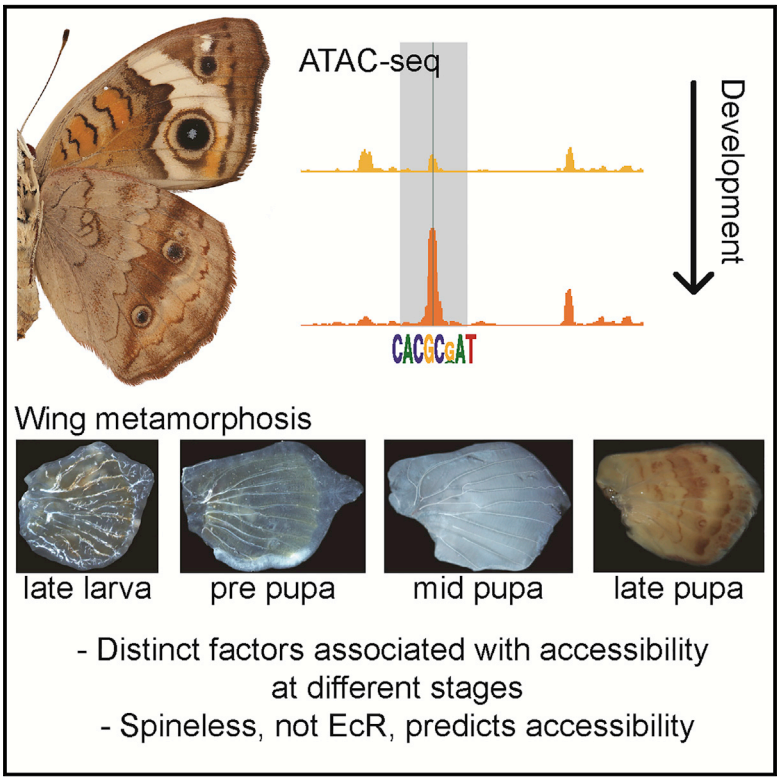
Cell Reports

Article

Contrasting Roles of Transcription Factors Spineless and EcR in the Highly Dynamic Chromatin Landscape of Butterfly Wing Metamorphosis

Graphical Abstract



Highlights

- Chromatin accessibility during butterfly wing development is highly dynamic
- Different sets of transcription factors associate with accessibility at different stages
- ChIP-seq indicates spineless binding predicts changes in accessibility
- Assembly and annotation of the *Junonia coenia* genome are provided

Authors

Karin R.L. van der Burg, James J. Lewis, Arnaud Martin, H. Frederik Nijhout, Charles G. Danko, Robert D. Reed

Correspondence

krv32@cornell.edu

In Brief

Using ATAC-seq, van der Burg et al. show that chromatin organization during butterfly wing development is highly dynamic. Motif enrichment analysis followed by ChIP-seq shows that DNA binding of transcription factor spineless, but not EcR, is strongly associated with opening chromatin.

van der Burg et al., 2019, Cell Reports *27*, 1027–1038
April 23, 2019
https://doi.org/10.1016/j.celrep.2019.03.092

CellPress

Cell Reports
Article

Contrasting Roles of Transcription Factors Spineless and EcR in the Highly Dynamic Chromatin Landscape of Butterfly Wing Metamorphosis

Karin R.L. van der Burg,[1,5,*] James J. Lewis,[1,2] Arnaud Martin,[3] H. Frederik Nijhout,[4] Charles G. Danko,[2] and Robert D. Reed[1]
[1]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA
[2]Baker Institute for Animal Health, Cornell University, Ithaca, NY 14853, USA
[3]Department of Biological Sciences, The George Washington University, Washington, DC 20052, USA
[4]Department of Biology, Duke University, Durham, NC 27708, USA
[5]Lead Contact
*Correspondence: krv32@cornell.edu
https://doi.org/10.1016/j.celrep.2019.03.092

SUMMARY

Development requires highly coordinated changes in chromatin accessibility in order for proper gene regulation to occur. Here, we identify factors associated with major, discrete changes in chromatin accessibility during butterfly wing metamorphosis. By combining mRNA sequencing (mRNA-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq), and machine learning analysis of motifs, we show that distinct sets of transcription factors are predictive of chromatin opening at different developmental stages. Our data suggest an important role for nuclear hormone receptors early in metamorphosis, whereas PAS-domain transcription factors are strongly associated with later chromatin opening. Chromatin immunoprecipitation sequencing (ChIP-seq) validation of select candidate factors showed spineless binding to be a major predictor of opening chromatin. Surprisingly, binding of ecdysone receptor (EcR), a candidate accessibility factor in *Drosophila*, was not predictive of opening but instead marked persistent sites. This work characterizes the chromatin dynamics of insect wing metamorphosis, identifies candidate chromatin remodeling factors in insects, and presents a genome assembly of the model butterfly *Junonia coenia.*

INTRODUCTION

Throughout tissue development and cell differentiation, constant changes in transcriptional activity are required. These changes in gene expression over developmental time require that different regions of the genome become accessible during development to ensure proper gene regulation (Li, 2002). Indeed, several recent studies show that chromatin structure is dynamic over developmental time (Daugherty et al., 2017; Thomas et al., 2011; Uyehara et al., 2017). Recent studies have shown that chromatin accessibility is initiated by low-level binding of pioneer factors across the genome. These weak-occupancy sites then attract cell-specific transcription factors (TFs), resulting in a removal of nucleosome barriers and thus an increase in chromatin accessibility (Donaghey et al., 2018; Mayran et al., 2018). While we know that these changes in accessibility are caused by complex interactions of multiple TFs, the identities of the suites of TFs that drive changes in accessibility are less well understood.

During insect metamorphosis, many changes in developmental gene regulation happen over a very short period of time, and it is therefore likely that changes in chromatin accessibility are occurring as well. Until recently, identifying TFs involved with changes in chromatin state had proven difficult; however, several data-driven case studies were recently published that made progress in identifying TFs associated with accessibility changes during development of complex tissues (Daugherty et al., 2017; Lamparter et al., 2017; Uyehara et al., 2017). For example, recent work on *Drosophila melanogaster* wing development showed that the ecdysone-induced TF E93 directly regulates chromatin accessibility. For this work Uyehara et al. (2017) assayed chromatin state changes over time in developing wings and identified sequence motifs associated with opening chromatin sites. Chromatin immunoprecipitation sequencing (ChIP-seq) of TFs corresponding to predicted binding motifs confirmed an association between these proteins and chromatin accessibility. A similar method was also used in *C. elegans* (Daugherty et al., 2017), in which authors identified several drivers of chromatin accessibility, including some previously known ones, thus further validating the value of this motif analysis approach. Since chromatin accessibility and gene network activity can be dynamic over time, we wondered whether the population of TFs driving accessibility would change over longer developmental sequences as well.

The butterfly *Junonia coenia* is an excellent system to study change in chromatin accessibility during development. Developing wings are relatively large and easy to dissect, making it possible to assay both chromatin state and mRNA levels in single wings from single individuals. Furthermore, during metamorphosis, butterfly wings undergo a predictable series of changes in gene expression, physiology, and chromatin landscape over

time as well (Connahs et al., 2016; Iwata et al., 2014; Lewis and Reed, 2019; Lewis et al., 2016; Zhang et al., 2017). Here we studied chromatin accessibility changes during *J. coenia* wing metamorphosis. We identified chromatin state changes over time and asked what sets of TFs are associated with changes in chromatin accessibility. We characterized sequence motifs associated with actively opening peaks across three developmental transitions and correlated motif enrichments with changes in mRNA levels between developmental stages, consistent with a scenario where expression of a TF is initiated and subsequently binds to weakly occupied sites and thus assists in increasing chromatin accessibility. We found that chromatin accessibility is highly dynamic over time and that distinct sets of TFs are associated with chromatin opening during each developmental transition. We present additional ChIP-seq validation for two candidate TFs—spineless and ecdysone receptor (EcR)—and show that binding of spineless, but not EcR, is individually predictive of onset of accessibility. In addition to these findings, we present a high-quality Pacific Biosciences (PacBio)-sequenced *J. coenia* genome assembly combined with a high-resolution annotation of wing development regulatory elements, setting the stage for more detailed chromatin level studies in this emerging model species.

RESULTS

J. coenia Genome Sequencing and Assembly

To assemble the *J. coenia* genome, we used PacBio RS II sequencing data from 58 single molecule, real-time (SMRT) cells representing 124X coverage, with a subread N50 size of 9.4 kb. After repeat masking, reads under 7,954 bp were removed for the initial FALCON pipeline error correction (https://github.com/PacificBiosciences/FALCON-integrate/tree/v0.4.2). Error correction resulted in 4.4 million reads, with an N50 of 6.9 kb, for 37X coverage of the genome. These reads were then processed by the overlap portion of the FALCON pipeline using a length cutoff of 5 kb. These were assembled into 1,139 primary contigs with a total assembly size was 585 Mb, with an N50 contig length of 1.57 Mb. Expected genome size was 525 Mb based on flowcell cytometry estimates. To annotate the genome, mRNA sequencing (mRNA-seq) data from multiple stages during egg, larval, and pupal development were assembled into a reference transcriptome containing 19,738 genes. These data were used in conjunction with protein annotations from *Melitaea cinxia* (Ahola et al., 2014) for annotation purposes. After three iterations of MAKER (Campbell et al., 2014), we annotated 19,234 genes. Benchmarking Universal Single-Copy Orthologs (BUSCO) scores (a measure of completeness [Simão et al., 2015]) show 98.2% (1,628 out of 1,658) of complete assembly of benchmark insect genes, of which 15.6% were duplicated. 0.5% of benchmark genes were fragmented, and 1.3% were missing. The genome is available on lepbase.org.

Chromatin Accessibility Is Highly Dynamic during Wing Metamorphosis

We collected three assay for transposase-accessible chromatin using sequencing (ATAC-seq) replicates from both forewing and hindwing tissue at four developmental stages: during the last day of the 5th instar (last) molting stage of the caterpillar (5th), 2–3 h before pupation; at the pre-pupa (pp) stage; 72 h after pupation (72h); and 168 h after pupation, when ommochrome pigments appear (d7) (Figure 1A). These developmental time points were chosen to represent distinct stages of wing development. At the 5th instar stage, many pre-pattern genes are expressed, marking it as an interesting time point for wing pattern development (Reed and Serfas, 2004). Around 24 h later, at the pp stage, wings undergo explosive growth associated with pupation. About 4 days later, at the 72h stage, endocrine signaling is at its peak (Rountree and Nijhout, 1995), making this an important stage for investigating endocrine effects on wing color. Lastly, 3 days later, at the d7 stage, scale structures are maturing and actual pigments are being made, marking this as an important stage for work on wing coloration. To estimate library complexity, we calculated PCR bottleneck coefficients according to ENCODE standards (ENCODE Project Consortium, 2012). Values ranged from 0.81 to 0.86, for all stages except pp, where the values ranged from 0.71 to 0.78 (Table S2). This suggested only mild PCR bottlenecking, moderate for pp stages. To avoid biasing our results, PCR duplicates were removed from the analysis. Pearson correlations of read depth under 1 kb bins showed that replicates were highly reproducible, with values of 0.95 (±0.014) and 0.93 (±0.026) in forewings and hindwings, respectively (Figure S1A). The fragment of reads in peaks (FRiP) scores range from 0.46 to 0.68, indicating minimal background noise in our samples. Lastly, reads were enriched in transcription start sites (Figure S1D), as is expected with ATAC-seq results. When we directly compared forewing and hindwing data, we found the two datasets to be highly similar, with a Pearson correlation of 0.97 (Figures S1B and S1C). Interestingly, in a direct differential peak height comparison between forewings and hindwings, there were only two significantly different ATAC-seq peaks, both within 50 kb of *Ubx*. This finding is consistent with ATAC-seq results in *Heliconius erato* (Lewis and Reed, 2019), where forewings and hindwings also only differed around *Ubx*. Thus, we view the nearly identical results for forewings and hindwings as powerful biological cross-validation of our ATAC-seq results. Between 271,775 and 362,834 accessible sites were called for each developmental stage and merged into 500,558 peaks for forewings and 488,675 peaks for hindwings across development. These results show that our dataset is of high quality, with high replication and low background noise.

Despite the similarity between forewings and hindwings, we decided to do all downstream analyses on forewings and hindwings separately. This effectively meant we would have independent biological validation for our findings, increasing confidence in the results. We compared chromatin accessibility between all pairs of succeeding stages for both forewings and hindwings for a total of six developmental transitions and identified three different types of sites: "closing," where accessibility peaks show significantly weaker signal in the later than the earlier stage; "opening," where peaks show stronger signal in the later stage; and "nc," non-changing sites that were not significantly different between stages (Figure 1B). Although the majority of peaks did not change significantly between subsequent stages, we found a large number of accessible sites that became either more or less accessible between stages (Figures 1B and 1C).

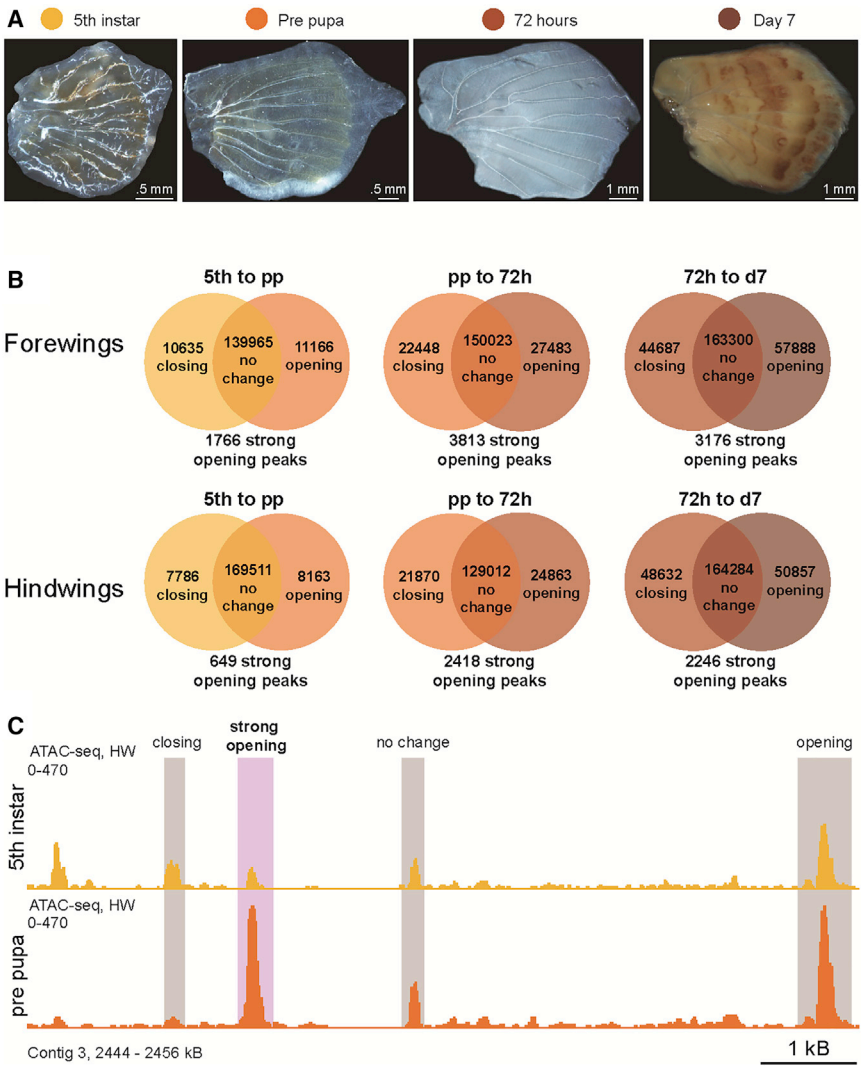


Figure 1. Chromatin Accessibility Is Highly Dynamic across Developmental Time

(A) We collected ATAC-seq and RNA-seq at four developmental stages, from both forewings (not shown) and hindwings.

(B) Chromatin accessibility changes over time, where a subset of peaks is downregulated between two stages ("closing") and upregulated ("opening"), and the majority of peaks do not significantly change. A subset of opening peaks is designated as strong opening on the basis of low height in the first stage and a >2.5 log fold change in the second.

(C) ATAC-seq screenshot containing all possible transitions.

For additional quality metrics, see Figure S1.

Pearson correlations between developmental stages show distinct patterns for each developmental stage (Figure S1A), where the 5th instar is more similar to both the pp and 72h stage and d7 is more different from other developmental stages. Finding extensive chromatin accessibility differences between developmental stages is consistent with previous studies done in *D. melanogaster* that show extensive dynamicity of chromatin organization over developmental time in whole embryos (Thomas et al., 2011) and pupal wings (Uyehara et al., 2017).

The goal of this study was to identify TFs associated with chromatin accessibility changes over development, so we focused our analysis on chromatin peaks absent at an early stage, which then showed a major increase in accessibility at a subsequent stage. These peaks were classified as "strongly opening" (SO), designated by especially stringent selection criteria that required a high positive log fold change (>2.5) and a low read count (bottom 50th percentile) in the early stage (Figures 1B and 1C; Table S3). Between 649 and 3,813 sites were designated as SO sites per developmental transition (Table S3), or 0.1%–0.7% of all peaks, respectively. Since we were interested in motifs enriched in SO sites versus the overall genome, we also identified a negative peak set of the same length of similar GC content while avoiding repeat regions as a negative control.

We next wanted to identify candidate TFs binding in these SO sites using motif analysis. We used LS-GKM, a support vector machine learning algorithm software package (Ghandi et al., 2014; Lee, 2016), to identify motifs enriched in 80% of SO sites versus negative sites for each of our developmental transitions. All possible 10-mers were scored, and the 100 10-mers that best described the difference between SO and negative control sites were clustered together by sequence similarity into 8 to 52 clusters per stage transition, with 1 to 50 10-mers per cluster, and used for further analysis (Figure 2A; Table S3). We found that in general, motifs identified for both forewings and hindwings during the 72h to d7 transition were more dissimilar than for earlier transitions, possibly due to a more diverse cell population later in development. This explains the increased number of clusters for these stages. To validate our results, we used the 20% of our SO sites not used in our motif identification analysis. For each developmental transition, the clusters from our model showed a significant enrichment in these remaining SO peaks (Fisher's exact test, p values range from 0.0462 to 5.09E−26; Figure 2B; Table S3). For each cluster with 2 or more 10-mers, 87 in total, we produced a weight matrix representing a motif, which we compared to an existing database of *D. melanogaster* motifs (Figures 2C and S2) (fly factor survey [Zhu et al., 2011]). We used the three most similar motifs for subsequent analyses if a TF homolog was found in our *J. coenia* gene set. If multiple motifs existed for a TF, we used the motif with the highest ratio difference between SO and negative control sites. Since many of our motifs were similar to each other and resulted in the same *D. melanogaster* TF motif, we used each fly factor survey motif only once. Thus, we ended up with 65 motifs total (Figure S2).

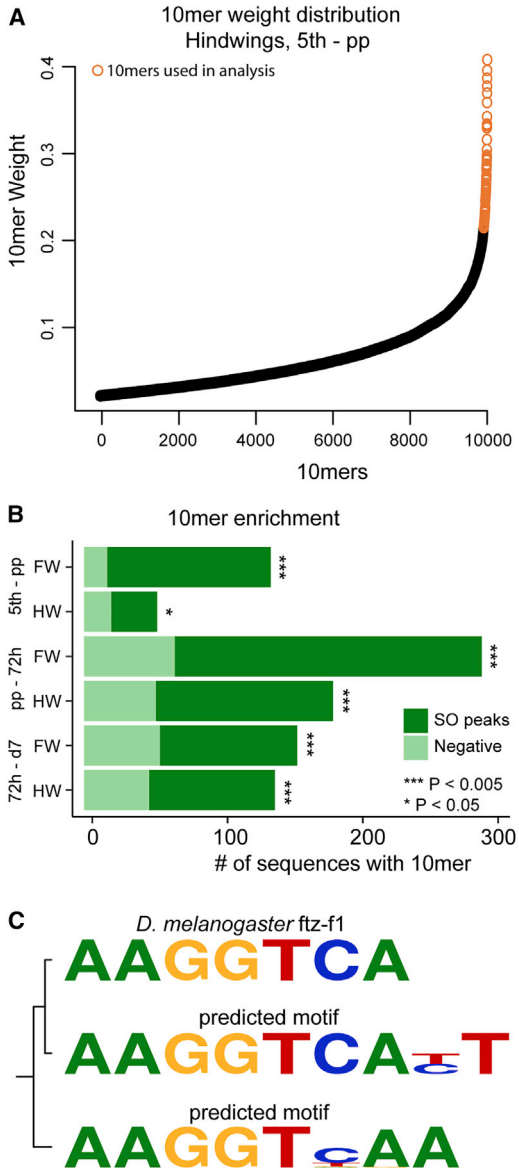


Figure 2. Top Motifs Are Significantly Enriched in Opening Sites and Correspond to Known *D. melanogaster* motifs

(A) We used the top 100 10-mers that best explained the difference between strong opening and negative (random sites of same length and GC content) peaks for subsequent analysis, as identified by the support vector machine model used in the analysis.

(B) Clusters of the top 10-mers are significantly enriched in test sequences (Fisher's exact, $p < 0.05$).

(C) An example of how each motif found in a cluster was compared to known *D. melanogaster* TF motifs.

For all motif comparisons, see Figure S2.

Distinct Stage-Specific Sets of TFs Associated with Chromatin Remodeling

If a TF is directly causing changes in chromatin state, we might also expect an increase in its mRNA levels concordant with an increase in SO sites containing the TF's binding motif. This is consistent with a scenario where a TF is expressed in certain cell types and subsequently increases chromatin accessibility of salient regulatory sites across the genome. To find such an association would also help rule out alternate candidate TFs that have similar binding motifs. Therefore, using the motifs we previously identified, we looked at the correlation between an increase in mRNA levels and the corresponding motif ratios between SO and negative sites (Figures 3 and S3). TFs whose binding motifs showed a large enrichment in SO peaks, along with a large change in mRNA levels between stages, were identified as top candidates for driving accessibility changes. We found that in both forewings and hindwings, nuclear hormone receptor TFs showed a high correlation between motif ratio and an increase in mRNA levels between the two larval stages (Figures 3A and S3A). Interestingly, between the pp and 72h stage, the hormone receptor signal is completely absent and is replaced by TFs with a basic-helix-loop-helix (bHLH) domain, associated with protein dimer formation, and PAS domains, used for interactions with chemical stimuli (i.e., environmental contaminants or heat shock proteins [Gu et al., 2000; Henry and Crosson, 2011]). In both forewings and hindwings, the *spineless-tango* complex showed a particularly strong association with opening chromatin (Figures 3B and S3B). Between the two pupal stages, TFs with a bHLH complex showed a strong association in both forewings and hindwings. In forewings alone, the nuclear hormone receptor gene *fushi tarazu-f1* (*ftz-f1*) showed the strongest association between 72h and d7 (Figures 3C and S3C). It was highly encouraging that both forewings and hindwings show similar patterns, since the analysis for both sets of wings was done independently. Thus, we have identified several candidate TFs driving accessibility during butterfly wing development and the motifs to which they were putatively bound.

Many Sites Show Strong Patterns of Oscillation in Accessibility

We next wanted to know whether specific TFs were associated with persistence in accessibility, so we investigated to what extent SO sites stayed open across multiple stages. Surprisingly, we found that the majority of SO peaks lost their accessibility in subsequent developmental stages. In hindwings, of the 649 SO peaks between 5th and pp, 62% (405 peaks) were closed at 72h, and of the 2,275 SO peaks between pp and 72h, 92% (2,096 peaks) closed by d7 (Figures 4A and S4A). We wondered whether the same TFs that were associated with an increase in accessibility were also associated with a decrease in accessibility—indicated by a decrease in expression levels. We found this to be the case for some genes (*ftz-f1*, *ultraspriracle* (*usp*), *tango* (*tgo*), *Blimp-1*) but not others (*Hormone receptor 4* (*Hr4*), *Ecdysone receptor* (*EcR*), *spineless, cycle*) (Figure S4B). These patterns could indicate a role for some TFs in maintaining accessibility; however, further work is required to test this idea.

Our data on closing of SO peaks demonstrated a high degree of chromatin dynamicity, so we decided to look closer to see if some of these peaks may reopen at a later time point. We were again surprised to find that a large number of SO peaks indeed showed a pattern where they open, close, and then reopen between development time points (Figure 4). A further surprise was that sets of peaks identified at different developmental transitions appear to show opposing patterns of

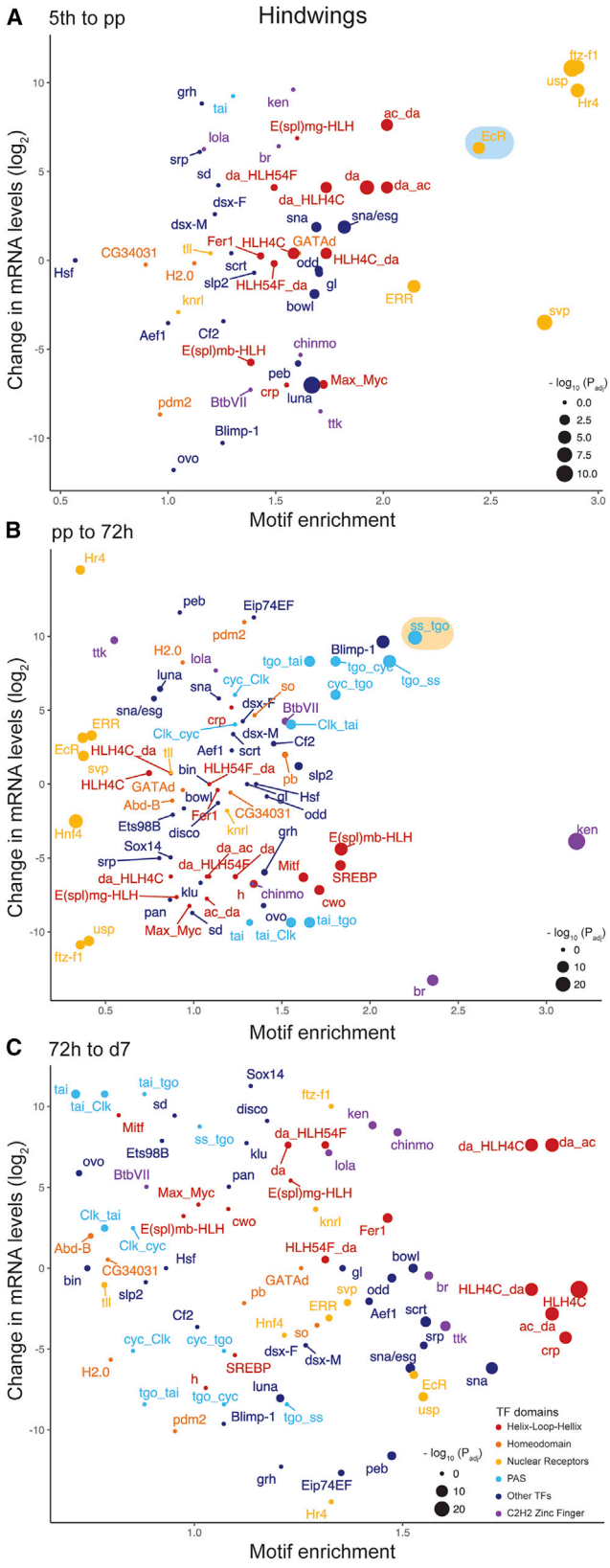


Figure 3. Correlation of mRNA Levels with Strong Opening Peaks Identifies Candidate Chromatin Accessibility Factors

(A–C) Change in TF mRNA levels between developmental stages is plotted against motif ratios in strong opening peaks between stages versus negative peaks for three developmental transitions: 5th to pp (A), pp to 72h (B), and 72h to d7 (C). Point size indicates adjusted p values of enrichment significance as determined by a Fisher's exact test. Data points are color coded by domain presence. TFs used for ChIP-seq follow-up are highlighted. For forewing data, see Figure S3.

oscillation in accessibility dynamics (Figure 4A). We were especially interested in the set of elements that showed distinct accessibility peaks at both the pp stage and d7 stage and wondered whether this pattern may be correlated with hormone signaling (Jindra et al., 1996; Rountree and Nijhout, 1995). We therefore asked if any sequence motifs, including hormone receptors, were able to predict this subset of twice-opening peaks; however, we could not identify any such motifs. Contrary to expectations, in both hindwings and forewings, for both the larval transition from 5th to pp and the pupal transition from 72h to d7, we found that for SO peaks that were *not* double opening, nuclear hormone receptors were enriched (Figures 4C and S4). Since *ftz-f1* showed a strong association with chromatin accessibility, as well as a two-time increase in expression levels similar to the two-time increase in accessibility, we wondered whether these peaks were enriched for the ftz-f1 motif specifically. However, again we found that the early 5th-pp transition and the late 72h-d7 SO peaks that were *not* double opening were enriched for the ftz-f1 motif. In sum, these results confirm that chromatin accessibility is indeed highly dynamic throughout development, that the binding of candidate remodeling TFs is not by itself predictive of persistent accessibility, and that a sizable subset of sites show a strong pattern of oscillating accessibility between stages.

Spineless Binding Is Strongly Predictive of Actively Opening Chromatin

We were intrigued by the strong association of *spineless* with opening chromatin, so we looked at this factor in more detail. spineless is a TF that plays a role in appendage and photoreceptor specification in *D. melanogaster* (Duncan et al., 1998; Wernet et al., 2006) and was also recently shown to play a role in wing pigmentation and photoreceptor determination in the butterfly *Vanessa cardui* (Perry et al., 2016). These known roles of spineless in developmental gene regulation and determining cell fate could possibly be explained by a molecular mechanism where spineless drives changes in accessibility. Thus, we wanted to see whether spineless binding was indeed enriched in SO sites, which would be further evidence for such a role. We performed spineless ChIP-seq in 72h pupal hindwings and found a Pearson correlation of 0.97 between replicates, with a set of 8,686 binding sites shared between biological replicates (Figures S5A and S5B; Table S4). We saw a strong signal of spineless binding in SO peaks that contained the known *D. melanogaster* spineless binding motif (Figures 5A–5C), where nearly half of SO sites with a spineless motif overlapped with a spineless ChIP peak call (150 out 317 peaks; Figure 5D). Furthermore, a broader analysis of all opening peaks (not just SO peaks) with spineless motifs showed that a third of these sites were also

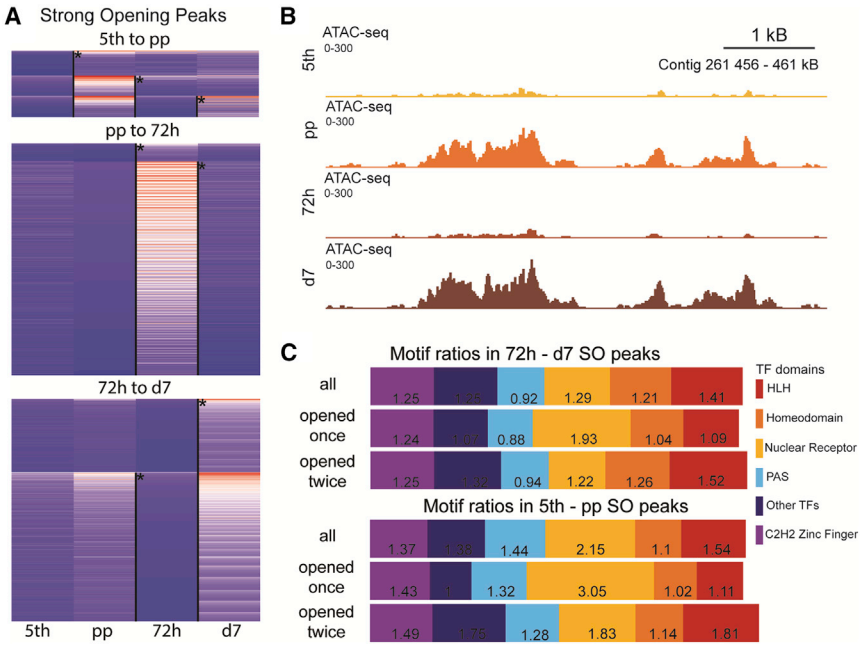


Figure 4. Strong Opening Peaks Oscillate in Accessibility over Time

(A) Heatmap showing reads per kilobase of transcript, per million mapped reads (RPKM) normalized read count in hindwing SO peaks over developmental stages, ranging from 0 (blue) to 300 (red). Asterisks indicate a significant transition up or down in the subsequent stage (Wald test, adjusted p value < 0.05). For each stage, almost all peaks show a decrease in accessibility in the next developmental stage, and for 5th to pp and 72h to d7, a subset increases in accessibility twice.

(B) Example screenshot of 5th to pp SO site, increasing again in d7.

(C) Motif ratios of TF domain groups for all 5th to pp and 72h to d7 peaks, SO peaks that do not oscillate, and SO peaks showing a double increase in accessibility.

For forewing data, see Figure S4.

bound by spineless, indicating that spineless is strongly correlated with chromatin remodeling (1,618 out of 4,961, or 33%; Figures 5D and S5C). Further visual inspection of many opening peaks centered on spineless motifs showed clear enrichment of spineless occupancy even though peaks were not statistically significant due to our conservative peak calling (Figure S5D), suggesting that the effect of spineless activity on chromatin state might be even greater than we detected statistically with our current data. Using multiple expectation maximization for motif elicitation for chromatin immunoprecipitation (MEME-ChIP), we determined the motif was enriched in spineless ChIP peaks and found it to be nearly identical to the motif independently identified by our machine learning model, as well as the experimentally determined *D. melanogaster* ss-tgo complex motif (fly factor survey [Zhu et al., 2011]) (Figure 5).

We wondered whether our results could be explained by a cofactor—a different TF could drive accessibility and spineless could subsequently bind to these sites. An obvious candidate would be *Blimp-1*, a known pioneer factor gene identified in our comparison (Figure 3B) that also showed a similar association with accessibility (Minnich et al., 2016). To test this, we determined Blimp-1 motif abundance in spineless occupancy sites. We found that only 28 out of 150 SO peaks with a spineless ChIP peak call had a Blimp-1 motif, suggesting that Blimp-1 and spineless do not consistently act as cofactors. We then used Find Individual Motif Occurrences (FIMO) to test for enrichment of all 65 previously identified motifs. We found only 3 TF binding motifs somewhat enriched: SPREB (73 versus 21 out of 150 SO peaks versus negative sites), hairy (69 versus 25), and Mitf (61 versus 29). However, all of these genes showed a decrease in gene expression from pp to 72h, suggesting they are not involved in driving accessibility. Furthermore, annotation of these three motifs often overlapped with annotation of spineless motifs. We thus conclude that correlated enrichment of these three motifs along with the spineless-bound SO sites is likely an artifact due to motif similarity, rather than an actual indication of cofactor activity. In sum, all of our analyses identify *spineless* as a prime candidate gene for driving chromatin accessibility during wing metamorphosis.

EcR Binding Marks Stable, Persistently Accessible Sites during Metamorphosis

Hormone signaling plays an important role in driving changes in developmental programs over time (Shlyueva et al., 2014; Uyehara et al., 2017). In our analysis, nuclear hormone receptors showed a strong association with opening chromatin, in both 5th-pp and 72h-d7 transitions. This is congruent with two ecdysone-signaling pulses that occur right before each transition, during the late 5th instar stage and at 72h (Bollenbacher et al., 1975; Rountree and Nijhout, 1995; Zitnan and Adams, 2000), and also with previous studies that have shown ecdysone signaling altering chromatin state in *D. melanogaster* (Shlyueva et al., 2014; Uyehara et al., 2017). There are several hormone receptors associated with ecdysone signaling that target very similar binding motifs, however, and it remains unclear whether all hormone receptors, or only a few, are associated with chromatin remodeling. Although the nuclear hormone receptors *ftz-f1*, *usp*, and *Hr4* showed a stronger association with chromatin accessibility (i.e., a higher correlation between gene expression and motif enrichment), we undertook EcR ChIP-seq thanks to the availability of a cross-reactive antibody and because the heterodimer of EcR and usp combined comprise the complete ecdysone receptor (Yao et al., 1993). We made EcR ChIP libraries at three developmental stages—pp, 72h, and d7—and found read depth Pearson correlations of 0.79, 0.88, and 0.82, respectively, indicating strong consistency between replicates (Figures S6A and S6B; Table S4). We called 5,107 EcR peaks in pp; 3,973 in 72h; and 3,951 in d6. Of these, the majority of peaks, 3,270 in total, overlapped in all stages, indicating relatively little change in EcR binding sites throughout development. Furthermore, we found no evidence of EcR occupancy enrichment in SO peaks

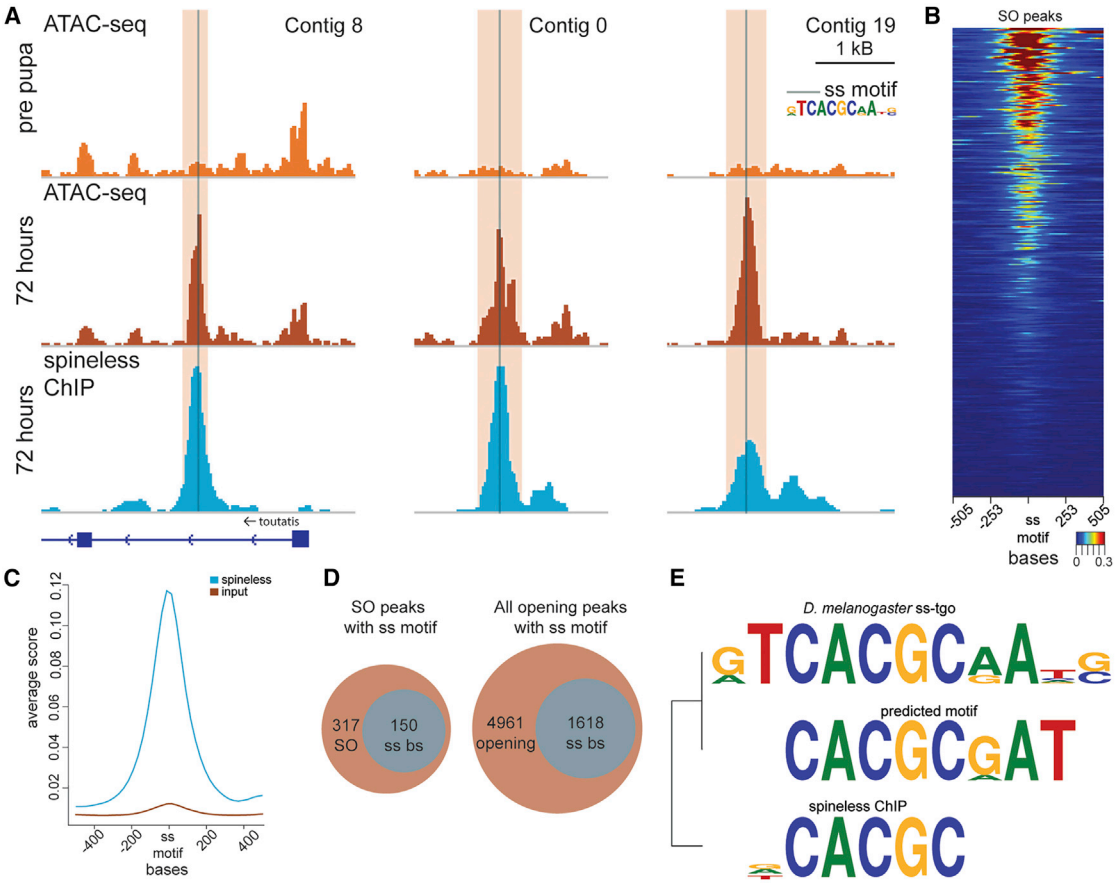


Figure 5. spineless Is Associated with Opening Chromatin

(A) ATAC-seq and spineless ChIP-seq tracks. Strong opening peaks between pp and 72h pupae are highlighted, as is the location of the spineless motif. spineless binding occurs in all three instances.

(B) Metaplot showing spineless TF binding at all pp to 72h SO sites with a spineless motif.

(C) Average read pileup of spineless ChIP-seq versus input around spineless motif.

(D) Number of SO sites and opening sites that have a spineless motif and overlap with a spineless peak call.

(E) Comparison of the *D. melanogaster* spineless-tango motif, our predicted motif, and the ChIP-seq inferred motif as determined by MEME-ChIP.

For additional quality metrics, see Figure S5.

that contained a *D. melanogaster* EcR motif. Using motif finding software MEME-ChIP (Machanick and Bailey, 2011), we determined which motif was enriched in EcR ChIP peaks at different developmental stages and found that the actual motif was very different from the known *D. melanogaster* EcR motif (Figure 6A). This finding was surprising, since many binding motifs are highly conserved between species (Nitta et al., 2015). We compared EcR amino acid sequences from both *J. coenia* and *D. melanogaster* and found relatively few amino acid changes (8 out of 90) within the EcR binding domain but many across the entire protein (only 52% similarity). Since the entire protein structure can affect binding site specificity (Nakagawa et al., 2013; Slattery et al., 2014), we hypothesize that these extensive sequence differences have contributed to evolution of the EcR binding motif between species. Interestingly, our ChIP-annotated EcR motif was significantly enriched in SO peaks in the pp to 72h transition (log odds ratio 2.39, p = 9.768E−15, Fisher's exact test) but not in 5th-pp or 72h-d7 SO peaks (log odds ratio 0.40 and 0.81, and p = 0.023 and 0.15, respectively; Figure 6B). However, there was no overlap between SO peaks and EcR ChIP peaks, nor was there any low-level binding of EcR in our SO peaks (Figures 6C and SD). Taken together, these results suggest that EcR binding by itself is not predictive of changes in chromatin accessibility. On the contrary, EcR tends to mark accessible sites that open early in development and are persistent and minimally dynamic through metamorphosis, perhaps indicative of maintaining chromatin accessibility.

DISCUSSION

Chromatin Accessibility Is Remarkably Dynamic during Butterfly Wing Metamorphosis

The goal of this study was to characterize the chromatin landscape of butterfly wing development through metamorphosis and pupal development and to identify factors associated with chromatin remodeling at specific stages of development. While

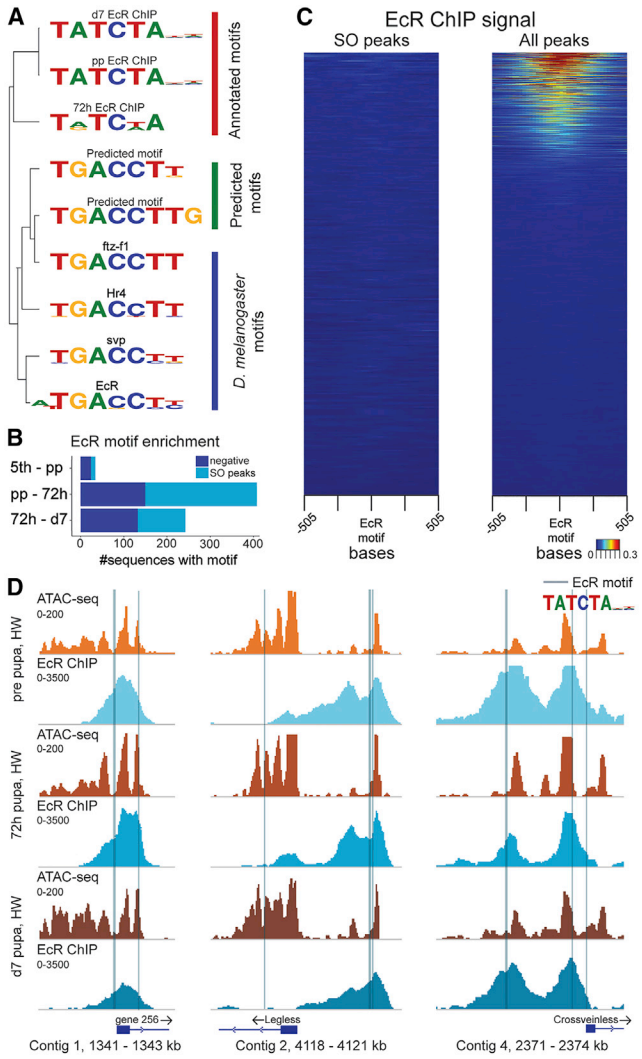


Figure 6. EcR Occupancy Is Not Associated with Opening Chromatin

(A) Tree depicting motif similarity between our predicted motifs, *D. melanogaster* nuclear hormone motifs, and our MEME-ChIP annotated EcR motif.

(B) Ratio of EcR motif occurrences in SO opening sites versus negative sites.

(C) Metaplot showing EcR ChIP reads centered around annotated EcR motif at 72 h, in both SO peaks and all ATAC peaks with an EcR motif.

(D) EcR ChIP signal overlaps with ATAC-seq peaks; both signals show very little change over developmental stages. Darker colors depict later developmental stages.

For additional quality metrics, see Figure S6.

we indeed found that the chromatin landscape was highly dynamic during wing development, one pattern we found to be of particular note was the distinct oscillation of accessibility in a sizable subset of peaks (Figures 4 and S4). Across the four stages we sampled, a large number of peaks showed either an open-closed-open-closed pattern or a closed-open-closed-open pattern, in opposition to each other. This observation of so many loci opening and closing repeatedly over a relatively short time span in a single discrete tissue suggests that it might be common for regulatory elements to be reused in different contexts throughout development. There is evidence that enhancers can be pleiotropic, meaning they activate expression in multiple tissues (Fish et al., 2017), and that they can be active across multiple developmental stages as well (Lewis et al., 2016; Stergachis et al., 2013). However, we are unaware of previous examples of accessibility data showing such dramatic oscillation over time. One may speculate that this oscillation could be driven by endocrine signaling, since many hormones, including ecdysteroids, occur in distinct pulses over time (Jindra et al., 1996; Rountree and Nijhout, 1995), and we suspect that hormone signals may be involved in chromatin remodeling. Interestingly, however, we found that known nuclear hormone receptor motifs were most enriched in sites that did not show double pulses of accessibility. Therefore, hormonal modulation of chromatin is, by itself, unlikely sufficient to explain our observations. In the future this system could prove to be a useful model for exploring how accessibility can be rapidly and strongly modulated during tissue development.

Spineless Is Strongly Associated with Initiation of Chromatin Accessibility

Recent studies have shown that chromatin accessibility is induced by genome-wide low-level binding of pioneer factors. At a subset of these sites, subsequent binding of one or more cell-type-specific non-pioneer TFs promotes stronger binding of these pioneer factors, resulting in cell-specific enhancers and thus a cell-specific developmental program (Donaghey et al., 2018; Mayran et al., 2018). Our results strongly suggest a role for spineless in promoting chromatin accessibility. Between the larval-pupal transition, (pp to 72h), we observed a high enrichment of the ss-tgo binding motif, combined with a large increase in mRNA levels for both proteins, which together provide evidence for a role of the ss-tgo complex in accessibility initiation. Additionally, our ChIP-seq data showed that spineless occupancy is strongly associated with opening chromatin, thus providing additional support for a role for spineless in the induction of accessibility. The question that follows is whether spineless increases accessibility directly as a pioneer factor or as a non-pioneer co-factor—future work must be done to assess this more rigorously.

spineless is a member of a family of TFs that contain both a PAS domain and a bHLH domain (Gu et al., 2000; Huang et al., 1993). The PAS domain functions as a surface for other molecules to bind to, such as other proteins with a PAS domain (like tango), heat shock factors, and in the case of the mammalian *spineless* homolog *AHR*, small-molecule ligands. This binding capacity is why the PAS domain is also called signal-sensing domain, as upon dimerization the complex can activate a downstream response pathway. Indeed, several members of the PAS domain family are involved in the response pathway to circadian signaling, hypoxia, or toxins (Gu et al., 2000). In *Drosophila*, apart from being involved in eye development, spineless has been to shown to be involved in the development of distal appendage structures (Duncan et al., 1998; Emmons et al., 2007). Recently, it was shown that in two other butterfly species, *V. cardui* and *Papilio xuthus*, stochastic *spineless* expression in developing eye cells underlies retinal cell fate determination, affecting color vision in butterflies (Perry et al., 2016). Importantly, *spineless* deletion by CRISPR/Cas9 also led to a complete loss of pigmentation in scale cells. Together, these previous results show many

different roles for spineless in insect development, even though pioneer cofactor activity of this TF has not been previously suggested. Interestingly, loss of AHR in human Tr1 cells leads to some loss of chromatin accessibility, although the effect is minor (Karwacz et al., 2017). Across many cell lines, AHRs and TFs of the same subfamily show very little effect of pioneer activity (Lamparter et al., 2017). We thus hypothesize that any role spineless may play in chromatin remodeling may be particular to invertebrates or arthropods.

The Role of Hormone Signaling in Chromatin Accessibility

Between the 5th instar and the pp stage, we noted strong associations between hormone receptors and chromatin accessibility. Several nuclear hormone receptor motifs showed a strong enrichment in SO peaks combined with an increase in mRNA levels (Figure 3). Our results are consistent with several recent studies examining the role of ecdysone in chromatin accessibility. A recent study found that during *D. melanogaster* development, several ecdysone-induced TFs showed an increase in gene expression, as well as enriched motif abundance in opening formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq peaks (Uyehara et al., 2017). The authors confirmed the role of ecdysone-induced protein 93F, a DNA-binding protein, in regulating chromatin accessibility. Another study showed many changes in enhancer activity in response to ecdysone treatment in *D. melanogaster* cell lines and showed some evidence that this ecdysone response was predominantly reflected in closed chromatin (Shlyueva et al., 2014). Contrary to our findings, this study showed an enrichment of EcR binding in opening chromatin, specifically in a third of the elements. Our data suggested that EcR binding was largely associated with non-dynamic, persistently accessible sites, and there was no overlap of EcR binding with actively opening SO peaks. Interestingly, a recent study showed the opposite—EcR binding is temporally dynamic in *D. melanogaster* developing wings (Uyehara and McKay, 2019). Taken together, these studies all support the general model that hormone signaling is likely driving chromatin remodeling during insect wing development, although probably not through EcR binding in the case of butterflies. Our data instead point toward other nuclear hormone receptor candidates such as ftz-f1 and Hr4. Both proteins are known factors in the ecdysone signaling pathway (Cruz et al., 2008; Ou et al., 2016). It is potentially significant to find that EcR binding does not show an association with ecdysone titer dynamics, which leads us to speculate that EcR deployment may be independent of ecdysone levels during wing metamorphosis. Possibly, EcR binds to DNA early in development but only induces gene expression upon ecdysone signaling. Another, non-mutually exclusive explanation of our findings could be that in the absence of ecdysone signal, EcR tends to bind to DNA and acts as a repressor (Dobens et al., 1991; Kreher et al., 2017). Taken together, these results show that further studies teasing apart the role of ecdysone signaling, EcR activity, and chromatin remodeling are warranted.

Evolution of TF-Chromatin Interactions

Another interesting result from our EcR ChIP-seq experiment is the evolution of motif recognition sites. The *D. melanogaster* EcR motif as determined by a bacterial one hybrid approach (Zhu et al., 2011) is very different from the *J. coenia* motif as determined by MEME-ChIP (Machanick and Bailey, 2011). Although the experimental bacterial one-hybrid (B1H) approach is generally considered the more reliable method for determining recognition sites, if the EcR motifs for both species were similar, our ChIP data should have reflected this. The most likely explanation of our result is evolution of TF binding motifs, a phenomenon described before in eukaryotes (Bourque et al., 2008; Nakagawa et al., 2013), where similar TFs could have very different motifs in widely diverged species, even though DNA binding domains were similar. Since *J. coenia* and *D. melanogaster* diverged 250–300 million years ago (Wiegmann et al., 2009), our observation of a change in EcR recognition sites is perhaps not surprising. It is entirely possible that other TF binding motifs have evolved as well. In fact, although many motifs were very similar to known *D. melanogaster* motifs, we found several motifs with high enrichment in SO peaks for which we could not determine a corresponding TF, since there were no similar motifs in the fly factor survey database.

In addition to TF-motif interactions being possibly lineage specific, our data also indicate that chromatin-remodeling roles of TFs could be lineage specific as well. As mentioned, *spineless* driving chromatin accessibility has not been shown in other species, and there are several known pioneer TFs that do not show a signal of inducing chromatin accessibility in our data. For example, GATAd is a known pioneer factor in several organisms (Corces et al., 2016; Daugherty et al., 2017) but does not show any sign of pioneering activity in developing wings. Another notable absentee TF is the GAGA factor, a well-known pioneer factor for which there is no known homolog in Lepidoptera (Fuda et al., 2015). Taken together, these results suggest that activity of TFs likely changes throughout evolutionary time, and further work examining chromatin accessibility changes outside of lab model organisms is warranted.

Conclusions

In this study, we show that the process of insect wing metamorphosis is associated with an extremely dynamic chromatin landscape over developmental time. We also observed that distinct, stage-specific families of TFs are associated with opening chromatin and that these sets of TFs are highly consistent between both forewings and hindwings. Focused ChIP-seq analyses of the accessibility-associated factor spineless supports a function for this protein promoting chromatin accessibility. Interestingly, however, ChIP-seq analysis of EcR binding shows that this factor marks persistently accessible sites and there was no overlap between EcR binding and actively opening chromatin at the stages we looked at. In addition to providing the first characterization of the chromatin dynamics of butterfly wing metamorphosis, this work also provides the first genome assembly of the model butterfly *J. coenia* including genome-wide annotation of wing development regulatory elements.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.03.092.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for their helpful and constructive comments on the manuscript. We thank Nipam Patel for supporting the *J. coenia* genome assembly, Mike Perry for the spineless antibody, Developmental Studies Hybridoma Bank for the EcR antibody, and members of the Reed Lab for comments on the manuscript. The work was supported by U.S. National Science Foundation grants IOS-1557341, IOS-1557443, IOS-1656514, IOS-1656553, and DEB-1546049 and U.S. Department of Agriculture grant NYC-183542.

AUTHOR CONTRIBUTIONS

Conceptualization, K.R.L.V., J.J.L., R.D.R., and C.G.D.; Methodology, K.R.L.V., J.J.L., R.D.R., and C.G.D.; Formal Analysis, K.R.L.V.; Investigation, K.R.L.V.; Writing – Original Draft, K.R.L.V. and R.D.R.; Writing – Review & Editing, K.R.L.V., J.J.L., A.M., H.F.N., C.G.D., and R.D.R.; Funding Acquisition, A.M., H.F.N., and R.D.R.; Resources, C.G.D. and R.D.R.; Supervision, K.R.L.V. and R.D.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 12, 2018
Revised: February 22, 2019
Accepted: March 25, 2019
Published: April 23, 2019

REFERENCES

Ahola, V., Lehtonen, R., Somervuo, P., Salmela, L., Koskinen, P., Rastas, P., Välimäki, N., Paulin, L., Kvist, J., Wahlberg, N., et al. (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nat. Commun. *5*, 4737.

Akalin, A., Franke, V., Vlahoviček, K., Mason, C.E., and Schübeler, D. (2015). Genomation: a toolkit to summarize, annotate and visualize genomic intervals. Bioinformatics *31*, 1127–1129.

Bollenbacher, W.E., Vedeckis, W.V., Gilbert, L.I., and O'Connor, J.D. (1975). Ecdysone titers and prothoracic gland activity during the larval-pupal development of *Manduca sexta*. Dev. Biol. *44*, 46–53.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.-L., Ruan, Y., Wei, C.-L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.

Boyle, A.P., Guinney, J., Crawford, G.E., and Furey, T.S. (2008). F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics *24*, 2537–2538.

Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. Curr. Protoc. Bioinformatics *48*, 4.11.1–4.11.39.

Challis, R.J., Kumar, S., Dasmahapatra, K.K.K., Jiggins, C.D., and Blaxter, M. (2016). Lepbase: the Lepidopteran genome database. bioRxiv. https://doi.org/10.1101/056994.

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods *13*, 1050–1054.

Connahs, H., Rhen, T., and Simmons, R.B. (2016). Transcriptome analysis of the painted lady butterfly, *Vanessa cardui* during wing color pattern development. BMC Genomics *17*, 270.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203.

Cruz, J., Nieva, C., Mané-Padrós, D., Martín, D., and Bellés, X. (2008). Nuclear receptor BgFTZ-F1 regulates molting and the timing of ecdysteroid production during nymphal development in the hemimetabolous insect *Blattella germanica*. Dev. Dyn. *237*, 3179–3191.

Daugherty, A.C., Yeo, R.W., Buenrostro, J.D., Greenleaf, W.J., Kundaje, A., and Brunet, A. (2017). Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. Genome Res. *27*, 2096–2107.

Dobens, L., Rudolph, K., and Berger, E.M. (1991). Ecdysterone regulatory elements function as both transcriptional activators and repressors. Mol. Cell. Biol. *11*, 1846–1853.

Donaghey, J., Thakurela, S., Charlton, J., Chen, J.S., Smith, Z.D., Gu, H., Pop, R., Clement, K., Stamenova, E.K., Karnik, R., et al. (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nat. Genet. *50*, 250–258.

Duncan, D.M., Burgess, E.A., and Duncan, I. (1998). Control of distal antennal identity and tarsal development in *Drosophila* by spineless-aristapedia, a homolog of the mammalian dioxin receptor. Genes Dev. *12*, 1290–1303.

Emmons, R.B., Duncan, D., and Duncan, I. (2007). Regulation of the *Drosophila* distal antennal determinant spineless. Dev. Biol. *302*, 412–426.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Fish, A., Chen, L., and Capra, J.A. (2017). Gene regulatory enhancers with evolutionarily conserved activity are more pleiotropic than those with species-specific activity. Genome Biol. Evol. *9*, 2615–2625.

Frith, M.C., Saunders, N.F.W., Kobe, B., and Bailey, T.L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput. Biol. *4*, e1000071.

Fuda, N.J., Guertin, M.J., Sharma, S., Danko, C.G., Martins, A.L., Siepel, A., and Lis, J.T. (2015). GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. PLoS Genet. *11*, e1005108.

Ghandi, M., Lee, D., Mohammad-Noori, M., and Beer, M.A. (2014). Enhanced regulatory sequence prediction using gapped k-mer features. PLoS Comput. Biol. *10*, e1003711.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Gu, Y.-Z., Hogenesch, J.B., and Bradfield, C.A. (2000). The PAS superfamily: sensors of environmental and developmental signals. Annu. Rev. Pharmacol. Toxicol. *40*, 519–561.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. *8*, R24.

Henry, J.T., and Crosson, S. (2011). Ligand-binding PAS domains in a genomic, cellular, and structural context. Annu. Rev. Microbiol. *65*, 261–286.

Huang, Z.J., Edery, I., and Rosbash, M. (1993). PAS is a dimerization domain common to *Drosophila* period and several transcription factors. Nature *364*, 259–262.

Iwata, M., Ohno, Y., and Otaki, J.M. (2014). Real-time *in vivo* imaging of butterfly wing development: revealing the cellular dynamics of the pupal wing tissue. PLoS ONE *9*, e89500.

Jindra, M., Malone, F., Hiruma, K., and Riddiford, L.M. (1996). Developmental profiles and ecdysteroid regulation of the mRNAs for two ecdysone receptor isoforms in the epidermis and wings of the tobacco hornworm, *Manduca sexta*. Dev. Biol. *180*, 258–272.

Karwacz, K., Miraldi, E.R., Pokrovskii, M., Madi, A., Yosef, N., Wortman, I., Chen, X., Watters, A., Carriero, N., Awasthi, A., et al. (2017). Critical role of IRF1 and BATF in forming chromatin landscape during type 1 regulatory cell differentiation. Nat. Immunol. *18*, 412–421.

Kreher, J., Kovač, K., Bouazoune, K., Mačinković, I., Ernst, A.L., Engelen, E., Pahl, R., Finkernagel, F., Murawska, M., Ullah, I., and Brehm, A. (2017). EcR recruits dMi-2 and increases efficiency of dMi-2-mediated remodelling to constrain transcription of hormone-regulated genes. Nat. Commun. *8*, 14806.

Lamparter, D., Marbach, D., Rueedi, R., Bergmann, S., and Kutalik, Z. (2017). Genome-Wide Association between Transcription Factor Expression and Chromatin Accessibility Reveals Regulators of Chromatin Accessibility. PLoS Comput. Biol. *13*, e1005311.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lee, D. (2016). LS-GKM: a new gkm-SVM for large-scale datasets. Bioinformatics *32*, 2196–2198.

Lewis, J.J., and Reed, R.D. (2019). Genome-wide regulatory adaptation shapes population-level genomic landscapes in *Heliconius*. Mol. Biol. Evol. *36*, 159–173.

Lewis, J.J., van der Burg, K.R.L., Mazo-Vargas, A., and Reed, R.D. (2016). ChIP-seq-annotated *Heliconius erato* genome highlights patterns of cis-regulatory evolution in Lepidoptera. Cell Rep. *16*, 2855–2863.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat. Rev. Genet. *3*, 662–673.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., and Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat. Genet. *50*, 259–269.

Minnich, M., Tagoh, H., Bönelt, P., Axelsson, E., Fischer, M., Cebolla, B., Tarakhovsky, A., Nutt, S.L., Jaritz, M., and Busslinger, M. (2016). Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. Nat. Immunol. *17*, 331–343.

Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L., and Bulyk, M.L. (2013). DNA-binding specificity changes in the evolution of forkhead transcription factors. Proc. Natl. Acad. Sci. USA *110*, 12349–12354.

Nijhout, H.F. (1980). Pattern formation on lepidopteran wings: determination of an eyespot. Dev. Biol. *80*, 267–274.

Nitta, K.R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E.E.M., and Taipale, J. (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. eLife *4*, e04837.

Ou, Q., Zeng, J., Yamanaka, N., Brakken-Thal, C., O'Connor, M.B., and King-Jones, K. (2016). The insect prothoracic gland as a model for steroid hormone biosynthesis and regulation. Cell Rep. *16*, 247–262.

Ou, J., Wolfe, S.A., Brodsky, M.H., and Zhu, L.J. (2018). motifStack for the analysis of transcription factor binding site evolution. Nat. Methods *15*, 8–9.

Perry, M., Kinoshita, M., Saldi, G., Huo, L., Arikawa, K., and Desplan, C. (2016). Molecular logic behind the three-way stochastic choices that expand butterfly colour vision. Nature *535*, 280–284.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics *47*, 11.12.1–11.12.34.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (*W1*), W160–W165.

Reed, R.D., and Serfas, M.S. (2004). Butterfly wing pattern evolution is associated with changes in a Notch/Distal-less temporal pattern formation process. Curr. Biol. *14*, 1159–1166.

Rountree, D.B., and Nijhout, H.F. (1995). Hormonal control of a seasonal polyphenism in *Precis coenia* (Lepidoptera: Nymphalidae). J. Insect Physiol. *41*, 987–992.

Shlyueva, D., Stelzer, C., Gerlach, D., Yáñez-Cuna, J.O., Rath, M., Boryń, Ł.M., Arnold, C.D., and Stark, A. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. Mol. Cell *54*, 180–192.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. Bioinformatics *21*, 3940–3941.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordân, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends Biochem. Sci. *39*, 381–399.

Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell *154*, 888–903.

Thomas, S., Li, X.-Y., Sabo, P.J., Sandstrom, R., Thurman, R.E., Canfield, T.K., Giste, E., Fisher, W., Hammonds, A., Celniker, S.E., et al. (2011). Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. Genome Biol. *12*, R43.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Uyehara, C.M., and McKay, D.J. (2019). A direct and widespread role for the nuclear receptor EcR in mediating the response to ecdysone in *Drosophila*. bioRxiv. https://doi.org/10.1101/517458.

Uyehara, C.M., Nystrom, S.L., Niederhuber, M.J., Leatham-Jensen, M., Ma, Y., Buttitta, L.A., and McKay, D.J. (2017). Hormone-dependent control of developmental timing through regulation of chromatin accessibility. Genes Dev. *31*, 862–875.

Wernet, M.F., Mazzoni, E.O., Çelik, A., Duncan, D.M., Duncan, I., and Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. Nature *440*, 174–180.

Wiegmann, B.M., Trautwein, M.D., Kim, J.-W., Cassel, B.K., Bertone, M.A., Winterton, S.L., and Yeates, D.K. (2009). Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. BMC Biol. *7*, 34.

Yao, T.-P., Forman, B.M., Jiang, Z., Cherbas, L., Chen, J.-D., McKeown, M., Cherbas, P., and Evans, R.M. (1993). Functional

ecdysone receptor is the product of *EcR* and *Ultraspiracle* genes. Nature *366*, 476–479.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, L., Martin, A., Perry, M.W., van der Burg, K.R.L., Matsuoka, Y., Monteiro, A., and Reed, R.D. (2017). Genetic basis of melanin pigmentation in butterfly wings. Genetics *205*, 1537–1550.

Zhu, L.J., Christensen, R.G., Kazemian, M., Hull, C.J., Enuameh, M.S., Basciotta, M.D., Brasefield, J.A., Zhu, C., Asriyan, Y., Lapointe, D.S., et al. (2011). FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. Nucleic Acids Res. *39*, D111–D117.

Zitnan, D., and Adams, M.E. (2000). Excitatory and inhibitory roles of central ganglia in initiation of the insect ecdysis behavioural sequence. J. Exp. Biol. *203*, 1329–1340.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
ecdysone receptor (EcR) B1-isoform specific	Developmental Studies Hybridoma Bank – University of Iowa	RRID:AB_528212
Spineless	Gift from M. Perry; Perry et al., 2016	N/A
Biological Samples		
Junonia coenia	F. Nijhout, Duke University Reared in-house	N/A
Critical Commercial Assays		
TRIzol Plus RNA Purification Kit	Invitrogen	12183555
Genomic-tip 100/G	QIAGEN	10243
NEBNext Ultra RNA Library Prep Kit for Illumina	NEB	E7530L
NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB	E7645S
Nextera DNA Library Preparation Kit	Nextera	15028211
Deposited Data		
ATAC-seq data	This paper	GEO:GSE121541
ChIP-seq data	This paper	GEO:GSE121542
RNA-seq data	This paper	GEO:GSE121734
Junonia coenia genome build JC v.1.0	This paper, lepbase.org	http://download.lepbase.org/v4/sequence/
Melitaea cinxia proteins	Ahola et al., 2014	LS-GKMGCA_000716385.1
Motif comparisons	This paper	https://doi.org/10.17632/99f765vwcp.1
Software and Algorithms		
MAKER	Campbell et al., 2014	http://gmod.org/wiki/MAKER
FALCON 0.4.2	Chin et al., 2016	https://github.com/PacificBiosciences/FALCON-integrate/tree/v0.4.2
deepTools 3.0.2.	Ramírez et al., 2016	https://deeptools.readthedocs.io/en/latest/index.html
R package DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
fseq 1.84	Boyle et al., 2008	http://fureylab.web.unc.edu/software/fseq/
FIMO	Grant et al., 2011	http://meme-suite.org/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
TopHat	Trapnell et al., 2012	https://ccb.jhu.edu/software/tophat/manual.shtml
Cufflinks	Trapnell et al., 2012	http://cole-trapnell-lab.github.io/cufflinks/
LS-GKM	Ghandi et al., 2014; Lee, 2016	https://github.com/Dongwon-Lee/lsgkm
R package ROCR	Sing et al., 2005	https://cran.r-project.org/web/packages/ROCR/ROCR.pdf
bedtools	Quinlan, 2014	https://bedtools.readthedocs.io/en/latest/
GLAM	Frith et al., 2008	http://meme-suite.org/index.html
Fly Factor Survey	Zhu et al., 2011	http://mccb.umassmed.edu/ffs/
Tomtom	Gupta et al., 2007	http://meme-suite.org/index.html
Blast	Challis et al., 2016	https://blast.ncbi.nlm.nih.gov/Blast.cgi
R package motifStack	Ou et al., 2018	https://www.bioconductor.org/packages/release/bioc/html/motifStack.html
MACS2	Zhang et al., 2008	http://liulab.dfci.harvard.edu/MACS/
MEME-ChIP	Machanick and Bailey, 2011	http://meme-suite.org/index.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Karin R.L. van der Burg (krv32@cornell.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Junonia coenia

For this study we used larvae and pupae from the butterfly *Junonia coenia*, derived from a Durham, North Carolina population. Larvae were fed an artificial diet as described in Nijhout (1980), and kept at 27°C with a 16:8 hour light/dark schedule. For DNA extractions for the butterfly genome we used a female whole pupa (the heterogametic sex) inbred for three generations. For ATAC-seq and mRNA-seq we collected three replicates of forewing and hindwing tissue at four developmental stages: wandering stage of a 5th (last) instar caterpillar (5th), prepupae (pp), 72 hours after pupation (72h), and six days after pupation, when ommochrome pigments appear (d7). For ChIP-seq, we collected two replicates per time point, and used 26 hindwings for pp per replicate, 21 hindwings for 72h pupae per replicate, and 26 hindwings for d7 pupae per replicate.

METHOD DETAILS

J. coenia genome assembly and annotation

We extracted high molecular weight DNA from an inbred *J. coenia* female using QIAGEN Genomic-tip 100/G. PacBio sequencing was done at the Duke Center for Genomic and Computational Biology. Genome assembly was outsourced to DNAnexus. Data from 58 SMRT cells, which represented a 124-fold coverage of the genome, were analyzed using the FALCON assembly pipeline (Chin et al., 2016) using a length cut-off of 7,954 bp during the initial error-correcting stage. The error-corrected reads were processed by the overlap portion of the FALCON pipeline using a length cut-off of 5,000 bp. The aligned reads were assembled in the third stage of FALCON into primary contigs containing 525 Mbp with an N50 contig length of 534 kbp. The assembly was then polished using PacBio's Quiver algorithm from SMRT Link 3.1, using the original raw reads. We generated a reference transcriptome using Tophat and Cufflinks (Trapnell et al., 2012) to assemble mRNA sequencing data collected from multiple stages of egg, larval, and pupal development (Table S1). This reference transcriptome was used in conjunction with protein annotations from *Melitaea cinxia* (Ahola et al., 2014) to perform gene annotation on the final *J. coenia* genome assembly using three iterations of MAKER (Campbell et al., 2014).

ATAC-seq and mRNA-seq

We collected three replicates of forewing and hindwing tissue at four developmental stages (see Experimental Methods and Subject Details). We sampled both left and right wings, and chose one wing for immediate use in the ATAC-seq assay, and stored the opposing wing in Trizol for later RNA extraction. For ATAC-seq, nuclei were extracted and processed as previously described (Lewis and Reed, 2019). For RNA extraction, we used the Invitrogen RNA PLUS kit to extract RNA. We used the Ultra RNA Library Prep Kit from NEBNext. We sequenced all ATAC-seq and mRNA-seq libraries at the Cornell Genomics Facility, on 2 by 36 bp lanes with an Illumina NextSeq500. For ATAC-seq, we required a minimum of 10 million aligned reads after removal of duplicate and non-uniquely aligning reads, for RNA-seq we required a minimum of 20 million aligned reads (Table S1).

ChIP-seq

For ChIP-seq, we followed the tissue sampling and processing protocol as described in Lewis et al. (2016), with a few notable exceptions. We used two replicates per time point, and used more tissue: 26 hindwings for pp per replicate, 21 hindwings for 72h pupae per replicate, and 26 hindwings for d7 pupae per replicate. After tissue lysis, nuclei were sheared with a diagenode bioRuptor on high, for 3x 5 min with 30 s on, 30 s off. Between 1.5 and 2.5 μg of sheared nuclei were set aside for input control. For EcR ChIP, for each sample, we added 10 μL of antibody 6B7 (Riddiford, DSHB) to 13 to 53 ug of chromatin. For spineless ChIP, we added 15 μL of spineless antibody (a generous gift from M. Perry [Perry et al., 2016]), to 49 to 53 μg of chromatin. For the pulldown, we used 20 μL of a mixture of 75% pG beads, and 25% pA. After reverse cross-linking and purification, library preparation was done with the NEB Ultra Library prep kit. We amplified our libraries for 13 cycles and purified them using AMPure XP beads.

QUANTIFICATION AND STATISTICAL ANALYSIS

ChIP-seq data analysis

Reads were aligned to our reference genome, and low quality alignments and non-uniquely aligning reads were removed using a custom script (Lewis et al., 2016). For each replicate, peaks were called using MACS2 'callpeak' command with –g set to 5.8e8 and default parameters (Zhang et al., 2008). Peaks were considered confirmed if they occurred in both replicates with a 33% overlap, as determined with bedtools function "intersect -u –f 0.33 –F 0.33 –e." To determine overlap of confirmed ChIP peaks with ATAC-seq peaks, the same function was used. Read alignment enrichment around regions of interest was determined with DeepTools, or the

R packages 'genomation' (Akalin et al., 2015; Ramírez et al., 2016). To determine TF binding motifs, we used MEME-ChIP on our confirmed ChIP peaks, using a similar negative peakset as described previously (Machanick and Bailey, 2011). Motif abundance of MEME-ChIP annotated motifs was determined using FIMO (Grant et al., 2011).

ATAC-seq data analysis

ATAC-seq reads were aligned to the *J. coenia* reference genome using Bowtie2 (Langmead and Salzberg, 2012). Duplicate and multiple aligning reads were removed from the analysis. We assessed read depth Pearson correlations between replicates within bins of 1 kB using the Deeptools software package (Ramírez et al., 2016). When correlations higher than 95% between replicates were verified, we merged replicate samples together, and called peaks on each assay using fseq (Boyle et al., 2008). Next, we merged all peak calls together and counted reads for each individual sample for differential peak analysis with DESeq2 (Love et al., 2014). We did pairwise comparisons between consecutive stages and determined how many peaks were significantly upregulated, downregulated, or not significantly different between stages. Peaks were considered 'strong opening' (SO) when there was a significant difference (adjusted p value < 0.05) between two stages, a log2 fold change greater than 2.5, and a read count number in the lowest 0.5 percentile in the earlier stage. These SO peaks were used for subsequent motif analysis. Thus, we had six peak sets for subsequent motif analysis. For both forewings and hindwings, we determined change in SO peaks across multiple developmental stage transitions with DESeq2. For 5^{th} to pp and 72h to d7 SO peaks that showed a significant increase in read abundance in d7 and pp, respectively, we determined abundance of TF groups associated motifs, and calculated the ratio between these SO peaks and corresponding negative peaks of the same size. Lastly, to compare forewings and hindwings directly, we merged forewing and hindwing peaks sets and counted reads for differential peak analysis. Using DESeq2 we did a pairwise comparison between forewings and hindwings, determining how many peaks were different between datasets. To call significance, we used a Benjamini-Hochberg adjusted p value < 0.05.

mRNA-seq data analysis

Next, we aligned mRNA sequencing reads to the *J. coenia* reference genome and used our MAKER transcriptome to count reads for each individual. Differential gene expression was analyzed with DeSEQ2. We determined delta log2 fold change by using the mean difference in normalized read count between two consecutive stages for all three replicates, and used the Benjamini-Hochberg adjusted p- value < 0.05.

Motif prediction

To predict motifs we used the LS-GKM software package as follows (Ghandi et al., 2014; Lee, 2016). First, for each peak set, we created a negative control random sequence file of similar GC content using bedtools (Quinlan, 2014). For each peak, we randomly selected a sequence of the same length elsewhere in the genome where ATAC-seq reads were aligned. Since multiple aligning reads were removed from the sequence dataset, we effectively avoided selecting repeat regions. We split off a fifth of our sites for future testing, and used the remainder to train the SVM model. We ran multiple rounds of training to determine the optimal parameters, and found the gkmrbf-kernel with a kmer size of 10, where the number of informative columns was 4, and gap length was 3. The training was iterated 5 times, each time a fifth of the training data was left out and to use later to determine accuracy. Area under the curve (AUC) scores, a measure of accuracy, were determined with R package ROCR (Sing et al., 2005). We then used this model to score all possible 10mers. The 100 10mers with the highest weight (i.e., that best explained the difference between the positive and negative dataset) were clustered together according to similarity. 10mers that did not cluster with any other 10mers were excluded from further analysis. To determine whether these 10mers were indeed associated with strong opening peaks we counted for each cluster how often a sequence occurred in our positive versus our negative test file set.

To determine what transcription factors might be binding to our 10mers, we used GLAM (Frith et al., 2008), to find motifs in each of our clusters. These motifs were compared to motifs from the fly factor survey (Zhu et al., 2011) database using tomtom (Gupta et al., 2007). For each motif, we took the top three most similar motifs, and used blast to find the homolog gene in our *J. coenia* gene database (Challis et al., 2016). Only gene-motif pairs with clear homologs were used in subsequent analyses. Motifs that showed no similarity to existing motifs were discarded. To account for motif similarity within TF families, we sorted our motifs according to similar protein domains into six groups: (1) helix-loop-helix TFs (HLH), (2) HLH TFs with a PAS domain, associated with signal sensors, (3) TFs with a homeodomain, (4) nuclear hormone receptors, (5) TFs with a C2H2 domain associated with transcriptional regulation, and (6) other transcription factors. Motif visualization was done with R package 'motifStack' (Ou et al., 2018). Motif abundance in SO and negative sites was determined with FIMO (Grant et al., 2011). The ratio of motif abundance between the two datasets was plotted against change in mRNA of corresponding TF between developmental stages.

DATA AND SOFTWARE AVAILABILITY

All data and scripts are available upon request. The accession number for the mRNA-seq, ATAC-seq, and ChIP-seq data reported in this paper is GEO: GSE121735. Genome build for JC v1.0 and annotations are available for download at lepbase.org (http://download.lepbase.org/v4/sequence/) and butterflygenome.org. Additional comparisons of computationally discovered motifs and *D. melanogaster* motifs were deposited on Mendeley at https://doi.org/10.17632/99f765vwcp.1.