A toolkit for high-throughput, cross-species gene engineering in *Drosophila*

Radoslaw K Ejsmont, Mihail Sarov, Sylke Winkler, Kamil A Lipinski & Pavel Tomancak

We generated two complementary genomic fosmid libraries for *Drosophila melanogaster* and *Drosophila pseudoobscura* that permit seamless modification of large genomic clones by high-throughput recombineering and direct transgenesis. The fosmid transgenes recapitulated endogenous gene expression patterns. These libraries, in combination with recombineering technology, will be useful to rescue mutant phenotypes, allow imaging of gene products in living flies and enable systematic analysis and manipulation of gene activity across species.

Complete genome sequences of model organisms together with libraries of molecular reagents enable systematic studies of gene function in cell biology and development. In *Drosophila melanogaster*, systematic approaches increasingly complement traditional random mutagenesis screens. Two technical advances dramatically expand the possibilities for manipulation of the *Drosophila* genome. First, the P[acman] system using φC31-mediated site-specific transgenesis[1] is capable of introducing large, bacterial artificial chromosome–sized, constructs into *Drosophila* genome, enabling functional analysis of any gene in the context of its intact *cis*-regulatory neighborhood[2]. Second, recombineering techniques are optimized for routine, high-throughput manipulation of large genomic clones by homologous recombination[3–5].

We combined both advances to create a flexible toolkit for high-throughput manipulation of drosophilid genomes. The centerpieces of our system are genomic fosmid libraries that cover the genomes of

D. melanogaster and *D. pseudoobscura*. We developed a fosmid vector (pFlyFos) containing the *attB* sequence recognized by φC31 integrase and the eye promoter–driven dominant selectable marker (*dsRed*) for isolation of transformants in a broad range of insect species[6] (**Fig. 1a**).

We constructed genomic fosmid libraries by ligating sheared genomic DNA isolated from *D. melanogaster* and *D. pseudoobscura* embryos into the pFlyFos vector between the *attB* and *dsRed* sequences. We isolated individual clones and mapped the inserts by BLAST analysis of the end-sequence reads. For the *D. melanogaster* library we mapped 88% of the clones; for 8.25% of the clones one of the sequencing reactions failed, and 3.65% of the clones are likely chimeric (**Fig. 1b**). The average size of the inserts was 36 kilobases (kb), with 95% of clone sizes between 20 kb and 50 kb (**Fig. 1c**). The library covered all chromosomes including heterochromatin (**Supplementary Fig. 1** online) with overall genomic sequence coverage of 3.3-fold. Among the 15,204 *D. melanogaster* clones we found clones for 89.27% of the annotated genes that contain at least 10 kb of noncoding sequence upstream and 5 kb downstream of the gene model (**Fig. 1d**). Seven percent of the genes in the *D. melanogaster* genome have gene models larger than the maximum size compatible with fosmid
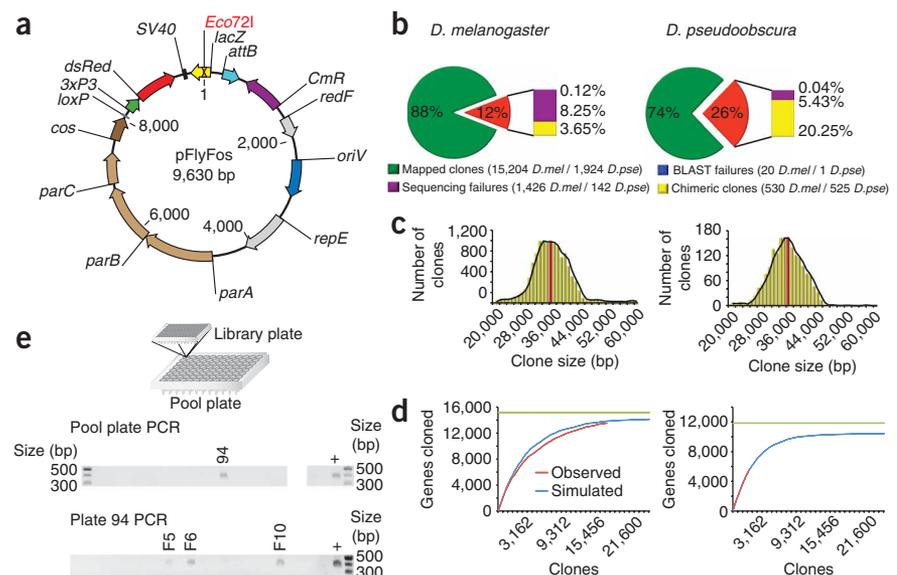
Figure 1 | Genomic fosmid libraries. (**a**) Schematic of the pFlyFos backbone containing inducible *oriV*, the *attB* sequence for φC31-mediated gene integration and eye promoter–driven *dsRed* selectable marker. (**b**) Mapping statistics for the *D. melanogaster* and *D. pseudoobscura* fosmid libraries. (**c**) Clone size distribution. (**d**) Comparison of observed and simulated gene coverage (green line shows total number of genes in the genome). (**e**) PCR-pool-screening strategy for identification of unsequenced clones in *D. pseudoobscura* library. F6 is a true, sequence-verified, *Mical*-positive clone (F5 is a contaminant from F6), and F10 is a false positive. "+", positive controls.
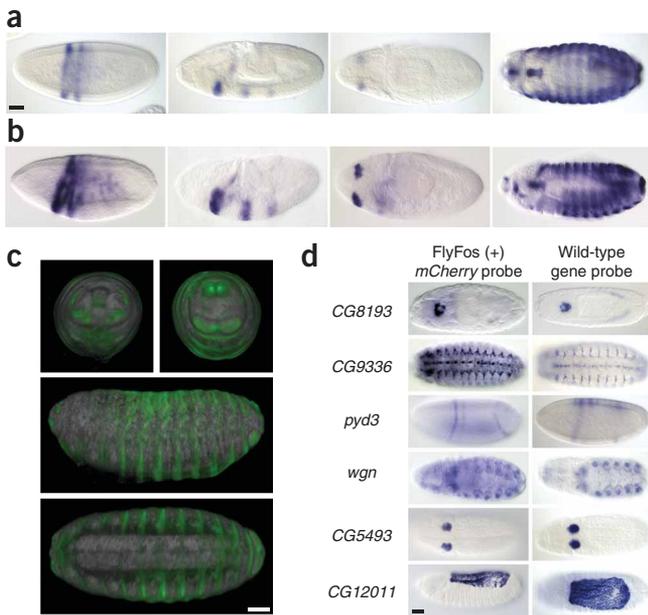
Figure 2 | Tagged fosmid clones recapitulate wild-type gene expression patterns. (**a,b**) Expression pattern of *CG4702* in wild-type embryos visualized by anti-*CG4702* probe (**a**) compared to transgene expression visualized by antisense *GFP* RNA probe in pFlyFos-*CG4702* embryos of similar age (**b**). (**c**) A three-dimensional rendering of a late-stage embryo probed with anti-GFP antibody and DRAQ5 nuclear marker imaged with a single plane illumination microscope[8] (shown are the frontal (top left), caudal (top right), lateral (middle) and ventral (bottom) views of the same embryo). (**d**) Examples of RNA *in situ* patterns recapitulated by the fosmid transgenes tagged with *mCherry*. Scale bars, 50 μm.

packaging. Thus, the library covers 96% of the genes that can in principle be included in our fosmids.

We sequenced a subset of the *D. pseudoobscura* fosmid clones (2,592 clones containing 37% of annotated *D. pseudoobscura* genes; **Fig. 1b**) and developed a pooling scheme for rapid PCR screening for clones containing any gene of interest (**Fig. 1e**). To validate the pooling approach, we screened pool plates with a primer pair (**Supplementary Table 1** online) targeting *D. pseudoobscura* ortholog of *D. melanogaster* gene *Mical* to identify the library plates harboring a potential *Mical* fosmid clone. With a second round of PCR on the single, *Mical*-positive library plate, we identified candidate clones that were end-sequenced and mapped to the *D. pseudoobscura* genome to reveal that one of them, indeed, includes the *Mical* locus (data not shown). By combining the initial high return sequencing with pool screening we circumvent the costly process of complete library mapping.

To introduce modifications, such as fluorescent tags, into the *D. melanogaster* FlyFos clones we adapted an established liquid culture recombineering pipeline for use with the FlyFos library[3,4]. As FlyFos clones already contain sequences necessary for transgenesis, tagging is achieved in a single step and the pipeline scales to a high-throughput 96-well format (**Supplementary Fig. 2a** online).

We designed three tagging cassettes (**Supplementary Fig. 3** online) and selected 48 tissue-specific marker genes covering most organ systems in *Drosophila* embryogenesis (**Supplementary Fig. 4** online). We identified suitable fosmids for all 48 *loci* and tagged them with the three tagging cassettes in 96-well format liquid culture recombineering reactions. We verified the structure of the modified fosmids by two sequencing reactions with primers from within the tag extending toward the tag-fosmid junction (**Supplementary Fig. 5** and **Supplementary Tables 2–4** online). Our success rate improved from 75% to 100% through three tagging experiments, suggesting that the method is easily scalable for genome-wide application (**Supplementary Fig. 2b**).

We transformed 12 fosmids into *D. melanogaster* to test whether the recombinant transgenes recapitulate wild-type gene expression patterns. The fosmid-driven expression of

GFP-tagged transgene was consistent with the wild-type pattern throughout embryogenesis (**Fig. 2a,b**) and permitted the detection of GFP by immunofluorescence analysis (**Fig. 2c**). We detected the GFP fluorescence only in the late-stage living embryo, possibly owing to the slow maturation time of the GFP reporter (**Supplementary Fig. 6** and **Supplementary Video 1** online). Eight out of twelve transgenes recapitulated the wild-type gene expression patterns suggesting that *cis*-regulatory elements required for control of gene expression were included (**Fig. 2d**). One tagged transgene showed no expression, and three were detected in the wrong tissues possibly because of the absence of insulator elements (data not shown). Furthermore, the tagged genes in the fosmids were biologically active as they could rescue mutant phenotypes, such as, for example, the oogenesis phenotype of *fat2* (I. Viktorinova *et al.*; unpublished data).

We described here an efficient method to produce genomic fosmid libraries that enable cross-species transgenesis and efficient clone manipulation by recombineering in liquid culture. The clones can be used directly for *in vivo* rescue experiments or can be recombinantly tagged to systematically generate HRP protein fusions, GAL4 enhancer traps or to introduce tags for *in vivo* ChIP-chip (chromatin immunoprecipitation–microarray) experiments. In particular, deletions are easily introduced into the clones by recombineering enabling a new 'top-down' approach to understand the role of *cis*-regulatory elements in the context of intact genomic neighborhoods.

As *D. pseudoobscura* cDNA sequences differ for many genes substantially from *D. melanogaster*, the *D. pseudoobscura* fosmids may be able to rescue RNA interference (RNAi)–induced gene knockdowns in *D. melanogaster in vivo*. Moreover, the ability to introduce targeted mutations, including large insertion and deletions, permits detailed structure-function analyses using RNAi-resistant mutant transgenes[7]. For evolution-development studies, tagged *D. pseudoobscura* clones can be introduced to *D. melanogaster* genetic background to assess the relative contribution of *cis*-regulatory sequences and *trans*-acting factors to expression pattern divergence.

Both FlyFos libraries are available to the *Drosophila* research community (http://transgeneome.mpi-cbg.de/). We plan to generate a large resource of tagged fosmid clones and invite the *Drosophila* researchers to participate in an online voting system to prioritize genes of interest. We believe that only a community effort can lead to the establishment of a genome-wide set of tagged fosmid reagents for the analysis of gene function in *Drosophila*.

METHODS

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

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

AUTHOR CONTRIBUTIONS
R.K.E. did the majority of the work presented in this manuscript. M.S. performed the high-throughput recombineering experiments. S.W. set up the fosmid sequencing protocol. K.A.L. generated the *D. pseudoobscura* library. P.T. conceived the whole project and wrote the manuscript.

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

1. Groth, A.C., Fish, M., Nusse, R. & Calos, M.P. *Genetics* **166**, 1775–1782 (2004).
2. Venken, K.J., He, Y., Hoskins, R.A. & Bellen, H.J. *Science* **314**, 1747–1751 (2006).
3. Sarov, M. *et al. Nat. Methods* **3**, 839–844 (2006).
4. Poser, I. *et al. Nat. Methods* **5**, 409–415 (2008).
5. Venken, K.J. *et al. Nucleic Acids Res.* **36**, e114 (2008).
6. Berghammer, A.J., Klingler, M. & Wimmer, E.A. *Nature* **402**, 370–371 (1999).
7. Bird, A.W. & Hyman, A.A. *J. Cell Biol.* **182**, 289–300 (2008).
8. Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J. & Stelzer, E.H. *Science* **305**, 1007–1009 (2004).

ONLINE METHODS

Fosmid backbone and validation of fly transgenesis. We developed a fosmid vector (pFlyFos) containing three features for efficient recombineering and fly transgenesis. An inducible origin of replication (*oriV*) ensures that the fosmid is maintained as a single copy during recombineering, but at a high copy number for isolation of the fosmid DNA for fly transformation. The *attB* sequence recognized by φC31 integrase is used for site-specific integration of the fosmid into the *Drosophila* genome. Finally, the *dsRed* driven by an eye promoter with broad species range functions as a dominant selectable marker for isolation of transformants in any species of insects.

The FlyFos vector used to prepare the genomic libraries was derived from pCC2fos (Epicentre). A 2.2 kb fragment containing *3xP3-dsRed-SV40* (from pSL-*FRT-EYFP-linotte-FRT3-3xP3-dsRed*[9]), *lacZ* (from pCC2fos) and *attB* (from *attB*-P[acman]-*CmR*-F-2[10]) was synthesized and cloned into *Apa*LI and *Sfi*I–digested pCC2fos. As the *3×P3* promoter drives expression in the eye of many insect species[11], it is likely that pFlyFos clones will be useful in any fly species carrying an *attP* landing site. We transformed the empty vector into a *Drosophila attp2* landing line carrying the φC31 integrase transgene under the control of a *nanos* promoter and obtained G1 transformants expressing easily observable levels of dsRed in the eyes in 10.3% of fertile G1 crosses. We introduced the empty pFlyFos vector into several landing lines marked with *yellow* gene and carrying vasdφC31 integrase marked with *3xP3-GFP* and *3xP3-dsRed* on the X chromosome. Only the male progeny of G0 and G1 generations could be used to screen for transformants due to the presence of *3xP3-dsRed* marker in the FlyFos clones. The transformation efficiencies with empty vector were as follows (dφC31 9727 Vk07, 37%; dφC31 9741 Vk23, 25%; dφC31 9748 Vk31, 18%; dφC31 9751 Vk36, 0%; dφC31 9738 Vk20, 41%). The transformation efficiencies with ∼45 kb large fosmid clone were comparatively poor (dφC31 9727 Vk07,13%; dφC31 9741 Vk23, 11%; dφC31 9748 Vk31, 17%; dφC31 9751 Vk36, 0%; dφC31 9738 Vk20, 0%). Due to the constraint to male lineage and due to the poor viability of many of the lines after injection we do not recommend the vasdφC31 integrase source for pFlyFos transgenesis.

Library construction. The pFlyFos vector was digested with *Eco*72I, dephosphorylated and gel-purified for library production. Both *D. melanogaster* and *D. pseudoobscura* libraries were constructed from sequenced strains (for *D. melanogaster* Berkeley *Drosophila* Genome Project Bloomington stock number 2057 and for *D. pseudoobscura Drosophila* Genome Project at the Baylor College of Medicine *Drosophila* Species Stock Center stock number 14011-0121.94). High molecular weight genomic DNA (gDNA) from both strains was isolated as described (**Supplementary Protocol 1** online) and sheared using a HydroShear device (Gene Machines) with speed-code 17 and 40 kb shearing assembly. We verified the size distribution of the gDNA by pulse-field gel electrophoresis and found it suitable for direct library production without further size selection. Sheared gDNA was end-repaired, ligated into pFlyFos and packaged following the EpiCentre Fosmid Library Production protocol. Phage particles were used to infect T1-resistant EPI300 cells of *E. coli* yielding approximately 100,000 clones for *D. melanogaster* and 20,000 clones for *D. pseudoobscura*.

The library of fosmid containing *E. coli* was plated on Luria agar plates with 12.5 µg ml^{-1} of choramphenicol and individual clones were manually arrayed into 96-well plates (REMP) with 200 µl of LB medium with 25 µg ml^{-1} of Chloramphenicol. The master library plate from REMP has the unique property that individual wells can be punched out of the plate and retrieved individually. The plates are housed in a robotic freezer system that enables automatic retrieval of any subset of the clones without error prone human intervention. To date, we have arrayed 17,280 clones for *D. melanogaster* and 9,600 for *D. pseudoobscura*.

The bacteria in REMP plates were cultured overnight at 37 °C with vigorous shaking. The primary culture was used to inoculate 384-well sequencing culture (arrayed into 4 quadrants from 4 library plates) and for *D. pseudoobscura* library also the 384-well pooling cultures. Next, 90 µl of 50% glycerol was added to the primary culture and 3 × 40 µl aliquots were distributed into 384-well backup plates stored at –80 °C. Induction cultures were incubated overnight at 37 °C in 100 µl LB medium with 25 µg ml^{-1} of chloramphenicol and 0.1% arabinose and used for high-throughput minipreps to isolate fosmid DNA for end sequencing.

Sequencing and mapping. Induced cultures were used for fosmid DNA isolation for sequencing. We downscaled Sambrook and Russell's alkaline lysis protocol to volumes of 15 µl P1, P2 and P3, 25 µl isopropanol and 75 µl ethanol without a phenol-chlorophorm extraction step. All steps were performed in 384-well format using Beckman-Coulter BiomekFX robot. Lysates were clarified by centrifugation performed at 6,000*g* for 45 min, 6,000*g* for 45 min for DNA precipitation and 6,000*g* for 15 min for DNA wash with 70% ethanol. Purified DNA was dissolved in 20 µl of nuclease-free water.

We used isolated fosmid DNA for end-sequencing with CC2_fwd and CC2_rev primers (EpiCentre) in a 10 µl 384-well sequencing reaction (5 µl template and 5 µl sequencing mix). The overall success rate of sequencing reactions from the high-throughput minipreps was 91.75% (1,426 out of 17,280 failed) permitting us to map 89.27% of the clones. Sequencing reads were BLASTed against genomic sequences to obtain exact locations of the respective clone inserts. As the fosmid packaging step constrains the size of the clones in the library, we considered clones with end-sequences mapped further then 60 kb apart chimeric (3.65%). We were unable to map clones for which one or both sequencing reactions failed (8.25%) or the sequencing read fell within repetitive DNA (0.1%).

We performed a computer simulation of genome shearing to estimate the number of clones one has to sequence and map to achieve inclusion of the majority of genes that can be cloned in a fosmid vector. We considered a gene cloned if at least 10 kb of noncoding DNA upstream and 5 kb downstream is included in the clone. Seven percent of the genes in the *Drosophila* genome have gene models larger than the maximum size of the fosmid clone under these constraints. For the remaining 93% of the genes we estimated by simulation that picking ∼18,000 clones randomly will result in coverage of more then 96% of the genes. The results of mapping of actual clones closely followed the simulation data, suggesting that given the genome size and approximate annotation we can predict the number of clones required to generate a high-coverage fosmid library.

Pooling strategy for clone identification in *D. pseudoobscura* library. We established a PCR pooling strategy to identify clones of interest in the *D. pseudoobscura* without mapping data. We have pooled DNA from each quarter of 384-well library plate (corresponding to 96-well library plate) into a single well in the 96-well pool plate. The DNA from this plate was then used as a template for PCR with gene-specific primers to identify the library plate containing the gene of interest. In the next step, DNA from the library plate was used for a second round of PCR to identify exact well positions of the clone. The primers used to identify gene of interest (*Mical*) were: dpse/MICAL_fwd (tatttcctaaagcaaaaatccc aca) and dpse/MICAL_rev (tacccggacagtgcattatactttt).

Recombineering. We adapted the *C. elegans* liquid culture recombineering pipeline described previously, to a 96-well format[12,13]. As our fosmid vector already contained a fly selectable marker, we skipped the second recombineering step, in which the selectable marker was introduced in the worm pipeline. A detailed protocol for recombineering is presented in the **Supplementary Protocol 2** online.

In short, the fosmid clone of interest was transformed with pRedFlp4 plasmid that carried inducible *recA* operon rendering the fosmid strain transiently homologous recombination competent after induction with L-rhamnose. Next, a PCR product carrying 50 bp homology arms surrounding the tagging cassette and *FRT* flanked positive selectable marker, kanamycin-resistance gene, is electroporated into the cells and the cells were grown in the presence of kanamycin, chloramphenicol and hygromycin (the latter two antibiotics positively select for the fosmid and the pRedFlp4 plasmid). It has been shown that only clones that underwent recombination between the homology arms and the homologous sequences on the fosmid are able to grow efficiently under such conditions[14]. Finally, the flippase operon residing on the pRedFlp4 plasmid is induced with anhydrotetracycline and the kanamycin-resistance gene is removed from the fosmid leaving behind a *FRT* scar sequence that is stop codon free. The flippout reaction is efficient and therefore does not require selection.

We developed an automated algorithm for selecting the best clone containing the gene to be tagged. We first picked all the clones that contain the complete gene model of interest including at least 2.5 kb of up- and downstream noncoding region. Next, we prioritized clones containing (in sorting order) either both, upstream only or downstream only neighboring genes. Finally we have ordered clones by a score *s* calculated using the formula listed below:

$$s = 10 \times \left(\log_2 \left(\frac{u}{d} \right) \right)^2 + \frac{(u+d)^2}{500 \times f^2} + \frac{50 \times f^2}{(u+d)^2},$$

where *u* is the length of the upstream sequence, *d* is the length of the downstream sequence, and *f* the length of the gene.

This scoring formula assigns the highest score to clones where the gene has more upstream than downstream sequence and where the clone size is proportional to the size of a given gene. For the tagging sites, we chose the start codon that contributes to the most protein isoforms (N-terminal tagging) or the stop codon of the largest protein isoform (C-terminal tagging). We identified suitable fosmids for 48 genes and designed primer pairs with homology arms targeting the C and N termini of these genes with the appropriate tagging cassette.

We selected genes for tagging experiments based on an annotated embryonic gene expression pattern in the Atlas of Patterns of Gene Expression (APOGEE)[15] and the availability of a suitable fosmid clone. The annotated patterns were extracted from the APOGEE database, collapsed to the level of organ systems and temporally grouped into three ranges covering the early (blastoderm), mid (stages 4–10) and late embryogenesis (stages 11–16). The patterns were organized by hierarchical clustering and cross-referenced with the database of FlyFos clones. Forty-eight genes that are included in fosmid clones and together cover most of the organ systems in the annotation hierarchy were selected for tagging (**Supplementary Fig. 3**).

Fosmid isolation and transgenesis. The DNA used for fly transgenesis was purified using FlyFos Clone Amplification protocol provided in **Supplementary Protocol 3** online. Purified fosmid DNA was injected into the *attP40* landing line[16] in case of 11 *mCherry*-tagged transgenes and into *attp2* landing site for the GFP-tagged *CG4702* gene[17]. In both cases, *nanos* promoter driven φC31 transgene on the X chromosome was used as a source of the integrase[18]. The transgenesis for GFP tagged *CG4702* was performed in-house. The remaining transgenic flies were generated by Genetic Services (http://www.geneticservices.com/). **Balanced stocks were established from flies exhibiting dsRed expression in the eyes and used for embryo collection, staining and imaging.**

RNA *in situ*. We generated *ubi-mCherry* probe template from tagged fosmid DNA by PCR using primers ubi_fwd (agaggct gatctttgctgg) and T7-mCherry_rev (taatacgactcactatagggatgtt gtctttcgaatttggc). The GFP probe was generated using primers EGFP_fwd (cataccaatcaggacccgctg) and T7-EGFP_rev (taatacgactc actatagggtcgtcgtcatccttgtagtcaatg). Probe production and *in situ* staining were performed as described[19].

Imaging. RNA *in situ* specimens were imaged as described previously[19]. For imaging of GFP antibody–stained specimens and live imaging of GFP fluorescence we used a prototype of a Single Plane Illumination Microscope (SPIM) developed by Zeiss[20].

In short, the fixed or living embryo specimens were mixed with low-melting-point agarose and loaded into a glass capillary using capillary forces. The capillary was mounted vertically in a SPIM sample holder and immersed in water filled SPIM sample chamber. The agarose was pushed out of the capillary using a plunger, so that the agarose with the specimen was facing the ×20 lens that seals one wall of the SPIM sample chamber. The fluorophores in the samples were excited using a thin laser light sheet (488 nm for GFP and 633 nm for DRAQ5 nuclear stain) entering the specimen at a 90° angle with respect to the axis of the lens and achieving optical sectioning. The embryo was moved through the light sheet at 2-μm steps resulting in a stack of acquisitions analogous to a confocal stack. The specimen was rotated and imaged from multiple angles (12 angles for the fixed sample and 6 angles for the live imaging) resulting in so-called mutliview SPIM acquisition. Fluorescent beads were included in the mounting medium to facilitate the registration of the multi-view SPIM data. The final isotropic

image of the specimen was reconstructed using a bead-based SPIM registration algorithm followed by entropy-based image fusion[21]. The three-dimensional rendering of the isotropic SPIM views was done using BioImageXD software package.

9. Horn, C. & Wimmer, E.A. *Dev. Genes Evol.* **210**, 630–637 (2000).
10. Venken, K.J. *et al. Science* **314**, 1747–1751 (2006).
11. Berghammer, A.J., Klingler, M. & Wimmer, E.A. *Nature* **402**, 370–371 (1999).
12. Sarov, M. *et al. Nat. Methods* **3**, 839–844 (2006).
13. Poser, I. *et al. Nat. Methods* **5**, 409–415 (2008).
14. Zhang, Y. *et al. Nat. Genet.* **20**, 123–128 (1998).
15. Tomancak, P. *et al. Genome Biol.* **8**, R145 (2007).
16. Markstein, M. *et al. Nat. Genet.* **40**, 476–483 (2008).
17. Groth, A.C. *et al. Genetics* **166**, 1775–1782 (2004).
18. Bischof, J. *et al. Proc. Natl. Acad. Sci. USA* **104**, 3312–3317 (2007).
19. Tomancak, P. *et al. Genome Biol.* **3**, 0088 (2002).
20. Huisken, J. *et al.* Rererer. *Science* **305**, 1007–1009 (2004).
21. Preibisch, S. *et al.* In *SPIE Medical Imaging 2009: Image Processing* (J.P.W. Pluim & B.M. Dawant, eds.) **7259**, 72592S (2009).