SUPPLEMENTAL INFORMATION

Supplementary Figures, Legends, and Tables.

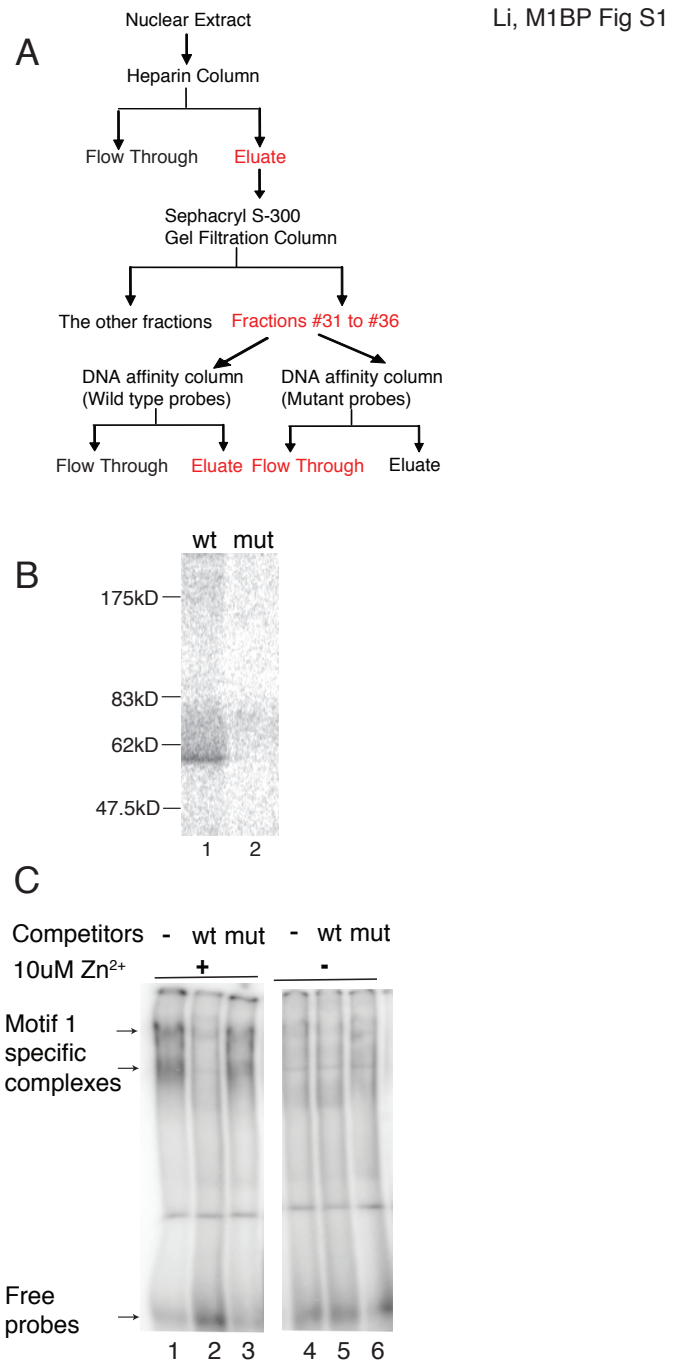

Figure S1. Purification and identification of M1BP

(A) Scheme for M1BP purification. Fractions containing Motif 1 binding activity are labeled in red.

(B) UV-crosslinking analysis of the eluate from the Motif 1 DNA affinity column. A protein with an apparent size of 55kD was specifically crosslinked to the radiolabeled wild-type (wt) Motif 1 (lane 1) but not the mutant (lane 2).

(C) Gel-shift assay with recombinant M1BP isolated from *E. coli*. The radiolabeled probe contains the Motif 1 consensus sequence from the Smo gene. The upper shifted bands may be due to dimerization of M1BP mediated by the ZAD domain.

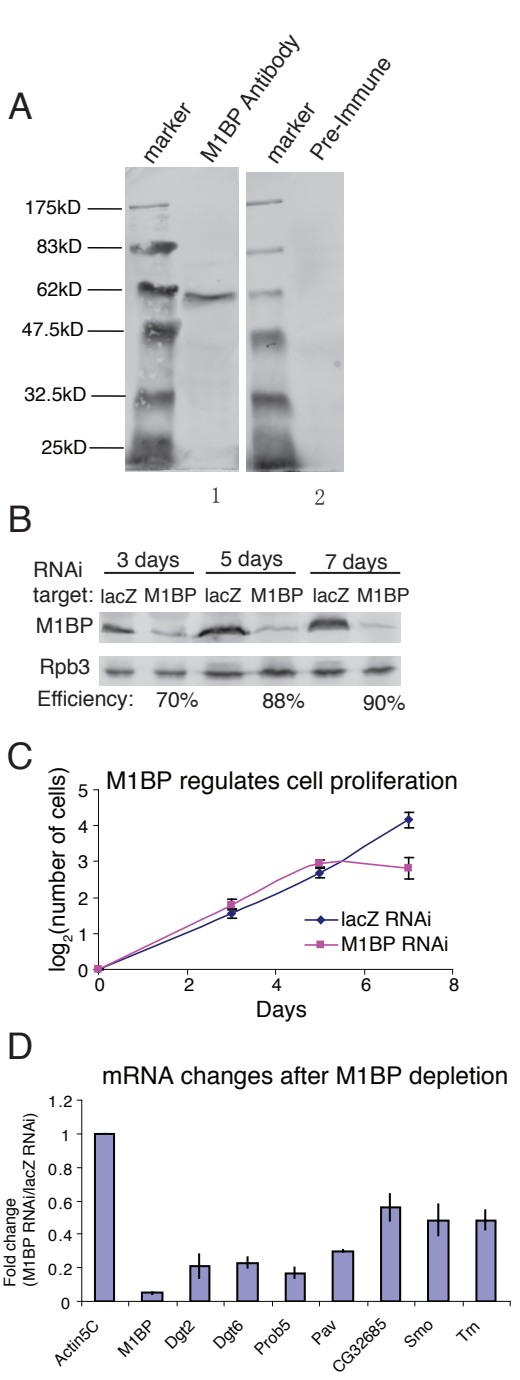

Figure S2. Functional analyses of M1BP

(A) Western blot showing the specificity of M1BP antibody. Sixty micrograms of embryo nuclear extract was first probed with M1BP anti-sera (1:5000, lane 1) or pre-immune sera

(1:2000, lane 2), and then with Cy5-conjugated anti-rabbit secondary antibodies (Invitrogen). The blot was visualized with a Typhoon 8600 (GE Healthcare).

(B) Western blots showing M1BP depletion at different times after RNAi treatment. The Rpb3 subunit of Pol II served as a loading control.

(C) Cell counts after various times of M1BP RNAi or control (lacZ) RNAi treatment. Error bars represent the standard deviation of biological triplicates.

(D) Expression of Motif 1 genes decreases after M1BP depletion. mRNA levels of Motif 1 genes were monitored by reverse-transcription followed by real-time PCR with gene specific primers. Results from different experiments were normalized to Actin5C, which lacks Motif 1. The expression changes caused by M1BP depletion are shown as the ratio of the mRNA level in M1BP RNAi-treated cells to lacZ RNAi-treated cells. Consistent with the microarray analysis, Dgt2, Dgt6, Prob5 and Pav show decreased expression. CG32685, smo and Trn did not show significant changes in the microarray analysis but the greater sensitivity provided by real-time PCR detected decreases of mRNA. Error bars represent the standard deviation of three independent RNAi experiments.

Li, M1BP FigS3

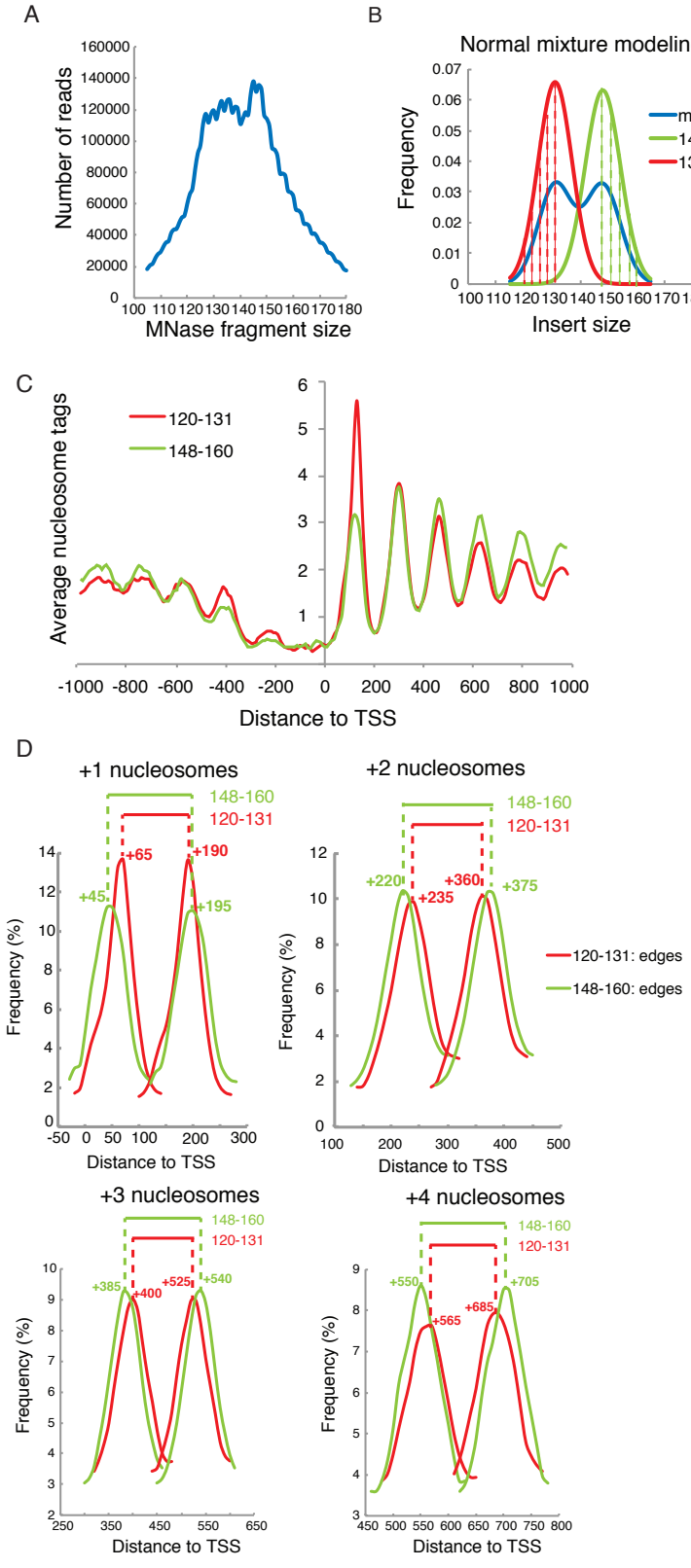

Figure S3. Asymmetric MNase digestion of the +1 nucleosome

(A) Size-distribution of MNase fragments derived from paired end reads that map within 1kb of the TSS of active promoters. This distribution suggests that at least two populations of nucleosomes with distinct sizes of MNase digested fragments are evident near promoters.

(B) Normal mixture modeling according to the distribution of fragment sizes in panel A. Green and red lines are the simulated size distribution of MNase-seq reads representing normal (148) and altered nucleosomes (131), respectively. The blue line simulates the distribution of the mixture of the two. The green and red shaded areas highlight the populations of reads that were selected to represent the normal and altered nucleosomes in further analyses. The size ranges of the selected reads were the mean to the mean + 2σ (148, 160) for normal nucleosomes and the mean - 2σ to the mean (120,131) for altered nucleosomes.

(C) Distribution of normal (red line) and altered nucleosomes (green line) at GAF-less, active genes. Altered nucleosomes with the smaller size of MNase digested fragments are enriched at the +1 position.

(D) Composite plots of each border of normal (148 to 160) and altered (120 to 131) nucleosomes at the +1, +2, +3 or +4 positions of GAF-less, active genes. These plots suggest the altered nucleosomes enriched at the +1 position are due to the preferential digestion by MNase of the first 20 bp DNA at the upstream border of the +1 nucleosome.

Li, M1BP Fig S4

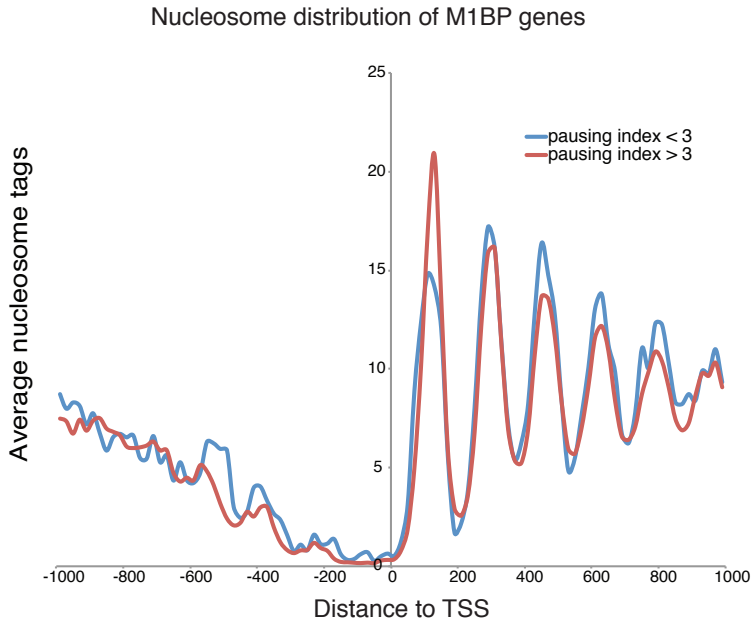

Figure S4. M1BP genes with high pausing efficiencies have greater nucleosome occupancy in the +1 position than M1BP genes with low pausing efficiencies. The pausing efficiency for M1BP genes was calculated as the ratio of the density of T-reads in the region from +1 to +100 to the density of reads from +150 to the end of the gene for genes that do not have neighbors within 500 bp. M1BP genes were separated into a group with a pausing index at 3 or more (high pausing efficiency) and those that were less than 3 (low pausing efficiency). The plot displays a composite nucleosome pattern for each group of genes.

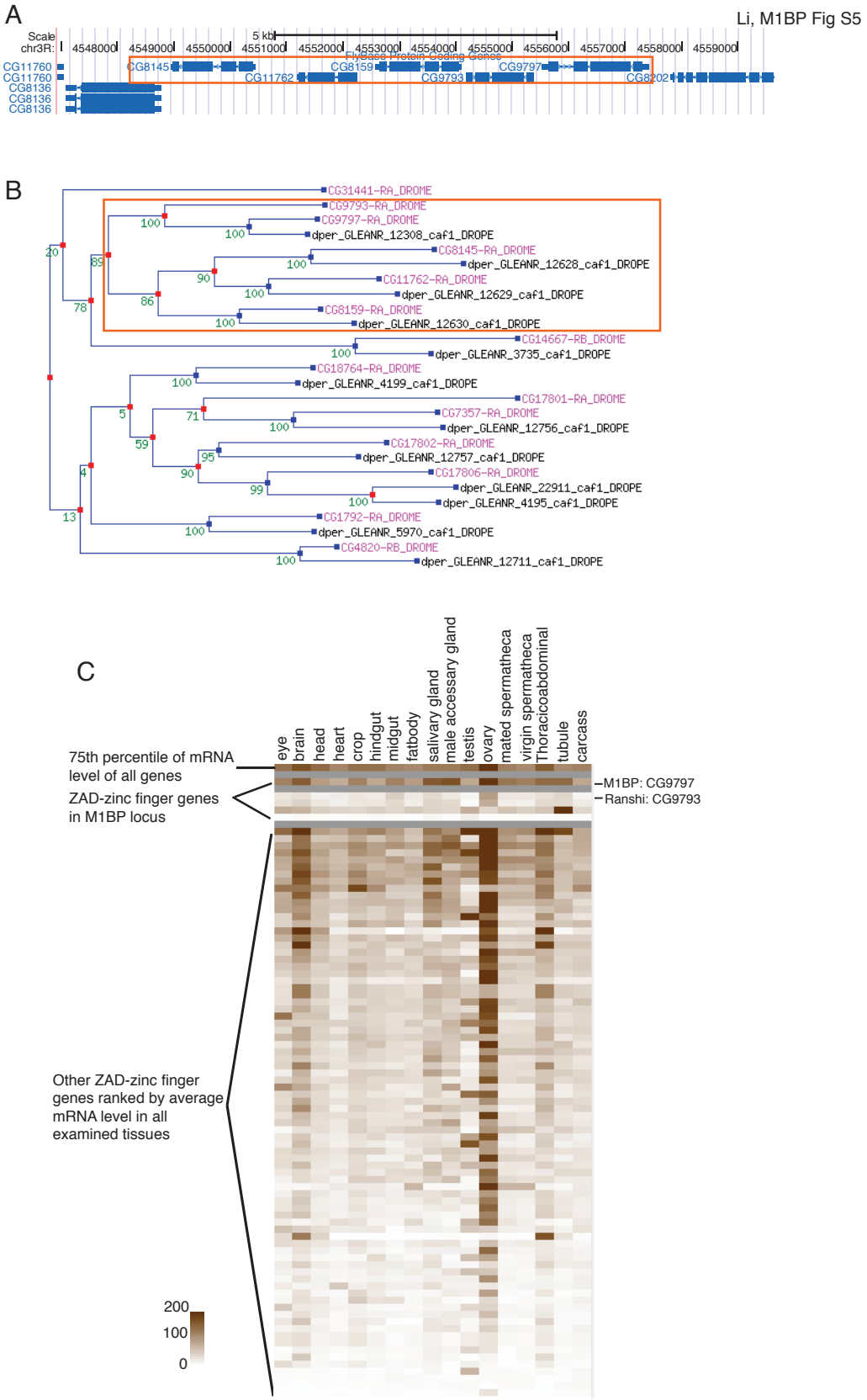
A
Li, M1BP Fig S5
Scale
chr3R:
5 kb
4548000
4549000
4550000
4551000
4552000
4553000
4554000
4555000
4556000
4557000
4558000
4559000
FlyBase Protein Coding Genes
CG11760
CG11760
CG8136
CG8136
CG8136
CG8145
CG11762
CG8159
CG9793
CG9797
CG8202
B
CG31441-RA_DROME
CG9793-RA_DROME
CG9797-RA_DROME
dper_GLEANR_12308_caf1_DROPE
CG8145-RA_DROME
dper_GLEANR_12628_caf1_DROPE
CG11762-RA_DROME
dper_GLEANR_12629_caf1_DROPE
CG8159-RA_DROME
dper_GLEANR_12630_caf1_DROPE
CG14667-RB_DROME
dper_GLEANR_3735_caf1_DROPE
CG18764-RA_DROME
dper_GLEANR_4199_caf1_DROPE
CG17801-RA_DROME
CG7357-RA_DROME
dper_GLEANR_12756_caf1_DROPE
CG17802-RA_DROME
dper_GLEANR_12757_caf1_DROPE
CG17806-RA_DROME
dper_GLEANR_22911_caf1_DROPE
dper_GLEANR_4195_caf1_DROPE
CG1792-RA_DROME
dper_GLEANR_5970_caf1_DROPE
CG4820-RB_DROME
dper_GLEANR_12711_caf1_DROPE
C
eye
brain
head
heart
crop
hindgut
midgut
fatbody
salivary gland
male accessary gland
testis
ovary
mated spermatheca
virgin spermatheca
Thoracicoabdominal
tubule
carcass
75th percentile of mRNA level of all genes
ZAD-zinc finger genes in M1BP locus
M1BP: CG9797
Ranshi: CG9793
Other ZAD-zinc finger genes ranked by average mRNA level in all examined tissues
200
100
0

Figure S5. M1BP is expressed in all tissues while most of the other ZAD-ZNF proteins show tissue-specific expression.

(A) UCSC browser view of the M1BP gene locus. The genes in the orange frame are the M1BP gene and its closest paralogs.

(B) Phylogenetic tree of the M1BP gene family was queried from http://www.treefam.org/. Genes in the orange frame with names in pink color are the M1BP gene (CG9797) and its closest paralogs that are located near the M1BP gene in the D. melanogaster genome.

(C) Heat map showing the expression profiles of ZAD-ZNF proteins in different tissues. Expression data are based on microarray results (Chintapalli et al. 2007). The color intensity of the cells in the first row corresponds to the level of the 75th percentile of mRNA for all genes in each tissue.

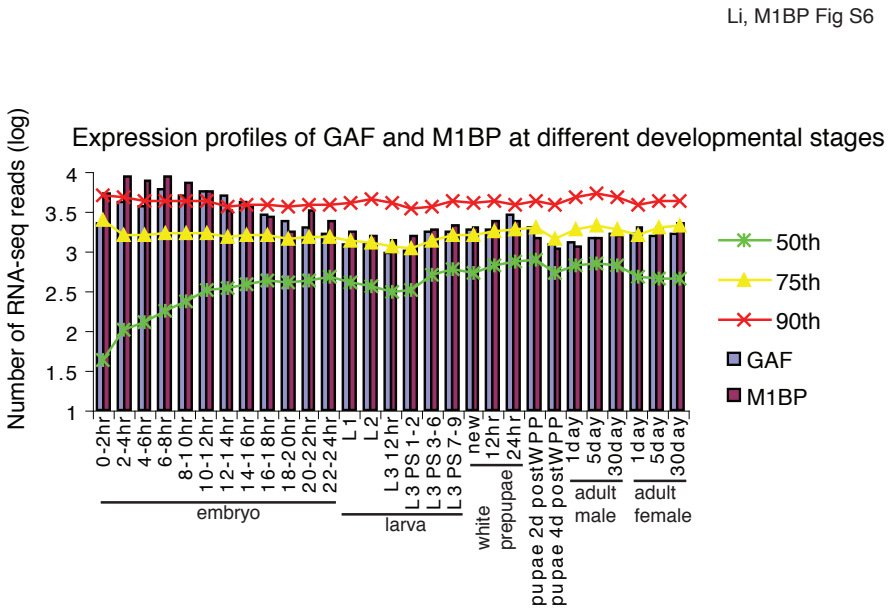

Figure S6. M1BP and GAF are expressed at all development stages.

The histogram displays the number of RNA-seq reads for M1BP and GAF mRNA at different stages of development. The lines indicate different percentile levels (50^{th}, 75^{th}, or 90^{th}) of expression of all genes for each developmental stage. The expression levels for GAF and M1BP are mostly in the range of the 75 to 90 percentile of expression throughout development.

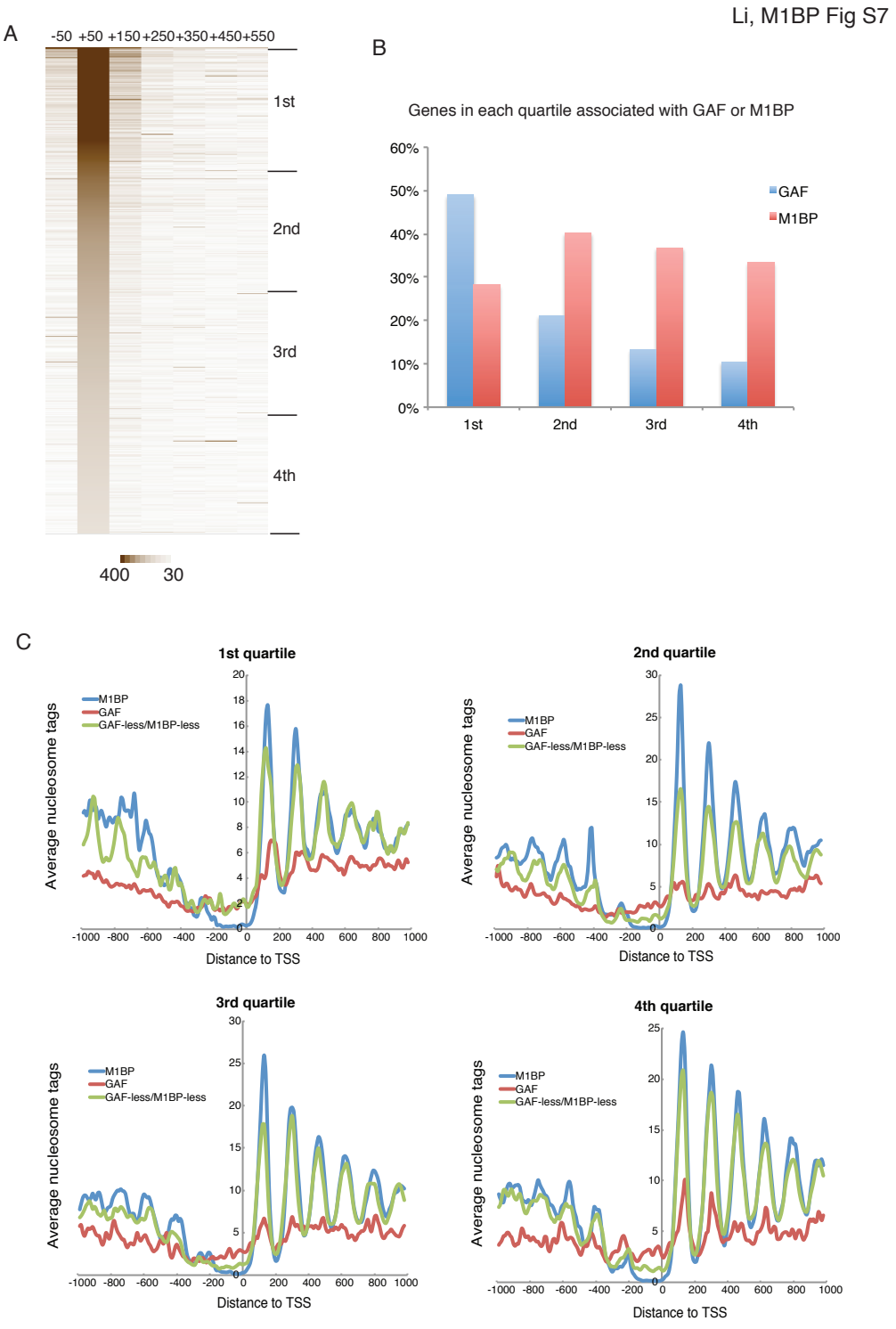

Figure S7. Distinct nucleosome distributions on M1BP and GAF genes are observed at promoters with different levels of paused Pol II.

(A) Heat map of T-reactivity between -100 to +600 for the top 4000 paused genes. Genes are ranked according to T-reactivity from +1 to +100, and then divided into quartiles.

(B) Histograms showing the portion of GAF and M1BP genes in each quartile. GAF genes are highly enriched in the first quartile.

(C) Composite plots of nucleosome distributions on M1BP, GAF and GAF-less/M1BP-less genes in each quartile. The midpoints of pair-end MNase-seq reads were used in mapping. Average nucleosome tags are total nucleosome tags divided by the number of genes.

Li, M1BP Fig S8

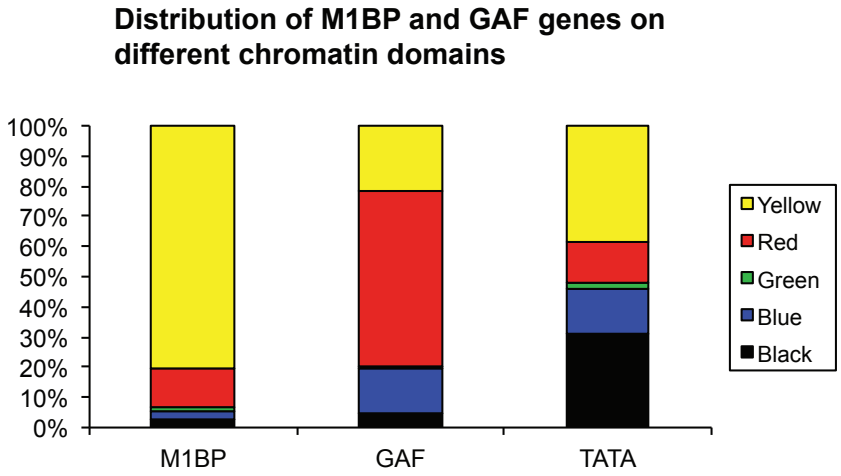

Figure S8. Distribution of M1BP, GAF and TATA genes on different chromatin types. Chromatin types were described by Filion et al. (Filion et al. 2010). YELLOW and RED chromatin types are transcriptionally active while the rest are largely inactive.

Supplemental table 1. Subset of M1BP genes.

Basal transcription factors	TAF1, TAF4, TAF5, TFIIA L, TBP-related factor
Mediator	Arc42, Med8, Med10, Med15, Med16, Med20, Med23
RNA polymerase	Pol II: Rpb4, Rpb8, Rpb10, RpII15; Pol III: CG17209
Transcription elongation	Spt4, Spt5, TFIIS, Eaf
Histone modification and Chromatin structure	Su(var)3-3,JHDM1, pr-set7, MRG15, Set2, Kdm4B, E(z) YL-1, Lid, non-stop, dSAP18 Su(z)12, Polycomblike, Hira, Ino80, domino, SSRP dRing, msl-3, PSR, Pontin, su(Hw), trithorax,trithorax-related, e(y)3
mRNA processing	CPSF subunit: CG7185, Sex lethal, Sex-lethal interactor, dicer-2 Stem-loop binding protein, female lethal d, hiiragi, smaug, tra, protein partner of snf
DNA binding factors	CrebB-17A, Dp, Gnf1, Hr78, CG5641 MTF-1, Myb, Max, CG2199, grauzone, tgo
Other	Notch, medea, rbf, CtBP, Smox, bunched, pygopus, spen

Supplemental table 2. LM-PCR primers for DNase I genomic footprinting.

Gene		Primer sequence	Anealing temperature(ºC)
ZAP3	LM1	GCCTATTAGCGAATAACACAC	50
	LM2	GCGAATAACACACTTGTACCG	54
	LM3	ACACTTGTACCGCTCGGTTG	58
Slmb	LM1	AATTGAGCGAATGTCGTAT	50
	LM2	GTCGTATGCAAGTCATTATTCTCA	54
	LM3	AGTCATTATTCTCACCGCTCCTG	58
Cib	LM1	GAAGCGGAGCTCTTGAC	50
	LM2	AGCTCTTGACGTCACAAATTTAATA	54
	LM3	CTCTTGACGTCACAAATTTAATACTGG	58
Smo	LM1	CATCGGTTATTATCGGTCA	50
	LM2	TATCGGTCATAGCTGCAACA	54
	LM3	GCAACAAGTCGAATGATATGCAA	58

Supplemental table 3. Oligonucleotides from the Smo promoter for Gel-shift and crosslinking.

Wt_38mer-311F	CGATACTCCGCACCCAGTGTGACCGTCGAGCGCATGGC
Wt_38mer-274R	GCCATGCGCTCGACGGTCACACTGGGTGCGGAGTATCG
Mut_38mer-311F	CGATACTCCGCACCCACTCTGACGGTCGAGCGCATGGC
Mut_38mer-274R	GCCATGCGCTCGACCGTCAGAGTGGGTGCGGAGTATCG
Smo_crosslinking-274R	GCCATGCGCTCGACGG

Supplemental table 4. Primers used to generate RNAi probes.

lacZ_RNAi_F	GAATTAATACGACTCACTATAGGGAGATGAAAGCTGGCTACAGGA
lacZ_RNAi_ R	GAATTAATACGACTCACTATAGGGAGAGCAGGCTTCTGCTTCAAT
M1BP_RNAi_F	GAATTAATACGACTCACTATAGGGAGAGCAGCCAAATTGCTTGTCC
M1BP_RNAi_R	GAATTAATACGACTCACTATAGGGAGAAGACGGTGAAGACGCCC

Supplemental table 5. qPCR primers

for RT-PCR

CG9797_RT_F	CGCTGGTGTGTGTGGATTTG
CG9797_RT_R	TTCAGCTCGGAAGTTGTGCAGAAGC
Actin5C_RT_F	TCAGTCGGTTTATTCCAGTCATTCC
Actin5C_RT_R	CCAGAGCAGCAACTTCTTCGTCA
Trn_RT_F	TGGATCGCCACATCGAGACATTG
Trn_RT_R	AATCACCCAGCAGGGCAAACG
CG32685_RT_F	CGATGAGAAGTGCCCCAAAACG
CG32685_RT_R	AATCCACGATCACGAAATCGTACAGA
Smo_RT_F	ACGCCAAGAAGGGCAGGGAC
Smo_RT_R	GGCAATTTTGAGCCGAAACAGG
Dgt2_RT_F	GCCCAAATCCGAGTGATGAA
Dgt2_RT_R	TGACCAGCTCCCTCCATCTC
Dgt6_RT_F	GAGAAGCACCAGCTGCATTG
Dgt6_RT_R	CGAACTCGAGCAGGAAGTTGA
Pav_RT_F	GGGATCCAGTGAATGTGTTCTGT
Pav_RT_R	GCGATCGTCGTGGAGTTCTT
Prosbeta5_RT_F	TCGTGGAGATCAATCAGTTCATG
Prosbeta5_RT_R	CGGCATTCCTTCGAAAGGA

for ChIP assays

Actin5C-65F	GCATTGCGGCTGATAAGGTT
Actin5C+36R	TCCACACAGCACAAGAACTCAA
CG3625+10F	CACCGCTCGCCCCTTT
CG3625+104R	CGGCACCAAAACACAAACAA
CG4609-10F	TTCGCACGCGGCATCT
CG4609+81R	CTCTCGGGAGTTTGAAAATTTGA
CG8224-65F	TCTAGCGGGTGGGTGTTTTT
CG8224+26R	CGCTGGTGTGTGTGGATTTG
Cib-151F	AAATGGACTGTGCCAAATTCG
Cib-50R	TTGCGTGGCTGTGTCATTTT
CG32685-89F	GAATATTCTAGGGTGCTTGCGATATA
CG32685+12R	TTGGCTTGAAAACTAGAGATGGAA
Smo-343F	CGATGCTTGGTGCGCTACTATC
Smo-217R	TGTTCCCCGCAACATTTTGACT
Trn-36F	CGTGACGACACGTTGCCATCT
Trn+63R	CACCTTTTCTTTGCACAATTAGCG
Nrx-94F	CCAGTTGGTGACTCCCTCGG
Nrx+38R	CCCACTTCGGAAGAGTTTCGTT
intergenic_F (chr3R_8399595F)	CAACCGTGGAAGTGACTATGGCG
intergenic_R (chr3R_8399701R)	TACGACTCTGCCGCTGACCTC
hsp70Bc-72F	GAGAGCGCGCCTCGAAT
hsp70Bc+29R	CGTGTTCACTTTGCTTGTTTGAA

Experimental Procedures

DNA Motif analysis

Gene lists: A set of non-redundant *Drosophila* promoters was used for the DNA sequence analysis. 21243 annotated flybase genes were queried from the UCSC table browser (Apr. 2006, BDGP R5/dm3). To avoid use of redundant annotated transcription start sites (TSS), multiple isoforms of the same gene that share an annotated TSS were compressed to one TSS, resulting in 17394 unique annotated TSSs.

We then refined the list of unique annotated TSSs with the sequencing data of short, capped nuclear RNA (Nechaev et al., 2010). 5' RNA reads that occurred only once at a particular position were eliminated. To better define the TSS of genes with multiple closely spaced 5' reads, we grouped 5'-RNA reads mapping within 10 bp of another 5'-RNA read on the same strand to define a cluster whose boundaries are set by the most upstream and downstream 5'-RNA reads. 75% of these clusters are smaller than 30 bp, but a few span over 100 bp. Any cluster with less than 7 reads was eliminated, and this resulted in the loss of only 1% of previously defined TSSs that have 5' reads (Nechaev et al., 2010). We mapped the 5'-RNA clusters to our list of unique TSSs in the region of ± 500 bp (± 50 bp for genes that are <1kb from the nearest TSS). The cluster containing the most reads in proximity to a unique annotated TSS was kept. Within this cluster, the location to which most reads mapped was called 'observed TSS'. If there was no cluster of 5'-RNA, the annotated TSS was used. When more than one annotated TSS was on the same strand within ± 50 bp of each other, only one of the corresponding genes was retained. Thus, we have 7808 TSSs defined by 5'-RNA and 8603 TSSs compressed from annotated TSSs. Due to the limit of resolution of Pol II ChIP-seq data, and considering the good correlation of short RNAs with Pol II on the promoters (Nechaev et al.,

2010), we define those 7808 genes with significant number of short RNAs as active genes and the remaining 8603 genes as inactive genes.

Motif discovery and scan: To discover conserved DNA motifs on Pol II associated promoters that lack GAF, we first filtered out 1227 active genes that have GAF ChIP-chip peaks (Lee et al., 2008) within 500 bp from the TSS. MEME analysis (Bailey and Elkan, 1994) was done with promoter sequences spanning from -100 to +100 of the remaining 6581 active promoters. The consensus sequence of Motif 1 was shown in the format of WebLogo (http://weblogo.berkeley.edu/logo.cgi).

Motif scans were done by adding position-specific scoring matrices to the program originally reported by Sojda and Nixon (Sojda et al., 1999). For the whole genome scan of Motif 1 in *Drosophila*, we queried the Dm3 genome assembly from the UCSC table browser and fetched the sequences accordingly (except unmapped extra sequences) on Galaxy (http://main.g2.bx.psu.edu/). The Motif1 scoring matrix from our MEME analysis was applied in Motif 1 scanning. For GAGA element analysis, a previously reported GAGA scoring matrix (Down et al., 2007) was used to scan the promoter sequences from -500 to +500 of all 16411 active and inactive genes. For analysis of TATA, the scoring matrix was taken from a previous study (Gershenzon et al., 2006). Since this core promoter motif has a conserved distance relative to the TSS (Gershenzon et al., 2006), we scanned the sequences in regions ± 50 bp around the conserved location of the motif. 80% similarity cutoff was applied to Motif 1 and the TATA box; 75% similarity cut off was applied to GAGA (75%).

For analysis of 39 gene-specific transcription factors (Figure 5), the binding motifs were queried from published work (Kulakovskiy et al., 2009). The “integrated” motifs based on both ChIP-chip/seq and footprinting data were used except for fkh, shn, slp1. For these three factors,

only 1 or 2 footprints were available and the motifs extracted solely from ChIP-chip data are more relevant (Kulakovskiy et al., 2009) and therefore were used in our analyses. 90% similarity cutoff was applied to scans of factors that have short and/or AT-rich motifs, including abd-A, antp, bcd, cad, slp1, dfd, ems, en, eve, ftz, hb, oc, sna, srp, vnd, fkh, shn and ubx. 80% similarity cutoff was applied to scans of the other factors.

In vitro protein-DNA crosslinking

Radiolabeled DNA was prepared by pulse-chase primer extension on a single strand DNA as previously described (Sypes and Gilmour, 1994). A 16mer oligonucleotide (Smo_crosslinking-274R) was annealed to the 3' end of a 38mer single strand DNA composed of the sequence from -311 to -274 of the smo promoter (Wt_38mer-311F, see supplementary table 3). Limited extension was done with the Klenow fragment (New England Biolabs) in the presence of BrdUTP, dATP and alpha-32P-dCTP. After the pulse, all four nucleotides were added in excess to limit incorporation of radioactive dCTP and to complete synthesis of the double stand probe. The double strand DNA probe was purified through a Bio-Spin 6 column (Bio-Rad). UV crosslinking of eluate from the DNA affinity column was performed as previously described (Gilmour et al., 1990). Binding reactions were assembled as described for the gel-shift assay but scaled up by 12-fold. Also, mutant probe was included as a competitor for nonspecific binding.

Expression and purification of recombinant M1BP

cDNA encoding full-length M1BP (LD30467) was obtained from the *Drosophila* Genomics Resource Center and cloned into NheI/EcoRI-cut pET28 so that the encoded fusion protein has a His-tag at the N-terminus. His-M1BP was expressed in BL21 DE3 E. coli and purified with TALON cobalt beads (Clontech) in 8M urea. Protein was then renatured via stepwise dialysis in

buffer (0.5 M NaCl, 20 mM Tris-HCl pH 7.9, 10% glycerol) containing 6M, 4M, 2M and finally no urea. Renatured His-M1BP was soluble and used for binding assays and for production of antibody in a rabbit.

Other bioinformatic analyses

Feature mapping: Mapping of motifs, ChIP-chip peaks, ChIP-seq reads and MNase-seq reads to transcription start sites was done with homemade scripts.

Expression variation analysis (Figure 4C & D): The list of 'single value' description for the expression level of each gene at different developmental stages or tissues was queried from FlyBase {McQuilton, 2012 #2984}, which is based on published RNA-seq (Graveley et al., 2010) and microarray data (Chintapalli et al., 2007). The relative standard deviation (standard deviation divided by average) of expression was calculated for each gene to describe the expression variation during development or in different tissues. 24 developmental stages from '0-2 h embryo' to '4 day pupae' or 15 somatic tissues from adult flies were included in this analysis, respectively.

Gene Ontology analysis: The program DAVID (http://david.abcc.ncifcrf.gov/) was used to determine which Gene Ontology (GO) terms of Biological Processes (BP) were overrepresented in GAF or M1BP genes(Dennis et al., 2003; Huang da et al., 2009). For Figure 6, specific GO terms were queried from the GO_BP_FAT list. Functional annotation clustering was done to collapse GO terms so that they present the biological meanings more clearly.

Clustering and heat map: Clustering analyses were done with Cluster 3.0 software (Eisen et al., 1998). Heat maps were generated with Java TreeView package (Saldanha, 2004).

Permanganate-ChIP-seq data analyses: The locations of transcription bubbles were determined by using Genetrack (Albert et al., 2008) to call 12 bp peaks based on T-reactivity.

Nucleosome distribution and structure analyses: Pair-end MNase-seq data were queried from the published work (Gilchrist et al., 2010). Raw data were mapped to the *Drosophila* genome with Bowtie 1.1.2 (Langmead, 2010) on Galaxy (Blankenberg et al., 2010; Goecks et al., 2010) using the built-in index for dm3. Plots of size distribution of MNase fragments were generated with homemade scripts. Normal mixture modeling was done with R-package mclust (Fraley et al 2006). Composite plots to anchor point were generated with Bedtools (Quinlan and Hall, 2010) and homemade scripts. The midpoints of the DNA fragments or the downstream cutting edges defined by pair-end sequencing reads were used in mapping as specified in the figure legends.

References

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2, 28-36.

Chintapalli, V.R., Wang, J., and Dow, J.A. (2007). Using FlyAtlas to identify better *Drosophila* melanogaster models of human disease. Nat Genet 39, 715-720.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4, P3.

Down, T.A., Bergman, C.M., Su, J., and Hubbard, T.J. (2007). Large-scale discovery of promoter motifs in *Drosophila* melanogaster. PLoS Comput Biol 3, e7.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95, 14863-14868.

Gershenzon, N.I., Trifonov, E.N., and Ioshikhes, I.P. (2006). The features of *Drosophila* core promoters revealed by statistical analysis. BMC Genomics 7, 161.

Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell 143, 540-551.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2010). The developmental transcriptome of *Drosophila* melanogaster. Nature.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4, 44-57.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458, 362-366.

Kulakovskiy, I.V., Favorov, A.V., and Makeev, V.J. (2009). Motif discovery and motif finding from genome-mapped DNase footprint data. Bioinformatics 25, 2318-2325.

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. Mol Cell Biol 28, 3290-3300.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. Science 327, 335-338.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics 20, 3246-3248.

Sojda, J., 3rd, Gu, B., Lee, J., Hoover, T.R., and Nixon, B.T. (1999). A rhizobial homolog of IHF stimulates transcription of dctA in Rhizobium leguminosarum but not in Sinorhizobium meliloti. Gene 238, 489-500.