Supporting Information

Papagianni et al. 10.1073/pnas.1713930115

SI Materials and Methods

***Drosophila* Genetics and Transgenic Lines.** The cic^5, cic^6, and dl^{28} alleles were generated via CRISPR-Cas9–based editing. Briefly, custom guide RNA expression constructs targeting *cic* and *dl* coding sequences were prepared in vector *pCFD3* (1) and inserted at the *attP40* landing site via PhiC31-mediated integration (2). The protospacer sequences targeted in each particular case were as follows:

cic^5: 5′-TGGCCCCCAGCTCAAAGTCC-3′ (lower strand)

cic^6: 5′-TAAGCACTGCAGATATAGTT-3′ (lower strand)

dl^{28}: 5′-GCTAAGCAGATTGCTGAGCGT-3′ (lower strand)

The molecular lesions associated with these alleles are described in Table S1; we found no evidence of off-target effects during the generation of these alleles. cic^1 is a hypomorphic allele affecting *cic* embryonic function (3). cic^{Q474X} is a presumed null affecting all known *cic* functions (4). cic^{fetU6} and cic^{fetE11} are, respectively, strong and medium hypomorphs also affecting all *cic* functions (5). cic^4 (6) behaves similar to cic^{fetE11}. cic^3 is a gain-of-function mutation interfering with MAPK-mediated down-regulation of Cic (7). gro^{MB41} is a strong hypomorph that does not affect the Gro-dependent activity of Cic in the early embryo (8, 9). It causes the amino acid substitution R483H mapping to the central pore of the β-propeller domain (8). gro^{MB41} embryos were obtained using the flippase-dominant female sterile (FLP-DFS) technique (10), which produces homozygous mutant clones in the female germline upon loss of the ovo^{D1} dominant sterile mutation. *dl* mutant embryos were derived from *MVD-Gal4 > UAS-shRNA-dl* females using the Transgenic RNAi Project (TRiP) insertion line *GL00676* (FlyBase). *cact* mutant embryos were obtained similarly, using the TRiP insertion line *HMS00084* (FlyBase). Embryos with uniform Torso activation were derived from *Tubulin-Gal4 > UAS-tsl* females, which express the Tsl determinant ectopically in all follicle cells of the ovary (11). The $cic^{\Delta C2}$ and cic^{eh1} transgenic lines have been described by Astigarraga et al. (12) and Forés et al. (9), respectively. *pipe* expression was visualized using the *M2 pipe-lacZ* reporter (13). The *VRE-lacZ*, VRE^{OPT}*-lacZ*, and $Sxl^{AT/Dl(0\text{-}2)}$*-lacZ* reporters were assembled using *pCaSpeR-hs43-lacZ*. The VRE^{OPT} enhancer contains a single base-pair substitution in each of the four AT sites (red nucleotides in Fig. 2*A*). The $Sxl^{AT/Dl(0\text{-}2)}$ synthetic module contains a cluster of AT and Dorsal binding sites replacing the sequence between positions –0.4 and –0.6 kb of the *Sxl* upstream region, a region devoid of regulatory sites required for *Sxl* expression (14). Wild-type and mutated tld^{CRM}*-lacZ* reporters were assembled in *placZattB*. Mutations in the *VRE* and tld^{CRM} regulatory fragments were introduced by recombinant and inverse PCR, respectively. Transgenic lines were established by P-element–mediated transformation or using the PhiC31-based integration system (2). All tld^{CRM}*-lacZ* constructs were inserted at cytological position 86FB.

In Situ Hybridization and Immunostaining. Embryos were fixed in 4% formaldehyde-PBS-heptane using standard procedures. Ovaries were dissected in PBS and fixed with 4% formaldehyde-PBS. Digoxigenin-UTP–labeled antisense RNA probes were synthesized using *dpp*, *lacZ*, *Sxl*, and *zen* cloned cDNA templates linearized at the 5′ end and transcribed with T3 or T7 polymerases. FISH analyses were carried out using similar probes labeled with digoxigenin-UTP (*kni* and *zen*) or biotin-UTP (*tll* and *twi*). Signals were obtained using antidigoxigenin antibody coupled to alkaline phosphatase (AP) (chromogenic detection) or with sheep antidigoxigenin or mouse antibiotin antibodies followed by incubation with appropriate secondary fluorochrome-conjugated antibodies (Molecular Probes) (fluorescent detection). Immunostaining signals were detected similarly using appropriate secondary fluorochrome-conjugated antibodies (Molecular Probes). Fluorescent and AP-stained embryos were mounted in Fluoromount and Permount, respectively. Cuticle preparations were mounted in 1:1 Hoyer's medium/lactic acid and cleared overnight at 60 °C. Wings were rinsed in isopropanol and mounted in Euparal.

Protein Expression and EMSAs. The HMG-C1, HMG-$C1^{mut}$, and HMG ^{mut}-C1 constructs carry a His tag at the C terminus and were expressed and purified from bacteria using the *Proteus* IMAC Mini Sample Kit. EMSAs were carried out using standard protocols. Briefly, DNA probes were synthesized as complementary oligonucleotides leaving 5′ GG overhangs, and were end-labeled using α-32P-dCTP and Klenow Fragment, exo- (Thermo Scientific). As a control (*CBS* probe), we used a CBS from the *ind* gene, a target of Cic during patterning of the neuroectoderm (15–17). The sequences of wild-type and mutant probes are as follows (intact and mutated CBSs are underlined):

CBS: 5′ GGAGACACTTCATGAATGAATACATCCTGACC 3′

VRE AT1: 5′ GGAAAACTTATATCAAAGAAAATAGGGGCACC 3′

VRE AT1 mut: 5′ GGAAAACTTATATCAGAGAAAATAGGGGCACC 3′

VRE AT2: 5′ GGGGGGCCTATATGAACGAATATTGATTGGCC 3′

VRE AT2 mut: 5′ GGGGGGCCTATATGAGCGAATATTGATTGGCC 3′

VRE AT2 opt: 5′ GGGGGGCCTATATGAATGAATATTGATTGGCC 3′

tld AT1: 5′ GGATCCGCCGCATGAACGAATCGTTTCGCGCC 3′

tld AT1 mut: 5′ GGATCCGCCGCATGAGCGAATCGTTTCGCGCC 3′

tld AT2: 5′ GGCTGTTGTTTGCATTCAATGGATTTTGATCC 3′

tld AT2 mut: 5′ GGCTGTTGTTTGCATTCCATGGATTTTGATCC 3′

dpp AT: 5′ GGAGCGCTTGCGTGAATGATATGAGGGGTGCC 3′

dpp AT mut: 5′ GGAGCGCTTGCGTGACTGATATGAGGGGTGCC 3′

dpp AT opt: 5′ GGAGCGCTTGCGTGAATGAAATGAGGGGTGCC 3′

Binding reactions were carried out in a total volume of 20 μL containing 60 mM Hepes (pH 7.9), 20 mM Tris·HCl (pH 7.9), 300 mM KCl, 5 mM EDTA, 5 mM DTT, 12% glycerol, 1 μg of poly(deoxyinosinic-deoxycytidylic) acid [poly(dI-dC)], 1 μg of BSA, 1 ng of DNA probe, and 1 ng of His-tagged protein. After incubation for 20 min at room temperature, protein–DNA complexes were separated on 5% nondenaturing polyacrylamide gels run in 0.5× TBE at 4 °C, and detected by autoradiography.

ChIP-nexus. Embryos aged between 2 and 4 h after egg deposition were dechorionated, washed with water and PBT (PBS/0.1% Triton), fixed for 15 min in formaldehyde/heptane with shaking, washed with PBT-glycine and PBT, and frozen in liquid nitrogen until used. ChIP-nexus experiments and data processing were carried out as described (18), except that the data were aligned to the dm6 (and not dm3) genome. A detailed ChIP-nexus protocol is available at research.stowers.org/zeitlingerlab/protocols.html.

Data Analysis. Dorsal ChIP-nexus data in $Toll^{10b}$ and Cic ChIP-nexus data in gd^{7} and $Toll^{10b}$ were plotted for selected enhancer regions using R. The zoom-in versions show Dorsal motifs (GGRWWTTCC with up to two mismatches) and Cic motifs: CBS (TSAATGAA with no mismatch) or AT (TSAATGAA with one mismatch). If there were multiple overlapping Dorsal motifs, only one (if possible, the most central) is shown. To identify additional enhancers repressed by Dorsal and Cic, the recently published putative dorsal ectodermal enhancers (19) were analyzed for Dorsal binding and Cic binding in $Toll^{10b}$, but not Cic binding in gd^{7}. AP genes previously found to have altered expression in *cic* mutants were manually analyzed, and no high Cic signal was found at known enhancers and surrounding regions (based on the Open Regulatory Annotation database). For each locus in the heat map in Fig. 4*B*, the enhancer with the highest Dorsal signal was selected, and the Dorsal and Cic ChIP-nexus signal was calculated in a 200-bp window centered on the Dorsal peak summit. To systematically identify Dorsal-dependent and Dorsal-independent Cic binding regions, Cic motifs (TSAATGAA with no or one mismatch) with high Cic binding in $Toll^{10b}$ were selected for downstream analysis. The ratio of Cic signal in gd^{7} versus $Toll^{10b}$ was used to separate bound Cic motifs into the "Dorsal-dependent Cic binding" and "Dorsal-independent Cic binding" sets. Cic motifs in the two sets were then analyzed for the presence or absence of a mismatch (CBS versus AT motifs). Then, Dorsal motifs (GGRWWTTCC with up to one mismatch) within a 50-bp distance of the Cic motif were scored. The code is available at github (https://github.com/zeitlingerlab/Papagianni_PNAS_2017).

1. Port F, Chen HM, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci USA* 111: E2967–E2976.
2. 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 USA* 104:3312–3317.
3. Jiménez G, Guichet A, Ephrussi A, Casanova J (2000) Relief of gene repression by Torso RTK signaling: Role of *capicua* in *Drosophila* terminal and dorsoventral patterning. *Genes Dev* 14:224–231.
4. Tseng AS, et al. (2007) Capicua regulates cell proliferation downstream of the receptor tyrosine kinase/ras signaling pathway. *Curr Biol* 17:728–733.
5. Goff DJ, Nilson LA, Morisato D (2001) Establishment of dorsal-ventral polarity of the *Drosophila* egg requires *capicua* action in ovarian follicle cells. *Development* 128: 4553–4562.
6. Forés M, et al. (2017) A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer. *PLoS Genet* 13:e1006622.
7. Forés M, Papagianni A, Rodríguez-Muñoz L, Jiménez G (2017) Using CRISPR-Cas9 to study ERK signaling in *Drosophila*. *ERK Signaling: Methods and Protocols*, ed Jiménez G (Springer, New York), pp 353–365.
8. Jennings BH, et al. (2006) Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol Cell* 22:645–655.
9. Forés M, et al. (2015) Origins of context-dependent gene repression by Capicua. *PLoS Genet* 11:e1004902.
10. Chou TB, Perrimon N (1996) The autosomal FLP-DFS technique for generating germline mosaics in *Drosophila* melanogaster. *Genetics* 144:1673–1679.
11. Stevens LM, Beuchle D, Jurcsak J, Tong X, Stein D (2003) The *Drosophila* embryonic patterning determinant Torsolike is a component of the eggshell. *Curr Biol* 13:1058–1063.
12. Astigarraga S, et al. (2007) A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. *EMBO J* 26:668–677.
13. Andreu MJ, et al. (2012) Mirror represses *pipe* expression in follicle cells to initiate dorsoventral axis formation in *Drosophila*. *Development* 139:1110–1114.
14. Yang D, et al. (2001) Interpretation of X chromosome dose at sex-lethal requires non-E-box sites for the basic helix-loop-helix proteins SISB and daughterless. *Mol Cell Biol* 21:1581–1592.
15. Ajuria L, et al. (2011) Capicua DNA-binding sites are general response elements for RTK signaling in *Drosophila*. *Development* 138:915–924.
16. Garcia M, Stathopoulos A (2011) Lateral gene expression in *Drosophila* early embryos is supported by grainyhead-mediated activation and tiers of dorsally-localized repression. *PLoS One* 6:e29172.
17. Lim B, et al. (2013) Kinetics of gene derepression by ERK signaling. *Proc Natl Acad Sci USA* 110:10330–10335.
18. He Q, Johnston J, Zeitlinger J (2015) ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat Biotechnol* 33:395–401.
19. Koenecke N, Johnston J, Gaertner B, Natarajan M, Zeitlinger J (2016) Genome-wide identification of *Drosophila* dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol* 17:196.

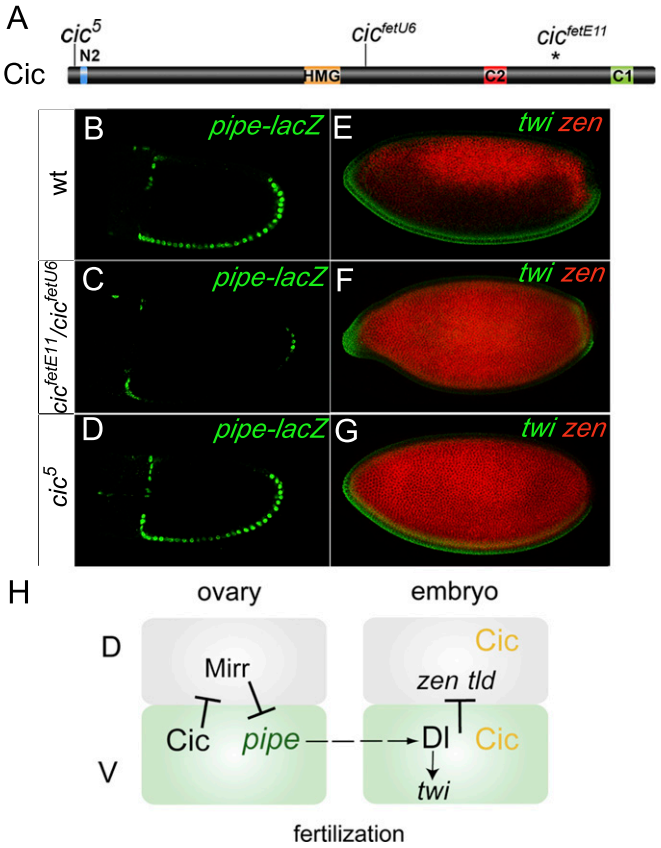

Fig. S1. Cic regulates embryonic DV patterning independently of its function in the ovary. (*A*) Diagram of the *Drosophila* Cic-S protein indicating its functional domains and the positions of mutations. The HMG-box and C1 domains are both required for binding of Cic to DNA (6). The N2 motif is required for interactions with the Gro corepressor (9), whereas C2 acts as a MAPK docking motif (12). The cic^{5}, cic^{fetU6}, and cic^{fetE11} mutations are described in Table S1 and *SI Materials and Methods*; cic^{fetE11} is caused by a transposon insertion (asterisk) (5). (*B–D*) Wild-type (wt) and mutant stage 10 egg chambers showing expression of a *pipe-lacZ* reporter; expression is lost in cic^{fetE11}/cic^{fetU6}, but not in cic^{5}, egg chambers. (*E–G*) Expression of *twi* and *zen* mRNA in embryos laid by wt and mutant females; *zen* is derepressed in both mutant backgrounds, but *twi* expression is normal in cic^{5} embryos. Note the abnormal morphology characteristic of cic^{fetE11}/cic^{fetU6} embryos, a phenotype we have not studied further. Images in *B–G* were obtained at 200× magnification. (*H*) Model of Cic regulatory functions in ovarian and embryonic DV patterning. The Cic-S protein active in the embryo (orange) is fully dispensable in the ovary and in other contexts (also Fig. S2). Mirr, Mirror.

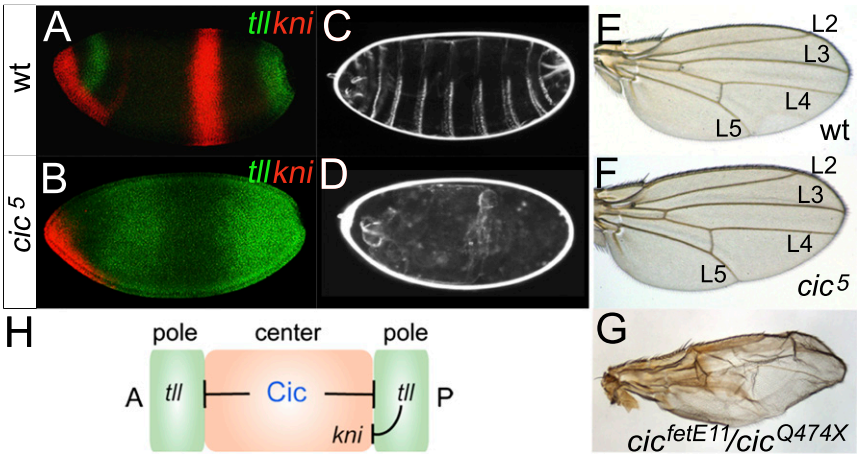

Fig. S2. N2-containing Cic-S isoform is essential for embryonic, but not wing, development. Patterns of *tll* and *knirps* (*kni*) mRNA expression in embryos produced by wild-type (*A*) and homozygous cic^5 (*B*) females are shown; the mutant embryo shows expanded *tll* expression, which then causes repression of the abdominal *kni* domain. The cuticles of embryos derived from wild-type (*C*) and homozygous cic^5 (*D*) females are shown; note the absence of segments in *D*. Wings from wild-type (wt) (*E*), homozygous cic^5 (*F*), and transheterozygous cic^{fetE11}/cic^{Q474X} (*G*) adult flies are shown; veins L2–L5 are indicated. Contrary to cic^{fetE11}/cic^{Q474X}, the cic^5 mutation does not affect wing vein patterning, indicating that the Cic-S isoform is dispensable for this process (1, 15). A detailed characterization of the Cic isoforms active in the wing and in the follicular epithelium will be presented elsewhere. Images in *A–D* and *E–G* were obtained at 200× and 40× magnification, respectively. (*H*) Model of Cic function in terminal patterning. Cic represses *tll* expression in central regions of the embryo, but not at the poles (where Cic is down-regulated by Torso RTK signaling). A, anterior; P, posterior.

1. Roch F, Jiménez G, Casanova J (2002) EGFR signalling inhibits Capicua-dependent repression during specification of *Drosophila* wing veins. *Development* 129:993–1002.

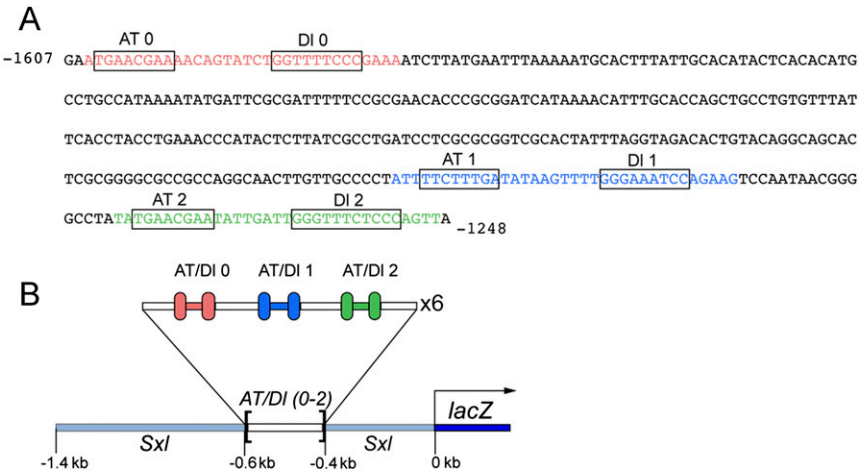

Fig. S3. Structure of the $SxI^{AT/DI(0-2)}$-*lacZ* reporter. (*A*) Partial sequence of the *VRE* enhancer indicating (in color) the three AT/DI binding site pairs included in the *AT/DI(0-2)* module (Fig. 2*F*). Numbers indicate the positions relative to the transcription start site of the endogenous *zen* gene. (*B*) Final structure of the reporter. Numbers indicate the positions relative to the transcription start site of the *SxI* gene.

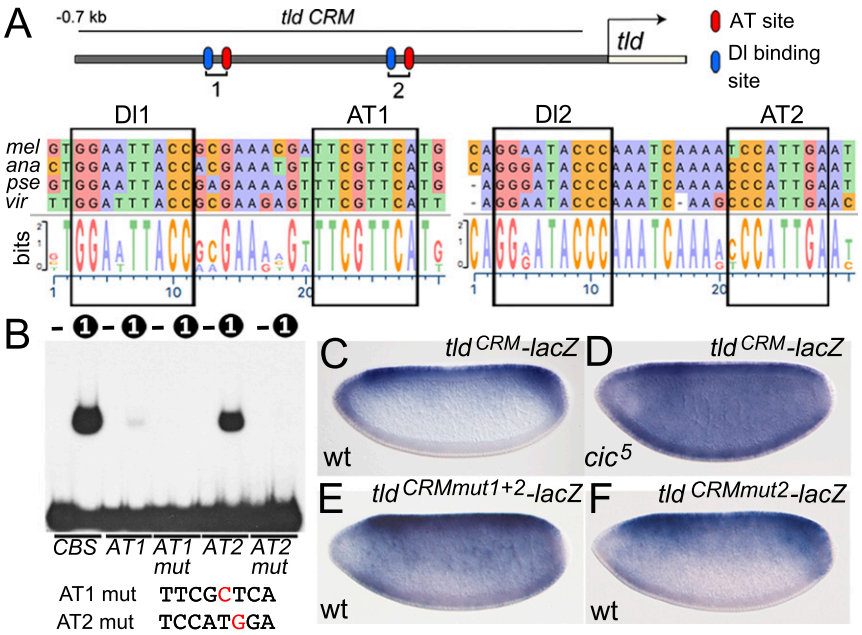

Fig. S4. Suboptimal DNA binding of Cic controls *tld* expression. (*A*) Diagram of the *tld* CRM containing linked DI and AT sites; the DI sites have been previously shown to be essential for repression in ventral regions (1). Alignments showing the conservation of these sites and their corresponding position weight matrices are also shown. *ana*, *Drosophila ananassae*; *mel*, *Drosophila melanogaster*; *pse*, *Drosophila pseudoobscura*; *vir*, *Drosophila virilis*. (*B*) EMSA using the HMG-C1 construct and labeled as in Fig. 2*D*. Wild-type and mutated *AT1* and *AT2* probes contain the corresponding AT sites from the *tld* CRM; the sequence of the mutant sites is shown below, with their respective substitutions marked in red. Note that the AT sites exhibit low (AT1) or intermediate (AT2) affinities for Cic. (*C*–*F*) Expression of intact and mutated *lacZ* reporters driven by the *tld* CRM in the indicated backgrounds. The $tld^{CRMmut1+2}$-*lacZ* carries the AT1 and AT2 mutations shown in *B*, whereas the $tld^{CRMmut2}$-*lacZ* contains only the AT2 mutation. Note the strong derepression of the intact tld^{CRM}-*lacZ* reporter in cic^5 embryos (compare *C* and *D*). Also, mutation of both AT sites in tld^{CRM}-*lacZ* causes clear derepression in ventral regions in an otherwise wild-type (wt) background (*E*), whereas mutating only the second site leads to a mild, partial effect (*F*). Thus, Cic regulates *zen* and *tld* expression through related AT sites in their CRMs. The embryos in *C*–*F* were photographed at 200× magnification.

1. Kirov N, Childs S, O'Connor M, Rushlow C (1994) The *Drosophila dorsal* morphogen represses the *tolloid* gene by interacting with a silencer element. *Mol Cell Biol* 14:713–722.

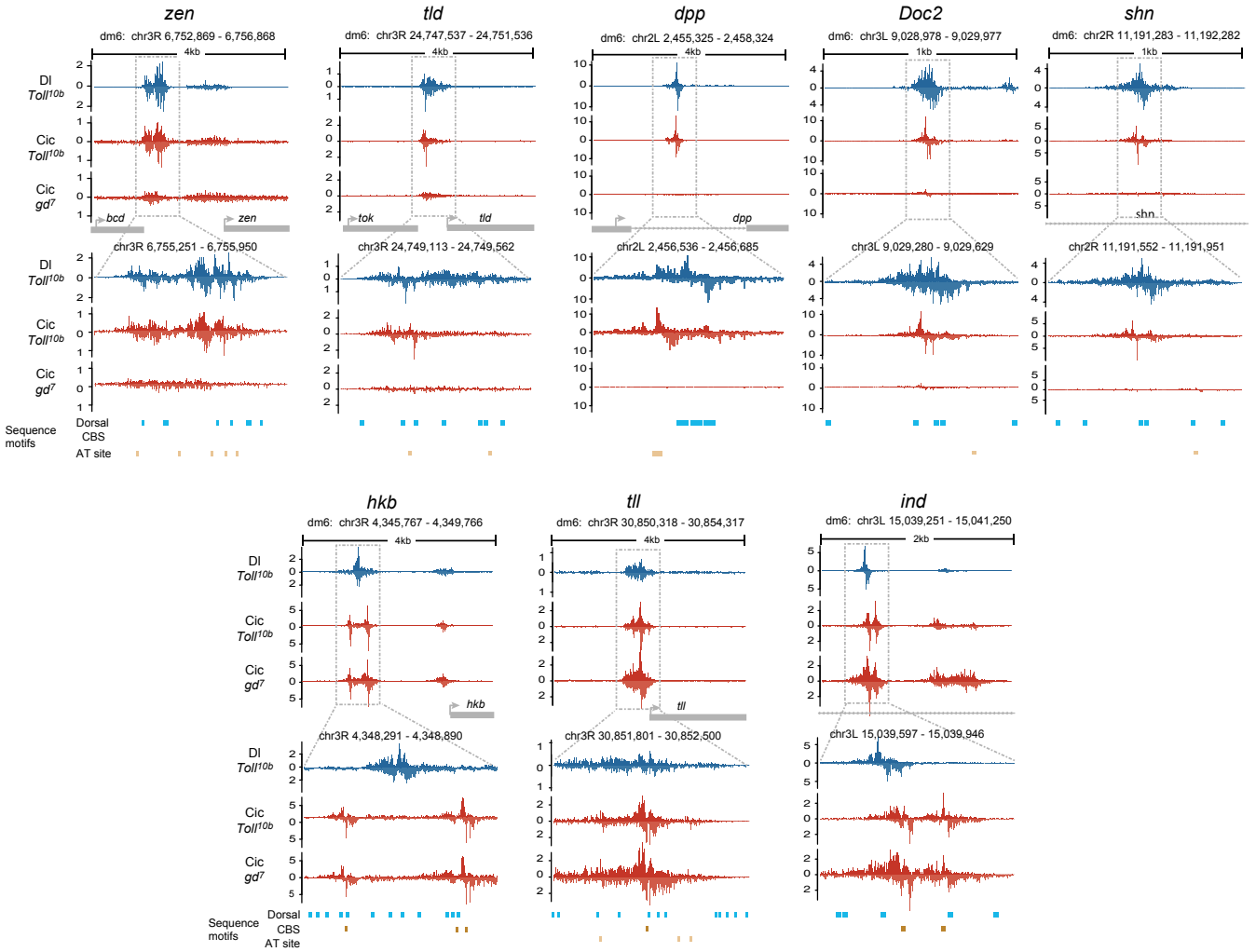

Fig. S5. ChIP-nexus signals of Dl and Cic at selected target genes. (*Upper*) Profiles corresponding to five dorsally expressed genes that are recognized by Cic in a Dl-dependent manner (i.e., Cic binding to these genes is lost in gd^7 embryos lacking Dl protein in the nucleus). (*Lower*) Signals obtained at three RTK-regulated genes that are bound by Cic independently of Dl. Zoom-in views also show Dl motifs (GGRWWTTCC with up to two mismatches) and Cic motifs: CBS (TSAATGAA with no mismatch) or AT (TSAATGAA with one mismatch), irrespective of whether the motif matches ChIP-nexus footprints. Note the prevalence of AT-like and CBS motifs in the Dl-dependent and -independent sets of genes, respectively.

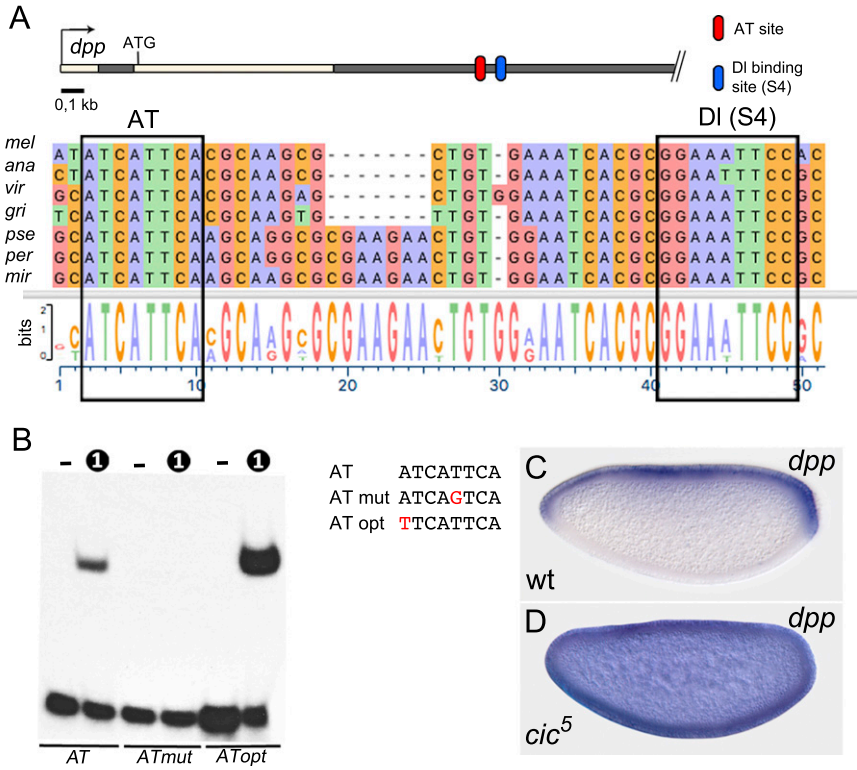

Fig. S6. Regulation of *dpp* expression by Cic. (*A*) Diagram of the *dpp* gene region containing the linked AT and DI binding sites identified by specific footprints in Fig. 3*A*. Open and gray boxes represent exons and introns, respectively. (*B*) EMSA using the HMG-C1 protein and labeled as in Fig. 2*D*; unlabeled lanes do not contain protein. Probes are indicated below the gels. The natural and mutated AT sequences are also shown, with substitutions indicated in red. Binding of HMG-C1 to the intact AT sites is about threefold weaker than to the optimized site. Patterns of *dpp* mRNA expression in embryos from wild-type (wt) (*C*) and cic^5 (*D*) females; note the strong derepression in the mutant background. Photographs in *C* and *D* were taken at 200× magnification.

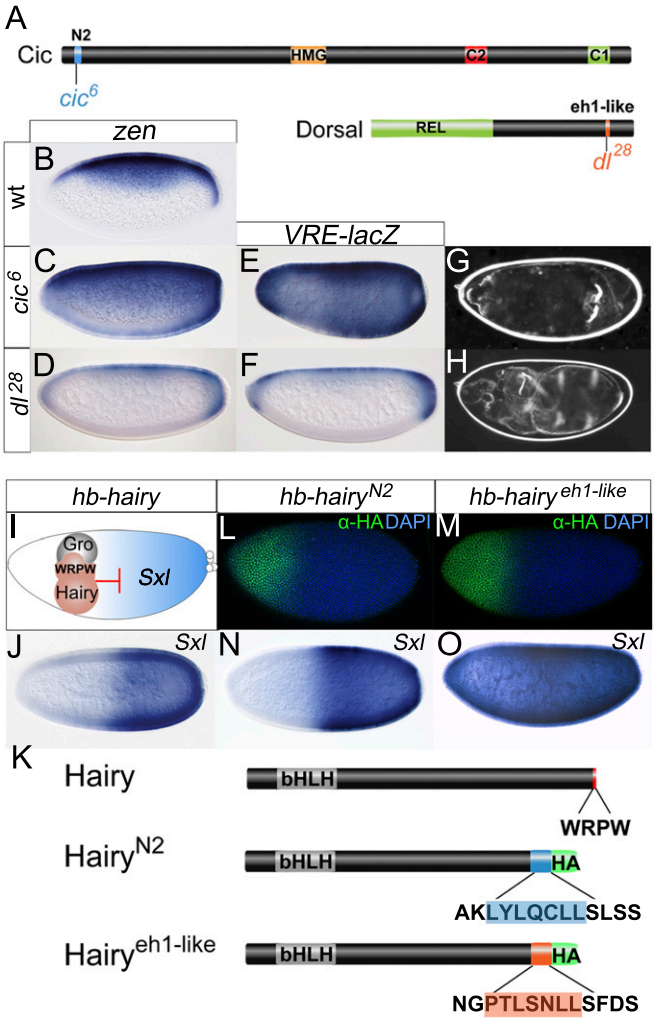

Fig. S7. Gro-dependent N2 motif of Cic is essential for repression of *zen*. (*A*) Diagram of Cic-S and Dorsal indicating the positions of the cic^6 and dl^{28} mutations. cic^6 eliminates the LY dipeptide from the N2 motif (LYLQCLL); dl^{28} removes all residues C-terminal to the threonine present in the eh1-like motif (PTLSNLL). REL, Rel homology domain. (*B–F*) Expression of *zen* and *VRE-lacZ* in embryos produced by wild-type (wt), cic^6, and dl^{28} females. Note the strong ventral derepression in cic^6 embryos. In contrast, expression of both targets in dl^{28} embryos appears narrower than in wild-type embryos, which is consistent with expansion of the Dorsal nuclear gradient upon truncation of its C-terminal region (1). (*G* and *H*) Cuticle phenotypes of cic^6 and dl^{28} embryos. The cic^6 embryos show a strong AP phenotype similar to that caused by cic^5. The dl^{28} embryos are lethal and exhibit defects in DV polarity. Together, these results indicate that the N2 motif of Cic is more important for *zen* repression than the eh1-like element of Dorsal. (*I*) Schematic representation of the *Sxl* repressor assay, an in vivo approach for analyzing the activity of repressor domains (ref. 2 and references therein). In this assay, expression of the Hairy segmentation protein under the control of the *hunchback* (*hb*) promoter leads to repression of *Sxl* in the anterior half of female embryos. Repression depends on the WRPW Gro-interacting motif of Hairy, but replacing this motif with autonomous repressor domains also leads to repression. (*J*) *Sxl* expression in an *hb-hairy* transgenic embryo. (*K*) Diagram of Hairy and two derivatives where the WRPW motif has been replaced with either N2 or the eh1-like motif of Dorsal. Both proteins are tagged with HA. The basic helix–loop–helix (bHLH) DNA binding domain of Hairy is also indicated. (*L* and *M*) Expression of the $Hairy^{N2}$ and $Hairy^{eh-1-like}$ fusions driven by the *hb* promoter; both proteins accumulate at similar levels in anterior regions of the embryo. (*N* and *O*) *Sxl* expression in embryos carrying the $hb\text{-}hairy^{N2}$ and $hb\text{-}hairy^{eh-1-like}$ transgenes; only the N2 motif is capable of effectively repressing *Sxl*. Images in *B–H*, *J*, and *L–O* were obtained at 200× magnification.

1. Isoda K, Roth S, Nüsslein-Volhard C (1992) The functional domains of the *Drosophila* morphogen *dorsal*: Evidence from the analysis of mutants. *Genes Dev* 6:619–630.
2. Jiménez G, Paroush Z, Ish-Horowicz D (1997) Groucho acts as a corepressor for a subset of negative regulators, including hairy and engrailed. *Genes Dev* 11:3072–3082.

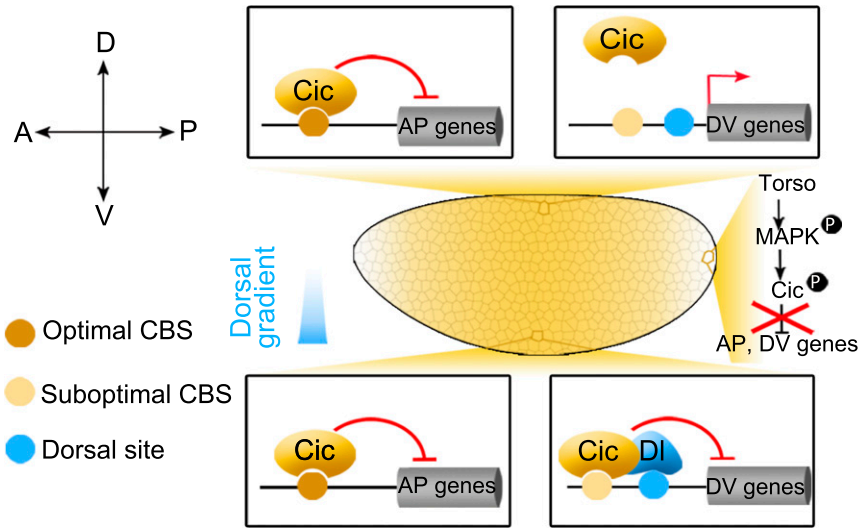

Fig. S8. Model of Cic regulatory functions in the AP and DV axes. Cic represses AP target genes by binding to high-affinity sites in a broad central domain of the embryo (yellow). Instead, repression of DV targets occurs through low-affinity sites and is restricted to ventral regions containing nuclear Dl protein. Both AP and DV targets are expressed at the poles, where Cic is directly down-regulated by Torso signaling. A, anterior; D, dorsal; P, posterior; V, ventral.

Table S1. *cic* and *dl* alleles generated via CRISPR-Cas9

Allele	Type of lesion	Base change*	Protein change
cic^{5}	Frameshift	A14T, Δ16–26	Frameshift from residue 5
cic^{6}	Deletion	Δ39–44	Δ13–14
dl^{28}	Frameshift	Δ5,819–5,829	L639A, frameshift from residue 640

*Bases numbered from the start of translation, corresponding to 3R:20293449 for *cic* and 2L:17437048 for *dl* in *Drosophila melanogaster* genome release 6.18.