THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

Supplementary Material

corresponding to:

Promoter strength delimits enhancer threshold in the early *Drosophila* embryo

MIROO HONG, JOUNG-WOO HONG

Full Text for this paper is available at: https://doi.org/10.1387/ijdb.240230jh

Table S1

Primer Name	Sequence (5' to 3')[1]	Usage
hsp70A promoter-F	GCAGAAAGAAAACTCGAGAA	Genomic PCR
hsp70A promote-R	CAGATCGATTCCAATAGCAG	Genomic PCR
bcd promoter-F	GCCGTACTGTTCGATTAAAA	Genomic PCR
bcd promoter-R	AGCGGATGATGGTAAAAGTT	Genomic PCR
abdA promoter-F	CAAATCGGGAAATTACTCAC	Genomic PCR
abdA promoter-R	CGTGTAATTGGTAATTCTTGC	Genomic PCR
AbdB promote-F	TGTATACAGTGTACGTGTGCATAA	Genomic PCR
AbdB promoter-R	TCCAAAAACGTTTCTTCTTTC	Genomic PCR
ftz promoter-F	GATAAAGTTGCCAGGACCTC	Genomic PCR
ftz promoter-R	GCGTAGCTGTAGTGGCTCTG	Genomic PCR
en promoter-F	CCAGCTACTTGTGGGATCT	Genomic PCR
en promoter-R	ATTGGTTTCGACTTGGT	Genomic PCR
rho promoter-F	GGAAACCAGTGAATAACCAG	Genomic PCR
rho promoter-R	AACTTAGTTTTGCTGCTCGT	Genomic PCR
GAL4DBD-CDK9-F	CAAAATGAAGCTACTGTCTTCTATCGAAC	PCR
GAL4DBD-CDK9-R	GGCTACCAAACCCGGTCAATCATAC	PCR
hsp70A-Bubble-1	CACTTTAACTTGCACTTTATTGCAG	LM-PCR[2]
hsp70A-Bubble-2	TTGCAGATTGTTTAGCTTGTTCAGCTGCGC	LM-PCR
hsp70A-Bubble-3	GCAGATTGTTTAGCTTGTTCAGCTGCGCTTGTTTG	LM-PCR
eve-Bubble-1	CAAGAGGTCCACAACCAG	LM-PCR
eve-Bubble-2	TTCTGGTCCACGGGACTG	LM-PCR
eve-Bubble-3	CACGGGACTGGCGTCGTG	LM-PCR
rho-Bubble-1	CATTGGTAACTTAGTTTTGC	LM-PCR
rho-Bubble-2	AACTTAGTTTTGCTGCTCGT	LM-PCR
rho-Bubble-3	TTTTGCTGCTCGTAAATCCAG	LM-PCR

Table S1. DNA oligonucleotide sequences used in this study. [1]All primer sequences are presented in 5' to 3' direction relative to the physiological orientation of transcription of Flybase (https://flybase.org/) gene model. [2]LM-PCR stands for ligation-mediated polymerase chain reaction and is a crucial step in the KMnO4 bubble assay (Hendrix *et al.*, 2008) for identifying transcription bubbles downstream of a TSS. F and R stand for forward and reverse, respectively.

Table S2

Promoter[1]	TATA	Inr	MTE	DPE	PB	DPR	*E*-value
	Start Position of the Core Promoter Element (relative to the transcription start site as a +1)[2]						
	DNA Sequence (5' to 3' relative to the sense strand orientation)						
SCP1	-31 TATATAAG	-2 TCAGTC	+17 CGAGCCGAGCAG	+28 AGACGT	+27 CAGACGT	+17 TCGAGCCGAGCAGACGTGC	2.1e-4
SCP1ΔTATA	-31 ACGTCCGT	-2 TCAGTC	+17 CGAGCCGAGCAG	+28 AGACGT	+27 CAGACGT	+17 TCGAGCCGAGCAGACGTGC	1.5e-2
SCP1ΔInr	-31 TATATAAG	-2 TGTGAC	+17 CGAGCCGAGCAG	+28 AGACGT	+27 CAGACGT	+17 TCGAGCCGAGCAGACGTGC	9.4e-3
SCP1ΔMTE	-31 TATATAAG	-2 TCAGTC	+17 ATCCACGAGCAG	+28 AGACGT	+27 CAGACGT	+17 TATCCACGAGCAGACGTGC	4.9e-3
SCP1ΔDPE	-31 TATATAAG	-2 TCAGTC	+17 CGAGCCGAGCAG	+28 AGCATA	+27 CAGCATA	+17 TCGAGCCGAGCAGCATAGC	4.9e-3
SCP1ΔAll	-31 ACGTCCGT	-2 TGTGAC	+17 ATCCACGAGCAG	+28 AGCATA	+27 CAGCATA	+17 TATCCACGAGCAGAGCATA	6.4e+0
hsp70	-32 TATAAATA	-2 TCAATT	+23 TGAACACATCGC	- -	+38 GCGAAAG	- -	1.2e-3
eve	-31 TATAAAAG	- -	- -	- -	- -	- -	1.4e-1
bcd	- -	- -	- -	- -	- -	- -	5.1e+0
abd-A	- -	- -	- -	- -	- -	- -	4.8e+0
Abd-B	- -	-2 TCAGTG	- -	- -	- -	- -	4.3e+0
rho	-33 TATAAAGG	-2 TCAGTT	- -	- -	- -	- -	7.9e-3
en	- -	-2 TCAGTC	- -	- -	+26 ACGTGCG	+12 TCGATGTGAACAGACGTGC	3.7e-2
ftz	-32 TATATAAG	-2 TCATTC	- -	+28 ACATCG	- -	- -	4.6e-2

Table S2. Statistical significance of the position weight matrix (PWM) scores for CPEs identified in the core promoters. [1]The size of all promoters used in this CPE analysis is approximately 85 bp surrounding the TSS. [2]In the presence of Inr, the third nucleotide of Inr was considered as the TSS. In the absence of Inr, the 5' end of the mRNA of the Flybase

(https://flybase.org/) gene model is taken as the TSS of the gene. The promoter activity of various mutant variants of SCP1 including the TATA box mutant was also experimentally determined (Juven-Gershon *et al.*, 2006). If the *E*-values and promoter strengths of the WT and variants of SCP1 are proportionally correlated, the least squares regression analysis produces the best-fit equations. These equations can be used as a standard line to estimate promoter strengths of *eve* and *hsp70* promoters.

Table S3

S4[1]	R-squared (R^2)	Pearson Correlation Coefficient (R)	Line of best fit[2]	Core Promoters					
				SCP1		*hsp70*		*eve*	
				variables (log_{10}, %)[3]					
				Predictor[4]	Response	Predictor[4]	Response[5]	Predictor[4]	Response[5]
A	0.8062	0.8979	$y = 7.16 - 20.994x$	-3.68	100[6]	-2.92	68.50	-0.85	25.00
B	0.6423	0.8014	$y = 3.13 - 19.776x$	-3.68	100	-2.92	60.88	-0.85	19.94
C	0.7714	0.8783	$y = 9.63 - 19.619x$	-3.68	100	-2.92	66.92	-0.85	26.31
D	0.7411	0.8609	$y = 7.85 - 19.437x$	-3.68	100	-2.92	64.61	-0.85	24.37

Table S3. The response variables of the *eve* and *hsp70* promoters generated by the best-fit equations (fig. S4). The best-fit equations (fig. S4) produced by the least squares regression analyses on the SCP1 promoters were used to statistically estimate the response variables of the *eve* and *hsp70* promoters. [1]The four rows in the first column, A, B, C and D, in the above table correspond the statistics of Fig. 6 A, B, C and D, respectively. [2]In least squares regression analyses, line of best fit is a trend line that best represents the relationship between a predictor variable and one or more response variables (Bremer and Doerge, 2009). Here, the line minimizes the difference between the observed and predicted response variables (promoter strength) corresponding to the predictor variables (combined *p*-value of CPEs). [3]Predictor (log_{10}) and response (%) variables for least squares regression analyses. [4]All of the predictor variables are expressed as *E*-values (log_{10}) of core promoters described in Table S3. [5]*hsp70* and *eve* response variables are predicted values obtained by substituting their predictor variable into the equation of best-fit line. [6]The amount of mRNA synthesized from the SCP1 by transient transfection (Juven-Gershon *et al.*, 2006) is considered 100 %.

Table S4

>PPM_GAGA										
A |	0.27	0.56	0.07	0.78	0.01	0.75	0.06	0.32	0.43	0.51
C |	0.14	0.16	0.10	0.03	0.01	0.20	0.04	0.42	0.16	0.13
G |	0.46	0.18	0.76	0.01	0.96	0.01	0.89	0.09	0.32	0.31
T |	0.13	0.10	0.07	0.18	0.02	0.04	0.01	0.17	0.09	0.05

>PPM_TATA												
A |	0.25	0.03	0.90	0.01	0.95	0.66	0.98	0.82	0.47	0.20	0.25	0.21
C |	0.27	0.06	0.02	0.01	0.01	0.00	0.00	0.02	0.06	0.35	0.34	0.30
G |	0.35	0.03	0.02	0.01	0.01	0.00	0.01	0.03	0.43	0.27	0.25	0.35
T |	0.13	0.88	0.06	0.97	0.03	0.34	0.01	0.13	0.04	0.18	0.16	0.14

>PPM_Inr												
A |	0.14	0.05	0.02	0.92	0.37	0.02	0.04	0.14	0.09	0.37	0.24	0.23
C |	0.25	0.07	0.84	0.01	0.04	0.02	0.20	0.38	0.35	0.14	0.19	0.18
G |	0.27	0.08	0.02	0.02	0.43	0.01	0.05	0.37	0.13	0.38	0.20	0.21
T |	0.34	0.80	0.12	0.05	0.16	0.95	0.71	0.11	0.43	0.11	0.37	0.38

>PPM_MTE												
A |	0.01	0.08	0.97	0.16	0.12	0.08	0.28	0.44	0.32	0.12	0.16	0.12
C |	0.97	0.01	0.01	0.24	0.60	0.24	0.28	0.12	0.04	0.48	0.16	0.16
G |	0.01	0.90	0.01	0.52	0.24	0.48	0.40	0.32	0.44	0.28	0.67	0.68
T |	0.01	0.01	0.01	0.08	0.04	0.20	0.04	0.12	0.20	0.12	0.01	0.04

>PPM_DPE						
A |	0.47	0.01	0.57	0.04	0.06	0.07
C |	0.06	0.20	0.04	0.49	0.26	0.17
G |	0.38	0.78	0.03	0.02	0.66	0.19
T |	0.09	0.01	0.36	0.45	0.02	0.57

>PPM_PB											
A |	0.27	0.27	0.10	0.01	0.09	0.30	0.49	0.01	0.01	0.22	0.23
C |	0.23	0.25	0.01	0.93	0.01	0.05	0.01	0.89	0.01	0.24	0.22
G |	0.32	0.25	0.49	0.01	0.89	0.60	0.01	0.01	0.97	0.26	0.28
T |	0.18	0.23	0.40	0.05	0.01	0.05	0.49	0.09	0.01	0.28	0.27

>PPM_DPR																			
A|	0.12	0.12	0.18	0.35	0.23	0.16	0.27	0.26	0.25	0.30	0.36	0.18	0.10	0.56	0.05	0.07	0.09	0.25	0.16
C|	0.13	0.22	0.12	0.09	0.23	0.24	0.15	0.06	0.06	0.2	0.34	0.01	0.08	0.12	0.53	0.17	0.27	0.16	0.32
G|	0.15	0.37	0.52	0.31	0.33	0.46	0.37	0.58	0.47	0.33	0.14	0.79	0.77	0.06	0.17	0.67	0.26	0.38	0.22
T|	0.60	0.29	0.18	0.25	0.21	0.14	0.21	0.10	0.22	0.17	0.16	0.02	0.05	0.26	0.25	0.09	0.38	0.21	0.30

Table S4. PPMs of core promoter elements. PPMs of GAGA, TATA, Inr, MTE, DPE, and PB were produced by modifying the reported PFMs of GAGA (Sloutskin *et al.*, 2015), TATA (Gershenzon *et al.*, 2006), Inr (Gershenzon *et al.*, 2006), MTE (Sloutskin *et al.*, 2015), DPE (Sloutskin *et al.*, 2015), and PB (Hendrix *et al.*, 2008) using Kadonaga's *Drosophila* core promoter database (Kutach *et al.*, 2000). A machine learning algorithm recently identified human DPR (Vo Ngoc *et al.*, 2020), which is composed of two consecutively connected elements, MTE and DPE. The *Drosophila* variant of DPR was isolated through comparative genomics analysis (Vo Ngoc *et al.*, 2023).

Figure S1

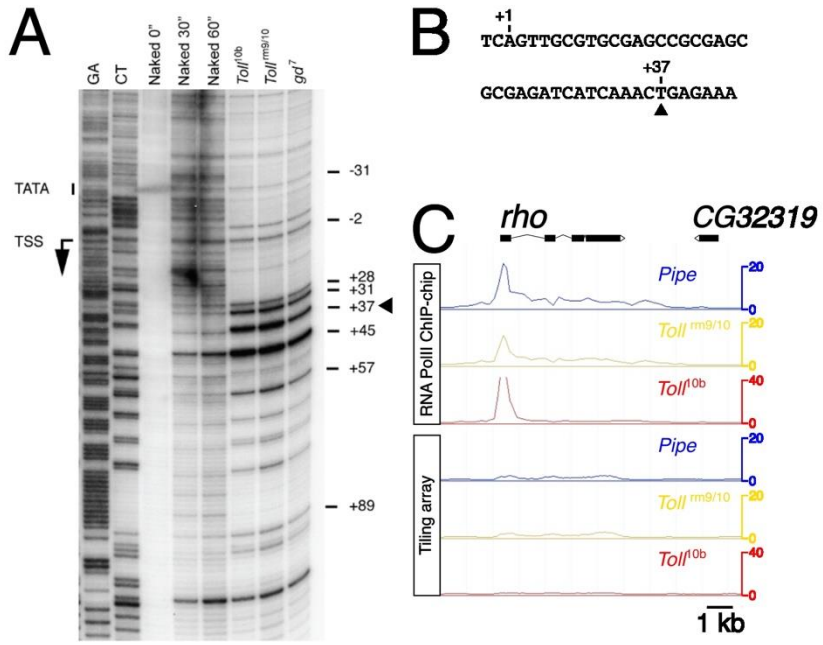

Figure S1. A potassium permanganate ($KMnO_4$) transcription bubble assay on the *rhomboid* (*rho)* locus strongly indicates the existence of the transcription bubble downstream of TSS. (A) The transcription bubble assay was performed in the same way as in Figure 3, except that the *rho* promoter was used. The TSS is labeled on the left and the location of prominent bands relative to the TSS (+1) is shown on the right of the autoradiograms. (B) Promoter sequences downstream of TSS is presented. (C) Results of Pol II ChIP-chip (the first three lines from the top) (Zeitlinger *et al.*, 2007) and tiling microarray (Biemar *et al.*, 2006) analyses on the *rho* locus are shown. Gene prediction models are displayed above each graphical presentation. Each 3' end is labeled with open triangle. The thirty-seventh pyrimidine T nucleotide (solid arrowheads) downstream of the *rho* TSS in all three mutant embryos was modified by $KMnO_4$ *in vivo* (A and B), which is entirely consistent with previous studies (Zeitlinger *et al.*, 2007).

Figure S2

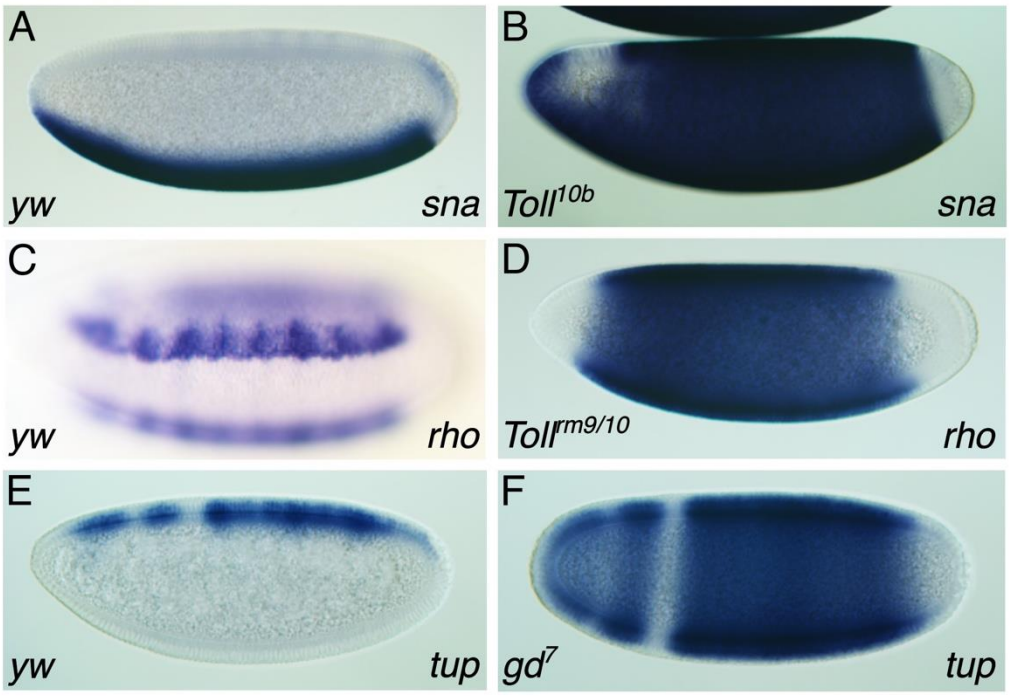

Figure S2. Generation of mutant embryos with three different germ layers. Three different mutant embryos were produced as reported previously (Hendrix *et al.*, 2008) and expression of genes [(*snail* (*sna*), *rho*, and *tailup* (*tup*)] representing three different germ layers was examined by *in situ* hybridization. *Toll*^{10b} mutant embryos contain uniformly high levels of Dorsal (Dl). As a result, early mesodermal genes, for instance *sna*, are expressed throughout the embryo (A and B). *Toll*^{rm9} and *Toll*^{rm10} are recessive alleles of the *Toll* gene that lead to low levels of nuclear Dorsal throughout the early embryo. In *gd*^{7} (*gd*^{7}/ *gd*^{7}) mutant embryos, Dorsal fails to translocate to the nucleus so that its target genes are not activated throughout the early embryo. Thus, *Toll*$^{rm9/rm10}$ and *gd*^{7} mutant embryos lack mesoderm and contain only neurogenic ectoderm (C and D) and dorsal ectoderm (E and F), respectively.

Supplementary References

BIEMAR F, NIX DA, PIEL J, PETERSON B, RONSHAUGEN M, SEMENTCHENKO V, BELL I, MANAK JR, LEVINE MS (2006). Comprehensive identification of *Drosophila* dorsal-ventral patterning genes using a whole-genome tiling array. *Proc Natl Acad Sci U S A* 103: 12763–12768.

BREMER M, DOERGE RW (2009). *Statistics at the bench: a step-by-step handbook for biologists*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

GERSHENZON NI, TRIFONOV EN, IOSHIKHES IP (2006). The features of *Drosophila* core promoters revealed by statistical analysis. *BMC Genomics* 7: 161.

HENDRIX DA, HONG J-W, ZEITLINGER J, ROKHSAR DS, LEVINE MS (2008). Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. *Proc Natl Acad Sci USA* 105: 7762–7767.

JUVEN-GERSHON T, CHENG S, KADONAGA JT (2006). Rational design of a super core promoter that enhances gene expression. *Nat Methods* 3: 917–922.

KUTACH A, IYAMA S, KADONAGA JT (2000). *Drosophila Core Promoter Database*. University of California San Diego. https://labs.biology.ucsd.edu/Kadonaga/DCPD.htm.

SLOUTSKIN A, DANINO YM, ORENSTEIN Y, ZEHAVI Y, DONIGER T, SHAMIR R, JUVEN-GERSHON T (2015). ElemeNT: a computational tool for detecting core promoter elements. *Transcription* 6: 41–50.

VO NGOC L, HUANG CY, CASSIDY CJ, MEDRANO C, KADONAGA JT (2020). Identification of the human DPR core promoter element using machine learning. *Nature* 585: 459–463.

VO NGOC L, RHYNE TE, KADONAGA JT (2023). Analysis of the *Drosophila* and human DPR elements reveals a distinct human variant whose specificity can be enhanced by machine learning. *Genes Dev* 37: 377–382.

ZEITLINGER J, STARK A, KELLIS M, HONG J-W, NECHAEV S, ADELMAN K, LEVINE M, YOUNG RA (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39: 1512–1516.