Interpreting the regulatory genome: the genomics of transcription factor function in *Drosophila melanogaster*

Matthew Slattery, Nicolas Nègre and Kevin P. White

**Abstract**

Researchers have now had access to the fully sequenced *Drosophila melanogaster* genome for over a decade, and the sequenced genomes of 11 additional *Drosophila* species have been available for almost 5 years, with more species' genomes becoming available every year [Adams MD, Celniker SE, Holt RA, et al. The genome sequence of *Drosophila melanogaster*. *Science* 2000;287:2185–95; Clark AG, Eisen MB, Smith DR, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 2007;450:203–18]. Although the best studied of the *D. melanogaster* transcription factors (TFs) were cloned before sequencing of the genome, the availability of sequence data promised to transform our understanding of TFs and gene regulatory networks. Sequenced genomes have allowed researchers to generate tools for high-throughput characterization of gene expression levels, genome-wide TF localization and analyses of evolutionary constraints on DNA elements across multiple species. With an estimated 700 DNA-binding proteins in the *Drosophila* genome, it will be many years before each potential sequence-specific TF is studied in detail, yet the last decade of functional genomics research has already impacted our view of gene regulatory networks and TF DNA recognition.

**Keywords:** *Drosophila; transcription factor; genomics; enhancer; Zelda*

**INTRODUCTION**

Although its myriad genetic tools make *Drosophila melanogaster* one of the preferred model organisms, *Drosophila*’s transcription factors (TFs) are a primary reason this fruit fly is known to scientists of all stripes. *Drosophila* biology and TF biology seemingly go hand-in-hand, for a variety of reasons. From the strikingly organized expression patterns of the TFs that comprise the segmentation network, to the dramatic homeotic phenotypes associated with selector proteins such as Eyeless, Antennapedia and Ultrabithorax, *Drosophila* TFs have long fascinated developmental biologists [3]. The polytene chromosomes of the larval salivary gland have provided a mechanistic glimpse into the interplay of chromosomal dynamics, protein–DNA interactions and gene expression long before the genomics revolution brought such studies to their current state of resolution [4–6]. The polytene studies have been particularly informative in studying transcriptional responses to hormonal signals and heat shock [7]. From cell-type specification to hormone signaling, to stress responses, *Drosophila* has been and will continue to be an important model for TF biology.

With the rapid decrease in sequencing costs and countless new genomes being sequenced every year, a significant challenge in genome science is to identify and characterize functional elements encoded within genomic DNA. Central to this challenge is the identification of DNA elements that are bound...
DNA motifs, or cis-regulatory elements. Regulatory elements for different TFs are often clustered into groups referred to as cis-regulatory modules or enhancers. Together, the elements in an enhancer can drive anything from relatively simple, ubiquitous patterns to exquisitely precise spatiotemporal patterns. Genes expressed in complex patterns throughout development are often regulated via multiple distinct enhancers. In general, enhancers are not constrained by location and can be located nearby or distal to both the target gene and additional enhancers with regulatory input into the target gene [8].

Many insights into enhancer function have come from pioneering studies on the transcriptional regulation of genes in the Drosophila segmentation network. For example, work on the regulation of gap genes by the TF Bicoid (Bcd) provided a model of dose-responsive gene regulation in which cis-regulatory elements with different affinities for Bcd responded differently to the various concentrations of Bcd along the anterior–posterior axis of the embryo [9]. Many of our current models on the combinatorial action of TFs and enhancer function are informed by now classical studies on the ‘stripe 2’ enhancer of even-skipped. The stripe 2 expression pattern is driven by combination of low and high affinity sites for four TFs—Hunchback (Hb), Kruppel (Kr), Giant (Gt) and Bcd—two with activating functions (Hb, Bcd) and two with repressive functions (Kr, Gt) [10]. The activating factors drive expression in the stripe 2 domain, and the repressor proteins prevent expression from extending beyond stripe 2 boundaries. This example, and many others, illustrate the vast potential for combinatorial regulation of gene expression on a single enhancer, potential that is expanded even more considering the combinatorics of multiple enhancers working together to drive gene expression [11–14].

To regulate transcription levels, TFs acting at enhancers must somehow transmit regulatory information to a gene’s core promoter and influence the rate at which RNA polymerase II (Pol II) transcribes the gene. Whereas there are many unanswered questions with regard to enhancer–promoter communication, much is known about the basal TFs acting at the core promoter (see [15] and references therein) For example, Transcription Factor II D (TFIID) is a multi-protein complex consisting of TATA-binding protein (TBP) and more than 10 TBP-associated factors (TAFs) [16]. TBP recognizes TATA box motifs in core promoter regions, and additional TFIID sub-units recognize the initiator (Inr) and downstream promoter element (DPE) motifs [15]. Another basal TF complex, TFIIIB, binds to the TFIIB recognition element (BRE) core promoter motif, but only when TBP is bound to the TATA box [17]. However, although multiple motifs and their binding factors have been characterized, not all promoters contain the same combination of motifs [15]. Thus, as with enhancers, it seems the core RNA Pol II promoter has the potential for combinatorial binding and regulation by basal TFs.

Targeted studies of individual enhancers and promoters have provided tremendous insight into the mechanisms by which TFs recognize their cognate DNA motifs and regulate gene expression. However, the sequenced Drosophila genome and the current era of Drosophila genomics are giving us the ability to test the generality of the principles dictated by the classical enhancer studies. In addition, comprehensive identification of TF target genes and TF-binding events throughout the genome has the potential to identify previously unrecognized interactions, and further our understanding of TF-binding specificity and the combinatorial regulation of gene expression.

**GENE EXPRESSION PROFILING**

Over a decade ago, with the availability of large collections of cDNAs, and eventually the fully sequenced Drosophila genome, researchers began generating expression microarrays as a tool for comprehensive characterization of gene regulatory networks (GRN) [18–20]. Gene expression profiling with microarrays and, more recently, RNA-sequencing (RNA-seq) experiments have been used to profile cell-, tissue- and region-specific gene expression profiles, and to characterize gene expression changes throughout all stages of Drosophila development.
In addition, expression profiling experiments are often used to monitor gene expression changes upon ectopic expression of a TF, upon RNA interference (RNAi)-based depletion of a TF, or in fly lines with TF mutations (described below). All of these approaches have identified downstream targets of TFs, based on either the factor that is genetically manipulated or the factors that are expressed in the tissues being profiled.

The first Drosophila developmental time course microarray study focused on gene expression changes during metamorphosis, and this was followed up with a more comprehensive study of all stages of fly development, from embryogenesis through adulthood [18, 20]. These studies identified gene sets correlated with tissue-specific developmental processes such as muscle development during embryogenesis and programmed cell death of larval muscles at metamorphosis, to sex-specific differences resulting from the germline tissues in male and female adult flies. Clustering of gene expression profiles during development also discovered a range of additional co-regulated gene batteries. These initial whole animal studies only provided a glimpse into the structure of TF-driven GRN. However, they successfully illustrated the potential power of carefully designed expression profiling experiments, coupled with computational analyses, to study TF-regulated genomic networks.

Gene expression profiling experiments subsequently became progressively more complex by testing expression differences across tissues or regions of tissues, and by incorporating genomic epistasis experiments [21–25]. For example, the Hox TF Ubx, which controls development of a small balancing organ called the haltere, has been the subject of multiple gene expression studies. The haltere and wing imaginal discs are serially homologous tissues, the main difference between them being Ubx expression. Ubx is expressed in the haltere and not the wing; when Ubx is lost from the haltere disc, the haltere transforms to wing fate, and when Ubx is misexpressed in the wing disc, the wing transforms to haltere fate [26]. A series of experiments testing expressing differences between the wing (Ubx−) and haltere (Ubx+), as well as experiments monitoring gene expression changes after ectopic expression of Ubx in the wing, have resulted in an increasingly refined view of the Ubx GRN in the haltere tissue [27–29]. Importantly, many of the genes identified as part of the Ubx regulatory network are consistent with a role for Ubx in limiting haltere growth by modifying morphogen signaling pathways [30, 31]. This approach has also been used for characterizing the GRNs for additional TFs. For example, a similar combination of tissue-specific expression profiling, and profiling in various genetic epistasis experiments has been used to define the regulatory network driven by the eye selector TF Eyeless (Ey) [32–35].

The main caveat with gene expression profiling, at least when taking a TF-centric point of view, is that it does not indicate which targets are direct and which targets are secondary to the direct targets. Nor do gene expression studies identify the cis-regulatory modules to which a TF is binding. It is important to note, however, that direct target genes and cis-regulatory modules can be identified and tested using both bioinformatics approaches and traditional enhancer analyses [36, 37]. Biologically validated direct targets of both Ubx and Ey were identified by careful analysis of the gene expression profiling [27, 34]. This approach of expression profiling and bioinformatic identification of putative cis-regulatory sequences has been tremendously successful in Saccharomyces cerevisiae, where the majority of cis-regulatory input for a given gene falls within the ~800 bp upstream of its transcription start site, but has proven more problematic in the complex cis-regulatory environment of Drosophila [38, 39]. For this reason, and to understand TF-binding events that might not lead to detectable gene expression outcomes, it is also necessary to study the genome-wide, in vivo binding patterns of TFs.

METHODS FOR STUDYING TF-BINDING IN VIVO

The most common method for mapping TF-binding sites in vivo is chromatin immunoprecipitation (ChIP). In a typical ChIP experiment, a TF of interest is immunoprecipitated from a formaldehyde-crosslinked chromatin extract using an antibody specific for the TF [40]. Immunoprecipitated DNA can then be monitored by microarray or deep sequencing analysis (ChIP-chip or ChIP-seq, respectively) [41–43]. DNA regions bound by a TF are identified as enriched relative to control DNA, and this provides a genome-wide map of TF-binding sites [44].

A similar approach, termed DamID (DNA adenine methyltransferase identification), uses a fusion protein consisting of a TF and Escherichia coli Dam. The fusion protein methylates adenines in regions of
TF binding [45, 46]. As a result, TF-bound DNA contains a unique methylation tag that can be separated from unmethylated DNA by digestion of genomic DNA with methylation-specific restriction enzymes (DpnI and DpnII). Recently, this technique has been modified to allow for immunoprecipitation of methylated DNA (DamIP) [47]. As with ChIP, DamID and DamIP fragments can be monitored by microarray or sequencing, and enrichment relative to background Dam methylation profiles is used to identify TF-binding sites [44, 46, 48].

**CORE RNA POL II PROMOTER**

One of the first ChIP-chip studies that interrogated the entire *Drosophila* genome had a significant impact on our view of transcriptional regulation at the core promoter of genes transcribed by Pol II [49]. The traditional view of gene regulation is that TFs activate gene expression by binding regulatory DNA elements, followed by recruitment of general TFs and Pol II [50]. Contrary to this ‘recruitment’ model, Zeitlinger et al. found significant Pol II enrichment near the transcriptional start sites of ~10% of *Drosophila* genes, often inactive genes that are highly responsive to environmental signals or genes destined to be expressed at later developmental stages. These findings are consistent with a model of transcriptional activation by release of paused Pol II originally put forth to explain the rapid and robust induction of heat shock genes in *Drosophila* [51–54]. Thus, a genome-wide investigation of Pol II localization suggested that a transcriptional activation mechanism first identified on a small number of genes might apply to a significant fraction of the fly genome. The Pol II pause and release mechanism does not apply to the majority of genes, however, so one must ask how certain TFs activate paused Pol II, and whether this mechanism uses the same machinery as used in recruitment models.

At the same time that bench scientists were generating data that would revise our views on TF activation by Pol II recruitment to the core promoter, bioinformatics studies on these regions were unveiling a variety of promoter DNA motifs. In total, 10 promoter motif elements, including the previously characterized TATA box, Inr and DPE motifs, were discovered in the initial genome-wide analysis of all *Drosophila* core promoters [55]. Further analysis of these motifs revealed that the three canonical motifs (TATA, Inr, DPE) often co-occur with a newly identified, and functionally validated, motif termed the MTE (motif ten element) at promoters with well-defined sites of transcription initiation [56]. On the other hand, broad promoters with multiple initiation sites, often associated with ‘housekeeping’ genes, lack these motifs. Computational analysis of promoters associated with paused Pol II identified an additional motif, the pause button (PB), which is often associated with Inr and DPE motifs at these promoters [57]. These poised, Inr + PB/DPE promoters also contain the GAGA motif and, indeed, GAGA factor is often associated with paused Pol II, though the mechanistic role of GAGA in Pol II pause and release remains unclear [58]. These studies illustrate, as has long been stressed by those studying transcription initiation, that the core promoter is much less generic than it is often perceived to be; understanding the core promoter on a genome-wide scale will be necessary for understanding mechanisms of Pol II recruitment, Pol II pausing and enhancer–promoter communication [15, 59–61].

**GENOME-WIDE BINDING OF TFs**

One of the more striking findings from the onslaught of ChIP-chip and ChIP-seq studies has been the seemingly widespread binding of many TFs. Although there were many opinions regarding the number of genome-wide binding events before the advent of ChIP-chip, it was not uncommon for people to view TFs as likely to have on the order of a few hundreds of direct target genes. However, ChIP analysis of dozens of *Drosophila* TFs has demonstrated that the number of genome-wide binding events is often on the order of 1000 to 10 000 [62–65]. It is important to note that the extensive genome-wide binding appeared widespread based on our ideas about the structure of GRNs, but only a fraction of a given TF’s target DNA motifs are actually bound throughout the genome (discussed below). This widespread binding phenomenon has also been observed in human ChIP studies [66, 67].

The widespread binding phenomenon was apparent in early genome-wide studies of single transcription factors, but another pattern emerged when researchers began to address multiple transcription factors in the same study [68, 69]. When looking at the patterns of multiple TFs it became apparent that binding was not only widespread, but highly overlapping [69]. This initial observation was based
on TF binding in Drosophila Kc cells, and a similar observation was made for patterning TFs in the blastoderm embryo [63]. The pattern held true when looking at a much broader panel of TFs as well; these regions of TF co-localization have been termed HOT (high-occupancy target) regions, or TF-binding ‘hotspots’ [64, 65]. In general, hotspots are found in proximal promoter regions of the genome and the DNA at these regions is highly accessible and subject to high nucleosome turnover. Whether the high TF occupancy is a cause of, or results from the low nucleosome occupancy at hotspots is yet to be determined.

Many questions remain on the issue of widespread and overlapping TF binding. First, can any of these binding signals be attributed to experimental artifacts due to crosslinking (or some other step in the ChIP procedure)? Or are some of the binding events measured by ChIP indirect? The fact that these hotspots of TF localization are also obvious in DamID experiments (no formaldehyde crosslinking) suggest that crosslinking per se is not the cause of the binding signal, but still leaves open the possibility that a significant fraction of called binding events are indirect. For a number of TFs, known target DNA-binding motifs are preferentially enriched in non-HOT regions bound by the factor, suggesting that binding at hotspots may be indirect, although this is not the case for all TFs [64]. Interestingly, pairwise comparisons of the binding profiles for a panel of TFs show many more distinct patterns of TF–TF overlap when HOT regions are removed from the comparison (Figure 1); a number of the interactions revealed upon subtraction of HOT region binding are consistent with existing literature (e.g. Tinman and Twist, Engrailed and Groucho) [65].

Ultimately, it seems that TF binding outside of HOT regions is more consistent with our traditional views of TF function, whereas HOT region binding might represent a previously unrecognized, possibly nonspecific, property of TFs.

A recent functional analysis of 108 HOT regions found that >90% of these putative regulatory modules control highly patterned, rather than ubiquitous, expression [70]. These cell-specific expression patterns generally tracked with neighboring genes, indicating that HOT regions often act as developmentally regulated enhancers. Interestingly, the cell-specific expression patterns directed by HOT enhancers were not correlated with expression patterns of the TFs bound to a given HOT region. Thus, although the patterned expression indicates that a fraction of TF–DNA interactions at HOT enhancers are true regulatory interactions, it seems that many TF–HOT region interactions are likely to be neutral, driving neither expression nor repression of the target gene [70]. Considering that the HOT regions tested were based on whole embryo ChIP data, which represents a spatial average of binding across many cell types, this lack of correlation between the TFs targeting HOT regions and the expression patterns driven by HOT regions also suggests that HOT regions are accessible to different TFs depending on the cell type. That is, the overlapping TF binding at HOT regions is the result of different TFs occupying the DNA in distinct cell

![Figure 1: Pairwise comparison of TF binding site overlap, with (left) or without (right) TF hotspots included. The stronger intensities in the matrix indicate a higher Z-score enrichment of TF binding site overlap. See [65] for details.](image-url)
types (Figure 2A), rather than all TFs occupying the same region in the same nucleus. How regulatory specificity is achieved within this landscape of neutral binding is an important question that is a direct result of genome-wide studies of TF function.

**DNA-BINDING SPECIFICITY AT TRANSCRIPTION FACTOR HOTSPOTS**

The large number of TF-binding events revealed by genome-wide ChIP studies is seemingly at odds with the notion of TF-DNA-binding specificity. However, careful analysis of TF-DNA-binding properties based on both *in vivo* and *in vitro* data suggests that the number of binding events revealed by ChIP studies is exactly what is expected for the *Drosophila* genome [71]. Of note, it was shown that metazoan TFs on average bind lower information content DNA motifs (fewer bases, more degenerate) than TFs of the unicellular eukaryote *S. cerevisiae*, and bind significantly lower information content motifs than prokaryotes. For this reason, even when taking the chromatin landscape of the genome into account, *Drosophila* TFs are predicted to have on the order of 1000–10,000 binding events throughout the genome, consisting of both regulatory and spurious binding events [71]. This is still in excess of the expected number of target genes for a given TF, however, so the mechanisms by which TFs achieve regulatory specificity within these thousands of binding events is not very clear. One possibility is that the degenerate binding motifs of *Drosophila* (and other metazoan) DNA-binding domains (DBDs) allow for greater combinatorial interactions amongst TFs, and it is these interactions that distinguish regulatory from nonregulatory binding events.

Is there any evidence for TF specificity within hotspots of TF co-localization? As described above, it appears that many, but not all, TFs bind hotspots in a relatively nonspecific fashion, at least based on the absence of expected DNA motifs [64]. Binding could also be indirect, driven by protein–protein interactions (PPIs), or the result of high local protein

![Figure 2: (A) Model for TF binding at HOT regions. The same theoretical HOT region is occupied by distinct TFs depending on cellular context (cells in head region, marked “blue”, versus cells in the trunk region, marked “red”). DNA targeting can be direct or indirect, and only a subset of these interactions directs gene expression. GAF and Zld likely bind HOT regions across multiple cell types. (B) Proposed model for Zld as a pioneer factor. Zld binding to a DNA region at the maternal-to-zygotic transition maintains DNA in a highly accessible state, whereas the rest of the genome becomes increasingly chromatinized. Zld-bound regions overlap significantly with TF hotspots.](image-url)
concentrations in proposed transcriptional regulatory ‘factories’ within the nucleus [72, 73]. This idea is supported by the study that first described TF hotspots [69]. In this study, the TF Bcd was found to localize to hotspots along with a variety of other transcriptional regulators. Interestingly, hotspot binding by Bcd was not dependent on direct DNA binding, as a mutant version of Bcd lacking its DBD still localized to hotspots [69]. This illustrates that for some TFs, hotspot binding may be driven primarily by interactions beyond direct protein–DNA interactions. Nevertheless, the modENCODE project identified multiple DNA motifs overrepresented in TF hotspots, suggesting that specific DNA-binding proteins could be establishing or maintain hotspots [64].

Two DNA motifs, the TAGteam and GAGA motifs, are highly overrepresented in HOT region [64, 70, 74]. Both motifs are bound by factors involved in the formation of a chromatin environment that is permissive to TF binding. The GAGA motif is targeted by GAGA factor (GAF, also known as Trithorax-like/Trl), a multifunctional DNA-binding protein that has been implicated in many aspects of transcriptional regulation, including nucleosome displacement and the formation of open chromatin [75–77]. The TAGteam element (CAGGTA) was first described as a DNA motif associated with genes activated at the maternal-to-zygotic (MTZ) transition; the zinc-finger protein Zelda (Zld) was identified as the TF that binds the TAGteam motif, and Zld was found to be an integral regulator of early zygotic transcription [78, 79]. The Drosophila HOT regions are based on embryonic TF-binding patterns, so the fact that Zld appears to be a key regulator of transcription at the earliest stages of embryogenesis raised the intriguing possibility that Zld establishes the formation of hotspots through its interaction with the CAGGTA sequence. A series of recent studies support this hypothesis. Indeed, there is significant overlap between Zld binding in vivo and TF hotspots [80, 81]. In fact, Zld is in a unique position to influence chromatin structure in the developing embryo in that it appears in zygotic nuclei before even Bcd and Dorsal, and binds to the majority of its target motifs genome-wide at cycle 8, a time when the genome is proposed to be relatively accessible overall, though the chromatin structure at this stage has yet to be experimentally tested [81, 82]. If the chromatin of the early MTZ genome is found to be more accessible than at later stages, this places Zld in a position to act as a ‘pioneer’ factor, of sorts. However, rather than opening closed chromatin in the fashion of canonical pioneer factors, Zld might maintain certain regions in a highly accessible state while the rest of the genome becomes increasingly chromatinized (Figure 2B) [82].

**TF SPECIFICITY**

Although many new concepts have emerged by monitoring a broad range of TFs, focused studies of individual TFs (or small groups of interacting TFs) have also informed our view of TF biology in Drosophila. One of the most highly studied groups of factors in the Drosophila genomics era is the TFs regulating mesoderm development, including mesoderm-specific regulators such as Twist (Twi), Tinman (Tin) and Mef2, as well as a number of TFs that play roles in both mesoderm and non-mesoderm tissues [83]. These studies have added significant depth to the specifics of mesoderm GRNs, and they have also provided more general models for TF regulatory function and specificity. For example, characterizing enhancers based on overlapping ChIP signals revealed a distinct lack of specific motif grammar, suggesting flexibility in spacing and significant role for tethering of certain TFs to enhancers via PPIs [84, 85]. Individual motifs are important for direct binding in this TF collective model, because binding appears to be direct for a subset of the factors, but TFs can also occupy functional enhancers in the absence of direct DNA binding. In this collective model, it is the presence of the specific group of TFs at an enhancer, whether direct or indirect, that determines regulatory output [85].

Twi, the basic helix–loop–helix TF sitting atop the mesoderm GRN referenced above, has also been studied at the genome-wide DNA targeting level in the context of additional Drosophila species. By comparing Twi binding patterns in six species, it was shown that overall binding is conserved across Drosophila species, and clustered Twi peaks (<5kb between peaks) are very highly conserved, relative to isolated peaks [86]. The strong conservation of clustered peaks suggests that distributed regulatory binding may be especially important for gene regulation by Twi, and additional TFs [12, 87, 88]. Despite high conservation overall, species-specific differences in binding do exist. Although most species-specific changes in Twi occupancy could be explained by the loss of a DNA-binding motif or
changes in the quality of a motif, a significant fraction of lost Twi peaks did not correlate with a lost Twi motif. In these cases, loss of Twi binding was associated with the loss of a potential regulatory partner (e.g., Snail, Dorsal, etc.) [86]. This finding highlights both the importance of combinatorial TF binding, and the utility of comparative genomic analyses in studying combinatorial TF relationships.

The comparison of TF-binding events across species described for Twi provide a nice template for what is becoming the next frontier in *Drosophila* TF genomics: comparison of TF-binding patterns across distinct cellular contexts. This can be carried out by comparing binding patterns across species, but also by comparison of TF binding across developmental stages, across cell or tissue types, etc. An interesting example of such a study comes from an investigation into the Clock/Cycle (Clk/Cyc)-mediated of circadian rhythm GRN. Activity of the Clk/Cyc transcriptional activators is repressed by Period (Per) and Timeless (Tim). These proteins are part of a feedback loop in which Clk/Cyc heterodimers activate transcription of *per* and *tim*, and Per/Tim in turn represses Clk/Cyc; as Per and Tim decrease (degradation, lack of transcription), Clk/Cyc activity increases again [89]. A genome-wide look at the binding patterns of these TFs revealed that, on the temporal side of regulation, hundreds of genes are directly targeted in a periodic fashion by Clk/Cyc, with Per binding at Clk/Cyc locations, followed by a decrease in transcription and then the loss of Clk/Cyc binding [90, 91]. These data, along with biochemical analysis of Clk’s interaction Per/Tim, illustrate the importance of PPIs in mediating the cyclical binding of a TF complex to DNA [90].

The above Clk/Cyc data are based on ChIP signal from adult *Drosophila* heads, and thus represent a spatial average of binding events across the many tissue types in the adult head. When the authors investigated Clk binding in the heads of flies lacking eye tissue, they found that whereas 20% of binding is unchanged relative to wild-type adult heads, ~40% of putative direct Clk target genes are no longer bound, and ~40% of Clk target genes are still bound but the binding profile is weaker or certain binding peaks had disappeared [91]. These data suggest that a significant fraction of the Clk-binding signal from whole *Drosophila* heads is tissue-specific, with 80% of putative direct Clk target genes subject to eye-specific modification of Clk regulatory input. The modification of Clk-DNA-binding patterns across tissue types is likely the result tissue-specific PPIs, tissue-specific changes in chromatin state, or both.

Similar tissue- and developmental stage-specific binding events have been identified for the Hox TF Ubx [92–95]. Significant stage-specific differences in Ubx target genes have also been identified by gene expression profiling experiments [96]. As with Clk, it is likely that a combination of tissue-specific differences in chromatin state and protein interaction partners is responsible for context-specific Ubx binding. Consistent with a role for protein interaction partners, it has recently been shown that the Hox proteins, which all bind a core TAAT motif as monomers, achieve differential DNA-binding specificity through heterodimerization with the homeodomain TF Extradenticle (Exd) [97]. In this case, Hox TF interaction with Exd reveals emergent DNA recognition properties that are not evident in the monomeric binding properties of Hox TFs. Similar stories are emerging in other organisms as well, suggesting that PPI-mediated alterations in DNA-binding specificity will likely represent a significant mechanism by which TFs regulate cell- or tissue-specific gene sets [98].

**PERSPECTIVE**

The studies described above illustrate both the significant advances in our understanding of TF biology that have been the result of genomic studies, and the many avenues that are only beginning to be explored. A significant issue that remains, especially in light of the thousands of genome-wide binding events for many TFs, is how functional specificity is achieved. Understanding combinatorial interactions between TFs will assuredly be important for understanding this issue, but so will additional measures of ‘functional’ binding. ChIP experiments, expression profiling data, combined with in vitro approaches and traditional enhancer bashing are needed to address the various models of combinatorial interaction and importance of motif grammar at enhancers [11, 13, 14, 85, 97–99].

The next frontier of *Drosophila* TF biology is that of differential network biology [100]. TF-binding profiles compared across multiple cell types, tissues and environmental conditions will begin to address many of the questions associated with DNA-binding specificity. Furthermore, the increasing number of sequenced *Drosophila* species, and even individual strains within *D. melanogaster* populations, now...
provide a comparative framework for population genetic studies related to TF function. The tools for these comparisons—between tissues, between populations, between species—are now within reach of most Drosophila researchers, and the next generation of TF genomics is upon us.

**Key Points**

- Drosophila TFs generally bind thousands of genomic regions, in excess of expected number of targets based on gene expression profiling.
- A significant fraction of TF binding occurs in hotspots of TF co-localization; hotspots are associated with open chromatin and direct cell-specific gene expression patterns.
- The TF Zfd plays a role in maintaining specific regions of open chromatin at the earliest stages of zygotic transcription and, along with GAF, may play a role in establishing hotspots.
- Recent comparative studies of TF-binding patterns in different tissues, developmental stages or Drosophila species suggest a significant role for PPIs in mediating TF regulatory specificity.

**FUNDING**

This work has been supported by an NIH grants U01HG004264 and P50GM081892 awarded to KPW.

**References**


