Case study: Finding a new DNA binding domain

Stockholm, November 8 2018

Jakub Orzechowski Westholm

Long-term bioinformatics support

NBIS, SciLifeLab, Stockholm University

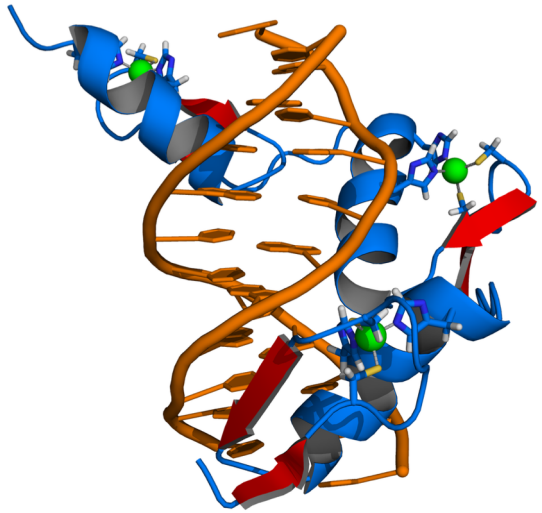
SciLifeLab

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

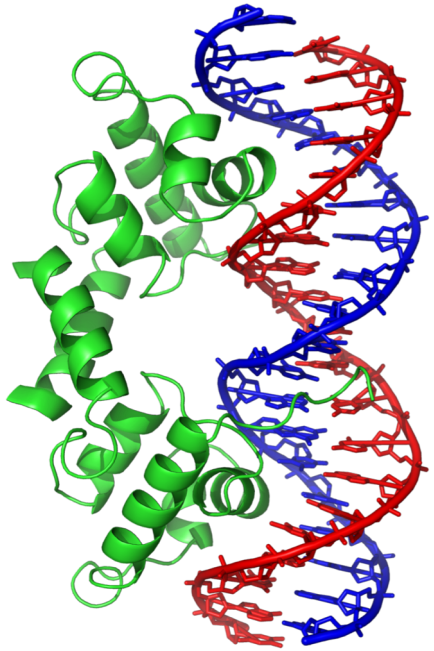
Transcription factors

- Transcription factors typically consist of
  - Activation/repression domains
  - A sequence specific DNA binding domain

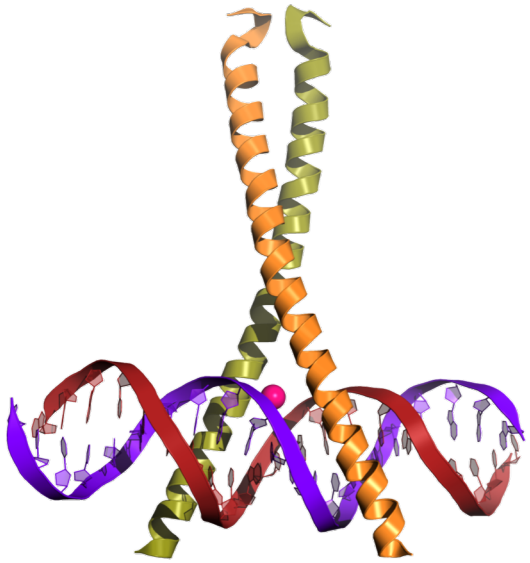
- The number of such DNA binding domains in eukaryotes is limited:
  - Less than 40 (**Yusuf** et al. *The Transcription Factor Encyclopedia*. Genome Biology 2012)



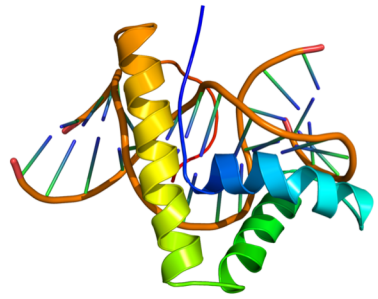
zinc finger



helix-turn-helix



basic leucine zipper



high mobility group box

BEN domains

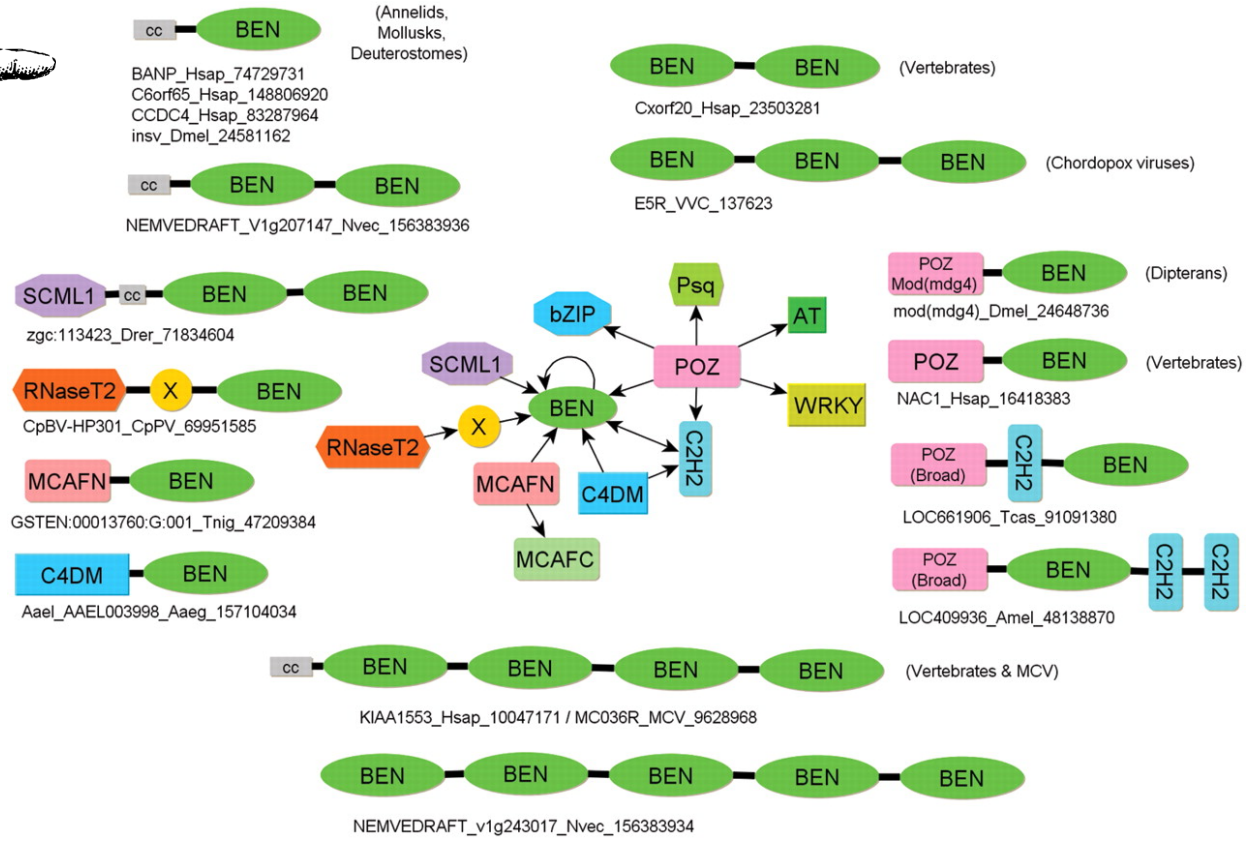
- Over 100 proteins across animals/metazoans and viruses have BEN domains.

"Prediction of the secondary structure using the multiple alignment indicated an all α-fold, with four conserved helices."

Abhiman et al. *BEN: A novel domain in chromatin factors and DNA viral proteins.* 2008, Bioinformatics

BEN domains, cont.

- The BEN domain sometimes co-occurs with chromatin remodeling domains (e.g for histone deacetylation).

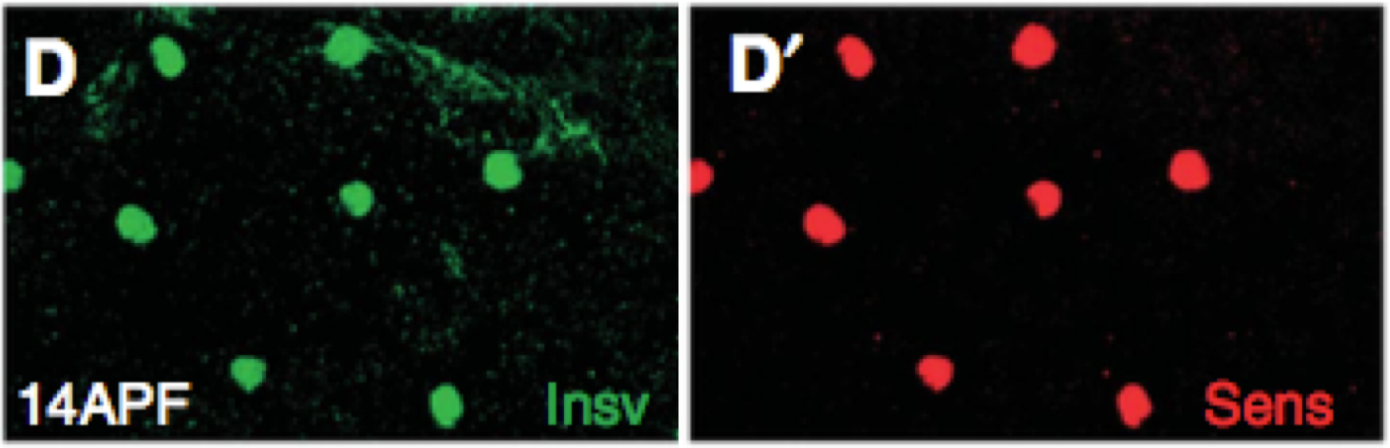


Insensitive protein

- We studied *Insensitive*, a *Drosophila* protein with a single BEN domain.
- *Insensitive* shows nuclear expression in the peripheral nervous system, and is involved in Notch signalling.
- *Insensitive* is expressed ubiquitously in the early embryo and later throughout the developing ectoderm but becomes highly restricted to the developing CNS and PNS. Peak expression at 2-4 hours.

Insensitive protein, cont.

- Previous studies suggested that *Insensitive* was a co-factor of a TF called *Suppressor of hairless*.
- We wanted to see where *Insensitive* bound to DNA, and determine possible targets.
- ChIP-seq from fly embryos, from two time points.
- IgG as control.



Duan et al. *Insensitive is a corepressor for Suppressor of Hairless and regulates Notch signalling during neural development.* 2011, EMBO J

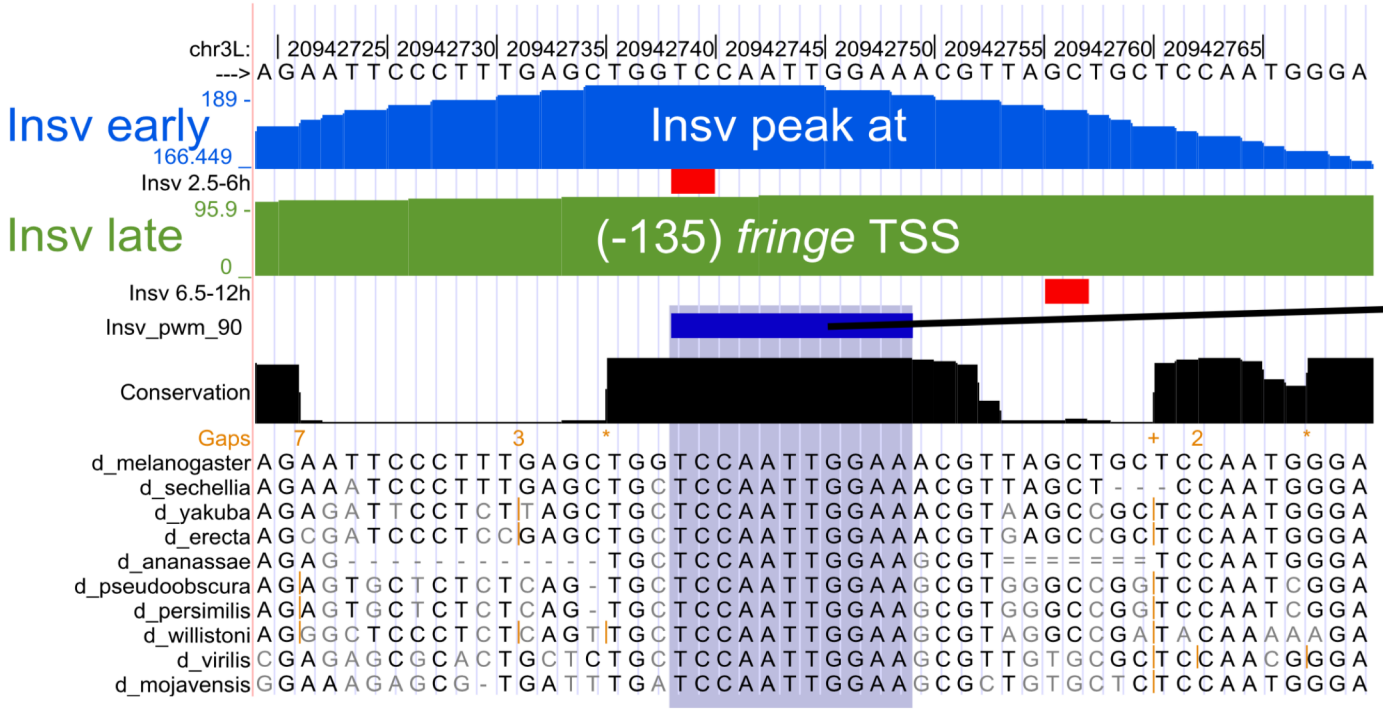
ChIP-seq experiment

- Analysis:
  - FastQC
  - Mapping: Bowtie
  - QC: Phantompeakqualtools
  - Peak calling: Quest (Valouev et al. *Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data*. Nature methods, 2008)
  - Peak annotation: chippeakanno
  - Motif finding: MEME, Weeder
  - Custom scripts..

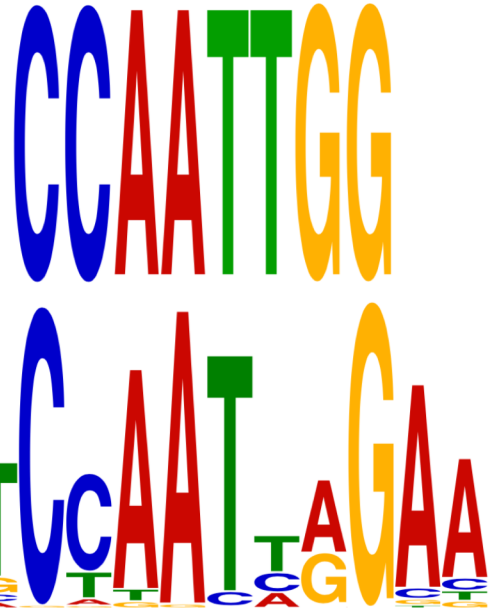
AB	Time	Unique reads mapping	Nr peaks
Insv	2.5-6h	7,473,521 (58%)	5364
Insv	6.5-12h	4,292,248 (61%)	2390

Insenstive seems to bind to a new motif

We were expecting to find the *Suppressor of Hairless* motif, but instead found a new site.



Weeder:

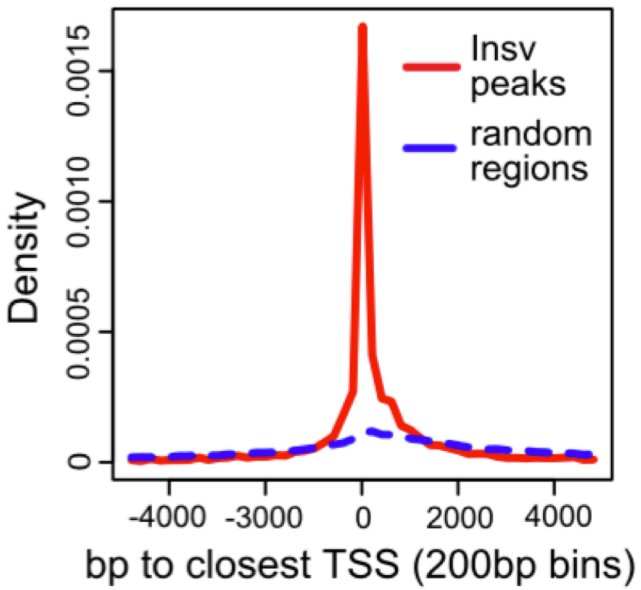


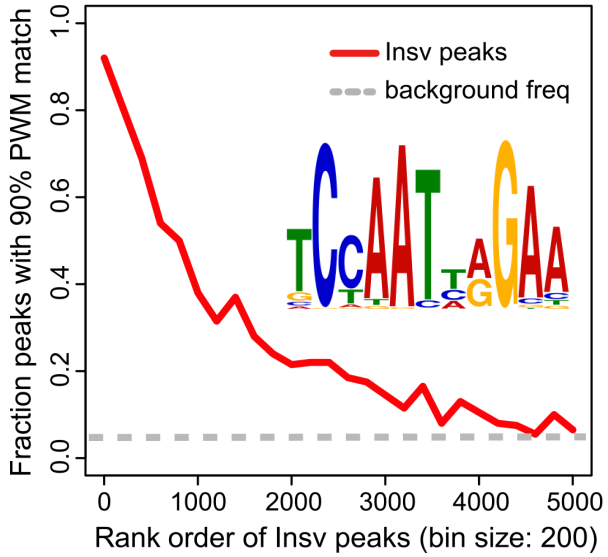
MEME:

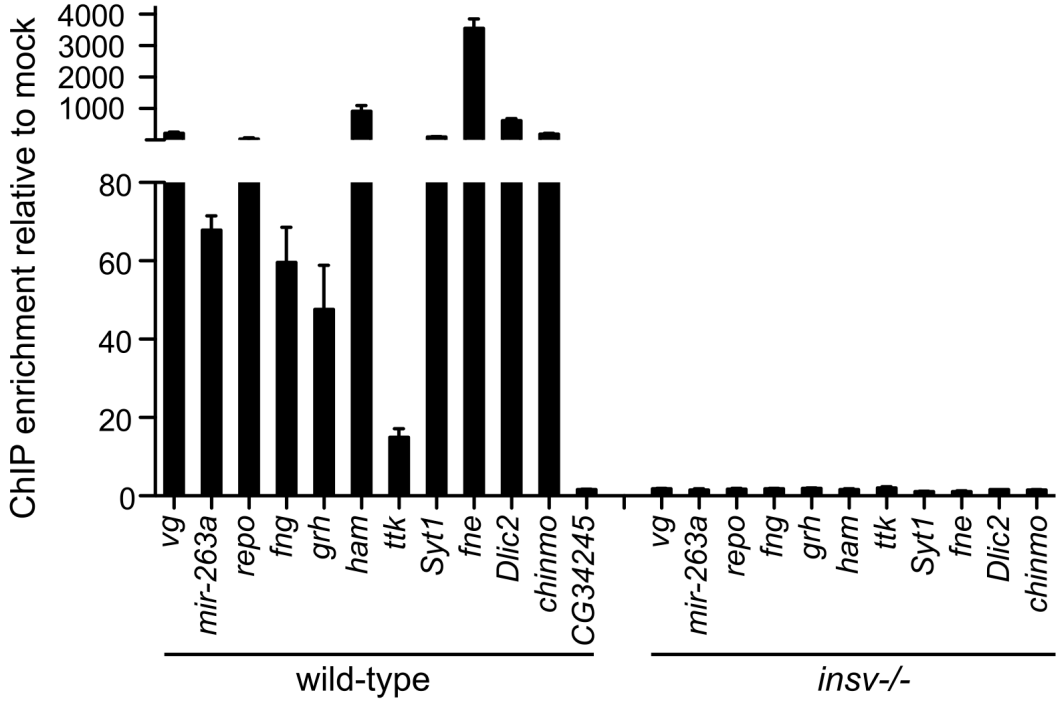
p-value = 6.4E-562

Dai et al. *The BEN domain is a novel sequence-specific DNA-binding domain conserved in neural transcriptional repressors*. Genes & Development, 2013.

Validating peaks



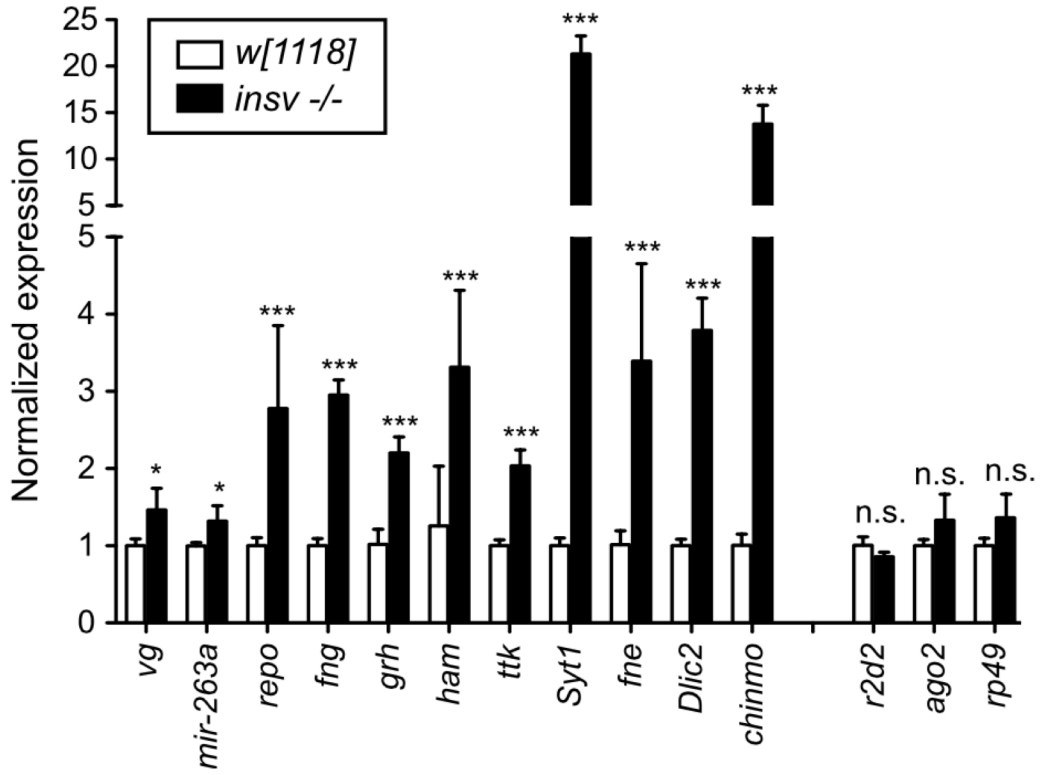




- *Insenstive* peaks are located at promotor regions
- Almost all the top *Insenstive* sites have the motif.
- ChIP-PCR validation of some peaks.

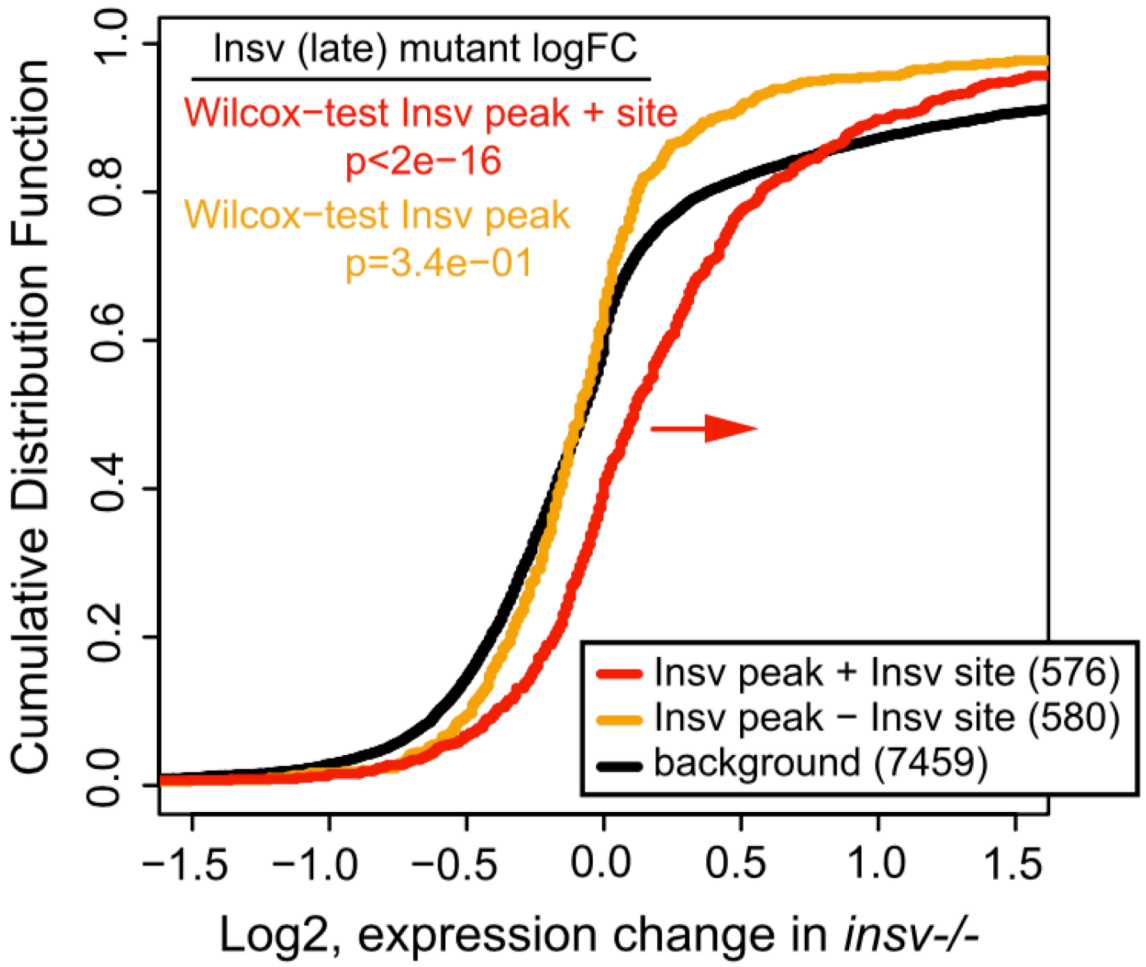
Gene expression

- rt-qPCR on selected genes → genes near Insensitive peaks have increased expression in an Insensitive mutant.



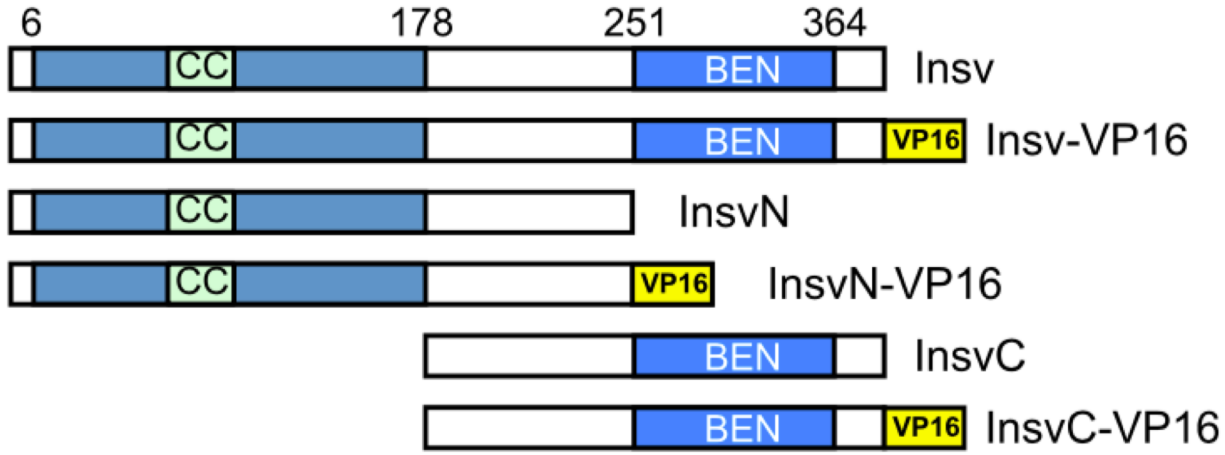
Gene expression, cont.

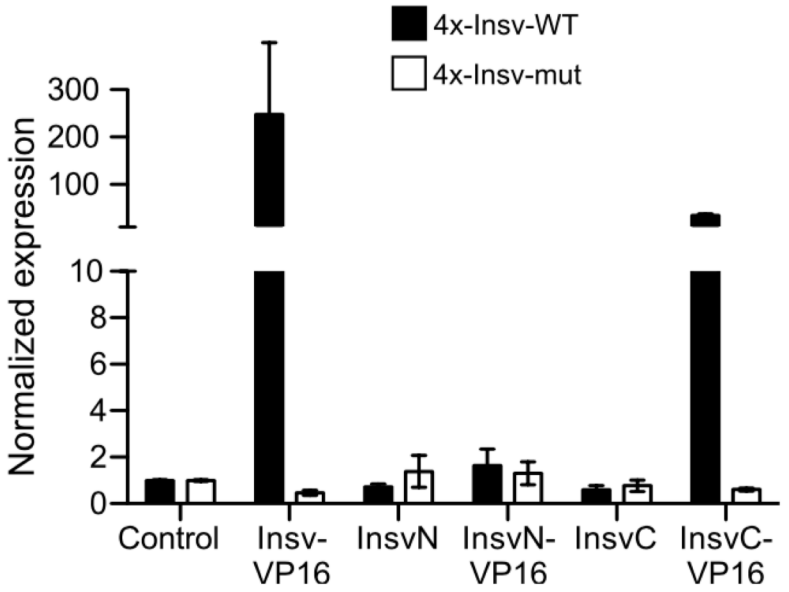
- We also looked a gene expression on a genome-wide scale.
- Genes near Insensitive peaks, that have an Insensitive site, have overall increased expression in an Insensitive mutant.



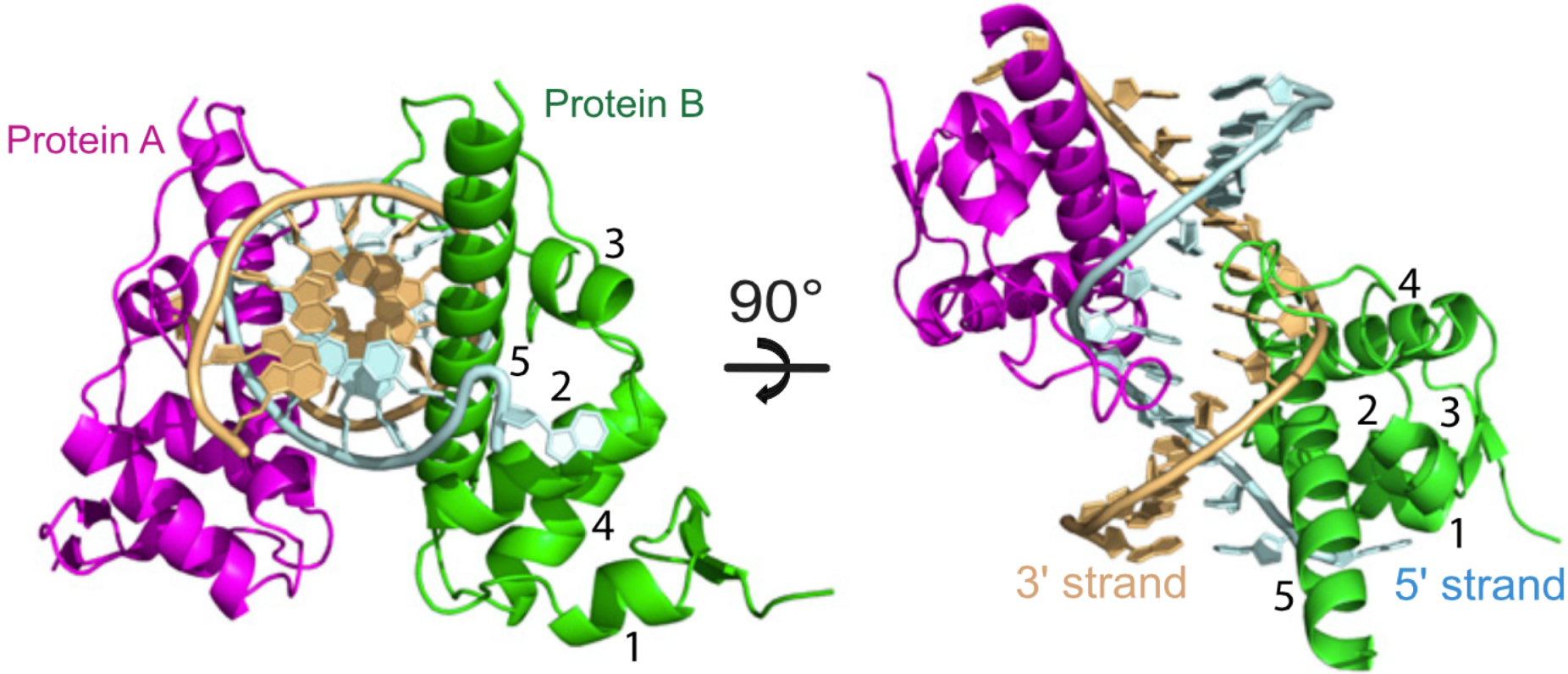
Structure-function experiments

- Actin-luciferase as read-out.
- 4 Insensitive sites in promoter or 4 mutated Insensitive sites
- Different parts of Insensitive, sometimes fused to the V16 activation domain.
- → the (C-terminal) BEN domain is necessary and sufficient for binding to the Insensitive site.



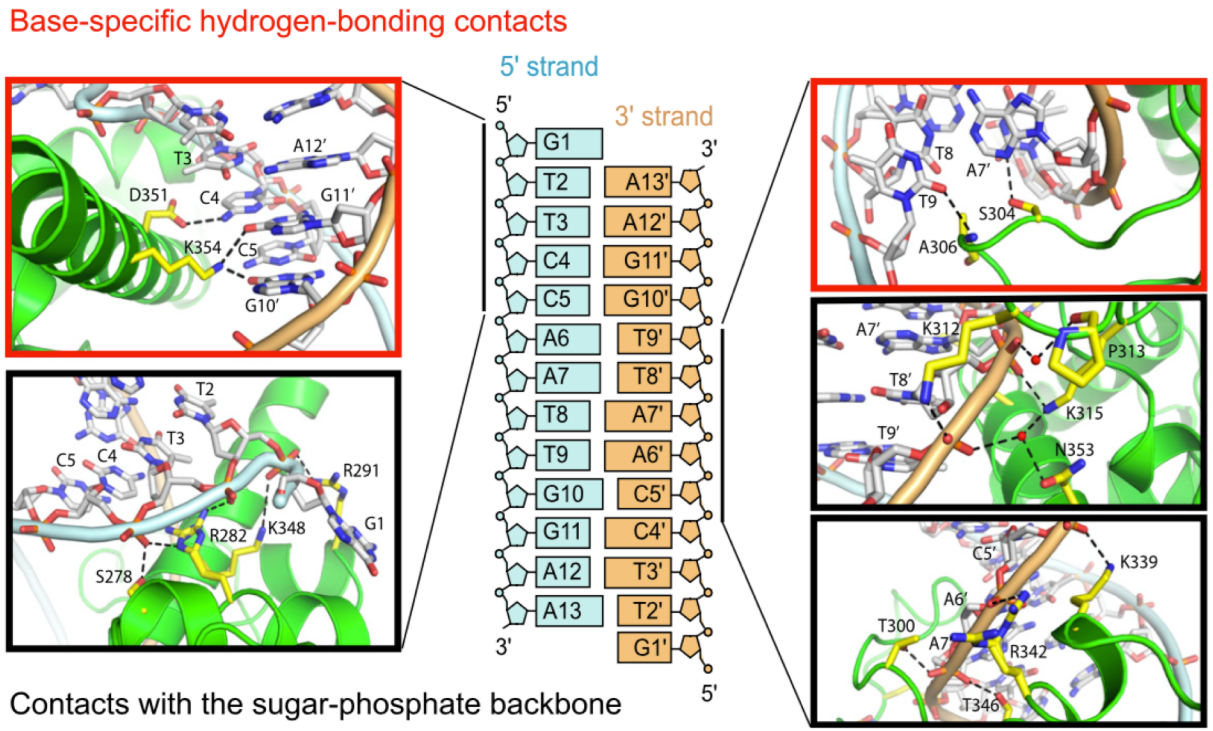


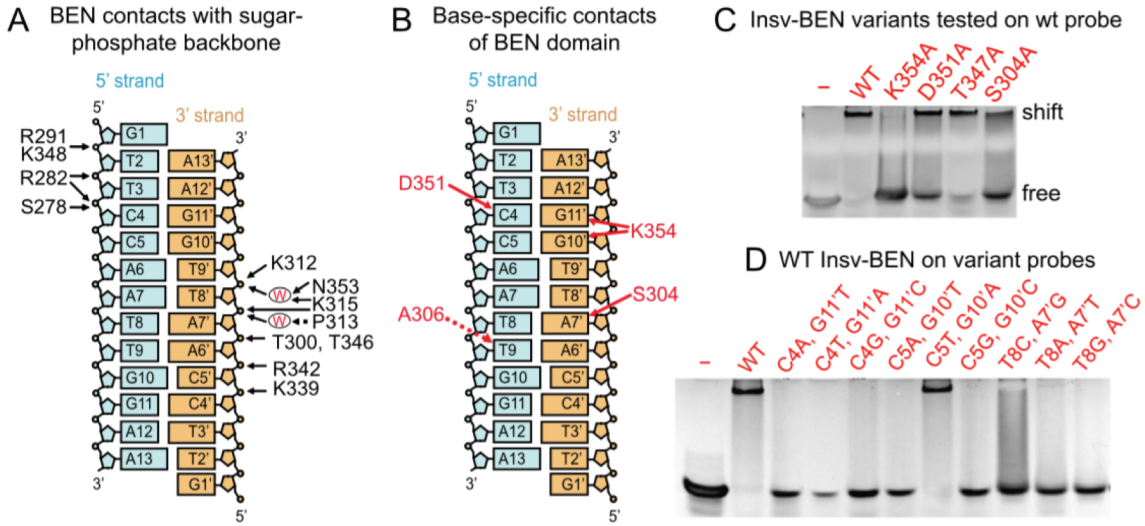
Crystal structure of BEN domain bound to DNA



Validating the structure

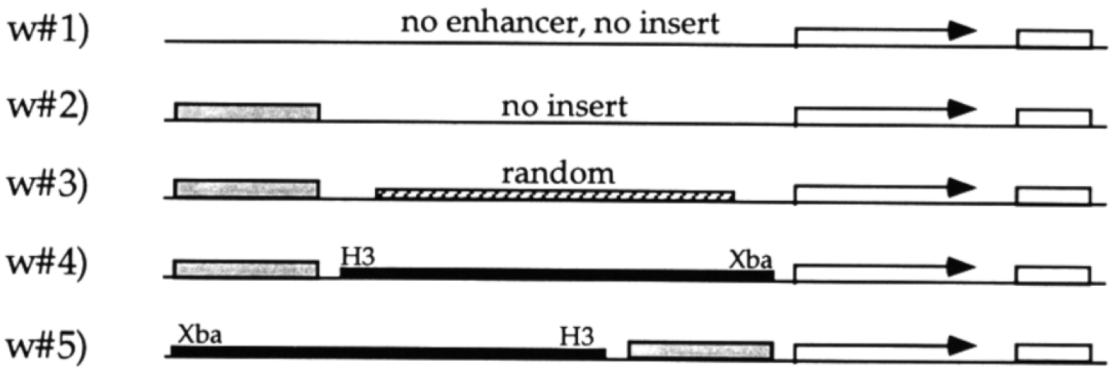
- From the structure, we can see with amino acids make contact with which nucleotides.
- We can make predictions about how amino acid and DNA mutations will affect binding, and test these predictions.



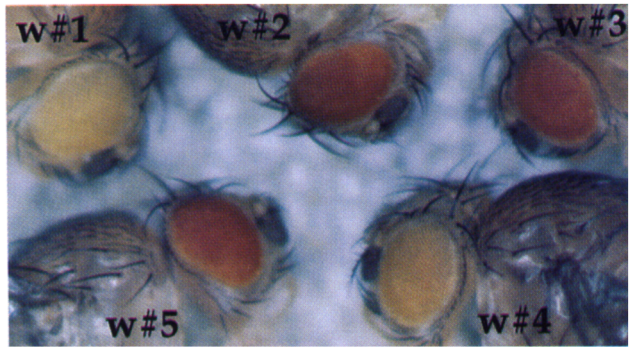


Insulator elements

- Insulator elements were first described as DNA elements that can restrict e.g. interactions between enhancers and target genes or the spread of heterochromatin.



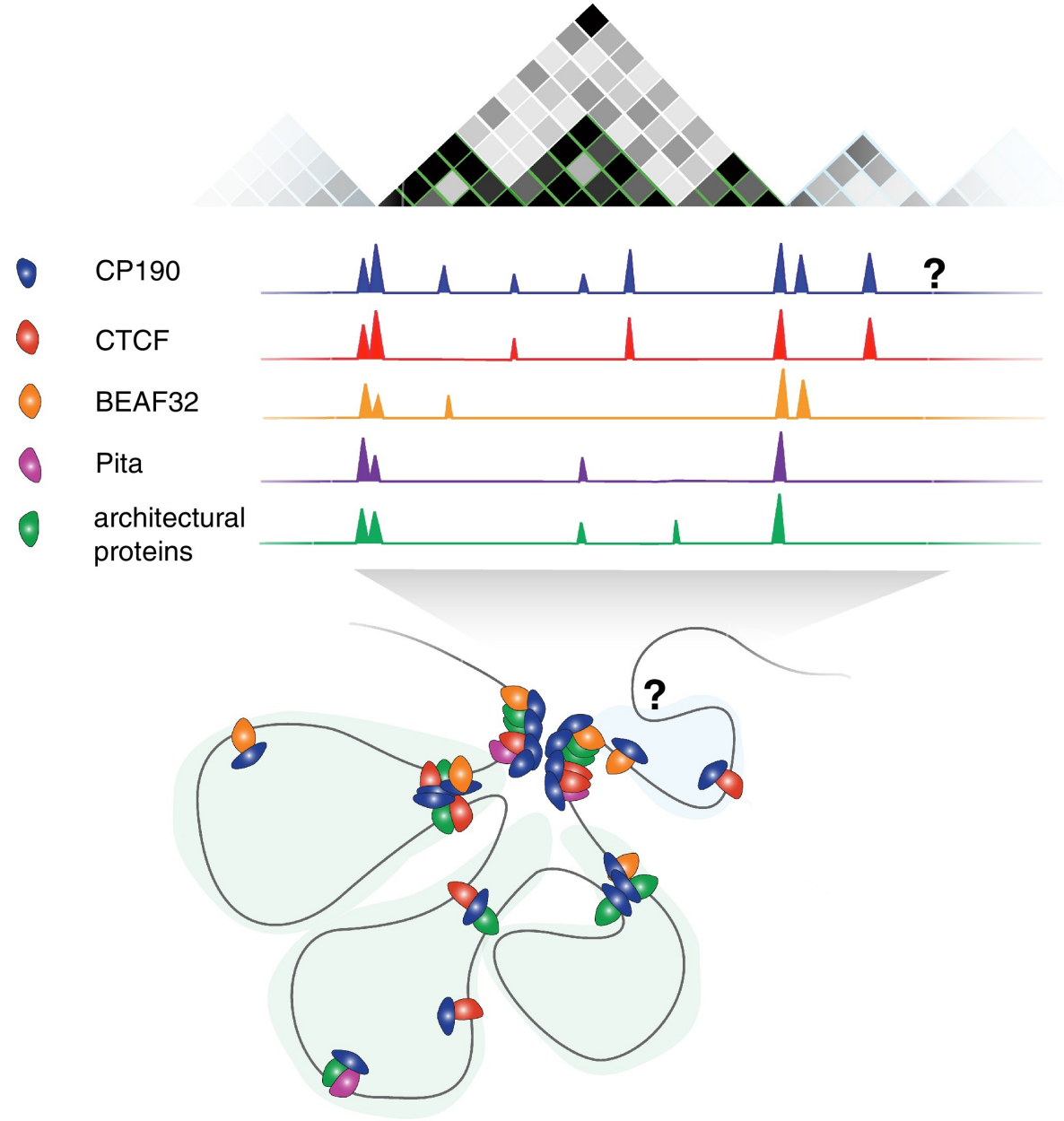
Enhancer blocking (#light eyes/total)	
—	(5/5)
—	(0/6)
—	(0/5)
+	(12/24)
—	(1/11)



Hagstrom et al. *Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the Drosophila bithorax complex*. Genes & Development 1996.

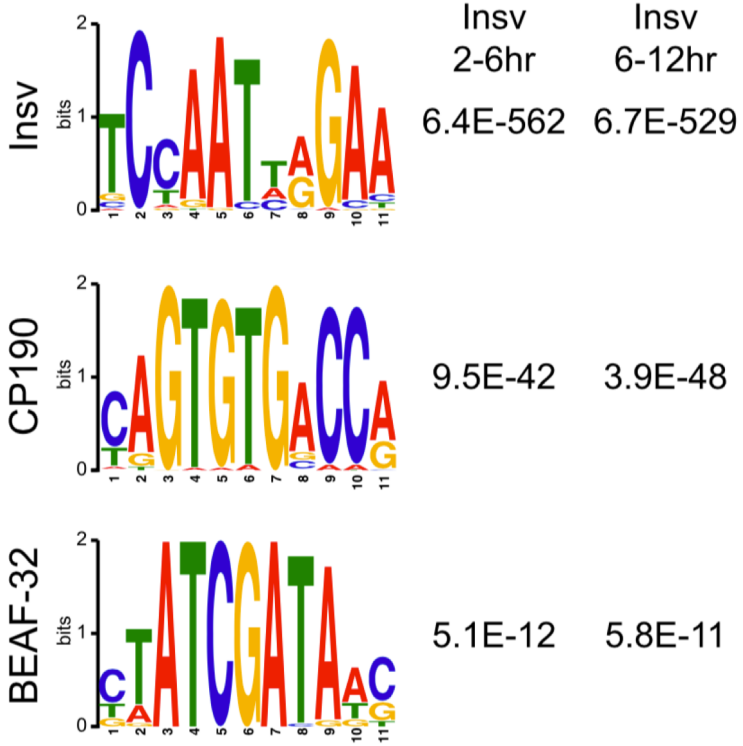
Insulator elements, cont.

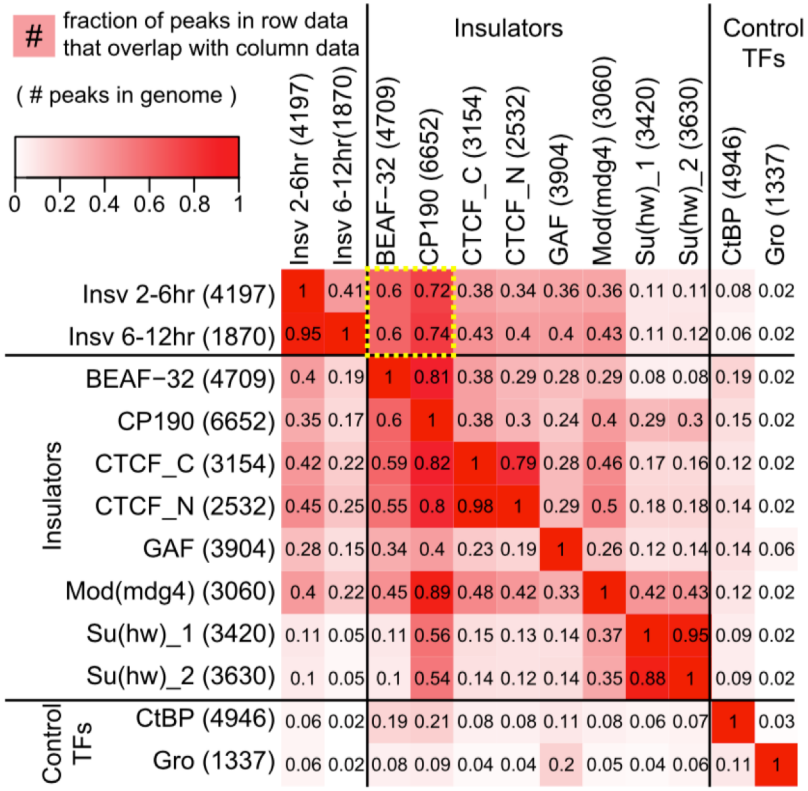
- Insulator elements control DNA looping.
- Enhancers and target genes can end up in different loop domains (≈ topologically associated domains, TADs)



Ali et al. *Insulators and domains of gene expression*. Current Opinion in Genetics & Development, 2016.

Insensitive binds at insulator elements

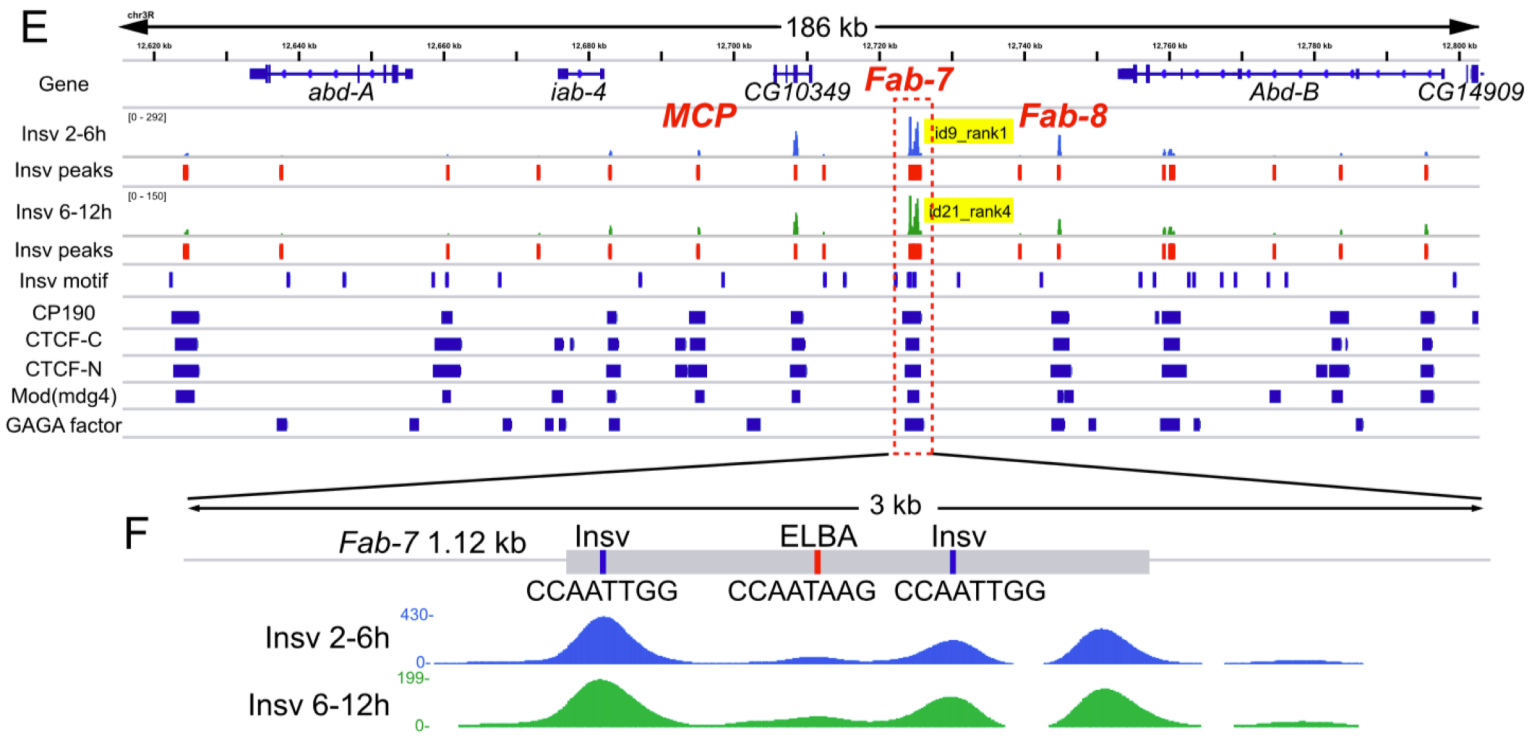


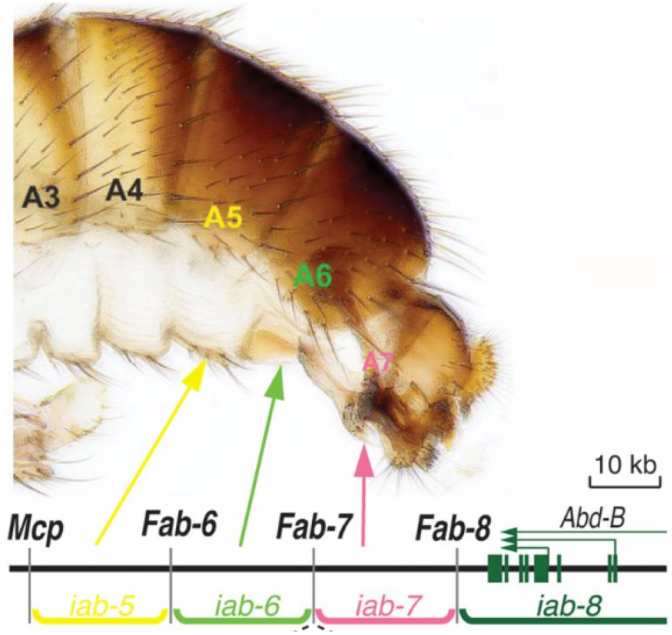


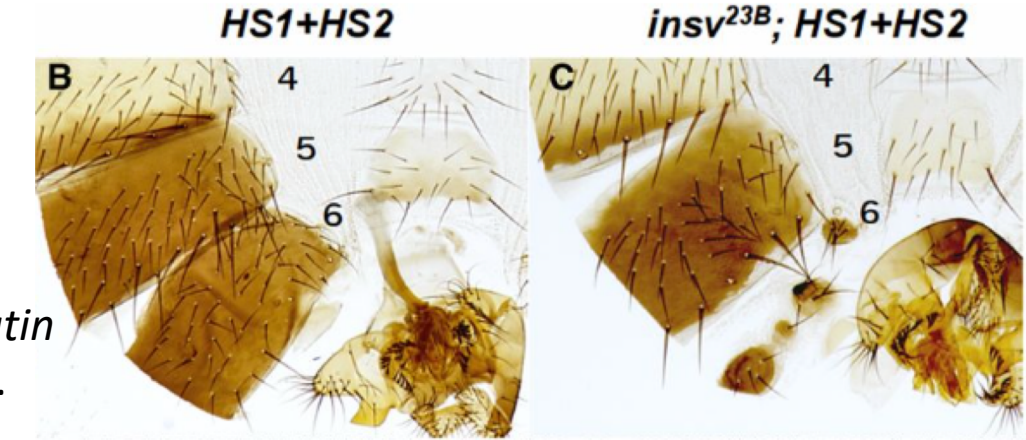
- *Insenstive* peaks are enriched for C190 and BEAF-32 motifs
- *Insenstive* peaks overlap C190, BEAF-32 and CTCF peaks

Dai et al. *Common and distinct DNA-binding and regulatory activities of the BEN-solo transcription factor family*. Genes & Development, 2015.

Insensitive binding at the Fab-7 insulator







Fedotova et al. *The BEN Domain Protein Insensitive Binds to the Fab-7 Chromatin Boundary To Establish Proper Segmental Identity in Drosophila*. Genetics 2018.

BEN domain protein function

- Insulators:
  - Elba1, Elba2, Elba3 (Aoki et al. *Elba, a novel developmentally regulated chromatin boundary factor is a hetero-tripartite DNA binding complex*. eLife, 2012)
- TFs:
  - BEND5 (Dai et al. *The BEN domain is a novel sequence-specific DNA-binding domain conserved in neural transcriptional repressors. Genes Dev*. 2013)
  - BEND6 (Dai. et al. *BEND6 is a nuclear antagonist of Notch signaling during self-renewal of neural stem cells*. Development, 2013)
- Chromatin remodelers:
  - BEND3 involved in heterochromatin formation (Saksouk et al. *Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation*. Mol Cell, 2014)
- Chromatin component?
  - Elba2 (Xu et al. *BEN domain protein Elba2 can functionally substitute for linker histone H1 in Drosophila in vivo*. Scientific Reports, 2016)

Some conclusions

- The BEN domain is a new DNA binding domain.
  - Gene annotation: clues about the function of over 100 genes with the BEN domain:
    - Transcription factors
    - Chromatin remodellers
    - insulator proteins etc.
- Insensitive is a transcriptional repressor
- Insensitive (and other BEN-proteins) have insulator activity.
- ChIP-seq was one (but important) method in this story

Acknowledgements

Eric Lai (Sloan-Kettering)

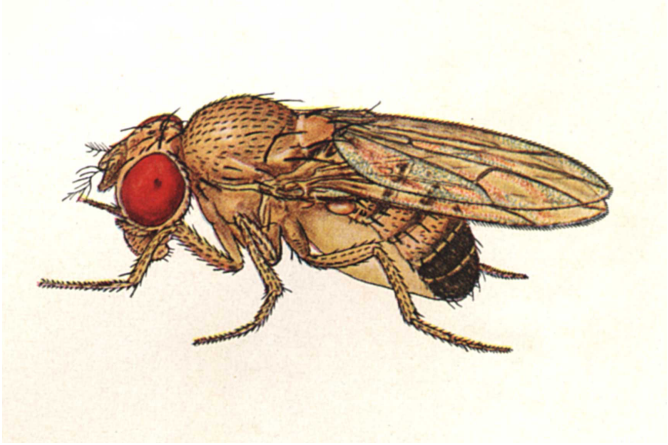
Qi Dai

Hong Duan

Dinshaw Palel (Sloan-Kettering)

Aiming Ren

Artem Serganov




The End