Supplementary Information

Supplementary Results

An approach to sorting nuclei based on the *Ubx* expression pattern

To sort embryonic nuclei, we created a 13 kb construct containing the *Ubx* promoter, several embryonic and larval enhancers and the *bxd* ME. *GFP,* with a nuclear localization signal (NLS), was cloned in frame at the 5' end of the *Ubx* ORF. This construct is essentially the same as those used in previous studies (Tillib et al., 1999), except that the *lacZ* gene is replaced by *NLS-GFP* (Figure S1A). A number of transgenic fly lines were generated and tested. We chose a line that showed a high level of specific *GFP* expression for further analysis. Expression of *GFP* was detected in nuclei, consistent with the presence of the NLS in our construct. The expression patterns of GFP protein in transgenic embryos mimic the expression patterns of endogenous *Ubx* (Figure 5A), with the exception of parasegments 5 and 6. In these parasegments, the expression of the *GFP* reporter gene is lower than that of endogenous *Ubx,* since the *abx/bx* regulatory region is not present in this construct (Tillib et al., 1999).

In sorting experiments, we successfully used nuclei prepared 7-9, 10-13, or 14-16 hr embryos. Using these nuclei, we developed appropriate conditions for using a Coulter cell sorter. We can obtain over one million *GFP*-positive and *GFP*-negative nuclei in each sorting experiment, sufficient for a number of biochemical tests, including the chromatin immunoprecipitation (ChIP) assay. Although nuclei are kept at 4°C, and sorting is usually complete in 1 hr, it was important to show that sorting did not change chromatin or nuclear integrity. We first determined the amounts of Trx and dCBP, both components of TAC1, in the nuclei before and after sorting. Figure S1D shows that the amounts of these proteins are essentially the same, suggesting that chromatin-associated transcription factors are not

significantly lost during sorting. To show that the sorting procedure did not disturb chromatin structure, leading to decreased association of proteins with chromatin, we tested the amounts of Trx and modified histone H3 (H3-meK4 and AcH3) associated with the promoter region of *Ubx* in freshly isolated nuclei, nuclei which were kept on ice for the approximate duration of the sorting procedure, and nuclei after sorting. Figure 1D shows that there is no difference in the amounts of these proteins, suggesting that neither the length of the procedure, nor the potential mechanical disturbance during sorting interfere with stable association of chromosomal proteins. Furthermore, with ChIP analysis of the sorted nuclei, we detect TAC1 components, as well as histones, with an efficiency that is similar to that seen in analysis where whole embryos are used (Figures 6,7). Together our control experiments suggest that chromatin structure and chromatin-associated regulatory complexes are not significantly altered during this procedure.

An important quality control for the sorting procedure was the analysis of *GFP* and *Ubx* RNA expression levels in isolated nuclei by RT-PCR to determine the relative enrichment obtained after nuclear sorting. Figure S1C shows that expression of *GFP* and endogenous *Ubx* RNA is about 10 and 16 times higher, respectively, in the *GFP*-positive compared to the *GFP*-negative nuclei. The enrichment of *GFP* is expected to be slightly higher than that of *Ubx* because of the lower expression of *GFP* in PS 5, which can lead to some contamination of the *Ubx*– pool by *Ubx*+ nuclei. This level of enrichment is reproducible, and provides sufficient purification to reliably detect differences between the structure of the activated and silenced *Ubx* gene in embryos. These results indicate that sorted nuclei can be used to detect chromatin-associated proteins, to detect RNA by RT-PCR, and for expression profiling.

Supplementary Experimental procedures

Drosophila Genetics and clone construction

The following mutant strains were used: *trx*^{B11} (Mazo et al., 1990); *spt16* (Shimojima et al., 2003); *spt4*, an insertion in *Dspt4* from the EP project – EP(2)2604. *Spt6:* Cytological and Southern blot analysis of the 5EF region indicated that the Dspt6 gene is covered by the Df(1)5D deficiency and the Dp(1;f)q[2] duplication (data not shown). This duplication and this deficiency were used to screen a collection of X-linked EMS-induced lethal mutations, generated and kindly provided by S. Noselli, CNRS, Nice. 23 lethal mutations that do not complement Df(1)5D but are complemented by the Dp(1;f)q[2] duplication were isolated. These mutations were rescreened for complementation by a small autosomal Dspt6+ duplication that was generated by P-element transformation of cosmid CosPeR 37-4. This cosmid was isolated by screening a CosPeR library (generated by C. Tamkun and kindly provided by M. Bourouis, CNRS, Nice) with a Dspt6 cDNA probe (LD3201). A single complementation group comprising two alleles, noted D1-10 and B1-26, were complemented by CosPeR 37-4. Genomic sequencing of the *Spt6* gene revealed that a 2 nucleotide CG deletion is associated with the B1-26 allele, leading to a frame shift and a premature stop codon (the sequence of the predicted mutant Dspt6 protein is, starting at R1397: RGEGKstop).

To construct *trx RNAi* transgenic line, a Trx cDNA fragment corresponding to nt 2665-3059 (M31617) was inserted into the pUAST vector in direct and inverted orientation separated by the 60 bp spacer. The *UAS-Trx* transgenic lines were obtained using standard procedures. To induce Trx RNAi we crossed the UAS-Trx transgenic line to the driver line that ubiquitously expresses GAL4 under the Hsp70 promoter (Bloomington Stock Center). The progeny of this cross survive till the late third instar stage and die at early pupal stages. Alternatively, the UAS Trx line was crossed to the driver line with conventional hsp70 promoter, and Trx RNAi was

induced by placing these larvae for 30 min at 37^0C during the second and the early third instar stages; this treatment leads to the complete lethality of the adults.

The construct for nuclei sorting was generated by cloning *GFP* into the Pst I site (1138 nt of the *Ubx* transcription start site) of the previously described N construct (Tillib et al., 1999). The *GFP* sequence was preceded by the sequence of the nuclear localization signal PKKKRKVE.

Preparation of nuclei

For each sorting experiment, 1-2 g of 7-13 hr embryos were collected, chorions were removed by treatment with 50 % Chlorox bleach for 2 min, followed by a thorough wash. Embryos were resuspended in buffer A (0.02M HEPES, pH7.3, 5mM $MgCl_2$, 60mM KCl, 15mM NaCl, 0.5mM EDTA, 0.01% NP40, protease inhibitor cocktail; Roche), transferred to a glass homogenizer (approximately 0.05g-0.2g of embryos per 1 ml of buffer A) and disrupted by 5 strokes of pestle type A followed by 10 strokes with pestle type B. To remove cellular debris, homogenate was centrifuged for 1 min at 500g, supernatant was transferred to another tube and nuclei were collected by centrifugation for 5-7min at 2,300g. The nuclear pellet was resuspended in 5 ml of buffer A, and loaded over 5 ml of buffer B (the same as buffer A with 0.3M sucrose), and centrifuged for 15 min at 2,300g. A hard yellow pellet at the bottom contains mostly yolk protein, while nuclei form a looser pellet at the top. The loose pellet was resuspended in 3 ml of Buffer C (0.02M HEPES, pH7.3, 0.15mM NaCl, 5mM KCl, 5mM $MgCl_2$, 0.5mM EDTA, 0.1M sucrose, 0.02% NP40, protease inhibitor cocktail) avoiding dispersion of the hard pellet. The resuspended pellet was centrifuged again for 1 min at 500g to eliminate remains of hard pellet. The supernatant was carefully collected, and the procedure was repeated if there was a significant amount of hard pellet.

Chromatin Immunoprecipitation

For ChIP experiments, embryos were collected, dechorionated manually and placed in *Drosophila*-SFM (Invitrogen, Carlsbad, CA) containing 0.1% Tween-20. Embryos were fixed in 2ml of fixing solution (50mM HEPES, pH7.6, 100mM NaCl, 0.1mM EDTA, 0.5mMEGTA, 1.8% formaldehyde) with 2ml of heptane for 15 min at room temperature. After fixation, embryos were washed several times in PBT and resuspended in the SDS Lysis buffer (Upstate Biotechnology). For ChIP experiments, isolated nuclei were fixed by incubation with 1.8% formaldehyde for 10 min, washed with PBS and resuspended in the SDS Lysis buffer. All subsequent steps were performed as described in Upstate Biotechnology protocol.

Before antibody precipitation, 5% of the chromatin solution was removed, crosslinks were reversed, and serial dilutions of purified DNA were analysed by PCR to ensure that equivalent amounts of starting material were used for each precipitation. Starting material was then divided equally for each antibody used (and no-antibody control) so that each sample contained the amounts of material indicated above. ChIP experiments were performed at least 3 times using the following antibodies: Polyclonal acetyl-histone H3 and polyclonal dimethyl-histone H3 (Lys 4), obtained from Upstate Biotechnology; Polyclonal rabbit antibodies against components of TAC1 (Trx, and Sbf1) and Spt16 were as described previously (Petruk et al., 2001; Shimojima et al., 2003). Immunoprecipitated chromatin was then assayed by semi-quantitative PCR using primers to sequences at the *Ubx* locus and the coding regions of *GFP* and *lacZ*. Approximately 1-5% of the immunoprecipitated material was used for PCR amplification. The efficiencies of PCR primers were tested by using genomic DNA from embryos.

Primers for RT-PCR

For testing of nuclei after sorting, PCR primers from the transcription start site were as follows:

rp49: +12 to +393 (5'-cccagcatacaggcccaagatc-3' and 5'-gaacgcaggcgaccgttgg-3');

GFP: +123 to +460 (5'-aagctgaccctgaagttcatc-3' and 5'-tgatatagacgttgtggctg-3');

Ubx : +2062 to +2260 (5'-agacatacacccgctaccag-3' and 5'-tgcttctcctgttcgttcag-3').

RT reaction to test the presence of non-coding RNAs in the vicinity of *Ubx* promoter was performed using primer specific to the proximal upstream regulatory region of *Ubx*:

5'-tgctagaatacggcttacgcaa-3' 242481 (U31961).

The sequence of PCR primers from the transcription start site were as follows:

5'-Ubx: +156 to +440 (5'-agattcgaggaaatccgtcag-3' and 5'-gtttgtacggtttcacaagtg-3');

3'-Ubx: +1798 to +2187 (5'-tggaatgccaattgcaccatctc-3' and 5'-agatcttgatctgccgctccgt-3').

Coordinates of primer sets for the *bxd* region of the *Ubx* locus is based on the sequence U31961 (from proximal to distal with respect to transcription start site):

exon 1: 5-'ctcagtatgagttccagttcc-3' 210247 and 5'-cttggcaagcacatctttaac-3' 210579;

exon 4: 5' tgcaccacagttagcactgat-3' 226109 and 5'-tatatcgcaacagatggcagt-3' 225599 ;

exon 5: 5'-tctggctgttggctgattgttg-3' 230923 and 5'-tccatcaatgaggcaatga-3' 231599;

exon 7: 5'- acattatcacttcccattggac-3' 232587 and 5'-tgagttatcgtcttgcggag-3' 232950;

exon 8: 5'- agcacgattccggtcataattc-3' 234538 and 5'- agtatggttcgcagcaaatgtc-3' 234840.

Bxd: 5'- taagtgcggtgataaggtcca-3' 218799 and 5'-agagatacggttagccttctc-3' 219279.

1: 5'- actgctttgcacacaaacac-3' 235393 and 5'-tctggcgatgaagtgcaattg-3' 235721;

2: 5'-ttcgcaattgttgagccatc-3' 236648 and 5'-attgccattgcagcttgac-3' 235989;

3: 5'- ttcgacttgcactgacacatg-3' 237862 and 5'-attgcaacggcaaatgcaacg-3' 238240;

4: 5'-taggaaactccaagcacttg-3' 239732 and 5'-agtaccagagtatactagttgc-3' 240050;

5: 5'-acttggacaggtgtgagctttgc-3' 240484 and 5'- ttctgcattcccgagttggctg-3' 240985

U: 5’acgaacactcaagagagagcg -3' 242712 and 5’-ttgccttggcactatcaaag -3' 243203;

Primers for ChIP analysis

Coordinates of the primer sets for the *bxd* region of the *Ubx* locus are given based on the sequence U31961 (from proximal to distal with respect to transcription start site):

5’-atgattgggttccttcttagc-3' 220275 and 5’-agcaaagagggtcaggatat-3' 220771;

5’-aaagagcgtgcgcatctctc-3' 219603 and 5’-attgttgctgttgtgggggc-3' 220101;

5’-taagtgcggtgataaggtcca-3' 218799 and 5’-ttacgaacgacagttatggcga-3' 219346;

5’-tatatcggatcgcttatggcg-3' 217669 and 5’-aaactgtcacggtttggggac-3' 218176;

5’-acagatagcaccggtacacg-3' 217100 and 5’-ttcggttctccagatactcc-3' 217538;

5’-cacatgcacatccagaagatg-3' 216306 and 5'-gatgcacacattccgaacgtg-3' 216840;

5'-catacatactgacagaggttg-3' 215264 and 5'-tgcgtcaggcttcatcgataatg-3' 215795.

Coordinates of the primer sets for the transcribed region of *Ubx* are given relative to the transcription start site:

Promoter: 5’acgaacactcaagagagagcg -3' –158 and 5’-ttgccttggcactatcaaag -3' +340;

‘2 kb downstream’: 5’-gacacaacaccttattggca-3' +2038 and 5’-agcgaataaactatggccac-3'+2380.

Coordinates of the primer sets inside the *Ubx* transcription unit are given based on the sequence U31961:

‘middle’: 5'-taccagactggctaggtaag-3' 266720 and 5'-caaagaagctcggttcaatgga-3' 267075;

‘3’-end’: 5'- agacatacacccgctaccag-3' 318254 and 5'-ctttgtatctgtctccgcgtc-3' 318603.

Coordinates of the primer set is given relative to the beginning of the coding region of *GFP*:

GFP: 5'- aagctgaccctgaagttcatc-3' +123 and 5'-ttctcgttggggtctttgc-3' +600.

Coordinates of probes for *in situ* hybridization

The coordinates of Ubx probe are relative to the transcription start site:

+808 to +1679 (5'-aggaacagcacagaaagcg-3' and 5'-gaatgtgtgattgctggcct-3').

The coordinates of the intergenic RNA probes are based on sequence U31961:

Exons 1-7: from 210423 in exon 1 to 232950 in exon 7: (5'- agtgtagtgtatcctatgcaacga-3' and 5' tgagttatcgtcttgcggag-3');

Exons 1-8: from 210423 in exon 1 to 234840 in exon 8: (5'- agtgtagtgtatcctatgcaacga-3' and 5'-agtatggttcgcagcaaatgtc-3');

Bxd: from 218330 to 220101: (5'- actgaaactgaatcgcacaa-3' and 5'-attgttgctgttgtgggggc-3').

Exons 4, 5 and 7: the coordinates of the probes are the same as primer sets used for RT-PCR (see above).

Probe 5: from 240484 to 241554: (5'-acttggacaggtgtgagctttgc-3' and 5'-gtatgtggtggtgcgtatctgt-3').

Supplementary Figures

Figure S1. Sorting of embryonic nuclei according to the *Ubx* expression pattern.

(A) Schematic presentation of the *Ubx* promoter and *bxd* region that contains a number of previously mapped TREs and PREs (top). The construct used to create a transgenic fly line used for nuclei sorting (bottom), nls, nuclear localization signal.

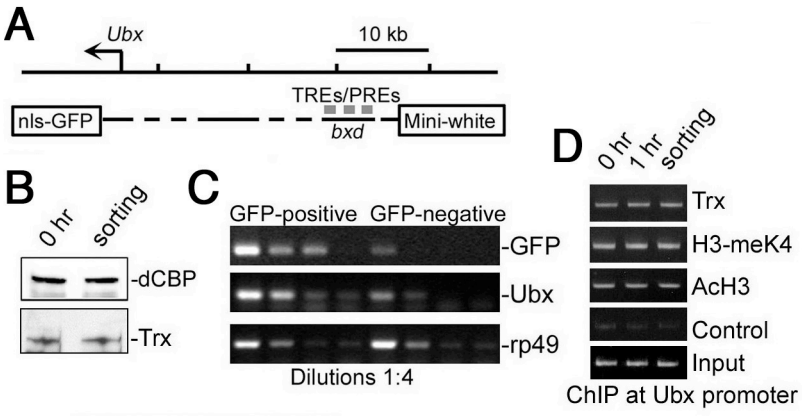
(B) Western blot analysis of the amounts of the dCBP and Trx proteins in embryonic nuclei before (0 hr) and after sorting.

(C) Analysis of the amounts of *GFP, Ubx* and *rp49* transcripts by RT-PCR in the RNA prepared from the GFP-positive and GFP-negative nuclei after sorting.

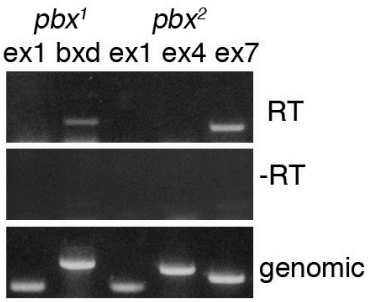
(D) ChIP analysis of association of Trx, H3-meK4 and AcH3 at the promoter of *Ubx* (the primer set is the same as shown for the *Ubx* promoter in Figure 6B) in freshly isolated nuclei (0 hr), nuclei kept at 4^0C for 1 hr, and nuclei after sorting. Control, no antibody.

Figure S2. Analysis of the ex1, bxd, ex4 and ex7 intergenic transcripts in the pbx^1 and pbx^2 homozygous mutant embryos. For coordinates of primer sets, see Supplementary experimental procedures.

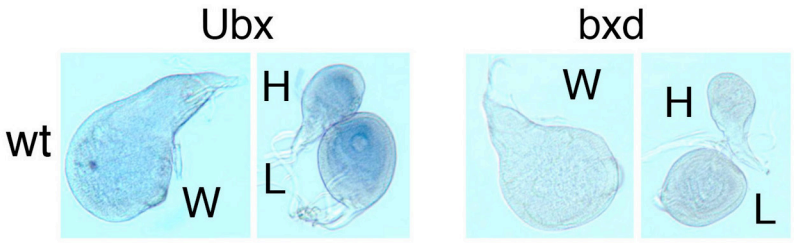
Figure S3. Detection of the *Ubx* and *bxd* transcripts in the wing (W), 3rd leg (L) and haltere (H) imaginal discs from the wild-type larvae by *in situ* hybridization. Probes are the same as in Figure 4B.



Figure_S1. Petruk et al.



Figure_S2. Petruk et al.



Figure_S3. Petruk et al.