

# Common and distinct DNA-binding and regulatory activities of the BEN-solo transcription factor family

Qi Dai,<sup>1</sup> Aiming Ren,<sup>2</sup> Jakub O. Westholm,<sup>1,3</sup> Hong Duan,<sup>1</sup> Dinshaw J. Patel,<sup>2</sup> and Eric C. Lai<sup>1</sup>

<sup>1</sup>Department of Developmental Biology, <sup>2</sup>Department of Structural Biology, Sloan-Kettering Institute New York, New York 10065, USA

Recently, the BEN (BANP, E5R, and NAC1) domain was recognized as a new class of conserved DNA-binding domain. The fly genome encodes three proteins that bear only a single BEN domain ("BEN-solo" factors); namely, Insensitive (Insv), Bsg25A (Elba1), and CG9883 (Elba2). Insv homodimers preferentially bind CCAATTGG palindromes throughout the genome to mediate transcriptional repression, whereas Bsg25A and Elba2 heterotrimerize with their obligate adaptor, Elba3 (i.e., the ELBA complex), to recognize a CCAATAAG motif in the Fab-7 insulator. While these data suggest distinct DNA-binding properties of BEN-solo proteins, we performed reporter assays that indicate that both Bsg25A and Elba2 can individually recognize Insv consensus sites efficiently. We confirmed this by solving the structure of Bsg25A complexed to the Insv site, which showed that key aspects of the BEN:DNA recognition strategy are similar between these proteins. We next show that both Insv and ELBA proteins are competent to mediate transcriptional repression via Insv consensus sequences but that the ELBA complex appears to be selective for the ELBA site. Reciprocally, genome-wide analysis reveals that Insv exhibits significant cobinding to class I insulator elements, indicating that it may also contribute to insulator function. Indeed, we observed abundant Insv binding within the Hox complexes with substantial overlaps with class I insulators, many of which bear Insv consensus sites. Moreover, Insv coimmunoprecipitates with the class I insulator factor CP190. Finally, we observed that Insv harbors exclusive activity among fly BEN-solo factors with respect to regulation of Notch-mediated cell fate choices in the peripheral nervous system. This *in vivo* activity is recapitulated by BEND6, a mammalian BEN-solo factor that conserves the Notch corepressor function of Insv but not its capacity to bind Insv consensus sites. Altogether, our data define an array of common and distinct biochemical and functional properties of this new family of transcription factors.

[Keywords: BEN-solo; insulator; repressor; transcription factor]

Supplemental material is available for this article.

Received September 2, 2014; revised version accepted November 20, 2014.

The BEN (BANP, E5R, and NAC1) domain was originally identified bioinformatically as a domain with  $\alpha$ -helical character that is present in a variety of metazoan and viral proteins (Abhiman et al. 2008). Several BEN-containing proteins were characterized to have chromatin-related functions, including mammalian BANP/SMAR1 (Kaul-Ghanekar et al. 2004; Rampalli et al. 2005), NAC1 (Korutla et al. 2005, 2007), BEND3 (Sathyan et al. 2011), and the C isoform of *Drosophila* mod(mdg4) (Gerasimova et al. 1995; Negre et al. 2010). Moreover, all of these proteins were linked to transcriptional silencing, albeit via different strategies such as interacting with matrix attachment sites (SMAR1), recruiting histone deacetylases or CoREST

(NAC1), or interacting with insulator sites [mod(mdg4)] or heterochromatin (BEND3). Altogether, these collected findings indicate an intimate connection between diverse BEN proteins and gene repression.

Previous studies of BEN proteins mostly concern factors that contain other characterized domains, and, in these cases, the function of the BEN domain has not received much attention (Abhiman et al. 2008). Recently, several studies indicate a new functionality for the BEN domain as an intrinsic sequence-specific DNA-binding domain. Mammalian RBB, a BEN and BTB domain protein, was found as a direct transcriptional repressor of

<sup>3</sup>Present address: Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, SE-17121 Solna, Sweden.  
Corresponding author: laie@mskcc.org  
Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.252122.114>.

© 2015 Dai et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

the HDM2 oncogene and an interactor of the nucleosome remodeling and deacetylase (NuRD) complex (Xuan et al. 2013). In vitro selection tests showed that RBB bound DNA as a homodimer to TGTCRSWWYYGWC-type sites. Two other studies in *Drosophila* provided the first functional analyses of the subclass of BEN proteins that contain only a single BEN domain and no other recognizable domains (i.e., “BEN-solo” proteins). Two of these proteins, Bsg25A (Elba1) and CG9883 (Elba2), along with the adaptor protein CG15634 (Elba3) associate with the asymmetric site “CCAATAAG” in the Fab-7 insulator element and are involved in insulator function (Aoki et al. 2012). In vitro gel shift assays demonstrated that DNA-binding activity on this site is a specific attribute of the heterotrimeric complex, with no binding being observed by any pairwise or single proteins tested. Genome-wide studies of the third *Drosophila* BEN-solo factor, Insensitive (Insv), showed its broad genomic association with TCYAATHRGAA sites, and in vitro studies demonstrated high-affinity binding of Insv homodimers to palindromic TCCAATTGGA sequences (Dai et al. 2013b). Reporter studies in culture cells and expression profiling of *insv* mutants indicate that Insv binding to its cognate site mediates transcriptional repression.

Curiously, Insv exhibits dual modes of transcriptional regulation. In addition to functioning as a direct repressor, it is also recruited to chromatin via a CSL-type transcription factor, which serves as the primary nuclear effector of the Notch signaling pathway (Lai 2004). In this context, Insv associates directly with CSL via its BEN domain, represses Notch pathway target genes, and is required phenotypically to oppose Notch signaling during neural cell fate specification and promote neurogenesis (Duan et al. 2011). Although a definitive mammalian ortholog of Insv is not apparent from the primary sequence, mammalian BEND6 has all of the key functional attributes of a CSL corepressor. In particular, BEND6 binds CSL directly via its BEN domain, associates with and represses Notch target genes, and restricts Notch signaling in neural stem cells, thereby opposing their self-renewal and promoting neurogenesis (Dai et al. 2013a).

These collected studies suggest that Insv and Elba exhibit distinct sequence preferences and strategies of DNA binding (homodimer vs. heterotrimer) and that this may be further associated with a distinct impact on gene regulation (repressor vs. insulator) and, ultimately, biological function (neural cell specification vs. segmental patterning). In this study, we characterize the three *Drosophila* BEN-solo factors in parallel. Unexpectedly, our molecular and structural analyses show that the BEN domains of all three proteins recognize the same high-affinity sequence as homodimers. In addition, Elba2 functions as a repressor whose activity via Insv-binding sites does not depend on the other ELBA factors. In contrast, binding to the ELBA site appears to be a property of the ELBA complex, and its capacity to mediate repression is limited. Reciprocally, we used genome-wide analysis to assign Insv as a novel component of class I insulators, including throughout the *Drosophila* Hox complexes. Finally, we found that Insv alone is endowed

with the capacity to regulate Notch signaling during peripheral neurogenesis. Indeed, mammalian BEND6 is better able to rescue neural phenotypes of flies compromised for CSL corepressor function than ELBA factors are. Altogether, we identified both distinct and overlapping functional properties of the three *Drosophila* BEN-solo proteins, introducing unexpected complexity in their contributions to gene regulation and development.

## Results

### *The neural BEN-solo factor Insv is deployed throughout the blastoderm with ELBA factors*

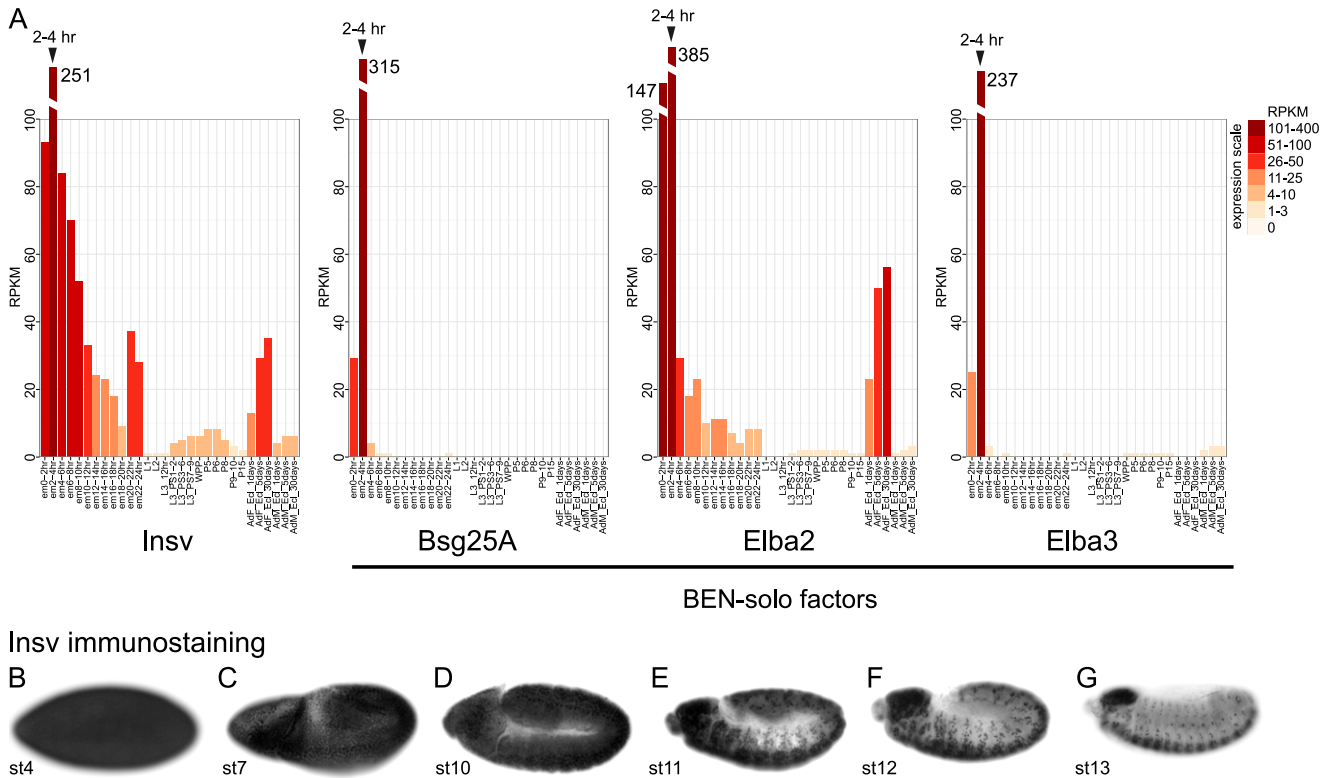
Of the three *Drosophila* BEN-solo factors, Elba1 (Bsg25A) and Elba2 (CG9883) reside in the ELBA heterotrimeric complex (Aoki et al. 2012). In this study, we chose to retain the Bsg25A nomenclature due to historical precedent (Singer and Lengyel 1997) and refer to the other Elba BEN-solo component as Elba2. As noted, Bsg25A and Elba3 reside in neighboring transcription units (with Bsg25A nested within CG11929), while the BEN-solo loci *Insv* and *Elba2* are also arranged as adjacent genomic loci (Aoki et al. 2012).

Our previous Insv immunostaining studies during larval and pupal stages highlighted its restricted spatial accumulation in the developing peripheral nervous system (PNS) for adult mechanosensory organs (Duan et al. 2011), consistent with studies of *insv* mRNA accumulation (Reeves and Posakony 2005). These patterns are seemingly distinct from the documented ubiquitous blastoderm expression of various ELBA factors. However, inspection of modENCODE developmental time-course RNA sequencing (RNA-seq) data (Graveley et al. 2011) indicated that *insv* mRNA levels are by far highest in 2- to 4-h embryos (Fig. 1A). Indeed, we confirmed strong ubiquitous accumulation of Insv protein in blastoderm stages (Fig. 1B). Thus, Insv exhibits two phases of expression: early/ubiquitous and later/neural-restricted (Fig. 1B–G). Notably, peak levels of Bsg25A, Elba2, and Elba3 are also found specifically in 2- to 4-h embryos (Fig. 1A). We confirmed the strong and specific blastoderm expression of Bsg25A and Elba3 proteins (Supplemental Fig. 1) and note that the presence of Elba2 in later embryonic stages corresponds mostly to its expression in the procephalic ectoderm primordium (Supplemental Fig. 1).

Overall, the *Drosophila* BEN-solo factors exhibit some distinct aspects of spatial deployment, but the observation of a common dominant expression pattern prompted us to investigate whether they potentially function in any common regulatory processes.

### *Three Drosophila BEN-solo proteins have a similar capacity to recognize Insv sites*

The heterotrimeric ELBA complex binds a sequence distinct from the Insv consensus, and individual ELBA BEN-solo proteins were reported not to bind DNA (Aoki et al. 2012). However, inspection of BEN domain alignments showed that Bsg25A and Elba2 share C-terminal tail residues with Insv that mediate its DNA-binding



**Figure 1.** Codeployment of *Drosophila* BEN-solo factors in the early embryo. (A) Developmental expression patterns of three BEN-solo factors—Insv, Bsg25A (also known as Elba1 or CG12205), and Elba2 (also known as CG9883)—and the ELBA complex cofactor Elba3. RNA-seq data from the modENCODE project: Expression values (in reads per kilobase per million mapped reads [RPKM]) were color-coded according to the key at the top right. Insv and Elba2 were detected at various developmental stages, but the predominant accumulation of all of these transcripts occurred during blastoderm stages in 2- to 4-h embryos. (B–G) Immunostaining of *Drosophila* embryos for Insv protein highlights its biphasic expression pattern. It is initially expressed ubiquitously in the early embryo (B) and later throughout the developing ectoderm (C,D) but becomes highly restricted to the developing CNS and PNS (E–G).

specificity (Dai et al. 2013b) (Fig. 2A, amino acids in alignment marked with blue asterisks). This motivated us to test whether Bsg25A or Elba2 might recognize Insv target sites. We addressed this in S2 cells using a luciferase reporter linked to the *actin* enhancer and four Insv-binding sites (act-4xInsv-luc) (Fig. 2B). We previously showed that this reporter is sensitive to both repression and activation (Dai et al. 2013b) and exploited it for both types of assays in this study.

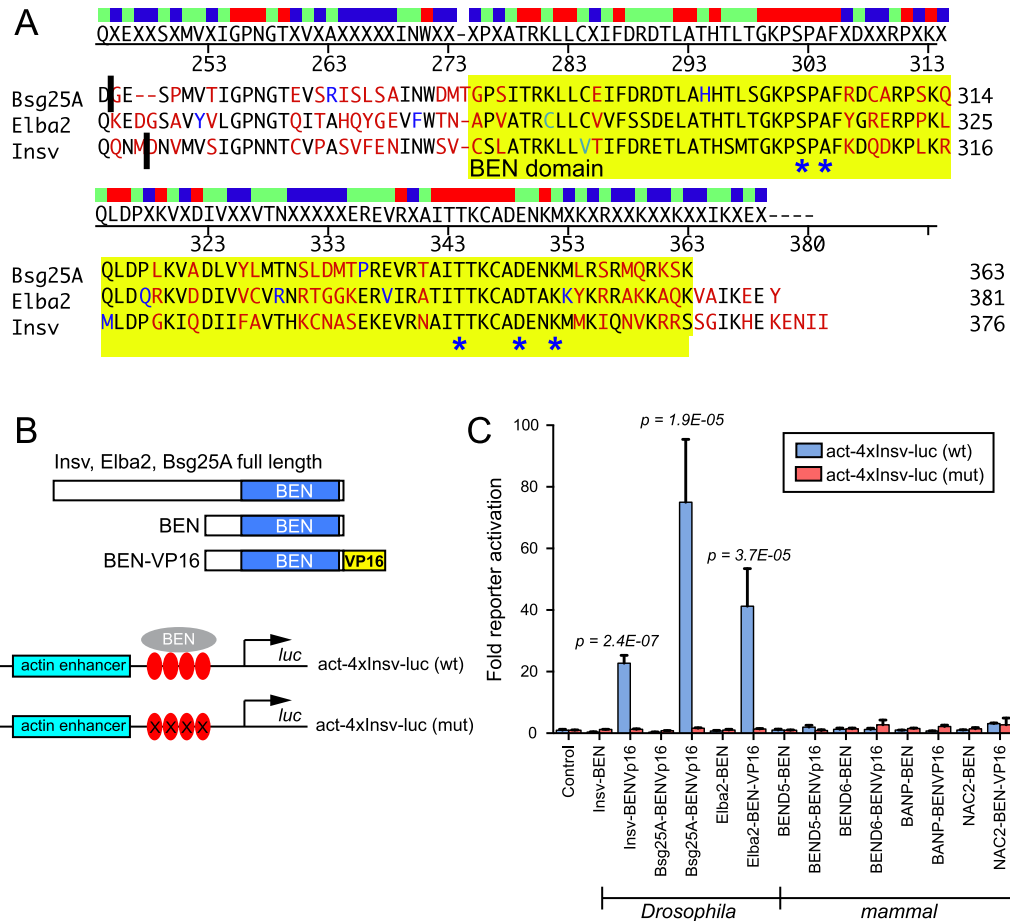
We divorced DNA-binding activity from other functionalities of these repressor/insulator proteins by fusing their BEN domains to the viral transactivation domain VP16 (Fig. 2B). In this manner, an Insv-BEN-VP16 fusion protein serves as a strong activator of act-4xInsv-luc (Dai et al. 2013b). Similarly, both Bsg25A-BEN-VP16 and Elba2-BEN-VP16 constructs strongly activated act-4xInsv-luc in S2 cells (Fig. 2C). In this experimental setup, the amount of reporter activation served as a rough gauge of the affinity of these BEN domains to the Insv palindrome. Therefore, these data suggested that all three BEN-solo proteins share affinities comparable with the Insv consensus site, since Elba2-BEN-VP16 activated the Insv reporter ~40-fold, while Bsg25A-BEN-VP16 exhibited ~80-fold reporter activation (Fig. 2C). The specificity of reporter regulation in these tests was shown by the fact

that all BEN-VP16 fusions were inert on mutant reporters containing base changes in the Insv palindrome (Fig. 2B,C).

By comparison, the BEN domain of mammalian BEND5 is detectably competent to bind the Insv site in vitro, but BEND5-BEN-VP16 achieved <1.7-fold activation of act-4xInsv-luc (Dai et al. 2013b), presumably reflecting that BEND5 prefers a different consensus sequence. We tested the BEN domains of three additional mammalian proteins (BEND6, NAC2, and BANP) as VP16 fusions, and these were incapable of activating act-4xInsv-luc (Fig. 2C), consistent with the lower identities of their BEN domains to these fly BEN domains (Supplemental Fig. 2). Therefore, the robust transcriptional activation capacity of Elba2-BEN-VP16 and Bsg25A-BEN-VP16 on this reporter is a specific attribute of the *Drosophila* BEN-solo factors. We conclude that the BEN domains of all three *Drosophila* BEN-solo factors specifically exhibit high affinity for the same palindromic sequence.

#### Determination of the crystal structure of Bsg25A on the palindromic Insv-binding site

We sought structural confirmation of the similar DNA-binding activities of Insv and ELBA factors. To this end,



**Figure 2.** All three *Drosophila* BEN-solo factors are intrinsic DNA-binding proteins. (A) Alignment of the BEN domains and immediate upstream regions from *Drosophila* Ins v, Bsg25A, and Elba2. The BEN domain is highlighted in yellow, and residues previously demonstrated to mediate nucleotide-specific recognition by the Ins v BEN domain are indicated with blue asterisks; note that these are conserved by all of the fly BEN-solo factors. The regions used for previous structural studies of Ins v and for current structural studies of Bsg25A are denoted by the black vertical lines and extend to their respective C termini. Red, green, and blue bars in the consensus indicate the residues that are identical in all three proteins, retained in two of the proteins, and present in only one of the three proteins, respectively. (B) The domain structure of *Drosophila* BEN-solo proteins and a strategy to convert them into BEN-VP16 transcriptional activator proteins. Shown *below* are reporters used to detect the regulatory activity of BEN factors. These contain a multimer of Ins v consensus sites, and specific regulation was judged with respect to a mutant Ins v reporter. (C) BEN-VP16 fusions of *Drosophila* Ins v, Bsg25A, and Elba2 all induce robust and specific activation of 4x-Ins v-luc, whereas none of four mammalian BEN-VP16 fusions had similar activity. Control refers to cotransfection of pAC5-STABLE2, a tricistronic expression construct expressing GFP, RFP, and Neo. Mean reporter values were averaged from three biological replicates, each in technical triplicate. Standard deviation error bars are shown, and *P*-values were calculated using *t*-test; two tails, one type.

we expressed and purified the BEN domain of Bsg25A (residues 248–362) and cocrystallized this bound to a self-complementary GTTCCAATTGGAA 13-mer DNA duplex that we previously crystallized with the Ins v BEN domain (Dai et al. 2013b). We obtained crystals that diffracted to 3.2 Å resolution, and the X-ray statistics are listed in Table 1.

We present an electrostatic surface representation of the complex in Figure 3A. Each asymmetric unit contains four BEN domains and two DNA duplexes such that a Bsg25A homodimer is bound to one DNA duplex. The DNA duplexes are involved in end-on-end stacking and are bound to the basic region of the Bsg25A homodimer surface. We also present two alternate views of the complex

in ribbon representations in Figure 3B. As with the Ins v-BEN:DNA complex, we detected a large number of intermolecular contacts in the complex, with each BEN domain interacting with both strands of the duplex. Helix  $\alpha 5$  (residues 335–360) of the BEN domain is positioned within the major groove, while elements of a long loop (residues 294–316) between helix  $\alpha 3$  and helix  $\alpha 4$  are positioned in the minor groove of the duplex.

Despite different packing interactions, we observed that the Bsg25A BEN domain uses recognition principles similar to those of the Ins v BEN domain (Dai et al. 2013b). The overall relationship of the Bsg25A and Ins v BEN domains are shown by their structurally annotated linear sequences (Fig. 3C). Their relationship is made more

**Table 1.** Crystallographic statistics for Bsg25A-BEN domain: DNA complexes

| Bsg25A-BEN(248–362)–DNA                             |                             |
|---|-----------------------------|
| Data collection                                     |                             |
| Wavelength  | 0.9792 Å                    |
| Space group   | P31                         |
| Cell dimensions                                     |                             |
| <i>a</i> , <i>b</i> , <i>c</i>                      | 92.6 Å, 92.6 Å, 63.0 Å      |
| $\alpha$ , $\beta$ , $\gamma$                       | 90°, 90°, 120°              |
| Resolution <sup>a</sup>                             | 50 Å–3.20 Å (3.30 Å–3.20 Å) |
| R <sub>sym</sub> or R <sub>merge</sub> <sup>a</sup> | 0.175 (0.616)               |
| <i>I</i> / $\sigma$ <i>I</i>                        | 7.1 (1.8)                   |
| Completeness <sup>a</sup>                           | 99.9% (100%)                |
| Redundancy <sup>a</sup>                             | 2.6 (2.5)                   |
| Unique reflections                                  | 9889                        |
| Refinement  |                             |
| Resolution  | 40.08 Å–3.21 Å              |
| Number of reflections                               | 9863                        |
| R <sub>work</sub> /R <sub>free</sub>                | 22.8%/27.7%                 |
| Number atoms  |                             |
| Protein   | 3324                        |
| DNA   | 1060                        |
| Water   | 1                           |
| B-factors   |                             |
| Protein   | 61.1                        |
| DNA   | 52.4                        |
| Water   | 36.6                        |
| RMSD  |                             |
| Bond lengths  | 0.002 Å                     |
| Bond angles   | 0.638°                      |

<sup>a</sup>Highest-resolution shell is shown in parentheses.

evident by the three-dimensional superimposition of their structures (Fig. 3D) as well as their DNA complexes (Fig. 3E). In these depictions, Insv is in green and Bsg25A is in pink, and we calculated a root mean square deviation (RMSD) as 0.631 Å for 82 backbone residues. Detailed inspection of the base-specific intermolecular contacts revealed conserved amino acid–nucleotide interactions between the Insv and Bsg25A BEN domains. For example, K352 and S302 of Bsg25A mediate strong base-specific interactions (heteroatom separations of ~2.8 Å), similar to the corresponding residues of Insv (Fig. 3F, Insv in green and Bsg25A in pink). On the other hand, the weaker interaction of Bsg25A A304 with its target base (3.6 Å separation) (Fig. 3G, dotted line) is also characteristic of Insv. Nevertheless, these structures are not identical, since the intermolecular interaction involving D349 in the Bsg25A-BEN:DNA complex is very weak (separation 4.0 Å) (Fig. 3G, dotted line), in contrast to the strong interaction of its counterpart residue, D351 (separation 2.8 Å), in the Insv-BEN:DNA complex (Dai et al. 2013b).

We performed EMSA-based gel shifts using native polyacrylamide gel (PAGE) on DNA base pair substitution mutants involving identified base-specific intermolecular interaction in the Bsg25A-BEN:DNA complex. We tested mutations in critical bases, including A7' (T8), C4 (G11'), and G10' (C5), all of which interact with the Bsg25A BEN domain (Fig. 3G). Consistent with the results of the Insv BEN domain (Dai et al. 2013b),

mutation of A7' (T8 in the opposite strand) to G, T, and C; mutation of C4 (G11') to A, T, or G; and mutation of G10' (C5) to T or C while retaining Watson-Crick pairing by compensatory changes on the partner strand all resulted in complete loss of binding by the Bsg25A BEN domain (Fig. 3H). Similar to the Insv BEN domain, mutation of G10' to A (C5 to T) was compatible with binding (Fig. 3H). Interestingly, the G10'A (C5T) change corresponds to the sequence of the asymmetric Elba site on the bottom strand, providing a rationale for how Bsg25A might associate with both Insv and Elba-type sites. Altogether, these structural and functional data provide strong evidence that the BEN domains of *Drosophila* BEN-solo proteins intrinsically recognize similar DNA sequences as homodimers.

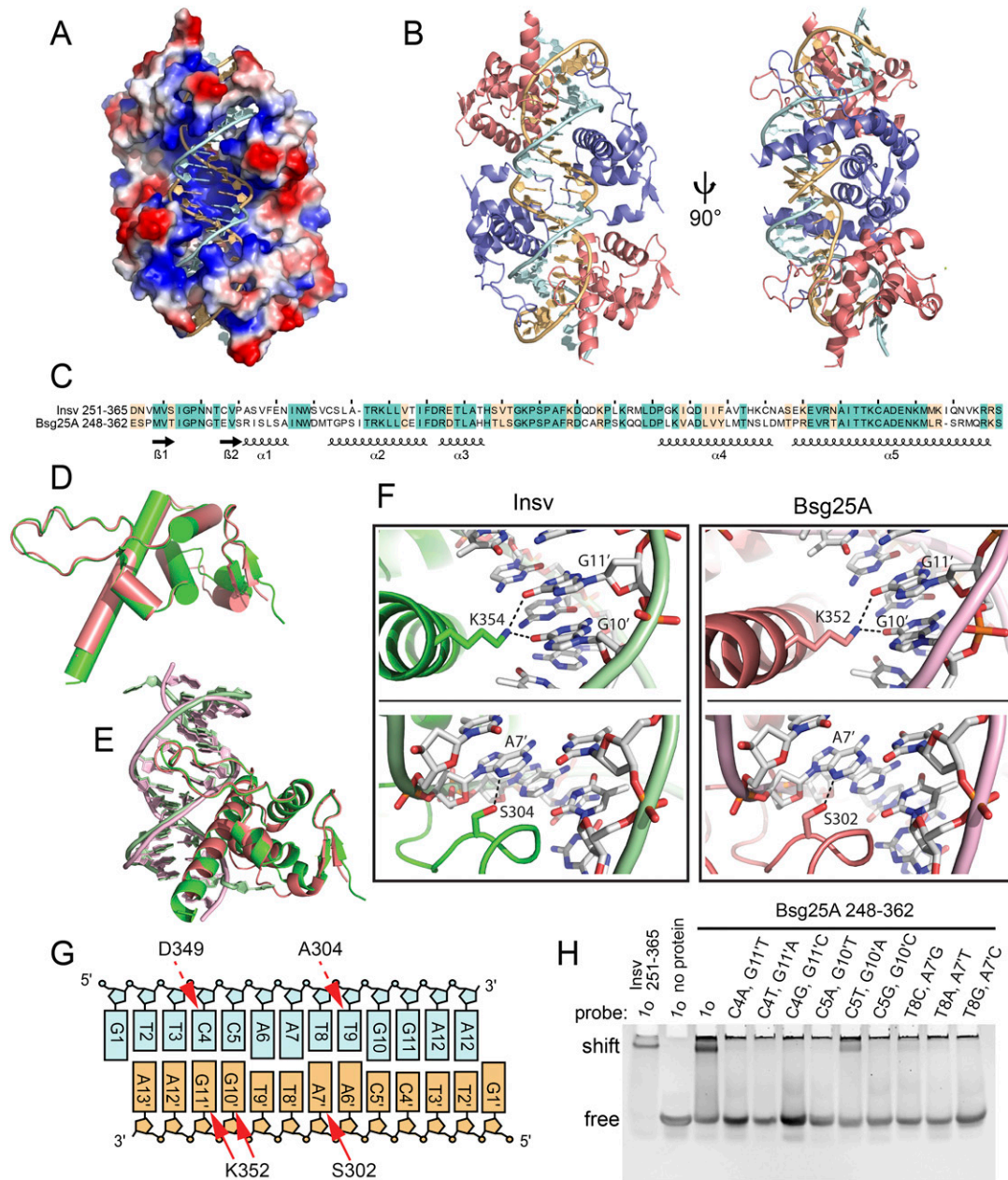
#### Regulatory capacities of the three BEN-solo factors on Insv and ELBA sites

Studies of the ELBA complex suggested its contribution to Fab-7 insulator activity (Aoki et al. 2012). On the other hand, Insv appeared to be a transcriptional repressor, based on the findings that genes located in proximity to its binding were up-regulated in *insv* mutants and that insertion of Insv-binding sites into reporter constructs conferred Insv-dependent decreases in reporter activity (Dai et al. 2013b). In general, insulator and boundary element proteins are not thought to be endowed with direct transcriptional repression activities, although it seems imaginable that such activities are not exclusive.

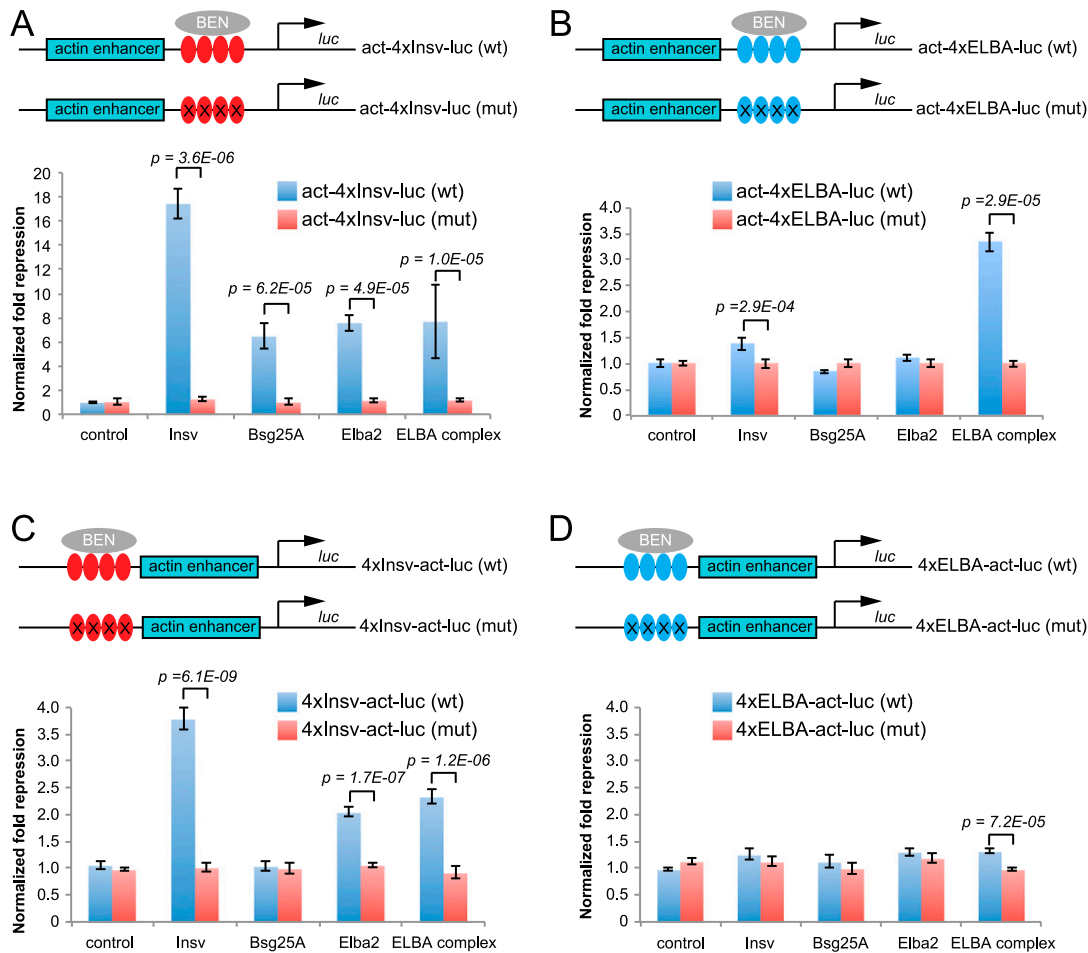
Having shown that all three BEN-solo factors have a similar capacity to recognize the cognate Insv consensus site, we compared the functional properties of the full-length proteins. S2 cells express negligible levels of any BEN-solo genes and only low levels of *Elba3* (Graveley et al. 2011) and thus serve as a “blank slate” for these tests.

We previously showed that full-length Insv can strongly repress act-4xInsv-luc (Dai et al. 2013b). Similarly, transfection of either Bsg25A or Elba2 alone repressed act-4xInsv-luc, and neither affected the expression of the corresponding mutant Insv reporter (Fig. 4A). These assays confirm that the ability of the BEN domains of all three fly proteins to associate with the Insv-binding site (Figs. 2,3) is recapitulated in the context of full-length proteins. The amount of repression conferred by Insv was twice that of either ELBA factor. Cotransfection of all three ELBA factors using a tricistronic 2A construct that ensured their coexpression did not potentiate repression relative to any individual BEN-solo Elba factor. This suggested that availability of the ELBA complex did not impart any unique regulatory capacity with respect to Insv-binding sites, perhaps due to homodimeric association of ELBA-class BEN-solo factors on Insv sites.

In reciprocal tests, we assessed whether Insv was capable of binding to the ELBA site. Consistent with previous observations that ELBA-class BEN-solo factors could not bind ELBA sites *in vitro* (Aoki et al. 2012), we found that neither Bsg25A nor Elba2 was able to repress an act-4xELBA-luc reporter *in vivo* (Fig. 4B). Coexpression



**Figure 3.** Structure of the Bsg25A BEN domain-DNA complex. (A) Electrostatic surface representation of the Bsg25A BEN-DNA complex consisting of two homodimeric BEN domain assemblies bound to two end-to-end stacked self-complementary 13-mer DNA duplexes. (B) Two views of the structure of the complex containing four BEN domains (colored dark blue and rust, in ribbon representation) bound to two end-to-end stacked self-complementary 13-mer DNA duplexes (strands colored in light blue and biscuit). (C) Comparison of amino acid sequences of the BEN domains of the Bsg25A and Insv proteins. Invariant residues are colored aqua, and the adoption of  $\alpha$ -helical and coiled regions is labeled *below*. (D) Superposition of the BEN domains of Bsg25A (rust) and Insv (green) [Protein Data Bank [PDB] ID: 4IX7]. (E) Superposition of the BEN-DNA complexes containing the BEN domains of Bsg25A (rust) and Insv (green) (PDB ID: 4IX7). (F) Common intermolecular contacts between a Lys side chain and the G10'-G11' step in the Insv complex (*left*) and the Bsg25A complex (*right*) and between a Ser side chain and the A7' base in the Insv complex and the Bsg25A complex. (G) Summary of specific structural nucleotide contacts in the complex. The solid lines indicate strong intermolecular contacts (heteroatom separation of  $\sim 2.8$  Å), while the dashed lines indicate weak intermolecular contact (heteroatom separation of 3.6-4.0 Å). (H) EMSA-based gel shift assays using the Insv consensus sequence (10) and visualized with ethidium bromide. The Insv BEN domain protein was used as a positive control. Note that substitutions were made at the base pair level. Of the mutant probes tested, only C5T (G10'A) retained binding to the Bsg25A BEN domain.



**Figure 4.** Regulatory capacities of BEN-solo proteins on Insv and ELBA targets. All assays used full-length BEN-solo or Elba3 proteins in transient transfection assays. Assays of the ELBA complex involved a multicistronic 2A vector that ensures their coexpression in individual cells. Control refers to cotransfection of pAC5-STABLE2. The wild-type (wt) and mutant (mut) reporter data were first normalized to their respective control transfection conditions, and the normalized mutant values were then used as a baseline for each normalized wild-type value. (A) All three BEN-solo proteins were able to repress act-4xInsv-luc, where the binding sites were located between the actin enhancer and the start of transcription. Combinations of Elba factors could also repress this reporter but may reflect the binding of homodimers. (B) Only heterotrimeric combinations of Elba factors could robustly repress act-4xELBA-luc, where the binding sites were located between the actin enhancer and the start of transcription. (C) Insv and Elba2 could repress a variant 4x-Insv-luc reporter (4xInsv-act-luc), where the binding sites were located distally to both the actin enhancer and the start of transcription. The demonstration of position-independent activity is a measure of true transcriptional repression activity. (D) No strong regulatory effects were observed on a 4x-ELBA-luc reporter where the sites were located distally (4xELBA-act-luc). In all graphs, mean reporter values were averaged from three biological replicates, each in technical triplicate. Standard deviation error bars are shown, and *P*-values were calculated using *t*-test; two tails, one type.

of the tripartite ELBA complex repressed act-4xELBA-luc, although this was modest compared with the repression achieved by any single BEN-solo factor on act-4xInsv-luc (Fig. 4B). Moreover, we found that Insv could not substantially repress act-4xELBA-luc. Therefore, the ELBA site appears to be uniquely regulated via the ELBA complex.

Since the binding sites in the above reporters lie between the actin enhancer and the transcription start site (TSS), they do not formally distinguish between transcriptional repression and insulator activity. However, Insv ChIP-seq (chromatin immunoprecipitation [ChIP] combined with deep sequencing) data revealed

that its direct binding sites are highly enriched near TSSs, a location perhaps not expected for insulator proteins, which are usually considered to reside between enhancers and/or transcription units (Negre et al. 2010). Nevertheless, a population of insulator proteins reside near TSSs and in 5' untranslated regions (UTRs) in *Drosophila* (Bushey et al. 2009; Negre et al. 2010) and mammalian (Barski et al. 2007; Kim et al. 2007; Vogelmann et al. 2011) cells, suggesting that insulator proteins might potentially influence transcription more directly.

To evaluate the positional dependency of BEN-solo proteins for gene regulation, we generated reporters in which Insv-binding sites were placed upstream of the

actin enhancer (4xInsv-act-luc) (Fig. 4C). According to classical views, insulator proteins should not influence gene transcription at these locations. However, transfection of full-length Insv yielded substantial repression of the 4xInsv-act-luc reporters bearing “upstream” sites (Fig. 4C), although the effects were weaker than with the original layout. This might be explained by a proximity effect, since Insv-BEN-VP16 activated 4xInsv-act-luc less robustly (Supplemental Fig. 3) relative to act-4xInsv-luc (Fig. 2C). Nevertheless, the activity that we detected was specific, since Insv did not affect the comparable reporters bearing mutations in the Insv sites (Fig. 4C). Therefore, Insv can directly repress transcription.

Similar tests provided evidence that Elba2 can also serve as a direct transcriptional repressor, since it was able to inhibit 4xInsv-act-luc via distal Insv sites (Fig. 4C). This broadens its regulatory capacity beyond its reported function in an insulator complex (Aoki et al. 2012). We note that Bsg25A was incapable of influencing this reporter and that coexpression of the three ELBA factors did not substantially alter regulation conferred by Elba2 alone (Fig. 4C). Finally, tests of a 4xELBA-act-luc reporter bearing “upstream” ELBA sites did not reveal substantial regulatory influences by any tested Insv/ELBA factors (Fig. 4D). Since Elba2 alone can mediate transcriptional repression, these data are consistent with a possibility that the ELBA site is less able than the Insv site to recruit Elba2 (via the ELBA complex). Reciprocally, these data are also consistent with the view that the ELBA site preferentially recruits the ELBA complex.

#### *Insv is a component of CP190/type I insulator complexes*

While mammalian insulators have principally been characterized solely by binding of CTCF protein, multiple proteins have collectively been found to associate with different functional *Drosophila* insulators. White and colleagues (Negre et al. 2010) investigated the genome-wide occupancy of six insulator-associated proteins—CTCF, CP190, BEAF-32, Su(Hw), Mod(mdg4), and GAGA factor (GAF)—using the ChIP-chip technique. They classified insulators into two major types: class I, typified by cobinding of BEAF-32/CP190/CTCF, and class II, defined by Su(Hw); Mod(mdg4) and GAF were less consistently associated with class I insulators (Negre et al. 2010).

We asked whether Insv genomic occupancies overlapped with those of insulator proteins by analyzing the 2- to 6-h and 6- to 12-h Insv ChIP-seq data (Dai et al. 2013b) and all of the available insulator protein ChIP-chip data. Figure 5A shows all pairwise comparisons of cobinding among these data sets. These analyses recapitulate the dominant signature of the class I insulator (Negre et al. 2010). Interestingly, Insv clearly showed substantial cobinding with class I insulator proteins, in particular with CP190 and BEAF-32. For example, 60% of Insv peaks overlapped with BEAF-32 peaks in both Insv data sets, and 72%–74% of Insv peaks overlapped with CP190 peaks (Fig. 5A). In general, the cobinding of Insv

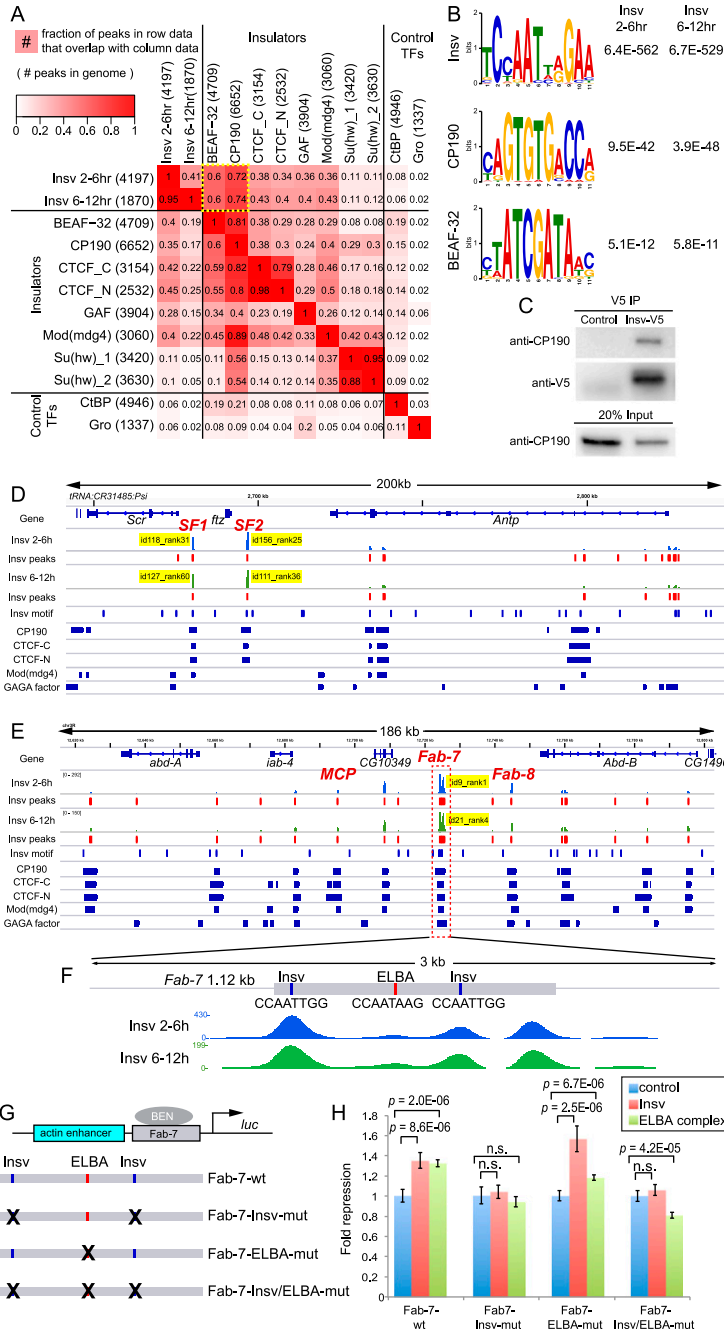
with various insulator proteins was more extensive at the earlier Insv time point (Fig. 5A), for which nearly twice as many peaks were called using parameters similar to the later Insv time point.

The overlaps of Insv with insulator proteins are more specific than between insulator proteins themselves, since the former are defined by relatively narrow ChIP-seq regions, whereas the latter are defined by broader ChIP-chip windows. As a gauge of specificity, we compared the overlap of Insv with a collection of 23 embryonic ChIP-chip data sets for 13 other factors, including a selection of neural-related transcription factors and transcriptional regulators analyzed in the modENCODE project (Negre et al. 2011; Slattery et al. 2014) and a set of muscle transcription factors (Zinzen et al. 2009). Collectively, these showed minimal overlaps with Insv (Fig. 5A; see also Supplemental Fig. 4). For example, Insv ChIP-seq peaks from either time point overlapped only 10% of peaks of Scute (Sc) (Supplemental Fig. 4), a transcriptional activator that is required for *insv* expression (Reeves and Posakony 2005). In addition, Groucho (Gro) and CtBP are both transcriptional corepressors associated with Notch signaling (Paroush et al. 1994; Morel et al. 2001; Barolo et al. 2002), similar to Insv (Duan et al. 2011), making them relevant comparisons. However, Insv ChIP-seq peaks overlapped only 2% of Gro peaks and only 6%–8% of CtBP peaks (Fig. 5A). Reciprocally, the collection of control factors analyzed exhibited generally minor or minimal cobinding with any insulator proteins despite the fact that these analyses involved overlaps between genomically broadly defined elements. Perhaps the only notable trend was the larger degree of overlaps detected between many muscle transcription factors and GAF than between GAF and several of the insulator proteins (Supplemental Fig. 4A).

We used COOCCUR (Huen and Russell 2010) to assess the significance of transcription factor occupancy overlaps and performed 10,000 iterations. Systematic pairwise comparisons showed that the only significant overlaps were between CP190 and the other insulator proteins—BEAF-32, CTCF, and Mod(mdg4)—as well as between the Insv early data and both CP190 and BEAF-32 (Supplemental Fig. 4B; Supplemental Table 1). The statistical framework of COOCCUR appears to be conservative, since, for example, it does not report significant overlap between BEAF-32 and CTCF or between developmental time points of some individual factors (Supplemental Table 1). Nevertheless, these analyses support the notion that association of Insv with class I insulator complexes is specific.

To further investigate the relationship between Insv and class I insulator factor binding, we performed de novo motif analyses. As reported earlier (Dai et al. 2013b), MEME reports the Insv consensus site as by far the most enriched motif amongst the top 500 ChIP-seq peaks in both the early and late data sets. However, MEME reports additional statistically significant motifs. Remarkably, the CP190- and BEAF-32-binding sites were among the top motifs in both Insv ChIP-seq data sets (Fig. 5B). Indeed, discounting the repetitive motifs CTCTCT...





**Figure 5.** Insv exhibits extensive cobinding with class I insulator elements. (A) Heat map summarizing the genomic cobinding of Insv, various insulator proteins, and the control corepressors CtBP and Gro. See Supplemental Figure 4A for additional control transcription factors analyzed. Since the number of peaks for each data set (noted in parentheses) is different, the heat map is not symmetric (e.g., the fraction of “A” peaks that overlap with “B” is different from the fraction of “B” peaks that overlap with “A”). Among insulator proteins, a dominant overlap signature of class I factors is evident, including CP190, BEAF-32, and CTCF. Insv similarly overlaps well with these factors. In contrast, CtBP and Gro overlap minimally with Insv and exhibit far less overlap with insulator proteins than Insv does. (B) De novo motif analysis of the top 500 Insv ChIP-seq peaks identified Insv-, CP190-, and BEAF-32-binding sites. (C) Coimmunoprecipitation tests demonstrated association of transfected Insv with the endogenous class I insulator factor CP190. (D) Extensive cobinding of Insv and class I insulator factors within the ANT-C. Highlighted in yellow are insulators that flank the *ftz* gene and are bound by CP190 and CTCF, both of which contain Insv consensus sequences and are among the top Insv-bound peaks genome-wide. (E) Extensive cobinding of Insv and class I insulator factors within the BX-C. The *Fab-7* region is the top Insv-bound region genome-wide in the early Insv ChIP-seq data set and fourth highest bound region in the late Insv data set. Note that BEAF-32 tracks are not shown in D and E, as there were few bound regions in the Hox clusters (Supplemental Fig. 4C) even though BEAF-32 generally exhibits extensive cobinding with CP190 and Insv. (F) Zoomed view of *Fab-7*, which contains two strong Insv ChIP-seq peaks that center on perfect consensus motifs, with weak binding detected on the ELBA site; the downstream Insv peak is not accounted for by a known site. (G) Schematics of the wild-type and mutant *Fab-7* reporters used in H. (H) Insv and the Elba complex repress a *Fab-7* reporter primarily via the Insv consensus motifs. Control is coexpression of pAc5-STABLE2. Mean reporter values were averaged from three biological replicates, each in technical triplicate. Standard deviation error bars are shown, and P-values were calculated using t-test; two tails, one type. Note that transfection of the ELBA complex yielded mild activation of the *Fab-7*-Insv/ELBA-mut reporter. While this effect was reproducible, we do not have an explanation for this as a direct effect on this reporter.

and TTTTTT... (and their complements), among the Insv-early peaks, CP190 was the second highest enriched motif, and BEAF-32 was the fourth highest enriched motif (the third best motif, TGGCARCNC, does not correspond to a known binding site). Similarly, in the Insv-late peaks, the top motifs after Insv were CP190, BEAF-32, and the above mystery site. These observations demonstrate an intimate connection between Insv-bound regions and the genomic occupancy of known class I insulator factors.

We sought physical evidence for Insv as a component of insulator complexes. Of note, we observed that Insv is

bound robustly at six out of seven classical functional insulators (Supplemental Fig. 4C), all of which are bound by CP190, but most of which lack BEAF-32 (Negre et al. 2010). We therefore focused on the potential interaction of Insv with CP190. We transfected S2 cells with a V5-tagged Insv construct and assessed potential interactions with endogenous CP190 using coimmunoprecipitation. Indeed, Insv was able to efficiently pull down CP190 in this assay (Fig. 5C). We conclude that Insv is a component of *Drosophila* class I insulator complexes and in particular forms a physical complex with CP190.

### Association of *Insv* with *Hox* insulator complexes

White and colleagues (Negre et al. 2010) observed that class I insulators generally separate transcription units and furthermore partition *cis*-regulatory regions of individual genes. They specifically noted this to be the case within the Antennapedia complex (ANT-C), a classically studied region known to harbor multiple insulator regions that are important for correct regulation of its resident homeobox genes, including HOX genes. We noted that the ANT-C contains an abundance of loci that were highly bound by *Insv*, many of which were associated with consensus *Insv*-binding sites and/or furthermore associated with class I insulator factors (Fig. 5D). For example, among ANT-C insulators previously cited (Negre et al. 2010), one (SF1) is located between *Scr* and *ftz* and was suggested to be important for proper segregation of *Scr* regulatory elements from activating *ftz* (Maeda and Karch 2003). This insulator is the 31st highest bound region in the *Insv* 2- to 6-h data, contains an *Insv* consensus match, and is cobound by CTCF, CP190, and Mod(mdg4) (Fig. 5D). Strikingly, another proposed insulator (SF2) located on the other side of *ftz* (Maeda and Karch 2003) is the 25th highest bound region in the *Insv* 2- to 6-h data, also contains an *Insv* consensus match, and is cobound by core class I insulator factors (Fig. 5D).

The other *Drosophila* HOX genes are located in the Bithorax complex (BX-C), which contains, among other regulatory elements, the Fab-7 insulator from which the ELBA complex was purified. We observed that the BX-C is also rich in *Insv*-bound regions and were struck by the realization that the region encompassing Fab-7 is actually the top bound region in the entire genome in the *Insv* 2- to 6-h data set (Fig. 5E). In fact, the Fab-7 region contains two high-affinity *Insv* palindrome sequences within 700 base pairs (bp) of each other that flank the ELBA site, and the peaks of *Insv* CHIP-seq signals lie precisely on top of these consensus motifs (Fig. 5F); we also observed *Insv* binding to an adjacent region that does not contain either an *Insv*-like or ELBA-like site.

We tested whether binding of *Insv* in an insulator had functional influence. We introduced a previously characterized 1.12-kb Fab-7 fragment (Li et al. 2008) between the actin enhancer and basal promoter in a luciferase reporter construct and also generated versions bearing mutations in both *Insv* sites, the ELBA site, or all three sites (Fig. 5G). We assayed the responses of these reporters to cotransfected *Insv* or tricistronic ELBA complex (Bsg25A + Elba2 + Elba3) expression vectors. As shown in Figure 5H, both *Insv* and the ELBA complex were able to repress the Fab-7 reporter, indicating their functional interaction with this genomic region. Notably, mutation of the *Insv* sites alone rendered the Fab-7 reporter insensitive to both constructs (Fig. 5H). This indicated that most of the observed repression by both types of BEN constructs was mediated via the *Insv* motifs. These data are straightforward to rationalize for *Insv*, which appears to interact very modestly with the ELBA motif (Figs. 4B, 5F), and furthermore suggest that the “downstream” binding of

*Insv* at Fab-7 (Fig. 5F) may not contribute substantially. However, these data are perhaps unexpected for the ELBA complex, since the cognate ELBA site remains in this reporter variant. It may be that a single ELBA site confers only mild regulatory properties, a notion that might be consistent with our reporter assays that indicate that even four copies of ELBA sites only confer modest repression by the ELBA complex (Fig. 4B).

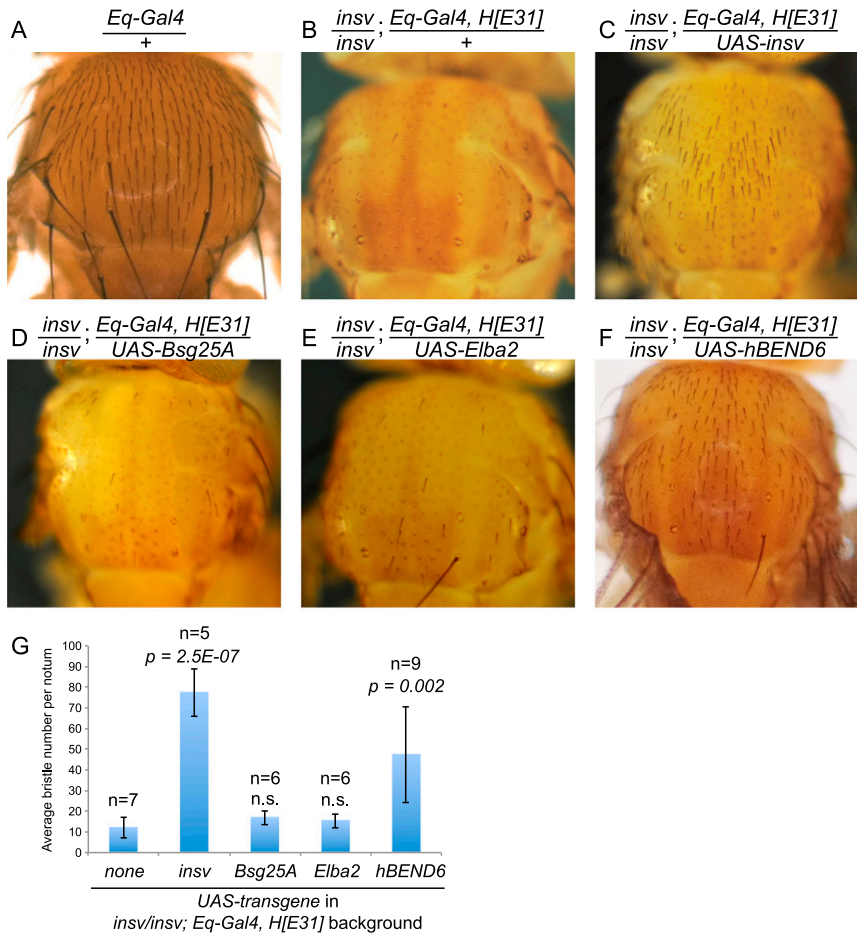
Reciprocally, both *Insv* and ELBA complex expression constructs could still repress the Fab-7-ELBA-mut reporters (Fig. 5H), consistent with the notion that both *Insv* and individual ELBA-class BEN-solo proteins can each recognize *Insv* sites that remain in this Fab-7 variant. Finally, neither *Insv* nor the ELBA complex were able to repress a triple mutant Fab-7 reporter that lacked both *Insv* and ELBA sites (Fig. 5H). The ELBA complex actually reproducibly activated this mutant reporter, although we presume this is due to some unrecognized nonspecific effect. Taken together, these results provide evidence that both *Insv* and the Elba complex can directly bind to Fab-7 to repress or block transcription.

### *Insv* and BEND6, but not Elba-class BEN-solo factors, repress Notch signaling during neural development

In our final functional comparisons, we returned to the PNS, the setting of our original studies of *Insv* function. We earlier reported that misexpression of *Insv* during adult PNS and wing development induced phenotypes consistent with inhibition of Notch signaling, including increased notum microchaete density and loss of wing margin (Duan et al. 2011). We generated inducible transgenes of the three ELBA factors and observed that Elba2 and Bsg25A did not have the capacity to influence mechanosensory bristle development as *Insv* did (data not shown). These data hinted at a unique function of *Insv* in regulating neurogenesis and/or Notch signaling.

A more stringent assay for endogenous *Insv* activity is the rescue of *insv* mutant phenotypes. On its own, *insv* mutants exhibit a mild sensory organ loss and double-socket organs that are indicative of loss of its activity as a corepressor for the Notch pathway transcription factor Suppressor of Hairless [Su(H)] (Duan et al. 2011). In particular, the double-socket organs are caused by gain of Notch signaling in shaft cells that convert them into their sister socket cell fates. The *insv* mutant phenotype is strongly enhanced by heterozygosity for another Su(H) corepressor encoded by *Hairless* (*H/+*) such that almost all notum sensory organs are double-socketed, with virtually no shaft structures present (Fig. 6A,B).

In the *insv/insv; H/+* background, general activation of *UAS-insv* in the notum prodomain using the *Eq-Gal4* driver results in substantial rescue of shaft structures (Fig. 6C), restoring approximately one-third of the normal number of shaft structures. This demonstrates that these mutant phenotypes are indeed due to loss of *Insv* function. In contrast, we observed that *UAS-Elba2* and *UAS-Bsg25A* were unable to provide substantial rescue of *insv<sup>-/-</sup>; H/+* microchaete fields (Fig. 6D,E). We ruled



**Figure 6.** *Insv* and mammalian BEND6 alone are able to regulate PNS development. Shown are adult notum from flies bearing the *Eq-Gal4* driver along with other UAS responder transgenes and/or mutations, as noted. (A) Normal pattern of bristle mechanosensory organs in *Eq-Gal4/+*. (B) *insv* homozygous flies lacking one copy of *Hairless* (*H*) exhibit severely aberrant notum bristle development, reflecting cell fate conversion of the shaft cell into its sister cell fate, the socket cell (i.e., “double-socket” organs). (C) Transgenic expression of *Insv* can substantially rescue shaft specification in *insv/insv; H/+* animals. Neither *Bsg25A* (D) nor *Elba2* (E) provides any rescue of double sockets, but expression of mammalian BEND6 substantially restores bristle development (F). Note that *Insv*/*Bsg25A*/*Elba2* exhibit similar high affinity for the *Insv* consensus sequence, whereas BEND6 does not recognize *Insv* sites (see Fig. 1C). (G) Quantification of notum bristle rescues in the indicated genotypes. For comparison, the average number of microchaete bristles in wild type is ~230. Error bars are standard deviation, and *P*-values were calculated using *t*-test; two tails, two types. (n.s.) Not significant.

out a trivial explanation that perhaps ELBA factors were not stably expressed or properly localized in these transgenic experiments, since immunostaining of wing imaginal discs demonstrated similar levels of nuclear-restricted proteins for all three BEN-solo factors (Supplemental Fig. 5). We were not able to test further higher-order combinations of ELBA factors due to the technical difficulty of expressing these in the requisite double-mutant backgrounds. However, it is clear that *Insv* did not require additional factors (e.g., ectopic *Elba3* or other BEN-solo proteins) to provide rescue in this assay.

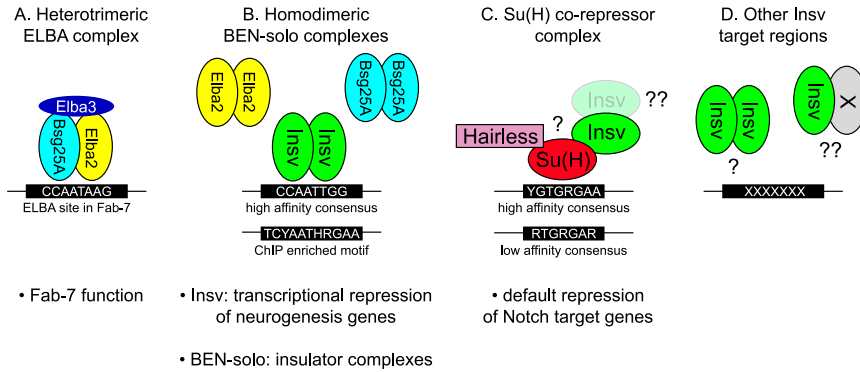
For comparison, we also misexpressed mammalian BEND6 using *Eq-Gal4* in this sensitized background. We previously demonstrated cross-species conservation of the ability of *Insv* and BEND6 to directly bind to the Notch transcription factors of *Drosophila* and mammals and further showed that BEND6 antagonizes Notch signaling in the mammalian brain (Dai et al. 2013a), similar to the action of *Insv* in the *Drosophila* nervous system. Remarkably, BEND6 was able to restore a majority of shaft structures to notum microchaetes in the *insv/insv; H/+* background (Fig. 6F), thus exhibiting a rescue capacity similar to that of *Insv* itself. We infer that this is due to its capacity to substitute for the Notch corepressor function of *Insv* because BEND6 has absolutely no capacity to regulate transcription via *Insv*-binding sites (Fig. 2C).

We quantified the degree of bristle rescues conferred by transgenic expression of the different BEN-solo proteins in the *insv<sup>-/-</sup>; H<sup>+/-</sup>* background (Fig. 6G). This analysis confirmed that only *Insv* and BEND6 could promote bristle development in this mutant condition. Altogether, these data provide striking evidence for a division of functional attributes of *Insv* that are shared with BEND6 (i.e., promotion of neurogenesis and inhibition of Notch signaling) and separately shared with other BEN-solo proteins (i.e., sequence-specific recognition of palindromic sites).

## Discussion

The recent literature that newly defined BEN-solo proteins as transcription factors presented contrasting views on the sequence recognition properties and functional activities of the homodimeric *Insv* complex (Dai et al. 2013b) and the tripartite ELBA complex (containing the BEN-solo proteins *Bsg25A* and *Elba2* and the adaptor protein *Elba3*) (Aoki et al. 2012). In particular, these complexes were reported to associate with distinct target sites and mediate distinct effects on gene regulation.

Our study provides a uniting view on several core properties of BEN-solo factors but also highlights their functional differences (Fig. 7). For example, we provide functional and structural evidence that all three *Dro-*



**Figure 7.** Overlapping and distinct functional complexes for BEN-solo proteins. (A) The heterotrimeric ELBA complex has been associated with a single target site thus far, located in the Fab-7 insulator. (B) All three *Drosophila* BEN-solo proteins efficiently recognize the binding site consensus originally established for Insv. Insv has been shown to broadly regulate neurogenesis-related genes through such binding sites and is shown here to also occupy a number of insulator regions through its cognate site. Since all BEN-solo proteins can recognize the same site, we predict that Bsg25A and Elba2 may prove to

occupy at least some of these Insv target regions. (C) Insv is recruited indirectly to DNA via a protein–protein interaction with the Notch pathway transcription factor Su(H), where it functions in a default repressor complex. Su(H) similarly recruits the corepressor H. It is not currently known whether H and Insv can co-bind Su(H) or occupy distinct Su(H) complexes (?) or whether Insv associates with the Su(H) corepressor complex as a monomer or homodimer (??). (D) As is true for many other transcription factors, Insv ChIP-seq data indicate binding to a number of genomic regions that lack either direct binding sites or Su(H)-binding sites. It remains to be elucidated whether any of these represent specific and/or functional recruitment events (?) through either nonconsensus binding sites or perhaps interactions with other chromatin proteins.

*sophila* BEN-solo factors are intrinsic DNA-binding proteins that efficiently recognize the same palindromic consensus sequence (Figs. 2, 3). We provide evidence that, when bound to DNA, Insv and Elba2 are capable of directly functioning as transcriptional repressors (Fig. 4), although their activity as such may have distance-dependent function, as has been documented for some other repressors. Our data are also consistent with the notion that Bsg25A may also harbor direct repression activity but that it is intrinsically weaker in this capacity than Insv or Elba2. This inference is based on the observation that the BEN domains from all three factors have comparable affinities for Insv-binding sites when measured using cellular reporter assays.

Although Insv can function directly as a repressor, we also provide evidence that Insv broadly occupies *Drosophila* insulator complexes (Fig. 5). Relatively few insulator proteins are known in most species; e.g., CTCF is the sole major insulator factor studied in mammals. In contrast, a number of proteins have been described to associate with various insulators in *Drosophila*. This apparent discrepancy may partially be attributed to the long history of genetic studies of diverse insulators and enhancer-blocking elements in intact *Drosophila*, whereas such elements are more difficult to assay specifically using cultured cells and transient methodologies. The heterotrimeric ELBA complex was previously localized to and functionally implicated at a single insulator (Fab-7) element (Aoki et al. 2012), but we provide here genome-wide evidence that Insv is generally bound at class I insulator elements. Indeed, the cobinding of Insv with core class I factors CP190 and BEAF-32 is nearly as extensive as that of CP190 with BEAF-32. Moreover, we found that a number of insulator elements, including previously characterized insulators in the ANT-C and BX-C complexes, are bound by Insv and contain consensus Insv sequences (Fig. 5). This implicates Insv as not only a transcriptional repressor frequently bound in proximity to core promoters

but also a component of insulator complexes that separate genes and enhancer elements.

Finally, we provide evidence for a major functional distinction of Insv and ELBA-class BEN-solo factors. Despite their documented functional similarities and broad coexpression in the early embryo, Insv alone is redeployed during specification of the nervous system (Fig. 1; Supplemental Fig. 1). There, it appears to have distinct functions in both the direct regulation of neurogenesis-related genes (Dai et al. 2013b) and the repression of Notch pathway genes as part of a Su(H) complex (Duan et al. 2011). Notably, Bsg25A and Elba2 can not substitute for the role of Insv in controlling Notch-mediated asymmetric cell divisions in the peripheral system (Fig. 6) even though they exhibit similar functional attributes in multiple other assays (e.g., Figs. 2–5). Instead, this particular activity of Insv can be fulfilled by transgenic expression of mammalian BEND6, a natural separation-of-function protein that conserves the Notch-regulating activity of Insv (Dai et al. 2013a) but not its capacity to recognize Insv-binding sites (Fig. 2C).

#### Evolutionary fluidity in functionalities in BEN proteins

Altogether, these findings provide a potentially useful perspective on thinking about the functions of other members of the relatively poorly studied BEN protein family (Abhiman et al. 2008). In particular, we hypothesize that many other BEN factors are likely to act as DNA-binding proteins but that different BEN proteins are also likely to have distinct influences on transcriptional regulation and that at least some BEN proteins are also likely to function as components of other protein complexes.

In the future, comparative structural studies will be valuable to elucidate whether the DNA-binding properties of the isolated BEN domains differ from the full-length proteins and how homodimeric and heterodimeric or heterotrimeric (e.g., with Elba3) complexes might differ.

Moreover, additional ChIP-seq studies should be valuable to dissect the regulatory attributes of BEN proteins, and our precedent of using sequential ChIP-seq reactions may help to distinguish the various functional complexes that individual BEN proteins may be involved in. For example, it would be highly informative to be able to segregate Insv-solo, Insv-ELBA, Insv-insulator, and Insv-Su(H) complexes along the genome at different times and places during development. Moreover, genome-wide comparisons of Insv and ELBA factor occupancies should be useful to understand the extent of their functional overlap, distinction, and possibly competition. For example, as we showed that all three BEN-solo proteins associate well with the Insv consensus site, we may expect binding maps of all three BEN-solo proteins to overlap at Insv sites. In this scenario, the genomic occupancy of Elba3, the adaptor protein in the ELBA complex, may be useful to distinguish locations of the ELBA complex from homodimeric or potentially heterodimeric binding of BEN-solo proteins.

## Materials and methods

### Molecular biology

Insv-BEN, Bsg25A-BEN, Elba2-BEN, BEND5-BEN, BEND6-BEN, BANP-BEN, and NAC2-BEN fragments were created by PCR amplification from Insv, Bsg25A, Elba2, human BEND5, mouse BEND6, mouse BANP, and human NAC2 cDNAs and cloned at EcoRI and XhoI sites (for Insv, BEND5, BEND6, BANP, and NAC2) or NotI and XhoI sites (for Bsg25A and Elba2) in pAcV5/HisA (Invitrogen). These plasmids were digested with XhoI and XbaI sites and ligated with a DNA fragment with the same sites encoding the VP16 activation domain to generate VP16 fusion constructs. The ORFs of Bsg25A and Elba2 were cloned to the NotI/XhoI sites of pAcV5/HisA to generate pAc-Bsg25A and pAc-Elba2\_V5.

We generated multicistronic 2A constructs as follows. For pAcElba1-T2A-Elba2-T2A-Elba3, the Elba1-T2A, T2A-Elba2-T2A, and T2A-Elba3 fragments were first amplified individually, and then a PCR product covering all three fragments was amplified and cloned into the NotI/XhoI sites of pAcV5/HisA. We also used pAC5-STABLE2 (Addgene, 32426), a multicistronic 2A construct expressing EGFP, RFP, and Neo, as a control.

Oligos containing two copies of wild-type or mutant Insv or Elba motifs that also cover 10-nucleotide upstream and downstream flanking sequences were synthesized in IDT (Integrated DNA Technology) and ligated to obtain four copies of Insv or Elba motifs (4xwt and 4xmut). These ligated fragments were then cloned into pAc-2T-Luc vector (Ryu and Arnosti 2003) at the AscI site to obtain pAc-4xwt-Luc and pAc-4xmut-Luc constructs or at the XhoI site to generate p-4xwt-Ac-Luc and p-4xmut-Ac-Luc. The wild-type Fab-7 element covering the 1.12-kb Fab-7 fragment (Li et al. 2008) was PCR-amplified and cloned to the pAc-2T-Luc vector via the AscI site. We then generated the panel of Insv/ELBA site mutant Fab-7 reporters using site-directed mutagenesis. Oligonucleotides used for cloning are provided in Supplemental Table 2.

### Protein expression and purification

The sequence corresponding to residues 248–362 of Bsg25A spanning the BEN domain was inserted into a modified pRSFDuet-1 vector (Novagen) in which the BEN domain was separated from a preceding His6-SUMO tag by a ubiquitin-like protease (ULP1)

cleavage site. Fusion proteins were overexpressed in *Escherichia coli* BL21(DE3) and purified through a Ni-NTA affinity column. The His6-SUMO tag was removed by ULP1 cleavage during dialysis against buffer containing 25 mM Tris (pH 8.0), 0.5 M NaCl, and 5 mM  $\beta$ -mercaptoethanol. After dialysis, the protein sample was further fractionated through a second Ni-NTA affinity column and a SP column (GE Healthcare) followed by gel filtration on a 16/60 G75 Superdex column (GE Healthcare). The final samples for crystallization were concentrated to ~20 mg/mL protein in buffer containing 20 mM HEPES, 150 mM NaCl, 5 mM DTT, and 20 mM MgCl<sub>2</sub> (pH 6.8). About 2 mg was usually produced from each liter of cell culture.

### Crystallization and structure determination

To generate the Bsg25A-BEN-DNA complex, a 13-mer self-complementary DNA duplex containing a central palindromic TCCAATTGGA-binding site was incubated with Bsg25A-BEN(248–362) in a 1:1.2–1.5 molar ratio for 1 h on ice. Crystals of the BEN-DNA complex were obtained in 0.1 M NaOAc (pH 4.6) and 25% PEG1000. The crystals were directly flash-frozen in liquid nitrogen for data collection.

X-ray diffraction data sets for the Bsg25A-BEN(248–362)-DNA complex were collected at NE-CAT beamlines at the Advanced Photon Source at Argonne National Laboratory and processed using the HKL2000 program (HKL Research). The structure of the complex was solved by a molecule replacement method using the Insv-BEN(251–365)-DNA complex (Protein Data Bank [PDB] ID code 4IX7) as the search model. Structure refinement was performed using PHENIX (Adams et al. 2002). The statistics for data collection and structural refinement of the Bsg25A-BEN(248–362):DNA complexes are summarized in Table 1. The atomic coordinates and structure factors were deposited to PDB (ID code 4X0G).

### Gel shift assay

The 13-mer self-complementary DNA duplex containing a central palindromic TCCAATTGGA-binding site or mutants in specific positions of the binding sites were used to characterize DNA binding of the Bsg25A-BEN(248–362) domain. Binding reactions contained 1 nM dsDNA and 5 nM protein dissolved in 5  $\mu$ L of binding buffer (20 mM HEPES, 150 mM NaCl, 5 mM DTT, 20 mM MgCl<sub>2</sub> at pH 6.8). Binding reaction mixtures were electrophoresed in 10% PAGE in buffer (50 mM Tris-acetate, 50 mM K-acetate, 5 mM Mg-acetate at pH 8.0) at 4°C. DNA was visualized by ethidium bromide staining.

### Cell culture and transient transfection

All transfections in this study were performed using *Drosophila* S2-R<sup>+</sup> cells grown in Schneider *Drosophila* medium containing 10% fetal calf serum. Cells were transfected in 96-well plate using the Effectene transfection kit (Invitrogen). Luciferase assays were performed as previously described (Dai et al. 2013b). Luciferase activity was measured using the dual-luciferase assay system (Promega), and expression was calculated as the ratio between the firefly and *Renilla* luciferase activities.

### Coimmunoprecipitation and Western blot

Three wells of  $2 \times 10^6$  cells in a six-well plate transfected with pAc-insv-V5 or the same amount of nontransfected control cells were harvested and lysed for protein extraction. Coimmunoprecipitation was performed as previously described (Dai et al. 2008). Mouse V5 antibody (Invitrogen) was used in immunopre-

precipitation. Rabbit anti-CP190 (gift of Elissa Lei, National Institutes of Health-National Institute of Diabetes and Digestive and Kidney Diseases) and mouse anti-V5 (Invitrogen) antibodies were applied in Western blotting and detected with HRP-coupled secondary antibodies, and the signals were developed with ECL Plus reagent (GE Healthcare).

#### Drosophila stocks and fly genetics

We used previously described transgenic stocks for *UAS-insv* (Duan et al. 2011) and *UAS-hBEND6* (Dai et al. 2013a). To generate transgenes for ELBA-class BEN-solo factors, we PCR-amplified the ORFs of Bsg25A and Elba2 and cloned them into Gateway Entry vector (pEntr-D-TOPO, Invitrogen). Following sequence verification, we transferred these into the destination vector pTMW (pUAST-6xMyc) using LR clonase (Life Technologies). These were injected according to standard methods (BestGene, Inc.).

For the bristle rescue assay, we recombined *Eq-Gal4* with *H[E31]* and then further crossed this to *insv[23B]* to obtain the *insv[23B]/CyO; Eq-Gal4, H[E31]/TM6B* tester stock. We similarly introduced *UAS-Insv*, *UAS-Bsg25A*, *UAS-Elba2*, and *UAS-hBEND6* transgenes into the *insv[23B]* background. We then crossed these to the tester stock and recovered *insv<sup>-/-</sup>; Eq-Gal4, H[E31]/UAS-BEN* animals for notum shaft quantifications. Normally, *insv<sup>-/-</sup>; H<sup>+/-</sup>* animals on their own are normally nearly completely double-socketed (Duan et al. 2011). All crosses were maintained at 25°C.

#### Computational analyses

We used sets of previous ChIP–chip regions for insulator proteins (Negre et al. 2010), modENCODE data (Negre et al. 2011; Slattery et al. 2014), or muscle transcription factors (Zinzen et al. 2009), as reported in these studies or downloaded from <http://data.modencode.org>. For *Insv*, we used previously reported ChIP-seq data (Dai et al. 2013b), which we analyzed previously using Quest (Valouev et al. 2008). This peak-calling program outputs regions with high ChIP signal and further subdivides these into individual peaks. In our previous analyses of these data, we used 200-bp ChIP-seq windows defined by extending 100 bp in either direction from the summits of Quest peaks (Dai et al. 2013b). De novo motif finding was carried out using MEME (Bailey and Elkan 1994) on the top 500 *Insv* peaks from each time point using these 200-bp *Insv* ChIP-seq windows.

The ChIP-seq peaks were generally much smaller than ChIP–chip windows and may have been biased to have lower overlap potential. In addition, there were frequently multiple ChIP-seq peaks contained within a single ChIP–chip window. Therefore, we decided to use the Quest “region” outputs in the overlap comparisons. These regions were still generally small: For the *Insv* 2- to 6-h data, the regions were 426 bp (SD 338), and for the *Insv* 6- to 12-h data, they were 444 bp (SD 316). We then analyzed the overlaps of *Insv* occupancy with Insulator proteins and control transcription factors as follows. We first computed the fraction of overlapping peaks for each pair of data sets (A and B) by counting the numbers <nr overlapping regions>/<nr regions in A> and <nr overlapping regions>/<nr regions in B>. We then estimated the statistical significance of the overlaps using COOCCUR (Huen and Russell 2010) to compute *P*-values using 10,000 resampling iterations.

#### Acknowledgments

We are grateful to Jiayu Wen for plotting the developmental expression profiles of *Insv* and ELBA factors, and Paul Schedl and

Elissa Lei for antibodies. D.J.P. was supported by funds from the Abby Rockefeller Mauze Trust and the Maloris Foundation. Q.D. was supported by a fellowship from the Swedish Research Council, and J.O.W. was supported by a fellowship from the Swedish Society for Medical Research. Work in E.C.L.’s group was supported by the Burroughs Wellcome Fund and the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award number R01-NS074037.

#### References

- Abhiman S, Iyer LM, Aravind L. 2008. BEN: a novel domain in chromatin factors and DNA viral proteins. *Bioinformatics* **24**: 458–461.
- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**: 1948–1954.
- Aoki T, Sarkeshik A, Yates J, Schedl P. 2012. Elba, a novel developmentally regulated chromatin boundary factor is a hetero-tripartite DNA binding complex. *eLife* **1**: e00171.
- Bailey TL, 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.
- Barolo S, Stone T, Bang AG, Posakony JW. 2002. Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless. *Genes Dev* **16**: 1964–1976.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837.
- Bushey AM, Ramos E, Corces VG. 2009. Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions. *Genes Dev* **23**: 1338–1350.
- Dai Q, Bergman M, Aihara H, Nibu Y, Mannervik M. 2008. *Drosophila* Ebi mediates Snail-dependent transcriptional repression through HDAC3-induced histone deacetylation. *EMBO J* **27**: 898–909.
- Dai Q, Andreu-Agullo C, Insolera R, Wong LC, Shi SH, Lai EC. 2013a. BEND6 is a nuclear antagonist of Notch signaling during self-renewal of neural stem cells. *Development* **140**: 1892–1902.
- Dai Q, Ren A, Westholm JO, Serganov A, Patel DJ, Lai EC. 2013b. The BEN domain is a novel sequence-specific DNA binding domain conserved in neural transcriptional repressors. *Genes Dev* **27**: 602–614.
- Duan H, Dai Q, Kavalier J, Bejarano F, Medranda G, Negre N, Lai EC. 2011. Insensitive is a novel corepressor of Suppressor of Hairless and regulates Notch-mediated cell fate decisions in the peripheral nervous system. *EMBO J* **30**: 3120–3133.
- Gerasimova TI, Gdula DA, Gerasimov DV, Simonova O, Corces VG. 1995. A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* **82**: 587–597.
- Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**: 473–479.
- Huen DS, Russell S. 2010. On the use of resampling tests for evaluating statistical significance of binding-site co-occurrence. *BMC Bioinformatics* **11**: 359.
- Kaul-Ghanekar R, Jalota A, Pavithra L, Tucker P, Chattopadhyay S. 2004. SMAR1 and Cux/CDP modulate chromatin and act as negative regulators of the TCR $\beta$  enhancer (E $\beta$ ). *Nucleic Acids Res* **32**: 4862–4875.

- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenko VV, Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**: 1231–1245.
- Korutla L, Wang PJ, Mackler SA. 2005. The POZ/BTB protein NAC1 interacts with two different histone deacetylases in neuronal-like cultures. *J Neurochem* **94**: 786–793.
- Korutla L, Degnan R, Wang P, Mackler SA. 2007. NAC1, a cocaine-regulated POZ/BTB protein interacts with CoREST. *J Neurochem* **101**: 611–618.
- Lai EC. 2004. Notch signaling: control of cell communication and cell fate. *Development* **131**: 965–973.
- Li M, Belozero V, Cai HN. 2008. Analysis of chromatin boundary activity in *Drosophila* cells. *BMC Mol Biol* **9**: 109.
- Maeda RK, Karch F. 2003. Ensuring enhancer fidelity. *Nat Genet* **34**: 360–361.
- Morel V, Lecourtois M, Massiani O, Maier D, Preiss A, Schweisguth F. 2001. Transcriptional repression by Suppressor of Hairless involves the binding of a Hairless-dCtBP complex in *Drosophila*. *Curr Biol* **11**: 789–792.
- Negre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG, Feng X, Ahmad K, Russell S, White RA, et al. 2010. A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet* **6**: e1000814.
- Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, et al. 2011. A cis-regulatory map of the *Drosophila* genome. *Nature* **471**: 527–531.
- Paroush Z, Finley RL Jr, Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D. 1994. Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**: 805–815.
- Rampalli S, Pavithra L, Bhatt A, Kundu TK, Chattopadhyay S. 2005. Tumor suppressor SMAR1 mediates cyclin D1 repression by recruitment of the SIN3/histone deacetylase 1 complex. *Mol Cell Biol* **25**: 8415–8429.
- Reeves N, Posakony JW. 2005. Genetic programs activated by proneural proteins in the developing *Drosophila* PNS. *Dev Cell* **8**: 413–425.
- Ryu JR, Arnosti DN. 2003. Functional similarity of Knirps CtBP-dependent and CtBP-independent transcriptional repressor activities. *Nucleic Acids Res* **31**: 4654–4662.
- Sathyan KM, Shen Z, Tripathi V, Prasanth KV, Prasanth SG. 2011. A BEN-domain-containing protein associates with heterochromatin and represses transcription. *J Cell Sci* **124**: 3149–3163.
- Singer JB, Lengyel JA. 1997. Expression and sequence analysis of the *Drosophila* blastoderm-specific gene bsg25A. *Gene* **197**: 379–382.
- Slattery M, Ma L, Spokony RF, Arthur RK, Kheradpour P, Kundaje A, Negre N, Crofts A, Ptashkin R, Zieba J, et al. 2014. Diverse patterns of genomic targeting by transcriptional regulators in *Drosophila melanogaster*. *Genome Res* **24**: 1224–1235.
- Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A. 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-seq data. *Nat Methods* **5**: 829–834.
- Vogelmann J, Valeri A, Guillou E, Cuvier O, Nollmann M. 2011. Roles of chromatin insulator proteins in higher-order chromatin organization and transcription regulation. *Nucleus* **2**: 358–369.
- Xuan C, Wang Q, Han X, Duan Y, Li L, Shi L, Wang Y, Shan L, Yao Z, Shang Y. 2013. RBB, a novel transcription repressor, represses the transcription of HDM2 oncogene. *Oncogene* **32**: 3711–3721.
- Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**: 65–70.