The relationship between long-range chromatin occupancy and polymerization of the Drosophila ETS family transcriptional repressor Yan

Jemma L. Webber1,2, Jie Zhang1,2,3, Lauren Cote1, Pavithra Vivekanand1,4, Xiaochun Ni5,6, Jie Zhou6, Nicolas Nègre7,8, Richard W. Carthew9, Kevin P. White5,6,7, Ilaria Rebay1,3

1Ben May Department for Cancer Research, The University of Chicago, Chicago, IL 60637, USA
2These authors contributed equally to this work
3Committee on Cancer Biology, The University of Chicago, Chicago, IL 60637, USA
4Current address: Biology Department, Dickinson College, Carlisle, PA 17013, USA
5Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637, USA
6Department of Human Genetics, The University of Chicago, Chicago, IL 60637, USA
7Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL 60637, USA
8Current address: INRA, UMR1333, Université de Montpellier II, Place Eugène Bataillon, 34095 Montpellier, France
9Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

Running title: Chromatin Spreading of the ETS repressor Yan

Keywords: Enhancer, Drosophila development, Signaling pathways

Corresponding author: Ilaria Rebay, Ben May Department for Cancer Research, The University of Chicago, Chicago, IL 60637, USA, irebay@uchicago.edu

Copyright 2012.
Abstract

ETS family transcription factors are evolutionarily conserved downstream effectors of Ras/MAPK signaling with critical roles in development and cancer. In *Drosophila*, the ETS repressor Yan regulates cell proliferation and differentiation in a variety of tissues; however the mechanisms of Yan-mediated repression are not well understood and only a few direct target genes have been identified. Yan, like its human ortholog TEL1, self-associates through an N-terminal sterile alpha motif (SAM), leading to speculation that Yan/TEL1 polymers may spread along chromatin to form large repressive domains. To test this hypothesis, we created a monomeric form of Yan by recombineering a point mutation that blocks SAM-mediated self-association into the *yan* genomic locus and compared its genome-wide chromatin occupancy profile to that of endogenous wild type Yan. Consistent with the spreading model predictions, wild type Yan-bound regions span multiple kilobases. Extended occupancy patterns appear most prominent at genes encoding crucial developmental regulators and signaling molecules, and are highly conserved between *D. melanogaster* and *D. virilis*, suggesting functional relevance. Surprisingly, although occupancy is reduced, the Yan monomer still makes extensive multi-kb contacts with chromatin, with an overall pattern similar to that of wild type Yan. Despite its near-normal chromatin recruitment, the repressive function of the Yan monomer is significantly impaired, as evidenced by elevated target gene expression and failure to rescue a *yan* null mutation. Together our data argue that SAM-mediated polymerization contributes to the functional output of the active Yan repressive complexes that assemble across extended stretches of chromatin, but does not directly mediate recruitment to DNA or chromatin spreading.

Introduction

Dynamic regulation of gene expression during development requires the combined and coordinated action of transcriptional activators and repressors across multiple cis-regulatory modules (CRMs). Research over the last decade has led to a growing appreciation of the existence and importance of both short-range linear and long-range three-dimensional chromatin interactions to overall regulation of gene expression (Bulger and Groudine 2011; Dunipace et al. 2011; Frankel et al. 2010; He et al. 2011; Hong et al. 2008; Perry et al. 2010; Perry et al. 2010; Hug et al. 2010; Shao et al. 2010; Wang et al. 2010; Wu et al. 2010; Zhang et al. 2010; Zhu et al. 2010;).
However, the molecular determinants underlying long-range transcriptional regulation remain poorly understood.

One long-standing hypothesis of transcriptional repression is that the biochemical ability of a factor to polymerize might drive spreading of repressive complexes along the chromatin, thereby providing a mechanism of long-range repression (COUREY and JIA 2001; ROSEMAN et al. 2001). Well-studied examples include multiple Polycomb Group (PcG) corepressors and the ETS family transcriptional repressors TEL1 (ETV6) and Yan, all of which carry a strong oligomerization domain termed the Sterile Alpha Motif (SAM) (KIM et al. 2002; KIM et al. 2001; KIM et al. 2005; QIAO and BOWIE 2005; QIAO et al. 2004; TRAN et al. 2002). In vitro, the isolated SAM domains from these proteins form helical, head-to-tail polymers whose overall structural homology suggests a common mode of function. In vivo, mutations that disrupt SAM-mediated self-association have been shown to reduce or ablate repression activity of both the PcG and ETS proteins in a variety of cultured cell and transgenic overexpression assays (ROBINSON et al. 2012; ROSEMAN et al. 2001; SONG et al. 2005; ZHANG et al. 2010). Genome-wide occupancy analysis of two polymerization competent PcG proteins in Drosophila, Polycomb (Pc) and Polyhomeotic (Ph), have shown that chromatin occupancy "spreads" over regions ranging from several to hundreds of kilobases (NEGRE et al. 2006; SCHWARTZ et al. 2006; TOLHUIS et al. 2006). Comparable studies have not been performed yet for either human TEL1 or Drosophila Yan, and although it is widely inferred, it has not been demonstrated that SAM-mediated oligomerization drives the long-range PcG chromatin occupancy patterns.

Here we have focused on the ETS family repressor Yan which acts downstream of receptor tyrosine kinase signaling in \textit{Drosophila} to orchestrate a proper balance between proliferation and differentiation in a variety of tissues. Thus depending on context, loss of \textit{yan} leads to overproliferation or inappropriate cell fate specification, while overexpression of a constitutively active form can block the induction of a variety of neural, epithelial and mesodermal cell fates (HALFON et al. 2000; HSU and SCHULZ 2000; REBAY and RUBIN 1995; ROGGE et al. 1995). In depth investigation of a small number of direct transcriptional targets identified from genetic studies has led to the suggestion that Yan functions as a short-range passive repressor that competes with the ETS family activator Pointed (Pnt) for access to GGA(A/T) ETS consensus binding motifs within specific cis-regulatory enhancers (KLAMBT 1993; O'NEILL et al. 1994; SCHOLZ et al. 1993). Competition between Yan and Pnt is regulated
by MAPK activation, which attenuates Yan-mediated repression while stimulating Pnt-mediated activation (GABAY et al. 1996). These regulatory interactions have been proposed to provide a bistable switch that must be flipped in order for a cell to commit to a fate (GRAHAM et al. 2010).

To test the model that Yan self-association through its SAM domain can induce spreading of repression complexes over extended stretches of chromatin and to gain further insight into Yan-mediated regulation of gene expression during development, we compared the global chromatin occupancy profile of endogenous wild type Yan to that of a recombineered genomic transgene carrying a missense mutation in the SAM domain that restricts the Yan protein to a monomeric form. Consistent with the starting chromatin spreading model, we find that wild-type Yan binds at developmentally important genes as clusters of densely packed peaks that span multiple kilobases, a pattern that is conserved between D. melanogaster and D. virilis. However, although binding is reduced, Yan monomers still exhibit prominent multi-kilobase chromatin occupancy patterns, an unexpected result in light of the limited ability of Yan monomers to rescue the null allele or repress gene expression. Based on these findings, we propose a revised model in which SAM-mediated polymerization of Yan does not provide the primary chromatin recruitment mechanism, but is instead required for the function, stability and/or maintenance of long-range repressive complexes. Given that SAM-mediated chromatin spreading has also been proposed to underlie the extensive chromatin interactions and long-range repression of PcG proteins, our findings may have broader implications with respect to the relationship between polymerization, long-range chromatin occupancy and transcriptional repression.

Materials and Methods

Fly strains, genotypic selection of yan mutant embryos and rescue experiments

For ChIP-qPCR verification of the ChIP-chip results, ~400 stage 11 GFP-negative yan null embryos were hand selected from the cross: yanER433/CyO, twist-Gal4, UAS-GFP (CTG) × yanER33/CTG. For YanV105R ChIP (~800 embryos per replicate) and rescue experiments, GFP-negative stage 11 embryos were selected from the cross: yanER433/CTG; YanV105R × yanER33/CTG; YanV105R. Controls were similarly selected from the cross: yanER433/CTG; YanWT x yanER33/CTG; YanWT. For the genetic rescue experiments, embryos that hatched were counted as rescued to
larval stage. Embryos that did not hatch were hand removed from their chorion and vitelline membrane, incubated in a 1:1 solution of glycerol:acetic acid overnight, mounted in Hoyers:Lactic acid and dried at 65°C for approximately 18 hours before imaging on a wide field microscope.

**Chromatin Immunoprecipitation**

**w**118 embryos were collected, aged to Stage 5-7 (2h10-3h10) or Stage 11 (5h20-7h20 D. melanogaster; 10-12h at room temperature D. virilis), dechorionated in 50% bleach, cross-linked in 10ml of cross-linking solution (50mM HEPES, pH 7.6, 1mM EDTA, 0.5mM EGTA, 100mM NaCl, 1.8% formaldehyde) and 30ml of n-Heptane for 15 minutes, washed in Stop Solution (PBS, 0.1% Triton X-100, 125mM Glycine), PBS/T (PBS, 0.1% Triton X-100) and Wash Solution (10mM HEPES, 10mM EDTA, 0.5mM EGTA, 0.25% Triton X-100), homogenized in ChIP lysis buffer (50mM HEPES, 140mM NaCl, 1mM EDTA, 1mM EGTA, 1% Triton X-100, 0.1% Sodium Deoxycholate, protease inhibitor tablet [Roche]) and chromatin sonicated to a final size of 500bp using a Fisher Scientific Sonic Dismembrator sonicator (Model 500) with 9 cycles at 15% amplitude for 15 seconds (0.9 seconds on/0.1 second off). Clarified lysates were incubated either with guinea pig anti-Yan (1:500), anti-GFP (1:200; Invitrogen) or mock-treated with no antibody overnight at 4°C. Gamma-bind sepharose beads (GE Healthcare) were added to the lysates and incubated for 4 hours at 4°C to precipitate the DNA-protein complexes. Beads were washed in ChIP lysis buffer, High Salt ChIP lysis buffer (500mM NaCl in ChIP lysis buffer) and TE (10mM Tris pH 8, 1mM EDTA). Chromatin was washed off beads with TE/SDS (10mM Tris pH 8, 1mM EDTA, 1% SDS) and reverse-cross-linked at 65°C overnight. ChIP DNA was purified using the QIAquick PCR purification kit (Qiagen).

Following reverse cross-linking, native ChIP DNA was amplified via the linker-mediated PCR (LM-PCR) method (Lee et al. 2006) using Klenow DNA Polymerase (NEB) and ligation with a linker produced by annealing two oligos: 5’-5Phos/AGAAGCTTGAATTGCAGCTAGTCAG-3’ and 5’-CTGCTCGAATTCAAGCTTCT-3’. After adding the linker, DNA was amplified using the 20-mer primer and QIAquick purified. The dNTP mixture used in the amplification reaction contained a 3:7 ratio of dUTP:dTTP so that the products could be fragmented by Uracil DNA Glycosylase and APE1 (Affymetrix). Fragmentation, labeling, and hybridization were performed as described in the Affymetrix ChIP Assay Protocol. For ChIP-seq, after purification of native
DNA, an adenine residue was added with Klenow [3’-5’ exo-] enzyme. Adaptors from Illumina for LM-PCR were ligated to the end of DNA molecules and the 200bp product of the reaction extracted and purified from a 2% agarose gel. 18 cycles of PCR were performed using phusion polymerase (Finnzyme F-530S) and the Illumina oligos. The product was purified by gel electrophoresis.

**Genome-wide binding profiles and data analysis**

Raw data are available at GSE34038 and GSE34030 and were mapped to the April 2006 *D.melanogaster* genome. Three biological replicates of immunoprecipitated versus no antibody control mock-treated ChIP samples hybridized on Drosophila Genomic Tiling Array 2.0R (Affymetrix) microarrays were analyzed with quantile normalization plus scaling and bandwidth of 200 bp using Affymetrix Tiling Analysis Software (TAS). Analysis using Model-Based Analysis of Tiling Arrays (MAT, (JOHNSON et al. 2006) was performed using the following parameters Bandwidth: 200; MaxGap: 100; Minprobe: 10 (ROY et al. 2010). Peak calling and visual data inspection from both TAS and MAT genome-wide binding profiles was performed in the Integrated Genome Browser (IGB, Affymetrix) using a top 3% or top 5% threshold with a minrun of 300 bp and maxgap of 100 bp (Table S1). Statistical co-occurrence of binding peaks was analyzed using Cooccur R package (HUEN and RUSSELL 2010). ChIP-chip data is shown as -10log10 p-value of the three biological replicates. The most significant peaks are expected to have a high consistency between replicates and thus a high p-value and conversely, peaks with low significance represent those with low consistency between replicates (LANDT et al. 2012).

Two biological replicates of ChIP and an Input sample were sequenced on one lane of Illumina each and high throughput sequencing was performed on an Illumina Genome Analyzer with standard Illumina 36 cycles. The quality-filtered 36bp reads were aligned using ELAND (Efficient Local Alignment of Nucleotide Data)., Following standard practice in the field (ROBERTSON et al. 2007; ROZOWSKY et al. 2009; ZHONG et al. 2010), the two deep-sequencing IP reads were combined to increase the number of reads, and analyzed using Model-based Analysis of ChIP-seq (MACS, (ZHANG et al. 2008)) versus input for peak-calling using genome size of 1.2x 10^8 and p-values 1x10^{-5} or 1x10^{-10}. MACS predicted Yan-bound regions and summits were used for transcription factor occupancy comparisons and motif analysis in Figures 4 & 6, respectively. Smoothed kernel density generated by the ChIP-seq analysis R package spp
(Kharchenko et al. 2008) with bandwidth=200, step=100 was used for visualization of peaks in Figures 1, 3 & 5.

**Identification of High-density Regions and Genomic assignments**

Using a cut-off of the top 3% of Yan-bound peaks called by TAS and a homemade clustering algorithm (available upon request), the genome was separated into Yan-bound or unbound regions for each chromosome. Yan chromatin occupancy (bound length/total length) was calculated for each Yan bound region. High-density regions (HDR) were defined as regions greater than 2kb in which the density of Yan occupancy was greater than 40%. The limits for each HDR were defined when the next Yan-bound peak was at a far enough distance that including it would decrease Yan occupancy levels below the 40% threshold (Figure S8). All peaks that did not fall into an HDR were classified as isolated.

For comparison between datasets, this analysis was carried out using thresholds of 30, 40, 50 and 60% Yan density. To calculate the percentage of total peak length within HDRs in Table S1, the total base-pair (bp) length of Yan bound peaks within HDRs was divided by the total bp length of all peaks in the genome; this number thus represents the fraction of total Yan chromatin occupancy which falls into HDRs. Yan bound-peaks that did not overlap with either 5’UTR, 3’UTR, exon or intron (Flybase r5.40) were annotated as intergenic. Gene assignments within 3kb of Yan peaks were made using BDGP5.25 RefSeq genes (15,842 genes including noncoding RNAs).

**Analysis of D. virilis data**

We utilized the UCSC genome browser LiftOver software to translate each uniquely mapped *D. virilis* solexa sequencing read into the *D. melanogaster* genome, using default parameters with the exception of match = 0.5 to account for the large evolutionary distance between *D. melanogaster* and *D. virilis* (Kuhn et al. 2009).

**Quantitative PCR**

ChIP signals were quantified using the QuantiTech SYBR Green PCR Kit (Qiagen). A standard curve for each primer pair (Table S10) was generated using serial dilutions of genomic DNA. The relative amounts of input, no antibody mock-treated and immunoprecipitated DNA were determined based on the standard curves, and the ChIP signals were calculated as IP/input ratios.
Transcription assays

Reporter constructs were generated by inserting putative Yan binding sequences (Table S9) upstream of the luciferase gene coding sequence. The inserts were chosen based on the ChIP-chip signals and predicted ETS binding sites using CISTER (Frith et al. 2001). Primer sequences used for PCR amplification are available upon request. Transfection of Drosophila S2 cells grown in Gibco Sf-900 serum-free medium (Invitrogen) and transcription assays were as previously described (Zhang et al. 2010).

GO analysis

For Stage 11 and Stage 5-7, the top 2000 genes based on maximum peak p-values were functionally classified with Gene Ontology terms using DAVID (Huang et al. 2009a; Huang et al. 2009b). For the D. virilis dataset, all genes assigned to Yan-bound regions were classified. Only functional clusters with a Bonferroni-corrected p-value of <0.001 were regarded as significantly enriched. Panther (Thomas et al. 2003) was used to assess enrichment of signaling pathways (p<0.05). Cytoscape (Cline et al. 2007) was used to map the interactions from String 9.0 (Szklarczyk et al. 2011) between members of the Notch, Wnt and RTK signaling pathways. Comparisons between single-peak, multiple-peak, or high-density genes were limited to the Stage 11 D. melanogaster dataset.

Motif identification

Positional weight matrices were derived from either a) all occurrences of GGAA/T and the flanking 6 residues found within the Top50 D. melanogaster ChIP-seq sequences (36bp either side of the MACS predicted summit) or b) all experimentally validated GGAA/T motifs within the Yan target genes eve, prospero, lozenge and D-pax2. CentriMo analysis (Bailey and Machanick 2012) was performed with the top scoring 600 ChIP-seq peaks, top 600 isolated and top 600 HDR sequences using PWM motif databases from Flyreg (based on Flyreg Drosophila DNAse I Footprint Database v2.0) and idmmpmm2009 and dmmpmm2009 (Kulkovskii and Makeev 2009; Kulkovskiy et al. 2009). MAST (Bailey and Gribskov 1998) was used to predict binding sites (p<0.0001) in a set of Yan- bound sequences derived from the top 600 ChIP-seq peaks versus a set of control sequences derived from shuffling the regions within the genome using BedTools (Quinlan and Hall 2010). Motif searches were also performed on the top 600 isolated peaks and top 600 peaks associated with HDRs. MEME analysis was carried out...
on 600 sequences consisting of 100bp around the MACS identified ChIP-seq summit (p<1x10^{-10}). DREME analysis (BAILEY 2011) was carried out on the top 600 isolated or HDR ChIP-seq peak sequences. Motif analysis of the genomic sequences used in transcription reporter assays was performed using MAST with a PWM derived from the top 50 Yan-bound sequences and a PWM for Mad from dmmpmm2009 (KULAKOVSKIY et al. 2009).

**Recombineering**

The V105R missense mutation was introduced into a BAC carrying the wildtype (WT) yan locus (2L:2156180-2188108) using the Counter-Selection BAC Modification Kit (Gene Bridges, Heidelberg, Germany) following manufacturer’s protocol. The YanWT and YanV105R BAC constructs were YFP or GFP-tagged at the C-terminus respectively. Subcloning via the BAC Subcloning Kit (Gene Bridges, Heidelberg, Germany) into the transformation vector and site-specific integration into the attP2 site on the 3rd chromosome (GROTH et al. 2004) was performed as described in (LUDWIG et al. 2011). YanWT was also integrated into the attP40 site on the 2nd chromosome to enable construction of the 6xYan stock that carries both copies of the endogenous yan gene plus four copies of YanWT, two on the 2nd and two on the 3rd.

**Embryo staining, quantification of eve expression and protein immunoblots**

Embryos were collected for 2 hours, aged to Stage 11 at 25°C, genotyped as described above using GFP fluorescence of the CTG balancer, and stained as previously described (ZHANG et al. 2010) with guinea-pig anti-Yan (1:2000) or mouse anti-Eve MAb 3C10 (1:10; Hybridoma Bank, Iowa). Using a Zeiss LSM 510 confocal microscope, 20 or 36 serial 0.8 μm z-sections were maximum projected using LSM software for analysis of Eve and Yan expression respectively.

For calculating the relative mean intensity of Eve expression in the mesodermal clusters of yan^{E833/ER443}; YanWT and yan^{E833/ER443}; YanV105R embryos, the mean background pixel intensity was subtracted from the mean pixel intensity for the cluster (measured using Image J with a box of defined size), and normalized by dividing by the average mean pixel intensity for Eve expression in wild-type embryos. For protein immunoblots, ~100 stage 11 embryos per genotype (6xYan or hand-selected yan^{E833}; YanWT and yan^{E833}; YanV105R) were homogenized in 50ml SDS sample buffer (250mM Tris-Cl pH 8, 10% SDS, 50% Glycerol, 5% b-mercaptoethanol, 0.04% bromophenol blue), passed through a 27G needle 10 times and boiled for 10 minutes prior to running on an 8%
SDS-PAGE gel and transfer to PVDF as previously described (ZHANG et al. 2010). The blots were probed with anti-GFP (1:1000; Invitrogen) and anti-tubulin (1:1000; Sigma).

Results

Yan chromatin occupancy spreads across multiple kilobases

To characterize the genome-wide binding profile of Yan and identify a more comprehensive list of targets, we performed a ChIP-on-chip analysis of endogenous Yan in stage 11 embryos, a developmental window in which Yan is both highly expressed and its signaling mediates critical developmental events (HALFON et al. 2000; O’KEEFE et al. 1997; PRICE and LAI 1999; RIEGGO-ESCOVAR and HAFEN 1997; SCHOLZ et al. 1997). High confidence bound regions (Tables S1, S2) were mapped to genes within 3kb, yielding 2901 potential Yan target genes. Approximately 10% of Yan-bound regions fall within known promoter regions, while 44% and 39% map to intronic or intergenic regions, respectively, consistent with the expected locations of cis-regulatory elements (Figure S1). Although the criterion for assigning Yan-bound regions to target genes will mis-assign CRMs that act over large distances and will generate false-positives in gene-dense regions, it is broadly accurate since the assignments included known Yan target genes such as eve, argos, mae, prospero and mir7 (FLORES et al. 2000; GABAY et al. 1996; GOLEMBO et al. 1996; HALFON et al. 2000; LI and CARTHEW 2005; QIAO et al. 2004; VIVEKANAND et al. 2004). Further, the putative targets include 104 genes previously shown by microarray analysis to be responsive to perturbation of the EGF pathway in either follicle cells or third instar wing discs (Table S3) (BUTCHAR et al. 2012; JORDAN et al. 2005).

To validate the ChIP-chip results, we performed ChIP followed by quantitative PCR (qPCR) at selected Yan-bound regions from stage 11 wild type or hand-selected yan null mutants. Although there is no evidence of maternally provided Yan (GABAY et al. 1996; ROGGE et al. 1995), yan null embryos do not die until after stage 16, permitting analysis at stage 11. All nine regions tested showed enriched Yan occupancy in wild type embryos, but not in yan null mutants (Figure 1A). Confirming antibody specificity, immunostaining of wild type and yan null embryos revealed only background level fluorescence in the mutants (Figure S2). Yan occupancy patterns were also independently verified at selected targets by ChIP qPCR analysis of embryos carrying GFP-tagged wild-type Yan using an anti-GFP antibody (Figure S3). Finally
we showed that Yan-bound regions contain elements capable of responding to both Yan and Pnt by cloning candidate CRMs with predicted ETS binding clusters (identified by Cister (Frith et al. 2001); see also Table S9) upstream of the firefly luciferase coding sequence and measuring activity in transfected cultured S2 cells. Sixteen of the 18 reporters tested were significantly activated by Pnt and repressed by Yan (Figure 1B).

In agreement with the starting prediction that Yan may spread linearly across the DNA, a prominent feature of the genome-wide occupancy profile was that Yan binding occurs not only as isolated peaks (Figure 1C) but also as clusters of densely packed peaks that span multiple kilobases (Figure 1D). Several controls were performed to ensure that these spreading patterns were not the result of experimental artifacts. First, we used PCR analysis to confirm efficient sonication of the genomic DNA into fragments of an average size of 500bp or less (Figure S4). Second, a follow-up experiment using ChIP-seq revealed similar binding patterns (Figure 1C, D), ruling out hybridization artifacts with the ChIP-chip platform. Thus, as our protocol is standard for the field (see methods) and the results reproducible between independent experiments and platforms, the observed patterns are unlikely to be the result of excessive cross-linking (Table S1).

To determine whether the phenomenon of multi-kb Yan chromatin occupancy is stage specific, we profiled stage 5-7 embryos, a developmental time-point in which the function of Yan remains largely uncharacterized despite its prominent expression (Price and Lai 1999). We separated putative target genes into two bins based on whether single or multiple Yan-bound peaks were detected, and then compared the assignments at stage 5-7 versus stage 11. Overall, Yan binding patterns were broadly similar across the two time points with approximately 60% of genes bound by Yan at both stages retaining their classification as either a single or multiple-peak gene. Thus multi-kb spreading appears to be a general characteristic of Yan occupancy, rather than a stage-specific anomaly (Figure 2A, B).

However, approximately 11% of genes that retained their classification of having a single isolated Yan-bound peak, and almost 50% of genes classified as having a clustered multiple peak Yan binding profile, displayed stage-specific differences in chromatin occupancy. A striking example of such potential differential regulation was observed at genes involved in axial patterning of the early embryo, a developmental context in which Yan function has not been
previously implicated. Thus, the majority of the anterior-posterior segmentation network, including maternal effect, terminal, gap, pair-rule, segment polarity, and homeotic selector genes were identified as putative Yan targets at both stage 5-7 and stage 11 (Figure 2C). Analysis of gene expression profiles available from the Berkley Drosophila Genome Project (TOMANCAK et al. 2002; TOMANCAK et al. 2007) revealed that with the exception of bcd and osk, these genes have dynamic expression profiles throughout the embryonic stages we have profiled and are expressed in regions of the embryo that overlap with yan’s expression profile (Table S4). At approximately 80% of these genes, temporally dynamic binding patterns of Yan chromatin association were detected (Figure 2C, asterisks). For example, at the knirps (kni) locus, Yan binding shifted from multiple peaks upstream of the transcriptional unit in early embryos, to multiple peaks across and downstream of the gene at stage 11 (Figure 2D). At the odd-skipped (odd) locus, we observed both loss and gain of individual peaks between stage 5-7 and 11 (Figure 2E). At each of these examples, stage-specific Yan binding appeared closely correlated with DNAse accessible chromatin (LI et al. 2011). Across the genome, Yan-bound peaks overlapped with stage 5 and stage 11 DNAse accessible chromatin by approximately 70 and 87% respectively, suggesting that the dynamic patterns of Yan recruitment to DNA may result from changes in chromatin conformation at different stages of development.

**Yan chromatin association patterns are conserved in Drosophila virilis**

If chromatin binding across multi-kilobase regions is important for Yan-mediated regulation of its target genes, then both this tendency and the specific pattern at a given target gene should be evolutionarily conserved. To address this, we performed ChIP-seq from stage 11 *D. virilis* embryos, a species that despite diverging from *D. melanogaster* more than 60 million years ago, exhibits a conserved pattern of embryonic Yan expression (PRICE and LAI 1999). Analysis of Yan chromatin occupancy patterns using LiftOver data mapped to the *D. melanogaster* genome (see methods) revealed similar patterns across both species, with conservation of both the putative target genes bound by Yan as well as the actual Yan binding footprint within multi-peak regions (Figure 3A).

To facilitate comparisons of these datasets, we developed an algorithm to quantify the prevalence of clusters of Yan-bound peaks spanning multiple kilobases, defining a region of
chromatin as a high-density region (HDR) if the length of Yan-bound peaks divided by the total length of the bound region was greater than a density threshold (see methods). An example of an HDR at the edl locus, defined at increasing density thresholds, is given in Figure 3B. Across the genome, a similar fraction of Yan-bound peaks are assigned to HDRs at each density threshold within both datasets. For example, at 40% Yan occupancy, 39% of Yan-bound peaks are within HDRs in the melanogaster dataset compared to 44% of Yan-bound peaks within the virilis dataset (Figure 3C). Further, 40% of Yan bound peaks present in the D. virilis dataset overlap with peaks called within the D. melanogaster dataset. Of these shared peaks, more than 60% are assigned to HDRs in the D. melanogaster dataset (Figure 3D). This relatively high degree of conservation of HDR-type Yan-bound regions suggests that these CRMs could be involved in regulation of conserved expression patterns that are subject to strong selection.

In agreement with this hypothesis, analysis of gene ontology (GO) terms associated with putative target genes assigned to Yan-bound regions from D. melanogaster and D. virilis revealed that while single-peak genes are not significantly enriched for GO terms, genes associated with multiple-peaks or HDRs are enriched for terms reflecting involvement in many aspects of development, with significant over-representation of several signal transduction pathways (Figure 3E, Table S5-7). Both single-peak and high-density genes were among those validated by qPCR and transcription assays (Figure 1A,B, S3), suggesting both categories include biologically relevant target genes. Thus the differences in complexity of Yan occupancy patterns likely reflect the requirement for more complicated regulation of key signaling and patterning genes.

Particularly striking pathway signatures were observed within the EGFR, Wingless and Notch signaling networks in both species (Figure 3E). To highlight this, and using just the melanogaster data, we mapped the interactions between the canonical members of these three pathways using a cytoscape STRINGWSClient plug-in (CLINE et al. 2007; SZKLARCZYK et al. 2011); genes with Yan-bound peaks are highlighted in orange (Figure 3F). Analysis of the network suggests that Yan influences the expression of many core components of each signaling pathway predominantly via high-density type binding (Figure 3F, orange diamonds). Further, genes associated with HDRs appear more highly connected within the EGFR-Wingless-Notch network, showing an average of 13.3 interactions relative to 7.4 for the remaining Yan-bound genes (Figure 3F, orange circles).
**Yan exceeds other transcription factors in high-density type chromatin binding**

Previous studies of other sequence-specific transcription factors have noted multi-peak regions analogous to the Yan-bound HDRs we have described (Li et al. 2008; NEGRE et al. 2011; SLATTERY et al. 2011; ZINZEN et al. 2009). To compare the extent and prevalence of HDR patterns of chromatin association, we carried out a meta-analysis of modENCODE ChIP-chip data for 19 other sequence-specific transcriptional regulators (MOD et al. 2010).

First, using our HDR algorithm, we determined for each transcription factor the frequency with which bound peaks cluster into HDRs and the length distribution of bound regions. To compare between datasets we assessed the top 3% of bound regions for each transcription factor. Across the genome, Yan exceeds all other transcription factors in both HDR frequency and length, with only Kruppel (Kr) approaching comparable levels (Figure 4A,B). Distalless (Dll) and Ultrabithorax (Ubx) exhibit prominent, but less prevalent and extensive HDR binding (Figure 4A,B).

Second, focusing on the EGFR-Notch-Wingless signaling network genes (Figure 3F), we asked whether multi-kb Yan occupancy might correlate with clustered binding of other transcription factors. However, unlike Yan, none of the modENCODE transcription factors analyzed, including those with prominent genome-wide HDR signatures such as Kr, Dll and Ubx (Figure 4A,B), exhibit extensive occupancy at the canonical EGFR, Notch and Wingless signaling network genes (Figure 4C).

**SAM-mediated oligomerization is not the primary determinant of high-density Yan chromatin association patterns**

The high-density chromatin association patterns we observe with Yan appear consistent with the hypothesis that SAM-mediated oligomerization could induce long-range spreading of repression complexes (QIAO and BOWIE 2005; QIAO et al. 2004). Two additional predictions of this chromatin-spreading model are that increasing Yan levels should increase polymer formation and lead to even more extensive chromatin occupancy, while blocking SAM-mediated self-association should reduce the extent of occupancy. To test these predictions, we generated
transgenic animals carrying either six copies of the wild type yan gene (two endogenous copies plus four genomic transgenes; referred to as 6xYan) or a version of yan into which the V105R missense mutation in the EH surface of the SAM domain was introduced. The V105R mutation blocks polymer formation in vitro and increases Yan diffusion rates in fluorescent recovery after photobleaching assays in Drosophila cells, consistent with loss of polymer formation (QIAO et al. 2004; ZHANG et al. 2010). To ensure correct spatio-temporal patterns and levels of expression, a recombineered BAC clone spanning the entire yan genomic locus was used to generate both the wild type and V015R transgenes (see methods). Immunostaining revealed normal patterns of Yan protein expression, with levels elevated in 6xYan embryos and reduced in Yan\textsuperscript{V105R} mutants; the latter showed diffuse staining in both the cytoplasm and nucleus (Figure S5), consistent with our previous finding that loss of SAM-mediated polymerization increases Yan nuclear export (ZHANG et al. 2010). Whereas the control Yan\textsuperscript{WT} transgene fully rescued the yan null mutant, over 50% of yan\textsuperscript{+/+};Yan\textsuperscript{V105R} embryos died just before hatching with an anterior open phenotype (Figure 5A,B). The few escaper Yan\textsuperscript{V105R} animals that survived to adulthood were infertile and died after a few days (data not shown). Elevated expression of eve and an increased number of Eve-positive mesodermal cells in yan\textsuperscript{+/+};Yan\textsuperscript{V105R} embryos relative to control yan\textsuperscript{+/+};Yan\textsuperscript{WT} embryos (Figure S6) confirmed impaired repression of Yan target gene expression. These results agree with previous experiments with overexpressed cDNA transgenes that showed the V105R mutation, while not compromising DNA binding, abrogates the ability of Yan to repress gene expression and to supply full in vivo function (QIAO et al. 2004; ZHANG et al. 2010).

Surprisingly, ChIP-chip analysis of both 6xYan and hand-selected yan\textsuperscript{+/+};Yan\textsuperscript{V105R} stage 11 embryos revealed broadly similar occupancy profiles to that of wild type embryos, suggesting that SAM-mediated polymerization is not the primary determinant of Yan occupancy patterns (Figure 5C). We quantified the prevalence of HDR type binding and found that while the 6xYan condition did not increase the frequency or extent of HDR occupancy genome-wide (Figure 5C, D and Table S1), quantitative changes in Yan\textsuperscript{V105R} binding patterns relative to wild type were apparent. Thus 20% of Yan\textsuperscript{V105R} peaks fall in dense occupancy regions at a 40% threshold, as compared to 41% of wild-type Yan peaks, with a reduced median HDR length of 3.3 kb compared to 4.1 kb for wild-type Yan (Figure 5D). While the amplitude of bound peaks was generally reduced in the V105R sample, presumably reflecting the lower Yan protein levels (Figure S5), and may account for some of the reduction in HDR occupancy, the loss of Yan-
bound peaks in the V105R dataset was significant across a range of bound thresholds (Figure S7). We therefore propose that SAM-mediated polymerization is not, as previously suggested, the primary determinant for Yan spreading over extended chromatin domains, but rather contributes to the formation, stabilization and function of the molecular complexes that occupy HDRs.

**ETS and MAD motifs are enriched in Yan-bound sequences**

Given the unexpected discovery that SAM-mediated self-association appears unlikely to be the primary determinant of Yan chromatin occupancy, an alternate hypothesis is that Yan could bind and spread along DNA via direct recognition of consensus ETS binding sites located throughout the Yan-bound regions. To explore how differences in the number and/or spacing of ETS motifs contribute to Yan chromatin binding patterns, we performed central motif enrichment analysis using CentriMo, a visualization and statistical analysis tool that identifies the region of maximum central enrichment in a set of ChIP-seq peak regions and displays the positional distributions of predicted sites (BAILEY and MACHANICK 2012). Analysis of the top 600 ChIP-seq sequences revealed that while all ETS motifs tested are centrally enriched over flanking regions, motifs derived from the top 50 MACS sorted ChIP-seq peaks (see methods for details) and the human (h)-ETS1 motif (WEI et al. 2010) show both a stronger central enrichment and occur more frequently than either the Drosophila ETS family Eip74EF motif (KULAKOVSKII and MAKEEV 2009), the experimentally derived TEL1 motif (h-ETV6 (WEI et al. 2010) or a motif derived from 16 experimentally validated ETS motifs within the Yan target genes prospero, eve, D-pax2 and lozenge (BEHAN et al. 2002; FLORES et al. 2000; HALFON et al. 2000; XU et al. 2000) (Figure 6A, B).

Analysis of the top 600 ChIP-seq peaks using a positional weight matrix derived from the top 50 sequences revealed a 1.5 fold enrichment of ETS binding sites over a control genomic shuffled sequence (QUINLAN and HALL 2010) (Figure 6C). Using a window of 55 base-pairs (bp), the maximum distance between ETS sites that supports cooperative *in vitro* binding of the mammalian homologue of Yan, TEL1 (GREEN et al. 2010), we observed a 1.8-fold enrichment in the total number of sequences containing two adjacent motifs and a 2.7-fold enrichment in the number of sequences containing 3 or more adjacent motifs in the experimental versus control
datasets (Figure 6D). In agreement with this finding, MEME, a motif discovery tool (BAILEY and ELKAN 1994), identified a tandem GGAA/T repeat motif from a randomized set of 600 sequences comprised of 100bp around the MACs predicted ChIP-seq summits (47 sites, \( p=1.5 \times 10^{-6} \); Figure 6B).

To identify DNA sequence characteristics specific to HDRs, we separated the top MACS sorted sequences from HDR and isolated Yan-bound peaks. Although the analysis reveals equivalent density of ETS motifs within high-density and isolated peak regions, there is increased clustering of ETS motifs in HDRs, with a 1.3 fold enrichment over the isolated-peak regions (Figure 6E). We speculate that the extent of clustering of ETS binding motifs may help stabilize Yan binding to chromatin in high-density clusters of peaks. However, the complexity of the chromatin occupancy patterns we have observed together with the temporal dynamics across development suggests additional modes of Yan recruitment to DNA are likely required.

Although our meta-analysis of modENCODE ChIP-chip datasets did not reveal likely candidates, recruitment of Yan to multi-kb regions may occur in conjunction with one or more DNA-binding transcription factors. If correct, then motifs for other transcription factors should be enriched in our datasets. Returning to the central motif enrichment analysis on our top 600 MACS sorted sequences, we noted that while ETS motifs have high central enrichment in both the HDR and isolated peak datasets (Figure 6F,G), the site-probability curve is broader in the HDR dataset, a feature that has been interpreted as indicative of cooperative binding to another transcription factor (BAILEY and MACHANICK 2012). Of the transcription factors analyzed in Figure 4A,B that have available motif matrices, namely, Kr, Ttk, Ubx, Bab1, H, En and Hkb, we did not observe significant enrichment in our datasets either by CentriMo or DREME analysis (Figure 6F,G, Table S8). Instead, the binding motif for Mad, which along with Medea, is the primary transcriptional effector of the Dpp signaling pathway (HUDSON et al. 1998; SEKELSKY et al. 1995), was not only enriched in both the Yan HDR and isolated peak datasets, but was also the most centrally enriched motif in the HDR dataset (Figure 6G). This correlation raises the possibility of a link between Mad binding and HDR type Yan recruitment. Supporting this, analysis of the genomic regions tested by transcription assay (Figure 1B) revealed putative MAD binding sites in 13/18 of the reporters tested. Further, clustered MAD and ETS sites were identified in 12/13 of these regions with overlapping sites found in neur, cib and Rapgap1 reporters (Table S7). Unfortunately, although Mad ChIP-assays have been carried out
(MacArthur et al. 2009), the dataset quality is low with few peaks called, preventing us from comparing actual Mad- and Yan-bound regions. However analysis of Medea ChIP data (MacArthur et al. 2009) reveals a 63% overlap of stage 11 Yan bound peaks with Medea bound regions, suggesting further investigation of Yan-Mad/Medea interactions might provide insight into mechanisms of Yan chromatin association.

Discussion

In this study, we have used genome-wide chromatin occupancy analysis to test the hypothesis that Yan’s ability to self-associate through its SAM domain enables it to spread across long distances on chromatin. Mechanistically, the chromatin spreading model predicts that the initial recruitment of Yan monomers, either individually or cooperatively, to clusters of high affinity ETS binding sites would nucleate polymer formation. The resulting increase in local Yan concentration would enable interactions with lower affinity sites in flanking DNA that would not normally be bound by Yan monomers, thereby allowing stable repression complexes to assemble across multiple kilobases of DNA (Courey and Jia 2001; Kim et al. 2002; Kim et al. 2001; Qiao and Bowie 2005; Qiao et al. 2004; Roseman et al. 2001; Song et al. 2005; Tran et al. 2002; Zhang et al. 2010). Such large polymeric complexes could repress transcription both by passively blocking access to large stretches of DNA and/or by actively recruiting associated factors. Consistent with the spreading model, a significant fraction of Yan occupancy occurs as multi-kb clusters of densely packed peaks.

Before considering the implications of this finding further, it is important to note that some of the complexity of Yan occupancy patterns undoubtedly results from distinct enhancers being bound in different cell types. However, both the general conservation of occupancy patterns across development, despite the presence of very different cell types, and the strength of the ChIP signals, suggest that extreme cell-to-cell heterogeneity in Yan binding is unlikely to explain the full profile. While formal confirmation will require cell type-specific ChIP analyses, our interpretation that complex transcription factor chromatin occupancy can occur within a single tissue or cell type and still be detected in a whole-embryo ChIP assay is supported by recent work from the Furlong and Mann labs (Agelopoulos et al. 2012; Bonn et al. 2012). Thus for the purpose of the ensuing discussion, we will assume this interpretation is correct.
Our finding that introduction of the V105R missense mutation into the Yan SAM domain only modestly reduced chromatin occupancy, despite drastically compromising function, leads us to propose that SAM-mediated polymerization is not the primary determinant of multi-kb chromatin spreading but instead contributes to active repression. Although it is formally possible that the modest chromatin occupancy differences detected between wild type and monomeric Yan<sup>V105R</sup> could fully explain the loss of functionality, the essentially identical multi-kb occupancy profiles observed across critical direct target genes such as aos (Figure 4C), argues against this. We also cannot rule out the possibility that Yan<sup>V105R</sup> can polymerize via an unknown, SAM-independent mechanism sufficiently to establish long-range chromatin contacts. However, our prior FRAP studies in Drosophila cells demonstrated that the Yan monomer diffuses significantly faster than wild type Yan (Zhang et al. 2010), consistent with a loss of polymer formation following mutation of the SAM interface.

A number of non-mutually exclusive mechanistic explanations can be envisioned for why Yan monomers appear significantly less functional, despite relatively wild type patterns of chromatin recruitment. One possibility is that although the DNA binding affinity of Yan<sup>V105R</sup> is comparable to that of wild type Yan in vitro (Qiao et al. 2004; Song et al. 2005; Tootle et al. 2003), in vivo, the absence of self-association might destabilize cooperative DNA binding sufficiently to abrogate function. Indeed, in vitro studies of Yan's mammalian homologue Tel1 have demonstrated that cooperative DNA binding allows dimers to bind to linear templates carrying tandem ETS sites more stably than monomers (Green et al. 2010). We also observe a preference for clustered ETS binding sites within Yan bound regions, suggesting that a proportion of Yan chromatin association could reflect cooperative binding to multiple ETS sites. Indeed, previous work has shown the importance of homotypic clustering of binding sites for the cooperative recruitment of transcription factors (Borok et al. 2010; Lebrecht et al. 2005; Segal et al. 2008). Thus although the Yan<sup>V105R</sup> monomers would still be recruited to high affinity sites, loss of self-association mediated cooperativity might destabilize DNA binding. The resulting reduced probability of occupancy could explain the lower but broadly similar ChIP signal obtained with Yan<sup>V105R</sup> relative to wild-type Yan, and the apparent failure to assemble a functional or stable repression complex.

Alternatively, oligomerization might promote or stabilize three-dimensional chromatin conformations that are critical for repression. In this scenario, we predict that abolishing SAM-
mediated self-association would only minimally reduce linear chromatin spreading but significantly impair Yan’s repressive function. Investigation into the correlation between Yan-bound HDRs and three-dimensional chromatin conformation will be required to explore this idea. Very speculatively, if such contacts exist, they might not only influence transcription of a single locus, but could also provide a mechanism for coordinate regulation across multiple genes. The particularly striking high-density Yan occupancy observed across components of multiple signaling pathways, including the Notch, Wingless and EGFR networks, make these appealing contexts for further exploration of such ideas.

Finally, heterotypic interactions with other corepressor proteins, chromatin remodeling factors, or transcription factors could explain both the recruitment and function of chromatin-bound Yan. For example, a protein-protein interaction that was critical to transcriptional repression but either specifically blocked by the V105R mutation or dependent on the presence of Yan oligomers could explain why monomeric Yan<sup>V105R</sup> is less able to repress transcription despite its relatively normal chromatin recruitment. Regardless of the mechanism of Yan’s recruitment to DNA, the similar occupancy profiles and distinct functions of Yan<sup>WT</sup> and Yan<sup>V105R</sup> suggest a predominantly active rather than passive mode of Yan-mediated repression.

What might be the functional significance of extensive Yan occupancy across a locus with respect to regulation of gene expression? While formally there could be none, the fact that such patterns occur predominantly at essential signaling factors or developmental regulators and are highly conserved at these loci in distantly related Drosophila species argues otherwise. The prevailing model of Yan function is that competition with Pnt for binding to clustered ETS motifs provides a binary switch that dictates whether a target gene is repressed or activated (Gabay <i>et al.</i> 1996; Graham <i>et al.</i> 2010; Rohrbaugh <i>et al.</i> 2002; Vivekanand <i>et al.</i> 2004). Consistent with this, we showed that 16/18 Yan-bound regions selected based on high probability of clustered ETS binding sites can be both activated by Pnt and repressed by Yan in cultured cell transcription assays. However genome-wide, the complexity of the Yan occupancy pattern together with the relatively modest enrichment of ETS sites in HDRs versus isolated peaks and the absence of ETS motifs in many high confidence Yan bound sequences suggests that there are other modes of Yan recruitment and function. Arguing for a more complex mode of regulation, the <i>yan</i> loss of function phenotype (Nusslein-Volhard and Wieschaus 1980; Olson <i>et al.</i> 2011; Rogge <i>et al.</i> 1995) is not consistent with broad “all or nothing” regulation of
the developmentally important genes that we identify as putative Yan targets in our ChIP datasets. However, since to date there is no available ChIP data for Pnt, we cannot rule out the possibility that Yan and Pnt act in combination at all targets.

One speculative idea is that broad binding of Yan across loci that encode critical signaling pathway components could provide a tuning mechanism that both prevents premature pathway activation in response to stochastic transcription factor binding or subthreshold signaling and primes the system for rapid response once signaling exceeds a critical threshold. Alternatively, HDRs might not confer direct regulation, but instead could provide staging grounds for preassembly, storage or sequestration of transcriptional complexes that are quickly mobilized to specific CRMs as needed. The fact that other transcriptional repressors, most strikingly Kr, also exhibit prominent HDR-type occupancy, suggests that future investigation into the functional significance of such patterns could provide new insight into conserved mechanisms of gene regulation.

Acknowledgements

We thank the University of Chicago Microarray Facility and Argonne National Laboratory for ChIP-chip hybridization and deep sequencing, respectively. We thank Alec Victorsen and Ralf Kittler for recombineering the Yan constructs, Aaron Mitchell-Dick for assistance in screening for transgenics, and Misha Ludwig, Manu and Rick Fehon for assistance with Eve-YFP quantification. We thank members of the Rebay, Fehon and White labs for helpful discussions, and Jean-Francois Boisclair-Lachance, Trevor Davis, Thomas Graham, Charlene Hoi, and Matthew Slattery for comments on the manuscript. K.P.M., R.W.C. and I.R. are supported by an NIH grant P50 GM081892. Additional funding for I.R. is provided by NIH R01 GM80372. J.Z. is a recipient of a Women's Board Fellowship of the U. of C. J.W is supported by an award from the American Heart Association.

References


KULAKOVSKII, I. V., and V. MAKEEV, 2009 [Integration of data obtained by different experimental methods to determine the motifs in DNA sequences recognized by transcription-regulating factors]. Biofizika 54: 965-974.


The yan gene is highly conserved in Drosophila and its expression suggests a complex role throughout development. Dev Genes Evol **209**: 207-217.

The many faces of SAM. Sci STKE **2005**: re7.

Derepression by depolymerization; structural insights into the regulation of Yan by Mae. Cell **118**: 163-173.


Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. Cell **81**: 857-866.


Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods **4**: 651-657.

The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. J Biol Chem **287**: 8702-8713.

The role of yan in mediating the choice between cell division and differentiation. Development **121**: 3947-3958.


Long-range repression by multiple polycomb group (PcG) proteins targeted by fusion to a defined DNA-binding domain in Drosophila. Genetics **158**: 291-307.

Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science **330**: 1787-1797.

PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol **27**: 66-75.

Genetic dissection of pointed, a Drosophila gene encoding two ETS-related proteins. Genetics **135**: 455-468.

Control of midline glia development in the embryonic Drosophila CNS. Mech Dev **64**: 137-151.


Figure Legends

Figure 1. Yan associates with chromatin at both isolated peaks and within clusters of densely packed peaks. A) ChIP enrichment is abolished in yan\textsuperscript{E833} \textsuperscript{ER433} null mutant embryos relative to wild-type embryos. NC: negative control sites. Error bars are standard deviations of three independent ChIP experiments. \textit{lace, CG3430, CG42390} and \textit{CG31368} are classified as single-peak genes; \textit{eve, argos, neur, jing} and \textit{cv-2} are multiple-peak genes. B) Transcription assays using luciferase reporters carrying Yan binding regions identified in the ChIP-chip experiment. Reporters can be activated by Pnt and repressed by Yan. \textit{CG3430, king-tubby, CG42390, wisp, ia2, CG31368} and \textit{mth} are single-peak genes; \textit{corn, cv-2, Rapgap1, trol, neur, spi, jing, cycB, ro, cib} and \textit{pnt} are multiple-peak genes. Error bars are standard deviations of three independent experiments. C-D) Occupancy patterns of wild-type Yan at the \textit{lace} and \textit{pnt} loci, provide examples of isolated peak and clustered peak binding respectively. ChIP-seq data is shown as smoothed tag density with a scale of number of reads per million. Gene structures are shown below ChIP patterns with gene names above or below the genome coordinates to depict + or – strands respectively.

Figure 2. Yan binding patterns are dynamic during development. A, B) ChIP-chip patterns of \textit{spi} and \textit{argos} provide representative examples of loci that retain their classification as multiple-peak genes and have identical binding profiles at stage 5-7 and stage 11. C) Yan putative targets include genes regulating axial patterning of the early embryo. Asterisks depict loci with temporally dynamic Yan chromatin occupancy patterns. D, E) ChIP-chip patterns of representative loci with changing patterns of Yan chromatin occupancy. Grey bars indicate DNAse accessible chromatin at stage 5 or 11 (Ll et al. 2011).

Figure 3. Conservation of Yan high-density regions in \textit{D. virilis}. A) ChIP-chip and ChIP-seq patterns at the \textit{yan} locus highlight the conservation between \textit{D. melanogaster} and \textit{D. virilis}. B) Defining high-density regions (HDRs) at the \textit{edl} locus. Boxes indicate HDRs at the different percent occupancies. C) Quantification of the prevalence of HDR occupancy across the genome in \textit{D. melanogaster} and \textit{D. virilis}. D) \textit{D. melanogaster} classification of peaks that are conserved
in *D. virilis* E) Yan binds core components of multiple signaling pathways. *genes assigned to Yan-bound peaks in both *D. melanogaster* and *D. virilis*. **peaks close to *csw* just missed our statistical cut-off. F) Cytoscape analysis of interactions within and between the EGFR, Notch and Wg signaling pathways using the *D. melanogaster* stage 11 dataset. Putative targets are colored orange, with orange diamonds representing genes associated with high-density Yan occupancy. In grey are genes not assigned to a Yan-bound peak. Interactions depicted are STRING 9.0 data, derived from experimentally determined protein-protein and genetic interactions as well as computationally predicted interactions.

**Figure 5. Meta-analysis of modENCODE transcription factors reveals high-density type binding is most prominent with Yan.** A) Cumulative frequency plot of transcription factor (TF) bound region lengths. Using a threshold of top 3% (MAT score), bound regions were defined for each transcription factor. Using our algorithm for defining HDRs, the lengths of bound regions occupied at 40% are plotted as a cumulative frequency distribution. Across the genome, Yan (red line) exceeds other transcription factors in the length of its bound regions. Kr, Jumu, H, GATAe and Hkb datasets are from 0-8hr embryo ChIP-chip, Sens data are ChIP-chip data from 4-8hr embryos and all other TF datasets, with the exception of Yan, are from 0-12hr embryo ChIP-chip. B) Percentage of TF-bound peaks that fall within HDRs, defined as regions of increasing length (bp) where TF occupancy ≥40%. C) Meta-analysis of TF chromatin occupancy across the EGFR-Notch-Wingless pathway genes reveals extensive occupancy by Yan, but not by other transcription factors. The maximum length of bound regions at each gene was determined and plotted as a heat map for each transcription factor. Grey indicates no binding.

**Figure 5. Yan self-association is not required for high-density occupancy patterns.** A) Yan^{V105R} has reduced genetic rescue ability relative to Yan^{WT}. The number of embryos that survived to hatching were counted for each genotype: 1) yan^{ER33}/yan^{ER433} 2) yan^{ER33}/yan^{ER433}; Yan^{WT} 3) yan^{ER33}/yan^{ER433}, Yan^{V105R}. N> 240. B) Cuticle preparations of 1) *W*^{118} 2) yan^{ER33}/yan^{ER433} 3) yan^{ER33}/yan^{ER433}, Yan^{V105R} embryos reveal that Yan^{V105R} embryos die with an anterior open phenotype. C) ChIP patterns of Yan^{V105R}, 6xYan and wild-type Yan at the *argos* locus. ChIP-seq data is shown as smoothed tag density with a scale of number of reads per million. Yan^{V105R} and 6xYan have broadly similar patterns to wild-type Yan, with high-density type binding to chromatin within clusters of multiple peaks. D) Quantification of the prevalence of HDR occupancy at defined thresholds within each dataset.
Figure 6. Motif analysis reveals clustering of ETS motifs in Yan-bound regions

A) CentriMo analysis of the top 600 Yan-bound regions reveals central enrichment of the h-ETS1 motif and a motif derived from regions flanking GGAA/T sequences located within 72bp around the top 50 ChIP-seq summits. B) The PWMs of ETS motifs used to perform CentriMo analysis. MEME analysis of ChIP-seq peaks reveals clustering of GGAA/T motifs. C) Scanning sequences bound by Yan at stage 11 (blue) with the PWM derived from the top 50 ChIP-seq sequences revealed enrichment of ETS motifs compared to a control set of sequences (orange). D) Analysis of the gap between ETS motifs in either Yan-bound sequences (blue) or control sequences (orange) revealed an enrichment in clustering of ETS motifs. A cluster is defined as the number of ETS motifs found within 55bp of each other (x-axis). E) Clustering of ETS motifs in peaks associated with HDRs (purple) over isolated peaks (yellow). Y-axis shows percentage of sequences with the number of motifs within 55bp of each other (x-axis). F, G) CentriMo analysis of known transcription factor PWMs using the FlyReg (based on Flyreg Drosophila DNAse I Footprint Database v2.0), idmmpmm2009 and dmmppmm2009 (KULAKOVSKII and MAKEEV 2009; KULAKOVSKII et al. 2009) databases found within the top 600 isolated peaks (F) or top 600 peaks associated with HDRs (G) revealed that with the exception of Mad central enrichment, only ETS motifs were centrally enriched in either dataset.
Figure 1

(A) ChIP-chip and ChIP-seq data for the *lace* gene. The yellow bars represent wild type, and the blue bars represent Yan mutant. The *pnt* region is marked with a red line.

(B) Relative luciferase activity for various genes. The black bars represent the reporter, the orange bars with Pnt P1 treatment, and the blue bars with Pnt P1 and Yan treatment.

(C) Comparison of ChIP-seq and ChIP-chip tag density for the *lace* gene.

(D) Comparison of ChIP-seq and ChIP-chip tag density for the *pnt* gene.
Figure 2

A

ChIP-chip St 5-7
ChIP-chip St 11

B

ChIP-chip St 5-7
ChIP-chip St 11

msb1l

19570k 19580k

argos

CG32158

16460k 16470k 16480k

spi

CG10268

19

0

ChIP-chip St 11

Figure 2
Figure 3

A

ChIP-seq D. melanogaster
ChIP-seq D. virilis
ChIP-chip D. melanogaster

-10 log10 p-value
2160k 2170k 2180k 2190k

B

ChIP-seq D. virilis
ChIP-seq D. melanogaster
ChIP-chip D. melanogaster

-10 log10 p-value
14550k 14560k 14570k

C

% of peaks in high-density region
30 40 50 60

D

% of region occupied by Yan

E

Yan

RTK pathway
EGFR* aop* ago*
bib*

Wingless pathway
EGFR* aop* ago*
bib*

Hedgehog pathway

BMP pathway
dpp brk sog*
cv-Z Med

Hippo pathway

Notch pathway
N* ago* bib*

N* ago* bib*

D. melanogaster
D. virilis

key
high-density
isolated
no binding

F

EGFR

Notch

Wingless
Figure 4

A

B

C
Figure 5

A

% of hatched embryos

yan null | WT | Yan V105R

C

-10log10 p-value

ChIP-seq

ChIP-chip Yan V105R

ChIP-chip 6x Yan

ChIP-chip

CG33158

16460k 16470k 16480k

D

% of peaks in high-density region

W1118 | yan null | yan null; Yan V105R

% of region occupied by Yan

30 40 50 60
Figure 6

A

B

h-ETS1
ETS_exp validated

C

Top600 sequences
Shuffled sequences

D

Top 600 sequences
Shuffled sequences

E

Top 600 Isolated
Top 600 Dense

F

h-ETS1
h-ETV6
ETS_exp validated
Top50_MACSsorted
Ets_homemade
Mad

G

Mad
h-ETS1
h-ETV6
Ets_exp validated
Top50_MACSsorted
Ets_homemade