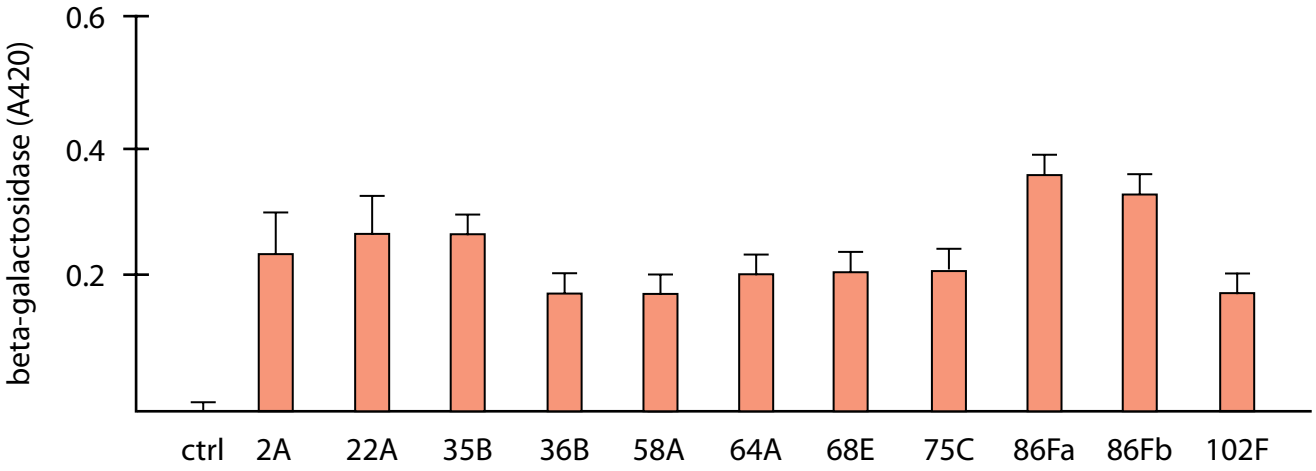


Fig. S1. Levels of induced expression at various *attP* sites. The induced β-galactosidase activity was measured from heterozygous *C765-Gal4* larvae (negative control) and from heterozygotes carrying both one copy of the *UAS-lacZ* reporter at the indicated *attP* site and one copy of the imaginal disc-specific driver *C765-Gal4*. Three pools of two dissected L3 larvae per *attP* or *C765-Gal4* line were measured. This assay was performed three times and a representative plot is shown. Among the 11 tested *attP* lines, we see at most about a twofold difference in the measured activity. Note, the line *ZH-attP-102F* is identical to line *ZH-attP-102D* (Bischof et al., 2007), but corrected for the cytological position, which is *102F* and not *102D*. This β-galactosidase assay to detect *lacZ* reporter activity was essentially carried out according to the procedure described by Viktorinová and Wimmer (Viktorinová and Wimmer, 2007). ONPG was used as a substrate.

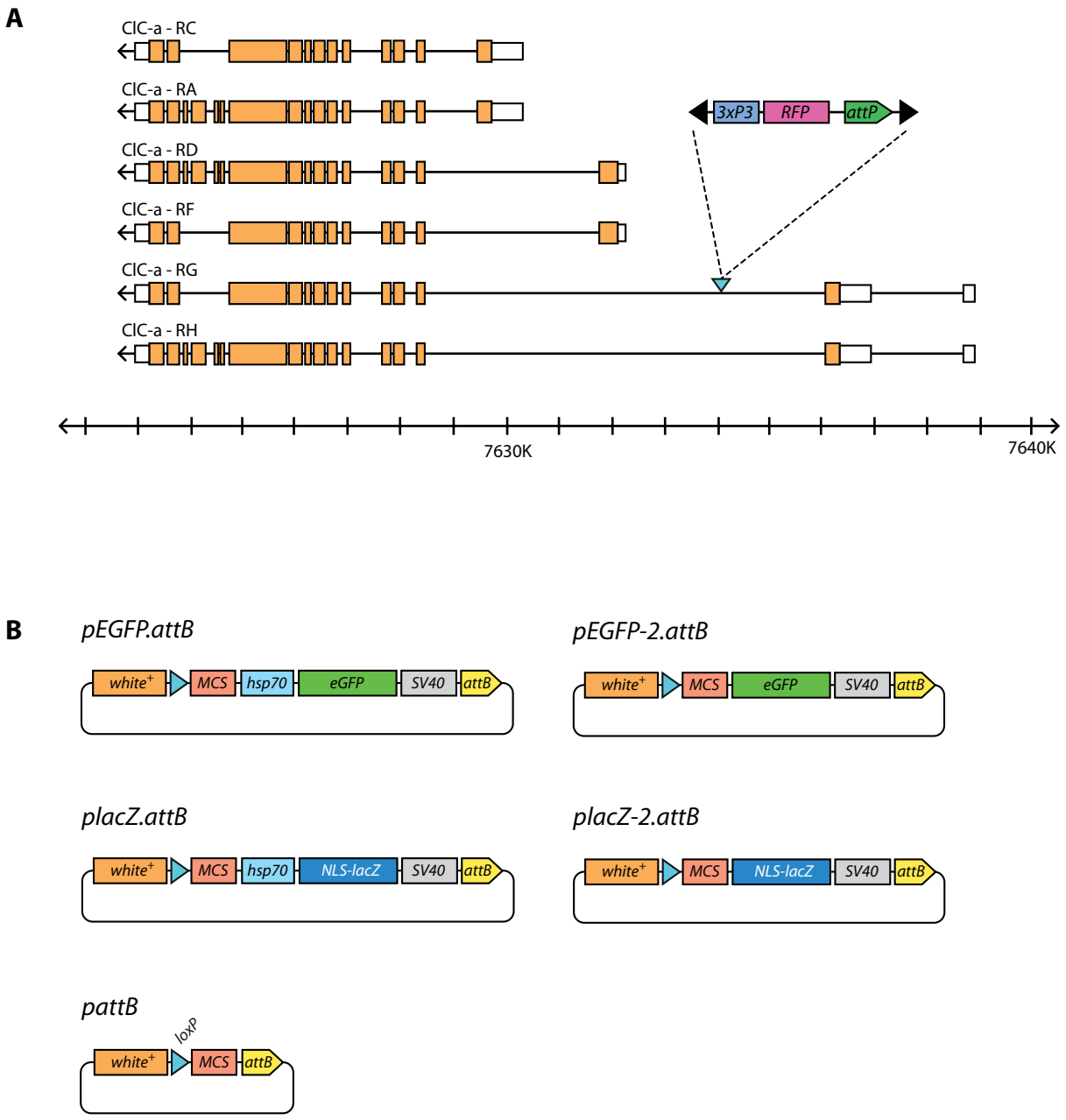


Fig. S2. Genomic region of *ZH-attP-86Fb* and ΦC31-based transformation vectors. (**A**) Genomic region at 86F containing the *ClC-a* gene with the six reported splice isoforms. The *attP* landing site construct pM{3xP3-RFPattP} is inserted in the 5′ intron of the two longer isoforms of the gene *chloride channel-a* (*ClC-a*; CG31116; orange boxes, protein coding). The insertion point is indicated (blue triangle) and the insertion plasmid is shown schematically. No loss-of-function phenotypes for this gene have been reported. (**B**) Additional transformation vectors used for ΦC31 integration. Vector pattB for genomic rescue constructs, and four reporters: placZ.attB, placZ-2.attB, pEGFP.attB, and pEGFP-2.attB. The indicated loxP site (blue triangle) can be used to flox out substantial parts of the vectors and the *ZH-attP* landing sites adjacent to the integrated vector (Bischof et al., 2007).

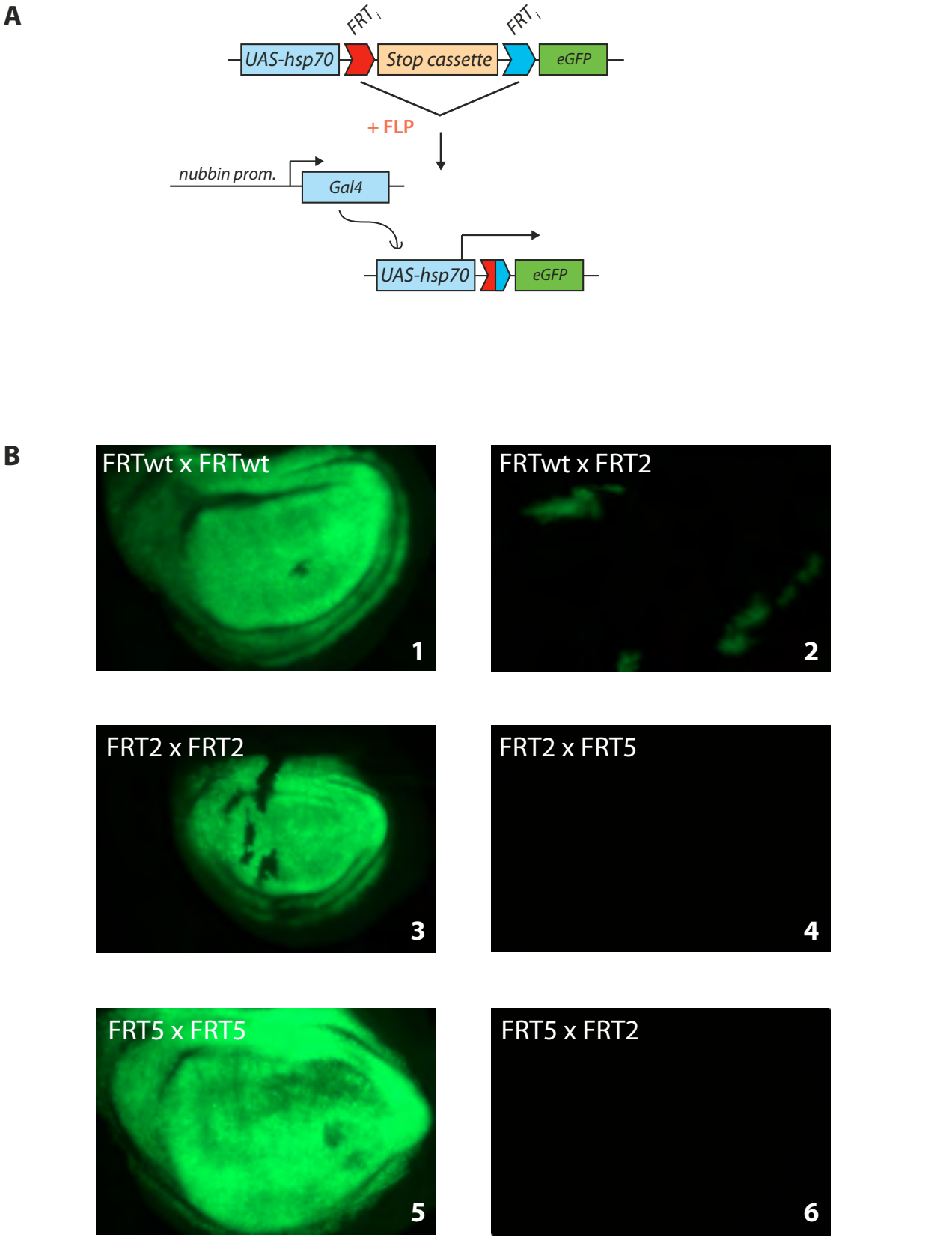


Fig. S3. Testing cross-reactivity in cis between *FRT* sites. Recombination between *FRT* variants (*FRT*$_i$) was tested in the pouch region of third instar wing discs. Cassette removal was assayed by *nub*-Gal4-driven *eGFP* expression. (**A**) Schematic of the test system used. Recombination between *FRTs* releases the stop cassette (a *hsp70* 3'UTR), resulting in *eGFP* expression in the presence of an active Gal4 driver. Expression of *eGFP* is only expected if the *FRT* sites cross-react to release the stop cassette. The *FRT* constructs were examined in the landing site *ZH-attP-86Fb*. (**B**) Strong expression of *eGFP* can be detected for homologous pairs of *FRTs* (1, 3, 5). The heterologous combination *FRTwt-FRT2* (2) shows some *eGFP* expressing clones, indicating weak cross-reactivity. The heterologous combinations between *FRT2* and *FRT5* showed no cis cross-reaction in this assay (4, 6). The *FRTwt-FRT5* combination revealed no cross-reaction at all (not shown).

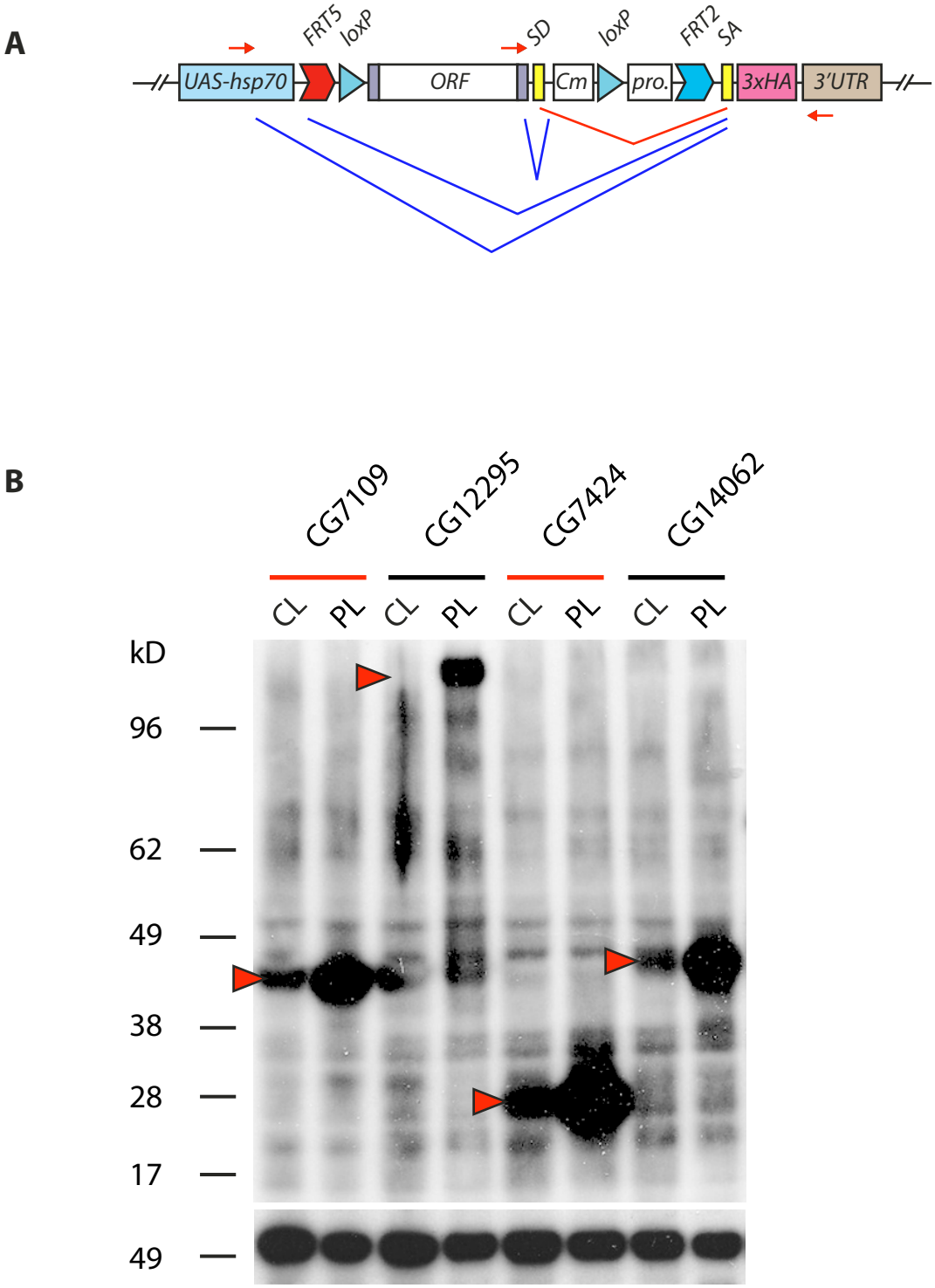
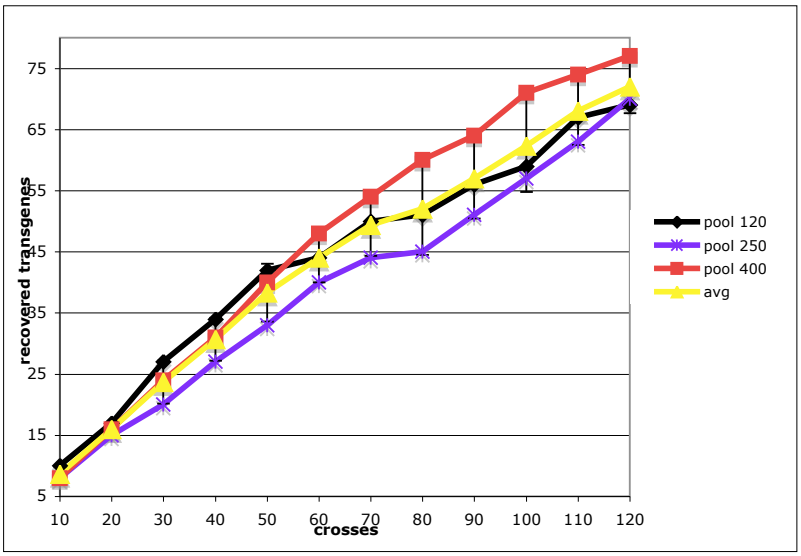


Fig. S4. Comparison of expression between direct HA lines and Creator HA lines. (**A**) Schematic of the verified splice events found in the Creator lines. Except for one case, the splicing specifically uses the acceptor site (SA). However, apart from the expected reaction between splice acceptor and donor sites (SD; indicated in red) we found various unspecific splice events (blue). *Cm*, chloramphenicol resistance; pro., promoter for *Cm*. (**B**) Western blot comparison between Creator lines (CL) and direct-tag lines (PL, pilot library). Indicated transgenes (labelled on top) were overexpressed with *C765-Gal4* driver and expression levels in wing disc samples (20 each) were compared. Creator lines displayed substantially lower amounts of the expressed protein than the directly tagged lines (indicated by red arrows).

Supp. Table S1 (A)

F1 outcrosses	10	20	30	40	50	60	70	80	90	100	110	120
pool 120	10	17	27	34	42	44	50	51	56	59	67	69
pool 250	8	15	20	27	33	40	44	45	51	57	63	70
pool 400	8	16	24	31	40	48	54	60	64	71	74	77
avg	8.7	16.0	23.7	30.7	38.3	44.0	49.3	52.0	57.0	62.3	68.0	72.0
sd	1.2	1.0	3.5	3.5	4.7	4.0	5.0	7.5	6.6	7.6	5.6	4.4
%	86.7	80.0	78.9	76.7	76.7	73.3	70.5	65.0	63.3	62.3	61.8	60.0



Supp. Table S1 (B)

crosses	10	20	30	40	50	60	70	80	90
pool 120	10	17	27	34	42	44	50	51	56
pool 120 single	10	20	28	35	41	44	51		
pool 250	8	15	20	27	33	40			
pool 250 single	9	19	24	32	38	45			
pool 400	8	16	24	31	40	48	54	60	64
pool 400b	7	14	20	26	32	38	45	48	57
pool 400 single	9	18	28	33	41	49	57	66	
pool 400b single	10	20	29	39	48	55	62	68	
avg	8.7	17.3	25.7	32.7	40.3	47.3	54.7	60.7	57.0

Table S1. Pool size versus transgene recovery. Pools containing 120-400 ORFs were injected, followed by outcrosses and determination of recovered transgenes. (**A**) On top is indicated the number of F1 outcrosses after injection (here from 10 to 120). Generally, we made two outcrosses from a positively scored F1 vial, as we observed that we can acquire multiple different ORF construct integrations in an injected embryo, eventually giving rise to multiple independent transgenic flies. In the three rows indicating pool sizes 120-400, the number of recovered transgenes is given. avg, average number of newly identified transgenes per number of outcrosses; sd, standard deviation; %, average percentage of recovered transgenes with respect to the indicated number of performed outcrosses. For three pools the recovery rate of different transgenes is plotted against the number of F1 outcrosses up to 120 outcrosses. With 10-20 outcrosses, recovery rates of 80% and more are achieved (see row with % value above). This rate slowly drops with the number of outcrosses. (**B**) Some more examples of different pool sizes, subsequent F1 outcrosses, and the averaged number of newly identified transgenes. (From pools indicated as 'single' only one male outcross per positively scored F1 vial was carried out.)

Table S2. List of untagged and tagged strains created. The table lists 547 untagged and the 602 HA-tagged lines with their CG numbers and gene names/symbols (PL, pilot library).

Download Table S2

Appendix S1. Required background information to the evaluation of the Creator cloning system

The Berkeley *Drosophila* Genome Project (BDGP) has generated thousands of sequence-verified ORFs, either with a native stop codon or without for C-terminal tagging (Yu et al., 2011). These ORFs are cloned into the donor vector pDNR-Dual (Clontech), from which the ORFs can be moved into appropriate destination vectors by a Cre/lox recombination reaction. The pDNR-Dual vector contains a splice donor (SD) site that is placed downstream of the ORF insertion site and, for carboxy-tagged ORFs; this SD site is placed upstream of a splice acceptor (SA) site after shuttling into the final destination vector. When expressed, the ORF fuses to the provided 3′ tag in eukaryotes by a splicing reaction between SD and SA. This cloning strategy is also known as the Creator System (for a visualisation see http://www.fruitfly.org/EST/proteomics.shtml).

Because of the huge number of high-quality clones offered by the BDGP and the inherent versatility of the system we wanted to make use of this resource. To evaluate its suitability for the ORFeome library we tested this system with a set of transgenic lines in vivo. We generated a slightly modified pDNR-Dual vector, i.e. we introduced a Gateway cassette into the multiple cloning cassette, thus allowing cloning of ORFs by Gateway. We further generated a UAS-acceptor vector containing all the required Creator system elements with a 3xHA tag as the C-terminal epitope. This UAS-expression vector is additionally equipped with a shortened *FRT* (FLP recognition target) sequence and an *attB* site for site-specific integration. We generated 41 different transgenic UAS-ORF 'Creator lines' and assayed them with both *ey-Gal4* and *MS1096-Gal4* drivers for overexpression phenotypes (at 25°C and 29°C). Only one Creator line out of 41 showed a phenotype (weak dorsalisation) with *MS1096-Gal4*. From the 41 Creator ORFs we already had 38 as 3xHA directly tagged lines in our pilot library. Of these 38 lines, 11 gave a phenotype in the *ey-Gal4* and *MS1096-Gal4* assays, including the ORF that led to a dorsalised phenotype with the Creator system. In summary, the Creator system cannot recapitulate the phenotypes generated with the direct fusion construct.

To test whether the Creator system leads to unspecific splicing we isolated mRNA from several test crosses, transcribed it to cDNA, PCR-amplified stretches between the promoter and the 3′UTR of the Creator vector, and sequenced them. This analysis confirmed the presence of both specific and unspecific splicing. We found variable splicing events, often taking out the whole sequence between the SA site and the *hsp70* promoter region, thus eliminating the ORFs completely (see supplementary material Fig. S4A). To further substantiate this result, we compared protein expression by

western blot from direct-tag and Creator-tag lines of the same ORFs when overexpressed with the imaginal disc specific *C765-Gal4* driver. The Creator lines showed substantially lower expression levels (see supplementary material Fig. S4B) indicating that the demonstrated erroneous splicing accounts for the reduced levels of expected protein products and thus results in the absence of misexpression phenotypes. One case resulted in no detectable expression.

The previously mentioned *FRT* site (a shortened 36-bp variant) was introduced upstream of the SA site in our destination/expression vector and we cannot absolutely exclude that this influences correct splicing. However, there is sufficient distance between the *FRT* and the exactly defined acceptor splicing cassette to not expect interference with the acceptor site. This assumption is supported by the fact that most of the PCR-confirmed mis-splicings correctly used the SA site, but not the provided SD site (one event did not use either site). In summary, the splicing machinery in our Creator cloning system acted unspecifically. Most importantly, essentially no phenotypes were detected with this system in overexpression experiments. As in vivo experiments and macroscopically observable phenotypes are the ultimate quality control for an expression system, we therefore abandoned this strategy.