

Case study: Finding a new DNA binding domain

Stockholm, November 8 2018

Jakub Orzechowski Westholm

Long-term bioinformatics support

NBIS, SciLifeLab, Stockholm University

SciLifeLab

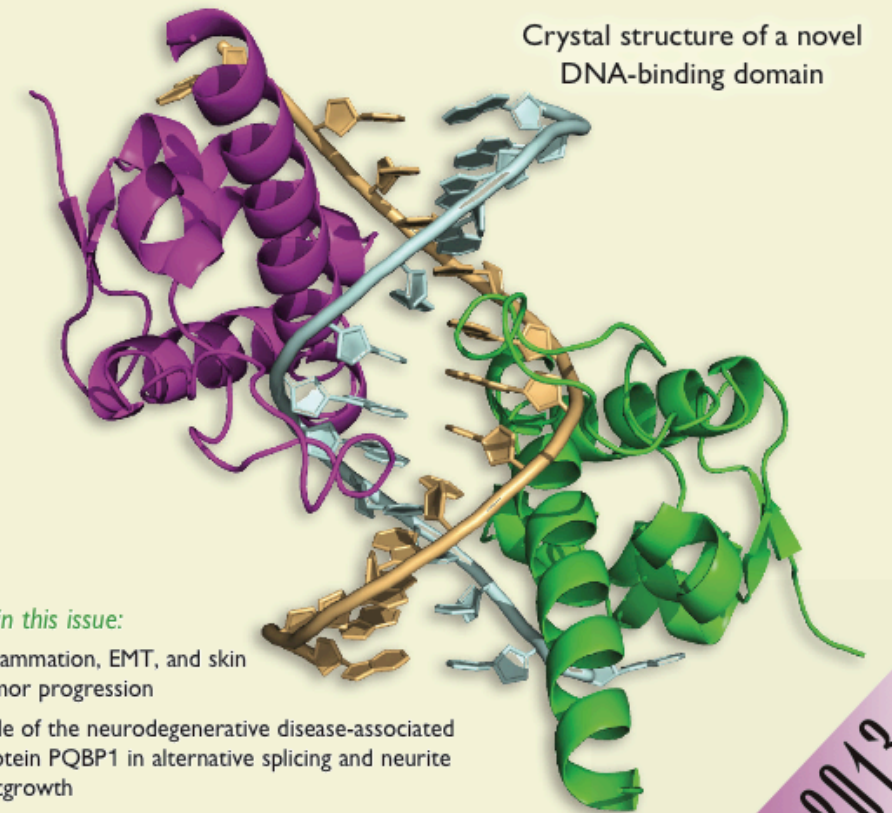
NBIS
NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

G e n e s
& **D e v e l o p m e n t**

Volume 27 No. 6

March 15, 2013

A JOURNAL OF CELLULAR AND MOLECULAR BIOLOGY



Also in this issue:

- Inflammation, EMT, and skin tumor progression
- Role of the neurodegenerative disease-associated protein PQBP1 in alternative splicing and neurite outgrowth

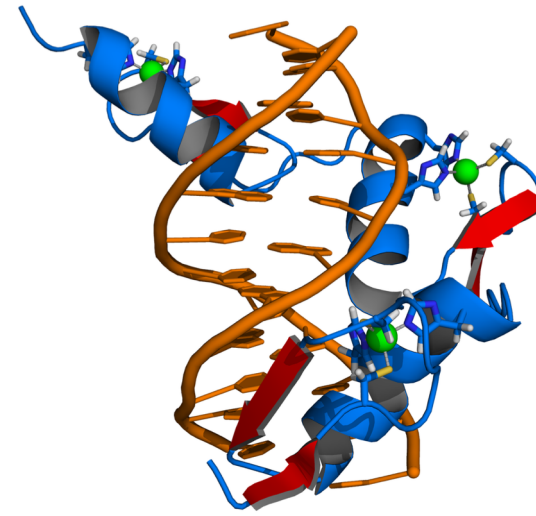


Cold Spring Harbor Laboratory Press

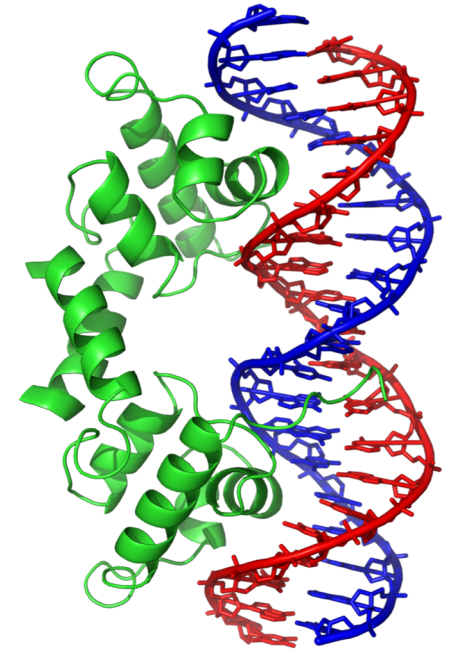
G&D 2013

Transcription factors

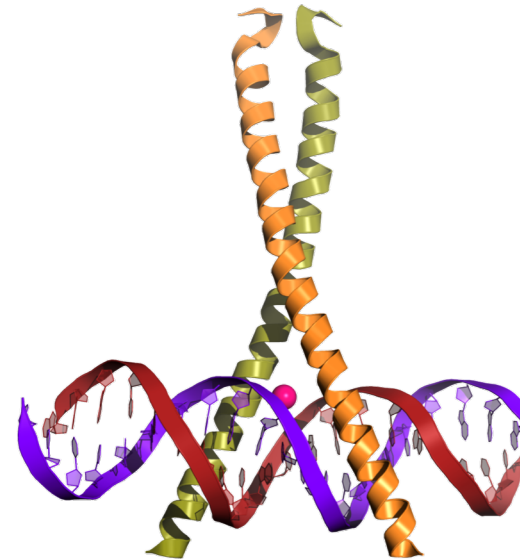
- Transcription factors typically consist of
 - Activation/repression domains
 - A sequence specific DNA binding domain
- The number of such DNA binding domains in eukaryotes is limited:
 - Less than 40 (Yusuf et al. *The Transcription Factor Encyclopedia*. Genome Biology 2012)



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.

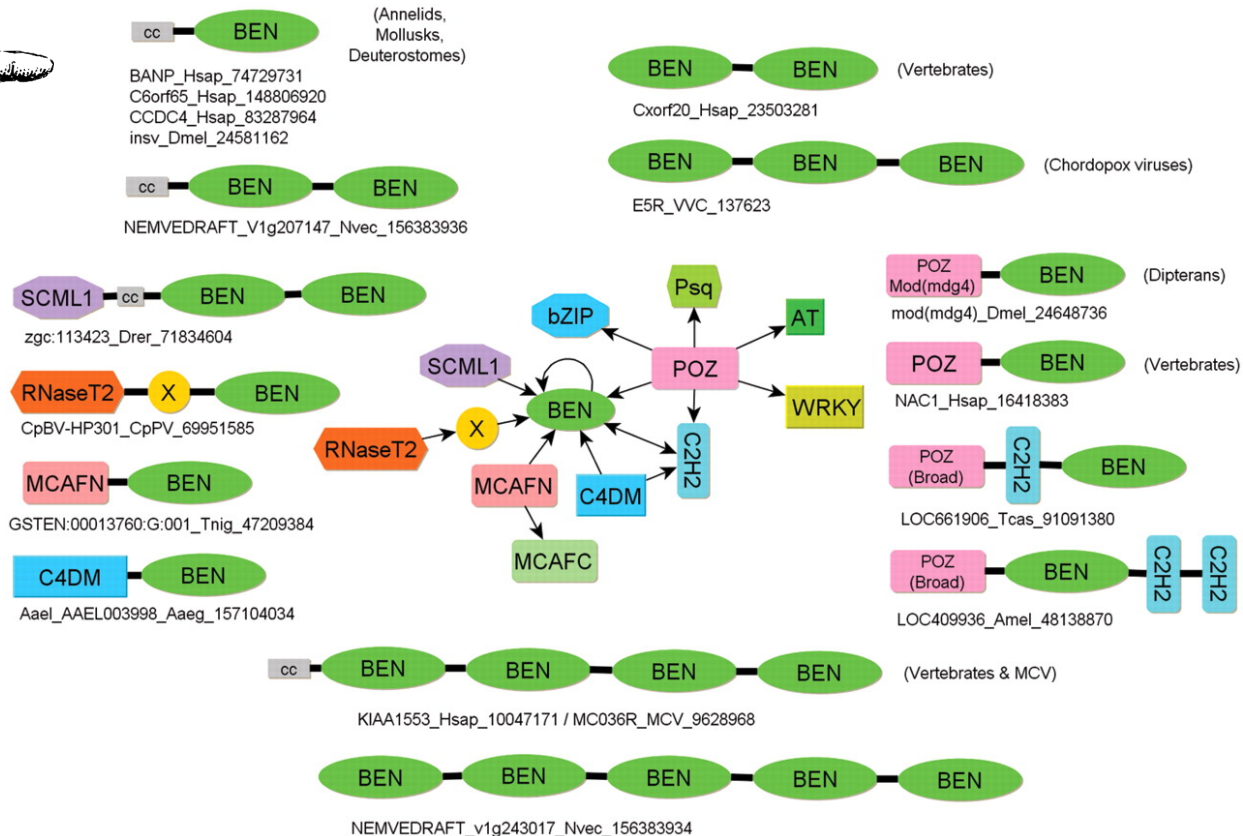
```
Secondary structure
insv_Dme1_24581162
Agap_ENSANGG00000025789_Agam_118791739
LOC724266_Amel_110759165
bsg25A_Dme1_1930012
Clorf165_Hsap_13375807
LOC566161_Drer_125823408
AaeL_AAEL003998_Aaeg_157104034
mod(mdg4)_Dme1_24648736
NAC1_Hsap_16418383
LOC495228_Xlae_148226339
CCDC4_Hsap_83287964
GSTEN:00029264;G:001_Tnig_47226171
LOC560711_Drer_125843107
C10orf30_Hsap_21618768
CebV_3.4_CebV_57753424
CebV6gpp3_CebV_57753417
GIP_L1_00580_Gind_117935419
MdbV_sBgp1_MdbV_66391199
C6orf65_Hsap_148806920
LOC794392_Drer_125831342
BANP_Hsap_74729731
SMAR1_Mmus_10312104
LOC575996_Spur_115728493
Capitella_spI
NEMVEDRAFT_vlg232490_Nvec_156390312
CXorf20_Hsap_23503281_2
LOC100003955_Drer_125851480
zgc:113423_Drer_71834604_1
LOC764357_Spur_115613065
NEMVEDRAFT_vlg243017_Nvec_156383934_4
KIAA1553_Hsap_10047171_1
KIAA1553_Hsap_10047171_2
KIAA1553_Hsap_10047171_3
KIAA1553_Hsap_10047171_4
GSTEN:00016974;G:001_Tnig_47220120_1
NEMVEDRAFT_vlg243810_Nvec_156379688_1
LOC584784_Spur_115651987
MC036R_MCV_9628968_1
MC036R_MCV_9628968_2
MC036R_MCV_9628968_3
MC036R_MCV_9628968_4
E5R_VVC_137623_2
E5R_VVC_137623_3
GSTEN:00013760;G:001_Tnig_47209384
xpat-A_Xlae_148222226
Daphnia_pulex
Branchiostoma_floridae
Consensus/80%
-----HHHHHHHHH-----HHHHHHHHHH-----HHHHHHHHHHHHHHHH-----HHHHHHHHHHHHHHHH-----
PNNTCVPASVFNINWS VC SLATRKQLVTTDRETLATH SMTGKPSP...QDKPLKMDPGKIQDIIFAVTHKCNASE ---- EVRNAITTKCADENKMM 259-356\1
SNNTLVPKRALEAVRWH SY KFGTRKQLQMLTRETLLASC SLSGRPCP...DRPVKGADPPKVVADIVEYVNGKCNVEE ---- HVRGVITNKCADENKML 619-717|
GEGIAICEQLRAVAKWS DY RKLTRGQAAILSPTELATC SVTQGRWS...ERPVKPADDKAKVQAIISYVTSRFPTVD ---- SVKQVLAYKCKENSTAL 27-126|
PNGTEVSRISLSAINWD MT PSITRKLCEEDRDTLAHH TLSGKPSP...CARPSKQDPLKQVADLVYLMNLSLDMTP ---- EVRTAITTKCADENKML 102-200/
GSGIWDDEEKWHQLQVT GD SKYTKNNAVMWGTDLVKNR SVTGVA TK...DAVPKP PPSRKLKIVRECLYDRIAQET ---- VDETEIAQRLSKVNKYI 133-228\2
GGGIWVDEEKWHQLQRT GD SKFTKNNAVMWGTETLKNR SVTGVA TK...DALPKP PPSPKLKIVRECLYDRVQSET ---- ADSAETQRLSKVNKYI 273-368|
VRDSLIPYQTMVDIDSV...KY LRFVSKCALALWGHMERLVA SVTGKRSN...NSTPSI QDSEPKLSFIKEKVYHRAMQET ---- RVQAMARFDDSRINRLL 364-466/
GSRVFSKVALAKAYIP MP MIYTCRVMDLVIGKDKL-VR IAQHEETT...-----DKDLIQDIITHVCKVFPALRG ---- AVQEFIDHKLSTLKLMP 441-529\3
GTNVYITRAQLMNCVHS...RH KULLRRDLASPDRLNTLANS CGTGIRSS...NDPRRK PDSRVLHAVKYQCNPAPNFK ---- EMNAIADMCTNARRVV 374-471\4
SSGVYITYQLELDLSHI...KP KLMTRRRLDYFHSRETLARS SATGQRIA...TMEKPL RPPDKVVTAKIAYVTRACGRGC ---- NFNNAVINSKCGTSRRAV 348-446/
NYPVYITSKQWDEAVNS...KD RRLRLRYDLRFVTTDELKYS CGLGKRKR...ETGPER PDPVKVTCLERFIRMHCTSNP DWMW PSEBQINKVPSDAVGH 386-490\5
NYPLFITNKQWDEAVNS...KD RRLRLRYDLRFVTTDELKYS CGLGKRKR...DSGLER PDPVKVTCLERFIRMHCANSP DWMW PSEBQINKVPSDAVGH 255-359|
DYDFVFPKAGLDSILLN...RS SLLFRKVCAPD DDTLLANS LPNGKRKR...LNDTRK GDDQNI VGAIKVFTEKYCTANG...RDWV QILQDQIKLARRLKRG 267-373|
GDFVMPKSQLDSILSN...RS SLLFRKVCAPD DDTLLANS LPNGKRKR...LNDNRK GDDQNI VGAIKVFTEKYCTANH...RDWV QILQDQIKLARRLKRG 239-345/
RTGVYVKKRKLKRCIRE...ND RTLARRDLTEVSNALSVCTWTGQKAK...NIDIRP GDDENARWVLTTFVBEQH-GKK...CGWSANTSAMVSTIRTKINDI 1112-1213\6
QRGVVYSYGDLYKCCQV...KD KSLARRDLAVNRRKALSVC LSITERAQ...GSNARP EDDHACTVLLNFVLEH-GLQ...RGWN TDIQPLNLTLSKIQEI 1083-1183|
QSDIYVSYGELYKCCQV...KD KSLARRDLTEVSNALSVCTWTGQKAK...GSNLRP EDEHASKVLLNFVIDY-GLQ...CGWN TDLKPLDLSHSKIQDI 955-1055|
HTNVYINAIKLSNCKRL...KD KSLARRDLTEVSNALSVCTWTGQKAK...GATIRP GDDETARTVLLTYVEEY-GRE...KGWI LDTQSIQNSIRNKMQEF 142-243 /
EKQFQIEKWIARC... KS QKFINDMQLVYTYNEYMATH SLTGAKSS...DKAVKP AMNQVETIIGVTKQLFPNTD...SIRR MIQKLN-NCTKPNLS 171-270\7
-YTEFITP-ELLERCNT...GT QKLTNDLGRGVERECLASH SISGVVYN...RGQPKP ADPTEEVQAILRTVQYFPFGKT...EIKG YIRQKLNQAKRLRKKP 202-300/
VRCIIIPS-DMLHISTN...RT EKMALTLDDYFHREVOAVS NLSGQGGK...---GKK QDDPLTYIGIRCHLFPYKFGITE SDWY RIKQSIDSKRTAWRRK 255-348\8
VRCIIIPS-DMLHISTN...RT EKMALTLDDYFHREVOAVS NLSGQGGK...---GKK QDDPLTYIGIRCHLFPYKFGITE SDWY RIKQSIDSKRTAWRRK 237-330|
VRCKINPT-EMVHINM...KT DKLALKLDDLDPKEMQAVS NLSGTGKH...---KKK KDDPLTYIGIRCHLFPYKFGITE EDWY RIRQNSIDSKRTAFRRK 278-371|
VRVPITPS-DLHISHN...RT EKMALTLDDYFHREVOAVS NLSGQGGK...---GKK QDDPLTYIGIRCHLFPYKFGITE QDWH RIKQNSIDSKRTAFRRR 228-321/
PHISDAELQSLRDEK...KP ENLAVVGLRRLTTRQERGR TVCGF...---GGS GDDNDVVQDIRRYFYPALPDPF...DKWG QCISAMNSYLRGTRRRK 285-375
WRNIRMP- SVLT LAK...KS SLSARYDLQKTKDVLVQS NVYGNLKH...---GIC QDDPNKISALREFYQENYICD...RDWK SCVTSINSGRSLRHVDV 667-765\9
LRKVWIPQ-CVYKVF...ET QKAVAPVLYSDFPISLTLSCS AVTGMPEK...---GLQ QDDPNKISALREFYQENYICD...VAMA QCLGVIN-SITKNLKKT 383-480|
ERKVFIS-FILQRAG...MT SAAVRYVSRNDDTKELSQS STTGNSPR...---CLL RQDNTKVDAREWAVKRYPKPD...DKWK VCLAVINSTARYRPM 239-337/
RIQMVMQDSRWEEMTP...GA RLAIALALARYC DGTGKILIRS SVTGRN...---SKN PDDFAGLRKIKHLFPQYGSRC...VIWK TSRESISQLCKRLRRKY 966-1064\10
YQDVTLPLEDFRQITV...ET SNAVAVAVRDPDEVELERA AAGE...---GTR SDDTLTKAIKAIWLGFRFAEK...LIWD NCLAAITQIRIRNPLLGG 604-699 /
PPEYQLTAAELKQIVDQ...LS GDLACRGLVQLPELFDV-DFSRCGSA...GFAAKR KVESLHLQLIRNYEVVYPSVK...AVWQ ECLPQLNDFPFRFVAQR 85-185 \11
ASDHVDTQDLTEPLE...SS GDFAVFLHRDPELFDHR- KLGEQYSC...GDGGKQ DDDPQLQIRIRNYTEIYFPDMQ...EAWL QCAQRINDELEGLGLDA 229-329 |
GADCLLSKEQLRSIY...LS GNFPASRLVHDPPELFTHE-NLRQYINC...GSLGKK QDDPNKILIRHYVQLLYPRAK...RVWT EFPVKLDERCRRRDTQ 392-492 |
PSPYLLSDKEVREIY...LS GNFAARGLVRLPELFTA-NLRLQYINC...GACNKK QDDPTRLIRHYVAVVYVVEK...EVWH ECIPSIDERCRRPNNRK 558-658 |
PQEYLLSREQLRNIY...LS GNFPASRLVHDPPELFTQE-NARRRYNC...GSLGKK QDDPNKILIRHYVQLLYPRAK...RVWM EFPVKLDERCRRRETEQ 529-629 |
RPQFASRS-AVMQIK...KS GNFPASRLVHDPPELFTQE-NARRRYNC...GSLGKK QDDPNKILIRHYVQLLYPRAK...RVWT EFPVKLDERCRRRDTQ 169-264 |
AWEKLSMGVICHLY...KG -NFPASRLVHDPPELFTQE-NARRRYNC...GSLGKK QDDPNKILIRHYVQLLYPRAK...RVWT EFPVKLDERCRRRDTQ 323-421 |
ALEMIPSPAELCHLAH...CS ADMARRVLLRYPEV-----VCGADR...---AE -IPALYFVAVRACSEYYPVLC...YVWQ EGLPLREFVLRCLVR 18-102 |
PAWAGPVTLDIYECAS...SV GELAVLGLHKVQELFDA--QLRRCYSC...GDGRTH QDDPARLQLIRHCVLCPSPMS...GEWV ECVSRVNSELTGEELMD 634-734 |
SCVPLPTRAHLRKMV...AS -YNFVAVRMLVYMPPELFTA--NLHTFNC...GSMGKR RDPDLRLRLRHYVQLLHPAAR...RVWI KFLACLDERCRRRCART 807-907 |
PAQYLISAKRVKELAR...RS GHFAAQITVMLPELFS--TERQKFS...GSDHL RDPVVRVRLIRHYVAVVCLPGA...RTWE ECVPSIDARCOQPLRR 935-1035 |
NQKTYKLFSDISAIG...AS -SKMVAQLLYMPLNFGD--HRFIRYM...SKIKKH IPSPKLNLIRILVEERFYNN...NKWR IIGTQVDMKLAIESDKY 102-204 |
IKGKSEED-TLPIKQMV...VT QELVEKVKLIDRDLFKS--GEYKAYRY...VENGFI GDDTLK-LNIVHDIPEPCMPV...PVAK LCKEMNKVYENPLHI 218-318 /
LRKINSHMEKILFENCK...GV DRYASYVFRVLPYNYKCEW VTKVNY...GLMGKE ADPTNVRRLRYIERFPPLS...DHWK EIRDAINEILRVKRPKE 221-322 \12
LPDIILNPLDGKLLVSM...SM HRFALDPOHVVHSLFQLWANKVNF--...GSRGKL GDFRNLMDILHQTSKRFV-LG...KEKR KIKTRNLNLLRTRQDRA 187-287 /
HMSSDEL--DYNMMA...NN TYMISLMLGKVTVEELTTK SLTGK...-RTTK ADPVDKVNVAKYILKRHPDKH...---EFNQKVTNYLRDQAAS 266-354
EQGVVTPYVILQAQNK...KT EQFPKIDMGIYTEELING LHGG...---GTHQ ADSPAISAILTETKQYGGVQ...---KLYRVNNEKGRMRA 140-229
.....ph..h.....s..h...Lh..hFsp..b...p.....Ls...h..lb..h....s.....l...h.p.....
```

“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).

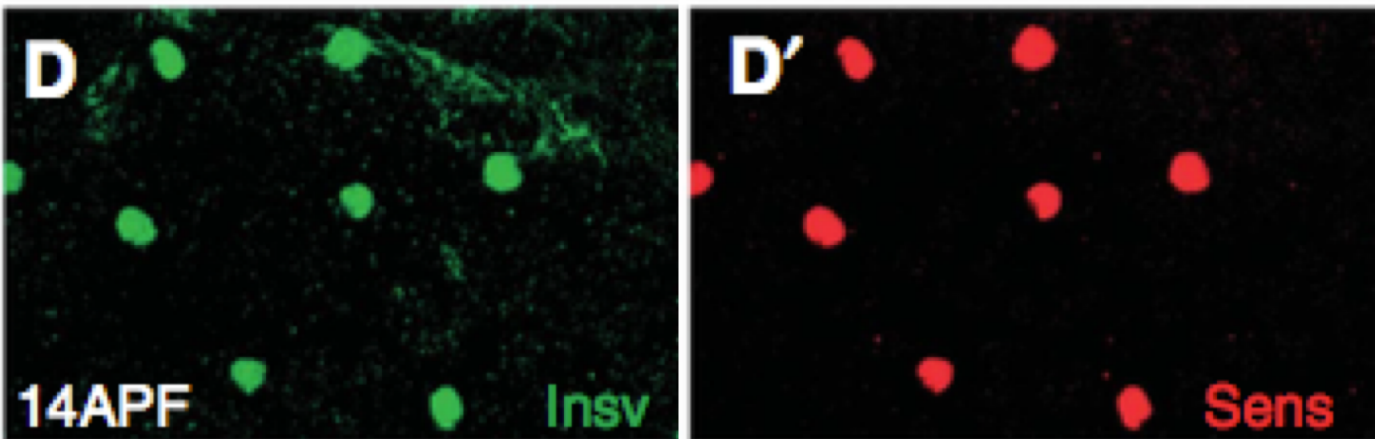


Insensitive protein

- 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.

Inensitive protein, cont.

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



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

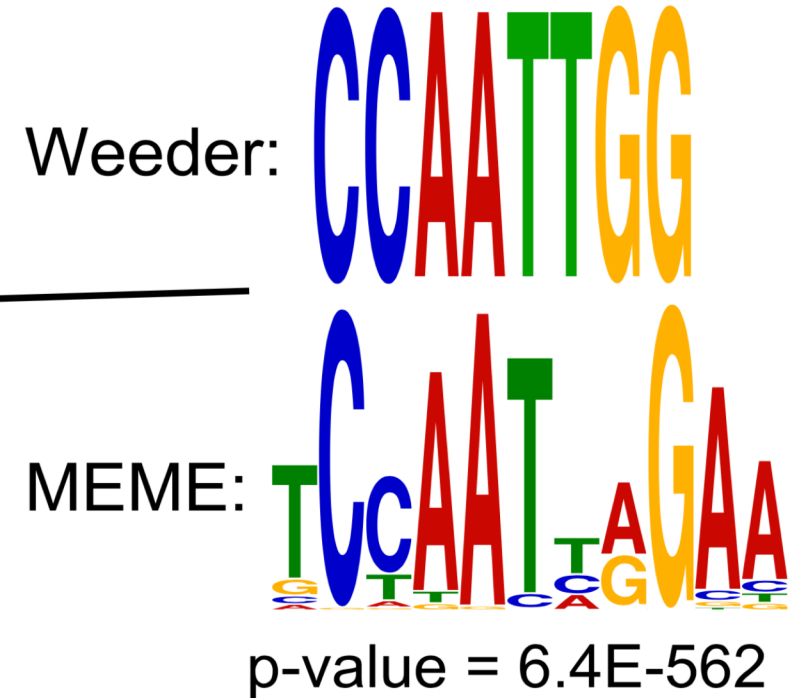
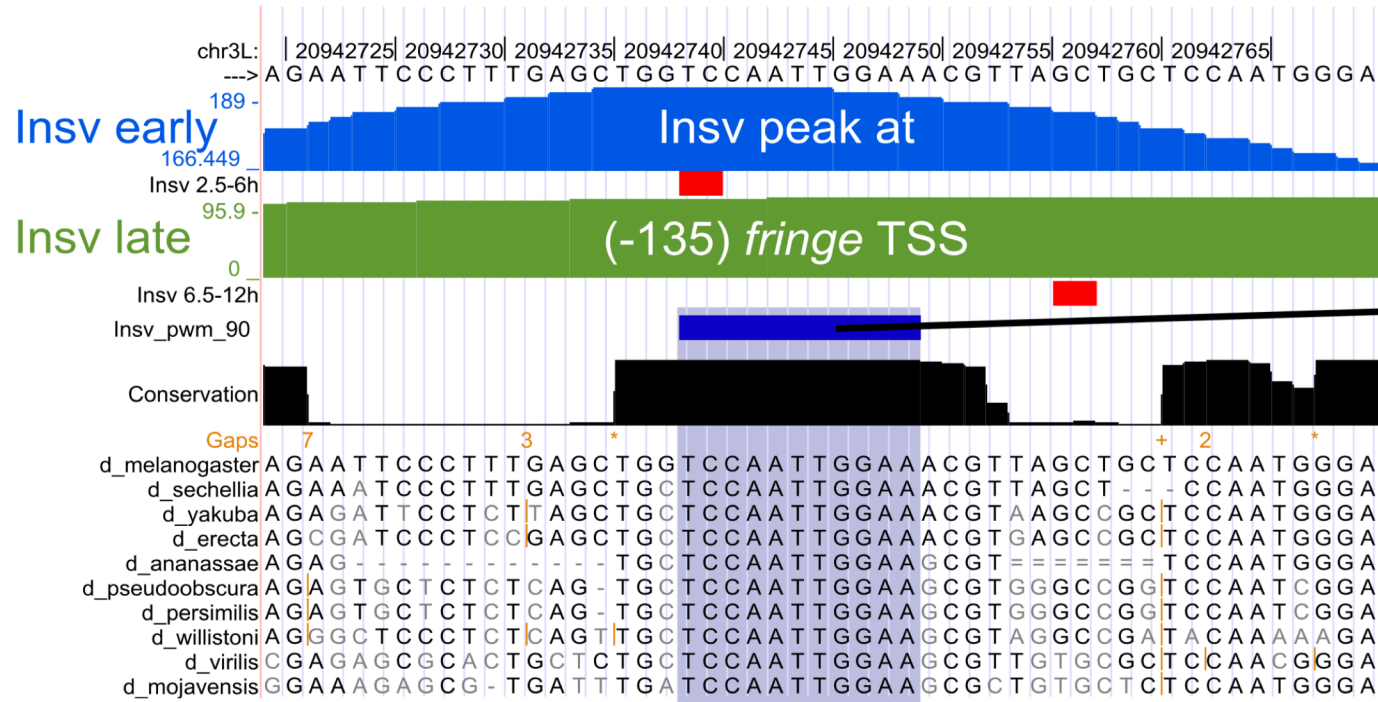
ChIP-seq experiment

- 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

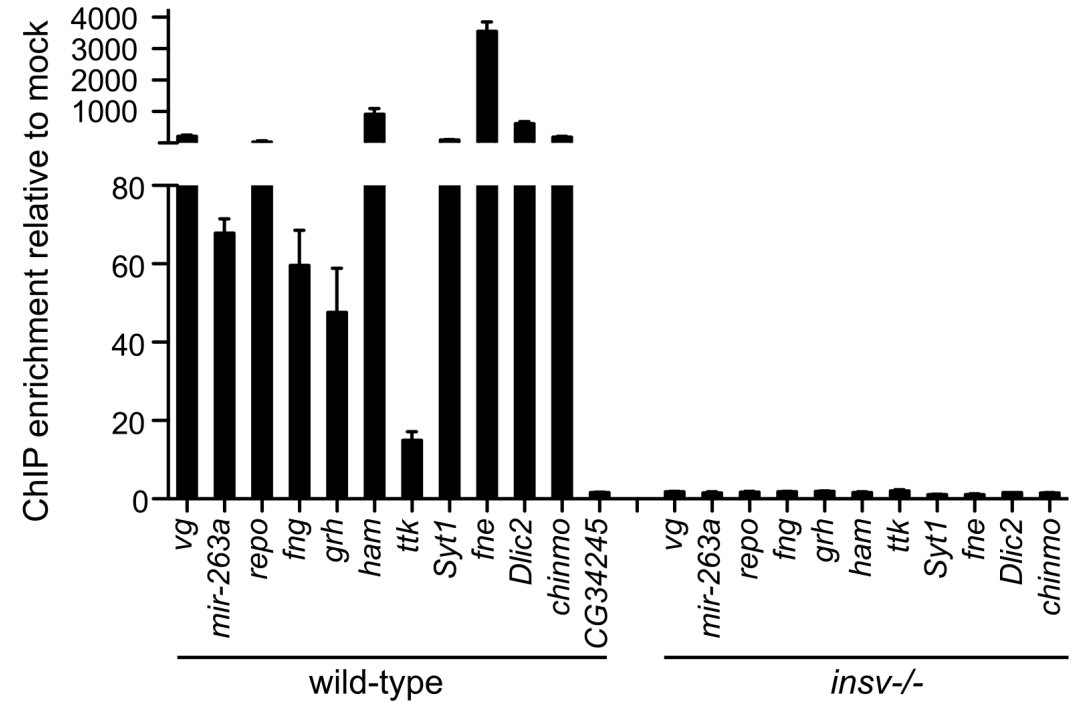
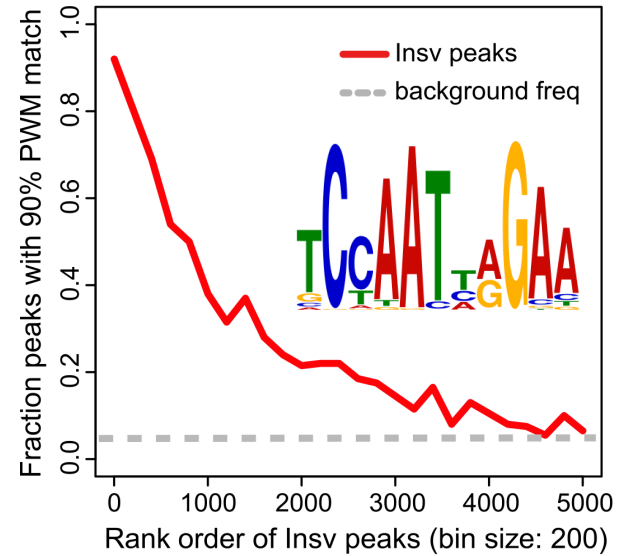
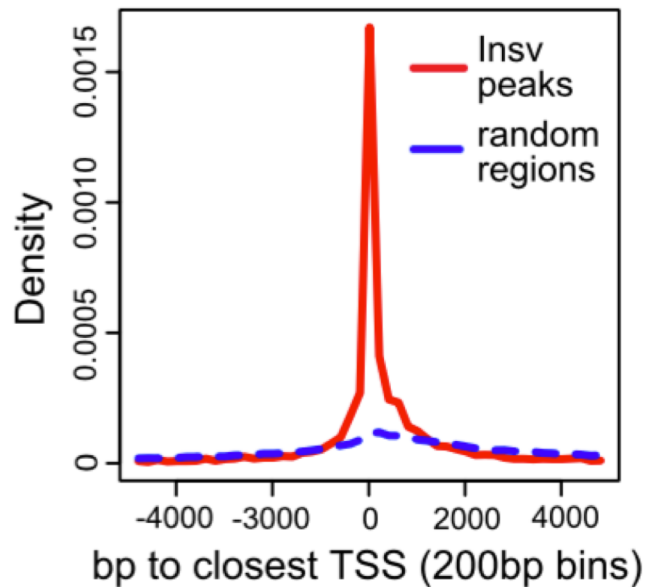
Insenstive seems to bind to a new motif

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



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

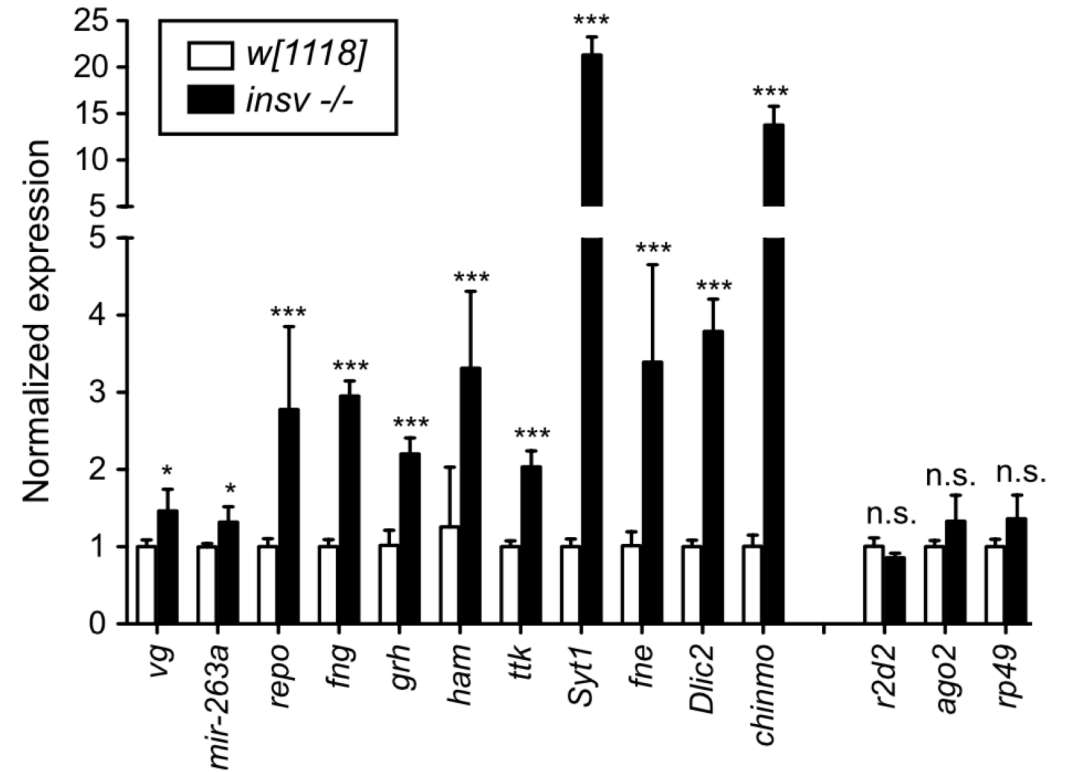
Validating peaks



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

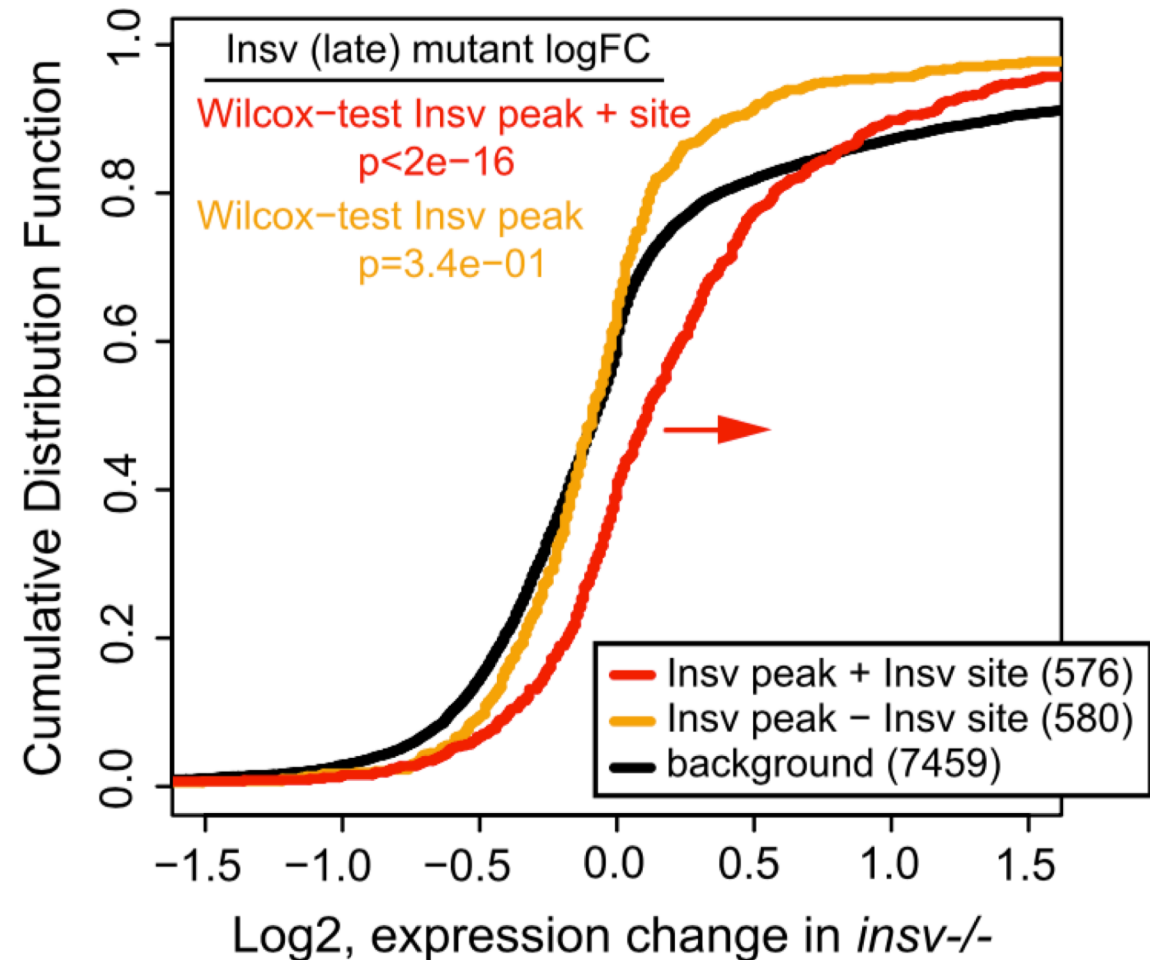
Gene expression

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



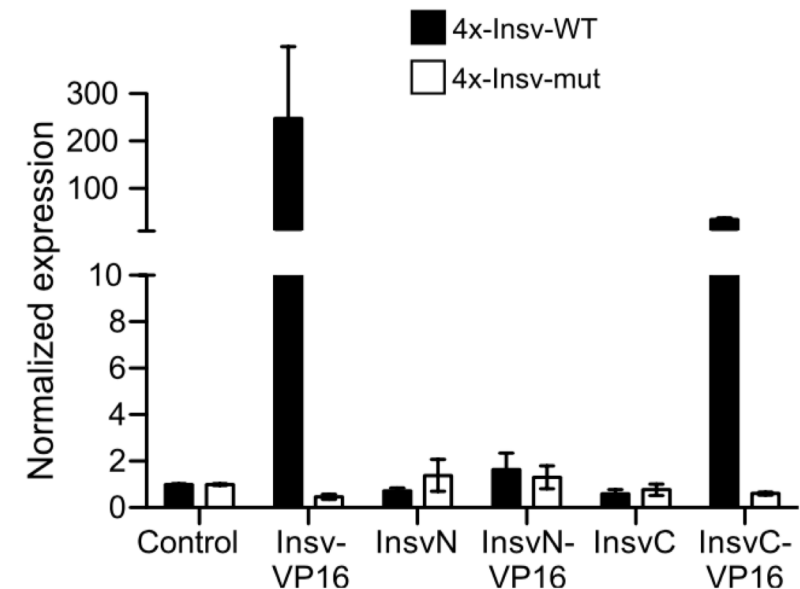
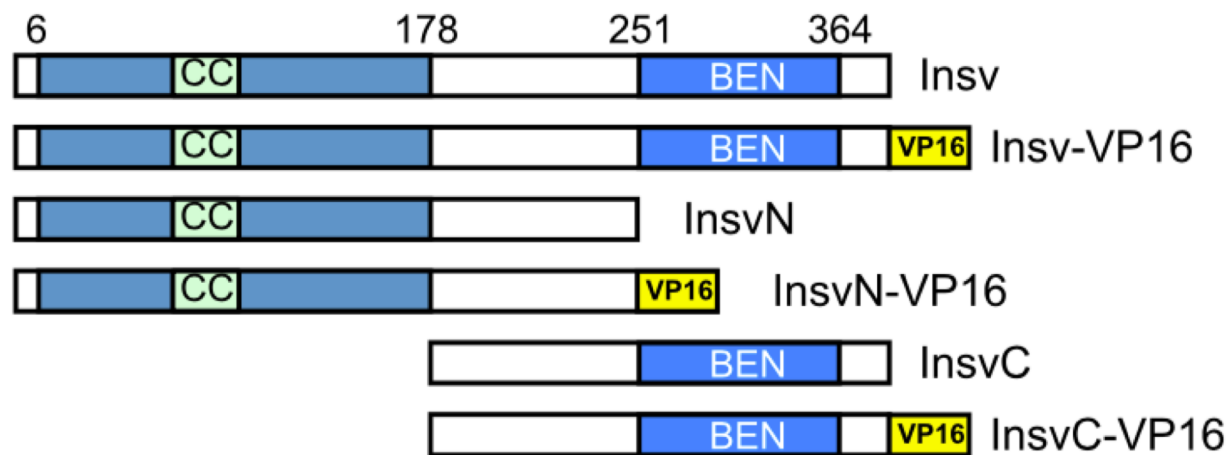
Gene expression, cont.

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

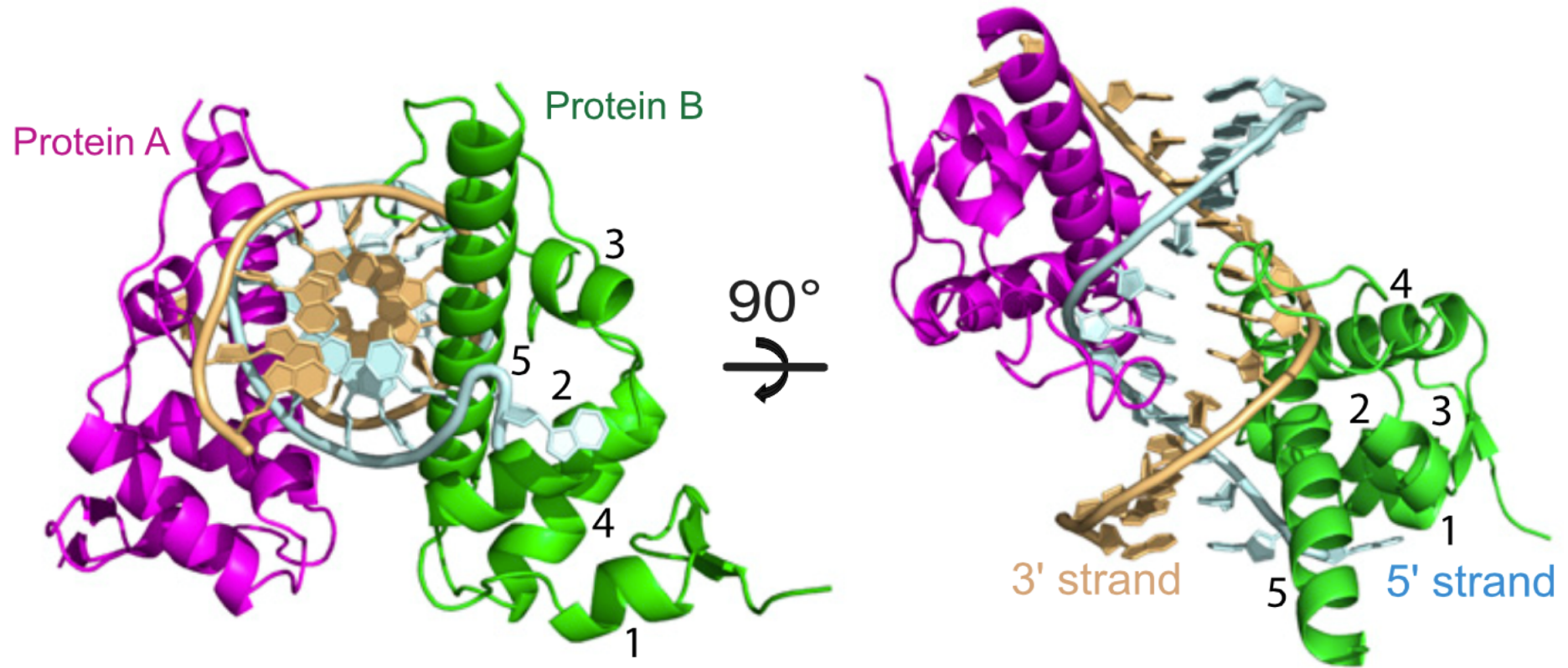


Structure-function experiments

- 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.



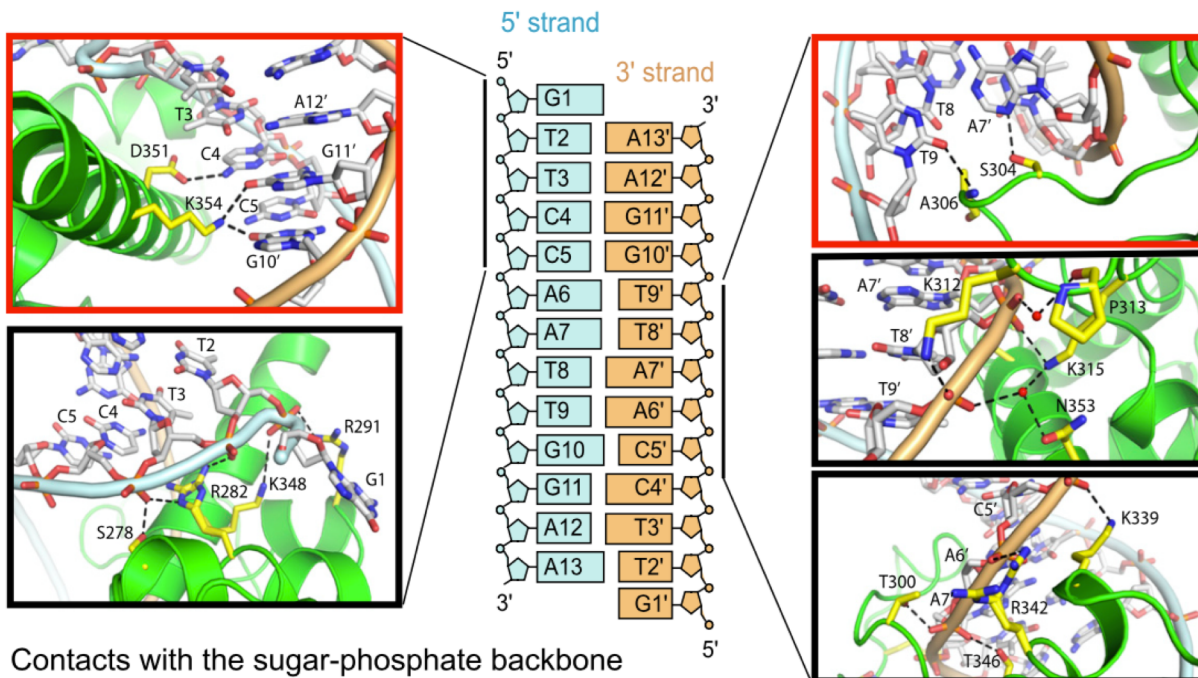
Crystal structure of BEN domain bound to DNA



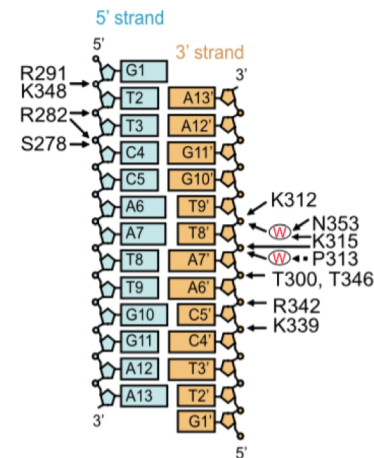
Validating the structure

- 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.

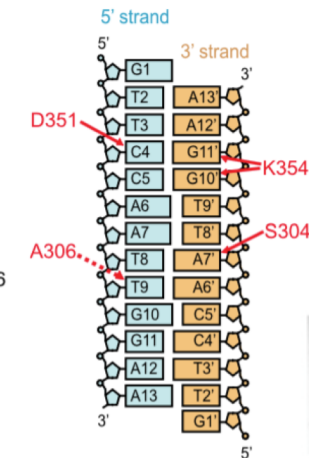
Base-specific hydrogen-bonding contacts



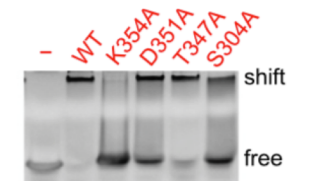
A BEN contacts with sugar-phosphate backbone



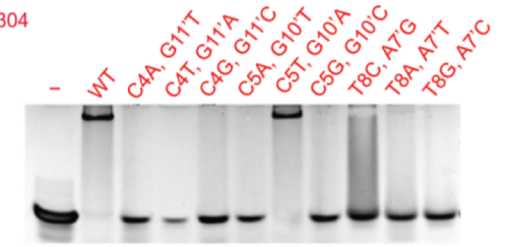
B Base-specific contacts of BEN domain



C Insv-BEN variants tested on wt probe

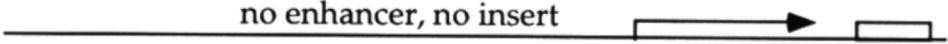
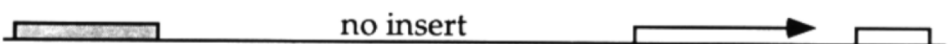
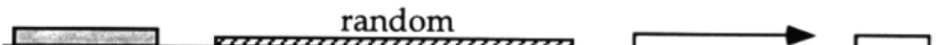




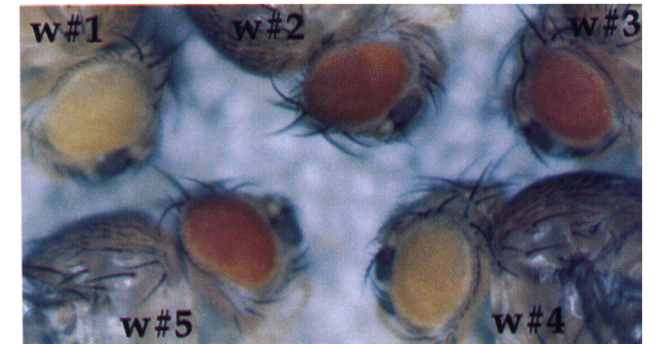
D WT Insv-BEN on variant probes



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.

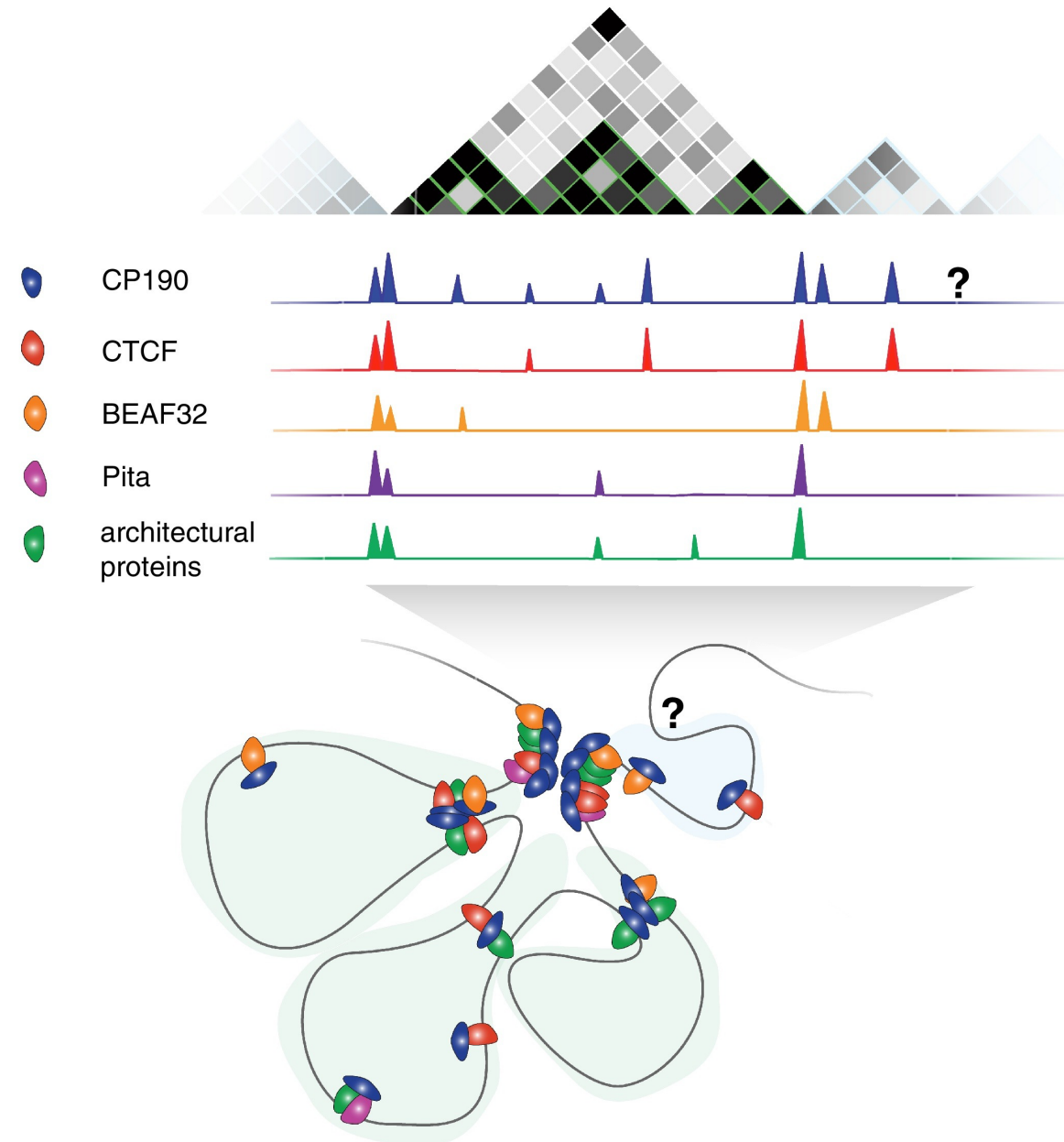
		Enhancer blocking (#light eyes/total)
w#1)	no enhancer, no insert 	— (5/5)
w#2)	no insert 	— (0/6)
w#3)	random 	— (0/5)
w#4)	H3 Xba 	+ (12/24)
w#5)	Xba H3 	— (1/11)



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

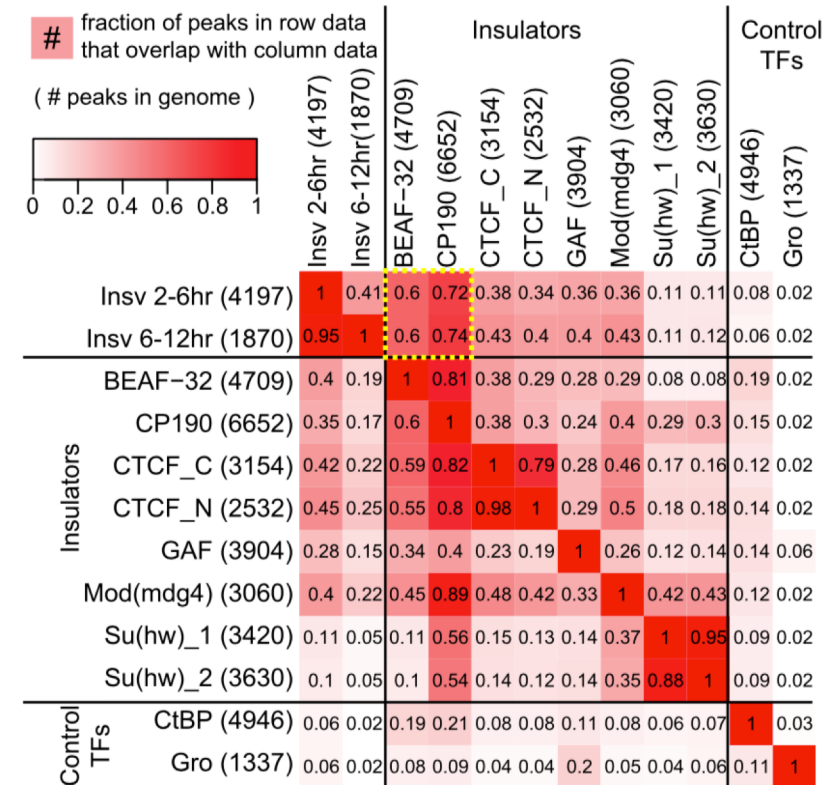
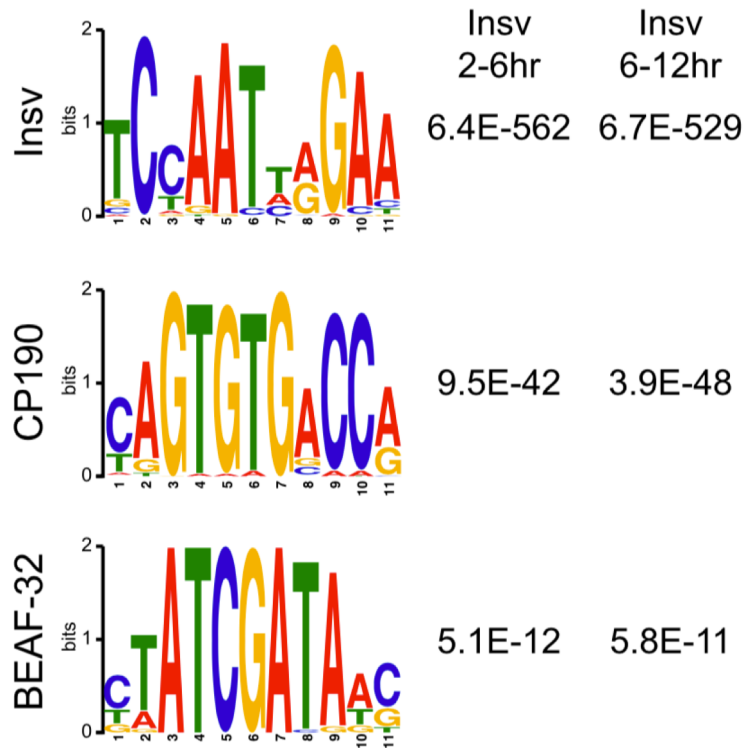
Insulator elements, cont.

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



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

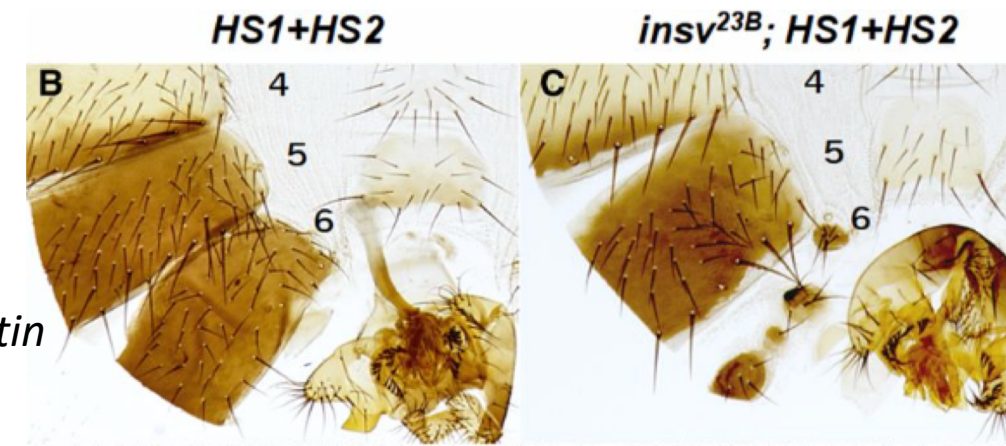
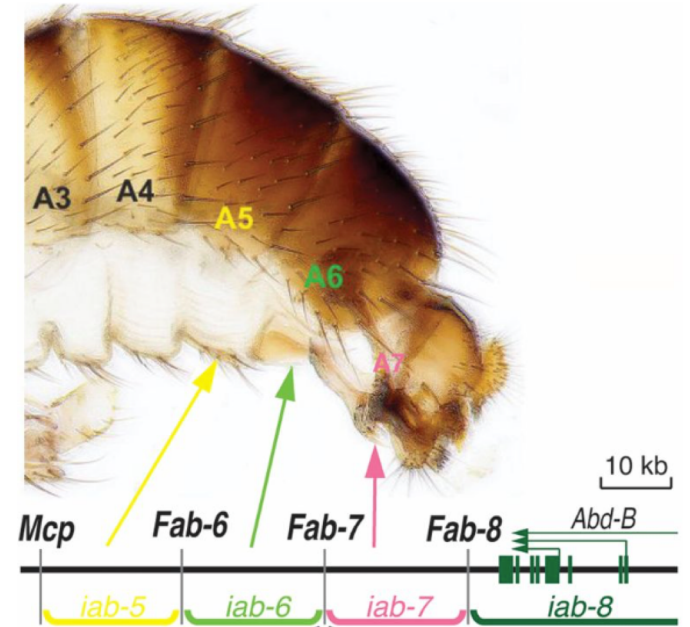
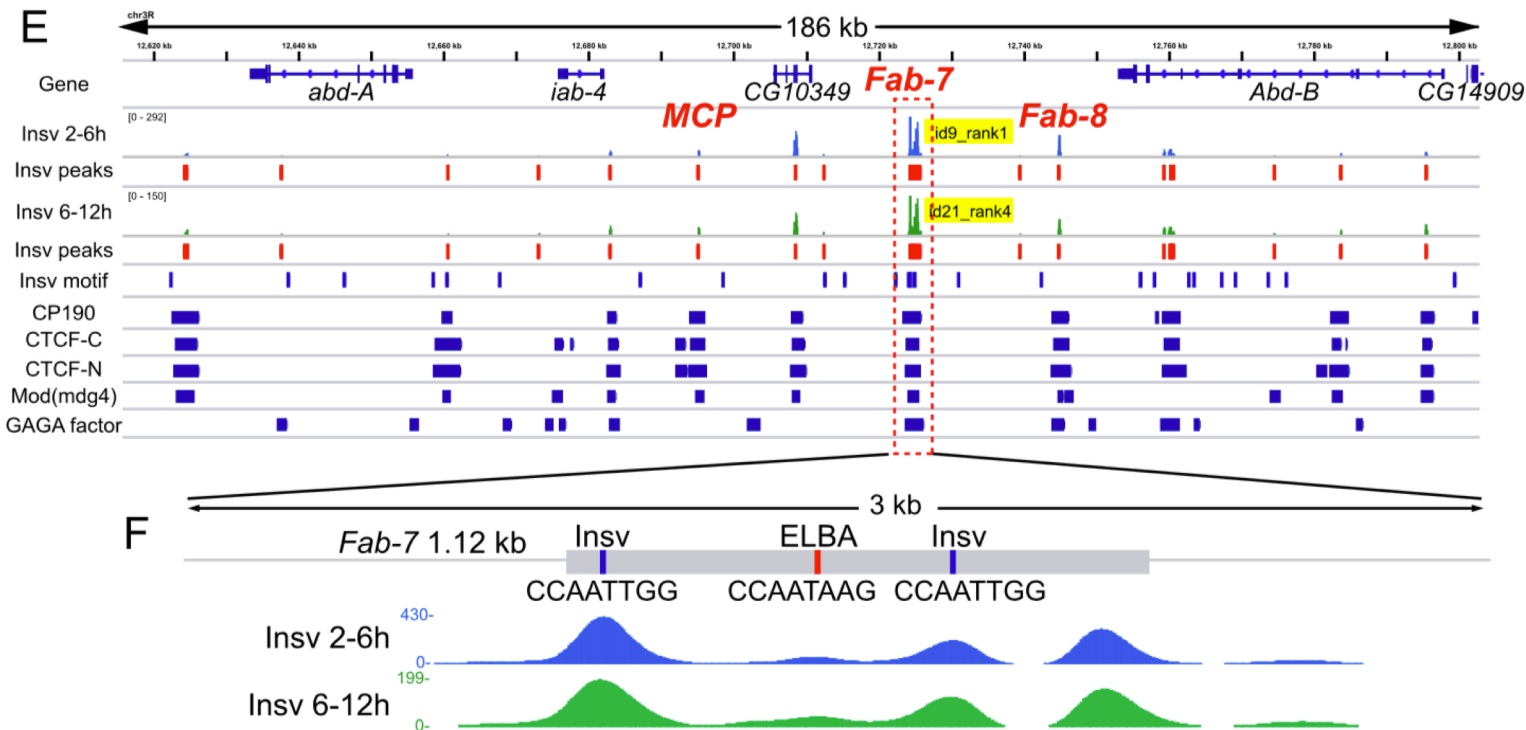
Inensitive binds at insulator elements



- *Inensitive* peaks are enriched for C190 and BEAF-32 motifs
- *Inensitive* 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.

In 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

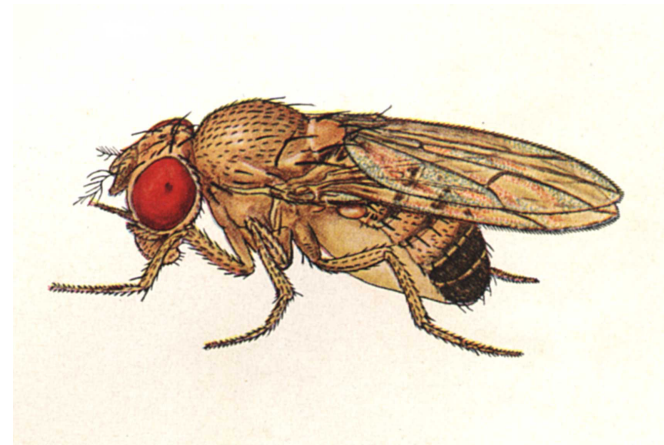
Acknowledgements

Eric Lai (Sloan-Kettering)

Qi Dai
Hong Duan

Dinshaw Palel (Sloan-Kettering)

Aiming Ren
Artem Serganov



Extensions of ChIP-seq

Stockholm, November 8 2018

Jakub Orzechowski Westholm

Long-term bioinformatics support

NBIS, SciLifeLab, Stockholm University

SciLifeLab

NBIS
NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

So far..

.. you have seen how to use ChIP-seq for

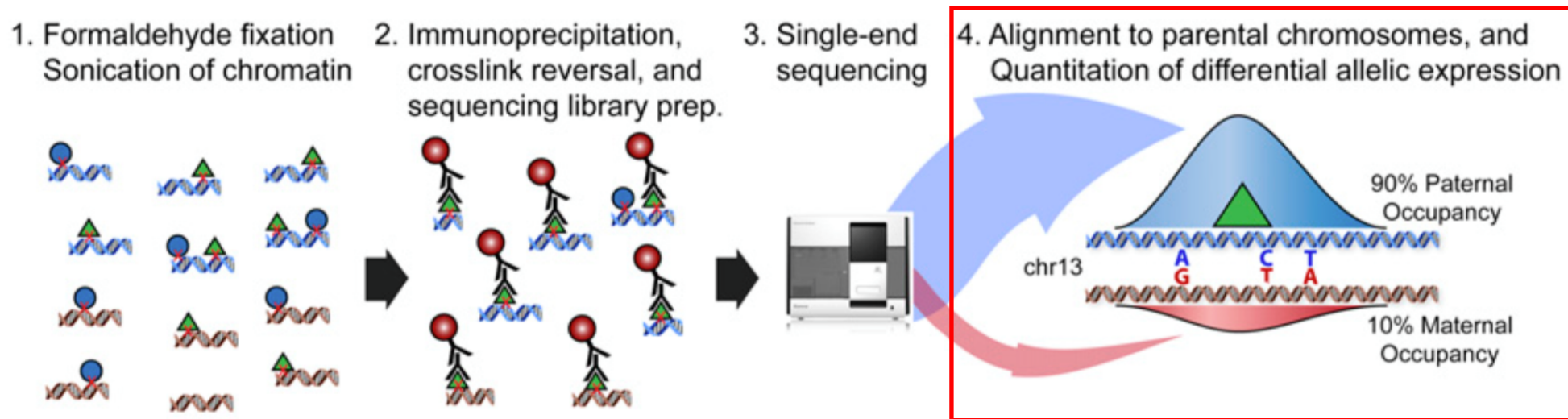
- analyzing which regions of the DNA a protein interacts with
- using a lot of material (millions of cells)

This lecture

- Allele-specific binding of transcription factors
- ChIP-seq from small numbers of cells
- Single cell ChIP-seq

Allele-specific binding

- Using ChIP-seq data it's possible to find variants that affect protein binding.
- If there are heterozygous sites, it's possible to see differences in binding to the two alleles.



Reddy et al. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Research* 2012.

Why is this interesting?

- GWAS studies have found many mutations involved in disease and other traits in non-coding regions.
- It's harder to figure out the effect of such mutations, compared to mutations in coding regions.
- But many non-coding mutations might influence DNA binding of transcription factors or other proteins.
- It's possible to use ChIP-seq data to see which transcription factors are affected, giving an mechanism to the mutations.

Early example:

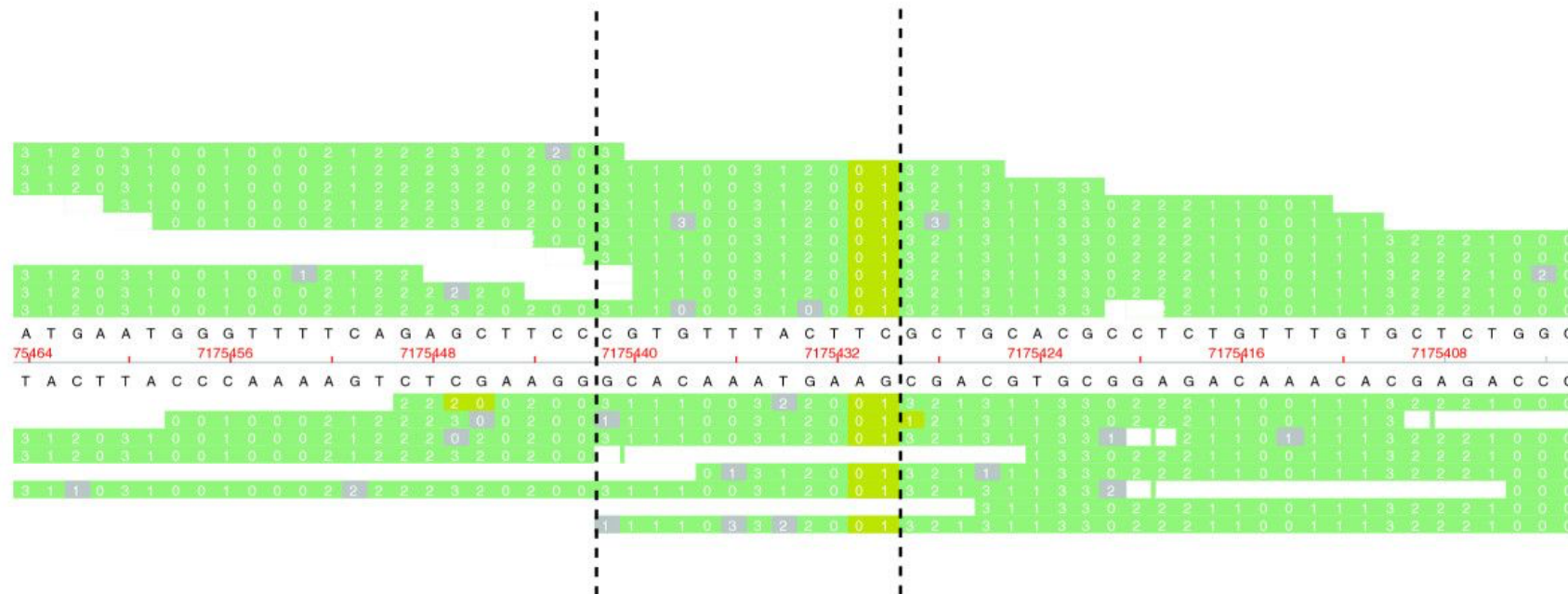
FOXA1 top motif:



HepG2 genomic DNA: CGTGTTTACTTT[T/C]

Reference genome: CGTGTTTACTTC

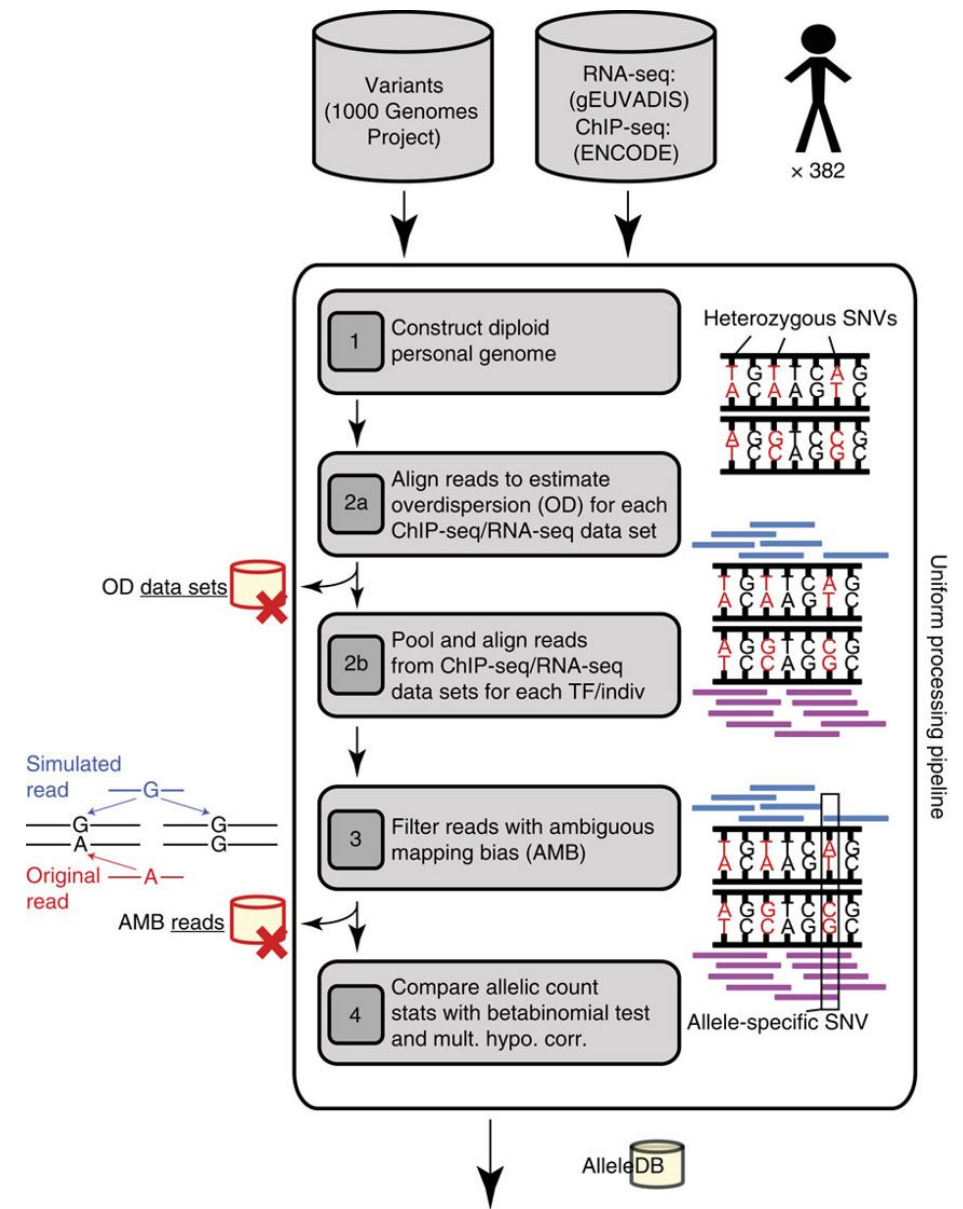
FOXA1 ChIP DNA: CGTGTTTACTTT



Motallebipour et al. Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP-seq. *Genome Biology* 2009.

Procedure

- Need reference genome. Otherwise heterozygous regions where the TF only binds to one allele are missed.
- Need good way to call variants and avoid biases when mapping reads
 - SNVs are easy
 - Small indels also quite easy
 - Large variations harder
- Binomial test for differential binding.



Chen et al. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. Nature Communications 2017.

Individual	SNV position	Functional decoration	Read counts
			A C G T
NA12878	chr3 238,290 ASE	Gene: CHL1	1 5 0 0
HG00096	chr3 238,500 ASB	TF: PAX5	9 23 0 1

Overall results:

- 1-11% of sites have been reported to have allele specific binding (MacDaniell 2010, Rozowski 2011, Reddy 2012)
- Resolution: enrichment for mutations within 50bp of highest point of peak (Reddy 2012)
- TF binding is strongly heritable, more than gene expression (MacDaniell 2010, Reddy 2012, Chen 2017)
- Sites with allele specific binding were significantly enriched for variants associated with disease. (Reddy 2012)
- Some mutations hit the transcription factor motif, but most do not. (Reddy, 2012)
 - other mechanisms for transcription factor recruitment. Co-factors?

Low input ChIP-seq

- Usually ChIP-seq requires a lot of starting material: around 1-10 million cells
- This is a problem when we want to study rare cell types/populations
 - Nervous system
 - Cancer
 - ..

Application with low cell numbers

- Rare neural cell populations:
 - Midbrain dopamine-producing neurons
 - 20,000–30,000 cells per mouse, yield when sorting cells is around 5000 cells
- If we need 1 millions cells per ChIP, it would take over 200 mice
- Now one mouse gives enough cells for 3 ChIPs + input + RNA-seq



Corrected: Author correction

ARTICLE

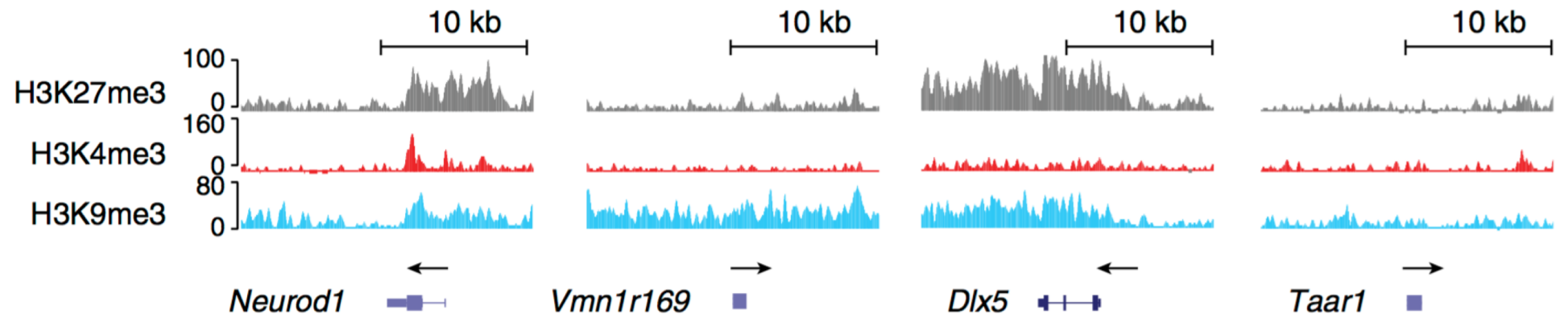
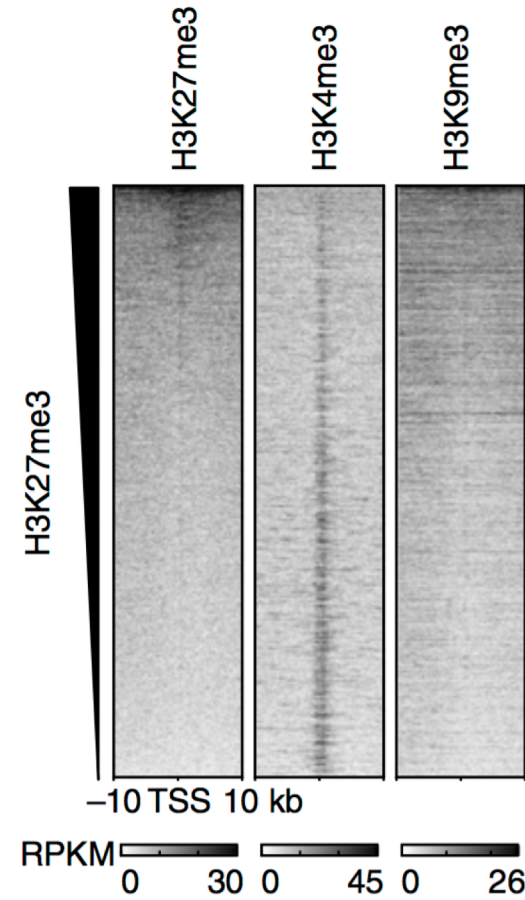
DOI: [10.1038/s41467-018-03538-9](https://doi.org/10.1038/s41467-018-03538-9)

[OPEN](#)

A comprehensive map coupling histone modifications with gene regulation in adult dopaminergic and serotonergic neurons

Erik Södersten¹, Konstantinos Toskas¹, Vilma Rraklli¹, Katarina Tiklova¹, Åsa K. Björklund², Markus Ringnér³, Thomas Perlmann¹ & Johan Holmberg¹

- They were able to get useful data for 3 histone marks.
- Also comparison with RNA-seq data.
- No big changes to analysis
 - Some quality measures might not look as good, e.g. duplication rates
 - QC even more important!



Single cell ChIP-seq

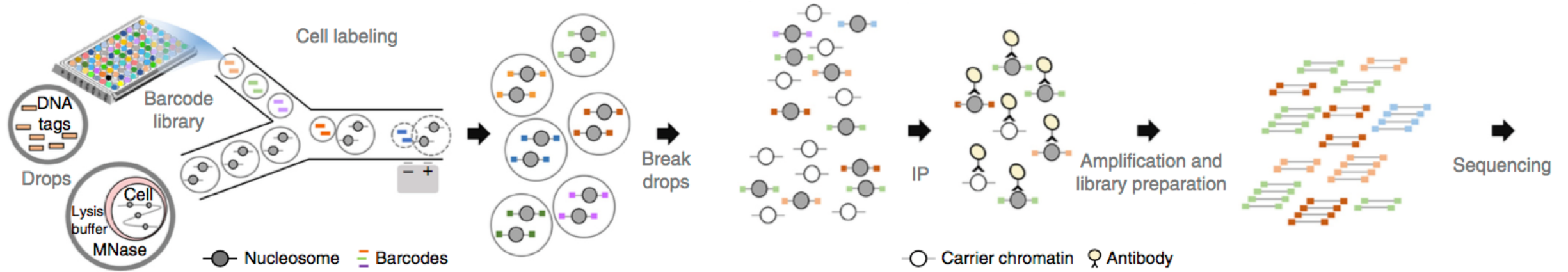
- The signal we get from normal ChIP-seq is an average over all cells in the sample
- This misses heterogeneity
 - Cell types
 - Primed vs unprimed cells
 - Response to stimuli
- With single cell ChIP-seq, we get data for each cell separately
- This is similar to single cell RNA-seq, but much harder (since we only have two chromosome copies, compared to many RNA molecules).

nature
biotechnology

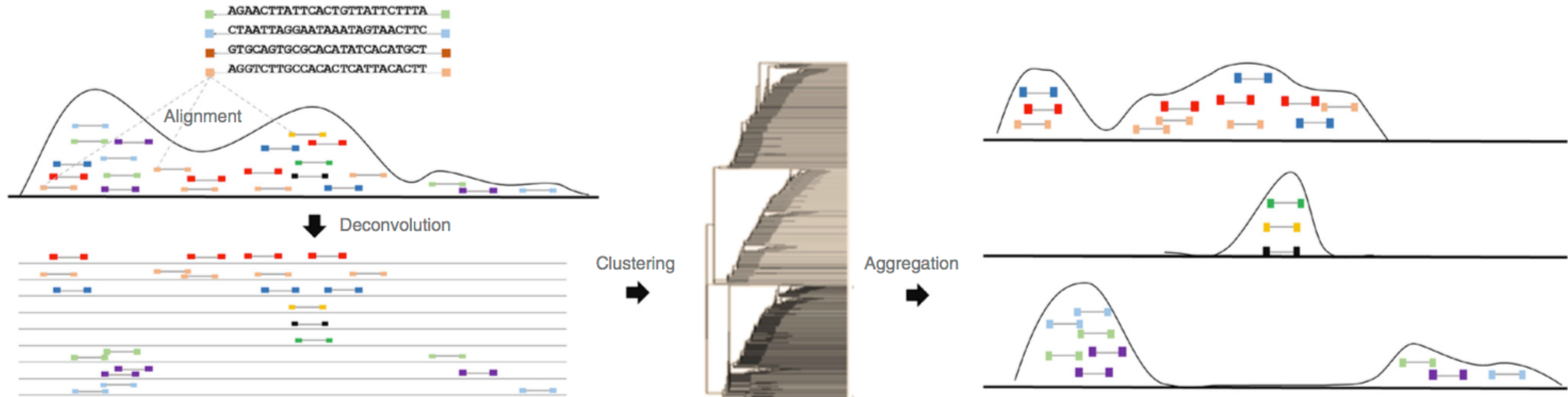
Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state

Assaf Rotem^{1,2,7}, Oren Ram^{2-4,7}, Noam Shoresh^{2,7}, Ralph A Sperling^{1,6}, Alon Goren⁵, David A Weitz¹ & Bradley E Bernstein²⁻⁴

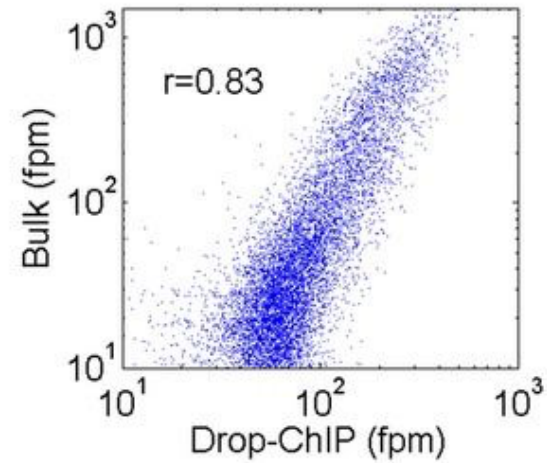
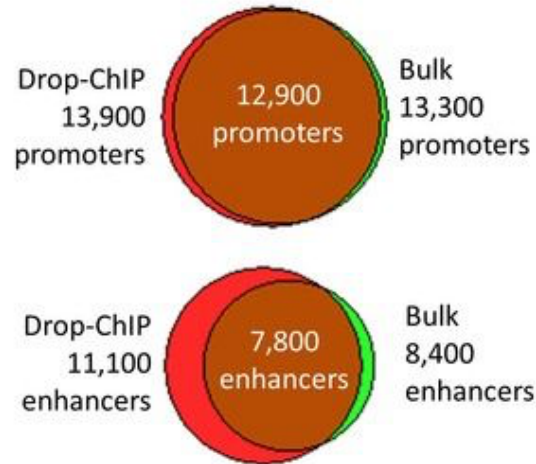
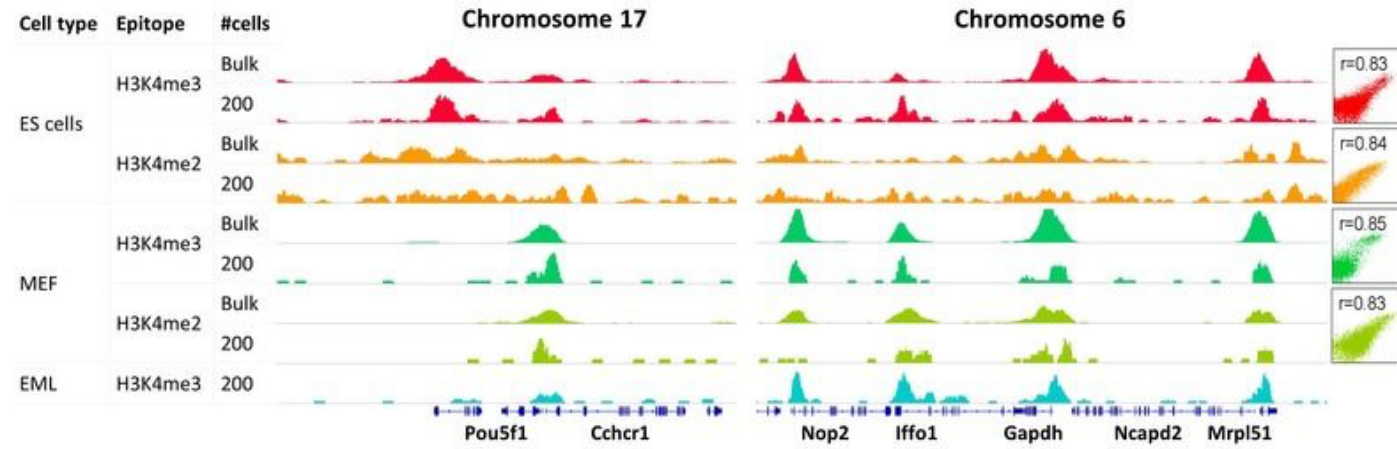
Experiment overview



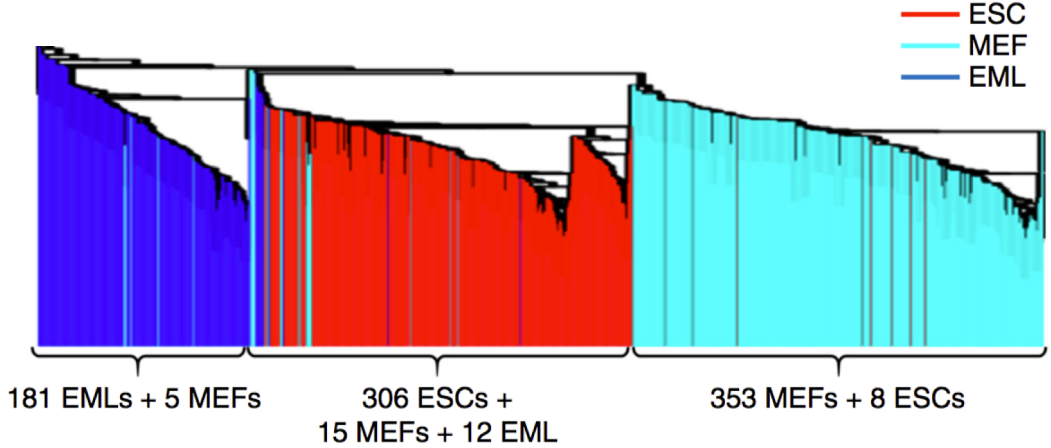
Analysis overview



Aggregated single cell vs bulk data

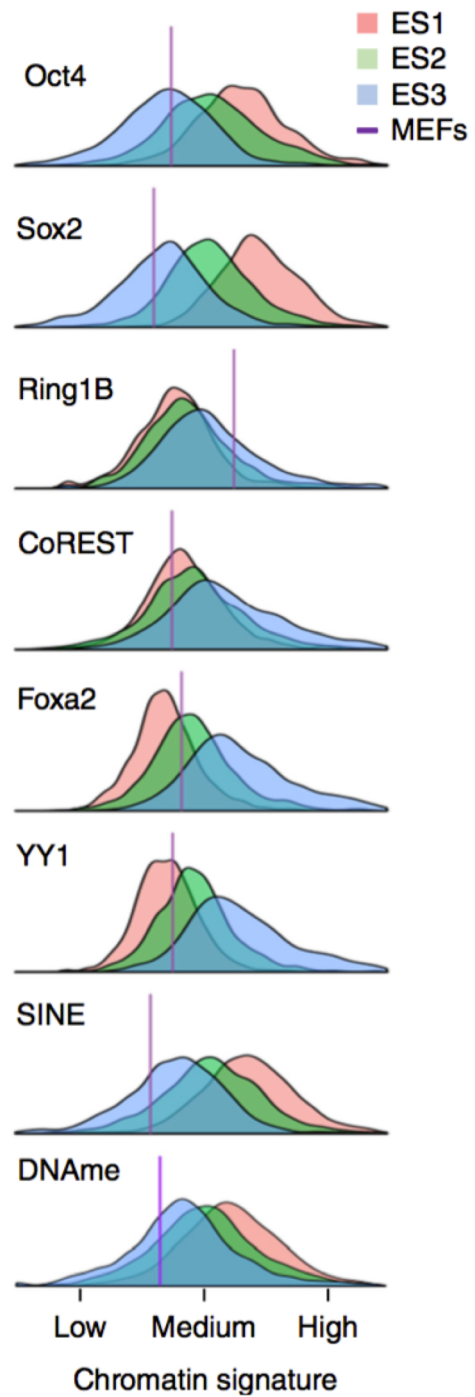
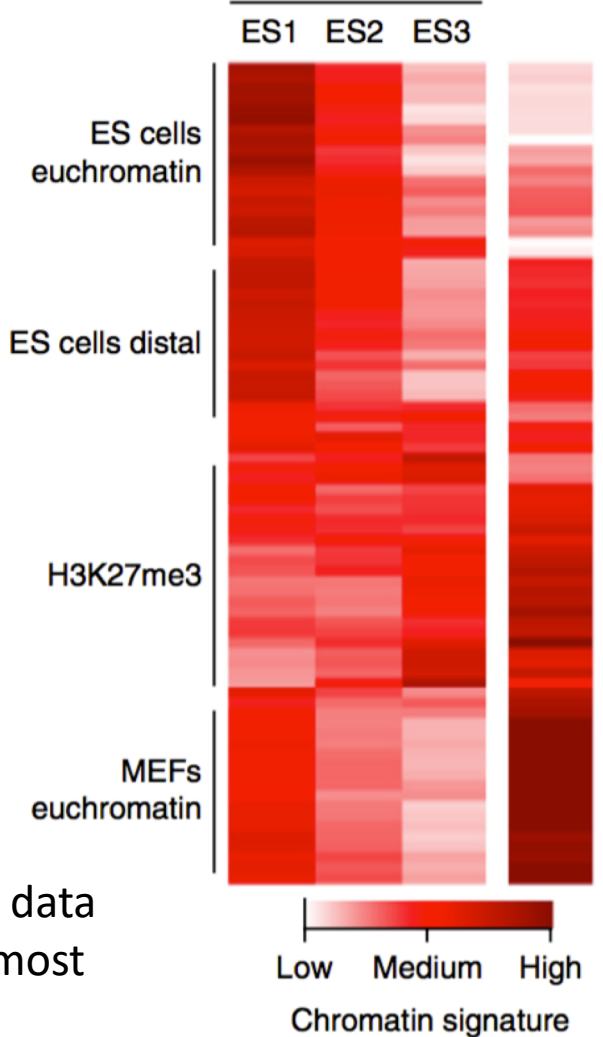
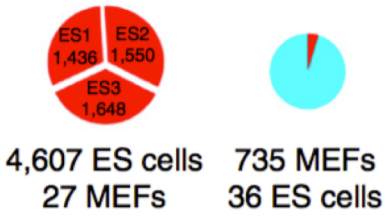


Clustering of single cells



Using promoters and enhancers
 → Possible to separate cell types

Using “chromatin signatures” derived from other data
 → Also possible to separate subpopulations (E1 most pluripotent, then E2m then E3)



Conclusions

- Works
 - Aggregate data look good
 - It's possible (but not easy!) to cluster cells, and find new cell types
- Data from each cell is very sparse
 - This is still enough to cluster cells
 - But this may not be good enough for studying rare cell types
- (Other single cell methods are getting more popular
 - ATAC-seq
 - Bisulphite seq, for DNA methylation).

The End