Generation of a transgenic ORFeome library in Drosophila

Bischof, Johannes; Sheils, Emma M.; Bjorklund, Mikael; Basler, Konrad

Published in:
Nature Protocols

DOI:
10.1038/nprot.2014.105

Publication date:
2014

Document Version
Peer reviewed version

Generation of a transgenic ORFeome library in *Drosophila*

Johannes Bischof[1], Emma M. Sheils[2], Mikael Björklund[2] & Konrad Basler[1]

[1]Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[2]College of Life Sciences, Division of Cell and Developmental Biology, University of Dundee, DD1 5EH, Scotland, UK

Correspondence should be addressed to M.B. (M.Bjorklund@dundee.ac.uk) or K.B. (konrad.basler@imls.uzh.ch).

ABSTRACT

Overexpression screens can be used to explore gene function in *Drosophila melanogaster*, but to demonstrate their full potential comprehensive and systematic collections of fly strains are required. Here we provide a protocol for high-throughput cloning of *Drosophila* open reading frames (ORFs) regulated by Upstream Activation Sequences (UAS sites); the resulting Gal4-inducible UAS-ORF plasmid library is then used to generate *Drosophila* strains by ΦC31 integrase-mediated site-specific integration. We also provide details for FLP/*FRT*-mediated *in vivo* exchange of epitope tags (or regulatory regions) in the ORF library strains, which further extends their potential applications. These transgenic UAS-ORF strains are a useful resource to complement and validate genetic experiments performed with loss-of-function mutants and RNAi lines. The duration of the complete protocol strongly depends on the number of ORFs required, but the procedure of injection and establishing balanced fly stocks can be completed within approx. 6-7 weeks for a few genes.

INTRODUCTION

Loss-of-function approaches have been very powerful for determining gene function in *Drosophila*; in particular, genetic screens based on random mutagenesis have been instrumental in elucidating biological processes and delineating signal transduction pathways. More recently, targeted approaches such as the genome-scale generation of RNAi libraries have gained popularity[1] and RNAi lines are now widely available from various sources such as the Vienna Drosophila RNAi Center (VDRC, http://stockcenter.vdrc.at/control/main), the Transgenic

RNAi Project (TRiP, http://www.flyrnai.org/TRiP-HOME.html) and Flystocks of National Institute of Genetics (NIG-Fly, http://www.shigen.nig.ac.jp/fly/nigfly/).

A drawback of the loss-of-function approach for elucidating gene function is the lack of observable phenotypes for the majority of genes in conventionally-used screening set-ups[2]. For many genes this may be due to functional redundancy between related genes; for others the reason may be the lack of suitably sensitive assays. Given that a large fraction of genes in *Drosophila* (and many other organisms) still have no known function, alternative approaches to characterize these genes are needed.

Gain-of-function analyses, such as over- and misexpression of genes, are a useful complement to loss-of-function studies. To date, such approaches have relied mainly on random insertions of transposons containing Gal4-inducible enhancer-promoter (EP) elements, which drive the expression of genes flanking the insertions[3,4,5]. A rare exception to this approach is the collection of transgenic strains specifically generated from a limited set of UAS-ORF constructs comprising 236 human genes[6]. A limitation of transposon-based approaches is that insertions do not target all available genes; therefore, screens using EP lines are typically far from saturating. Furthermore, it usually requires considerable effort to firmly establish whether the phenotype is caused by: overexpression of nearby genes, an effect of integration into that particular genomic locus, or a combination of both.

To overcome these limitations we have designed a strategy (Fig. 1) to generate a comprehensive UAS-ORFeome library to control expression of transgenes in *Drosophila*[7]*.* The main features of this approach are:

- the use of sequence-verified ORFs cloned using the 'Gateway' recombination cloning system;
- the presence of a DNA barcode in the transgenic construct, facilitating injection of pooled plasmids, for the generation and subsequent identification of transgenic fly strains;
- the use of the site-specific ΦC31 integrase method;
- the presence of specific *FRT* sites flanking the ORFs to allow *in vivo* 'swapping' of promoter and/or tags.

A pilot library consisting of an untagged ORF set and a 3xHA(hemagglutinin)-tagged set, altogether comprising approx. 650 unique genes, was used to screen for novel Wnt/Wg pathway components and led to the identification of three kinases as regulators of Wg signaling[8]. To increase the versatility of the library, we have further modified the expression vector and introduced 'swapping' elements (i.e. mutated *FRT* sites) to facilitate *in vivo* exchange of epitope tags and regulatory regions. This led to the development of the currently-used destination vector, pGW-HA.attB[7] (see Fig. 2a).

Here, we provide a detailed protocol for the high-throughput creation of UAS-ORF (i.e. UAS-*hsp70*-ORF) *Drosophila* strains for *in vivo* expression. Currently available fly strains can be searched and ordered through the Zurich ORFeome Project (www.*flyorf.ch)*.

Applications

The approach presented in this paper can be used to generate individual transgenic strains and is scalable for the generation of genome-scale libraries of UAS-ORF strains. The most obvious application for such a large ORF library is its use in systematic overexpression screens to characterize gene functions through phenotypic readouts[8]. To increase output and specificity, many of these screens will likely be performed in sensitized genetic backgrounds, similar to loss-of-function screens. However, compared to loss-of-function screens, ORF expressing fly strains potentially have greatly increased versatility through an expanded range of applications and we have attempted to include features that would maximize the range of current and potential future applications. In addition to phenotypic characterization, the use of 3xHA epitope-tagged ORFs already allow for the analysis of protein localization, and analysis of interaction partners by mass-spectrometry and chromatin immunoprecipitation (ChIP) assays. Furthermore, the inclusion of specifically altered *FRT* recombination sites allows any sequence-encoded N- or C-terminal modification to be made. Lines can thus be customized for an even wider range of applications simply by exchanging epitope tags or promoter regions. Some potential applications are BiFC (bimolecular fluorescence complementation) analysis and introduction of "degron tags"[7,9,10,]. Although the cloned genes introduced into our destination vector pGW-HA.attB are intended for expression *in vivo*, they can be also used in cultured *Drosophila* cells through co-transfection with a suitable GAL4 expression plasmid such as *tubulin*-GAL4 or by transferring the ORFs into other *Drosophila* cell culture expression vectors[8]. The flexibility of the Gateway cloning system further allows for the transfer of ORFs into other vectors, such as those used for bacterial or mammalian expression.

Limitations

Phenotypes obtained from expressing epitope-tagged transgenes must be evaluated carefully and ideally validated by other methods[11], including loss-of-function analyses. Notably, expression of transgenes can result in dominant-negative effects, yielding the same phenotype as loss-of-function approaches. Despite the benefit of epitope tags for protein detection, tags may generate false positive or false negative phenotypes when compared to untagged proteins[7]. Overexpression of transgenes can cause mis-localization of proteins, affect protein folding and protein complex assembly, which could potentially have far-reaching regulatory effects (for a recent review, see[12]). Approaches that tag proteins in their endogenous loci are likely to be more reliable strategies for reflecting accurate subcellular localization[13], but these endogenously tagged constructs cannot be used to screen for overexpression phenotypes, which we consider

the main application of our ORF library. A limitation of our approach is that we are currently generating only a single transcript variant per gene, but the protocol provided here would allow other users to generate additional transcript variants. However, despite extensive transcriptional variation through alternative promoter usage and splicing, more than half of the fly genes produce only a single polypeptide[14].

Experimental design

ORF cloning systems. No complete genome-wide ORF library for *Drosophila* is currently available; however the Berkeley *Drosophila* Genome Project (BDGP) has generated a large collection of ORFs using the "Creator" recombination cloning system[15]. Side by side comparison of fly strains generated using either Creator or Gateway cloned ORFs showed that the protein expression levels of the Gateway clones were generally higher and consequently *in vivo* phenotypes, upon overexpression, were more reliably observed in the Gateway fly strains than in those using the Creator system[7]. In light of this, we opted to use the Gateway system to clone the *Drosophila* ORFs.

ORF template sequences. Coding sequences can be found and downloaded individually or in batches from various sources such as FlyBase (http://flybase.org/), Ensembl (http://www.ensembl.org) and NCBI (http://www.ncbi.nlm.nih.gov) The Drosophila Gold collection contains a large collection of cDNAs whose coding sequences perfectly match reference sequences. Alternatively, many ORFs can be cloned using cDNA prepared from embryos or adult tissues. Many of these resources are readily available from the *Drosophila* Genomics Resource Center (https://dgrc.cgb.indiana.edu/). Additionally, for *Drosophila* ORFs that are single-exon genomic DNA can be used as a convenient template for amplification.

PCR strategy. We use a two-step PCR strategy (see Fig. 3a) to amplify ORF sequences for cloning. The first PCR reaction uses gene-specific primers (see Reagent Setup) to introduce the *Drosophila* Kozak sequence CAAC before the initiator methionine codon[16]. A second amplification round with primers pDONR221_attB1 and pDONR221_attB2 primers introduces full-length attB1 and attB2 recombination sites into the PCR product as instructed in the Gateway cloning manual. To improve the amplification efficiency, betaine is routinely included.

Purification and verification of PCR products. We purify PCR products using paramagnetic beads, which we have found to be suitable for high-throughput cloning applications. These beads can be prepared and used as previously described[17] or commercially available kits such as the Agencourt AMPure XP beads (Becton Dickinson) can be used. Gateway cloning efficiency is reduced with increasing insert size, but gel purification of large ORFs (>3 kb) may increase success rate compared to bead purification. Purified PCR products encoding full-length ORFs are cloned into the pDONR221 entry vector and sequence verified (see Fig. 3)**.** Our pilot library was sequenced with a combination of Sanger sequencing and Roche 454 sequencing. More recently,

we have moved to Illumina MiSeq for full insert sequencing using dsDNA fragmentase (NEB) as a convenient and scalable method to fragment plasmid DNA (Fig. 3c). High-throughput sequence analysis can be carried out using published ORF assembly software[18].

Key features of the pGW-HA.attB expression vector. Sequence-confirmed ORF entry clones are transferred to the destination vector pGW-HA.attB using a Gateway LR reaction. Key features that we have incorporated into the pGW-HA.attB expression vector to maximise the potential applications of the transgenic flies (see Fig. 2a) are:

- A C-terminal tag: The collection is currently being generated without a stop codon and with a C-terminal 3xHA tag, which is useful for determining that the full-length protein is expressed using western blotting. The HA epitope tag was chosen for its small size, the availability of affordable reagents to detect it and the experience many researchers have with it. C-terminal tagging is used as it has been shown by others that N-terminal tags more commonly cause mis-localization of proteins[19].
- Mutated FRT sites to facilitate domain swapping: This functionality enables researchers to easily replace the HA tag with other tags, such as fluorescent proteins (see below, Tag or promoter swapping). Promoters can also be exchanged using this approach.

Barcoding the pGW-HA.attB expression vector. Molecular barcoding using randomized duplex oligonucleotides is used to facilitate the identification of plasmids and fly lines when applying pool injections. Introduction of a barcode into other regularly-used expression vectors is an attractive option for any lab working with *Drosophila,* as pool injections would circumvent the more common and laborious single injection approaches. The barcoding approach is conceptually similar to that used in yeast genomic screens[20]. Efficient barcoding is dependent on minimizing 'insertless' vector background. We achieve this by using a pGW-HA.attB vector containing a lacZ "stuffer" gene in the place of the barcode; when the vector is digested with appropriate enzymes (XhoI-HindIII in our case) to release the stuffer fragment, the separation of single-cut from double-cut vector on an agarose gel is maximized – thereby facilitating efficient barcoding, which is based on ligation of randomized oligonucleotides.

Selection of a suitable *attP* integration site. ΦC31 integrase mediated site-specific integration of genetic elements into the *Drosophila* genome has rapidly become a standard method for transgenesis and considerable resources are already available (see "*Drosophila* reagents" section below). The ΦC31 integrase method offers (i) an efficient way to generate transgenic flies due to high integration rate, (ii) eliminates the need for mapping of the insertions, and (iii) allows for better control of expression by using characterized *attP* integration sites[21, 22, 23, 24, 25]. For our ORF library we prefer to use the integration site at cytological position 86F, using the previously generated *attP* insertion *ZH-attP-86Fb*[24] (see Fig. 2b). We use this *attP* integration site combined with a codon-optimized ΦC31 integrase transgene inserted on the X chromosome. The resulting line, *ΦX-86Fb* displays a high transgene integration rate, relatively strong transgene expression,

ease of scoring transgenic offspring due to strong *white* marker expression as well as good overall fitness and high fertility. We have detected some background expression during early embryonic development from this landing site and note that this could potentially affect phenotypic characterization with some ORFs.

Embryo injections with pooled plasmids. Injection of embryos with individual transgenes is time consuming and we have circumvented this limitation by pooling expression plasmids (typically one 96-well plate at a time). This pool size typically allows recovery of ~60% of the injected ORFs from 100-120 F1 outcrosses per injection round of ~1000 embryos. Accurate quantification of expression plasmids used for pooling is important to ensure equal representation of the transgenes in the offspring. Constructs not recovered in the offspring initially are pooled again, often together with other plasmids to keep the pool size optimal (~100 or more for our large-scale approach). Two or three rounds are typically needed to obtain most transgenes at which point missing genes can be individually injected. The barcode information is used to identify the transgenes using single-fly PCR and Sanger sequencing (Fig. 2b).

Tag or promoter swapping. We have included mutated *FRT* sites[26] in the pGW-HA.attB expression vector, which flank the Gateway cassette/ORF sequence (Fig. 2a). This allows facile exchange of either the promoter or, more commonly, the epitope tag by crossing the original UAS-ORF-3xHA line with either epitope tag or promoter 'swapping' lines without the need of re-cloning[7] (see Fig. 4 for examples). The promoter/epitope tag lines have alternative sequences inserted at the same locus as the UAS-ORF transgenes (86F), and these lines also carry an *hs-flp* transgene necessary for the exchange reaction. The swapping events can be easily tracked by screening for specific marker situations, e.g. w^+y^+ for the C-terminal exchange (see Fig. 4b). As already mentioned, the swapping option substantially increases the versatility of the ORF library and we are in the process of expanding the repertoire of alternative tags (see Fig. 4c).

MATERIALS

REAGENTS:

Molecular Cloning

- Gateway entry vector pDONR221 (Life Technologies, cat. no. 12536-017)
- Gateway destination vector pGW-HA.attB (Fig. 2; accession no. KC896838)
- Tris, (Formedium, cat. no. TRIS01)
- Hydrochloric Acid (VWR, cat. no. 20252.335)
- EDTA (Fisher Scientific, cat. no 1021-3570)
- Sodium chloride (VWR, cat. no 27810.634)
- 10 mM Tris-HCl, pH 8.5 (pH adjusted with HCl)
- TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0)

- 1x oligonucleotide annealing buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM, EDTA pH 8.0)
- Betaine (Sigma-Aldrich, cat. no. 14300, prepared as 5M stock in ultrapure water)
- 10mM dNTPs (Fermentas, cat. no. R0481)
- HindIII (Fermentas, cat. no. ER0505),
- XhoI (Fermentas, cat. no. ER0695)
- Bsp1407I (BsrGI) (Fermentas, cat. no. ER0932)
- *D.melanogaster* nucleic acid template for amplifying ORF (genomic DNA, total RNA and/or cDNA clone). This can be either DNA or RNA purified from *Drosophila* embryos, adults or cultured cells (e.g., Kc167 or S2 cells) or cDNA clones from The Drosophila Genomics Resource Center (DGRC, https://dgrc.cgb.indiana.edu)
- Any high-fidelity polymerase, such as Phusion (Thermo Scientific, cat. no. F-530L), and the corresponding PCR buffer. CRITICAL We do not recommend standard Pfu polymerase due to its low processivity.
- GeneRuler DNA ladder or similar (Thermo Scientific, cat. no. SM0331)
- Carboxylated M-PVA C11 Magnetic Beads (0.5-1.0µm size), 500mg unit (Chemagen, Product Code: 203)
- PEG8000 (Sigma Aldrich, cat. no. P5413-500g)
- Sodium azide (Sigma Aldrich, cat. no. S8032-25g)

! Caution Sodium azide is toxic; exposure via skin contact can be very harmful.

- 15% PEG8000, 2.5 M NaCl solution for Magnetic Beads. Add 0.1% sodium azide for long-term storage. Store at 4 °C.
- 70% Ethanol (Fisher Scientific, cat. no 1010-0332)

! CautIon Ethanol is highly flammable and volatile.

- *E. coli* DB3.1 or ccdB survival (Invitrogen, cat. no. A10460) for maintenance of destination vectors.
- Max Efficiency DH5α-T1R Competent Cells (Invitrogen, cat. no. 12034-013, or similar cloning strain) for transformation of the BP and LR recombination reactions.
- LB medium (Merck Millipore, cat. no 71753-6). Store at room temperature.
- Terrific Broth medium (Merck Millipore, cat. no 71754-4). Store at room temperature.
- SOB medium (Sigma-Adrich, cat. no H8302-500g). Store at room temperature.
- LB medium agar (Merck Millipore, cat. no 1102830500). Store at room temperature.
- 100 mg/ml Ampicillin (Melfords, cat. no. A0104). Store at -20°C.
- 50 mg/ml Kanamycin (Formedium, cat. no. KAN0025). Store at -20°C.
- Gateway BP Clonase II Enzyme mix (Life Technologies, cat. no. 11789-020)
- Gateway LR Clonase II Enzyme mix (Life Technologies, cat. no. 11791-020)
- pDONR221 (Life Technologies, cat. no. 12536-017)

- NEBNext dsDNA Fragmentase (NEB, cat. no. M0348S)
- NEBNext End Repair Module (NEB, cat. no. E6050S)
- NEBNext dA-Tailing Module (NEB, cat. no. E6053S)
- T4 DNA ligase (Fermentas, cat. no. EL0011)
- Wizard SV 96 Plasmid DNA Purification System (Promega, cat. no. A2250)
- Plasmid midi kit (Qiagen, cat. no. 12143) for cleaning up pooled plasmid DNA
- ORF-specific primer pairs for first-round PCR amplification (Sigma-Aldrich) (see Reagent Setup)
- Barcode oligo pair, both 5' phosphorylated (IDT). When annealed (see Reagent Setup) will have sticky ends (bold nucleotides) compatible with XhoI and HindIII digested vector.
  barcode_up: **TCGAG**ANNTGNNNACNNNNTGANNNNACNNNATNNNGANN**G**
  barcode_down: **AGCTC**NNTCNNNATNNNGTNNNNTCANNNNGTNNNCANNT**C**.
- Second-round PCR primer pair for adding att sites (IDT).
  pDONR221_attB1 GGGGACAAGTTTGTACAAAAAAGCAGGC*T
  and pDONR221_attB2 GGGGACCACTTTGTACAAGAAAGCTGGG*T
  (* denotes phosphorothioate modification between the last two nucleotides. The purpose of this modification is to reduce primer degradation leading to non-specific priming. Order as HPLC purified oligonucleotide). Can also be used without phosphorothioate and desalted without HPLC purification.
- Sequencing primers. For single ORF sequencing, use M13F (GTAAAACGACGGCCAG) and M13R (CAGGAAACAGCTATGAC) sequencing primers. For high-throughput sequencing of an ORF collection use any suitable library preparation kit for Next Generation Sequencing.
- Barcode sequencing primer, tub-F2new (5'-ATTTATGTGACTATGGTAGGTCG)
- Primers for amplifying and sequencing the barcode region in the fly strain HL1-F 5'-TGGACCTGCGGGGTTAATTTACC and attP-R 5'-TTGTGTCATGTCGGCGACCCTACG.
- Generic primers for amplification and sequencing the ORF in the fly strain hsp-GW-F (GCAACTACTGAAATCTGCCAAG) and HL1-R (AATTAACCCCGCAGGTCCACCGG).

Drosophila reagents

- Fly strain containing a mapped insertion site and an integrase source. In this protocol we use *ΦX-86Fb* fly strain (Bloomington *Drosophila* Stock Center, stock no. 24749)
- Bleach solution (7% sodium hypochlorite (NaClO); Erne-Chemie, Art. Nr. 82365)
- "Heptane-Scotch" solution (see Reagent Setup)
- Oil 10S (Prolabo, prod. No. 24627.188)
- Proteinase K (for example, Sigma Aldrich)
- Squishing buffer (see Reagent Setup)

- GoTaq G2 DNA Polymerase (Promega, cat. no. M7845)
- A suitable fly strain for swapping epitope tags or promoters. In this protocol we use *TSeGFP-86Fb* as an example. All available swapping lines can be obtained upon request from FlyORF, https://www.flyorf.ch/. Cloning constructs that facilitate the generation of additional swapping lines are also available upon request.
- Other useful resources for the ΦC31 integrase method:
  Bloomington Drosophila Stock Center
  (http://flystocks.bio.indiana.edu/Browse/phiC31/phiC31home.htm);
  FlyC31 (http://www.frontiers-in-genetics.org/flyc31/);
  Fly RMCE (http://genepath.med.harvard.edu/WuLab/RMCE/news.html);
  P[acman] Resources (http://www.pacmanfly.org/index.html)

EQUIPMENT:

Molecular Cloning

- Skirted 96 well PCR plates (VWR, cat.no.82006-704)
- Adhesive PCR Plate Seals (Thermo Scienific, cat. no. AB-0558)
- Manual or electronic 8 or 12-channel pipette and tips (Biohit, cat. no. 73049X; Biohit, cat. no. 791201)
- Manual or robotic 96-well pipetting system and tips (for example Anachem, cat. no. LIQ-96-200; Anachem, cat. no. LIQ-200; Anachem, cat. no. LIQ-20)
- A centrifuge suitable for microtiter plates (for example, ThermoFisher Scientific, Heraeus Multifuge X3R cat. no. 75004515)
- 15 ml polypropylene tubes (Greiner Bio-one, cat. no. 188271)
- Reservoir trays (Sigma Aldrich, cat. no. R9259)
- 1.5 ml microcentrifuge tubes (Sarstedt, cat. no. 72.690.001)
- 96- or 384-well Thermocycler, depending on project scale (for example, Life Technologies, 96-well cat. no. 4375786 or 384 well, cat. no. 4388444)
- 96- or 384-well (ThermoScientific, cat. no. AB-0937) PCR plates
- E-gel High-throughput DNA electrophoresis System (Invitrogen, Mother E-Base Device, cat. no. EB-M03; 1% and 2% E-Gel 96 Gels cat. nos. G700801 and G720802)

! Caution DNA stains are typically mutagenic and must be handled with care and disposed in accordance with institutional regulations.

- 96-well Super Magnet plate (Alpaqua, cat. no. A001322)
- 24-well plates (Greiner Bio-One, cat. no. 662160)
- 2.0 ml 96-well deep-well plates (Greiner Bio-One, cat. no. 780270)
- 37°C rotary incubator (for example, Infors HT Multitron Standard)

- 37°C static incubator (for example, LTE Scientific IP Series)
- Humidity control chambers for incubator (Ermis Medizinetechnik, cat. no. ER210.100E)
- 96-well UV star microplates (Greiner Bio-One, cat. no. 655801)
- Spectrophotometer suitable for measuring DNA concentrations on 96-well microplate (for example BMG Labtech, FluoStar Optima)

Drosophila equipment

- Apple juice agar plates (In-house media service; also see[27])
- Bench-top centrifuge
- Desiccation chambers (see Equipment Setup)
- Dissecting microscope
- Egg-laying cage (fly cage; in-house made; also see[27])
- Electrical needle tip grinder (H. Saur Laborbedarf, Reutlingen, Germany)
- FemtoJet (Eppendorf, order no. 5247 000.013)
- Filter flask with water jet pump
- Fly food vials
- Humid chambers (= Petri dishes with Whatman paper fixed to the lid)
- Inverted microscope Axio Observer.A1 (Zeiss)
- Membrane filters (Whatman, ME 26/31 ST, Ref. 10 409 870)
- Microcapillaries (World Precision Instruments, item no. TW100-3)
- Microloader (Eppendorf, order no. 5242956.003)
- Micromanipulator TransferMan NK2 (Eppendorf, order no. 5188 000.012)
- Microscope slides (Thermo Scientific, 76x26 mm)
- Narishige needle puller (Model PN-30, Narishige International, Japan)
- Scotch adhesive tape (from 3M, double sided)
- Syringe (10 ml; e.g. B. Braun, Melsungen, Germany, Ref. 460108V)

Software

- Primer3 for primer design (open source available from http://primer3.sourceforge.net)
- PrinSeS-C (web version available at http://webprinses.epfl.ch and available from the developers[18])

REAGENT SETUP

Primer design As cloning of the ORF requires specific primer binding sites, you will need to allow flexibility in primer design quality parameters. Set optimum size as 22 nt (allowed range 19-32) and melting temperature 62°C (allowed range 58-72, using SantaLucia salt correction for Tm

determination). For other options, set "PRIMER_TASK=pick_cloning_primers" so that Primer3 program designs the primers to amplify the whole coding sequence, set "PRIMER_NUM_RETURN=1" so that only one set of primers for each gene is designed and set "PRIMER_PICK_ANYWAY=1" to pick primer even if optimal primer sequences are not found. Provide template sequences without stop codon as an input file. Amend forward primer with a sequence AAAAAGCAGGCTTCAAC before methionine codon and gene specific sequence and reverse primer with AGAAAGCTGGGTC flanking sequence. Order these primers as plate oligonucleotides synthesised at 0.025 µmole scale, desalted, lyophilised and normalised to 5-10 nmol/well (Sigma-Aldrich). Oligonucleotides are then dissolved to 50 µM in 10 mM Tris-HCl, pH 8.5. Mix fw and rev primer to 1 µM final concentration for the amplication of the ORFs. Store primers at -20°C.

Annealing of barcode oligonucleotides Mix 1 µM forward and reverse 5'-phosphorylated barcode oligos in 1x oligonucleotide annealing buffer (see Reagents) and heat at 95°C for 5 min followed by slow cooling to room temperature (<1°C/sec). This results in appropriate flanking restriction sites being created as overhangs suitable for ligation.

"Heptane-Scotch" solution Incubate double-sided adhesive tape (e.g. Scotch from 3M, 1-2 meters scrunched up to a ball) in a 100 ml volume glass bottle, which contains ~50-70 ml heptane, and shake for ~1 day at RT to dissolve the glue from the tape.

Bleach solution Dilute a 14% sodium hypochlorite solution 1:1 in ddH_2O, and keep the stock solution at 4 °C. Approximately 100 ml aliquots are kept at ~18 °C as ready-to-use bleach to remove the chorion from the embryos.

Pooled plasmid DNA Centrifuge plasmid DNA pool in a bench-top centrifuge at 4°C at full speed for 20 min or longer to sediment particles that could clog the injection needle. Immediately after centrifugation most of the upper volume can be transferred to a clean tube to be used for loading the injection needles.

Squishing buffer This is the lysis buffer needed for single-fly PCR and has the following composition: 10 mM Tris-HCl, pH 8.2; 1 mM EDTA; 25 mM NaCl; 200 µg/ml Proteinase K.

EQUIPMENT SETUP

Egg-laying cages Establish a collection cage with approx. 600-700 *ΦX-86Fb* (or equivalent) flies at least two days before embryo injections. Close the cage with an apple juice plate with some dry yeast paste and keep at 25 °C. Change the apple plates at least once a day. On the day of injection change plates twice before collecting eggs for injection. Alternatively, transfer the flies to fresh fly food vials/bottles about three days before injection and transfer them to the cage one day before injection at the latest.

Injection needle preparation We use the Narishige PN-30 needle puller to generate suitably formed needles from microcapillaries. The various settings should result into a slender taper while

ensuring an appropriate stiffness. After pulling, the needles are opened and sharpened on a continuously watered tip grinder. Press the needle briefly on the rotating grinder until some backflow into the needle is observed. A sharp beveled tip is desired.

Preparation of glass injection slides These are prepared immediately before use. Lean the slides against a raised object, take a Pasteur pipette filled with the heptane-scotch solution, and cover the glass slide twice in the middle region.

Desiccation chambers Use conventional glass jars filled to one-third with silica gel as desiccation chambers and keep the filled jars in an oven set at 80°C. About one hour before use take the jar out from the oven, close with a lid, and keep at about 18 °C to let it cool down. When used for drying the embryos the glass jar should no longer feel warm to the touch.

PROCEDURE

Barcoding the library expression vector pGW-HA.attB TIMING 3 d

CRITICAL Ligation must be carefully optimised to minimise background from vector with no barcode and to maximise barcode diversity if preparing a large library. Avoid or minimise exposure of the vector to UV light as this dramatically reduces cloning efficiency. Competent DB3.1 or *ccdB* survival cells are required for transformation as *ccdB* gene is toxic to normal cloning strains such as DH5α. We routinely obtain competencies of $5x10^7$ colony forming units/µg pUC19 DNA with *ccdB* survival cells and this is sufficient for barcoding.

1. Digest 15 µg pGW-HA.attB containing a lacZ stuffer, with HindIII/XhoI at 37°C for ~2 hours.

2. Separate fragments on a 0.8% agarose gel in TBE buffer. The digested vector is approximately 10 Kb and the lacZ stuffer is approximately 3.1 Kb).

3. Gel purify vector using standard methods and quantify DNA spectrophotometrically.

PAUSE POINT DNA can be stored at -20 °C for an indefinite period.

4. Set up a series of ligation reactions, each using a different insert:vector ratio, as tabulated below (including a negative control without barcode duplex oligo). Incubate ligations for 1 h at 16°C.

Component	Amount per reaction	Final
Cut vector	~700 ng	~700 ng
T4 DNA ligase buffer (10x)	12.5 µl	1x
Barcode duplex oligo (1 µM)	0-2 µl*	0-16 nM

H_2O	to 125 µl	
T4 DNA ligase (5U/µl)	3 µl	15U

*use for example 0, 0.2, 0.5, 1 and 2 µl of the annealed duplex oligo.

5. Determine the best insert:vector ratio by setting up test transformations on separate aliquots of DB3.1 or *ccdB* survival cells with 1 µl of the different ligation mixes. Plate the transformed cells on LB agar plates containing 100 µg/ml ampicillin and incubate the plates overnight at 37 °C. Incubate the remaining ligation reactions overnight at 16°C.

6. Next day, check the LB plates to determine which ligation gives the best signal to background. Use this ligation to transform 50 samples of competent DB3.1 or *ccdB* survival cells (50 µl aliquots) with 2.5 µl ligation mix each; combine the heat-shocked bacteria, add to 100 ml SOC medium and shake for 1 hour at 37°C. If preferred, electroporation can be used in place of heat-shock based transformation.

7. Plate 10 µl and 100 µl of transformed bacteria in duplicate on LB agar plates containing 100 µg/ml ampicillin. Incubate the plates overnight at 37 °C. To the remainder of the transformed bacteria, add 500 ml LB medium containing 75 µg/ml ampicillin and 10 µg/ml chloramphenicol and grow for another 16-19 hrs at 37°C.

8. Next day, score colony numbers on the LB plates to estimate the diversity of the barcoded plasmid library. If the diversity is acceptable, use the 500 ml overnight cultures to purify plasmid DNA using a maxi prep kit.

   CRITICAL STEP At least 50 times higher barcode diversity than the planned ORF clone number is recommended to minimize occurrence of identical barcodes.

PAUSE POINT DNA can be stored at -20 °C for an indefinite period.

Cloning and sequence verification of ORFs **TIMING** **several days to weeks depending on the number of ORFs required**

9. Set up the following 25 µl PCR mix to amplify ORFs using gene specific primer pairs and appropriate template DNA (see Experimental Design and Reagent Setup).

Component	Amount per reaction (µl)	Final
5xPCR buffer	5	1x
5 M Betaine	5	1 M

10 mM dNTPs	0.5	200 nM
DNA template	20 ng	20 ng
H_2O	to 25 µl	
Proof-reading polymerase (2U/µl)	0.5	1U
1 µM each gene-specific Fw + Rv primers	2.5 each	100 nM each

10. Amplify using the following cycling conditions as a guide. Note that the optimal annealing temperature may vary depending on the primer pairs, and that the extension time is dependent on the length of the expected PCR product.

Cycle Number	Denature	Anneal	Extend	Hold
1	98°C, 10 sec		9	
2-6	98°C, 15 sec	64°C-1°C/cycle, 5 sec	68°C, 1 min/kb + 1 min	
7-27	98°C, 15 sec	60°C, 25 sec	68°C, 1 min/kb + 1 min	
28				4°C

11. Without purifying the PCR products, perform a second round of PCR using primers pDONR221_attB1 and pDONR221_attB2 to add the complete recombination sites into the PCR product. Prepare a 25 µl reaction as tabulated below.

Component	Amount per reaction (µl)	Final
5xPCR buffer	5	1x
5 M Betaine	5	1 M
10 mM dNTPs	0.5	200 nM
Template from previous PCR	3	
H_2O	to 25 µl	
Polymerase (2U/µl)	0.5	1U
2.5 µM each pDONR221_attb1 + pDONR221_attB2	2.5 each	250 nM each

12. Amplify using the following cycling conditions as a guide. Note that the extension time is dependent on the length of the expected PCR product.

Cycle Number	Denature	Anneal	Extend	Hold

1	98°C, 10 sec		9	
2-6	98°C, 15 sec	50°C, 25 sec	68°C, 1 min/kb + 1 min	
7-17	98°C, 10 sec	60°C, 15 sec	68°C, 1 min/kb + 1 min	
18				4°C

13. Purify PCR products using solid-phase reversible immobilization[17] as follows: add two volumes of magnetic bead/PEG/NaCl solution to each PCR; incubate for 10 min at room temperature; capture the beads with a magnetic plate and remove the solution; wash the beads twice with 70% ethanol (capturing beads on the magnetic plate when removing each wash); and air-dry the beads for 10 min. Elute the purified PCR product from the beads by adding 20 µl 10 mM Tris-HCl, pH 8.5, capturing the beads on the magnetic plate and transferring the solution to a clean well/tube.

PAUSE POINT DNA can be stored at -20°C for an indefinite period.

14. Perform Gateway BP recombination: add 1.5 µl of purified PCR product to Gateway BP cloning reaction containing pDONR221 vector (total volume 7.5 µl) in a 96-well or 384-well PCR plate. Seal the plate and incubate at 25°C for at least one hour, preferably overnight.

    CRITICAL STEP To avoid evaporation of the small reaction volume, place the 384-well reaction plates into reclosable polyethylene bags.

    CRITICAL STEP Proteinase K treatment before transformation as detailed in the Gateway kit manual is optional (but improves transformation efficiency) and we omit it in the high-throughput protocol.

15. Perform transformations in 96-well conical PCR plates using 20 µl competent DH5α T1R cells (Invitrogen) and 2 µl BP recombination reaction. Heatshock for 1 min at 37°C in a thermal cycler, place on ice and add 100 µl SOC medium. Recover at 37°C for 60 minutes before plating 10-30 µl on 24-well plates containing 500 µl LB agar containing 50 µg/ml kanamycin. To avoid excess drying of the 24-well agar plates pack them into reclosable polyethylene bags and incubate overnight at 37°C.

16. Next day, pick single clones into 96-well deep well plates containing 1.3 ml Terrific Broth medium containing 50 µg/ml kanamycin and culture for 22-24 h at 37°C in a humidity controlled environment to minimize evaporation.

PAUSE POINT For storage of ORF clones as bacterial cultures, add sterile glycerol to a final concentration of 15% and store at -80°C

17. Purify plasmid DNA using the Wizard SV 96 Plasmid DNA Purification System, according to the manufacturer's instructions. Elute plasmids with 100 µl 10 mM Tris-HCl, pH 8.5 and measure plasmid concentrations on 96-well UV compatible plates using known concentrations of control plasmid DNA as standards.

PAUSE POINT DNA can be stored at -20 °C for an indefinite period.

18. Pool approximately equimolar amounts of each plasmid to ensure even sequencing coverage and fragment 0.5-1 µg plasmid DNA in a 20 µl dsDNA fragmentase reaction containing 1 µl fragmentase enzyme mix for 45 minutes at 37°C (see Fig. 3).

19. Depending on the sequencing sample preparation kit used, perform end repair, A tailing and adapter ligation according to the manufacturer's instructions. Alternatively, plasmids can be Sanger sequenced using M13F and M13R primers. Presence of an insert can be verified using Bsp1407I (BsrGI) digestion. The sequencing library preparation is typically done by the sequencing service provider, but for a general protocol see[28]. Note that due to the att recombination sites in the entry clone, Sanger sequencing does not often yield ideal sequencing quality.

PAUSE POINT Sequencing library can be stored at -20 °C for an indefinite period.

20. Assemble clone sequences using PrinSeS-C software[18].

21. Transfer sequence-verified clones to barcoded pGW-HA.attB vector by adding 1.5 µl of entry clone miniprep from step 17 (concentration ~50 µg/ml) to Gateway LR cloning reaction (total volume 7.5 µl) and transform as described in step 15. Select transformants on LB agar containing 100 µg/ml ampicillin.

22. Pick single colonies into 96-well deep well plates containing 1.3 ml Terrific Broth medium and 100 µg/ml ampicillin. Culture for 20-24 h at 37°C in a humidity controlled environment to minimise evaporation.

23. Purify plasmids and measure concentrations as in step 17.

Barcode-sequence determination of the plasmids TIMING **2 d**

24. Organise cloned and sequence verified ORF minipreps in 96-well plates and normalise concentrations to ~100 ng/µl. Take a 3 µl aliquot from each well position of a 96-well plasmid miniprep plate and transfer to sequencing wells.

25. Sequence barcodes using Sanger sequencing with primer tub-F2new to associate the specific barcode sequence with an individual ORF and a respective well position.

Pooling and purification of ORFs for microinjection TIMING **3 h**

26. Pool ~5 µl miniprep DNA from each well of a 96-well plate and clean using a QIAfilter Midi cartridge (Qiagen) as per the manufacturer's instructions to obtain injection-quality DNA. Adjust pool concentration to 100 ng/µl for microinjection.

CRITICAL STEP Equal representation of the clones in the pool and midiprep purification maximizes the recovery of all the transgenes.

PAUSE POINT DNA Pools can be stored at -20°C for an indefinite period.

Injections TIMING **~1 d**

CRITICAL The number of embryos that have to be injected to subsequently recover a substantial number of different transgenic flies depends on several criteria, mainly the integration rate of the used attP site and the survival rate of the injected fly embryos. The line ΦX-86Fb scores well in both these aspects and as a rule of thumb we inject about 10 "injection slides" per pool consisting of 100 constructs (on one slide we have approx. 100 embryos aligned, thus we inject about 1000 embryos). Generally, after the G0 outcrosses, we recover around 60% of the pooled constructs as transgenic flies from approx. 100 F1 outcrosses. (Note, we use only injected males in G0 and F1 crosses to remove the integrase transgene located on the X chromosome*.)*

27. Change the plate of the fly cage with a fresh, prewarmed apple agar plate containing a fingertip-sized smear of yeast paste. Leave cage for approx. 30 min at 25°C. Repeat this step so that plates are changed (and discarded) 1-2 times before proceeding to step 28 for embryo collection. The plate for embryo collection should be left on the cage for approx. 30 min.

28. Harvest and then dechorionate embryos in the injection room (set at ~18 °C) by adding bleach to embryos on the agar plate. Incubate for ~4 min, then briefly agitate to see whether embryos detach from the agar.

29. Start the water jet pump, which is connected to a conventional filter flask, place a membrane filter onto the permeable plate on top of the flask, and rinse the filter with tap water, which will be sucked into the flask.

30. Pour dechorinated embryos from the agar plate onto the filter membrane by repeated rinsing. Wash those embryos on the membrane filter thoroughly with water. When embryos start to clump together, stop washing, take the filter membrane with the embryos and start aligning.

31. Align embryos directly on the filter membrane with a wet brush, side by side, about two embryo widths apart, in the same anterior-posterior orientation. After finishing one row (~50 embryos), start a second next to the first with opposite anterior-posterior orientation. In this "double row" the micropyles of all embryos should point to the midline, i.e. the posterior poles point to the outside, where they are accessible to the injection needle.

CRITICAL STEP Leave enough space between the aligned embryos to prevent oxygen depletion.

32. Cover one side of the glass injection slides with the sticky heptane-scotch solution (e.g. with a Pasteur pipette). Once dry, take the prepared slides and gently press the glue-coated side against the double rows of aligned embryos, until the embryos stick to the glass slides (one double row per glass slide).

33. Place the slides into a cooled desiccation chamber, close the lid and incubate for ~6 min or longer.

CRITICAL STEP The appropriate desiccation time is highly critical for the success of an injection experiment. If it is too short, embryos will almost immediately leak out after retracting the needle. We recommend covering a few embryos of the first slide with some oil and injecting these embryos to evaluate the desiccation. If an appropriate drying period has been achieved, overlay all the properly dried embryos.

34. Meanwhile sharpen a needle on the grinder, load with a small amount of DNA and fix onto the needle holder. Stick the needle into a drop of oil, focus the tip of the needle and press the clear button of the FemtoJet. This allows for a rough evaluation of needle

quality before injection (for example, is the tip open wide enough, is the tip sharp or blunt).

35. After drying the embryos, cover them with some halocarbon oil. To start injecting, focus the embryos, then the needle. Insert the needle into the embryo, retract the tip quickly back into the very posterior region to ensure that the injection mixture mainly flows into that part. We use continuous flow while injecting. After one row is injected, turn the slide to inject the opposite row.

Post-injection procedures TIMING **~12 d**

36. After injection, place the embryo slides into humid chambers (a petri dish containing a Whatman paper soaked with water) and incubate the chambers at 18°C for two days.

37. After ~48 hrs collect crawling larvae with a thin wire: heap up ~20 larvae to a clump-like aggregate in the covering oil, then carefully reach with the wire under the clump and transfer larvae to the yeast plug of a fly food vial. We generally put 80-120 larvae into one vial.

TROUBLESHOOTING

38. Transfer the vials with collected larvae to 25°C. After about 10 days flies will start to eclose.

Identifying and establishing transgenic strains TIMING **~4 weeks**

39. Collect adult males from vials and discard the females. Cross two males with 3-4 *y*⁻*w*⁻ virgin females per vial (=G0 outcross; see Fig. 5).

CRITICAL STEP We use males in order to remove the ΦC31 integrase transgene, located on the X chromosome, with this and the subsequent F1 outcross. Note, here we use two males per cross to increase the rate of fertile crosses (with two males almost all crosses will produce offspring).

40. After 10-12 days screen the offspring for orange-eyed males. From each positively scored vial establish at least one outcross (=F1 outcross) of a single male with virgin females from an appropriate balancer strain (for our library we usually use: *y*⁻*w*⁻*; D gl3 /*

TM3 Sb Ser; see Fig. 5). Frequently, we set up two outcrosses as we can often acquire two different transgenic lines from these.

41. About 3 days after setting up the F1 outcrosses, isolate the single males from these vials while keeping the females in the crossing vials. Subject each single male to a single-fly PCR protocol[29] using the HL1-F/attP-R primer pair (**Box 1**).

Box 1: Single-fly PCR protocol

1. Mash single fly with pipette tip and 50µl SB (squishing buffer, see Reagent Setup).
2. Incubate at 37°C for 1 hour.
3. Inactivate Proteinase K by placing at 95°C for 5 min.
4. Set up the following PCR mix to amplify the barcode sequence region using the HL1-F/attP-R primer pair.

Component	Amount per reaction (µl)	Final
single-fly DNA template	1	
Fw + Rv primers (10 µM each)	0.5 each	250 nM
dNTPs (2 mM)	1.5	150 nM
GoTaq buffer (5x)	4	1x
GoTaq polymerase (5U/µl)	0.1	0.5U
H_2O	12.4	

5. Amplify using the following cycling conditions.

Cycle Number	Denature	Anneal	Extend	Hold
1	95°C, 2 min			
2-36	95°C, 30 sec	60°C, 30 sec	72°C, 40s	
37			72°C, 2 min	
38				12°C

(END OF BOX)

42. Subject PCR products to Sanger sequencing with the reverse primer attP-R, which binds downstream of the barcode sequence (see Fig. 2b).

43. After determining the exact barcode composition of the processed males, continue balancing only one fly strain per specific barcode. Recurring barcodes, which represent recurring transgenic lines, are discarded, i.e. we only process one strain per barcode (step 39).

CRITICAL STEP When working with 'high priority' ORFs, we recommend balancing at least two independently obtained transformants to ensure establishment of a homozygous transgenic strain. Occasionally, despite an identical transgene and an identical *attP* integration site, one strain cannot be made homozygous, whereas another can.

44. After the molecular determination of the transgenic flies and sorting out recurring strains, establish balanced stocks by sibling crosses between flies of the genotype w^+*/ TM3 Sb Ser* (w^+ indicating the UAS-ORF transgene) using a few flies per gender (see Fig. 5). After a few days these crosses are moved to 18°C, without deliberately establishing homozygous stocks.

(Optional) ORF confirmation of transgenic strains **TIMING ~2-3 d**

CRITICAL: In previous steps, we presume that a specific ORF is present in a strain based on the barcode sequence. To formally demonstrate this we need to amplify and sequence the corresponding ORF from the strain, though this is not obligatory.

45. Several generations after balanced stocks are established, take one fly per strain and subject it to a single-fly PCR protocol using the hsp-GW-F/HL1-R primer pair. The extension time for the polymerase is adjusted according to the assumed ORF length.

TROUBLESHOOTING

46. Subject PCR products to Sanger sequencing with the forward primer hsp-GW-F, which binds upstream of the ORF in the *hsp70* promoter region and allows sequencing into the 5' region of the ORF.

(Optional) Swapping the 3xHA tag for other tags **TIMING ~up to several weeks**

CRITICAL As an example of a further application of the UAS-ORF lines, we describe here the exchange of the standard 3xHA tag for an eGFP tag (Fig. 4b). However, the same procedure can be used to swap in other tags or promoters.

47. Cross 5-8 males carrying the ORF-3xHA transgene (homo- or heterozygous) with 10-15 virgins of the strain *TSeGFP-86Fb* (TS= tag swapping) per fly vial. The strain *TSeGFP-86Fb* provides the *eGFP* tag at 86Fb plus the necessary *hs*-FLP on the X chromosome.

48. After a period of egg-laying (~2 days), discard adults to avoid later overcrowding. You can also transfer them to a fresh vial and thus create a second round of "reagents".

49. Three days after setting up the crosses, place the vial into a 37 °C water bath and subject the larvae to a 40 min heat shock. You can repeat this heat shock treatment on the following day to likely increase efficiency of swapping.

50. Allow larvae to develop to adulthood, collect ~7 males from the offspring (y^+w^+) and mate them with 10-12 y^-w^- virgin females.

51. After egg-laying, transfer the flies to a fresh vial to generate a second screening vial.
CRITICAL STEP Generally, we find only a very few y^+w^+ males in a vial, thus we recommend to create at least two vials for screening.

52. Raise the progeny to adulthood and screen for y^+w^+ males. Cross a single male with 3rd chromosome balancer virgin females and subsequently establish a balanced or homozygous stock.

TROUBLESHOOTING

TROUBLESHOOTING

For troubleshooting of individual steps of PCR cloning (steps 9-13) and Gateway recombination (Steps 14-16), consult the troubleshooting sections in the corresponding manuals. Use gene expression information on FlyBase for selection of a suitable life-stage or tissue to maximise the likelihood of successful amplification of your gene of interest. If PCR remains unsuccessful, a synthetic ORF construct can be purchased. Helpful information on the fly procedures can be found elsewhere[25,27]. Further Troubleshooting advice can be found in Table 1.

Table 1. Troubleshooting

Step	Problem	Possible reason	Solution
37	Few surviving larvae	DNA is toxic	Use high-quality purification kit (e.g. Qiagen Midi kit)
		Chosen *attP* strain has low survival rate	Change to another *attP* line (Note, *ΦX-86Fb* has a high larval survival rate)
		Harmful injection	Prepare thinner needles
45	Single-fly PCR does not work	Very long ORF sequences	Use internal gene-specific primers to reduce the length of the PCR product
		Poor quality DNA	Purify genomic DNA from individual flies
52	No correct marker combination found	Heat shock conditions insufficient	Apply a longer heat shock or/and perform a second heat shock next day (step 49)
		Insufficient number of flies to screen	Increase number of vials in step 48 as well as in step 51

ANTICIPATED RESULTS

This procedure should produce viable flies with specific ORFs stably integrated. Cloning success when starting with individual cDNA clones is greater than 85% and mainly depends on the ORF length. The use of a high fidelity polymerase does not typically generate mutations. It is desirable to obtain specific PCR products without nonspecific bands (Fig. 3b). ORF cloning from pooled cDNA libraries has more variable success and may be dependent on variables such as the expression level of the gene of interest, the size of the gene or GC content. For sequencing library preparation, a time course with the dsDNA fragmentase is useful for optimal digestion (Fig. 3c). As the strain *ΦX-86Fb* exhibits very good overall fitness with high integration rates injections

should be straightforward. Approx. every second outcross is expected to lead to transgenic offspring, which can be easily scored due to the moderately strong *white* expression.

FIGURES

Figure 1| Flowchart for creating an UAS-ORFeome library.

The flowchart lists the main steps involved in cloning the ORFs into the barcoded Gateway destination vector pGW-HA.attB (top four boxes, steps 1-25) and in generating UAS-ORF transgenic fly strains (bottom four boxes, steps 26-46).

Figure 2| Schematic representation of the ORF library vector pGW-HA.attB and the ΦC31 integration.

(a) Gateway destination vector pGW-HA.attB. The vector contains the Gateway cassette with the genes *Cm* (Chloramphenicol) and *ccdB* (control of cell death B). The final expression vector is generated by replacing the Gateway cassette with a specific ORF sequence. The cassette is flanked by two differently altered *FRT* sites, *FRT5* and *FRT2*. The vector further contains a Gal4-inducible 5xUAS-*hsp70* promoter region (the *hsp70* fragment provides a TATA box and a transcriptional start site), a *loxP* site and a *white*$^{+}$ marker. The ORFs will be tagged by a triple HA epitope, which is separated from the gene region by a flexible linker (FL) region of 24 aa. The 3' UTR is a *tubulin alpha 1* trailer. Note, the illustration shows an already barcoded destination vector, i.e. the randomized barcode is inserted between the trailer and the following attB sequence.

(b) Integration of the UAS-ORF expression construct into an *attP* landing site. For our ORF library we use the fly strain *ΦX-86Fb*, which provides an *attP* integration site on the 3rd chromosome and the ΦC31 integrase on the X-chromosome. The *attP* site is marked with *DsRFP*, which is expressed in the eyes by the 3xP3 promoter. The *in vivo* situation after the integration is indicated in detail (except for the presence of Amp). The individual transgenic strains are determined by identifying the ORF-specific barcodes (the primer pair is indicated by red arrows).

Figure 3| ORF cloning and sequencing strategy.

(a) Illustration of the two-step PCR for amplification of ORFs using Act5C gene as an example. (b) Anticipated PCR results from ORF cloning. Example of eight different ORFs (1.2-1.5 kb) amplified using the two-step PCR strategy; a 5 μl aliquot of the final PCR product for each ORF was run on a 1.2% agarose gel. Each ORF is visible as single bright band without additional non-specific bands. Note that some genes may produce more than one specific band due to alternative transcripts.

(c) Fragmentation of plasmids for high-throughput sequencing. Time-scale of ORF entry clone plasmid pool digestion using dsDNA fragmentase enzyme mixture. In this case, 45 minute digestion yields ideal fragmentation of the plasmids, with the majority of the plasmid pool being fragmented into small molecular weight fragments.
(d) Strategy for high-throughput sequencing of ORFs. Individual ORF entry clones are pooled and fragmented followed by high-throughput sequencing library preparation. We prefer to use a "beads-in" protocol where paramagnetic beads used to purify the DNA are kept in the reaction mix to increase the final yield of the library.
(e) Illustration of the Illumina sequencing library preparation. Inclusion of barcoded sequencing adapters (optional) during library preparation allows multiplexing of sequencing libraries or association of different plasmid pools with specific plates or wells.

Figure 4| Schematic for N- and C-terminal *in vivo* swapping.
(a) All currently created ORF-3xHA strains will contain *FRT*s to exchange N- or C-terminal regions. Here we depict an N-terminal *in vivo* exchange of the UAS-*hsp70* promoter region for an *actin5C* promoter. The *actin5C* line (or promoter swapping line) provides additionally a heat shock-inducible Flipase (FLP, on the X chromosome) required for executing the recombination. The desired swapping event can be identified, after outcrossing, in the progeny by scoring for flies with y^+ only.
(b) The illustration depicts a C-terminal *in vivo* exchange of the 3xHA epitope for an *eGFP* tag. The *eGFP* tag line again provides a heat-inducible Flipase on the X chromosome. The desired swapping events can later be identified by scoring for flies with the co-occurrence of w^+y^+.
(c) Indicated are other currently existing N- and C-terminal swapping strains compatible with our ORF library. All these strains are equipped with a *hs-flp* transgene (not indicated). The 3xStop strain introduces stop codons immediately after the *FRT*, thus removing the 3xHA epitope and leaving the ORFs with only an *FRT* tag. The VNm9 and VC155 lines are used for BiFC analysis experiments (VNm9 and VC155 are the N- and C-terminal halves of the YFP variant VENUS[30]). The 2xTY1 represents an alternative epitope to the triple HA tag. This swapping line includes a TEV site, allowing for the enzymatic cleavage, which may be advantageous in purification procedures.

Figure 5| Crossing scheme for establishing UAS-ORF strains.
Two injected males are crossed with several y^-w^- females because some males are sterile due to the injection procedure. Use of males also removes the ΦC31 integrase. In the F1 outcross clonality is obtained by crossing one orange-eyed male with an appropriate balancer strain (i.e, TM3 Sb Ser). To produce the final stock, UAS-ORF balanced siblings are crossed *inter se*.

Acknowledgements

We gratefully acknowledge Jussi Taipale, who co-initiated the pilot library project. We thank E. Furger and C. Schertel for substantial contributions to molecular cloning, library construction and overexpression analysis, C. Bastos, E. Escher, A. McLeod, S. Miettinen, and N. Wang for technical assistance, J.-P. Vincent for *FRT* details, R. Baumgartner for BiFC reagents, and K. Hens and B. Deplancke for some specific pGW-HA.attB expression clones.

This work was supported by the National Center of Competence in Research "Frontiers in Genetics", the Swiss National Science Foundation, the Kanton of Zürich, the European Research Council and the Scottish Universities Life Sciences Alliance and from the Biotechnology and Biological Sciences Research Council (BB/J006424/1).

Author Contributions

K.B. conceived and coordinated the project, M.B. coordinated and supervised the molecular cloning and data analysis, and J.B. coordinated and supervised the creation of the fly library. The experiments were performed by J.B., E.S. and M.B., who together wrote the paper.

References

1. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, Couto A, Marra V, Keleman K, Dickson BJ. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448:151-6.

2. Miklos GL, Rubin GM. (1996) The role of the genome project in determining gene function: insights from model organisms. Cell 86(4):521-9.

3. Rorth P. (1996) A modular misexpression screen in Drosophila detecting tissue-specific phenotypes. Proc Natl Acad Sci USA 93:12418-12422.

4. Toba G, Ohsako T, Miyata N, Ohtsuka T, Seong KH, Aigaki T. (1999) The gene search system. A method for efficient detection and rapid molecular identification of genes in Drosophila melanogaster. Genetics 151(2):725-37.

5. Bellen HJ. et al. (2011). The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics 188,731-743.

6. Xu R, Deng K, Zhu Y, Wu Y, Ren J, Wan M, Zhao S, Wu X, Han M, Zhuang Y, Xu T. (2008) A large-scale functional approach to uncover human genes and pathways in Drosophila. Cell Res.18(11):1114-27.

7. Bischof J, Björklund M, Furger E, Schertel C, Taipale J, Basler K. (2013) A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development 140(11):2434-42.

8. Schertel C et al. (2013) Systematic screening of a Drosophila ORF library in vivo uncovers Wnt/Wg pathway components. Dev Cell 25(2):207-19.

9. Taxis C, Stier G, Spadaccini R, Knop M. (2009) Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. Mol Syst Biol. 5:267.

10. Hudry B, Viala S, Graba Y, Merabet S. (2011) Visualization of protein interactions in living Drosophila embryos by the bimolecular fluorescence complementation assay. BMC Biol. 9:5.

11. Prelich, G. (2012). Gene overexpression: uses, mechanisms, and interpretation. Genetics 190, 841-854.

12. Gibson TJ, Seiler M, Veitia RA. (2013) The transience of transient overexpression. Nat Methods. 10(8):715-21.

13. Venken, K.J., Simpson, J.H. and Bellen, H.J. (2011). Genetic manipulation of genes and cells in the nervous system of the fruit fly. Neuron 72, 202-30.

14. Brown JB et al. (2014) Diversity and dynamics of the Drosophila transcriptome. doi:10.1038/nature12962.

15. Yu C, Wan KH, Hammonds AS, Stapleton M, Carlson JW, Celniker SE. (2011) Development of expression-ready constructs for generation of proteomics libraries. Methods Mol Biol. 723:257-72.

16. Cavener DR. (1987) Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucl Acid Res. 15/4, 1353-1361.

17. DeAngelis MM, Wang DG, Hawkins TL. (1995) Solid-phase reversible immobilization for the

isolation of PCR products. Nucleic Acids Res. 23(22):4742-3.

18. Massouras A, Decouttere F, Hens K, Deplancke B. (2010) WebPrInSeS: automated full-length clone sequence identification and verification using high-throughput sequencing data. Nucleic Acids Res 38:W378-84.

19. Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, Uhlén M, Simpson JC, Lundberg E. (2013) Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nat Methods. 10(4):315-23.

20. Giaever G. et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418(6896): 387-91.

21. Groth AC, Fish M, Nusse R, Calos MP. (2004) Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics 166:1775-1782.

22. Bateman JR, Lee AM, Wu CT. (2006) Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. Genetics 173:769-777.

23. Venken KJ, He Y, Hoskins RA, Bellen HJ. (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science 314(5806):1747-51.

24. Bischof, J, Maeda, RK, Hediger, M, Karch, F & Basler, K. (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A 104(9):3312-3317.

25. Fish MP, Groth AC, Calos MP, Nusse R. (2007) Creating transgenic Drosophila by microinjecting the site-specific phiC31 integrase mRNA and a transgene-containing donor plasmid. Nat Protoc 2(10):2325-31.

26. Schlake T, Bode J. (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry 33(43):12746-51.

27. Dahman C. (2008) Drosophila. Methods and Protocols. Vol. 420 (Humana Press).

28. Meyer M, Kircher M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010(6):pdb.prot5448.

29. Gloor GB, Preston CR, Johnson-Schlitz DM, Nassif NA, Phillis RW, Benz WK, Robertson HM, Engels WR. (1993) Type I repressors of P element mobility. Genetics 135(1):81-95.

30. Saka Y, Hagemann AI, Piepenburg O, Smith JC. (2007). Nuclear accumulation of Smad complexes occurs only after the midblastula transition in Xenopus. Development 134: 4209-18.