ΦC31 (**Supplementary Table 2**). We tested 11 tagged constructs for expression of the fusion protein. As this EGFP does not fold efficiently in embryos before stage 15, we performed immunohistochemistry analyses on embryos with an antibody to GFP (**Fig. 2** and **Supplementary Fig. 5a,b** online). 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* were reproduced by the transgenic fusion constructs (**Supplementary Discussion** online). 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]. Notably, 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 was expressed in the expected pattern. Similarly, we only observed *en* expression 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.

Here we described a versatile P[acman] BAC library resource for functional analysis of transgenes in *D. melanogaster* (**Supplementary Discussion**). We conservatively estimate that the new resource allows *in vivo* analysis of more than 95% of *D. melanogaster* genes, including large genes, gene complexes and heterochromatic genes (**Supplementary Fig. 6** online). 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].

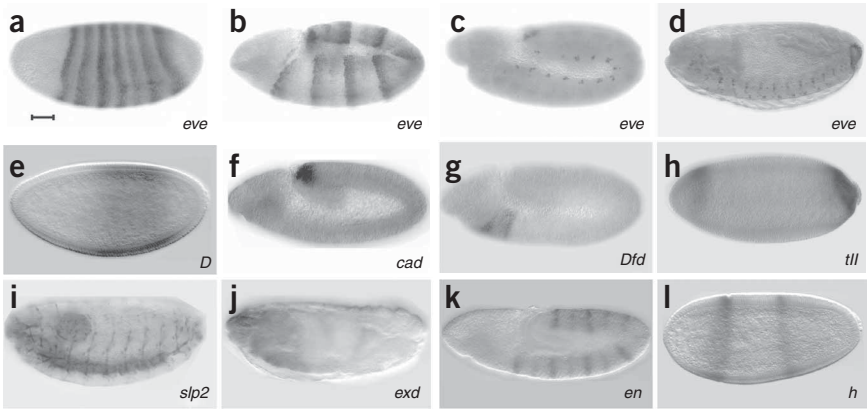

Figure 2 | Expression of EGFP fusion proteins in transgenic embryos. (**a**–**d**) Even skipped, embryonic stage 5 (**a**), embryonic stage 9 (**b**), embryonic stage 11 (**c**) and embryonic stage 15 (**d**). (**e**) Dichaete, embryonic stage 5. (**f**) Caudal, embryonic stage 9. (**g**) Deformed, embryonic stage 11. (**h**) Tailless, embryonic stage 5. (**i**) Sloppy paired 2, embryonic stage 16. (**j**) Extradenticle, embryonic stage 15. (**k**) Engrailed, embryonic stage 9. (**l**) Hairy, embryonic stage 6. Fusion proteins were detected using an antibody to GFP and peroxidase staining. Scale bar, 50 μm.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Accession codes GenBank: attB-P[acman]-CmR-BW vector sequence, FJ931533; P[acman] BAC end sequences, FI329972–FI494724.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank members of the Washington University Genome Sequencing Center for their excellent BAC end-sequencing services. We thank members of the Bloomington *Drosophila* Stock Center for providing flies, members of US National Cancer Institute (NCI) Frederick for recombineering reagents, A. Hyman (Max Planck Institute, Dresden) for the LAP-tag plasmid, R. Karess (Centre National de la Recherche Scientifique) for flies, R. Ordway (Penn State University) for flies, J. Reinitz (Stony Brook University) for antibodies, D. Schmucker (Harvard Medical School) for flies, T. Schwarz (Children's Hospital, Boston) for flies, B. Wakimoto (University of Washington) for flies, L. Zipursky (University of California Los Angeles) for flies, and J. Bischof, K. Basler (University of Zurich) and F. Karch (University of Geneva) for providing germline FC31 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, and B. Wakimoto for critical reading of the manuscript. Confocal microscopy was supported by the Baylor College of Medicine 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 US National Institutes of Health modENCODE project in collaboration with K.P.W.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Bier, E. *Nat. Rev. Genet.* **6**, 9–23 (2005).
2. Venken, K.J. & Bellen, H.J. *Development* **134**, 3571–3584 (2007).
3. Venken, K.J., He, Y., Hoskins, R.A. & Bellen, H.J. *Science* **314**, 1747–1751 (2006).
4. Wild, J., Hradecna, Z. & Szybalski, W. *Genome Res.* **12**, 1434–1444 (2002).
5. Sawitzke, J.A. *et al. Methods Enzymol.* **421**, 171–199 (2007).
6. Groth, A.C., Fish, M., Nusse, R. & Calos, M.P. *Genetics* **166**, 1775–1782 (2004).
7. Bischof, J., Maeda, R.K., Hediger, M., Karch, F. & Basler, K. *Proc. Natl. Acad. Sci. USA* **104**, 3312–3317 (2007).
8. Venken, K.J. *et al. Nucleic Acids Res.* **36**, e114 (2008).
9. Pepple, K.L. *et al. Development* **135**, 4071–4079 (2008).
10. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E. & Perrimon, N. *Nat. Genet.* **40**, 476–483 (2008).
11. Ni, J.Q. *et al. Nat. Methods* **5**, 49–51 (2008).
12. Pfeiffer, B.D. *et al. Proc. Natl. Acad. Sci. USA* **105**, 9715–9720 (2008).
13. Yasuhara, J.C. & Wakimoto, B.T. *Trends Genet.* **22**, 330–338 (2006).
14. Poser, I. *et al. Nat. Methods* **5**, 409–415 (2008).
15. Hooper, K.L., Parkhurst, S.M. & Ish-Horowicz, D. *Development* **107**, 489–504 (1989).
16. Howard, K.R. & Struhl, G. *Development* **110**, 1223–1231 (1990).
17. DiNardo, S., Kuner, J.M., Theis, J. & O'Farrell, P.H. *Cell* **43**, 59–69 (1985).
18. Giepmans, B.N., Adams, S.R., Ellisman, M.H. & Tsien, R.Y. *Science* **312**, 217–224 (2006).
19. Adams, M.D. *et al. Science* **287**, 2185–2195 (2000).
20. Hoskins, R.A. *et al. Science* **316**, 1625–1628 (2007).

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 members of NCI-Frederick)[21]. The multiple cloning site of vector *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** online). 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 members of NCI Frederick)[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 transgenesis mediated by P transposase[25] and ΦC31 integrase[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^{-1}) 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. After transformation of the two libraries into EPI300-T1R cells (Epicentre), clones with inserts (white colonies) were selected on LB plates (12.5 μg ml^{-1} chloramphenicol, 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^{-1} chloramphenicol), and plasmid copy number was induced using CopyControl 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-3BVK00037*) located on chromosome 2 and *VK33* (*PBac{y$^+$-attP-3BVK00033*) on chromosome 3 (ref. 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 (ref. 3) (~50%) and the 18 kb insert containing *sens-L* clone[3] (~30%) and receptivity for the large insert clones *Dscam-1*, *E(Spl)-C* and/or *Dscam-2* (ref. 3) (data not shown). Chromosomes bearing each docking site were brought together with the X chromosome–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*, $wg^{Gla\text{-}1}$/*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" (for example, *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 members of NCI Frederick) with modifications (**Supplementary Fig. 4**). Tags containing EGFP followed by a kanamycin resistance marker[14] were PCR-amplified from the plasmid containing a gene encoding a C-terminal localization and affinity purification (CAP) tag (gift from A. Hyman) using the recombineering primers, GOI-F and GOI-R (**Supplementary Table 4** online). 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** online). 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^{-1} chloramphenicol, 10 μg ml^{-1} tetracycline) at 32 °C. Recombineering was performed using the PCR-amplified tags. Recombinants were isolated on LB plates (with 12.5 μg ml^{-1} chloramphenicol, 10 μg ml^{-1} tetracycline and 25 μg ml^{-1} kanamycin) 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.

21. Lee, E.C. *et al. Genomics* **73**, 56–65 (2001).
22. Hobert, O. *Biotechniques* **32**, 728–730 (2002).
23. Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. & Copeland, N.G. *Nucleic Acids Res.* **33**, e36 (2005).
24. Kim, U.J. *et al. Genomics* **34**, 213–218 (1996).
25. Rubin, G.M. & Spradling, A.C. *Science* **218**, 348–353 (1982).
26. Osoegawa, K. *et al. Genomics* **52**, 1–8 (1998).
27. Brizuela, B.J., Elfring, L., Ballard, J., Tamkun, J.W. & Kennison, J.A. *Genetics* **137**, 803–813 (1994).
28. Hoskins, R.A. *et al. Science* **287**, 2271–2274 (2000).
29. Altschul, S.F. *et al. Nucleic Acids Res.* **25**, 3389–3402 (1997).
30. Stein, L.D. *et al. Genome Res.* **12**, 1599–1610 (2002.).

doi:10.1038/nmeth.1331