Genome-wide mapping and characterisation of protein expression and interaction in *Drosophila melanogaster,* using a hybrid piggyBac/P-element YFP gene trap system with tandem affinity tags.

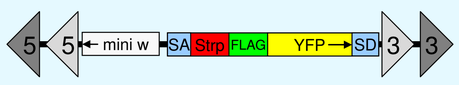
E. Ryder[1], H. Spriggs[1], G. Johnson[1], E. Drummond[1], J. Drummond[1], J. Webster[1], J. Roote[1], N. Lowe[2], K. Lilley[3], S. Hester[3], J. Howard[3], J. Rees[3], S. Russell[1,3], D. St. Johnston[2].

1. Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH. UK. 2. The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2. 1QN.
3. Cambridge Systems Biology Centre, Tennis Court Road, Cambridge, CB2 1QR

We have initiated a screen to generate and characterise protein trap lines in *Drosophila* using a piggyBac transposon-based strategy. The ability to generate *in vivo* tagged proteins has tremendous potential for furthering our understanding of developmental processes by allowing the characterisation of sub-cellular protein localisation and facilitating the isolation of multi-protein complexes. This is a large project involving a collaboration with over thirty UK laboratories.

The piggyBac protein tag construct

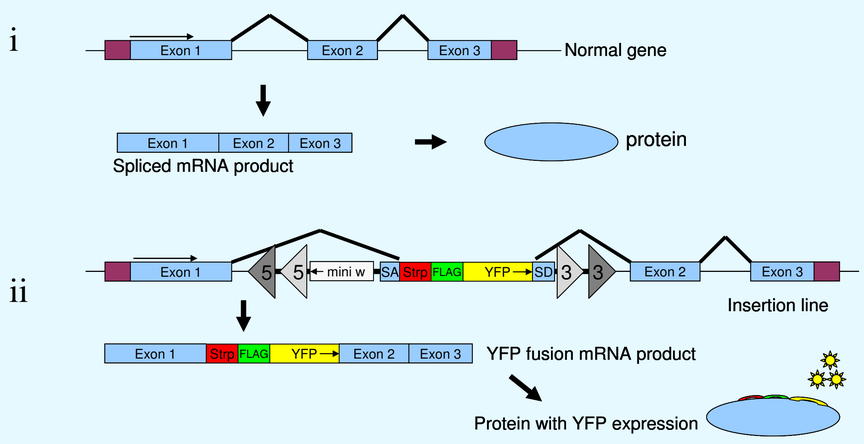
Figure 1. piggyBac construct



- piggyBac element to maximise insertions into introns.
- Internal P-element ends for future gene disruption and P replacement experiments.
- Mini-white gene for tracking the element in stocks.
- Protein tag cassette containing splice acceptor and donor sites, two affinity purification tags (StrepII and FLAG) and a functional YFP exon.
- Three versions: one in each reading frame for maximum potential gene coverage.

The creation of YFP-gene fusions is summarised in Figure 2. In normal mRNA production (2i) the gene is transcribed and the introns spliced out before translation. The YFP construct contains splice acceptor (SA) and splice donor (SD) sites which incorporates it into the spliced mRNA product. If the piggyBac element transposes into the intron of a gene in the correct orientation and the correct frame a functional YFP fusion will be created (2ii) which can be detected under a fluorescence microscope.

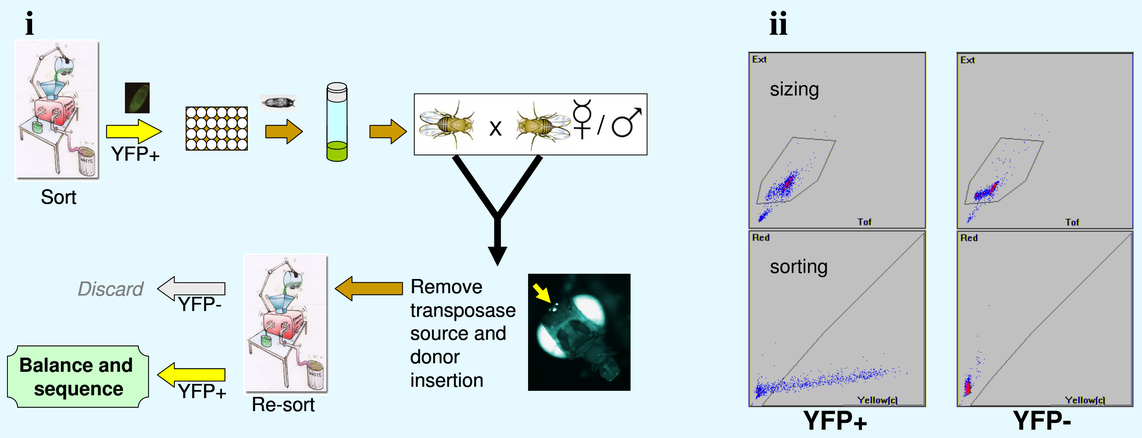
Figure 2. Production of YFP-gene fusions



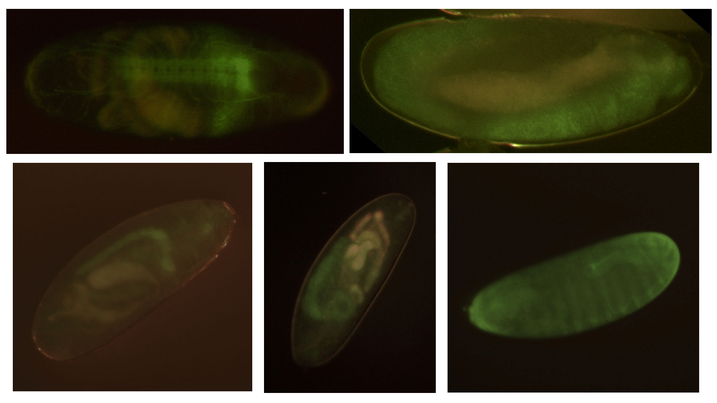
Recovery of YFP fusion stocks

Recovery of YFP fusion lines is summarised in Figure 3i. For the initial screen, stocks containing the transposase source and donor element are setup in cages and approximately 250,000 embryos collected. These are then analysed for any YFP signal using an embryo sorter (Union Biometrica), and putative positives dispensed into a 24well apple-agar plate. After transfer to a standard tube and media during pupation they are then crossed with w^- males or w^- virgin females. Any transposase source or donor insert chromosomes present in the stock are removed (the transposase is tagged with Pax-3 promoter CFP and can be seen in the ocelli; the donor element is on a marked chromosome) and the lines resorted individually (Figure 3ii). Those lines still expressing YFP are then balanced and sequenced. Examples of expression patterns observed are shown in Figure 3iii

Figure 3. Isolation of YFP trap stocks.



iii. Examples of expression patterns observed



Sequence mapping of insertions

DNA from fifteen adults from a new insertion line is isolated and the flanking regions of the piggyBac element amplified via inverse PCR. Purified products are sequenced with dye terminator v3.1 chemistry (ABI) and visualised with an ABI3100 automated sequencer. Analysed sequences are mapped on to the *Drosophila* genome using BLAST*n* and processed using custom software developed at Cambridge University. Examples of genes trapped are shown in Table 1.

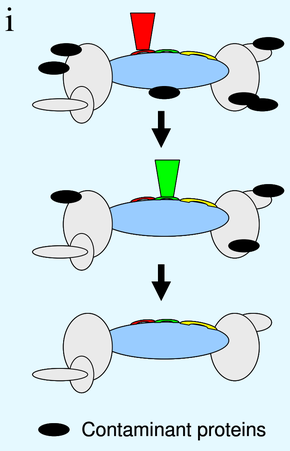
Table 1. Examples of gene trap lines

Stock	Loc	Chr	Gene CG	Gene Name	Feature ID	YFP Status
CPTI-000023	2507266	X	CG2621	sgg	intron_CG2621:12_CG2621:3	Confirmed - Yes
CPTI-000030	27691116	3R	CG31000	heph	intron_CG31000:8_CG31000:9	Confirmed - Yes
CPTI-000031	13555967	2L	CG7147	kuz	intron_CG7147:2_CG7147:3	Confirmed - Yes
CPTI-000037	18316849	3R	CG5374	T-cp1	intron_CG5374:2_CG5374:3	Confirmed - Yes
CPTI-000056	7584482	3R	CG17342	Lk6	intron_CG17342:1_CG17342:2	Confirmed - Yes
CPTI-000076	259348	3R	CG9805	eIF3-S10	intron_CG9805:1_CG9805:2	Confirmed - Yes
CPTI-000077	13114793	X	CG1770	HDAC4	intron_CG1770:14_CG1770:15	Confirmed - Yes
CPTI-000091	3633496	2L	CG10033	for	intron_CG10033:3_CG10033:4	Confirmed - Yes
CPTI-000106	19838059	3L	CG8103	Mi-2	intron_CG8103:1_CG8103:2	Confirmed - Yes
CPTI-000110	3629601	2L	CG10033	for	intron_CG10033:3_CG10033:4	Confirmed - Yes
CPTI-000130	15003094	2L	CG4140	CG4140	CG4140-RA-in	Confirmed - Yes
CPTI-000155	17678593	3R	CG6575	glec	CG6575-RA-in	Confirmed - Yes

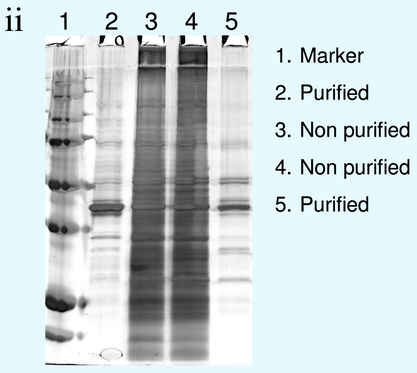
Protein complex purification and analysis

The dual affinity tags in the piggyBac construct allow for the purification and detailed analysis of protein complexes associated with the YFP fusion protein (Figure 4i). Analysis of complexes are performed using liquid chromatography-mass spectrometry (LC-MS), on either whole complexes or sub-complexes separated using SDS-PAGE (Figure 4ii). Analysis is performed using the Mascot search engine on the *Drosophila* transcriptome (Figure 5)

Figure 4. Protein complex purification

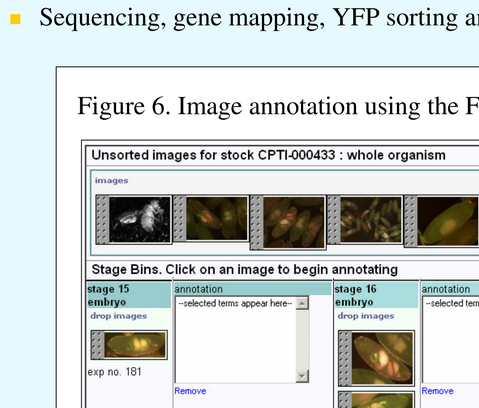


- Bind protein complex to a Strep-Tactin column with the **Strep tag II** site.
- Elute with addition of desthiobiotin.
- The **FLAG tag** can be purified using an M2 monoclonal antibody.
- The purified protein is now ready for analysis by LC-MS



Raw vs. purified protein extraction (coomassie stained SDS-PAGE gel)

Figure 5. Mascot analysis of purified protein complexes



Analysis performed on earlier version of the construct which contains a GFP tag instead of YFP

Annotation of expression patterns

To aid in the analysis of YFP-trap lines, we have written software (The Flannotator) which allows annotation of gene expression at all stages of development and all tissue types (including sub cellular location) using the standard *Drosophila* anatomy controlled vocabulary and gene ontology (Figure 6).

- Annotation of gene expression at all stages of development and tissue types (including sub cellular location).
- Each user can customise their annotation tools so they only see what is relevant to them.
- Uses the *Drosophila* anatomy controlled vocabulary and gene ontology to ensure data integrity.
- Menus and tick-boxes remove all manual input apart from comments.
- The web-based input and retrieval system allows multiple groups to work in collaboration, whilst still protecting the original data.
- Stock management (with full history)
- Sequencing, gene mapping, YFP sorting and affinity tag purification data available as a stock report

Figure 6. Image annotation using the Flannotator