Dissecting regulatory and expression dynamics during ecdysone signaling in *Drosophila melanogaster*

Andrea Ennio Storti

München 2018

Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Dissecting regulatory and expression dynamics during ecdysone signaling in *Drosophila melanogaster*

Andrea Ennio Storti

aus

Brindisi, Italien

2018

Erklärung:

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Karl-Peter Hopfner betreut.

Eidesstattliche Versicherung:

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 20/04/2018

_____

Andrea Ennio Storti

Dissertation eingereicht am 06.02.2018

1. Gutachter: Prof. Dr. Karl-Peter Hopfner

2. Gutachter: Prof. Dr. Andreas Ladurner

Mündliche Prüfung am 12.04.2018

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to begin my list of "thank yous" by expressing my sincere gratitude to my supervisor, Ulrike Gaul. She gave me the opportunity to work in her lab, and trusted me from the very first moment I started this project. She supported me in every aspect, scientifically and financially. Her name and contribution will always be fully and strongly part of this thesis, and nothing will change it.

I am very grateful to Prof. Dr. Karl-Peter Hopfner for his availability in stepping in during a difficult period, and for his time spent on my thesis. I really appreciate it. I also would like to thank my TAC members, Prof. Dr. Andreas Ladurner and Dr. Jürg Müller, for their commitment in following my project for 4 years, and their useful suggestions and guidance. Moreover, I would like to thank all the other members of the examination board: Prof. Julian Stingele, Dr. Stefan Canzar, Prof. Veit Hornung, and Prof. Roland Beckmann for their support and time.

For the accomplishment of this thesis, I am extremely grateful to Marta Bozek, Roberto Cortini, Alessio Renna and Ulrich Unnerstall. Either experimentally or through fruitful discussions, their help was fundamental. I would not be here writing these words without their support. A huge thank you also goes to all the previous and current members of the Gaul lab. In some way, all of them participated to my personal growth during the past years. And at the end of the day, we also had good fun together.

Many thanks to the IRTG SFB1064 Graduate School "Chromatin Dynamics", for giving me the opportunity to share great scientific and non-scientific moments with other Ph.D. students, some of whom are now very good friends of mine. A special thanks to the IRTG coordinator Dr. Elizabeth Schroeder-Reiter, for her professionalism, kindness and support that make her a terrific person.

Now, there is bunch of guys I have to thank. We took different directions, following our paths that led us far apart from each other. And so all the daily moments we used to share and that made us what we are, could have become just a lost memory. Nevertheless, that has not happened to us. I love how, every time we meet, it feels like time has stopped, and we are back again to our daily moments, with the only difference of stronger hangovers. This is why I have to thank the friends of a lifetime: Cristian, Giaime, Francesco, Luis, Marco, Michele, Niccolò, Simone. All in

different cities or countries, all living new experiences (ehm.. randomly pointing to you two, future grooms), but all still keeping the same bond that has characterized us so far.

Speaking of new experiences, Munich was definitively a challenging one. But when you get to meet beautiful people, any city feels like your hometown. In this sense, I consider myself very lucky. Firstly, I would like to thank George. He hosted me at the beginning of my experience as if I were part of his family. I will always be grateful to him for that. Then, in the next months in Munich, I got to know some guys, with whom a natural friendship started almost instantly. All of sudden, it felt like I have known those guys forever. It was a great sensation. I really hope that we will keep it rolling in the future. This is why I have to thank the Munich friends: Addie, Alessandro, Alessio (I have mentioned you twice, you owe me a beer for that..), Andrea, Elena, Julius, Marco, Niels, Silvia. With all of them I have shared moments that made my experience in Munich a great one.

A very big thanks goes to the Munich Irish Rovers. A bunch of drunk people (but occasionally also good football players) that are as committed to football as they are to a couple of beers altogether, which is the perfect combination. And during a Ph.D., it is great to have a place in which a conversation does not begin with "How is your project going?".

I will never be enough grateful to the people that, more than anyone, helped me, encouraged me, supported me in any decisions of my life. Ettore, Flavia and Rosaria: my family. A constant presence in all my acknowledgments: they have been with me in the past, they are with me now, and they will always be with me in the future. If I achieved this fantastic result, it is all thanks to them.

Finally, Elena. Elena knew how to handle all the moody up and downs I had during my Ph.D. She understands what I feel with no need for words. She knows me. She comforts me. She is the solid foundation on which everything during this experience could rely. Without her, everything would have fallen apart. And so, day by day, we have slowly built a lot together in Munich, and we will keep doing that. Thank you for all.

ABSTRACT

The information for transcriptional regulation of gene expression is located in *cis*-regulatory regions, such as enhancers and promoters. Chromatin accessibility is a hallmark of active regulatory regions, characterized by nucleosome depletion and transcription factor binding events. Despite decades of investigations, the mechanisms that connect regulatory regions, transcription factor binding and transcript synthesis still need further elucidation.

In *Drosophila melanogaster*, the steroid hormone ecdysone is a key player that dictates developmental processes. Ecdysone binds to a nuclear receptor and triggers a multi-tiered transcriptional cascade that ultimately regulates thousands of genes in the genome. Interestingly, the response to ecdysone is highly context-dependent, leading to distinct molecular and morphological outputs.

Here, we investigated the role of ecdysone in regulating chromatin and transcriptional dynamics in S2 cells and larval development. A multi-pronged experimental and computational approach was performed in a genome-wide quantitative fashion. We integrated: (1) chromatin accessibility assays to assess activity of regulatory regions and the contribution of transcription factor binding events; (2) nascent RNA levels to correlate regulatory regions activity and expression output; (3) differential MNase-seq to measure nucleosome occupancy in dynamic regulatory regions.

Firstly, ecdysone-stimulated S2 cells were monitored over a time course. Dynamic regulatory regions quantitatively correlated with nascent RNA levels, with increasingly active regions promoting gene expression and vice versa. However, a small set of regions that showed decreased activity, targeted up-regulated genes. Those genes execute the ecdysone-dependent morphological changes occurring in S2 cells, indicating that complex regulatory mechanisms control key developmental genes. Additionally, a comprehensive transcription factor lexicon provided novel candidates and modes of action involved in the ecdysone transcriptional cascade.

Furthermore, nucleosomes with different sensitivity to MNase digestion were mapped over dynamic regulatory regions in S2 cells. Striking changes in nucleosome sensitivity were detected in relation to regulatory regions activity. Knockdown experiments indicate that two main ecdysone-pathway components, EcR and br, may play a fundamental role in these activity-driven changes.

In parallel, we studied the ecdysone paradigm also in the whole organism. Three stages of larval-to-pupal development were investigated. Tissues representative of different context-dependent ecdysone responses were selected, and subjected to chromatin accessibility probing. Our approach provided excellent recovery of tissue-specific accessibility landscapes, and demonstrates that distinct activities of regulatory regions shape cell identity. Moreover, a motif enrichment analysis was conducted in a tissue-specific manner, indicating novel transcription factor candidates that may determine the diverse outputs in response to ecdysone.

In summary, we developed a comprehensive set of experimental and computational strategies to investigate gene regulation mechanisms in *Drosophila* development. Arguably, our extensive datasets provide the largest genome-wide characterization of the ecdysone cascade, as chromatin structure, activity of regulatory regions, contribution of transcription factors, and expression outputs were measured and integrated with a very detailed spatio-temporal resolution.

1 INTRODUCTION

1.1 Mechanisms of transcriptional regulation

Transcriptional regulation of gene expression is a fundamental cellular process that controls the level of transcript synthesis in a correct spatio-temporal fashion (Tautz, 2000). In developmental biology, gene regulation has been shown to play an essential role in body patterning and cell fate determination in response to intra-organismic or environmental stimuli (Levine, 2010). Gene expression is regulated by non-coding DNA *cis*-regulatory elements (CREs): enhancers, which are non-directional sequences with transcription factor (TF) binding sites for spatio-temporal control of the expression, and core promoters, which are directional sequences for the control of expression levels. Enhancers and promoters are tightly connected due to the role of enhancers in driving the assembly and initiation of the transcriptional machinery, the RNA polymerase II (Pol II) complex, on promoters. The accessibility of CREs to DNA-binding proteins is restricted by the local structure of the chromatin, which is determined by nucleosome occupancy, positioning and epigenetic post-translational modifications of histones. However, chromatin accessibility is a highly dynamic process, which involves the recruitment of chromatin remodelers to enhance TF or Pol II machinery binding on active CREs (Figure 1).

Due to its central role in gene regulation, the organization of regulatory regions is of particular interest for biological investigation. However, CREs are much more flexible in encoding information than coding sequences (Istrail and Davidson, 2005): (1) regulatory regions such as enhancers are orientation- and distance-independent; (2) typically a single regulatory region receives input from multiple TFs and contains multiple binding sites for each factor (Li et al., 2008); (3) the spacing between binding sites can be fairly flexible (mainly in developmental enhancers) without affecting the transcriptional output (Kulkarni and Arnosti, 2003). Therefore the identification of regulatory regions is challenging and their behavior difficult to predict.
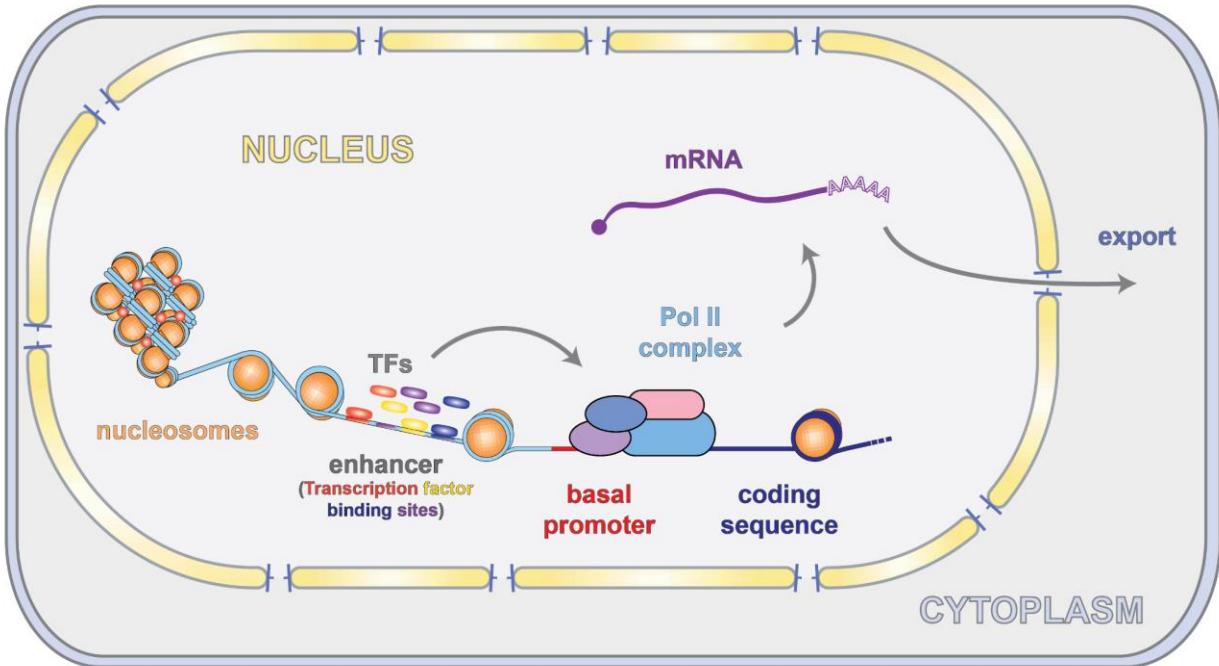
Figure 1: Schematic illustration of transcriptional regulation. Local structure of the chromatin in active enhancers and promoters is characterized by diminished nucleosome occupancy. Thus, TFs and Pol II complex can bind DNA and regulate transcript synthesis.

1.1.1 Promoters

A core promoter is often defined as the ~150 bp of sequence that surrounds the transcription start site (TSS). Several core promoter consensus motifs were characterized in the last decades and their patterns were linked to the expression level of the downstream genes (Lubliner et al., 2015). Historically, three motifs are mostly cited when referring to promoter structure: a TATA sequence, otherwise called TATA box, centered at around -30 (that is, 30 bp upstream of the TSS), an Initiator (Inr) sequence at +1, and downstream promoter elements (DPE) at around +30.

Those elements provide a platform on which the Pol II machinery assembles. Pol II machinery is composed of the Pol II enzyme, which transcribes mostly protein-coding genes, and numerous general transcription factors (GTFs) (Roeder, 1996). Among the GTFs, TFIID is the main component that permits the initiation of Pol II machinery nucleation (Smale and Kadonaga, 2003). The two most characterized core promoter elements, TATA box and Inr, serve as binding sites for TFIID: specifically, TATA box is recognized by the TFIID-subunit TATA binding protein (TBP) with consequent recruitment of Pol II, whereas Inr interaction with TFIID is dependent on TBP-associated factor 1 and 2 (TAF1 and 2). This redundancy of motif recognition by TFIID is

explained by genome-wide studies of core promoter architecture: although TATA box and Inr can be found together in some promoters, they often occur separately and in different functional families of genes (Frith et al., 2008).

The heterogeneity of promoter structures has led to thorough investigations on promoter responsiveness to enhancer inputs based on different promoter features (Arnold et al., 2016; Juven-Gershon et al., 2008). It is clear now that the diversity in structure and function of core promoters significantly contributes to developmental processes, however how all the regulatory input from TFs that bind to distal or proximal enhancers is integrated into the core promoter is still matter of investigation.

1.1.2   Enhancers

Enhancers are referred to as discrete elements that stimulate transcription in an orientation- and distance-independent manner. They harbor several binding sites for specific TFs, and usually more than one single type of TF binds to an enhancer, refining the definition of an enhancer as a functional cluster of TF binding sites (TFBSs).

Enhancers are responsible for cell type-specific gene expression in response to intra- or extra-cellular stimuli, raising the question on how only a small subset of enhancers are activated in a very specific spatio-temporal manner among the vast repertoire of possible CREs. With the advent of genome-wide techniques, this question has been partly addressed by providing evidence of specific chromatin mark combinations associated with inactive, poised or active CRE states (Ernst and Kellis, 2010). In this regard, the model organism Encyclopedia of DNA Elements (modENCODE) project mapped chromatin features of enhancers in *Drosophila melanogaster* (hereafter referred to as *Drosophila*) (The modENCODE Consortium, 2011; Kharchenko et al., 2011; Negre et al., 2011), which are discussed in greater details in paragraph 1.1.4.

In the past, enhancer direct identification and activity measurement were always difficult, mainly due to their unspecific location along the genome (as opposed to promoters) and to their variable number controlling any individual gene during developmental processes. Recently, in the post-genomic era, those tasks have become more feasible. In *Drosophila*, few studies carried out genome-wide mapping of enhancer activity during development or in different cell types. Worthy of note, the Stamatoyannopoulos lab utilized DNase-seq (discussed in greater details in paragraph 1.1.6) to obtain a landscape of active CREs in five *Drosophila* embryonic stages (Thomas et al.,

2011), whereas the Stark lab established the self-transcribing active regulatory region sequencing (STARR-seq) to annotate cell-type specific enhancers independently from their chromatin state (Arnold et al., 2013). Both studies reported that tissue- or stage-specific enhancer activity directly shapes cell identity, highlighting the importance of enhancer plasticity in development.

With regard to the distribution of TFBSs within enhancers, two models were proposed to depict their architecture: (1) in the enhanceosome model, the precise arrangement of TFBSs is a critical parameter to achieve a high cooperativity among the enhancer-binding proteins, which can therefore assemble in a well-defined nucleoprotein complex to direct a single output to the Pol II machinery (Merika and Thanos, 2001). Any alterations in individual binding sites or in the spacing between adjacent sites can disrupt the enhanceosome function, as demonstrated in the mammalian IFN-β regulatory element, the best characterized example of enhanceosome (Panne et al., 2007); (2) the billboard model, instead, illustrates a more flexible organization of TFBSs within an enhancers. As opposed to the enhanceosome, the spacing between TFBSs is less critical, and the entire element does not necessarily need to act as a single unit, but rather as a composition of separate subelements that can independently regulate gene expression (Arnosti et al., 1996). Therefore, a billboard enhancer functions as an information display which is interpreted by consecutive and discrete interactions with the Pol II machinery, also in the simultaneous presence of activating and repressive states (Kulkarni and Arnosti, 2003). The billboard model describes the mode of action of many developmental enhancers, which can generate complex patterns of gene expression during development (Papatsenko et al., 2009).

Very interestingly, combinations of binding sites of some TFs occur more often than others, especially in developmental enhancers (Dogan et al., 2015). These TFBS patterns were identified as very conserved in different species and therefore used to improve enhancers computational detection (Arnold et al., 2014). TFBS patterns lead to cooperative TF binding events (TFBEs), a key factor for a functional enhancer. As a general rule of thumb, inducible inactive enhancers are covered by a condensed nucleosomal array that prevents access to TFBSs. The expression of a so called pioneer TF (described in paragraph 1.1.3) in response to a stimulus permits chromatin remodeling and exposure of TFBSs. Consequently, TFBEs occur, often through a hierarchical manner in which an initial TFBE stabilizes the binding of successive TFs, eventually leading to enhancer activation (Biddie et al., 2011). Nevertheless, this is simplistic model, as other features may determine enhancer activation: (1) despite occurring TFBEs, the enhancer could be marked

by repressive chromatin states that hold it in a poised state (Bonn et al., 2012); (2) the enhancer activity could be tightly dependent on the concentration of activators and repressors in overlapping spatial domains, as demonstrated for the *Drosophila* segmentation paradigm (Stanojevic et al., 1991); (3) how TFs cooperate and the recruitment of additional tissue-specific co-activators is often context-dependent and is not applicable to all the enhancer sets (Spitz and Furlong, 2012).

Despite the advances in technology and the big efforts in systematically identifying enhancers, the understanding of their mechanisms in regulating gene expression is still far from being accomplished. It is clear that an integrated analysis of enhancer architecture and TFBEs is essential to decipher enhancer functions, and a comprehensive but still context-dependent investigation constitutes the best strategy to pursue.

1.1.3 TFs

TFs are proteins containing at least one DNA-binding domain (DBD) and that recognize a small 6-12 bp long DNA sequence called motif. Typically, the sequence specificity of TFs *in vivo* is fairly low compared to other DNA-binding proteins (i.e. restriction enzymes). Therefore, the main question regarding TFs is how they can precisely target such a small degenerated sequence only in specific genomic regions.

In the context of development and cell differentiation, chromatin structure plays an intrinsic repressive role: nucleosome condensation forms higher levels of chromatin compaction that limits the amount of free DNA available for other binding proteins (such as TFs), preventing undesired gene expression. This physical barrier on CREs is overcome by pioneer factors. Pioneer factors are able to recognize their cognate TFBSs in the context of nucleosomal DNA (that is, DNA bound to nucleosomes) (Cirillo et al., 2002). It seems that pioneer factors have a longer residence time on nucleosomal DNA than other TFs, favoring the recruitment of chromatin remodelers and/or establishing cooperation with other TFs. This mechanism would permit nucleosome displacement on CREs, exposing TFBS-containing nucleosome-free DNA that would be then occupied by the recruited TFs, strengthening the complex (Li et al., 2012; Sekiya et al., 2009). Thus, a mode of action for TF specific binding is characterized by a limited exposure of TFBSs due to local nucleosome displacement and by a functional cooperative binding among cell-specific TFs (Carroll et al., 2005). TF cooperativity in a context-dependent environment is also a determinant

to distinguish functional from non-functional individual TFBEs (Li et al., 2008; Stampfel et al., 2015).

In parallel, many studies focused on the identification and prediction of TFBSs genome-wide, often with the intent of modeling gene expression. However, the low specificity of the motifs has made this task challenging, especially computationally. Therefore, a position weight matrix (PWM) has emerged as a popular instrument to model the motif variability (Stormo et al., 1982). PWMs specify the frequency distribution of nucleotides at each position of the TFBS, and their individual contribution to the binding affinity (Sinha, 2006). Generally, PWMs are visualized as sequence logos. In the *Drosophila* segmentation paradigm, the involved TFs represent one of the best examples in terms of motif characterization. Consequently, their PWMs were used to feed algorithms aimed to predict TFBSs. Blatti and colleagues applied a motif- and chromatin accessibility-based approach to reliably identify the regulatory mechanisms in *Drosophila* embryonic development, almost as accurately as experimental assays (Blatti et al., 2015). Instead, Segal and colleagues developed a thermodynamic model that integrates CRE sequences, motifs and TF expression information to predict expression patterns at spatial resolution along the embryo (Segal et al., 2008).

PWMs certainly improved the computational identification of TFBSs, however they are dependent on the type of experimental data that provide the binding information. In the recent years, such data have been obtained by various high-throughput methods. Arguably, the most popular is Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) (Johnson et al., 2007). ChIP-seq not only provides information on TF motifs, but also allows genome-wide TFBE mapping *in vivo*. Nevertheless, ChIP-seq carries some experimental limitations: (1) it strongly relies on the antibody quality for TF pull down; (2) low TF expression levels could affect the signal-to-noise ratio (SNR), making difficult to distinguish true TFBEs; (3) only one TF can be investigated per experiment. Alternatively, *in vitro* high-throughput methods were established to measure TF-DNA interactions, such as protein-binding microarrays (PBM) (Badis et al., 2009) , bacterial one-hybrid assay (B1H) (Noyes et al., 2008), and high-throughput SELEX (HT-SELEX) (Jolma et al., 2013) . Although those methods vary with respect to their experimental setups, they share a common drawback: TF-DNA interactions undergo stringent washes, resulting in rigid PWMs and excluding weak binding information, which are thought to be essential for CRE activity *in vivo* (Segal et al., 2008; Tanay, 2006).

Regardless its limits, this enormous characterization of TF motifs through different methods constitutes a milestone in transcriptional regulation research. For *Drosophila*, several databases were created to provide community access to hundreds of TF binding information, such as FlyFactorSurvey (Zhu et al., 2011) (http://mccb.umassmed.edu/ffs/), RedFly (Gallo et al., 2006) (http://redfly.ccr.buffalo.edu/index.php), and Jaspar (Sandelin et al., 2004) (http://jaspar.genereg.net/).

1.1.4 Chromatin features of CREs

Arguably, the modENCODE project provided the largest genome-wide dataset of functional element identification, CRE mapping, and chromatin landscape in *Drosophila*. The project generated datasets that profile transcripts, histone modifications, nucleosome properties and TFs in cell lines and tissues (or whole organism) during development (The modENCODE Consortium, 2011). Generally, the study reported that very few genes showed constant repressive or activating marks throughout development, whereas most genes locate within dynamically marked regions, as confirmation of the high plasticity of the genome.

With respect to CREs, specific chromatin characteristic signatures were detected: enrichment of H3K4me3, H3K9ac and H3K27ac histone modifications, depletion of nucleosome density and increased nucleosome turnover were found to mark TSS-proximal regions (that is, active promoters); whereas active enhancers were associated with enrichment of H3K4me1, H3K27ac and higher occupancy of CREB-binding protein (CBP, a transcriptional co-activator known to interact with a large number of developmental TFs). Interestingly, both promoters and enhancers showed higher presence of chromatin remodelers compared to other genomic regions, further highlighting the dynamic processes that govern chromatin structure in CREs (Kharchenko et al., 2011). By integrating all those data in two unsupervised hidden Markov models (HMM), the modENCODE authors captured the overall complexity of chromatin signatures with 9 combinatorial states, associating each genomic location with a particular state. Promoter- and enhancer-like signatures were identified with state 1 and 3, respectively.

Nucleosome organization around CREs is also a key feature for gene regulation mechanisms. Nucleosome organization is a dynamic process in which chromatin remodeler complexes act by moving nucleosomes through ATP hydrolysis. This could result in nucleosome sliding, partial or complete nucleosome eviction, or exchange of histones with histone variants,

depending on the remodeling complex (Moshkin et al., 2012). A canonical nucleosome pattern in promoters is characterized by a nucleosome-depleted region (NDR) of around 150bp around the TSS with low nucleosome occupancy, surrounded by two well positioned nucleosomes: the -1 nucleosome located upstream of the TSS and the +1 nucleosome located downstream. *In vivo*, it was shown that the +1 nucleosome represents the main obstacle for transcription elongation, at least 2-3 fold higher compared to the +2 and further downstream nucleosomes (Weber et al., 2014). Several mechanisms exist to modulate or overcome the nucleosome barrier, including chromatin remodelers themselves.

Around enhancers, chromatin structure is characterized by two well positioned nucleosomes that surround a NDR fairly located within the enhancer center, which co-localizes with a higher presence of TFBS clusters (Barozzi et al., 2014). As TFBSs are generally located in regions with predicted high nucleosome occupancy (due to high GC content of the underlying DNA, which favors nucleosome-DNA affinity), TFs have to introduce relevant changes in the local chromatin structure in order to have access to their own binding sites (Tillo and Hughes, 2009). The NDR within enhancers partly derives by the affinity and residence time of TFs (Vierstra et al., 2014), and consequently the surrounding nucleosome positioning is driven by a boundary effect introduced by TFs themselves. However, a great contribution on nucleosome organization in enhancers is attributed to chromatin remodelers and pioneer factors, whose dynamic interactions determine nucleosome displacement to support enhancer functions throughout differentiation and development (King and Klose, 2017). Notably, when comparing cells at different developmental stages, drastic variations of nucleosome occupancy occur only locally, and mostly affect a single nucleosome which is often located at the enhancer center and covers TFBSs (West et al., 2014). Those results demonstrated the importance of nucleosomes as gatekeeper of TFBSs, and how their dynamics are restricted to precise regions involved in gene regulation.

1.1.5 Chromatin accessibility

As already mentioned, TFs, Pol II machinery and nucleosomes compete for binding the DNA in CREs. Generally, in active CREs the chromatin acquires an 'open' state, which is characterized by a depletion of nucleosomes to permit TF and Pol II binding. Therefore, it is common to refer to active CREs as open or accessible (chromatin) regions, as opposed to the compact 'closed' chromatin structure typical of heterochromatin or inactive CREs. Over the years, this particular

feature of active CREs has been adapted for promoter and enhancer detection: as nucleosome-free DNA is accessible to DNA-binding proteins, it is also accessible to small unspecific endonucleases, such as DNase I. The result is a targeted digestion of the active CRE DNA, as opposed to nucleosomal DNA which remains intact (Wu, 1980). Therefore, open CREs are also referred to as DNase Hypersensitive Sites (DHSs).

DHS mapping has recently emerged as an essential tool in genome-wide chromatin landscape characterization (Boyle et al., 2008). In the modENCODE project, DHSs were identified in two different cell types. Around 90% of the total number of DHSs were detected in state 1 and 3 regions (promoters and enhancers, respectively). Moreover, more than half of state 3 DHSs occurred within annotated expressed genes, which is in strong agreement with many evidence of enhancer localization in intronic regions. Interestingly, although in general most of the DHSs are in state 1 regions, 91% of cell-type specific DHSs overlapped with state 3 signatures, demonstrating the importance of enhancer plasticity in regulating cell identity (Kharchenko et al., 2011; Song et al., 2011). Overall, these findings presented a fundamental role of chromatin accessibility in chromatin structure investigation and, more specifically, CRE activity assessment.

Subsequently, many studies conducted chromatin accessibility analyses in various biological paradigms: from DHS mapping of 125 human cell lines (Thurman et al., 2012), to the investigation of chromatin accessibility dynamics in response to glucocorticoid stimuli (Stavreva et al., 2015), until the elucidation of chromatin accessibility regulation in mouse cerebellum development (Frank et al., 2015). DHS profiling resulted critical also in the understanding of different cancers progression (Gomez et al., 2016; Qu et al., 2017).

In *Drosophila*, chromatin accessibility dynamics were followed over developmental stages *in vivo* or in response to developmental stimuli *in vitro*. McKay and Lieb mapped DHSs in the fly appendages from embryonic to late larval stages and concluded that sets of DHSs vary across tissues and stages, determining cell fate (McKay and Lieb, 2013); Uyehara and colleagues investigated chromatin accessibility in the wing disc during larval-to-pupal transition, and proposed new regulatory mechanisms involved in DHS dynamics of that particular paradigm (Uyehara et al., 2017); Shlyueva and colleagues complemented their STARR-seq data with a DHS landscape of hormone-responsive CREs *in vitro*, and further confirmed that a stimulated enhancer activity coincides with a closed-to-open state of the chromatin (Shlyueva et al., 2014).

DHS probing provides a comprehensive analysis of functional CREs, but it also contributes to the characterization of novel TFs involved in regulatory mechanisms. Indeed, the information provided by chromatin accessibility is broad: as DHSs and TFBEs are highly correlated, genomic regions that show increased accessibility are incorporated into TFBE prediction algorithms to restrict biologically-relevant binding sites. This approach reliably identifies TFBEs as good as other biochemical assays *in vivo*, such as ChIP-seq (Kaplan et al., 2011; Pique-regi et al., 2011), and has the great advantage of targeting many TFs in a single DHS-probing assay.

Due to its central role in regulating gene expression and shaping cell identity, as well as in TFBE identification, chromatin accessibility has become an essential feature in any high-throughput genomic studies. Therefore, several chromatin accessibility assays have been developed, all of them with advantages and disadvantages according to the paradigm to investigate. Next, I will describe the most common chromatin accessibility assays, which were also fundamental for the accomplishment of this thesis project.

1.1.6 Chromatin accessibility assays

Since the early 1980s, low-throughput studies in *Drosophila* demonstrated that active open chromatin coincides with nuclease hypersensitivity (Wu, 1980). Nowadays, chromatin accessibility assays are coupled to next-generation sequencing (NGS) to generate high-throughput mapping of active CREs. Those assays share a common principle: separation of accessible regions from the rest of the genome can be achieved through limited digestion of the chromatin by enzymatic means. The greatest advantage is the independence from any antibodies or epitope tags. However, the used enzymes seem to have intrinsic sequence biases that could affect the final results (Koohy et al., 2013). Nevertheless, at what level of the analysis enzyme biases significantly mislead the outcome is still debated (Allan et al., 2012; He et al., 2014).

In this paragraph, I report two accessibility assays that have been mostly used in recent years: DNase-seq, which relies on the unspecific endonuclease DNase I, and the assay for transposase-accessible chromatin using sequencing (ATAC-seq). Moreover, I describe MNase-seq, which is the best complementary approach to accessibility assays, as it maps nucleosome positioning and occupancy (Figure 2).
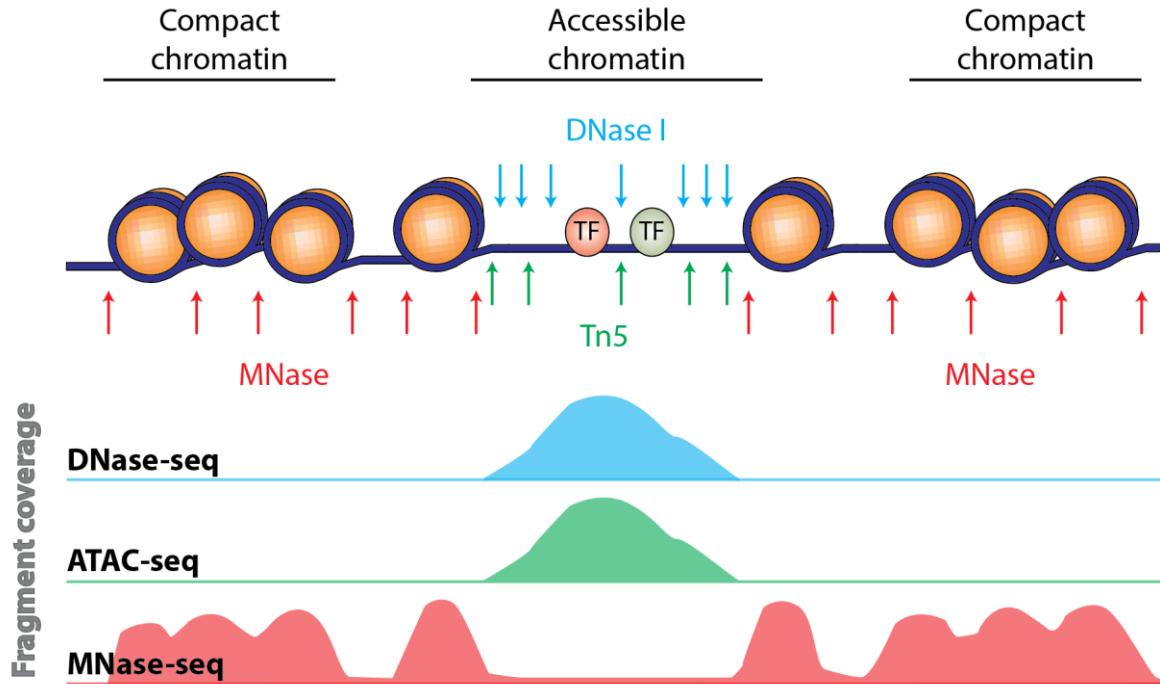
Figure 2: Schematic illustration of chromatin accessibility assays. DNase I (blue arrows) and Tn5 (green arrows) preferentially target regions of accessible chromatin (active CREs). After sequencing, fragment coverage tracks are visible on a genome browser and result in peaks in open regions, whereas compact chromatin shows low flat signal. MNase (red arrows) preferentially cuts within linker DNA. In mono-nucleosomal fragments coverage track of a typical digestion, sharp peaks are visible at the level of well-positioned nucleosomes, which mostly surround active CREs. Nucleosome occupancy in active CREs is low. Regions of compact chromatin show fuzzy nucleosome positioning.

DNase-seq can be considered as the natural follow-up to the early low-throughput accessibility assays after the advent of NGS. It relies on the small unspecific endonuclease DNase I, as initial pioneering studies on open CREs did. In a typical DNase-seq protocol, a nuclei preparation is carried out, followed by a limited digestion of the chromatin by the enzyme (Vierstra et al., 2014). In the context of a general compact structure of the chromatin, DNase I preferentially cuts accessible DNA, therefore targeting active CREs. The digestion step is crucial, as under- or over-digestion lead to low SNR or detection of false positive open regions, respectively. Optimal digestion levels are estimated on agarose gel, and fragments smaller than 500 bp are isolated through sucrose gradient. The rationale is to select fragments released by two cutting events occurred in close proximity, enriching for DNA that derives only from open regions. Isolated fragments are then subject to library preparation and high-throughput sequencing.

The main limitation of DNase-seq is the high number of initial cells required for one experiment. In human cells, at least 10 million cells are needed for a successful assay. In

Drosophila, the number raises to 50 million. This limits DNase-seq applications *in vivo*, where the starting material is usually scarce. Other concerns come from the long protocol and the many steps involved, including a laborious nuclei preparation that could disrupt the native chromatin structure. Moreover, initial small-scale preliminary experiments are necessary to ascertain the optimal digestion level, which is usually dependent on the DNase I lot and the type and number of cells. A big controversy is represented by the intrinsic sequence bias of DNase I, which however seems to be distinguishable only at high sequencing depth (around 200 million pair-end reads), and not at the level of general DHS landscape detection (obtained at around 40 million pair-end reads) (Sung et al., 2014).

Overall, DNase-seq is a powerful and robust technique to identify open active CREs. It generates data with high resolution, SNR and reproducibility. Additionally, a qPCR quality control can be performed to assess the enrichment of known open regions compared to closed loci prior to library preparation. It was extensively used in the modENCODE project, and for all those reasons it is considered as the 'golden standard' for probing chromatin accessibility.

ATAC-seq is a more recent technique compared to DNase-seq, nevertheless it has gained popularity especially for *in vivo* applications. It relies on a hyperactive Tn5 transposase which is pre-loaded with adapters for Illumina sequencing (Buenrostro et al., 2013). After nuclei preparation, chromatin is treated with the Tn5, which inserts the adapters in accessible regions, resulting in 'tagmented' DNA. Subsequent to a quick DNA purification, PCR is performed to amplify tagmented regions, leading to a ready-to-use library for sequencing.

ATAC-seq brought evident improvements in chromatin accessibility probing. In human cells, the starting material can be as low as 500 cells, also due to the absence of a size selection step, although it could be introduced in a customized protocol. This has attracted a growing interest particularly for *in vivo* paradigms. Moreover, the protocol is a simple two-step process that generates a library for sequencing within a day. Finally, the sensitivity and specificity are similar to DNase-seq data.

ATAC-seq drawbacks come from its recent establishment. Tn5 biases and mode of action are still not well characterized. In addition, the protocol lacks a reliable quality control before sequencing to assess open regions enrichment, thus implicating several sequencing test runs to establish perfect tagmentation conditions.

MNase-seq is a complementary assay to accessibility probing, as it detects nucleosome positioning and occupancy genome-wide. It has been implemented in a number of organisms, and it was the first approach that resolved the canonical chromatin structure around TSSs (Mavrich et al., 2008). As opposed to chromatin accessibility assays, its range is not restricted to few open regions, but fairly all the genome is subject to analysis. It relies on MNase, an endonuclease with a pseudo-exonuclease activity. MNase preferentially cuts within linker DNA and thanks to the exonuclease activity, it digests the DNA until a barrier prevents it. This barrier is mainly represented by nucleosomes. Therefore a typical MNase-seq protocol implies a digestion of the chromatin until ~80% of the DNA is between 140 and 160 bp of length, that is, mono-nucleosomal size. Mono-nucleosome bands are extracted from agarose gel, and fragments subject to library preparation and sequencing.

Similarly to DNase- and ATAC-seq, optimal digestion levels must be obtained through titration tests. Interestingly, in this regard, low digestion level experiments have recently showed that certain nucleosomes are more susceptible to MNase digestion (Weiner et al., 2010). Those MNase-sensitive nucleosomes most likely map around promoters and within enhancers, raising the question on their biological relevance in transcriptional regulation. The term MNase-sensitivity mainly indicates a bias in nucleosome occupancy introduced by MNase-based experiments, which nevertheless can be considered as a tool for studying chromatin-related features. More precisely, the terms nucleosome fragility and resistance indicate nucleosomes with a differential stability along the genome due to different DNA sequence content, biophysical properties, and active mechanisms (e.g. chromatin remodeling), with fragile nucleosomes being more susceptible to MNase activity and enriched in active CREs. Thus, differential MNase-seq (that is, MNase-seq conducted with different digestion levels) is used as a probe for measuring nucleosome fragility and resistance. MNase-sensitive (fragile) nucleosomes are characterized by decreased occupancy with the increase of digestion levels throughout a differential MNase-seq (Chereji et al., 2015).

Finally, Tillo and Hughes reported that DNA GC content is also major component of nucleosome sequence preference and is a determinant of intrinsic nucleosome occupancy, which needs to be accounted for when investigating genome organization through MNase-seq (Tillo and Hughes, 2009).

Overall, DNase-, ATAC- and MNase-seq provide robust data on chromatin structure genome-wide. Their wide range and longstanding applications are proofs of their power, however

their limits and enzyme- or digestion-dependent biases must be considered. The easiest control is to perform the same experimental protocol on naked genomic DNA (gDNA), which simultaneously corrects both the intrinsic enzyme sequence specificity and the detection of chromatin structure-independent features.

1.2 The steroid hormone ecdysone

Steroid hormones regulate the development, maturation and metabolism of higher eukaryotes. In *Drosophila*, the steroid hormone 20-hydroxyecdysone (hereafter referred to as ecdysone) is the main component that dictates the timing of developmental processes in response to physiological and environmental cues (Baehrecke, 1996). Periodic pulses of ecdysone are released from the prothoracic glands to the hemolymp, targeting all the peripheral tissues that integrate the biological information carried by the hormone.

During metamorphosis (that is, the transition from larval to pupal stage), two major pulses of ecdysone occur, leading to drastic re-arrangements of the body structures (Figure 3). The first pulse dictates the end of third instar larva (3rd IL) and triggers the initiation of prepupal morphogenesis, a transition characterized by the secretion of glue proteins required for the larvae to attach itself onto a surface. The second pulse is released 10-12 hours later and promotes transition to pupal stage, visible by darkening and hardening of the cuticle and head eversion. At the same time, an extensive range of tissues respond differently to the ecdysone pulses. The most extreme cases are represented by larval tissues and imaginal discs. Larval tissues such as salivary glands (SG) are no longer required in the adult fly, therefore they are removed by histolysis and undergo programmed cell death (PCD). On the contrary, imaginal discs such as the wing disc (WD) and eye disc (ED) complete an additional cell cycle and then begin their differentiation into the future adult appendages (Handler, 1982). Therefore, the cellular processes controlled by ecdysone are vast and include cell proliferation, differentiation and death, along with the regulation of metabolic and behavioral mechanisms. All these processes need to follow a strict developmental timing and are adjusted in a cell and tissue specific manner, implicating complex levels of regulation. Consequently, it is of main interest to understand how a single hormone can lead to such a broad range of morphological responses, and how different tissues individually interpret the ecdysone signal.
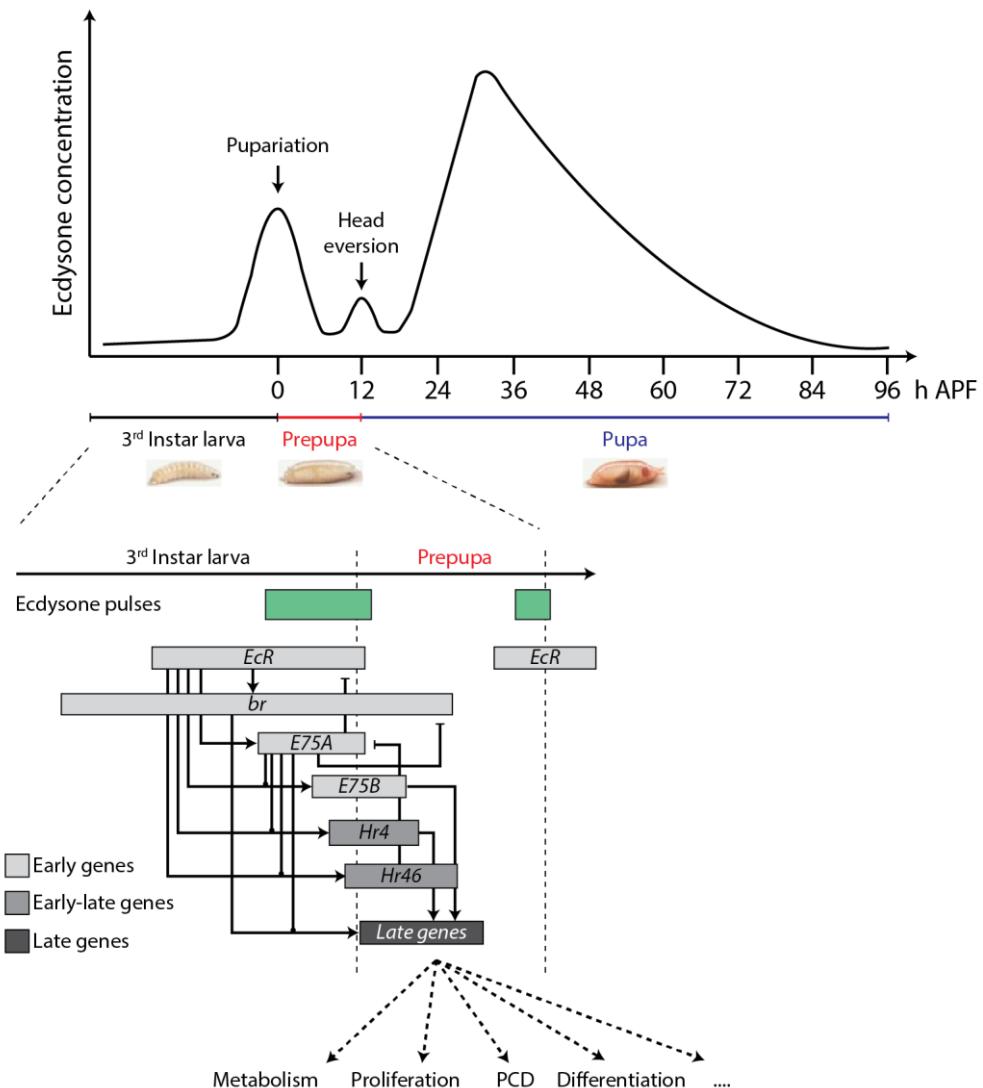
Figure 3: Schematic illustration of ecdysone pulses during metamorphosis. Ecdysone dictates the timing of developmental stages. During metamorphosis, a first ecdysone pulse triggers larval-to-prepupal transition (pupariation). 10-12 hours later, a second pulse triggers prepupa-to-pupal transition (characterized by head eversion). At the molecular level, ecdysone triggers a hierarchical transcriptional cascade, which involves numerous TFs (only few are illustrated here). The final targets of the cascade are the effector late genes, which execute multiple cellular processes in a context-dependent fashion, resulting in drastic tissue-specific morphological changes. The transcriptional cascade scheme is adapted from (Ou and King-Jones, 2013).

1.2.1   The ecdysone signaling pathway

During the 1970s, it was already known that transcriptional mechanisms could be studied *ex vivo* in cultured polytene chromosomes of dipteran SG. Such mechanisms manifested in chromosome puffs as a result of local alterations of chromatin structure to enhance transcriptional activation. Ashburner observed hierarchical puffing patterns in cultured *Drosophila* SG chromosomes upon

ecdysone treatment. He monitored the timing and the location of the puffs, and reported that a relative small number of initial puffs were visible after 10 minutes of ecdysone exposure in specific genomic loci, followed by a greater number of secondary puffs at later time points. He eventually postulated a model in which the early puffs are direct gene targets of ecdysone, encoding regulatory proteins that induce expression in the later puffs (Ashburner, 1973). The so called Ashburner model is a milestone in research on mechanisms regulating gene expression during insect development.

Since then, the ecdysone pathway has been intensively dissected, and all the main players characterized. Initial molecular investigations on the early puffs 2B, 74EF and 75B identified three TF coding regions, namely *Broad Complex* (*BR-C* or *br*), *E74* and *E75*, consistent with the predicted regulatory role of the early puffs postulated in the model (Burtis et al., 1990; Chao and Guild, 1986). Particularly, E75 was discovered to be a member of the nuclear hormone receptor superfamily, raising the possibility of being the ecdysone receptor itself (Segraves and Hogness, 1990). However, experiments on purified E75 proteins could not prove its ability of binding radiolabeled ecdysone. Subsequently, a genomic screen identified the protein encoded by the *DHR23* locus as capable of binding ecdysone in cultured *Drosophila* cells. Accordingly, the corresponding gene was designated as *Ecdysone Receptor* (*EcR*) (Koelle et al., 1991).

EcR is a nuclear hormone receptor, orthologous of the vertebrate farnesoid X receptor (FXR) that comprises a transcriptional activation domain, a DBD and a ligand-binding domain (LBD). However, its ligand-binding activity is dependent on another nuclear receptor, ultraspiracle (USP), which is the orthologous of the mammalian retinoid X receptor (RXR). The EcR-USP heterodimer is the fully functional ecdysone receptor complex in *Drosophila*. In the presence of the hormone, EcR dimerizes with USP allowing efficient binding to ecdysone-responsive elements (EcREs), which co-localize with functional CREs, thus triggering transcriptional activation of ecdysone-responsive genes.

1.2.2 Molecular mechanisms of the cascade

With the identification of the main components involved in the ecdysone transcriptional cascade, a clearer picture on the molecular mechanisms controlling the ecdysone response during development was provided (Figure 3). At the beginning of the larval-to-prepula transition, the first ecdysone pulse induces the expression of a small group of early genes, targeted by the functional

dimer EcR-USP. Those early genes can be classified according to the hormone concentration required for their activation. Class I early genes consist of *EcR* itself and *E74B* and respond to very low concentrations of ecdysone towards the end of the 3rd IL, when the titer of the hormone is increasing. At the beginning of the pupariation, when the hormone concentrations are at their peak, class I genes are repressed and class II early gene transcripts accumulate. The latter consist of *E75A, E75B* and *E74A*. The repression of class I gene is attributed to a feedback loop that involves both class I and class II gene products to limit the duration of the response (Karim and Thummel, 1992).

BR-C shows an unusual behavior, as it appears to have both class I and II characteristics. Its transcription is activated at low ecdysone concentrations as class I early genes, but the maximal activity is reached at the peak of the ecdysone titer as class II early genes, thus responding to a much broader ecdysone dose (Karim and Thummel, 1992). Therefore, br is thought to play a pivotal role both in the initiation and progression of the ecdysone response through metamorphosis (Mugat et al., 2000).

Early gene TFs target late genes, whose puffs (visible if expressed in SG) appear later as they do not directly respond to the hormone signal. Late genes are so called "effector genes", as they execute the developmental changes during larval-to-prepupal transition, thus their expression patterns are more context-dependent. Late genes are numerous: to mention few examples, in SG they vary from the glue genes to the death activators *reaper* (*rpr*), *head involution defective* (*hid*) and *dronc* (Beckendorf and Kafatos, 1976; Jiang et al., 2000), whereas in imaginal discs they can include genes involved in cell proliferation such as *Cyclin B*, *Cyclin D* and the caspase inhibitor *Diap1* (Cranna et al., 2009). Effector genes can also be other TFs or members of non-systemic signaling pathways, such as Wingless (Wg) or decapentaplegic (Dpp) pathways, known to interact locally with the ecdysone cascade (Li and White, 2003).

An additional level of cascade control is attributed to the "early-late genes", consisting of *Hr4, Hr46* and *E78*. Early-late genes respond directly to ecdysone stimulus similarly to the early genes, however their puffs appear with a temporal dynamic comparable to the late effector genes, probably due to a delay in transcript accumulation (Huet et al., 1995). Early-late genes act as early genes by activating late genes and repressing class I early genes, but their different expression timing is thought to provide temporal heterogeneity in both repression and induction mechanisms.

The second pulse of ecdysone during prepupal-to-pupal transition uses the same hierarchy of regulatory early genes but triggers a different set of late effector genes to remodel body plan. In summary, the ecdysone signal is propagated within the nucleus by the interaction of the hormone with EcR-USP heterodimer to induce a cascade of primary regulatory and secondary effector genes. Nevertheless, while the Ashburner model explains the temporal gene expression cascade triggered by an ecdysone pulse, it does not necessarily contribute to the understanding of how individual tissues respond differently to the same developmental stimulus.

1.2.3   Spatio-temporal expression of ecdysone-pathway regulators

The tissue- and stage-specific interpretation of the ecdysone signaling is a combination of multiple factors which, among many, include varying sensitivities of ecdysone-responsive genes to ecdysone concentrations, recruitment of distinct sets of co-activators and co-repressors (Arbeitman and Hogness, 2000), local interaction with other signaling pathways, and differential expression patterns of early gene isoforms. With regard to the latter, it is noteworthy to mention that the genomic loci of the early genes are particularly large and show complex regulatory modules, characterized by large intronic regions that harbor multiple ecdysone-responsive enhancers (Karim et al., 1993). Additionally, splicing variants and alternative promoter usage produce several protein isoforms.

The protein product of *EcR* is present in three isoforms (EcR-A, EcR-B1 and EcR-B2) that share the same DBD and LBD, but differ in their amino-terminal domain (Talbot et al., 1993). The expression patterns of those isoforms have been characterized and seem to play a role in determining cell fate. In 3rd IL, isoform B1 immunohistochemistry predominates in larval tissues that will go through PCD, while isoform A predominates in imaginal discs (Truman et al., 1994). On the contrary, the ventral cord of the central nervous system (CNS) shows approximately 300 neurons with higher levels of isoform A compared to other neurons: those neurons undergo rapid degeneration after the adult emerges from the pupal case (Robinow et al., 1993). Additionally, isoform-specific or all-isoforms mutations lead to distinct stage lethality (Cherbas et al., 2003). As all the isoforms bind the same DNA motif and the hormone equally, differences in their signal propagation must reside in the transcriptional activation through their N-terminal domains, probably due to recruitment of diverse additional co-factors.

BR-C presents 14 transcript variants that are translated in four protein isoforms, distinguished by their zinc finger modules (Z1 to Z4). The extensive usage of alternative promoters to transcribe all those variants is one explanation to the *br* wide response to different ecdysone concentrations, as described above. At the onset of metamorphosis, br isoforms accumulate in the nuclei of all larval and imaginal tissues, however with unique kinetics of induction and repression (von Kalm et al., 1994). Z1 isoform is predominant in SG, whereas in imaginal discs an isoform switch occurs, where an initial Z2 isoform synthesis is followed by Z1 synthesis 4 hours after puparium formation (APF). In CNS, complex isoform combinations are detected along metamorphosis, suggesting that the relative isoform ratio is fundamental for the re-organization of larval neurons (Emery et al., 1994).

Therefore, it is clear that the physiological responses to the ecdysone stimulus are coordinated also, but not necessarily only, by a selective usage of the main pathway regulators. As also other ecdysone-primary responsive genes, such as *E74* and *E75*, show distinct isoform accumulation in a tissue-specific manner (Thummel et al., 1990), it has been proposed that the diversity of ecdysone responses may originate from different combinations of the total early gene products.

1.2.4 Ecdysone-regulated binding events and CRE dynamics

In recent years, investigations on the vast transcriptional cascade triggered by ecdysone have been carried out with the support of high-throughput approaches in different contexts. Those studies provided new insights on the global ecdysone response both at chromatin and expression levels, as well as on different EcR interactions. Nevertheless, at the same time, new questions emerged on the mode of action of the hormone and its cell-specific functions.

Gauhar and colleagues carried out a genome-wide DamID-based identification of the EcR-USP heterodimer binding sites in *Drosophila* Kc167 cells (Gauhar et al., 2009). They reported that only 42% of EcR-USP sites localized near known ecdysone target genes of the cells. A great portion (44%) of binding sites were found to be located near known tissue-specific ecdysone-responsive genes involved in metamorphic processes that did not represent Kc167 cells. Those results indicated that EcR-USP binds to a large proportion of the biologically-relevant genomic targets, but the cells respond only to a part of them, suggesting additional mechanisms to control ecdysone-specific response rather than the solely EcR-USP binding events.

The Stark lab extensively used ecdysone-stimulated Schneider 2 (S2) cells, the most common cell line in *Drosophila*, to establish its STARR-seq procedure. In (Shlyueva et al., 2014), they reported a motif enrichment analysis on ecdysone-induced enhancers that identified TF interplay necessary to establish the ecdysone response. Particularly, the TF serpent (srp) showed high enrichment in all ecdysone-induced enhancers as well as in regions with enriched EcR motif. Those results were also confirmed in luciferase reporter assays that carried mutated EcR and srp motifs of highly inducible enhancers. Notably, in another cell type, mutated srp motifs did not result in loss of reporter activity, strongly suggesting that cell-type-specific partner motifs may define target enhancers in each cell type.

In vivo, chromatin accessibility dynamics were monitored in the WD during metamorphosis (Uyehara et al., 2017). A motif enrichment analysis in dynamic DHSs resulted in the identification of E93 TF as putative regulator of accessibility induction and repression. Although *E93* is an ecdysone-responsive gene mainly implicated in SG histolysis (Lee et al., 2000), the authors could demonstrate its role in regulating ecdysone-dependent accessibility dynamics in WD through DHS probing of *E93* mutants.

Finally, EcR-dependent chromatin remodeling on EcREs was demonstrated to be essential in larval-to-prepupal transition (Badenhorst et al., 2005). Specifically, a whole genome expression analysis on null mutants of *Nurf301* (a subunit of the ISWI-containing chromatin remodeling complex NURF) showed clear down-regulation of all the ecdysone-responsive regulators and resembled the phenotypes of mutants in key downstream regulatory targets of EcR. Biochemical assays revealed that EcR and Nurf301 physically interact *in vivo* and their binding is ecdysone-dependent, providing new insights on the function of EcR in CRE chromatin structure through the recruitment of co-activators.

The ecdysone transcriptional cascade has always been a fundamental biological paradigm not only to study insect development, but also to better ascertain the role of nuclear hormone receptors in mammals. Traditional methods could not entirely cover the large transcriptional mechanisms and responses controlled by ecdysone, which involve a great number of TFs, effector genes, and co-factors in a context-dependent manner. Therefore, genome-wide approaches were essential to further decipher such a cascade and its impact on CRE structure and expression output. However, to date, a real comprehensive study that combined multiple high-throughput datasets in a very detailed spatio-temporal fashion is still missing: (1) studies that focused on chromatin

accessibility, transcript levels or TFBEs in cell lines were mostly conducted by analyzing one of those factors singularly, and only in very few cases those were combined. Furthermore, they all lacked a thorough time course of the ecdysone response, or monitored the cascade at very late time points, missing the early dynamics which represent the key mechanisms executed by ecdysone-responsive regulators; (2) *in vivo*, studies investigated CRE dynamics during metamorphosis either by using low-resolution assays, such as Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq), or by focusing on a single tissue, missing a reliable characterization of the tissue-specific ecdysone response or a comparative analysis of TFBEs that determine distinct morphological outputs.

Thus, despite much progress achieved in recent years, the ecdysone regulatory cascade still needs further elucidation. Taking into account the fundamental role of multiple TFs in propagating the ecdysone signal, and the well-established connection between TFBEs and chromatin accessibility, it is clear that a deep characterization of the ecdysone-induced DHS landscape and its underlying TF motifs is the best approach to pursue. Additionally, a integrated analysis of ecdysone-regulated CRE activity, chromatin structure and expression output would better clarify how those mechanisms correlate and are finely controlled by the hormone. Finally, dissection of the ecdysone cascade has to be carried out with a very high spatio-temporal resolution, in order to comprehensively identify all the dynamics that participate to the process.

2 AIM OF THE THESIS

The ecdysone transcriptional cascade is a great paradigm to study gene regulation mechanisms with regard to developmental biology. However, considering the numerous TFs and target genes involved, the optimal strategy to investigate such a cascade is a systems biology approach. Moreover, as the response to ecdysone triggers a hierarchical activation and repression of regulators, as well as a context-dependent interpretation of the hormonal signal, a detailed spatio-temporal resolution must be achieved.

Therefore, this study aims at thoroughly dissecting the ecdysone cascade by pursuing a combinatorial strategy. Experimental and computational approaches were integrated, with focus on three levels of transcriptional regulation studied in a genome-wide fashion: (1) chromatin accessibility, which provides information on CRE activity and TFBEs; (2) nascent RNA levels, to measure expression output; (3) differential MNase-seq, to detect the occupancy of sensitive nucleosomes over CREs. In addition, to achieve a comprehensive spatio-temporal resolution, two complementary systems were chosen: (1) a time course of ecdysone-treated S2 cells, which represents a great paradigm to characterize the cascade at very early time points and with precise timing; (2) selection of four tissues during three stages of larval-to-prepupal transition, which show distinct morphological changes occurring during metamorphosis, and represent the best approach to study context-dependent response to ecdysone *in vivo*.

By integrating these experimental and computational approaches, this thesis investigates three main aspects regarding the ecdysone transcriptional cascade:

I.   Quantitative analysis of ecdysone-triggered dynamics of chromatin accessibility, expression output, and TFs in S2 cells.

II.   Characterization of ecdysone-triggered chromatin structure changes and their association to the cascade components in S2 cells.

III.   Context-dependent chromatin accessibility dynamics and TF lexicon *in vivo.*

3   MATERIAL AND METHODS

3.1   Material

3.1.1   Cell line and culture

Name	Specification	Source (Catalog #)
Drosophila S2 cells	Single clone derived from late embryos	Förstemann lab
Express Five SFM	Protein-free, serum-free	Gibco (10486-025)
L-Glutamine	200 mM stock	Gibco (25030-081)
20-Hydroxyecdysone	10 mM stock	Roche (11376497001)

3.1.2   Enzymes

Name	Specification	Source (Catalog #)
DNase I	10 U/µl stock	Sigma-Aldrich (D4527)
MNase	0.5 U/µl stock	Sigma-Aldrich (N3755)
Tn5	Part of the Nextera DNA Library Preparation kit	Illumina (FC-121-1030)
Proteinase K	1 or 10 mg/ml stocks	Sigma-Aldrich (P2308)
RNase cocktail	Mix of RNase enzymes	Ambion (AM2286)
RQ1 RNase-Free DNase		Promega (M610A)

3.1.3   Antibodies

Name	Specification	Source (Catalog #)
Anti-EcR	Common for all the isoforms. 18 µg/ml stock. Host species: mouse	Developmental Studies Hybridoma Bank (Ag10.2)
Anti-br	Common for all the isoforms. 36 µg/ml stock. Host species: mouse	Developmental Studies Hybridoma Bank (25E9.D7)
Anti-Actin	Host species: mouse	Abcam (AB3280)
Anti-mouse IgG-HRP	HRP conjugated	Santa Cruz (sc-2055)

3.1.4 Additional commercial material

Name	Specification	Source (Catalog #)
cOmplete Protease Inhibitor cocktail	EDTA-free, tablets	Roche (04693159001)
cOmplete Lysis-M buffer		Roche (04719956001)
NuPAGE LDS Sample buffer	4x stock	ThermoFisher (NP0007)
2-Log DNA Ladder (0.1 - 10 kb)		NEB (N3200S)
SSO-fast Evagreen Supermix	2x stock	Bio-Rad (1725200)
QG buffer	Part of the MinElute PCR Purification kit	Qiagen (28004)
MinElute columns	Part of the MinElute PCR Purification kit	Qiagen (28004)
Elution buffer	Part of the MinElute PCR Purification kit	Qiagen (28004)
NEBNext Q5 Hot Start HiFi PCR Master Mix		NEB (M0543S)
AMPure XP beads		Beckman Coulter (A63880)
TRI Reagent		Sigma-Aldrich (T9424)
2x TD buffer	Part of the Nextera DNA Library Preparation kit	Illumina (FC-121-1030)

3.1.5 Kits

Name	Source (Catalog #)
iProof High-Fidelity PCR kit	Bio-Rad (1725330)
HiScribe T7 High Yield RNA Synthesis kit	NEB (E2040S)
Gel Extraction kit	Qiagen (28704)
PCR Purification kit	Qiagen (28104)
MinElute PCR Purification kit	Qiagen (28004)
Amersham ECL Prime Western Blotting Detection Reagent kit	GE Healthcare (RPN2232)
Pierce BCA Protein Assay kit	ThermoFisher (23225)
NEBNext Ultra DNA Library Prep kit	NEB (E7370S)
Nextera DNA Library Preparation kit	Illumina (FC-121-1030)
TURBO DNA-free DNase kit	Ambion (AM1907)

Ovation Human FFPE RNA-seq Library Systems   NuGEN (7150-08)

Bioanalyzer High Sensitivity DNA kit   Agilent (5067-4626)

Bioanalyzer DNA 1000 kit   Agilent (5067-1504)

3.1.6   Buffers

Name	Composition	Application
Nuclei Lysis buffer	10 mM Tris pH 8; 400 mM NaCl; 2 mM EDTA	gDNA extraction
Imaginal Disc buffer	60 mM KCl; 15 mM NaCl; 1 mM EDTA pH 8; 0.1 mM EGTA; 15 mM Tris pH 7.4; 0.15 mM spermine; 0.5 mM spermidine; 1 mM PMSF; 0.5% NP-40; Protease inhibitor cocktail	Nuclei prep from WD and ED
NPB buffer	20 mM MOPS; 40 mM NaCl; 90 mM KCl; 2 mM EDTA; 0.5 mM EGTA; 0.5% NP-40; 0.2 mM spermine; 0.5 mM spermidine; Protease inhibitor cocktail	Nuclei prep from CNS
Ringer's solution	0.123 M NaCl; 1.5 mM $CaCl_2$; 5 mM KCl; 0.2% sodium deoxycholate; 0.5% Triton X-100; Protease inhibitor cocktail	Nuclei prep from SG
NP-40 Lysis buffer	10 mM Tris pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$; 0.5% NP-40; 0.15 mM spermine; 0.5 mM spermidine; 0.5 mM PMSF; 2 mM benzamidine	Nuclei prep from S2 cells
DNase buffer A	15 mM Tris pH 8; 15 mM NaCl; 60 mM KCl; 1 mM EDTA pH 8; 0.5 mM EGTA pH 8; 0.5 mM spermidine; Protease inhibitor cocktail	DNase-seq
10x DNase I Digestion buffer	60 mM $CaCl_2$; 750 mM NaCl. Dilute to 1x in DNase buffer A	DNase-seq
DNase Stop buffer	50 mM Tris pH 8; 100 mM NaCl; 0.1% SDS; 100 mM EDTA pH 8; 1 mM spermidine; 0.3 mM spermine; 200 µg/ml Proteinase K	DNase-seq
MNase wash buffer	10 mM Tris pH 7.4; 15 mM NaCl; 60 mM KCl; 0.5% NP-40; 0.15 mM spermine; 0.5 mM spermidine; 0.5 mM PMSF; 2 mM benzamidine	MNase-seq

MNase digestion buffer	10 mM Tris pH 7.4; 15 mM NaCl; 60 mM KCl; 1 mM $CaCl_2$; 0.5% NP-40; 0.15 mM spermine; 0.5 mM spermidine; 0.5 mM PMSF; 2 mM benzamidine	MNase-seq
MNase stop buffer	50% (v/v) of 0.5 M EDTA pH 8.5; 50% (v/v) of 10% SDS	MNase-seq
Crush and Soak buffer	0.5 M ammonium acetate; 0.3 M sodium acetate pH 4.5; 0.1% SDS; 1 mM EDTA pH 8.5	MNase-seq

3.1.7   Primers

Primers for dsRNAs synthesis. T7 promoter sequence is in lower case

Name	Sequence (5' > 3')
EcR KD fw	taatacgactcactatagggTACGAAGAGCGCCGTCTACT
EcR KD rv	taatacgactcactatagggGCTCGCATGTCATAAGGTCA
br KD fw	taatacgactcactatagggGAATCTCCATCAGCGACAAG
br KD rv	taatacgactcactatagggACTGCTGCAACTGTGTGTTG

Primers for DNase-seq qPCR quality control.

Name	Sequence (5' > 3')
Act5c TSS fw	GGCTGCGGGACCAGTTTTCATATC
Act5c TSS rv	CGGCTTTGTGTCGGGAGGAGTATC
aTub84B TSS fw	CAAGCAAAGATTCACGCCCTGGTT
aTub84B TSS rv	CGCCGCATAACCGATAACTGAAGTG
Edg84A 3' UTR fw	GCCAGCGAAATCATCTGGAAGTGA
Edg84A 3' UTR rv	CCGAGACTCCGACTGGGACTT
Ems 3' UTR fw	GAATGCAGTCCAGTTCCAGTTATCG
Ems 3' UTR rv	CTAACGCCTTGGGATCGCTCTA

3.2 Methods

3.2.1 Cell biology, molecular biology, and biochemical procedures

Cell culture

Drosophila S2 cells were cultured in synthetic, serum-free Express Five medium (Gibco). 1 liter of Express Five medium was supplemented with 90 ml of 200 mM L-Glutamine (Gibco). Cells were thawed at passage 13 and cultivated until passage 20. During cultivation cells were grown at 25°C without CO_2 as semi-adherent monolayer in tissue culture flasks. When 90% confluent, cells were split into fresh flasks by means of seeding 0.8 x 10^6 cells/ml. Cell counting and assessment of cell viability were performed using the Cell Counter and Analyzer System (CASY; Roche).

Cell harvesting

S2 cells were harvested by centrifugation (500 g, 4°C, 5 min), then washed with 2.5 ml ice-cold 1x PBS and centrifuged again with the same parameters. Cell pellets were kept and supernatant discarded. If pellets were not directly utilized for experiments, they were shock frozen in liquid nitrogen and then stored at -80°C. The amount of initial cells to generate cell pellets was calculated according to the subsequent experiment.

Genomic DNA extraction

Genomic DNA (gDNA) was extracted from 25 x 10^6 S2 cells pellets. Briefly, cells were resuspended in 600 µl of Nuclei Lysis buffer. Then, 200 µl of 1 mg/ml Proteinase K (Sigma-Aldrich) and 80 µl of 10% SDS were added and mixed, followed by incubation at 55°C overnight. On the next day, 460 µl of 6 M NaCl were added, and samples were centrifuged (11,000 rpm, 30 min). Supernatant was recovered and centrifuged again (11,000 rpm, 10 min). Then, 1 volume of 100% ethanol was added and samples stored at -20°C for 1 hour. After, samples were centrifuged (11,000 rpm, 4°C, 30 min) and the resulting pellets were washed with 500 µl of 70% ethanol. Another centrifugation was carried out (11,000 rpm, 4°C, 10 min), then pellets were air dried and gDNA was resuspended in 75 µl of 0.1x TE buffer. Finally, 4 µl of RNase cocktail (Ambion) were added followed by incubation at 37°C for 30 min.

Ecdysone treatment

S2 cells were seeded 24 hours prior to cell treatment. On the next day, monolayer confluence was assessed, and if confluence was at least 80% cell treatment was performed. Cells were treated with

10 µM ecdysone (Sigma-Aldrich) and timing was started. For the DNase- and DTA-seq experiments (paragraph 4.1), treated cells were collected at 1, 2, 4, 8, and 12 hours after stimulus. For the MNase-, ATAC-, and RNA-seq experiments, including RNAi-based knockdowns (paragraph 4.2), treated cells were collected at 4 hours after stimulus. Untreated controls (UTC) were collected at the beginning of the time course.

Synthesis of dsRNA molecules for RNAi-based knockdown

Synthesis of dsRNA molecules was accomplished thanks to the suggestions of the Förstemann lab members. Briefly, primers were designed to target exonic regions and produce an amplicon of at least 250 bp. Amplification of target regions was performed by PCR using the iProof High-Fidelity PCR kit (Bio-Rad) following manufacturer's instructions. As DNA templates, either S2 cells gDNA or cDNA obtained from 3rd IL CNS were used, depending on whether the primer pair spanned an intronic region. PCR products were run on agarose gel and extracted with the Gel Extraction kit (Qiagen) according to manufacturer's instructions. *In vitro* transcription was performed with the HiScribe T7 High Yield RNA Synthesis kit (NEB) following manufacturer's instructions. Isolated PCR products were used as DNA templates up to 1 µg per reaction. After overnight incubation at 37°C, samples were treated with 1 µl of RQ1 RNase-free DNase (Promega), followed by serial incubations at 37°C for 30 min, 95°C for 10 min, and 65°C for 20 min. The obtained dsRNAs were diluted 1:10, 1:25, 1:50 and 1:100, and run on agarose gel, in order to quantify their concentrations. The 2-Log DNA Ladder (0.1 – 10 kb, NEB) was used as standard for quantification.

S2 cells treatment with dsRNAs for RNAi-based knockdown

Three days prior to collection, S2 cells were seeded and treated with 50 µg/ml dsRNA. Twenty-four hours prior to collection, S2 cells were seeded again and treated once more with 50 µg/ml dsRNA. On the next day, S2 cells were either collected for western blot validation or stimulated with ecdysone for high-throughput experiments as described above. For western blot validation, a dsRNA molecule targeting the luciferase gene (not present in wild type S2 cells) was kindly donated by the Förstemann lab, and used to treat cells for mock RNAi control.

Protein extraction and western blot

Pellets of 1×10^6 S2 cells were lysed with 300 µl of cOmplete Lysis-M buffer (Roche) implemented with cOmplete Protease Inhibitor cocktail (Roche). Lysed cells were incubated on

ice for 10 min and then centrifuged (14,000 rpm, 4°C, 10 min). Protein-containing supernatants were recovered. Protein concentration was assessed with the Pierce BCA Protein Assay kit (ThermoFisher) according to manufacturer's instructions. Western blot was performed following standard procedures. Briefly, 100 µg of proteins were mixed with 1x NuPAGE LDS Sample buffer (ThermoFisher) and 0.1 M DTT, and incubated at 99°C for 5 min. Protein separation was performed by SDS-PAGE electrophoresis using a 10% polyacrylamide gel. After run, proteins were transferred to a nitrocellulose membrane in transfer buffer for 1 hour at 300 mA. Membranes were blocked with 5% milk in TBS-T for 2 hours. Primary antibodies were diluted in blocking solution and incubated at 4°C overnight. On the next day, washes were performed with TBS-T. Subsequently, secondary antibodies were diluted in blocking solution and incubated for 1 hour. Then, additional washes were performed with TBS-T, followed by a final washing step in 1x PBS. Chemiluminescence was triggered with the Amersham ECL Prime Western Blotting Detection Reagent kit (GE Healthcare), and signal detected on films. Primary antibodies working dilutions: anti-EcR (DSHB Ag10.2, 1:50); anti-br (DSHB 25E9.D7, 1:50); anti-Actin (Abcam AB3280, 1:2500). Secondary antibody working dilutions: goat anti-mouse IgG-HRP (Santa Cruz sc-2055, 1:2500 for EcR and br, 1:5000 for Actin).

qPCR

qPCR was performed as quality control prior to library preparation during the DNase-seq protocol. 100 pg of DNA were used to set up a 10 µl PCR reaction containing 5 µl of 2x SSO-fast Evagreen Supermix (Bio-Rad) and 0.3 µl of both forward and reverse 20 µM primers. PCR was performed in a Bio-Rad CFX96 Real-Time System (Bio-Rad) using a 30 sec denaturation step at 95°C, followed by 40 cycles of 5 sec at 95°C and 5 sec at 58°C. Finally, a melting curve was generated in 0.5°C increments for 5 sec from 65 to 95°C.

3.2.2   Fly procedures

Larvae and prepupae staging

Wandering Early 3rd IL and Late 3rd IL were distinguished by adding 0.05% blue bromophenol in their food. In this way, Early 3rd IL guts are colored with high-intensity blue, whereas Late 3rd IL guts are basically colorless. Steady WPP were selected based on their anterior spiracle eversion and white, soft cuticle. Larvae were kept at 18°C to facilitate precise staging and tissue collection.

Tissue collection and nuclei preparation for ATAC-seq

Tissue collection and homogenization were performed by following distinct procedures. Each tissue had to be treated differently in order to obtain a preparation of intact nuclei for ATAC-seq applications.

40 pairs of WD or ED were dissected at each stage in cold 1x PBS. After dissection, tissues were collected in 1 ml of cold Imaginal Disc buffer. Then, tissues were homogenized by 5 strokes with a 25g needle, followed by 10 strokes with a 27g needle. Samples were centrifuged (500 g, 4°C, 7 min), and supernatant discarded. Nuclei pellets were kept on ice until resuspension in ATAC-seq transposase mix.

35 CNS were dissected at each stage in cold 1x PBS. After dissection, tissues were collected in 1 ml of cold NPB buffer. After a 5 min incubation on ice, tissues were transferred into a dounce homogenizer. 20 strokes with a loose pestle were applied, followed by a 10 min incubation on ice. Then, 15 strokes with a tight pestle were applied. After transferring the samples into a clean Eppendorf tube, 5 additional strokes with a 27g needle were applied. Samples were centrifuged (500 g, 4°C, 7 min), and supernatant discarded. Nuclei pellets were kept on ice until resuspension in ATAC-seq transposase mix.

18 pairs of SG were dissected at each stage in cold Ringer's solution. After dissection, tissues were collected in 200 µl of cold Ringer's solution. Tissues were incubated at room temperature for 5 min with gentle shaking. Then, tissues were homogenized by pipetting 20 times with a 1 ml tip. 800 µl of Ringer's solution were added, and samples were filtered through a 60 µm membrane. Samples were centrifuged (500 g, 4°C, 7 min), and supernatant discarded. Nuclei pellets were kept on ice until resuspension in ATAC-seq transposase mix.

3.2.3   High-throughput genome-wide procedures

DNase-seq

DNase-seq protocol was performed as described previously (Vierstra et al., 2014), with minor modifications. Briefly, nuclei were isolated from 50 x 10^6 S2 cells pellets. Pellets were washed with 10 ml of cold 1x PBS and centrifuged (500 g, 4°C, 7 min). Then, pellets were resuspended in 2 ml of NP-40 lysis buffer, incubated 5 min on ice, and centrifuged again (500 g, 4°C, 7 min). The obtained nuclei pellets were resuspended in 5 ml of DNase buffer A and centrifuged (500 g, 4°C, 5 min). Supernatant was discarded and nuclei pellets were kept on ice until DNase I treatment.

DNase I (Sigma-Aldrich) was diluted in 2.5 ml of DNase I digestion buffer to a final concentration of 25 U/ml. Nuclei were treated for 3 min at 37°C, and immediately after 2.5 ml of Stop buffer were added, followed by incubation at 55°C for 1 hour. Subsequently, 30 µl of RNase cocktail (Ambion) were added and samples were incubated at 37°C for 1 hour. After controlling the proper digestion level on agarose gel, samples were loaded on top of a 10-40% sucrose gradient and centrifuged at high speed (34,000 rpm, 20°C, 24 hours) in a SW40Ti rotor (Beckman Coulter). Fractions from the gradient were recovered by a fractionation machine (500 µl per fraction). DNA fragments size in each fraction was assessed by agarose gel, and all the fractions containing DNA fragments <500 bp were pooled. Three volumes of QG buffer (Qiagen) and 1 volume of isopropanol were added, DNA purified on MinElute columns (Qiagen) and finally eluted in 24 µl of Elution buffer (Qiagen). At this point, a quality control was performed by qPCR, in order to assess the enrichment of recovered fragments released from known open regions (TSS of *Actin5c* and *αTub84B* loci) over known closed regions (3' UTR of *Edg84A* and *ems* loci). Library preparation for Illumina sequencing was performed with the NEBNext Ultra DNA Library Prep kit (NEB) according to manufacturer's instructions, starting with 150 ng of DNA. After the adapter ligation step, a size selection was performed with the AMPure XP beads (Beckman Coulter) in order to enrich for fragments shorter than 150 bp. PCR amplification was carried out with 8 cycles. Final library purification was performed with AMPure XP beads. Library concentration and fragment size distribution were assessed by Bioanalyzer High Sensitivity DNA kit (Agilent). Biological duplicates were carried out.

MNase-seq

For MNase-seq, nuclei were isolated from 25 x 10^6 S2 cells pellets. Pellets were washed with 10 ml of cold 1x PBS and centrifuged (500 g, 4°C, 7 min). Then, pellets were resuspended in 1 ml of NP-40 lysis buffer, incubated 5 min on ice, and centrifuged again (500 g, 4°C, 7 min). The obtained nuclei pellets were washed once with MNase wash buffer without resuspending, and centrifuged (500 g, 4°C, 7 min). Then, pellets were resuspended in 4.8 ml of MNase digestion buffer and warmed up at 25°C for 5 minutes. For bulk MNase digest, 4 ml of resuspended nuclei were used. 7.5 U of MNase (Sigma-Aldrich) were added, and samples were incubated at 25°C for either 1 min (short digestion) or 3 min (typical digestion). The reaction was stopped with 400 µl of Stop buffer. Immediately, NaCl and sodium acetate pH 5.2 were added to a final concentration of 400 mM and 300 mM, respectively. DNA was isolated with the PCR purification kit (Qiagen)

according to manufacturer's instructions, and eluted with 30 µl of 0.1x TE. RNase treatment was performed with 12 µl of RNase cocktail (Ambion) at 37°C for 30 minutes. Samples were loaded on 3% agarose gel, and mono- and sub-nucleosomal fragments cut out. The collected gel pieces were smashed and covered with the Crush and Soak buffer and incubated at 37°C overnight. The solution was collected and DNA was purified with the PCR purification kit. Library preparation for Illumina sequencing was performed with the NEBNext Ultra DNA Library Prep kit (NEB) according to manufacturer's instructions. PCR amplification was carried out with 7 cycles. Final library purification was performed with AMPure XP beads. Library concentration and fragment size distribution were assessed by Bioanalyzer High Sensitivity DNA kit (Agilent).

ATAC-seq

ATAC-seq was performed as described previously (Buenrostro et al., 2013), with minor modifications. Briefly, nuclei were isolated from 3×10^5 S2 cells pellets. Pellets were washed with 50 µl of cold 1x PBS and centrifuged (500 g, 4°C, 7 min). Then, pellets were resuspended in 50 µl of NP-40 lysis buffer, incubated 5 min on ice, and centrifuged again (500 g, 4°C, 7 min). Supernatant was discarded and nuclei pellets kept on ice. At this point, procedure for ATAC-seq in S2 cells or *in vivo* tissues proceeded in the same way. Nuclei pellets were resuspended in 25 µl of Transposase mix, which included 6.25 µl of the Tn5 enzyme (Illumina) and 12.5 µl of 2x TD buffer (Illumina). Samples were incubated at 37°C for 30 min. Tagmented DNA was purified with the MinElute PCR purification kit (Qiagen), according to manufacturer's instructions, and eluted in 11 µl of water. 10 µl of tagmented DNA were used to amplify the library in a 50 µl PCR reaction. PCR reaction included also 3.125 µl of both barcoded and non-barcoded 20 µM customized primers (Buenrostro et al., 2013), and 25 µl of NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB). PCR parameters were as follows: 5 min at 72°C, 30 sec at 98°C, 12 cycles of 10 sec at 98°C, 30 sec at 63°C and 1 min at 65°C. Final library purification was performed with AMPure XP beads with a size selection to enrich for fragments shorter than 700 bp. Library concentration and fragment size distribution were assessed by Bioanalyzer High Sensitivity DNA kit (Agilent). In S2 cells, biological duplicates were carried out.

Total RNA-seq

Total RNA was isolated from 5×10^6 S2 cells pellets by using the TRI reagent (Sigma-Aldrich) according to manufacturer's instructions. DNase treatment was performed with the TURBO DNA-

free DNase kit (Ambion) according to manufacturer's instructions. Libraries were generated using the Ovation Human FFPE RNA-seq Library Systems (NuGEN) according to manufacturer's instructions. 100 ng of total RNA were used as starting material. For ribosomal RNA depletion, customized InDA-C primers specific for *Drosophila* were generated by Katja Frühauf as part of her PhD thesis (available in the faculty's archives), and used in this protocol. Libraries were amplified with 20 PCR cycles and assessed by Bioanalyzer DNA 1000 kit (Agilent). Biological duplicates were carried out.

DTA-seq

Any experimental procedures concerning DTA-seq on S2 cells were performed by Katja Frühauf as fundamental part of her PhD thesis. Briefly, nascent RNA was labeled using 200 µM 4sU, which was added to the cell culture medium for the last 60 min of each treatment time point. Then, 80 µg total RNA were biotinylated. Labeled (nascent) RNA was isolated with streptavidin-coated magnetic beads and subjected to high-throughput assays. For further details, I refer to Katja Frühauf's thesis available in the faculty's archives.

Next-generation sequencing

All the libraries were sequenced on an Illumina GenomeAnalyzer IIx in order to produce 50 bp pair-end reads. DNase-, MNase-, and ATAC-seq libraries resulted in 80 to 100 x 10^6 reads, whereas total RNA-seq libraries resulted in about 40 x 10^6 reads. Next-generation sequencing was performed by the LAFUGA sequencing facility at the Gene Center LMU Munich.

3.2.4 Computational procedures

Reads mapping

Unless specified otherwise, sequencing raw data were processed as follows, using a customized version of Galaxy (Giardine et al., 2005), available in the server of the Gene Center LMU Munich. After sequencing, reads were demultiplexed using the provided barcodes, the Illumina index read and the tool "Illumina Demultiplex". For each obtained sample, adaptors were trimmed using the tool "Clip adaptor sequence", with settings "Seed 5, Mismatches in adaptor 0, Minimum length after clipping 0, Output clipped and non-clipped one file". The files for each sample were downloaded from the Gene Center Galaxy and mapped locally using Bowtie2 version 2.2.9 (Langmead and Salzberg, 2012) with the following settings: "bowtie2 --quiet --local --very-

sensitive-local --threads 16 --mm -x /opt/bowtie2-2.2.9/indexes/dm3". Mapped reads were filtered for mapping quality and proper pairing using SAMtools 1.3.1 (Li et al., 2009) with the following settings: "samtools view -f 0x3 -q 10". Filtered reads were sorted and indexed using SAMtools 1.3.1.

DNase- and ATAC-seq peak calling in S2 cells

Peaks were called on each sample using MACS2 version 2.1.1 (Zhang et al., 2008), using a gDNA sample as control, with the following settings: "macs2 callpeak --keep-dup all -q 0.01 --nomodel --shift -100 --extsize 200 -f BAM -g dm –B". Cut sites in each peak for each time point were counted using "bedtools coverage" version 2.26.0, with the following settings: "bedtools coverage -sorted –counts". Cut sites were used to determine $\log_2$FC and adjusted p-value (p-adj) using the R/Bioconductor package DESeq2 (Love et al., 2014). Peaks with p-adj <0.01 compared to UTC were considered as differential.

Peak dynamics with ImpusleDE2

A common set of peaks for the entire time series was derived firstly by taking only peaks that were present in both duplicates for the same time point, using "bedtools intersect" version 2.26.0 (Quinlan and Hall, 2010). Then, all the remaining peaks in each time point were merged if an overlap occurred, using "bedtools merge" version 2.26.0. Cut sites in each peak for each time point were counted using "bedtools coverage" version 2.26.0, with the following settings: "bedtools coverage -sorted –counts". Differential peaks were called using the R/Bioconductor package ImpulseDE2 (Fischer et al., 2017) with an FDR threshold of 0.01. Differential peaks were classified into "Transition Up", "Transition Down", "Transient Up", "Transient Down" by ImpulseDE2.

DTA-seq and total RNA-seq data

Reads mapping was performed by Thomas Walzthöni. Subsequently, read counts were used to determine $\log_2$FC and adjusted p-value (p-adj) using the R/Bioconductor package DESeq2. Genes with p-adj <0.01 compared to UTC were considered as differential. For DTA-seq individual dynamics, differential genes were determined using the R/Bioconductor package ImpulseDE2 with an FDR threshold of 0.01. Differential genes were classified into "Transition Up", "Transition Down", "Transient Up", "Transient Down" by ImpulseDE2.

Peak-target gene association

Association between differential peaks and differential target genes was performed based on the distance of a differential peak from TSSs. A differential peak was assigned to the gene whose TSS had the shortest distance in bp. This operation was performed using the R/Bioconductor package ChIPseeker (Yu et al., 2015).

TF motif enrichment

Sets of *Drosophila* TFs were determined as follows. Firstly, all the TFs listed in (Pfreundt et al., 2009) were considered. Then, all the TFs belonging to the GO term "GO:0003700", which stands for "sequence-specific DNA binding transcription factor activity", were considered. Subsequently, those two sets were merged and used for TF selection. In S2 cells, TFs were selected according to the presence of a differential peak on their promoter and simultaneous differential expression at any time point. In larvae, TFs were selected according to the presence of a differential peak on their promoter at any stage, but keeping tissue-specificity. The presence of their PWMs was assessed in FlyFactorSurvey (Zhu et al., 2011) (http://mccb.umassmed.edu/ffs/), and Jaspar (Sandelin et al., 2004) (http://jaspar.genereg.net/). For sequence inputs, FASTA files containing the nucleotide sequences of differential peaks were obtained using bedtools getfasta version 2.26.0, using ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/dmel_r5.53_FB2013_05/fasta/dmel-all-chromosome-r5.53.fasta.gz as reference genome. FASTA files containing the control sequences were obtained using "fasta-shuffle-letters" from the MEME Suite version 4.11.2 (Bailey et al., 2009), with the following settings: "fasta-shuffle-letters -kmer 2 –dna". Motif enrichment was performed using AME from the MEME Suite version 4.11.2, with the following settings: "ame --method ranksum --scoring avg".

Gene Ontology

Gene Ontology (GO) analysis was performed by using a customized R script generated by Ivo Zeller as part of his Master's internship in the Gaul lab. The analysis is based on hypergeometric distribution and two-tailed hypergeometric test to assess under- and over-represented ontological gene categories compared to an equally sized random set. Ontological categories (GO terms) were obtained from databases such as KEGG, PANTHER or GO Consortium. For further details, I refer to Ivo Zeller's internship report.

MNase-seq data and mean coverage over differential DHSs

MNase-seq data were processed with a personalized script generated by Roberto Cortini. Briefly, data were normalized to 1x coverage. Then, all the fragments between 140 and 160 bp in length were considered as mono-nucleosomes, whereas all the fragments between 75 and 139 bp in length were considered as sub-nucleosomes. Dyads were calculated as the center of each fragment. Mean dyad coverage was calculated with a moving average of ±45 bp. Log_2 of mean dyad coverage ratios and mean coverage of ATAC-seq signal over differential DHSs (discussed in paragraph 4.2) were calculated and visualized with DeepTools (Ramírez et al., 2014), using the "computeMatrix" and "plotProfile" tools.

ATAC-seq peak calling in larvae

Peaks were called on each sample using MACS2 version 2.1.1 using a gDNA sample as control, with the following settings: "macs2 callpeak --keep-dup all -q 0.01 --nomodel --shift -100 --extsize 200 -f BAM -g dm –B". Differential peaks were called using the "macs2 bdgdiff" tool and following the instructions on the webpage https://github.com/taoliu/MACS/wiki/Call-differential-binding-events.

4 RESULTS

4.1 PART I: Genome-wide characterization of chromatin accessibility and expression output dynamics in ecdysone-stimulated S2 cells

Time series of *Drosophila* S2 cells upon ecdysone treatment represent an interesting and suitable paradigm to investigate the ecdysone cascade. S2 cells derive from a macrophage-like lineage of late-stage embryos. In culture, S2 cells are in proliferative state and undergo cell division every ~20 hours. When treated with ecdysone, the cells immediately respond by ceasing their proliferation, exiting the cell cycle and beginning a differentiation process, which morphologically comprises an increase in cell size and outgrowth of filopodia. At molecular level, ecdysone-treated S2 cells show the canonical cascade of regulators. Particularly, EcR protein levels are significantly induced already after 2 hours; similarly, br-Z1 protein levels increase after 2 hours, followed by an isoform switch at around 4-6 hours. Considering the very early time points at which ecdysone-regulated events occur in S2 cells, the Gaul lab established a time course to study transcriptional mechanisms in ecdysone-stimulated S2 cells, which goes from 1 to 12 hours. Thereby, to characterize the ecdysone-regulated dynamics of chromatin accessibility and expression levels, we also followed the same strategy (Figure 4A). Time points included 1, 2, 4, 8 and 12 hours after ecdysone treatment, together with an untreated control (UTC). At each time point, accessibility and expression were detected in a genome-wide fashion by carrying out DNase-seq and DTA-seq, respectively. Dynamic Transcriptome Analysis (DTA) is a recent approach to study RNA levels: cells are treated with 4-thiouridine (4sU), which is incorporated in the nucleoside salvage pathway (Miller et al., 2014). As eukaryotic RNAs do not contain thiol-groups, it is possible to isolate newly synthesized (nascent) 4sU-labeled RNAs, achieving a higher resolution of expression dynamics than total RNA sequencing. The DTA-seq protocol in S2 cells was established and performed by Katja Frühauf, a former PhD student in the Gaul lab. Nevertheless, the DTA-seq raw data were analyzed and handled differently in this thesis compared to her PhD thesis. On the contrary, the DNase-seq protocol in the Gaul lab was established in this study.

4.1.1 DNase-seq reliably identifies CRE activity and DHS dynamics in S2 cells

To investigate CRE activity induced by ecdysone genome-wide, DNase-seq protocol in S2 cells was established and its capability of detecting changes in DHSs assessed. In UTC, DNase-seq showed a high SNR, visible by a clear presence of restricted peaks compared to the genomic DNA control. SNR was high also upon ecdysone treatment. Additionally, after stimulus, changes in the DHS landscape occurred as well, with some peaks appearing and others disappearing when compared to UTC (Figure 4B). Those dynamic peaks are referred to as differential DHSs, which can be divided in opening and closing DHSs according to their signal at each time point with respect to UTC. To confirm the quality of the data, our DNase tracks were compared to enhancers detected with STARR-seq and DHS peaks obtained in the same study (Shlyueva et al., 2014). Our DNase-seq data identified activity of characterized enhancers with greater resolution (Figure S1), demonstrating that obtained DHS landscapes reliably captured true CRE activity.

As we were interested in ecdysone-responsive CREs, we then focused our attention on differential DHSs. Number of opening and closing DHSs increased almost linearly along the time course and with similar rates (Figure 4C). These findings indicate the involvement of new regulatory mechanisms as the cascade progresses, and a double role of ecdysone in governing chromatin accessibility, which equally comprises opening and closing of the chromatin. When differential DHSs were associated to specific genomic regions, significant differences between opening and closing DHSs manifested (Figure 4D). Opening DHSs mostly located in intronic regions (from ~60% at 1 hour to ~38% at 12 hours of the total opening DHSs), with lower localization in promoters (less than 20% at very early time points and up to ~30% at 12 hours). On the contrary, closing DHSs showed higher localization in promoters, especially at early time points (over 40% at 1 hour). As expected, very few differential DHSs located in exons or UTRs, but interestingly a similar small percentage located in intergenic regions. These findings are in agreement with the architecture of early genes loci, which is characterized by large intronic regions that harbor several ecdysone-responsive enhancers. The higher percentage of promoters in closing DHSs could be due to the switch off of alternative promoters necessary to transcribe early genes transcript variants after stimulus.

In summary, DNase-seq is a powerful and reliable technique to detect CRE activity and dynamics over a time course. Analysis on differential DHSs can be conducted in order to unravel the regulatory mechanisms triggered by ecdysone.
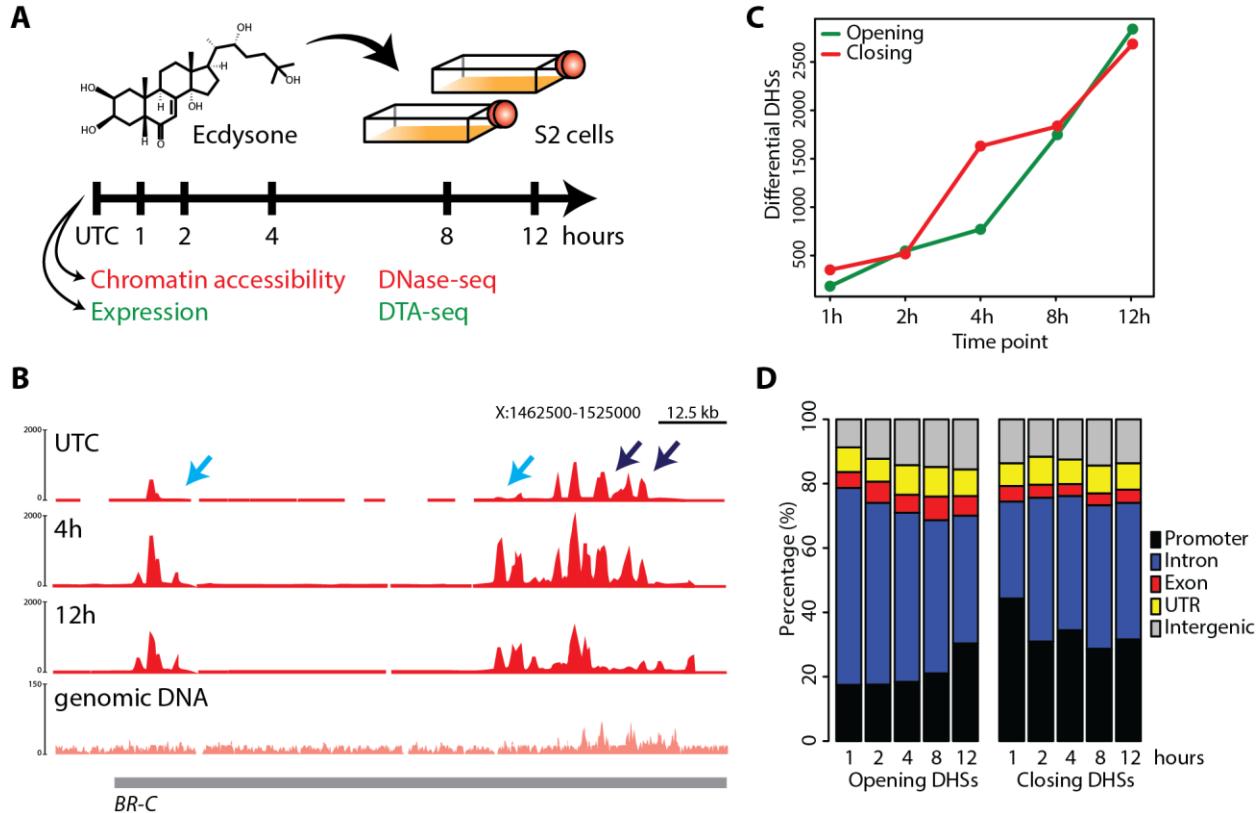
Figure 4: S2 cells are a valuable system for studying ecdysone transcriptional cascade in a genome-wide fashion. (A) Experimental strategy to dissect chromatin accessibility and expression dynamics in ecdysone-treated S2 cells. Cells were treated with 10 µM ecdysone and a time course was monitored. At each time point (including UTC), DNase-seq and DTA-seq were performed. (B) Genome browser screenshot of DNase-seq coverage tracks over the *br* locus. Tracks from UTC, 4 hours after stimulus (4h), 12 hours after stimulus (12h), and genomic DNA control are illustrated. Opening and closing peaks (DHSs) are indicated by light blue and dark blue arrows, respectively. (C) Number of differential DHSs over time course. Opening DHSs are indicated in green, whereas closing DHSs are indicated in red. (D) Genomic locations of differential DHSs over time course. Percentages are based on the total number of opening or closing DHSs at individual time points.

4.1.2 Differential DHSs and nascent RNAs correlate quantitatively

It is well known that changes in DHS activity affect transcription of their target genes. However, a systematic quantitative association between DHS and target gene fold changes (FC) has never been accomplished. This task is fundamental to better ascertain the role of activating and repressing CREs in relation to gene expression. The S2 cells experimental plan performed in this study is a perfect paradigm to study such correlation. Differential DHS FC can be quantified according to the DNase I cleavage events that occurred within a peak, directly reflecting the degree of single DHS openness (and consequently the differential levels of CRE activity). DTA-seq provides gene expression FC of nascent RNAs, without the noise of previously synthesized RNA molecules. Here, FC was assessed with respect to UTC. A DHS or gene was considered as

differential when its adjusted p-value (p-adj) was <0.01, independently from the $\log_2FC$ value. Only differential DHSs and genes will be considered hereafter.

Firstly, a comparison of FC between time points was conducted (Figure 5). When looking at adjacent time points, strong correlations were observed, whereas distant time points exhibited a more scattered FC distribution. Moreover, adjacent early time points displayed greater $\log_2FC$ for positive values, which correspond to opening DHSs or up-regulated genes. With the progression of the cascade, also negative FC (closing DHSs or down-regulated genes) increased. These results indicate that the ecdysone response is initially translated into fast activation of DHSs and gene expression. Along the time course, new mechanisms are involved, accompanied by presumably repressive actions characterized by increasing negative FC. Strikingly, those dynamics behaved similarly both in DHSs and genes, therefore that similarity was further examined.
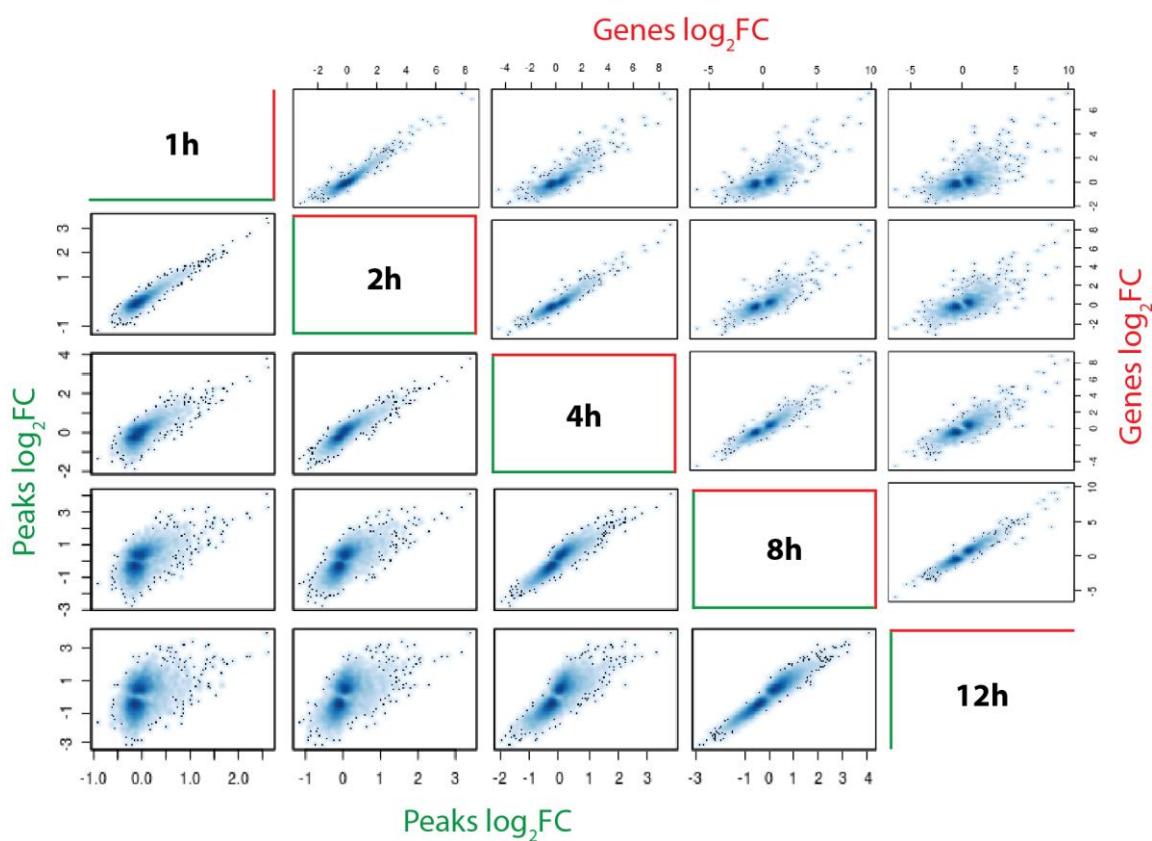
Figure 5: Ecdysone induces similar dynamics at accessibility and expression levels over time. Scatter plots of $\log_2FC$ comparisons between time points are shown. Peaks and genes $\log_2FC$ are illustrated. Peaks $\log_2Fc$ are indicated in green (bottom-left scatter plots). Genes $\log_2FC$ are indicated in red (up-right scatter plots). $\log_2FC$ are calculated with respect to UTC. Positive values represent opening DHSs or up-regulated genes. Negative values represent closing DHSs or down-regulated genes.
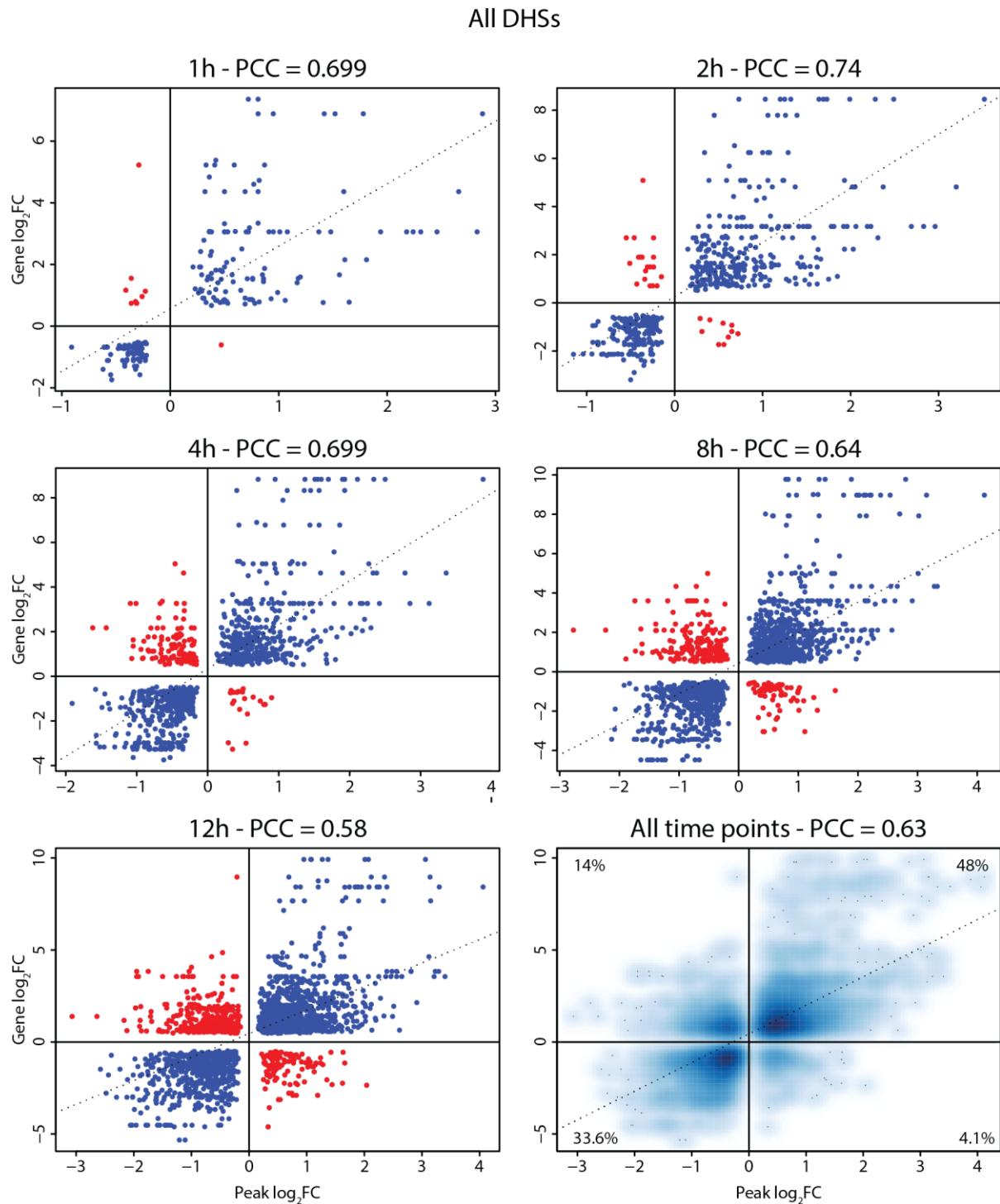
All DHSs

Figure 6: DHSs and target genes FC correlate quantitatively. Log$_2$FC of all differential DHSs (x-axis) and their associated target genes (y-axis) are illustrated. Each dot represents a DHS-target gene association. Scatter plots for individual time points and all time points (merged time points) are displayed. Dashed lines represent linear regression. Blue dots indicate opening DHSs associated to up-regulated genes (positive values), or closing DHSs associated to down-regulated genes (negative values). Red dots represent opening DHS associated to down-regulated genes, or closing DHSs associated to up-regulated genes. PCC values of correlation between DHS and gene log$_2$FC are illustrated.

Next, DHSs and target genes were associated and their FC correlated quantitatively. To associate DHSs to target genes, a common genomic distance approach was used: a differential DHS was linked to the nearest TSS, thereby assigning the corresponding gene as target. Interestingly, DHSs and target genes showed high Pearson Correlation Coefficient (PCC) in terms of $\log_2$FC values, when considering both time points singularly and all the time points together (for the latter PCC = 0.63) (Figure 6). PCC had similar values when only enhancers were taken into account (all time points PCC = 0.6) (Figure S2), and greater when only promoters were analyzed (all time points PCC = 0.75) (Figure S3). These data suggest a quantitative role of chromatin accessibility in regulating RNA synthesis of target genes, in which generally higher CRE activity leads to more transcription and vice versa. Particularly, this quantitative role seems to be governed mostly by promoters, as expected. In this regard, it is interesting to point out the distribution of closing DHSs with respect to their target genes. In promoters (all time points), the percentage of closing DHSs associated to up-regulated genes was 20% of the total closing DHSs, and their $\log_2$FC was never <-1.4. In enhancers (all time points), ~33% of total closing DHSs were associated to up-regulated genes, with $\log_2$FC up to -3.1, and consequent lower PCC than promoters only. Therefore, it seems that closing enhancers play a dual role in gene regulation.

Overall, a quantitative correlation between CRE activity and gene expression over a time course is demonstrated. Nevertheless, ecdysone-responsive closing enhancers that target up-regulated genes are a curious exception, which requires further elucidation.

4.1.3 Multiple opening or closing DHSs distinctly regulate gene expression

As many regulatory genes of the ecdysone pathway contain multiple CREs, the contribution of DHS number to gene expression FC was analyzed. Genes associated to 1, more than 1 and more than 2 differential DHSs were taken into account. Target genes associated to opening DHSs showed enhanced $\log_2$FC with the increase of DHS number (Figure 7A). More specifically, significant higher expression (p-value <0.05) was detected between genes linked to one opening DHSs and genes linked to >2 in all the time points. At 1 and 12 hours, significant $\log_2$FC increase was observed also in genes linked to >1 opening DHSs. When only enhancers were considered, a significant higher expression FC was detected in all the cases, except when comparing genes with one and more than one DHSs at 2 hours (Figure 7B).
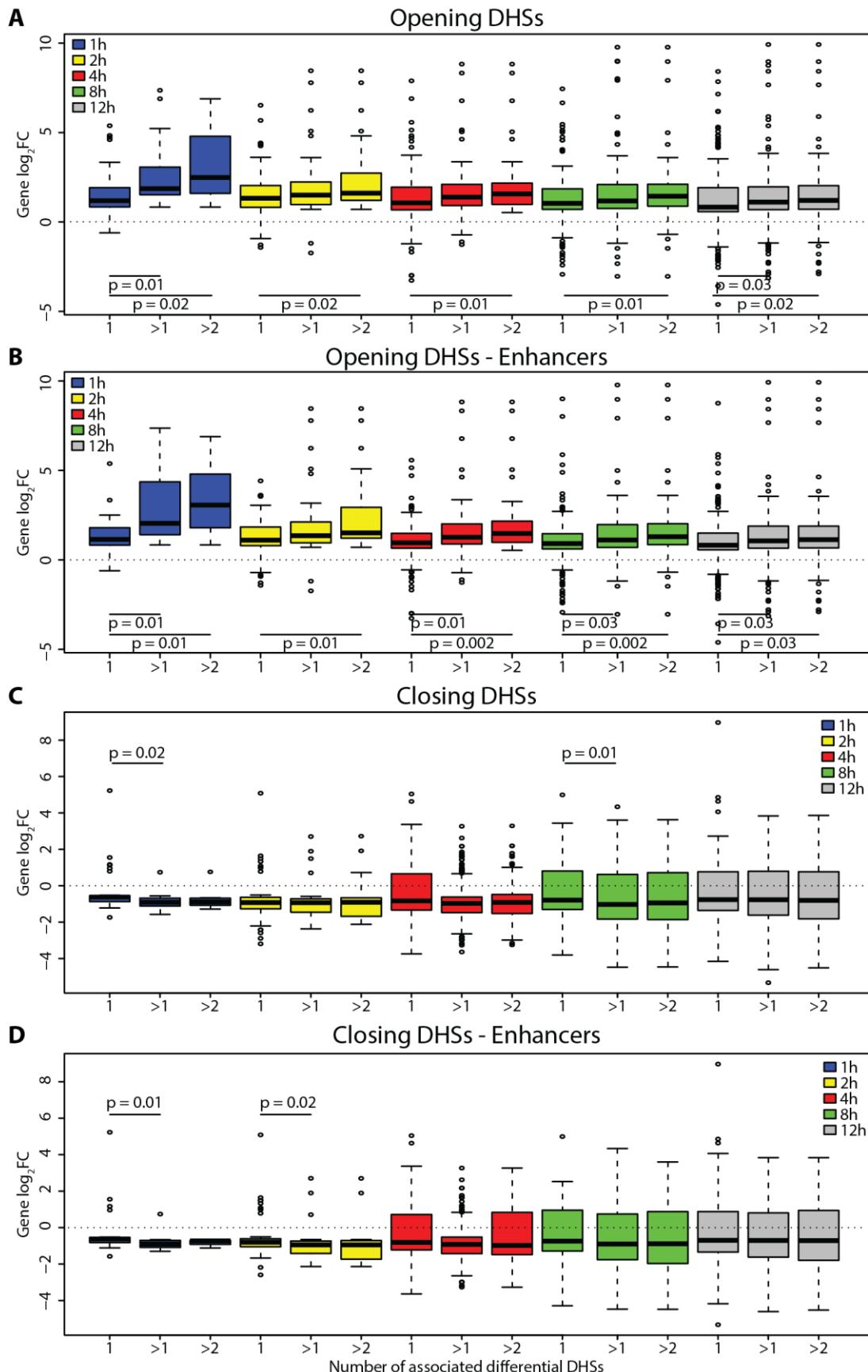
Figure 7 (previous page): Genes controlled by multiple opening DHSs show increased induction. Boxplots represent genes $\log_2$FC with respect to their association to 1, more than 1 (>1), or more than 2 (>2) differential DHSs along time course. (A) All opening DHSs are considered. (B) Only opening enhancers are considered. (C) All closing DHSs are considered. (D) Only closing enhancers are considered. Bottom and top of the boxes are 25th and 75th percentile, respectively. Thick lines in the boxes represent the median. P-values were calculated with respect to genes associated to 1 DHS (Wilcoxon rank-sum test). Only p-values <0.05 are illustrated.

A completely different scenario manifested when closing DHSs were examined (Figure 7C). A greater number of associated closing DHSs did not generally affect expression FC. Slight lower $\log_2$FC was observed between genes with one and more than one closing DHSs at 1 and 8 hours, as well as at 1 and 2 hours in enhancers only (Figure 7D).

In summary, these findings indicate that multiple opening DHSs have a synergetic function that enhances gene expression. On the contrary, more closing DHSs do not diminish (or increase) expression FC, also when eliminating promoters from the analysis, which rules out the repressed alternative promoters for transcript variants synthesis. This may be due to a more fine-tuning role of closing DHSs in response to ecdysone rather than mere transcriptional repression.

4.1.4 Individual DHS and gene dynamics manifest in four distinct behaviors and demonstrate multiple regulatory levels

To further decipher the dynamics of individual CREs and genes, the ImpulseDE2 algorithm was applied (Fischer et al., 2017). ImpulseDE2 models read count trajectories of each peak or gene with a descriptive single-pulse function over a time course. The outcome is a fit of read counts into 4 distinct categories: Transition up (Tn-U), Transition down (Tn-D), Transient up (Tt-U) and Transient down (Tt-D). Tn-D and Tn-U behaviors show continuous decrease or increase of counts levels, respectively. Alternatively, an initial decrease could be followed by return to initial or higher levels (Tt-D), or vice versa (Tt-U) (Figure 8A). Total sets of differential DHSs or genes were used to feed the algorithm (that is, peaks or genes from all the time points were merged into single larger datasets). This approach permitted a clustering of individual differential DHSs and genes according to their patterns. The largest part of them followed Tn-D or Tn-U dynamics (46% and 45% for DHSs; 41% and 50% for genes, respectively) (Figure 8B). When the dynamics were visualized in a heatmap, striking similarity emerged between accessibility and expression (Figure 8C). Firstly, Tn-U and Tn-D behaved with an analogous linear increment or reduction in z-scores. Secondly, and more notably, dynamics in Tt-U and Tt-D switched their trends at 4 hours, when looking both at DHSs and genes. This is in line with the isoform switch occurring in the *br* locus

in S2 cells, which is one of the main processes that determines the propagation of the ecdysone cascade. Therefore, it is possible that the cascade progression is mirrored in an inversion of CRE and expression activity, although limited to a small group.
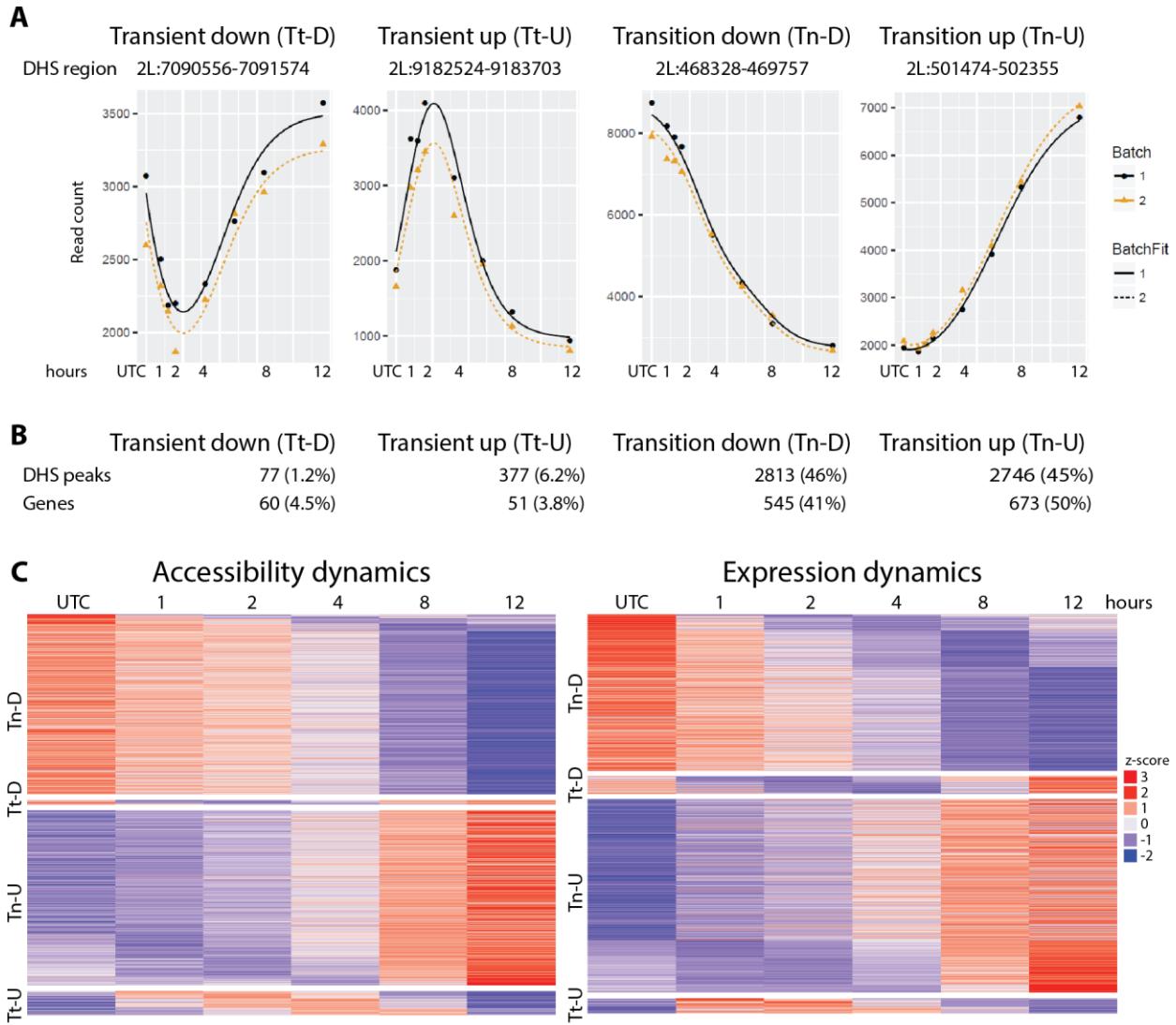
Figure 8: ImpulseDE2 categorizes individual accessibility and expression dynamics into four distinct patterns. (A) Examples of the fits modeled by the algorithm, which are characteristic of the four patterns. From left to right: Transient down (Tt-D); Transient up (Tt-U); Transition down (Tn-D); Transition up (Tn-U). Each plot displays read counts (y-axis) of a differential DHS (genomic region indicated on top) over time (x-axis). Read counts per time point (dots) and their fits (lines) are shown for two DNase-seq replicates (black and yellow marks). (B) Distribution of total differential DHSs or genes into the four categories. Absolute numbers and relative percentages compared to the entire set are shown. The largest part of identified differential DHSs and genes follows Tn-D and Tn-U dynamics. (C) Z-score heatmaps of individual DHSs (left panel) and genes (right panel) over time.

Intrigued by the analogy of clustered dynamics occurring in DHSs and genes, we next proceeded with a quantitative measurement of similarity among categories. For this purpose, the

Jaccard index was employed. Jaccard index measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. To compare all the sets, DHSs were associated to target genes as described before (see paragraph 4.1.2). Visualized heatmaps of Jaccard index values further confirmed a general high correlation between accessibility and expression (Figure 9A-B). In other words, each accessibility category exhibited higher Jaccard index values when compared to its expression counterpart. This was true when observing both all the DHSs (Figure 9A) and enhancers only (Figure 9B). Noteworthy, Tn-U genes showed fair similarity also with Tn-D and Tt-U DHSs. This finding means that a set of closing and temporary opening DHSs controls up-regulated genes. Thus, up-regulated genes undergo several levels of gene regulation that comprise various temporal dynamics of opening DHSs (Tt-U and Tn-U) and a still undefined role executed by closing DHSs. The same analysis was conducted after selecting only differentially expressed TFs or differential DHSs associated to target TFs (Figure 9C-D). Very interestingly, the highest Jaccard index value came from Tt-U accessibility with respect to Tn-U genes (Figure 9C). Furthermore, values between same categories appeared to be less predominant than in the global analysis. A very similar picture emerged for enhancers only (Figure 9D). These results can be explained by different inputs that ecdysone-responsive TFs require during the cascade, which include initial usage of ecdysone-responsive enhancers that subsequently are shut down to permit successive regulatory feedbacks.

To further investigate which genes represented each category, a gene ontology (GO) analysis based on hypergeometric test was performed. Terms regarding biological processes were visualized (Figure S4). As expected, continuously up-regulated genes (Tn-U) were enriched for terms indicating morphological changes (such as anatomical structure morphogenesis, system development, and developmental processes) and terms associated to ecdysone pathway (such as response to stimulus and signal transduction). On the other hand, repressed Tn-D genes were highly enriched for terms that indicated re-arrangements of metabolism as the cells stop their proliferation state, such as glycolytic processes and generation of precursor metabolites and energy, even though their percentages compared to the total number of Tn-D genes were low. In parallel, terms corresponding to development and stimulus were significantly depleted in Tn-D. Interestingly, Tt-D genes showed terms enrichment similar to Tn-U genes, although with lower $\log_{10}$ p-values (Figure S4).
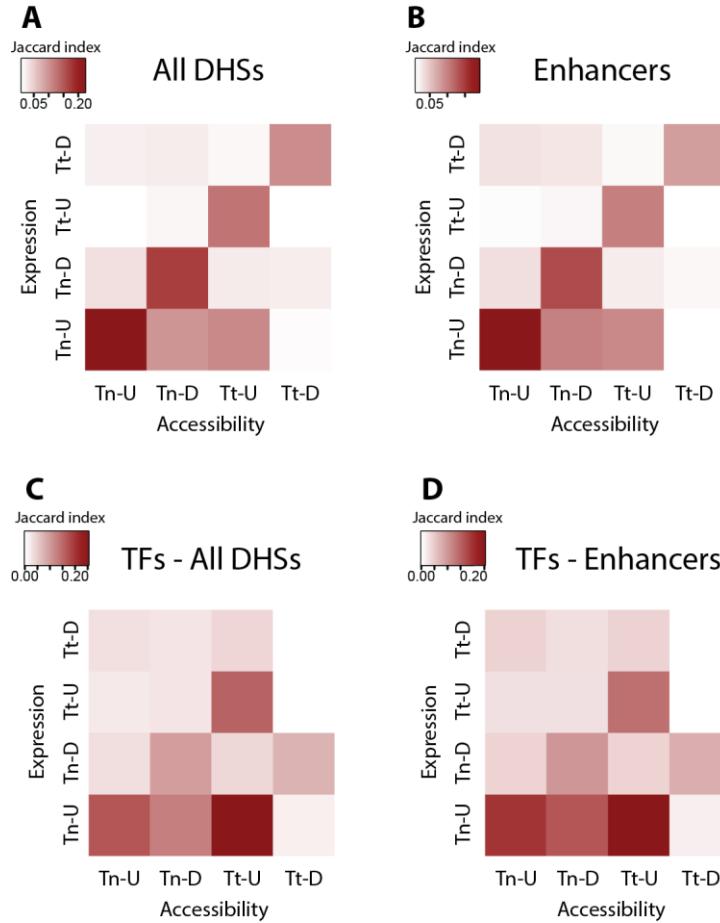
Figure 9: Dynamics measured by ImpulseDE2 demonstrate correlation between accessibility and expression, as well as complex regulatory mechanisms for Tn-U genes. DHSs were associated to target genes, and similarity between genes and DHSs categories were measured by Jaccard index. Jaccard index measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. Heatmaps of Jaccard index values are shown. DHSs dynamics (x-axis) were compared to genes dynamics (y-axis). (A) All DHSs are considered. (B) Only enhancers are considered. (C) TFs and all DHSs associated to a TF locus are considered. (D) TFs and enhancers associated to a TF locus are considered.

GO for DHSs revealed a less strict divergence (Figure S5). Terms regarding morphological changes and response to a stimulus were highly enriched in Tn-U, comparable to what was observed in Tn-U genes. Additionally, the same terms were also fairly represented in Tn-D and Tt-U, in line with the Jaccard matrix values. However, terms regarding metabolic processes were neither strikingly enriched nor depleted in any category. These results further confirm the involvement of different DHS dynamics in regulating up-regulating genes, which most likely correspond to ecdysone-responsive regulatory and effector genes, as shown by GO analysis.

Taken together, these findings illustrate that individual DHSs and genes follow four distinct behaviors along the time course, implied by ImpulseDE2. As discussed in the previous paragraphs,

a general correlation between accessibility and expression in terms of dynamics similarity is demonstrated. However, up-regulated genes exhibit a more complex regulation, as they are targeted by CREs with different temporal and functional activity (Tn-U, Tt-U and Tn-D). Those genes correspond to the effectors of morphological changes occurring in S2 cells, hence suggesting that the developmental processes triggered by ecdysone require complex regulatory mechanisms.

4.1.5 TF motif characterization of the ecdysone transcriptional cascade

After thoroughly dissecting ecdysone-regulated dynamics, we next asked which are the main TFs involved in their regulation. To address this question, a deep motif enrichment analysis of TFs with known PWMs was applied. TFs that had differential expression and simultaneously showed a differential DHS on their promoters in any time point were selected. Then, the presence of their PWMs in various databases was checked. The final set is listed in (Table 1). Motif enrichment was run with the AME algorithm (McLeay and Bailey, 2010). AME identifies known motifs that are relatively enriched in the input sequences compared to a shuffled control. Input sequences were DHSs computed by ImpulseDE2, maintaining the same 4 categories (see previous paragraph).

Notably, only 3 TFs were enriched in all 4 categories: hinge3 (hng3), a putative repressor of the Wg pathway involved in the wing-hinge development, br-Z1 and srp. Hng3 and br-Z1 in Tn-D revealed the two most significant values of all the analysis. Instead, srp displayed more enrichment in Tn-U (Figure 10). Remarkably, two other isoforms of br, br-Z2 and br-Z3, showed enrichment in closing Tn-D DHSs, with Z2 isoform showing greater significance than Z3. The presence of br in closing peaks was demonstrated also when a distribution of TFBSs in Tn-U and Tn-D DHSs was carried out (Figure S6).

Not surprisingly, the heterodimer EcR-USP was present only in Tt-U, as its regulatory function acts at the very beginning of the cascade. Other interesting outcomes were represented by pannier (pnr), a TF involved in imaginal discs and neural developmental processes, foxo, a regulator of the insulin pathway, and the uncharacterized TF CG5953. More specifically, pnr showed the highest enrichment in Tn-U, whereas foxo and CG5953 were more present in Tn-D. Several other TFs were identified in the analysis, for a total number of 20.
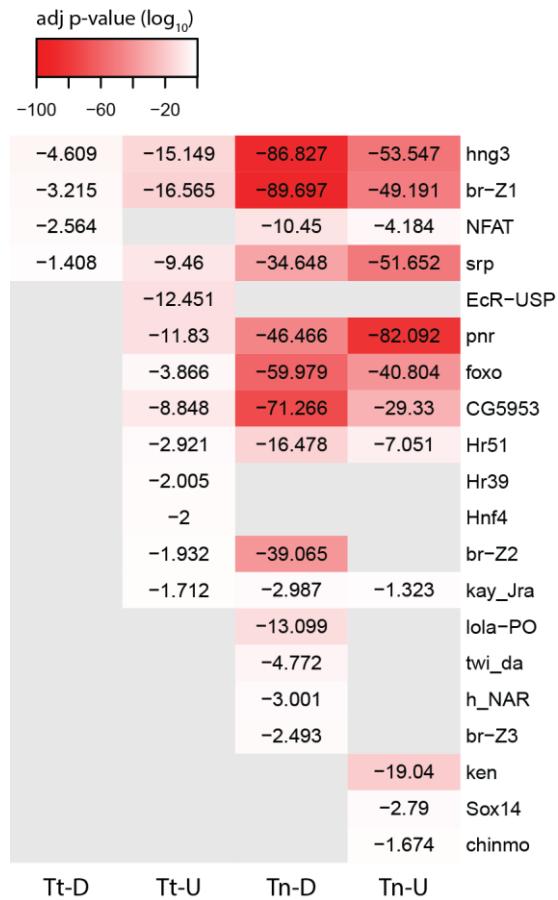
Figure 10: TF motif enrichment analysis reveals involvement of novel TFs and peculiar regulatory dynamics in the ecdysone cascade. TFs that had differential expression and simultaneously showed a differential DHS on their promoters in any time point, were selected for the analysis. As genomic inputs, differential DHSs categorized by ImpulseDE2 were used. Heatmap shows TF adjusted p-values ($\log_{10}$) as measure of motif enrichment in each DHS set. Corresponding values are displayed.

These data provide new insights on TFs involved in the ecdysone transcriptional cascade. Firstly, TFs that were not previously associated to ecdysone were characterized here. Among them, pnr, hng3 and CG5953 seem to be fundamental for the pathway dynamics, as they manifest high enrichment in at least 3 distinct DHS actions. Furthermore, the role of srp is highlighted. Srp was previously reported to bind target sites flanking EcR-USP motifs in an *Fbp1* enhancer in the fat body (Brodu et al., 1999), and to be required for activation of EcREs in S2 cells (Shlyueva et al., 2014). Our results not only confirm its contribution to the ecdysone response, but also propose that srp may be a key factor, considering its very significant presence in all CRE dynamics. Finally, our findings illustrate a putative function of br in closing chromatin. Particularly, Z2 and Z3 isoforms are enriched in closing peaks, whereas Z1 reveals enrichment in all dynamics, although with greater presence in Tn-D, suggesting multiple functions. Br is well known to be a key

regulator of the ecdysone cascade, and was already reported to function as a repressor (Karim et al., 1993; Lin et al., 2011). Additionally, br-Z2 was shown to directly act as enhancer repressor in larval fat body (Mugat et al., 2000). Thus, this interesting role of br deserves further elucidation.

4.2   PART II: Chromatin structure in ecdysone-responsive DHSs

Nucleosomes cover TFBSs on CREs, thus competing with TFs for DNA binding and determining the degree of chromatin accessibility. MNase-seq is the standard experiment to map nucleosomes genome-wide. Recently, it was shown that different nucleosome landscapes are obtained depending on the MNase digestion level. This information was used to characterize nucleosomes by their differential sensitivity to MNase titration in *Drosophila* (Chereji et al., 2015; Mieczkowski et al., 2016) (see paragraph 1.1.6). Hence, differential MNase-seq is used as a probe for measuring nucleosome sensitivity, where MNase-sensitive nucleosomes are enriched in active CREs. Nevertheless, this approach is highly dependent on the general accessibility of nucleosomes and chromatin environment to MNase activity, which could include longer linker DNA and higher-order chromatin structures.

Alessio Renna, a PhD student in the Gaul lab, established a new MNase-seq approach to study nucleosome fragility. It implied the selection of sub-nucleosomal (>75 bp and <140 bp of length) and mono-nucleosomal fragments in a typical digestion condition (Renna et al., in preparation) (Figure S7A). In this manner, the strength of nucleosome binding was measured through the transition of each nucleosome into sub-nucleosomal fragments, which mostly reflected the accessibility of nucleosomal DNA to MNase due to biophysical properties, DNA features and TFs/chromatin remodelers. Renna and colleagues applied a score on single called nucleosomes that compared sub- and mono-nucleosomal fractions, and subsequently measured nucleosome fragility. Additionally, they reported that fragile nucleosomes on active enhancers mapped at the center of the DHS peak in an activity-driven fashion.

Intrigued by the compelling evidence on MNase-sensitive nucleosomes in DHSs and their role in transcriptional regulation, we investigated this topic in the ecdysone-stimulated S2 cells paradigm. As initial step, DNase-seq tracks at UTC and 4 hours after ecdysone stimulus were compared to tracks produced by Renna of sub- and mono-nucleosomal fragments in typical digestion ("Sub typ" and "Mono typ", respectively), and mono-nucleosomal fragments in lower digestion ("Mono short") (Figure 11A). Active constitutive DHSs showed greater Sub typ and Mono short fragment coverage detected at UTC. On the contrary, opening DHSs had higher Mono typ coverage detected at UTC.
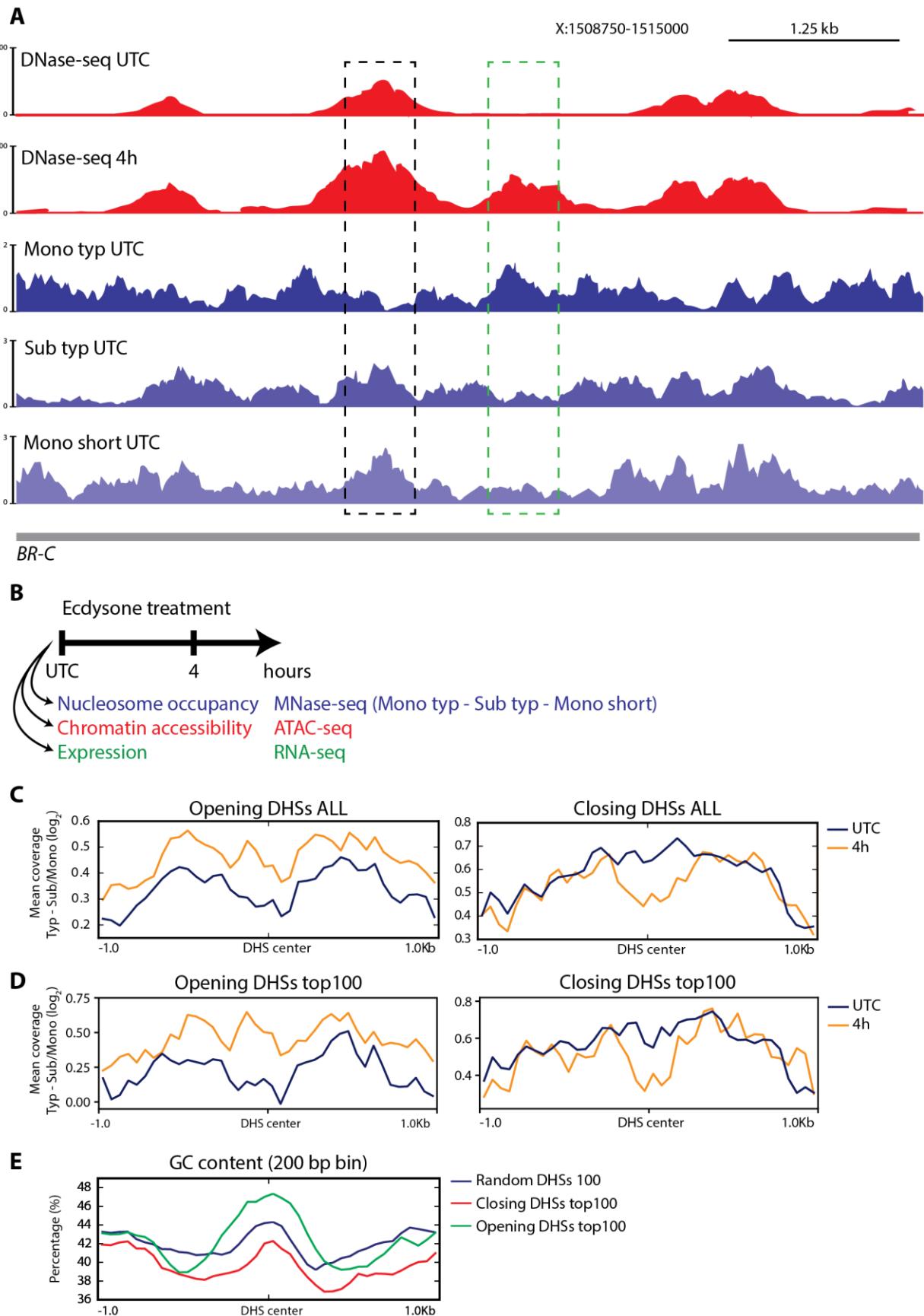
Figure 11 (previous page): Ecdysone triggers changes in MNase-sensitivity over inducible DHSs. (A) Genome browser screenshot of DNase-seq and MNase-seq coverage tracks over the *br* locus. From top to bottom: DNase-seq at UTC; DNase-seq at 4 hours after stimulus (4h); Mono typ at UTC; Sub typ at UTC; Mono short at UTC. Dashed black box indicates constitutive active CRE that shows higher Sub typ and Mono short occupancy at UTC. Dashed green box indicates opening DHSs that shows higher Mono typ occupancy, as well as Sub typ and Mono short depletion at UTC. (B) Experimental strategy to investigate ecdysone-triggered MNase-sensitivity changes and their relations to chromatin accessibility and expression. Two time points were selected (UTC and 4h), and MNase-seq, ATAC-seq and RNA-seq were performed at each time point. (C-D) Log_2 ratio between Sub typ and Mono typ mean dyad coverage in opening DHSs (left panels) and closing DHSs (right panels). Ratios were plotted over opening all DHSs and closing all DHSs (C), and over top100 opening DHSs and top100 closing DHSs (D). Ratios at UTC (blue lines) and 4h (yellow lines) are shown. Regions ±1 kb away from the DHS center were considered. (E) GC content in top100 opening DHSs (green line), top100 closing DHSs (red line), and 100 random constitutively active DHSs (blue line).

This observation further suggests that active CREs are characterized by the presence of MNase-sensitive nucleosomes, whereas closed CREs, although potentially highly inducible, contain non-sensitive nucleosomes.

Thus, what happens to the MNase-sensitivity levels in differential DHSs after treatment? As the ecdysone cascade is known to change chromatin structure in EcREs, are also MNase-sensitive nucleosomes affected by the stimulus in those regions?

4.2.1 Nucleosomes change their MNase-sensitivity on differential DHSs in an activity-driven manner

To address the question regarding MNase-sensitivity in differential DHSs, S2 cells were treated with ecdysone and two time points were selected. Together with UTC, we opted for 4 hours after stimulus (4h), as this time point captures both very early and progressive dynamics. At each time point, three genome-wide approaches were performed (Figure 11B): ATAC-seq to probe chromatin accessibility; MNase-seq with the selection of Sub typ, Mono typ and Mono short fragments; total RNA-seq. ATAC-seq was chosen over DNase-seq for its faster procedure. Thereby, ATAC-seq protocol was established in S2 cells and its reliability compared to DNase-seq both at the coverage and peak detection levels (Figure S7B), with successful results. Differential DHSs were assessed and divided in opening and closing.

To measure the degree of MNase-sensitivity, two different log_2 ratios of mean dyad coverages were applied: (1) ratio between Sub typ and Mono typ dyad coverage, which resembles the approach applied by Renna and colleagues. This procedure reduced the contribution of features mostly driven by the accessibility of the chromatin environment to MNase; (2) ratio between Mono short and Mono typ dyad coverage, which is based on the more widely applied differential MNase-

seq to probe MNase-sensitive nucleosomes (Mieczkowski et al., 2016; Xi et al., 2011). This procedure resembles the classification of MNase-sensitive regions established in a previous study in *Drosophila* (Chereji et al., 2015). Few computational corrections compared to Renna's fragility score and Chereji's classification had to be applied, due to the lower number of analyzed genomic regions. That led to the establishment of our $\log_2$ ratios. Nevertheless, our procedures are valid approximations for measuring susceptibility to the enzyme, and reflect the properties that characterize MNase-sensitive nucleosome mapping. Thus, we strongly assume that higher $\log_2$ ratio values identify sensitive nucleosomes. As the term fragility indicates nucleosomes with a differential stability due to multiple features (not analyzed in this study), hereafter I will only use the term MNase-sensitivity, which reflects the bias introduced by differential MNase-seq to study chromatin related features.

Firstly, Typ Sub/Mono coverage ratio was plotted over opening and closing DHS regions (Figure 11C). In opening DHSs (illustrated in left panels hereafter), enhanced sensitivity appeared at the center of the peak upon stimulus, very likely indicating the formation of an MNase-sensitive nucleosome, where sensitivity was not detected before treatment. Noteworthy, a higher sensitivity was detected also in regions directly flanking the peak center, both before and after the treatment, with a slight increase after ecdysone. In closing DHSs (illustrated in right panels hereafter), the opposite situation was observed. Greater sensitivity was detected on active DHSs before the stimulus. However, as the DHSs closed in response to ecdysone, a nucleosome with lower sensitivity was measured at the center of the peak. In this case, higher values in flanking regions compared to the center were not present.

Next, to better ascertain the association between nucleosome sensitivity and CRE activity, we restricted our analysis to the most inducible DHSs by selecting the top100 opening and top100 closing DHSs (that is, opening and closing DHSs that had the 100 greatest FC). The outcome was a larger sensitivity difference in the DHS center upon treatment, with the same patterns observed in all DHSs (Figure 11D).

Thus, these data demonstrate changes in MNase-sensitivity, which very likely affects a nucleosome at the center of EcREs and is driven by ecdysone. The level of susceptibility seems to be activity-dependent, as very inducible regions display greater difference in Sub/Mono ratio upon treatment.

To exclude possible biases, GC content was examined. It is known that underlying DNA GC content well correlates with nucleosome occupancy, most likely affecting histone-DNA interactions (Tillo and Hughes, 2009). Generally higher GC content (within a certain degree) favors nucleosome packaging. Additionally, MNase-sensitivity was reported to be associated with underlying DNA features, such as low GC content. Therefore, GC content in top100 differential DHSs and in 100 constitutive active random DHSs was calculated (Figure 11E). A general pattern was observed, characterized by a peak of GC content at the center of DHSs flanked by a GC depression. However, higher GC content was found at the DHS center in top100 opening DHSs. The size of GC peak in those regions was evidently larger than in random 100 and top100 closing DHSs, in accordance with the intrinsic favoring of nucleosome occupancy, most likely to mask TFBSs and prevent undesired binding events in absence of ecdysone. Regions flanking the opening DHS center, instead, had lower GC content than random DHSs, which could explain the higher MNase-sensitivity observed in flanking regions both before and after treatment. On the other hand, in top100 closing DHSs, the GC level was lower all across the examined genomic regions. This is in line with the observed increased MNase-sensitivity at UTC (Figure 11D), and characteristic of the active state of those CREs in untreated conditions. Finally, constitutive active DHSs (random 100) showed a GC content more similar to the average *Drosophila* genome (43%).

Overall, these results indicate that underlying GC content varies according to CRE dynamics. In absence of ecdysone, higher GC content favors nucleosome occupancy and diminished MNase-sensitivity in opening DHSs, whereas in closing DHSs a lower GC content drives susceptibility to MNase and most likely TFBEs. After stimulus, these intrinsic DNA features are overcome: very likely, nucleosomes located at the DHS center undergo a change towards higher MNase-sensitivity in opening DHSs, and vice versa in closing DHSs. Those mechanisms are driven in an activity-dependent manner. As the underlying DNA favors MNase-sensitivity levels detected in UTC, the changes occurring at 4 hours must be due to *trans*-acting factors, which could include TFs.

4.2.2   TF motif enrichment reveals EcR and br as candidates for changes in MNase-sensitivity

To investigate whether any components of the ecdysone cascade were involved in the changes in MNase-sensitive nucleosomes, we applied a TF motif enrichment analysis as described before

(paragraph 4.1.5). The same set of TFs was used. As genomic inputs, opening and closing DHSs were divided in 4 distinct sets to deeply dissect activity-correlation: all DHSs, top100 all DHSs, top100 enhancers, and bottom100 DHSs.

In opening DHSs, a striking pattern was visible (Figure 12A). All DHSs showed high enrichment for few TFs, in particular hng3, pnr, srp (as observed in Figure 10), and the heterodimer EcR-USP. Nevertheless, as the analysis focused on top100 DHSs or enhancers, EcR-USP clearly emerged as the most enriched TF. Additionally, the second most enriched TF in highly inducible regions was EcR alone. Notably, the presence of EcR-USP in bottom100 DHSs was not detected. Thus, it is possible that additional factors may play a role in chromatin structure changes, although through a more general action. In highly inducible CREs, however, the mechanisms that drive such response could be attributable mostly to EcR.

In closing DHSs, br-Z2 was predominant in 3 distinct DHS sets, being the most enriched TF in all DHSs, top100 all DHSs and top100 enhancers (Figure 12B). Its motif was detected also in bottom100 DHSs, although to a lesser extent. Again, these findings are in accordance with data illustrated in Figure 10. Thus, br, and most likely its Z2 isoform, plays a role in closing chromatin, presumably by promoting lower nucleosome sensitivity levels. In addition, it is possible that TF cooperativity is required for chromatin structure changes also in closing DHSs, as srp and pnr manifested enrichment in all the sets but bottom100 DHSs. Remarkably, the EcR-USP heterodimer was not detected in closing DHSs.

In summary, TF enrichment analysis reveals putative candidates that may regulate chromatin structure in response to ecdysone. As additional evidence to our hypothesis, it was previously reported that EcR participates in chromatin structure regulation through recruitment of additional factors (Badenhorst et al., 2005; Kreher et al., 2017). Similarly, br has a pivotal role as regulator of the cascade, and its action as a repressor was already mentioned (paragraph 4.1.5). Considering all these observations, EcR and br were further investigated.
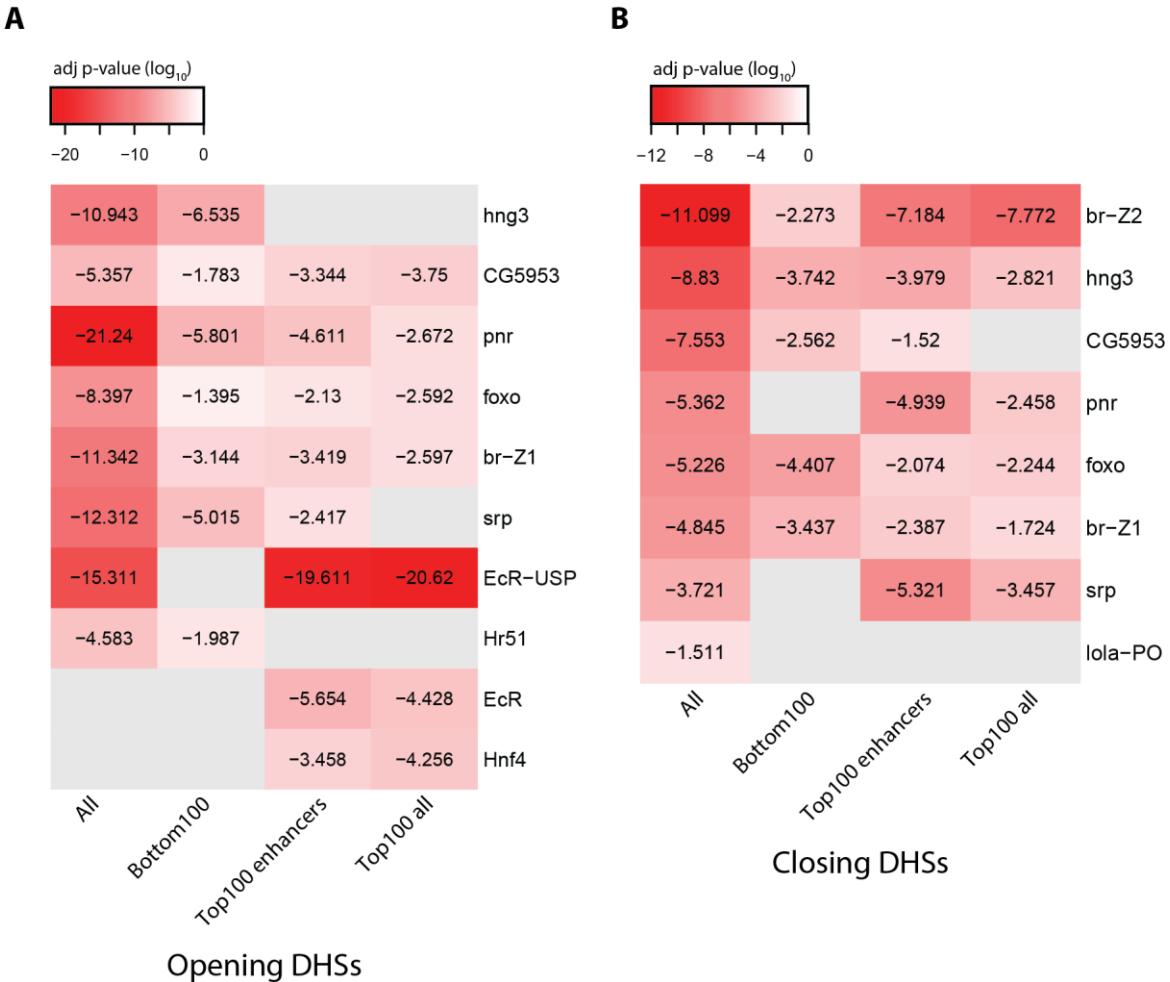
Figure 12: TF motif enrichment analysis reveals EcR and br as putative regulators of MNase-sensitivity changes. (A) TF motif enrichment was assessed in opening DHSs divided in all DHSs, bottom100 DHSs, top100 enhancers, and top100 all DHSs. (B) TF motif enrichment was assessed in closing DHSs divided in all DHSs, bottom100 DHSs, top100 enhancers, and top100 all DHSs. Heatmaps show TF adjusted p-values (log_{10}) as measure of motif enrichment in each DHS set. Corresponding values are displayed.

4.2.3 EcR and br knockdowns result in MNase-sensitive nucleosomes and chromatin accessibility alterations

Prompted by the data obtained in the enrichment analysis, we performed an RNAi-based knockdown in S2 cells of EcR and br. S2 cells are highly susceptible to RNAi after treatment with large double-stranded RNAs (dsRNAs) (Rogers and Rogers, 2008). Thus, dsRNA molecules were designed to knockdown all the isoforms. For each gene, at least four independent dsRNA molecules were generated, targeting different exonic regions. The efficiency of knockdown was validated by western blot in comparison to a mock RNAi in untreated and treated conditions. The most efficient dsRNAs are shown in Figure 13A. Those were used to ascertain EcR and br roles

in chromatin structure changes upon ecdysone stimulus, following the same experimental and computational strategy described before (Figure 11B). Dyad coverage ratios illustrated in Figure 11 will be also shown as comparison (no RNAi control – CTL).

Firstly, EcR RNAi will be discussed. Interestingly, we observed several alterations in MNase-sensitivity (Figure 13B). In top100 opening DHSs, increased sensitivity was noticed in EcR RNAi compared to CTL before treatment (UTC), specifically at the level of the central nucleosome and the flanking nucleosome upstream of the DHS center. At 4 hours after treatment, the levels of MNase-sensitivity in the central nucleosome were clearly lower than CTL, whereas the flanking nucleosomes were as susceptible to MNase as in CTL. In top100 closing DHSs, MNase-sensitivity alterations revealed in UTC, where lower levels were detected at the center of the peak. Notably, at 4 hours no differences were identified between CTL and EcR RNAi. As the accessibility of the surrounding environment is also a feature that determines nucleosome sensitivity, the degree of openness was investigated. Thus, the mean coverage of ATAC-seq signal was analyzed (Figure 13D). Chromatin accessibility mirrored what depicted by differential MNase-seq: in top100 opening DHSs, regions showed decreased accessibility in EcR RNAi compared to CTL at 4 hours. In top100 closing DHSs, EcR RNAi resulted in lower ATAC-seq signal than CTL at UTC, whereas DHSs could close as well as CTL at 4 hours. These findings suggest a triple role of EcR with regard to MNase-sensitive nucleosomes: (1) in inactive opening DHSs, knockdown of EcR leads to higher sensitivity of the non-sensitive nucleosome at the DHS center, which could be explained by the repressive role of EcR reported in absence of ecdysone (Gauhar et al., 2009). The higher sensitivity seen at the level of the upstream flanking nucleosome is surprising and requires further investigation; (2) after stimulus, in opening DHSs and without EcR, the MNase-sensitive nucleosome is not detected, indicating that EcR may be involved in the mechanisms that drive it, probably through the recruitment of chromatin remodelers; (3) in closing DHSs, EcR RNAi leads to decreased MNase-sensitivity before stimulus. However, EcR motif was not detected at all in closing DHSs, suggesting that this must be an indirect effect. Additionally, since sensitivity levels in closing DHSs are comparable to CTL, the corresponding action must be EcR-indipendent. These data are supported also by chromatin accessibility probing, further proving a direct correlation between MNase-sensitive nucleosome occupancy and DHS activity.
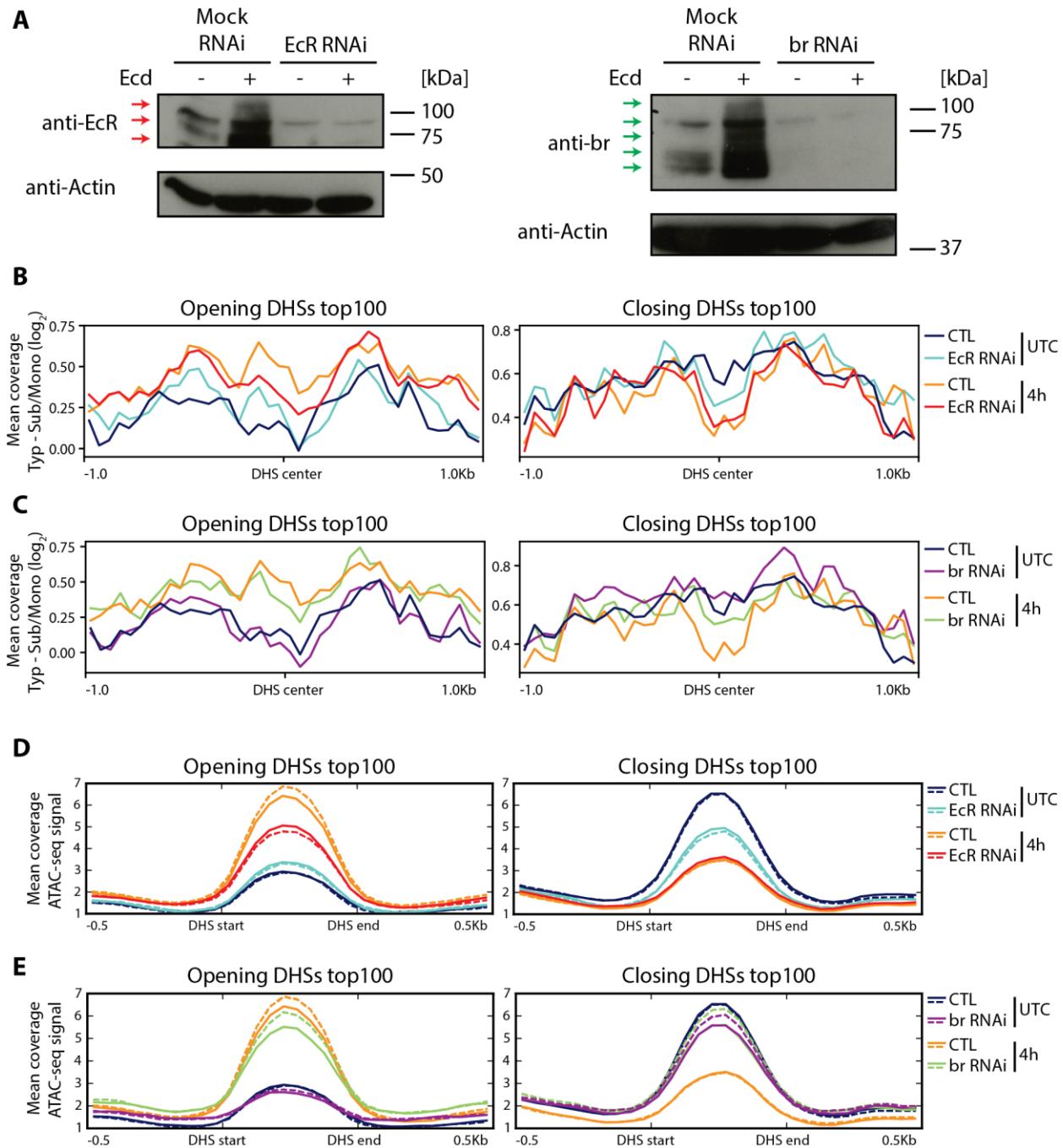
Figure 13: EcR and br knockdowns reveal alterations in MNase-sensitivity changes and chromatin accessibility in highly inducible CREs. (A) Western blot of EcR (left panel) and br (right panel). In each blot, RNAi-based knockdowns were compared to a mock RNAi in untreated (Ecd-) and treated (Ecd+) conditions. Red arrows indicate predicted and apparent EcR isoforms molecular weights (105, 80, and 73 kDa). Green arrows indicate predicted and apparent br isoforms molecular weights (118, 81, 72, 64, and 57 kDa). Actin was used as loading control. (B-C) Log$_2$ ratio between Sub typ and Mono typ mean dyad coverage in CTL and EcR RNAi (B), and in CTL and br RNAi (C). Ratios were plotted over top100 opening DHSs (left panels) and top100 closing DHSs (right panels). Ratios at UTC and 4h are shown. For CTL, dark blue and yellow lines indicate UTC and 4h, respectively. For EcR RNAi, light blue and red lines indicate UTC and 4h, respectively. For br RNAi, violet and green lines indicate UTC and 4h, respectively. Regions ±1 kb away from the DHS center were considered. (D-E) Mean coverage of ATAC-seq signal in CTL and

EcR RNAi (D), and in CTL and br RNAi (E). ATAC-seq signals were plotted over top100 opening DHSs (left panels) and top100 closing DHSs (right panels). Ratios at UTC and 4h are shown. For CTL, dark blue and yellow lines indicate UTC and 4h, respectively. For EcR RNAi, light blue and red lines indicate UTC and 4h, respectively. For br RNAi, violet and green lines indicate UTC and 4h, respectively. Dashed lines represent corresponding biological replicates. Regions 0.5 kb away from either DHS start or DHS end were considered.

When br RNAi was examined, a different scenario came out (Figure 13C). In top100 opening DHSs, br knockdown did not reveal alterations in MNase-sensitivity. However, a striking effect was visible in top100 closing DHSs, where at 4 hours a change towards lower MNase-sensitive levels in the peak center did not occur. As for EcR RNAi, ATAC-seq signal mirrored MNase-sensitivity (Figure 13E): in br RNAi, closing DHSs after stimulus showed accessibility levels comparable to UTC. These findings indicate that br may be directly required in the mechanism that brings the central nucleosome in EcREs from higher to lower sensitivity levels, along with closing chromatin. This is in accordance with the br motif enrichment discussed above, further supporting our hypothesis of br as main player for closing CREs upon ecdysone stimulus.

4.2.4 Distinct CRE activities drive MNase-sensitivity changes as well as EcR- and br-dependent mechanisms

Next, we further investigated the role of DHS activity with regard to MNase sensitivity, and its link to EcR and br. In previous studies, the relation between MNase-sensitive nucleosomes and activity was mainly assessed in promoters and only to a lesser extent in enhancers. Thus, we focused on the latter, along with regions that showed low changes in activity. Therefore, Sub/Mono $\log_2$ ratios and ATAC-seq signals from CTL, EcR RNAi and br RNAi were plotted over top100 enhancers and bottom100 DHSs (opening and closing). Notably, with regard to enhancers, the same patterns identified in top100 all DHSs were observed for all three conditions, both at the levels of MNase-sensitivity (Figure 14A-B) and accessibility (Figure 14C-D). In bottom100 DHSs, differences in MNase-sensitivity (Figure 15A-B) and chromatin accessibility (Figure 15C-D) upon treatment were minimal. More specifically, small reduction in sensitivity occurred only in closing DHSs in CTL, although to a much lesser extent than top100 closing DHSs. Remarkably, levels in EcR or br knockdowns did not show any clear patterns and were comparable to CTL. These results indicate that also enhancers show activity-driven changes at the level of MNase-sensitive nucleosomes, and such changes are detectable only in highly inducible DHSs. Consequently, EcR and br roles in MNase-sensitivity are also limited to very inducible CREs.
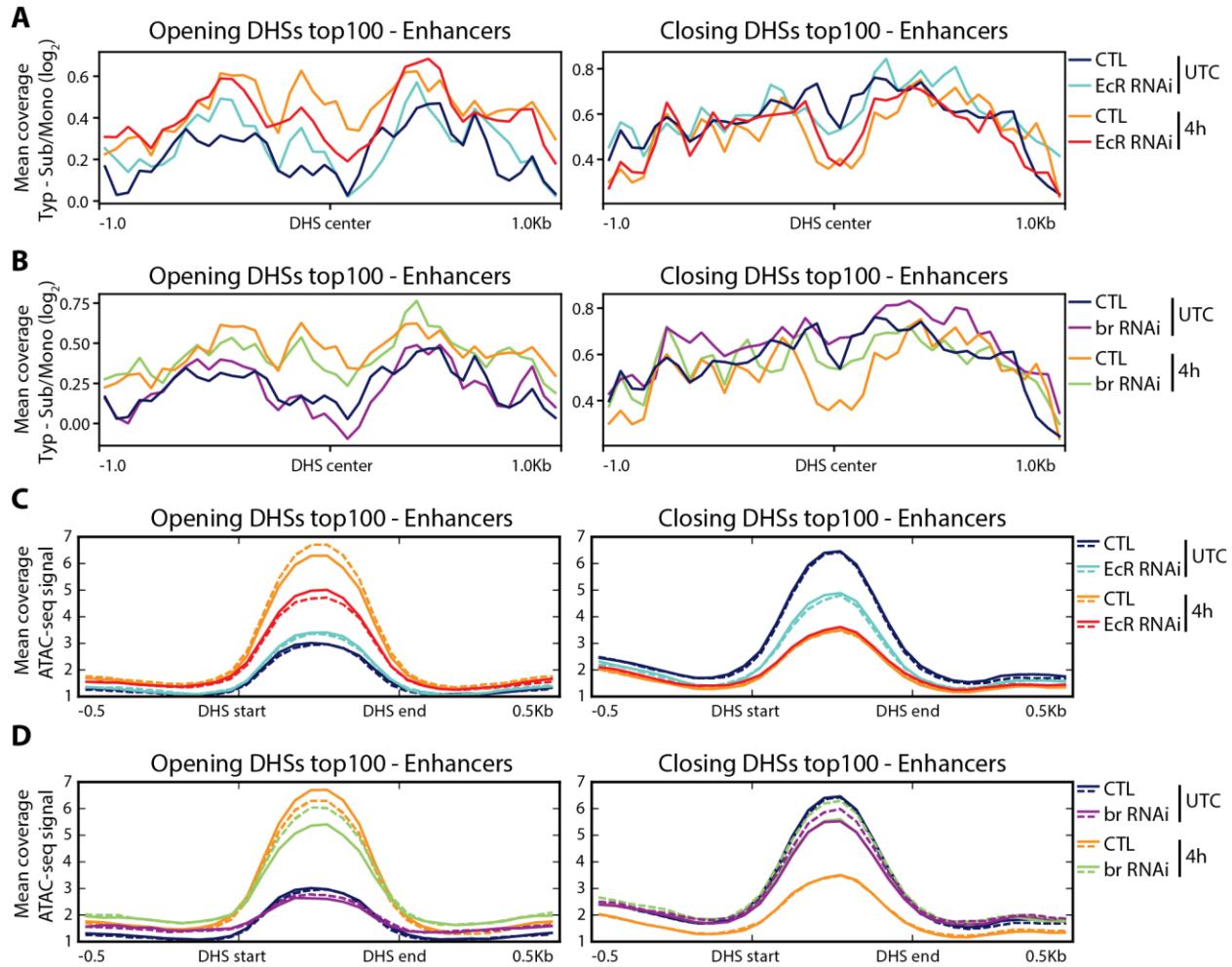
Figure 14: Highly inducible enhancers show changes in MNase-sensitivity as well as EcR and br knockdowns alterations. (A-B) Log$_2$ ratio between Sub typ and Mono typ mean dyad coverage in CTL and EcR RNAi (A), and in CTL and br RNAi (B). Ratios were plotted over top100 opening enhancers (left panels) and top100 closing enhancers (right panels). Ratios at UTC and 4h are shown. For CTL, dark blue and yellow lines indicate UTC and 4h, respectively. For EcR RNAi, light blue and red lines indicate UTC and 4h, respectively. For br RNAi, violet and green lines indicate UTC and 4h, respectively. Regions ±1 kb away from the DHS center were considered. (C-D) Mean coverage of ATAC-seq signal in CTL and EcR RNAi (C), and in CTL and br RNAi (D). ATAC-seq signals were plotted over top100 opening enhancers (left panels) and top100 closing enhancers (right panels). Ratios at UTC and 4h are shown. For CTL, dark blue and yellow lines indicate UTC and 4h, respectively. For EcR RNAi, light blue and red lines indicate UTC and 4h, respectively. For br RNAi, violet and green lines indicate UTC and 4h, respectively. Dashed lines represent corresponding biological replicates. Regions 0.5 kb away from either DHS start or DHS end were considered.

Figure 15: Changes in MNase-sensitivity as well as EcR and br knockdowns alterations are activity-dependent, as they are less visible in bottom100 DHSs. (A-B) Log_2 ratio between Sub typ and Mono typ mean dyad coverage in CTL and EcR RNAi (A), and in CTL and br RNAi (B). Ratios were plotted over bottom100 opening DHSs (left panels) and bottom100 closing enhancers (right panels). Ratios at UTC and 4h are shown. For CTL, dark blue and yellow lines indicate UTC and 4h, respectively. For EcR RNAi, light blue and red lines indicate UTC and 4h, respectively. For br RNAi, violet and green lines indicate UTC and 4h, respectively. Regions ±1 kb away from the DHS center were considered. (C-D) Mean coverage of ATAC-seq signal in CTL and EcR RNAi (C), and in CTL and br RNAi (D). ATAC-seq signals were plotted over bottom100 opening enhancers (left panels) and bottom100 closing enhancers (right panels). Ratios at UTC and 4h are shown. For CTL, dark blue and yellow lines indicate UTC and 4h, respectively. For EcR RNAi, light blue and red lines indicate UTC and 4h, respectively. For br RNAi, violet and green lines indicate UTC and 4h, respectively. Dashed lines represent corresponding biological replicates. Regions 0.5 kb away from either DHS start or DHS end were considered.

Finally, we asked whether our observations would have been confirmed also by Short/Typ log_2 ratio of mono-nucleosomal fragments, which represents an additional measure of MNase-sensitivity. With regard to CTL, not surprisingly, sensitivity at the center of the peak increased or diminished in top100 opening or closing DHSs, respectively (Figure S8A-B). Likewise, the same

changes were observed in top100 enhancers (Figure S8E-F). Interestingly, in top100 opening DHSs or enhancers only, greater sensitivity of flanking nucleosomes appeared only after stimulus. This is dissimilar to Typ Sub/Mono ratio, which showed sensitivity of those nucleosomes regardless of hormone induction. Bottom100 opening DHSs showed no difference upon stimulus, whereas bottom100 closing DHSs had slight decrease in sensitivity (Figure S8C-D), as already observed in Sub/Mono ratio. As expected, EcR and br knockdowns revealed exactly the same alterations discussed above: EcR RNAi led to slight higher sensitivity of the central nucleosome in opening DHSs at UTC, lack of increased sensitivity after stimulus in the same regions, and lower levels in closing DHSs in the absence of ecdysone (Figure S8A and S8E); br RNAi led to alterations in closing DHSs at 4 hours (Figure S8B and S8F).

Overall, these data further demonstrate that MNase-sensitivity changes of nucleosomes located at DHS center are activity-driven, also occur in enhancers, and restricted to very inducible CREs. This is shown by two alternative and reliable approaches to measure MNase-sensitivity, plotted over sets of differential DHSs that reflect distinct CRE activities. Furthermore, our results suggest an opposite role of EcR and br in regulating MNase-sensitivity of such nucleosomes. Hence, EcR and br are strong candidates for executing this *trans*-factor-dependent action postulated above. This hypothesis is supported by three different assays, but it does not exclude the contribution of additional factors.

4.2.5 Chromatin structure alterations lead to misregulated gene expression only with regard to up-regulation

Highly inducible CREs showed drastic changes in MNase-sensitivity, most likely driven by EcR and br. As we also characterized a strong association between CRE activity and gene expression (Figure 6), we aimed at studying gene expression FC in relation to MNase-sensitivity changes. Therefore, opening and closing top100 all DHSs, top100 enhancers and bottom100 DHSs were associated to target genes. Gene expression FC of those targets were calculated at 4 hours after ecdysone treatment, both in CTL, EcR and br knockdowns conditions.

In opening top100 all DHSs (Figure 16A), a clear up-regulation of gene expression occurred in CTL, as expected. However, significant (p-value <0.05) reduction of expression FC was detected only in EcR RNAi. By contrast, expression in br RNAi was as induced as in CTL. Also with regard to top100 enhancers, significant reduction in gene expression was recorded only

in EcR RNAi. With regard to bottom100 opening DHSs (Figure 16A), gene expression up-regulation was generally lower, however no significant differences were detected among the three conditions. These data indicate that alterations in nucleosome sensitivity levels and diminished accessibility (as for EcR RNAi) result in misregulated gene expression induction. When looking at conditions that do not present striking differences in sensitivity and accessibility, also gene expression induction is not affected.
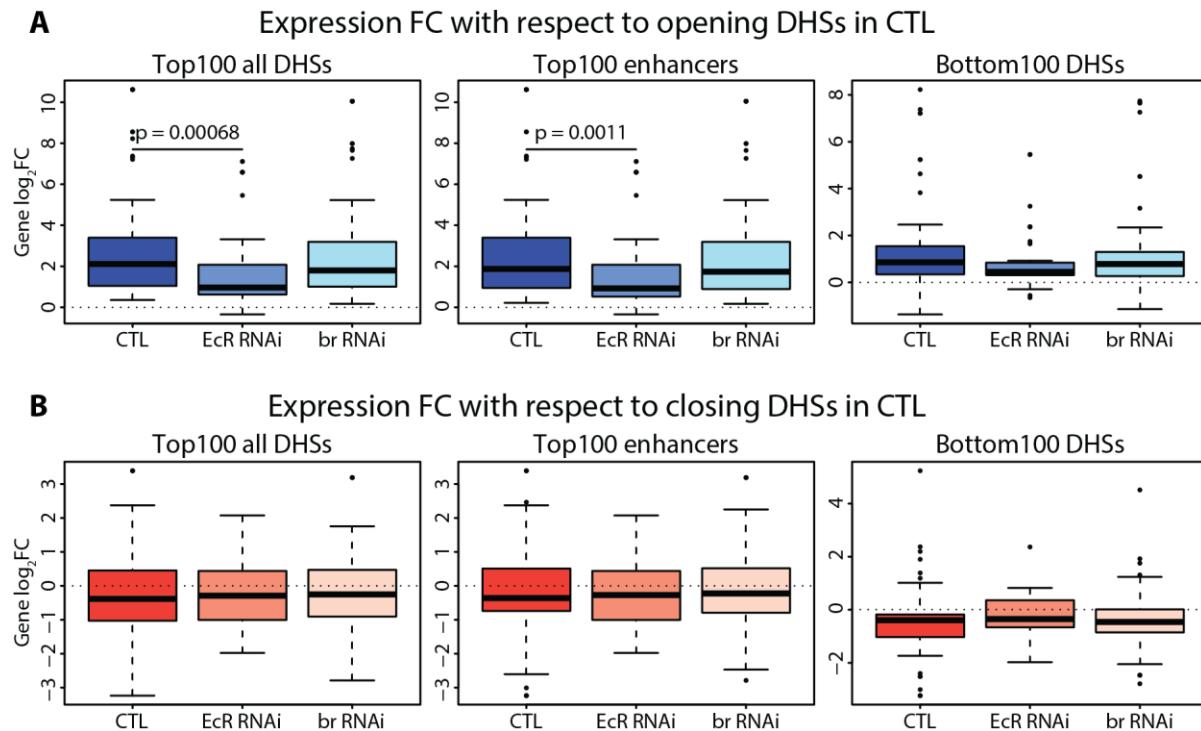
Figure 16: Chromatin structure alterations result in misregulated gene expression only with regard to EcR-dependent opening DHSs. (A-B) Opening (A) and closing (B) top100 all DHSs (left panels), top100 enhancers (middle panels), and bottom100 DHSs (right panels) were associated to target genes. Boxplots represent expression $\log_2 FC$ of those genes calculated at 4h. $\log_2 FC$ in CTL, EcR RNAi and br RNAi conditions are illustrated. Bottom and top of the boxes are 25th and 75th percentile, respectively. Thick lines in the boxes represent the median. P-values were calculated with respect to CTL $\log_2 FC$ (Wilcoxon rank-sum test). Only p-values <0.05 are illustrated.

Surprisingly, in closing DHSs unexpected outcomes revealed (Figure 16B). Although EcR RNAi led to lower sensitivity at UTC, and br RNAi to closing alterations at 4 hours, gene expression was neither significantly induced nor repressed. That was true both in closing top100 all DHSs and enhancers, as well as in bottom100 DHSs. This peculiar finding further implicates a double role of closing DHSs, which is not necessarily linked to gene down-regulation.

Nevertheless, it is curious how, in those regions, evident irregularities in chromatin structure do not result in gene expression misregulation. This topic certainly needs further investigation.

4.3 PART III: A large genome-wide characterization of CRE dynamics during *Drosophila* metamorphosis with great spatio-temporal resolution

S2 cells represent a powerful system to study the transcriptional cascade triggered by ecdsyone, as temporal dynamics can be deeply dissected. However, S2 cells respond all identically to the stimulus. Thus, a proper tissue-specific investigation must be conducted *in vivo*.

The first ecdysone pulse during *Drosophila* metamorphosis triggers larval-to-prepupal transition. During this phase, drastic morphological changes occur. Tissues can go towards completely opposite paths: imaginal discs (such as WD and ED) survive the pulse and begin differentiation. Instead, obsolete larval tissues (such as SG) undergo PCD and will be eventually removed. A complex tissue such CNS shows both cell fates: some neurons of the ventral cord die, but multiple cell types differentiate. These mechanisms are all triggered by ecdysone, but the response is tissue-specific, due to combinatorial actions of the hormone with local signaling pathways and TFs.

4.3.1 ATAC-seq provides excellent recovery of tissues-specific accessibility landscapes

To unravel the tissue-specific ecdysone response during metamorphosis, a large genome-wide characterization of CREs was conducted with great spatio-temporal resolution (Figure 17A). Firstly, 3 stages of the larval-to-prepupal transition were selected: early 3rd instar larva (Early 3IL), which is the stage prior to the ecdysone pulse; late 3rd instar larva (Late 3IL), which is almost at the peak of the pulse; white prepupa (WPP), which represents prepupal stage right after the end of the pulse. Then, larvae from those stages were dissected and four tissues were isolated, representative of the different morphological outcomes occurring during metamorphosis: WD and ED (surviving tissues), SG (dying tissue), and CNS (complex tissue with both survival and death). Finally, ATAC-seq was performed on all the tissues for each stage.

The chromatin landscapes obtained with this strategy were extremely large and with a great SNR (Figure 17B). When we looked at ATAC-seq tracks over the *EcR* locus, striking patterns appeared. WD and ED showed the same chromatin landscapes over time, and also their dynamic peaks coincided. CNS and SG, instead, had distinct landscapes, different from all the others. The obtained ATAC-seq *in vivo* tracks demonstrate that chromatin accessibility can be reliably probed

with spatio-temporal resolution, produces high quality data, and its dynamics can be detected in a tissue-specific manner.
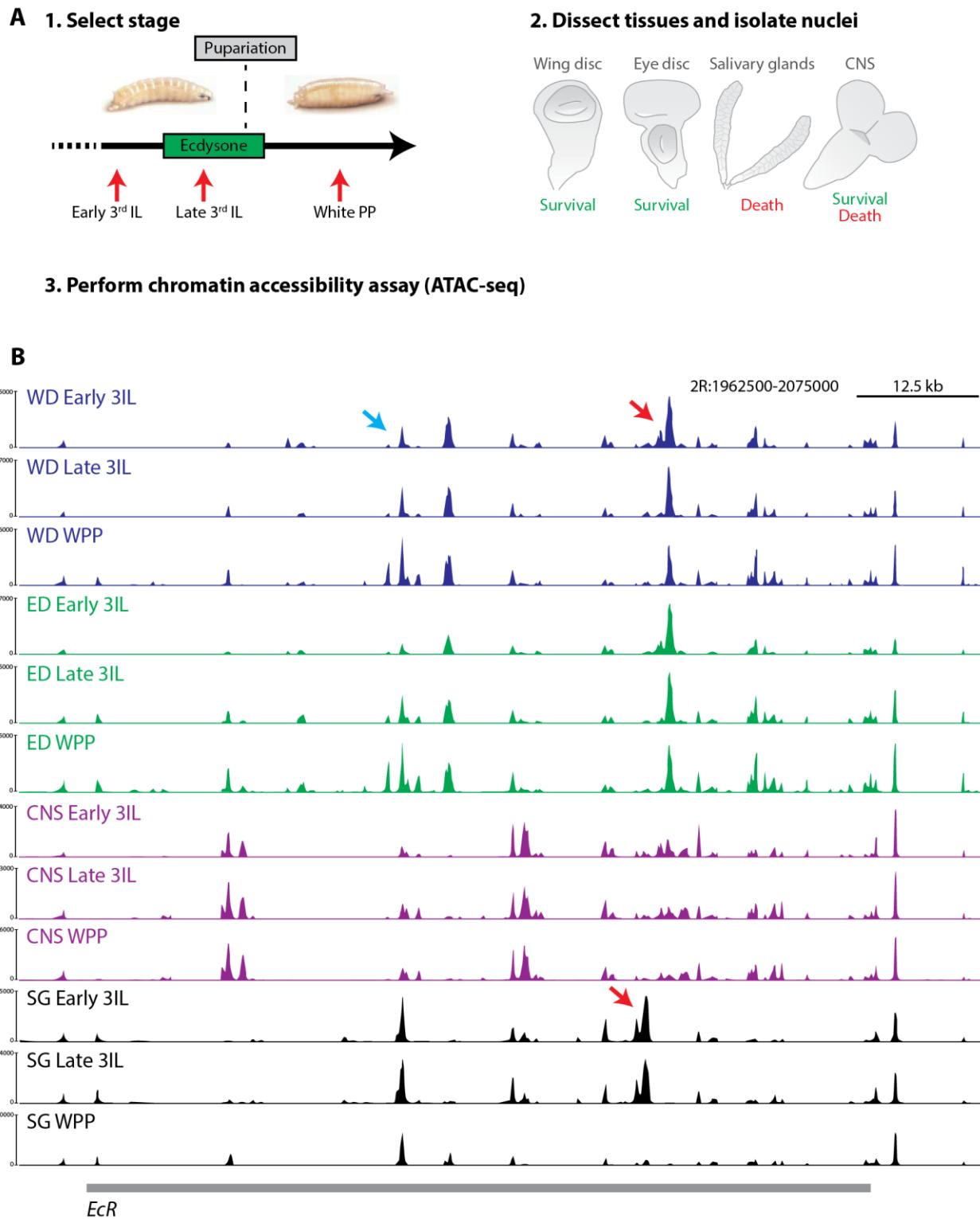
Figure 17 (previous page): A large chromatin accessibility landscape with great spatio-temporal resolution during *Drosophila* metamorphosis. (A) Experimental strategy to dissect tissue-specific ecdysone-response during metamorphosis. 1: Early 3IL, Late 3IL, and WPP stages were carefully timed. 2. Larvae from those stages were selected and WD, ED, CNS, and SG dissected. Tissues were homogenized, followed by nuclei isolation. 3: ATAC-seq was performed on the selected tissues. (B) Genome browser screenshot of ATAC-seq tracks over the *EcR* locus. For each tissue, all the stages are shown. Blue arrow indicates opening DHSs in WD and ED over time. Red arrows indicate closing DHSs in SG over time.

4.3.2 *In vivo* DHS landscapes and dynamics reflect cell fates and shape cell identity

Next, a more comprehensive evaluation of accessibility differences among tissues was conducted. Individual whole landscapes were considered, and sample-to-sample distances of similarity were calculated through the "dist" function of the DESeq2 R package (Love et al., 2014). Similarities were illustrated in heatmaps, in which samples were clustered based on the distances between rows/columns of the distance matrix (Figure 18A-C): similar samples were placed closer in the clustering and showed lower values. When Early and Late 3IL were compared, three distinct clusters of similarity were visible (Figure 18A). The largest cluster comprised WD and ED both in Early and Late 3IL, in accordance with the similar ecdysone-response they share. Interestingly, the two tissues at the same stage showed more similarity than the single tissue at different stages. A second cluster was represented by Early and Late 3IL CNS. Consequently, the third cluster included SG at the two stages. Notably, SG landscapes are far apart from the imaginal discs, in line with the two very opposite fates they go through. Similarly, the same three clusters were depicted when comparing Early 3IL to WPP (Figure 18B), and Late 3IL to WPP (Figure 18C). With regard to the latter, it is noteworthy to mention that similarity among the three clusters diminished even more, and that WD and ED landscapes were now closer in the same tissue between different stages than vice versa. This is in line with the differentiation and cell fate program that the different tissues are undergoing at later stages.

Subsequently, we focused on the DHS dynamics occurring in this paradigm. Differential DHSs were identified and divided in opening or closing with respect to Early 3IL (Figure 18D). Strikingly, the numbers of differential DHSs were very high compared to S2 cells, with an average of 8602 opening peaks and 3653 closing peaks. All the tissues generally seem to respond to ecdysone with chromatin opening rather than closing. Interestingly, CNS had greatest changes in opening DHSs during Late 3IL transition, whereas SG during WPP transition. When differential DHSs were compared among tissues, results demonstrated that also DHS dynamics shape cell identity (Figure S9). As representative examples, open DHSs in WD or in SG were measured with

respect to the other tissues over time. WD (left panel) showed less differential peaks compared to ED than CNS or SG (that is, WD shared more open peaks with ED than with CNS or SG over time). On the contrary, SG (right panel) revealed high numbers of differential DHSs compared to the other tissues in all the stages.
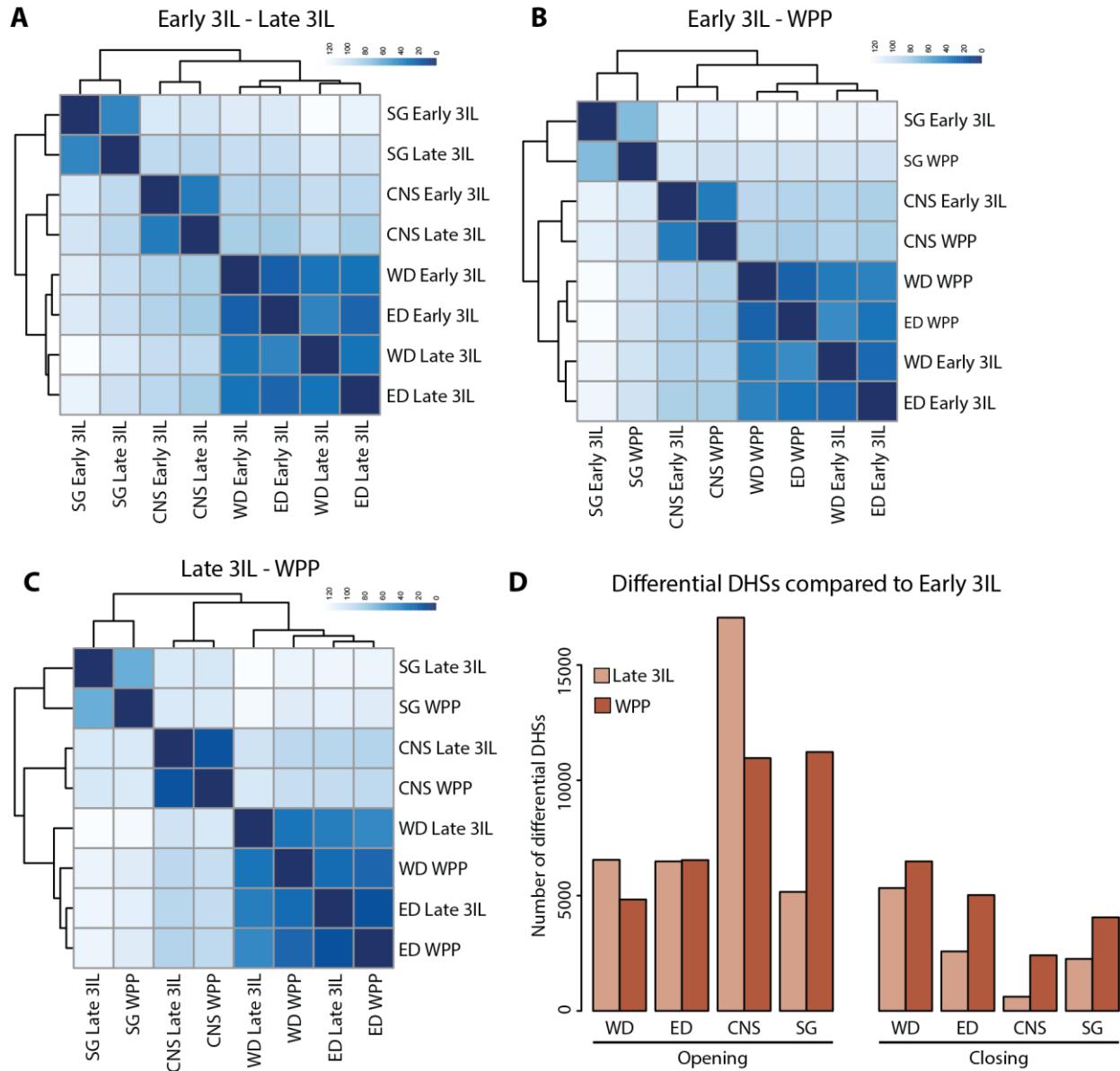
Figure 18: Chromatin accessibility landscapes mirror cell fates and shape cell identity. (A-C) Heatmaps showing distances of similarity among tissues with regard to their accessibility landscapes. (A) Early 3IL and Late 3IL stages are compared. (B) Early 3IL and WPP stages are compared. (C) Late 3IL and WPP stages are compared. (D) Number of differential DHSs compared to Early 3IL in each tissue. Differential DHSs were divided in opening (left) and closing (right) DHSs.

Overall, these results demonstrate that chromatin accessibility landscapes intrinsically determine cell identities, and similar cell fates over metamorphosis are accompanied by analogous DHS dynamics.

4.3.3 An in-depth tissue-specific analysis of TF motif enrichment during metamorphosis

Ecdysone tissue-specific response is often attributed to the interplay of the hormone-triggered cascade with locally expressed TFs and their cognate signaling pathways. Here, a comprehensive mapping of CREs during metamorphosis was obtained in a tissue-specific fashion. Hence, it was natural to perform a TF enrichment analysis on differential DHSs to characterize the most important regulators that determine distinct outcomes.

Firstly, differential DHSs were identified in each tissue always in comparison to Early 3IL. Then, they were divided in opening and closing DHSs during Late 3IL and WPP transitions. Next, differentially expressed TFs had to be selected, in order to create tissue-specific TF sets to feed the AME algorithm. As we have already demonstrated an almost linear correlation between promoters FC and gene expression FC (Figure S3), TFs were selected according to the presence of a differential DHS on their promoters in any stage. Then, their PWMs were searched for in several databases. The results were four sets of differentially expressed TFs, one set per tissue (final sets listed in Table 2-5).

Motif enrichment was run over opening and closing DHSs along time. It revealed significant values of several TFs in each tissue (Table 6-9). At first glance, it also seemed that many TFs were present in multiple tissues and in different stages. Intrigued by this redundancy, we selected TFs that: (1) in each tissue, were enriched both in opening and closing DHSs at each stage; (2) displayed such broad enrichment in at least 3 tissues. Fifteen TFs matched those conditions in all the tissues; sixteen in 3 tissues. Remarkably, TFs present in all tissues included br-Z1, which was already hypothesized to have multiple and wide functions when S2 cells enrichment was discussed (Figure 10). Additionally, two final effectors of local signaling pathways were recorded: cubitus interruptus (ci), which acts as activator or repressor at the end of Hedgehog pathway, and Mothers against dpp (Mad), which mediates the cellular response to Dpp pathway. In the list of TFs present in 3 tissues, two are noteworthy: Adh transcription factor 1 (Adf1), a transcriptional activator that regulates several developmental mechanisms, such as dendrite growth (Timmerman et al., 2013), and Trithorax-like (Trl), also known as GAGA factor

(GAF), a putative *Drosophila* pioneer factor associated to multiple contradictory functions, such as NDR generation and heterochromatin silencing, and that interacts with several chromatin remodelers (Lomaev et al., 2017). Notably, both TFs were previously characterized in an ecdysone-dependent motif enrichment analysis in S2 cells (Shlyueva et al., 2014). Although these data are certainly useful, they mostly depict TFs which are broadly expressed and execute many functional actions.

As we aimed at higher specificity, a second step was conducted, in which only TFs expressed either in WD or SG were considered, discarding all the shared ones (Figure S10). In this case, known tissue-specific TFs began to appear as mostly enriched. In WD (Figure S10A), apterous (ap) was the second most significant TF in opening DHSs in Late 3IL, and the most significant in opening DHSs in WPP. Ap is a fundamental regulator for the dorsal pattern development in WD. Interestingly, retained (retn), a TF require for glial cells development, was highly enriched in all differential DHSs throughout metamorphosis. In SG (Figure S10B), the master regulator fork head (fkh), fundamental for SG formation, showed high enrichment in all differential DHSs. Noteworthy, klumpfuss (klu), previously associated to differentiation of neural progenitors, revealed also very significant values. These data confirm the reliability of our analysis, which reported not only known tissue-specific TFs, but also regulators previously characterized in other contexts.

Finally, another level of specificity was achieved, as also ED and CNS were taken into account. Thereby, for each tissue, only TFs which were not detected in any of the other three tissues were considered (Figures 19-22). The outcomes were extremely interesting. At first glance, it was immediately visible how WD and ED showed very few specific TFs (in total 6 and 4, respectively). This is not surprising, as the two imaginal discs share the same cell fate (even though they differentiate in different appendages). Thus, it is very likely that many TFs function in both tissues, thereby being discarded in the analysis. However, this observation also means that this computational approach provides high tissue and stage specificity.

Notably, in WD, no specific TFs were enriched in opening DHSs during Late 3IL transition (Figure 19). Nevertheless, three considerations must be pointed out: (1) br-Z2 was found in closing DHSs both at Late 3IL and WPP. Once more, this br isoform is detected in closing chromatin, as we have already observed, strengthening our hypothesis on its role in such a mechanism. Interestingly, br-Z2 action seems to be WD specific; (2) the heterodimer EcR-USP came out only

in WPP opening DHSs of WD, with low enrichment compared to other TFs. This suggests that *in vivo* EcR-USP may regulate few specific targets to trigger tissue-specific response; (3) invected (inv) and ladybird early (lbe), enriched in WPP opening DHSs, are tightly related to engrailed (en), a TF fundamental for WD posterior pattern development. En regulates the Hedgehog pathway by controlling many of its targets and components expression in WD, suggesting a contribution of such a pathway in WD-specific ecdysone-response.

In ED (Figure 20), Homodomain protein 2.0 (H2.0) was highly enriched in all the differential DHSs. H2.0 has been poorly characterized previously and detailed information on its action are lacking. Therefore, it could be an interesting candidate to better understand the ED differentiation triggered by ecdysone.

On the contrary, CNS-specific TFs were numerous (Figure 21). In Late 3IL opening DHSs, a total of 78 TFs were recorded (only top20 visualized). In this regard, many components of the ecdysone-cascade were reported, such as E75B, E78C and different br motifs (including motifs named after its transcript variants). Noteworthy, the EcR monomer was highly enriched, as well as several uncharacterized TFs. These data are in line with the greater complexity of CNS differentiation. In agreement with (Figure 18D), Late 3IL opening DHSs seem to activate many genes targeted by multiple TFs, creating a complex regulatory network at that stage. In Late 3IL closing DHSs, and with the progression of metamorphosis, the number of CNS-specific TFs diminished. Bric-à-brac 1 (Bab1) and CG7368 represented the most significant TFs also in those contexts. Bab1 was mainly reported in antenna and leg proximal-distal axis formation (Couderc et al., 2002).

Finally, SG motif enrichment displayed biniou (bin) as most significant TF (Figure 22). Bin is a forkhead domain-containing TF involved in the visceral mesoderm formation, but never attributed to SG. In addition, pnr and NFAT, which we identified in S2 cells motif enrichment analysis, were mostly present in WPP opening DHSs, suggesting that their actions in response to ecdysone can be related to SG cell fate.

Overall, we provided a comprehensive tissue- and stage-specific characterization of TFs during *Drosophila* metamorphosis. Strikingly, thanks to the great quality of our data and to a computational approach that eliminates redundancy, new putative regulators involved in the context-dependent ecdysone-response *in vivo* were identified. Particularly, each tissue (WD, ED, CNS and SG) displays at least one highly enriched TF whose function was never previously

reported in the cognate context. These vast TF datasets represent a great source for further investigations on the ecdysone role in developmental processes.
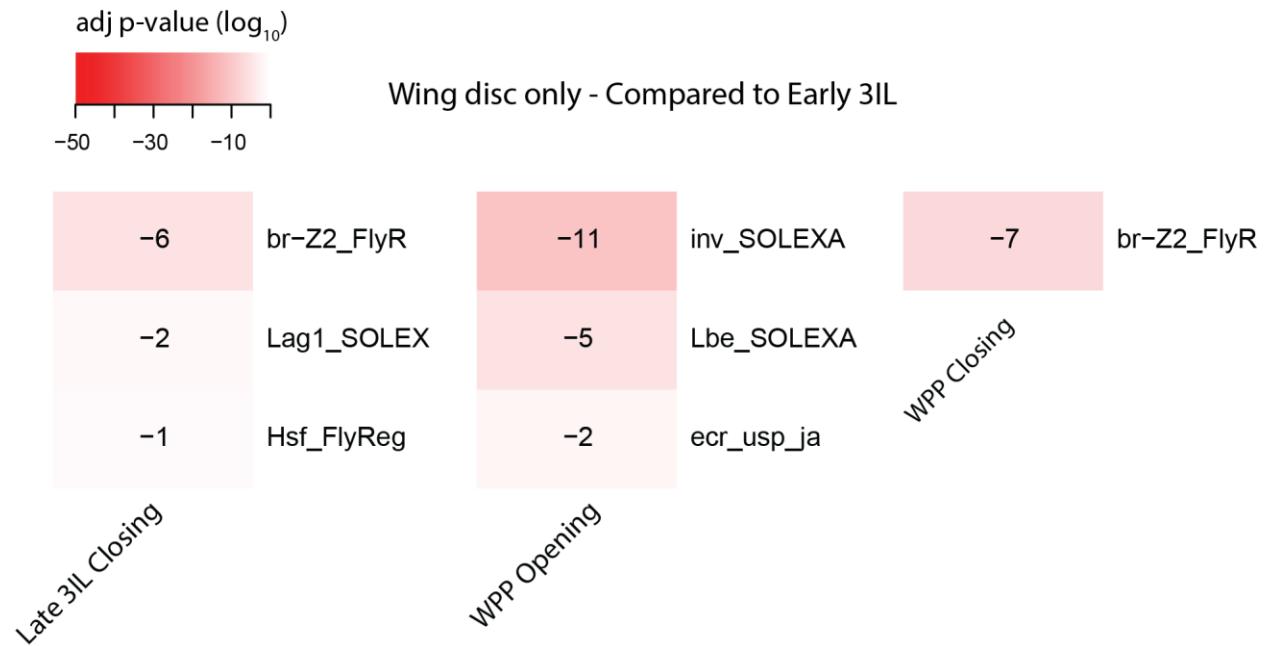
Figure 19: WD-specific motif enrichment in opening and closing DHSs over time. Differential DHSs were detected in comparison to Early 3IL. Late 3IL opening DHSs did not show any WD-specific TF enrichment. Heatmaps show TF adjusted p-values (log$_{10}$) as measure of motif enrichment in each DHS set. Corresponding values are displayed.
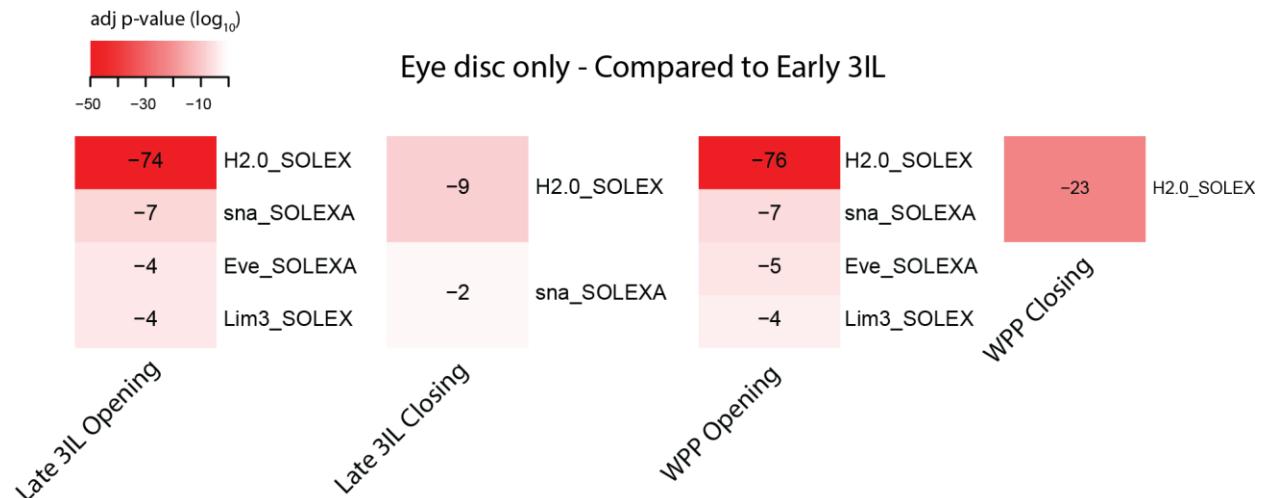
Figure 20: ED-specific motif enrichment in opening and closing DHSs over time. Differential DHSs were detected in comparison to Early 3IL. Heatmaps show TF adjusted p-values (log$_{10}$) as measure of motif enrichment in each DHS set. Corresponding values are displayed.
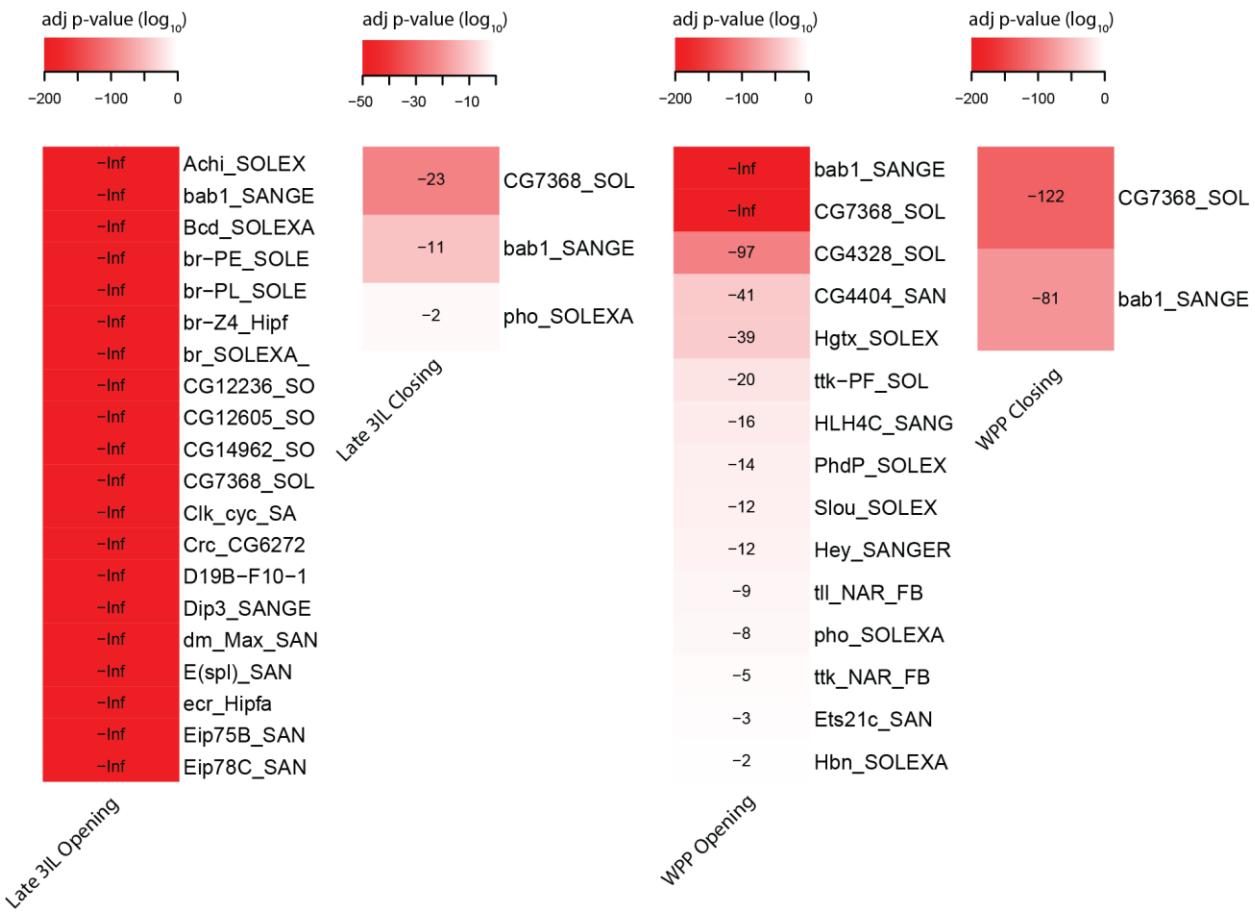
Figure 21: CNS-specific motif enrichment in opening and closing DHSs over time. Differential DHSs were detected in comparison to Early 3IL. Late 3IL opening DHSs showed 78 CNS-specific enriched TFs. Here only the top20 are displayed. Heatmaps show TF adjusted p-values ($\log_{10}$) as measure of motif enrichment in each DHS set. Corresponding values are displayed.
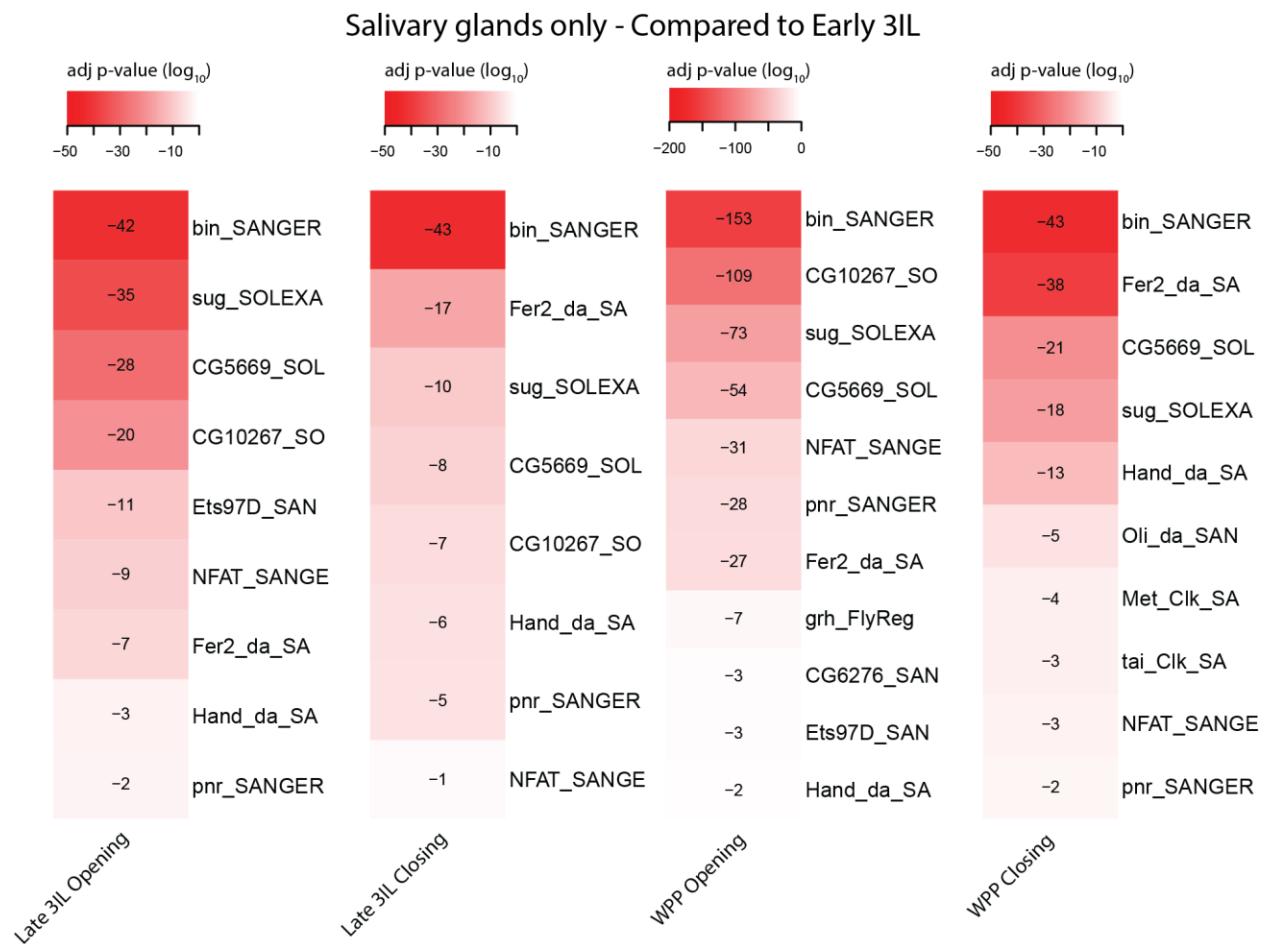
Figure 22: SG-specific motif enrichment in opening and closing DHSs over time. Differential DHSs were detected in comparison to Early 3IL. Heatmaps show TF adjusted p-values ($\log_{10}$) as measure of motif enrichment in each DHS set. Corresponding values are displayed.

5 DISCUSSION

Since the postulation of the Ashburner model, the ecdysone transcriptional cascade has been investigated at many levels. Ecdysone plays a pivotal role in dictating developmental processes in *Drosophila*. Its binding to the nuclear receptor heterodimer EcR-USP triggers a vast response that includes the activation and repression of several regulators in a hierarchical fashion. The ultimate targets of those TFs are effector genes, which execute multiple morphological changes in the fly. Although ecdysone acts systemically, different tissues interpret the hormonal signal in a context-dependent manner, leading to extremely opposite outcomes between tissues (as in the case of imaginal discs and SG), or even within the same tissue (as in the case of CNS, where distinct cell populations undergo either differentiation or PCD).

With the advent of NGS, the examination of the response to ecdysone in a genome-wide fashion was a natural step, both in cells and *in vivo* (Beckstead et al., 2005; Gauhar et al., 2009; Li and White, 2003; McKay and Lieb, 2013; Shlyueva et al., 2014). However, previous studies either focused on single aspects of the genomic response, or provided poor characterization of the dynamics occurring during the cascade, mostly missing the very early time points in which the regulators precisely govern the progression of the cascade. In other cases, the ecdysone context-specificity was not considered (Uyehara et al., 2017).

In this study, we thoroughly dissected the dynamics triggered by the hormone at the level of transcriptional regulation with unprecedented spatio-temporal resolution. We applied a multi-pronged experimental and computational approach that permitted an integrated analysis of the mechanisms controlling gene expression in response to ecdysone. The integration of data obtained from chromatin accessibility, nascent RNAs, and MNase-sensitivity provide the deepest characterization of the ecdysone cascade at genomic and transcriptional levels. By inspecting these data in S2 cells, we present novel insights on ecdysone-triggered dynamics regarding CRE activity, transcriptional output, and chromatin structure. *In vivo*, we highlight the role of accessibility landscape in determining context-dependent response. Furthermore, extensive TF lexica were produced, showing new TFs clearly involved in the cascade progression, as well as new regulators participating in the tissue-specific outputs governed by the hormone.

5.1 Ecdysone-triggered CRE dynamics can be reliably mapped, correlate with gene expression and suggest different modes of action

The ecdysone cascade is triggered by binding of EcR-USP on multiple EcREs, indicating that drastic changes in CRE activity occur already at the very first step. In Kc167 cells, it was demonstrated that gene expression is activated and induced at very early time points after ecdysone stimulus (Gauhar et al., 2009). Similarly, Katja Frühauf from the Gaul lab monitored early transcriptional events in S2 cells by applying DTA-seq, which provides greater sensitivity than total RNA-seq, and showed that most of the genes are differentially expressed within the first 12 hours. Therefore, it is clear that the ecdysone response in cells must be immediate also at the regulatory level. Here, we characterized early dynamics of ecdysone-induced CRE activity in S2 cells with remarkable resolution. We showed that already at 1 hour after ecdysone stimulus, hundreds of CREs are detected as differential. We also demonstrated that the number of differential DHSs increased along the time course, in line with the progressive activation and repression of new regulators that target additional CREs. When DHS activity was quantitatively assessed, higher FC values were mostly observed in opening DHSs at early time points, in agreement with the changes in chromatin structure visible as chromatin puffs *in vivo.*

Additionally, a step further was achieved: a deep quantitative analysis of the correlation between DHS activity and nascent RNA levels (data by Katja Frühauf) was accomplished. Although similar correlations have been run in previous studies in *Drosophila* (Shlyueva et al., 2014), also with the aim to model gene expression (Blatti et al., 2015), we present a direct quantitative association between individual CRE activity and differential target genes upon a developmental stimulus. Our results indicate strong correlation between DHS and gene FC at any time point both for promoters and enhancers, indicating that opening or closing chromatin immediately control gene expression by activating or repressing it, respectively. To our knowledge this is the first study that systematically addressed this topic in *Drosophila*.

Interestingly, despite this general correlation, distinct modes of action between opening and closing DHSs were observed. Genes controlled by multiple closing DHSs did not show greater repression, unlike the synergic expression induction executed by multiple opening DHSs. It is possible that the inactivation of a single DHS (such as a promoter) is sufficient to repress gene expression. However, when multiple closing DHSs target a single gene, they may play a more

fine-tuning role. The shutting down of alternative promoters in induced genes could have biased our analysis. Nevertheless, we observed a higher association of closing enhancers to up-regulated genes than closing promoters. Additionally, we also showed a lack of greater repression by multiple closing DHSs when only enhancers were considered. Thus, closing alternative promoters are ruled out. We therefore postulate a role of closing enhancers in regulating developmental genes that cloud be a fine-tuning control of induced genes, other than mere repression. Interestingly, a previous study in *Drosophila* demonstrated that genes with longer intronic regions have a higher number of TFBSs than other genes, also when normalized by their length (Stark et al., 2007). Thus, it is possible that genes with such a genomic architecture, as the ecdysone-induced genes (Karim et al., 1993), require very complex regulatory mechanisms that may involve a synergic function of opening DHSs and/or a fine-tuning role of closing enhancers.

In summary, our strategy allowed us to reliably map DHS dynamics in ecdysone-stimulated S2 cells, and when DTA-seq data were integrated, it provided interesting considerations on gene regulation mechanisms.

5.2 Individual dynamics are modeled in four behaviors and display complex regulation for S2 cells developmental genes

By taking advantage of our high quality data on ecdysone-triggered dynamics, we applied ImpulseDE2, which models individual DHS or gene dynamics in four distinct patterns. Notably, we demonstrated that the biggest portion of DHSs or genes underwent a continuous increase or decrease in their read count (Tn-U and Tn-D). Only few DHSs and genes fell into categories characterized by a switch in their patterns (Tt-U and Tt-D). Interestingly, also through this approach, we could illustrate a general correlation between differential DHS and target gene dynamics in each category. However, a couple of points are noteworthy. Firstly, we observed a quite striking correlation between Tn-U genes and other three DHS categories (Tn-D, Tn-U, and Tt-U), recorded also in the enhancer set. Additionally, with regard to TFs only, the correlation between Tn-U genes and different DHS dynamics was higher, especially when considering enhancers only. These findings demonstrate that induced genes are controlled by complex CRE dynamics, as postulated above. In particular, induced TFs, that during the ecdysone cascade receive different inputs of regulation, require more complex CRE control.

Furthermore, induced genes execute the morphological and developmental changes occurring in S2 cells, as shown in the GO analysis, thus they are very likely effector genes. Remarkably, also transient genes showed high enrichment in the same GO terms. These data are in agreement with previous studies in *Drosophila* embryo (Arbeitman et al., 2002; Hooper et al., 2007). Indeed, it was reported that most developmental genes are highly regulated at the level of transcription and are expressed as highly dynamic. Especially, effector genes at late stages of embryogenesis showed continuously increased but spatially tight expression, suggesting complex regulation. Our results depict the same scenario in the ecdysone-triggered regulatory mechanisms. Surely, our approach already selected only dynamic genes at the top of the analysis. Nevertheless, three out of four dynamic categories are enriched for developmental terms, and Tn-U genes, which mostly represent S2 cells development, undergo complex levels of regulation.

Taken together, our data provide novel insights on the dynamics of developmental gene regulation, which were never observed in the context of the ecdysone cascade.

5.3 TF lexicon in ecdysone-stimulated S2 cells

The comprehensive characterization of genomic and transcriptional dynamics in ecdysone-stimulated S2 cells permitted us the identification of TFs involved in the cascade. We focused on differentially regulated TFs, selected by strict parameters, which led to a set of forty TFs with known PWMs to be scanned over dynamic DHSs for motif enrichment. The reliability of this approach can be already assessed by looking at the heterodimer EcR-USP, which is specifically enriched in Tt-U, in agreement with its very early function (Koelle et al., 1991). Strikingly, although many components of the hormone pathway were part of the TF set, only six showed a significant enrichment. Thus, our strategy could detect a great number of TFs other than known ecdysone-signaling components.

Furthermore, our data highlighted TFs that seem to be fundamental in the cascade progression. Srp, pnr, hng3, foxo, and the uncharacterized CG5953 showed great enrichment in at least three dynamic DHS categories. Srp and foxo were already reported to be associated to ecdysone. Srp is an activator required for the ecdysone response in fat body (Brodu et al., 1999, 2001), and its motif was already identified in S2 cells induced enhancers at 24 hours after ecdysone stimulus (Shlyueva et al., 2014). However, neither srp involvement in early dynamics nor its presence in closing DHSs have been previously characterized. Foxo is a regulator of the insulin

pathway, and controls ecdysone biosynthesis to ensure proper larval growth (Koyama et al., 2014), but our enrichment analysis further suggests its direct implication in the ecdysone response. On the other hand, pnr, hng3 and CG5953 have never been associated to ecdysone.

In addition, we observed a very interesting enrichment distribution of the br isoforms. The Z1 isoform showed high enrichment in all four DHS dynamics, but predominantly in Tn-D. The Z2 and Z3 isoforms, instead, displayed Tn-D specificity. This finding is intriguing: it is known that the relative ratio of br isoforms accumulation is necessary for a context-dependent response to ecdysone (Emery et al., 1994), however our data suggest that they may also have distinct roles at the regulatory level. Z1 isoform could function more broadly, whereas Z2 and Z3 isoforms seem to specifically target closing DHSs. With regard to the latter, it is difficult to distinguish whether Z2 and Z3 isoforms directly act as repressors, or if their TFBSs simply need to be masked to avoid binding events. Br-Z2 was reported to have a repressive function in the fat body through biochemical experiments (Mugat et al., 2000), however in the same study br-Z3 was claimed to activate gene expression. In our knockdown experiment (Paragraph 4.2.3, also discussed in this chapter in paragraph 5.5) we show that the absence of all br isoforms indeed leads to lack of chromatin closing, pointing to the repressive action of br. Nevertheless, isoform-specific knockdowns of br would further elucidate its different roles in closing chromatin.

In conclusion, we present an extensive analysis on TFs that participate to ecdysone-triggered dynamics with a great degree of resolution. Our data focus on differentially expressed TFs, to specifically highlight regulators that strongly respond to the stimulus. We show not only novel TFs which were never previously associated to the ecdysone cascade, but also postulate new roles of known regulators of the pathway. A next step would be to include constitutively active TFs and investigate how they interplay with those characterized in our lexicon.

5.4 MNase-sensitive nucleosomes correlate with ecdysone-regulated CRE activity

Nucleosome positioning is typically determined by MNase-seq experiments. However, nucleosome maps are affected by the degree of MNase digestion. Recent studies reported the existence of MNase-sensitive nucleosomes located within or flanking NDRs in several organisms, suggesting that CREs are associated with easily digested, unstable nucleosomes that may play a central role in gene regulation. Although the origin of the particles deriving from differential

Discussion

MNase-seq experiments are still matter of debate (Chereji et al., 2017), compelling evidence strongly demonstrate that MNase-sensitive particles are in fact due to nucleosomes more susceptible to enzymatic digestion (Kubik et al., 2015; Weiner et al., 2010; Xi et al., 2011).

Our integrated analysis on chromatin structure in ecdysone-regulated CREs is based on the data regarding MNase-sensitive nucleosomes detectable by different MNase digestion levels. Although we did not directly assess whether our MNase-sensitive signals are indeed due to nucleosome particles (for example, through ChIP-seq experiments), we carried out a strategy which resembles two established approaches for MNase-sensitivity probing (with small computational differences due to the reduced number of analyzed regions): (1) ratio between sub- and mono-nucleosomal fragments from the same digestion level (Typical), similar to the approach established in the Gaul lab (Renna et al., in preparation). This procedure mostly captures the ease of nucleosome eviction and is less affected by the accessibility of the surrounding environment. Renna and colleagues could identify fragile nucleosomes, and compared them to previous salt fractionation data (Henikoff et al., 2009), demonstrating that they are significantly enriched in unstable H3.3 or H2Av nucleosomes in promoters; (2) ratio between mono-nucleosomal fragments from two different digestion levels (Short and Typical). Comparison between mono-nucleosomal tracks from short and typical digestion has been widely used (Chereji et al., 2015; Kubik et al., 2015; Mieczkowski et al., 2016; Xi et al., 2011). Those studies reported that nucleosomes released in low digestion experiments are mostly enriched in the -1 position of active promoters, and coincide with unstable nucleosomes in accessible regions. In addition, the size of the MNase-sensitive regions we identified is always longer than 150 bp and shorter than 300 bp (data not shown), thereby very likely accommodating only one nucleosome, and such regions are specifically restricted to the peak center. The length of the sensitive regions has been proved to be a determinant for MNase-sensitive nucleosomes to assemble in NDRs of active promoters (Kubik et al., 2017). Hence, for all these reasons, we strongly assume that our strategy indeed captured MNase-sensitive nucleosomes in CREs.

Consequently, we demonstrated that changes in MNase-sensitivity occur in differential DHSs upon ecdysone stimulus. At the center of opening peaks, MNase-sensitive nucleosomes are enriched after stimulus, and the contrary occurs in closing peaks. These mechanisms are highly activity-dependent, however restricted to very inducible DHSs. Notably, we showed not only that known GC content biases very unlikely determine our analysis, but also that the intrinsic barriers

given by the DNA underlying sequence are overcome after the stimulus. This is of particular interest, as we could then ask which other factors are responsible for the observed changes in sensitivity, focusing on ecdysone-regulated TFs.

In conclusion, we present evidence of drastic changes occurring at the chromatin structure level in response to ecdysone. Ecdysone-triggered chromatin remodeling has been already shown in other contexts (Ables and Drummond-Barbosa, 2010; Badenhorst et al., 2005). However, to our knowledge, this the first study that correlates the levels of activity in inducible DHSs and changes at the level of MNase-sensitive nucleosomes upon developmental stimulus in *Drosophila*. Furthermore, we also showed that such a correlation also applies for enhancers, whereas previous studies mainly assessed activity-related sensitivity in promoters of expressed genes. A question that needs further investigation is whether the changes of MNase-sensitive nucleosomes are due to a general increase or decrease in accessibility of the surrounding chromatin environment, or due to changes in biophysical properties that alter the affinity to nucleosomal DNA. With regard to the former, we observed a tight correlation between MNase-sensitivity and accessibility, suggesting that indeed an enhanced openness of the regulatory regions determines the characterization of sensitive nucleosomes. Nevertheless, by using the ration between sub- and mono-nucleosomal fragments, we aimed at detecting nucleosome sensitivity at the level of nucleosomal DNA (as shown in Renna et al., in preparation). It is likely, then, that the two assumptions do not exclude each other.

5.5   EcR and br could function as key players in chromatin structure changes

The dependency of chromatin accessibility and MNase-sensitivity on activity includes different mechanisms, such as involvement of TFs and recruitment of chromatin remodelers (Kubik et al., 2015). Through motif enrichment analysis and knockdown experiments, we postulate that EcR and br could be involved in these mechanisms.

In the case of EcR, we propose its role in driving the changes of MNase-sensitive nucleosomes at the center of opening DHSs, and in maintaining an open state in active DHSs before stimulus. Three types of data point to this assumption: Sub/Mono ratio in typical digestion; ratio of mono-nucleosomal signals from different digestion levels; chromatin accessibility probing. Very likely, EcR does not act alone. In mammals, the pioneer factor FoxA1 dictates the binding location of hormone receptors (Hurtado et al., 2011). Additionally, FoxA1 was proven to

keep nucleosomes in a MNase-accessible status through linker histone displacement in tissue-specific enhancers (Iwafuchi-Doi et al., 2016). It is possible that a similar scenario occurs in EcR-targeted inducible CREs, resulting in accessible chromatin and sensitive nucleosomes. EcR could bind closed regions through the help of a pioneer factor, and subsequent recruitment of chromatin remodelers could promote an accessible chromatin structure. Many studies, already cited in previous paragraphs of this thesis, demonstrated the cooperation of EcR with different chromatin remodelers. However, less is known about its interaction with pioneer factors. In our enrichment analysis, we did not identify any putative pioneer factor. However that could be due to the strict parameters we applied for the TF selection. One putative pioneer factor in *Drosophila*, Trl, is not differentially expressed in ecdysone-stimulated S2 cells, nevertheless it shows high constitutive levels of expression. Trl function was mostly investigated in promoter-proximal regions (Slattery et al., 2014) and never directly associated to EcR. However, their motifs scored high correlation in a pair-wise comparison of TFBSs (Negre et al., 2011), and Trl was detected to physically interact with the NURF-subunit Nurf301 (Xiao et al., 2001), the same chromatin remodeling subunit showed to interplay with EcR (Badenhorst et al., 2005). This topic certainly deserves further elucidation.

With regard to br, its knockdown led to lack of closing chromatin and altered MNase-sensitivity levels in closing DHSs. This is in agreement with all the results obtained in this study that point to the repressive role of br, a function that could lead to closing chromatin. As already mentioned, the roles of the br isoforms vary, and include a repressive action of the Z2 isoform. Nonetheless, we could not find previous evidence that directly associate br to mechanisms for chromatin structure regulation. Our findings about br represent an interesting field to be explored.

Taken together, our data open new scenarios on the functions of ecdysone pathway components in regulating chromatin structure.

5.6   A deep characterization of regulatory dynamics during metamorphosis

Arguably, the most interesting question regarding the ecdysone cascade is how the hormone can lead to context-specific morphological changes *in vivo*. Here, we addressed this question through chromatin accessibility landscape and TF lexicon characterization with unprecedented spatio-temporal resolution. Previous studies also followed a similar strategy, but we believe that our approach succeeded in capturing tissue-specific inputs that others missed.

With regard to ecdysone response, Li and White performed a tissue-specific gene expression profiling, followed by a computational identification of motifs, in order to construct genomic networks (Li and White, 2003). However, motif enrichment was run only in SG, and the CRE selection was simply based on a fixed distance from promoters of differentially expressed genes. Uyehara and colleagues defined distinct combinations of hormone-induced TFs that dictate the timing of WD development (Uyehara et al., 2017). Nevertheless, their TF set includes mainly known hormone-induced early TFs, completely lacking context-dependent resolution and contribution of additional regulators.

Finally, McKay and Lieb carried out a very extensive analysis of accessibility landscapes in the fly appendages, from their progenitor cells in embryo to late larval stages (McKay and Lieb, 2013). They claimed that appendages share the same DHS landscapes and dynamics, with the exception of their master regulator loci. As generally the fly appendages share the same cell fate, meaning that they all go through a process of proliferation and differentiation, this is in accordance with our findings, which demonstrate strong accessibility similarities between tissues that share the same survival path (WD and ED). Nevertheless, we could additionally provide landscapes with higher resolution (thanks to ATAC-seq as opposed to FAIRE-seq), as well as a more comprehensive comparison with tissues that do not share the same cell fate.

Therefore, our data represent the deepest and most comprehensive analysis on accessibility landscapes and TF inputs upon metamorphosis. Thanks to the intercrossed examination of four different tissues, we highlighted distinct TFs that were never previously considered as part of the context-dependent response *in vivo*.

In conclusion, our *in vivo* dissection of genomic dynamics during metamorphosis represents a great resource for the scientific community. We are confident that the produced data will be essential for subsequent and more mechanistic investigations to further elucidate the ecdysone tissue-specific response.

5.7 Outlook

5.7.1 Footprinting

Originally, this project aimed at using DNase-seq and ATAC-seq not only to detect CRE dynamics, but also and more importantly to identify TFBEs. When a TF binds the DNA, it prevents the DNase I or Tn5 from accessing its binding site. Thus, when looking at an open region with a

high frequency of cleavages, it is possible to distinguish small regions (8-20 bp in length) with a reduced number of cutting or tagmenting events, which coincide with TFBEs occurring at that time. Those are called TF footprints. DNase footprinting was already popular in the 70s to characterize TFBSs *in vitro* (Galas and Schmitz, 1978), although with many experimental limitations. With the advent of DNase-seq, it became immediately clear that a genome-wide identification of footprints *in vivo* would have started a new era for TFBE characterization, by quantitatively assessing multiple TFBEs in a single experiment with no need for good-quality antibodies or *a priori* knowledge, such as PWMs (Neph et al., 2012). Despite the initial enthusiasm, DNase- and ATAC-seq footprinting revealed many drawbacks: (1) the intrinsic biases of the enzymes, which do not affect CRE identification, become evident when reaching the sequencing depth necessary for the footprinting (~250 million pair-end reads), and can be incorrectly interpreted as patterns induced by TF binding (He et al., 2014; Sung et al., 2014); (2) TFs with shorter dwell time do not produce a detectable footprint (Sung et al., 2014); (3) most of the studies showed a so called metafootprint, which is a composite plot of the cleavage frequencies in thousands of binding sites of a specific TF, as opposed to the more informative footprints at single binding sites.

Marta Bozek (a PhD student in the Gaul lab) and I produced many sequencing DNase-seq and ATAC-seq libraries with the intent of characterizing footprints in S2 cells. Along the aforementioned issues, we observed a general difficulty in detecting footprints at single binding sites. To overcome these problems, we are currently improving the experimental and computational procedures to reliably detect footprints: (1) we are testing different fixation conditions to maximize the recovery of binding events, also in case of TFs with short dwell time; (2) we are improving the recovery of small fragments from open regions, to increase the resolution given by the reduced number of cleavages in binding sites as opposed to flanking regions; (3) together with Roberto Cortini, we are working on a computational workflow to eliminate the known intrinsic biases of the enzymes.

With these improvements, we aim at establishing a better protocol for DNase- and ATAC-seq footprinting that will hopefully provide unprecedented information on occurring TFBEs without relying on PWMs.

5.7.2   Modeling gene expression

The quantity, quality and diversity of data produced in this study are a great source for many follow up projects. Arguably, an approach for gene expression modeling can fully take advantage by the generated data, thereby representing the natural next step to be accomplished. Deciphering the system behind the complex regulation of gene expression is a challenging task in computational biology. Nevertheless, the information provided by this project is unique: (1) TFBEs are characterized and the regions in which they occur well defined, resulting in high resolution input information. Notably, we can also assume a "negative" effect of some TFs by including PWMs detected in closing DHSs, or by assessing their association to down-regulated genes; (2) the expression levels provide a biological readout on which the model can be based. In addition, we can use the dynamics of the S2 cell paradigm (thousands of differential DHSs, TFBEs and genes) to extend the source of information and attempt to predict gene expression behaviors upon a developmental stimulus.

Therefore, Roberto Cortini, as part of his PhD thesis and with my help, is working on a strategy to model gene expression. The modeling approach we are using takes inspiration by and is built upon Schmidt et al., 2017. Taken two different time points, we try to model the relationship between FC of genes and TFs affinity ratios. For each gene and for each time point, we take the peaks that are associated to a gene and use them to compute a TF affinity for each TF. Then we compute the ratios of such affinities and use them in a regularized linear regression to fit a linear model to the gene FC.

In this way, for each TF we have a fitted coefficient that gives us a hint on the effect of that TF on the system: a positive coefficient suggests an activating effect, whereas a negative coefficient suggests a repressive effect. In conclusion, we aim at providing a model for gene expression regulation built on an unprecedented set of data.

6 APPENDIX

6.1 Tables

Table 1: List of TFs for motif enrichment analysis in S2 cells.

br-Z1	Eip74EF	hng3	Max_Mnt
br-Z2	Eip75B	Hr39	NFAT
br-Z3	Eip78C	Hr46	pnr
br-Z4	Eip93F	Hr4	schlank_Lag1
Btn	Ets21c	Hr51	shn-F1-2
CG5953	foxo	kay_Jra	Sox14
chinmo	fru	ken	srp
dl	ftz-f1	lola-PC	tai_Clk
EcR	h	lola-PO	twi_da
EcR-USP	Hnf4	lola-PY	vri

Table 2: List of TFs for motif enrichment analysis in WD.

abd-A	da	Hsf	Poxn
Abd-B	Dfd	hth	prd
ac	disco	inv	Ptx1
Adf1	disco-r	jim	Rel
Aef1	Dll	Kah	retn
al	Doc2	kay	rib
aop	dpn	ken	salr
ap	Dr	kni	sc
ara	dsx	knrl	schlank
Atf6	dysf	Kr	Scr
ato	E(spl)m3-HLH	lab	sd
Awh	E(spl)m8-HLH	lbe	sens
bcd	E(spl)mbeta-HLH	lbl	shn
Bgb	E(spl)mdelta-HLH	Lim1	sim
B-H1	E5	lola	slp1
B-H2	ecr_usp_ja	Mad	sob
Blimp-1	eg	Max	Sox14
bowl	Eip74EF	Med	Sox15
br	Eip75B	mirr	Sp1
brk	Eip78C	Mnt	sqz
btd	ems	Mondo	sr
byn	en	net	ss
cad	esg	NK7.1	su(Hw)
caup	Ets21C	nub	sv
CG12236	Ets98B	oc	svp
CG15812	exex	odd	tai
CG3838	foxo	opa	toy
CG3919	fru	Optix	Trl
CG4360	ftz-f1	otp	ttk
CG5953	GATAd	ovo	tup
CG8765	gl	pad	twi
chinmo	grh	pan	unpg
ci	gsb	pb	Usf
crc	gsb-n	pdm2	vri
crol	h	peb	vvl
crp	Hand	pho	wor
cwo	Hnf4	phol	Xrp1
cyc	hng3	pnr	z
D	Hr3	pnt	ZIPIC
D19B	Hr78	Poxm	zld

Appendix

Table 3: List of TFs for motif enrichment analysis in ED.

abd-A	Dlip3	Hr78	retn
ac	Dll	hth	rib
Adf1	Doc2	jim	sc
al	dsx	kay	Scr
ara	dysf	klu	sd
Atf6	E(spl)m3-HLH	knrl	sens
ato	E(spl)m8-HLH	Kr	shn
Awh	E(spl)mbeta-HLH	lab	sim
Bgb	E(spl)mdelta-HLH	lbl	slp1
B-H1	eg	Lim1	sna
B-H2	Eip74EF	Lim3	Sox14
Blimp-1	Eip75B	Lmx1a	Sox15
bowl	Eip78C	lola	sr
br	ems	Mad	ss
brk	en	Max	Su(H)
byn	esg	Mnt	su(Hw)
cad	Ets21C	NK7.1	sv
caup	Ets65A	nub	svp
CG3838	eve	oc	tai
CG3919	exex	odd	tin
CG5953	fru	opa	tj
CG8765	ftz-f1	otp	toy
chinmo	grh	ovo	ttk
ci	gsb	pan	tup
crc	gsb-n	pnr	Ubx
cwo	h	pnt	unpg
D	H2.0	Poxm	vri
da	Hand	Poxn	Xrp1
Deaf1	Hnf4	prd	zld
Dfd	hng3	Rel	
disco	Hr3	repo	

Table 4: List of TFs for motif enrichment analysis in CNS.

abd-A	da	HGTX	Rel
Abd-B	Deaf1	HLH4C	repo
ac	Dfd	Hnf4	retn
achi	dimm	hng3	rib
Adf1	Dlip3	Hr3	run
Aef1	Dll	Hr78	salr
al	dpn	Hsf	sc
aop	Dr	hth	schlank
ap	dsx	ind	Scr
ara	E(spl)m3-HLH	inv	scrt
Asciz	E(spl)m8-HLH	jim	sd
ato	E(spl)mbeta-HLH	kay	sens
Awh	E(spl)mdelta-HLH	ken	shn
bab1	E(spl)mgamma-HLH	klu	Sidpn
bap	EcR	Kr	sim
bcd	eg	lab	Six4
Bgb	Eip74EF	lbl	slou
B-H1	Eip75B	Lim1	slp1
B-H2	Eip78C	Lim3	slp2
Blimp-1	ems	Lmx1a	sna
bowl	en	lola	sob
br	erm	luna	Sox14
brk	ERR	Mad	Sp1
bsh	esg	Max	sr
btd	Ets21C	mirr	ss
cad	Ets65A	Mnt	Su(H)
caup	eve	Mondo	sv
CG12236	exd	net	svp
CG12605	exex	NK7.1	tai
CG15812	ey	nub	tap
CG3838	Fer1	oc	tin
CG3919	Fer2	odd	tj
CG4328	Fer3	Oli	tll
CG4360	fkh	onecut	toy
CG4404	foxo	opa	Trl
CG5953	fru	Optix	ttk
CG7368	ftz-f1	otp	tup
CG8765	GATAd	ovo	Ubx
chinmo	gl	pad	unpg
ci	gsb	pan	Usf
crc	gsb-n	pb	vri
crol	gt	PHDP	vvl
crp	h	pho	wor
cyc	Hand	pnr	Xrp1
D	hb	pnt	ZIPIC
D19A	hbn	Poxn	zld
D19B	Hey	Ptx1	

Table 5: List of TFs for motif enrichment analysis in SG.

abd-A	D19A	hng3	Poxn
Abd-B	D19B	Hr3	prd
ac	da	Hr78	Ptx1
achi	Deaf1	Hsf	Rel
Adf1	Dfd	hth	repo
Aef1	disco	inv	rib
al	disco-r	jim	salr
aop	Dlip3	Jra	sc
ap	Dll	kay	schlank
ara	Doc2	ken	Scr
Asciz	Dr	klu	sd
Atf6	dsx	kni	sens
Awh	dysf	knrl	shn
bcd	E(spl)m3-HLH	Kr	slbo
Bgb	E(spl)m7-HLH	lab	slp1
B-H1	E(spl)mbeta-HLH	lbe	slp2
B-H2	E5	lbl	sob
bin	EcR	Lim1	Sox14
Blimp-1	eg	Lim3	Sox15
bowl	Eip74EF	lola	Sp1
br	Eip75B	luna	Spps
brk	Eip78C	Mad	sqz
btd	ems	Max	sr
cad	en	Med	ss
caup	erm	mirr	Su(H)
CG11617	ERR	Mnt	su(Hw)
CG12236	esg	Mondo	sug
CG15812	Ets21C	net	sv
CG3407	Ets97D	NFAT	svp
CG3838	Ets98B	NK7.1	tai
CG3919	eve	nub	tin
CG4360	exd	oc	toy
CG5953	exex	odd	Trl
CG6272	ey	opa	ttk
CG6276	fkh	Optix	tup
CG8765	foxo	otp	Ubx
chinmo	fru	ovo	unpg
ci	ftz-f1	pad	vri
Clk	GATAd	pan	vvl
crc	gl	pb	wor
CrebA	grh	pdm2	Xrp1
crol	gsb	peb	z
crp	gsb-n	pho	Zif
cwo	h	pnr	zld
cyc	hb	pnt	
D	Hnf4	Poxm	

Table 6: Adjusted p-values (log$_{10}$) of TFs in WD. NA = not enriched.

TF	Late 3IL opening DHSs	Late 3IL closing DHSs	WPP opening DHSs	WPP closing DHSs
AbdA_SOLEX	-27.1	NA	-36.2	NA
AbdB_SOLEX	-91.9	-8.4	-66.7	-28.4
ac_da_SANG	-30.4	-28.2	-32.4	-21.2
Adf1_SANGE	-92.0	-66.9	-52.7	-141.5
Aef1_SOLEX	-249.8	-153.4	-153.9	-316.9
Al_SOLEXA_	-2.3	NA	-16.5	NA
amos_da_SA	-8.5	-15.6	-9.8	-10.9
Ap_SOLEXA_	-12.2	NA	-31.6	NA
Ara_SOLEXA	-8.8	-3.2	NA	-7.6
ase_da_SAN	-18.5	-29.0	-28.9	-15.0
ato_da_SAN	-4.7	-16.8	-4.1	-15.3
Awh_SOLEXA	NA	NA	-11.9	NA
BH1_SOLEXA	-21.9	NA	-45.0	NA
BH2_SOLEXA	-8.3	NA	-17.5	NA
Blimp-1_SO	-149.0	-56.4	-102.3	-172.6
bowl_SOLEX	-2.3	-7.5	-12.9	NA
brk_FlyReg	-5.2	-10.4	NA	-15.0
br-Z1_FlyR	-62.6	-40.7	-33.2	-99.3
br-Z2_FlyR	NA	-5.5	NA	-6.8
btd_NAR_FB	-73.3	-13.2	-26.3	-60.1
Cad_SOLEXA	-97.8	-16.8	-74.9	-41.2
cato_da_SA	NA	-17.5	-1.6	-10.3
Caup_SOLEX	-1.6	NA	NA	NA
CG13897_SA	-37.7	-39.4	-11.3	-102.9
CG33557_da	-71.7	-45.5	-39.4	-72.8
CG3838_SAN	-3.4	NA	NA	-4.9
CG4360-F1-	-219.8	-141.5	-143.7	-277.4
CG5953_SAN	-17.4	-24.6	NA	-92.2
CG8765_SAN	-39.7	-21.7	-17.4	-59.6
chinmo_SOL	NA	NA	-9.9	NA
ci_SOLEXA_	-47.8	-51.1	-10.1	-99.6
crol-F7-16	-174.9	-27.2	-54.2	-146.9
crp_SANGER	-1.7	-19.5	-8.3	-10.3
D_NAR_FBgn	-5.4	-19.0	NA	-31.8
da_SANGER_	-7.4	-18.9	-4.8	-15.8
dei_da_SAN	NA	-2.9	NA	-2.9
dimm_da_SA	NA	NA	NA	-3.3
Dll_SOLEXA	-5.1	NA	-25.9	NA

Dr_SOLEXA_	-3.1	NA	-25.6	NA
E5_SOLEXA_	-21.7	NA	-34.4	NA
ecr_usp_ja	NA	NA	-2.1	NA
eg_SANGER_	-4.7	-13.2	-3.0	-11.6
Ems_SOLEXA	-3.0	NA	-18.0	NA
En_SOLEXA_	-18.5	NA	-37.4	NA
Exex_SOLEX	-6.7	NA	-27.4	NA
Fer1_da_SA	-1.3	-16.0	-3.6	-7.5
Fer3_da_SA	-1.4	NA	-5.3	NA
foxo_SANGE	-37.5	-41.2	-8.4	-58.9
fru_SOLEXA	-14.3	NA	-28.6	NA
HLH4C_da_S	-46.4	-34.8	-37.4	-32.8
HLH54F_da_	-5.8	-10.4	-3.9	-10.7
Hsf_FlyReg	NA	-1.4	NA	NA
inv_SOLEXA	NA	NA	-10.8	NA
jim_F1-9_S	-260.0	-142.0	-110.5	-inf
ken_SOLEXA	NA	NA	-10.4	NA
kni_SANGER	-3.1	-9.8	-6.2	-6.4
Kr_SOLEXA_	-12.0	NA	NA	NA
l(1)sc_da_	-18.9	-14.7	-36.8	-6.2
Lab_SOLEXA	-16.1	NA	-33.8	NA
Lag1_SOLEX	NA	-1.7	NA	NA
Lbe_SOLEXA	NA	NA	-5.4	NA
Lbl_SOLEXA	-3.3	NA	-16.5	NA
Lim1_SOLEX	-17.8	NA	-36.0	NA
lola-PD_SO	-95.9	-21.2	-28.3	-79.6
lola-PF_SO	-8.4	NA	-11.6	NA
lola-PL_SO	-35.2	-13.0	-19.8	-22.8
lola-PO_SO	NA	-12.4	NA	-4.0
lola-PQ_SO	-2.0	-12.8	NA	-35.0
Mad_FlyReg	-34.3	-26.2	-24.9	-60.3
Med_FlyReg	-23.7	-27.6	-24.4	-51.9
Mirr_SOLEX	-11.8	-2.4	NA	-6.3
nau_da_SAN	NA	-3.6	NA	NA
net_da_SAN	-7.9	-22.8	-14.6	-13.5
NK7.1_SOLE	-5.3	NA	-23.1	NA
nub_SOLEXA	-10.1	-5.2	-3.3	-15.3
odd_NBT_5_	-14.2	-7.5	-30.2	-1.4
opa_SOLEXA	-40.6	-9.7	-2.6	-51.8
Otp_SOLEXA	-10.3	NA	-28.9	NA
pad_SOLEXA	-40.4	-12.6	-13.6	-36.8

Pb_SOLEXA_	-4.4	NA	-19.9	NA
pdm2_SOLEX	NA	NA	NA	-3.8
pfk_SANGER	-4.9	NA	-1.9	NA
pnt_SANGER	NA	-1.8	-2.8	NA
Rel_SANGER	NA	-6.7	NA	-12.2
retn_SANGE	-28.3	-14.8	-12.9	-18.7
run_Bgb_NB	-10.6	-5.7	-26.9	-6.8
sage_da_SA	-13.6	-21.0	-15.9	-15.9
sc_da_SANG	-27.6	-25.9	-36.5	-17.8
Scr_SOLEXA	-21.2	NA	-37.0	NA
slp1_NAR_F	-152.7	-169.6	-54.0	-251.4
sob_SOLEXA	-8.0	-6.4	-21.4	NA
Sox14_SANG	NA	-2.7	NA	-1.7
Sox15_SANG	NA	-9.2	NA	-17.7
Sp1_SOLEXA	-49.4	-11.0	-11.1	-52.4
sqz_SOLEXA	-86.8	-27.4	-22.2	-110.3
sr_SOLEXA_	-97.1	-36.0	-28.2	-101.0
suHw_FlyRe	-4.4	NA	NA	-3.5
tap_da_SAN	-12.9	-4.6	-12.4	-6.5
toy_FlyReg	-9.1	NA	NA	-16.2
Trl_FlyReg	-210.2	-83.1	-59.5	-313.6
ttk-PA_SOL	-8.3	NA	NA	-7.0
Tup_SOLEXA	NA	NA	-14.1	NA
Unpg_SOLEX	-5.0	NA	-24.3	NA
vvl_SOLEXA	-8.3	NA	-2.6	-5.6
z_FlyReg_F	-13.9	NA	NA	-20.9

Table 7: Adjusted p-values (log₁₀) of TFs in ED. NA = not enriched.

TF	Late 3IL opening DHSs	Late 3IL closing DHSs	WPP opening DHSs	WPP closing DHSs
AbdA_SOLEX	-34.6	NA	-40.0	-4.1
ac_da_SANG	-27.5	-7.5	-27.7	-18.4
Adf1_SANGE	-83.7	-44.0	-78.2	-79.8
Al_SOLEXA_	-10.6	NA	-11.0	NA
amos_da_SA	-10.6	-3.6	-13.7	-10.7
Ara_SOLEXA	NA	NA	NA	-2.7
ase_da_SAN	-22.6	-7.4	-24.2	-15.3
ato_da_SAN	-7.3	-6.8	-12.9	-10.4
Awh_SOLEXA	-6.3	NA	-7.2	NA
BH1_SOLEXA	-29.4	NA	-27.3	NA
BH2_SOLEXA	-10.2	NA	-10.7	NA
Blimp-1_SO	-180.2	-35.3	-192.5	-89.6
bowl_SOLEX	-10.7	NA	-9.4	NA
brk_FlyReg	NA	-3.6	NA	-4.9
br-Z1_FlyR	-56.4	-20.4	-49.6	-43.5
Cad_SOLEXA	-91.8	-15.1	-95.3	-34.8
cato_da_SA	-4.5	-3.4	-9.0	-4.7
CG13897_SA	-30.7	-41.3	-31.7	-67.0
CG32105_SO	-66.6	-6.7	-67.7	-19.6
CG33557_da	-65.1	-23.8	-66.8	-48.3
CG3838_SAN	NA	NA	NA	-1.4
CG5953_SAN	NA	-31.8	NA	-56.8
CG8765_SAN	-45.8	-16.8	-45.0	-38.0
chinmo_SOL	-9.9	NA	-6.7	NA
ci_SOLEXA_	-28.7	-17.6	-25.1	-54.4
D_NAR_FBgn	NA	-7.8	NA	-15.6
da_SANGER_	-3.1	-4.7	-1.8	-12.5
Deaf1_FlyR	-29.5	-9.6	-19.8	-15.1
Dll_SOLEXA	-18.5	NA	-18.6	NA
eg_SANGER_	-2.6	-3.6	-6.1	-7.6
Ems_SOLEXA	-12.1	NA	-15.6	NA
En_SOLEXA_	-31.0	NA	-31.9	-3.3
Eve_SOLEXA	-4.4	NA	-4.8	NA
Exex_SOLEX	-21.2	NA	-21.5	NA
Fer1_da_SA	-3.1	-1.8	-4.0	-5.9
Fer3_da_SA	-3.5	NA	-5.5	NA
fru_SOLEXA	-33.0	NA	-34.0	NA
H2.0_SOLEX	-73.6	-8.8	-76.3	-23.0

HLH4C_da_S	-39.2	-11.9	-38.9	-26.2
HLH54F_da_	-1.5	-1.4	NA	-8.6
jim_F1-9_S	-183.2	-119.0	-163.2	-236.4
klu_SOLEXA	-163.8	-70.9	-131.8	-214.2
Kr_SOLEXA_	-8.9	NA	-4.3	NA
l(1)sc_da_	-18.7	NA	-23.5	-4.4
Lab_SOLEXA	-28.2	NA	-31.1	-3.4
Lbl_SOLEXA	-13.3	NA	-12.0	NA
Lim1_SOLEX	-30.3	NA	-29.4	-3.2
Lim3_SOLEX	-4.4	NA	-3.6	NA
lola-PC_SO	-1.3	NA	-7.6	NA
lola-PD_SO	-62.2	-14.7	-49.6	-37.6
lola-PF_SO	-8.6	NA	-8.6	NA
lola-PL_SO	-25.8	-3.5	-30.7	-7.0
lola-PO_SO	NA	NA	NA	-6.3
lola-PQ_SO	NA	-9.0	NA	-20.5
Mad_FlyReg	-26.7	-11.5	-30.6	-22.3
nau_da_SAN	NA	-1.5	NA	NA
net_da_SAN	-7.9	-2.8	-10.1	-6.4
NK7.1_SOLE	-12.2	NA	-18.4	NA
nub_SOLEXA	-42.5	-6.2	-30.6	-17.7
odd_NBT_5_	-16.6	NA	-19.5	-1.4
opa_SOLEXA	-18.0	-3.7	-10.4	-24.9
Otp_SOLEXA	-23.1	NA	-23.4	-1.4
pnt_SANGER	NA	NA	-5.0	NA
Repo_SOLEX	-32.1	NA	-32.4	-4.0
retn_SANGE	-15.1	-7.1	-14.1	-20.0
run_Bgb_NB	-47.5	-3.4	-60.9	-8.4
sage_da_SA	-12.5	-4.3	-14.1	-7.0
sc_da_SANG	-28.7	-7.1	-25.9	-16.3
Scr_SOLEXA	-34.9	NA	-40.2	-2.4
slp1_NAR_F	-111.6	-64.7	-105.2	-149.9
sna_SOLEXA	-7.5	-1.8	-6.7	NA
Sox14_SANG	NA	-2.0	NA	-1.8
Sox15_SANG	NA	-4.6	NA	-5.9
sr_SOLEXA_	-49.8	-24.0	-40.8	-69.8
suHw_FlyRe	-3.5	NA	-3.0	-2.1
tap_da_SAN	-10.5	NA	-15.5	-3.0
Tin_SOLEXA	-6.0	NA	-8.2	NA
tj_SANGER_	-9.0	-2.3	-11.3	-2.2
toy_FlyReg	-5.3	NA	NA	-7.5

Tup_SOLEXA	-7.4	NA	-11.5	NA
Ubx_SOLEXA	-67.3	-6.5	-70.2	-16.4
Unpg_SOLEX	-16.9	NA	-17.3	NA

Table 8: Adjusted p-values (log₁₀) of TFs in CNS. NA = not enriched.

TF	Late 3IL opening DHSs	Late 3IL closing DHSs	WPP opening DHSs	WPP closing DHSs
AbdA_SOLEX	-55.0	NA	-22.6	NA
AbdB_SOLEX	-185.8	NA	-88.9	NA
ac_da_SANG	-50.9	NA	-35.9	NA
Achi_SOLEX	-inf	NA	NA	NA
Adf1_SANGE	-225.4	NA	-147.5	-14.3
Aef1_SOLEX	-inf	NA	-inf	-34.1
Al_SOLEXA_	-9.1	NA	NA	NA
amos_da_SA	-10.9	-2.3	-2.6	-6.2
aop_SANGER	NA	NA	-6.4	NA
Ap_SOLEXA_	-26.8	NA	-8.4	NA
Ara_SOLEXA	-10.2	NA	-4.5	NA
ase_da_SAN	-31.6	NA	-14.9	NA
ato_da_SAN	-18.4	-2.6	-6.3	-3.6
bab1_SANGE	-inf	-11.0	-inf	-81.1
Bcd_SOLEXA	-inf	NA	NA	NA
BH1_SOLEXA	-47.6	NA	-21.5	NA
BH2_SOLEXA	-20.9	NA	-6.1	NA
Blimp-1_SO	-inf	-15.8	-304.2	-96.9
bowl_SOLEX	-39.8	NA	-43.6	NA
br_SOLEXA_	-inf	NA	NA	NA
brk_FlyReg	-18.3	-4.7	-13.0	-6.2
br-PE_SOLE	-inf	NA	NA	NA
br-PL_SOLE	-inf	NA	NA	NA
br-Z1_FlyR	-276.4	-2.9	-173.5	-18.0
br-Z4_Hipf	-inf	NA	NA	NA
btd_NAR_FB	-inf	NA	-193.3	-34.9
Cad_SOLEXA	-225.1	NA	-108.0	NA
cato_da_SA	-13.4	NA	-1.6	NA
CG12236_SO	-inf	NA	NA	NA
CG12605_SO	-inf	NA	NA	NA
CG13897_SA	-146.8	-2.0	-91.1	-18.0
CG14962_SO	-inf	NA	NA	NA
CG31670_SO	-100.6	NA	-68.1	-24.3
CG32105_SO	-151.9	NA	-70.7	NA
CG33557_da	-152.1	-2.6	-102.8	-20.8
CG3838_SAN	-4.2	NA	-7.3	NA
CG4328_SOL	-200.8	NA	-96.5	NA
CG4360-F1-	-inf	NA	-308.4	-28.1

CG4404_SAN	-42.6	NA	-40.9	NA
CG5953_SAN	-65.1	-2.1	-25.2	-25.9
CG7368_SOL	-inf	-23.2	-inf	-121.9
CG8765_SAN	-84.0	NA	-48.2	-6.3
chinmo_SOL	NA	NA	-4.3	NA
ci_SOLEXA_	-267.5	-5.0	-156.0	-56.8
Clk_cyc_SA	-inf	NA	NA	NA
Crc_CG6272	-inf	NA	NA	NA
crol-F7-16	-inf	-6.1	-217.9	-57.9
crp_SANGER	-4.5	-3.0	NA	-3.0
D_NAR_FBgn	-95.4	NA	-63.9	NA
D19A_F10-1	-6.0	NA	NA	NA
D19B-F10-1	-inf	NA	NA	NA
da_SANGER_	-23.0	NA	-14.5	-2.1
Deaf1_FlyR	-60.9	NA	-40.6	NA
Dip3_SANGE	-inf	NA	NA	NA
Dll_SOLEXA	-16.1	NA	-4.3	NA
dm_Max_SAN	-inf	NA	NA	NA
Dr_SOLEXA_	-6.2	NA	NA	NA
E(spl)_SAN	-inf	NA	NA	NA
ecr_Hipfa	-inf	NA	NA	NA
eg_SANGER_	-24.7	NA	-13.1	-8.6
Eip75B_SAN	-inf	NA	NA	NA
Eip78C_SAN	-inf	NA	NA	NA
Ems_SOLEXA	-11.9	NA	-3.0	NA
En_SOLEXA_	-52.7	NA	-18.4	NA
ERR_SANGER	-inf	NA	NA	NA
esg-F3-5_S	-inf	NA	NA	NA
Ets21c_SAN	NA	NA	-2.8	NA
Exd_SOLEXA	-inf	NA	NA	NA
Exex_SOLEX	-14.2	NA	-2.8	NA
ey_SOLEXA_	-144.5	-3.0	-63.2	-47.7
Fer1_SANGE	-14.5	NA	NA	NA
Fer3_da_SA	-3.8	NA	NA	NA
fkh_NAR_FB	-176.9	NA	-101.8	-10.1
foxo_SANGE	-108.2	NA	-72.7	-12.1
fru_SOLEXA	-19.0	NA	-19.3	NA
ftz-f1_SAN	-inf	NA	NA	NA
GATAd_SANG	-inf	NA	NA	NA
gsb-n_SOLE	-inf	NA	NA	NA
gt_NAR_FBg	-inf	NA	NA	NA

hb_SOLEXA_	-45.0	-1.6	-21.6	-21.3
Hbn_SOLEXA	-10.7	NA	-1.6	NA
Hey_SANGER	-14.4	NA	-11.9	NA
Hgtx_SOLEX	-73.4	NA	-39.4	NA
HLH4C_SANG	-31.2	NA	-15.8	NA
HLH54F_da_	-2.5	NA	-1.6	NA
HLHm3_SANG	-inf	NA	NA	NA
HLHmbeta_S	-inf	NA	NA	NA
HLHmdelta_	-inf	NA	NA	NA
HLHmgamma_	-inf	NA	NA	NA
Hnf4_SANGE	-inf	NA	NA	NA
Hr46_SANGE	-inf	NA	NA	NA
Hr78_SANGE	-inf	NA	NA	NA
Hth_SOLEXA	-inf	NA	NA	NA
jim_F1-9_S	-inf	-4.9	-291.2	-56.8
kay_Jra_SA	-inf	NA	NA	NA
ken_SOLEXA	-23.3	NA	-17.7	-6.9
klu_SOLEXA	-inf	-12.0	-inf	-156.2
Kr_SOLEXA_	-57.3	NA	-43.7	-7.0
l(1)sc_da_	-13.1	NA	-11.9	NA
Lab_SOLEXA	-33.4	NA	-11.8	NA
Lbl_SOLEXA	-2.3	NA	NA	NA
Lim1_SOLEX	-43.5	NA	-15.3	NA
lola_SANGE	-inf	NA	NA	NA
lola_SOLEX	-inf	NA	NA	NA
lola-PA_SO	-inf	NA	NA	NA
lola-PC_SO	-150.9	-1.7	-115.4	-4.9
lola-PD_SO	-77.5	-3.0	-31.3	-28.6
lola-PF_SO	-18.4	NA	-12.2	NA
lola-PG_SO	-inf	NA	NA	NA
lola-PK_SO	-inf	NA	NA	NA
lola-PL_SO	-76.2	-10.2	-43.4	-27.8
lola-PQ_SO	-34.7	NA	-18.8	NA
lola-PT_SO	-inf	NA	NA	NA
lola-PU_SO	-inf	NA	NA	NA
luna_SOLEX	-133.6	-3.4	-71.1	-24.1
Mad_FlyReg	-111.3	-1.9	-89.4	-2.7
Max_Mnt_SA	-inf	NA	NA	NA
Mio_bigmax	-inf	NA	NA	NA
Mirr_SOLEX	-9.9	NA	-5.8	NA
net_da_SAN	-17.7	-2.8	-9.4	-3.5

NK7.1_SOLE	-19.8	NA	-7.0	NA
nub_SOLEXA	-133.4	NA	-95.8	-9.7
Oc_SOLEXA_	-inf	NA	NA	NA
odd_NBT_5_	-31.8	NA	-35.2	NA
opa_SOLEXA	-167.8	-5.2	-97.4	-34.4
Optix_SOLE	-inf	NA	NA	NA
Otp_SOLEXA	-23.8	NA	-6.2	NA
ovo_SOLEXA	-inf	NA	NA	NA
pad_SOLEXA	-129.2	NA	-97.4	-7.1
pan_FlyReg	-inf	NA	NA	NA
Pb_SOLEXA_	-7.7	NA	NA	NA
pfk_SANGER	-22.4	NA	-6.1	NA
PhdP_SOLEX	-44.1	NA	-14.4	NA
pho_SOLEXA	-12.8	-1.7	-7.9	NA
pnt_SANGER	-4.7	NA	-6.7	NA
Poxn_SOLEX	-inf	NA	NA	NA
Ptx1_SOLEX	-inf	NA	NA	NA
Rel_SANGER	-19.6	NA	-13.7	-4.4
Repo_SOLEX	-55.8	NA	-21.5	NA
retn_SANGE	-69.6	NA	-37.3	NA
run_Bgb_NB	-13.3	NA	-5.8	NA
sage_da_SA	-21.6	NA	-15.3	-1.6
sc_da_SANG	-42.3	-2.0	-33.0	NA
Scr_SOLEXA	-43.2	NA	-17.1	NA
scrt_SOLEX	-inf	NA	NA	NA
sens_SOLEX	-inf	NA	NA	NA
Six4_SOLEX	-inf	NA	NA	NA
Slou_SOLEX	-34.0	NA	-12.4	NA
slp1_NAR_F	-inf	-2.4	-313.4	-51.7
slp2_SANGE	-220.9	NA	-124.5	-8.0
sob_SOLEXA	-25.3	NA	-31.7	NA
Sox14_SANG	-31.0	NA	-22.4	NA
Sp1_SOLEXA	-287.8	-2.4	-144.5	-34.8
sr_SOLEXA_	-inf	-5.2	-261.4	-67.4
ss_tgo_SAN	-inf	NA	NA	NA
SuH_FlyReg	-inf	NA	NA	NA
sv_SOLEXA_	-inf	NA	NA	NA
svp_SANGER	-inf	NA	NA	NA
tap_da_SAN	-35.2	NA	-27.9	-2.0
tgo_cyc_SA	-inf	NA	NA	NA
tgo_sim_SA	-inf	NA	NA	NA

tgo_ss_SAN	-inf	NA	NA	NA
tgo_tai_SA	-inf	NA	NA	NA
Tin_SOLEXA	NA	NA	-2.4	NA
tj_SANGER_	-15.8	NA	-10.0	NA
tll_NAR_FB	-23.7	NA	-8.6	NA
toy_FlyReg	-25.8	-1.3	-8.7	-20.7
Trl_FlyReg	-inf	-9.4	-228.8	-120.9
ttk_NAR_FB	-3.1	NA	-4.7	NA
ttk-PA_SOL	-102.1	-1.6	-66.9	-24.0
ttk-PF_SOL	-17.3	NA	-20.4	NA
Tup_SOLEXA	-9.0	NA	NA	NA
twi_da_SAN	-inf	NA	NA	NA
Ubx_SOLEXA	-128.7	NA	-60.1	NA
Unpg_SOLEX	-14.2	NA	-2.1	NA
Usf_SANGER	-inf	NA	NA	NA
vfl_SOLEXA	-inf	NA	NA	NA
vri_SANGER	-inf	NA	NA	NA
vvl_SOLEXA	-92.4	NA	-84.3	-2.6
wor_SOLEXA	-inf	NA	NA	NA
Xrp1_CG627	-inf	NA	NA	NA

Table 9: Adjusted p-values (log₁₀) of TFs in SG. NA = not enriched.

TF	Late 3IL opening DHSs	Late 3IL closing DHSs	WPP opening DHSs	WPP closing DHSs
AbdA_SOLEX	NA	NA	-13.3	NA
AbdB_SOLEX	-8.3	NA	-58.2	-5.5
ac_da_SANG	-72.4	-47.4	-149.0	-110.9
Adf1_SANGE	-54.4	-39.7	-195.0	-40.3
Aef1_SOLEX	-129.3	-71.7	-inf	-61.3
amos_da_SA	-41.6	-39.1	-72.9	-70.2
aop_SANGER	-2.1	NA	NA	NA
Ara_SOLEXA	-1.8	-4.8	-51.4	NA
ase_da_SAN	-56.1	-46.0	-115.0	-102.9
ato_da_SAN	-45.0	-43.9	-74.9	-89.7
BH1_SOLEXA	NA	NA	-3.9	NA
bin_SANGER	-42.2	-43.4	-152.6	-43.1
Blimp-1_SO	-135.5	-29.7	-inf	-55.2
bowl_SOLEX	-1.3	NA	-13.0	-3.0
brk_FlyReg	-12.6	-1.5	-31.6	-3.1
br-Z1_FlyR	-26.0	-16.6	-76.5	-20.6
btd_NAR_FB	-60.6	-17.1	-121.9	-27.6
Cad_SOLEXA	-15.2	-4.4	-91.8	-11.9
cato_da_SA	-34.0	-45.0	-86.1	-85.0
Caup_SOLEX	NA	NA	-5.5	NA
CG10267_SO	-20.3	-6.6	-108.8	NA
CG13897_SA	-27.8	-27.9	-108.3	-33.1
CG31670_SO	-21.7	-18.2	-83.7	-7.4
CG33557_da	-56.0	-36.3	-132.0	-72.9
CG3838_SAN	NA	NA	-3.6	NA
CG4360-F1-	-108.7	-61.1	-inf	-49.7
CG5669_SOL	-28.1	-8.2	-54.2	-21.0
CG5953_SAN	-19.6	-16.3	-23.5	-15.3
CG6276_SAN	NA	NA	-2.7	NA
CG8765_SAN	-10.4	-9.6	-59.2	-5.0
ci_SOLEXA_	-81.5	-38.1	-174.5	-53.9
crol-F7-16	-92.3	-31.6	-232.7	-36.3
crp_SANGER	-32.4	-40.6	-92.8	-66.4
D_NAR_FBgn	-2.6	-7.3	-209.9	NA
da_SANGER_	-55.3	-39.4	-84.1	-81.4
Deaf1_FlyR	NA	-7.4	-5.5	-12.4
dei_da_SAN	-22.9	-21.6	-18.0	-36.0
dimm_da_SA	-13.8	-11.7	-9.2	-13.9

E5_SOLEXA_	NA	NA	-4.3	NA
eg_SANGER_	-11.6	-10.3	-19.3	-5.3
En_SOLEXA_	NA	NA	-12.6	NA
esg-F3-5_S	NA	NA	NA	-1.8
Ets97D_SAN	-10.6	NA	-2.7	NA
ey_SOLEXA_	-42.0	-8.8	-65.8	-10.5
Fer1_da_SA	-48.6	-49.2	-99.4	-102.0
Fer2_da_SA	-7.4	-16.6	-26.9	-38.3
Fer3_da_SA	-15.8	-29.4	-47.2	-66.9
fkh_NAR_FB	-29.6	-98.1	-104.2	-95.1
foxo_SANGE	-46.9	-32.9	-137.0	-31.3
fru_SOLEXA	NA	NA	-11.7	NA
grh_FlyReg	NA	NA	-7.3	NA
Hand_da_SA	-2.7	-5.7	-2.0	-12.7
hb_SOLEXA_	-21.7	-18.0	-56.4	-12.2
HLH4C_da_S	-75.3	-51.4	-158.5	-114.5
HLH54F_da_	-38.1	-23.9	-74.9	-53.2
jim_F1-9_S	-147.2	-82.1	-inf	-64.5
ken_SOLEXA	-2.1	NA	-9.3	NA
klu_SOLEXA	-211.1	-78.5	-inf	-107.5
kni_SANGER	-5.6	-8.3	-19.5	-5.4
l(1)sc_da_	-42.2	-18.6	-86.6	-58.4
Lab_SOLEXA	NA	NA	-5.6	NA
Lim1_SOLEX	NA	NA	-8.6	NA
lola-PC_SO	-25.0	NA	-41.6	-6.9
lola-PD_SO	-56.5	-25.1	-120.1	-30.4
lola-PF_SO	NA	NA	NA	-2.5
lola-PL_SO	-40.9	-17.1	-105.0	-33.8
lola-PO_SO	NA	-4.8	-4.8	-5.3
lola-PQ_SO	-1.7	-9.8	-81.5	NA
luna_SOLEX	-32.9	-4.2	-46.5	-18.3
Mad_FlyReg	-19.9	-14.6	-86.8	-18.8
Max_Mnt_SA	NA	NA	NA	-2.9
Med_FlyReg	-15.7	-8.1	-69.7	-17.3
Met_Clk_SA	NA	NA	NA	-3.6
Mirr_SOLEX	-2.3	-5.9	-56.8	NA
nau_da_SAN	-20.1	-19.3	-29.3	-58.2
net_da_SAN	-68.8	-63.3	-139.6	-124.6
NFAT_SANGE	-8.7	-1.4	-30.6	-2.7
nub_SOLEXA	-10.5	-7.0	-19.8	-16.4
odd_NBT_5_	-4.3	NA	-20.9	NA

|				
Oli_da_SAN	NA	NA	NA	-5.3
opa_SOLEXA	-52.3	-13.5	-123.8	-26.1
Otp_SOLEXA	NA	NA	-1.7	NA
pad_SOLEXA	-21.0	-4.2	-56.0	-7.5
pdm2_SOLEX	-1.8	-2.4	-3.7	-8.6
pnr_SANGER	-2.5	-5.5	-28.0	-2.2
pnt_SANGER	-1.4	NA	NA	NA
Rel_SANGER	-10.4	-1.5	-18.0	-6.5
Repo_SOLEX	NA	NA	-12.1	NA
run_Bgb_NB	NA	NA	-9.6	NA
sage_da_SA	-56.0	-60.6	-116.4	-110.0
sc_da_SANG	-71.9	-53.6	-134.7	-118.5
Scr_SOLEXA	NA	NA	-1.8	NA
slp1_NAR_F	-134.2	-112.9	-inf	-114.1
slp2_SANGE	-53.3	-67.6	-207.3	-57.8
sob_SOLEXA	NA	-2.2	-6.6	-2.2
Sox14_SANG	NA	NA	-40.9	NA
Sox15_SANG	-3.8	-10.3	-168.7	NA
Sp1_SOLEXA	-55.9	-11.3	-92.7	-32.3
sqz_SOLEXA	-43.0	-30.1	-103.7	-23.5
sr_SOLEXA_	-98.3	-40.4	-221.0	-55.8
sug_SOLEXA	-35.4	-10.0	-72.8	-18.0
suHw_FlyRe	-8.7	NA	-4.1	-1.5
tai_Clk_SA	NA	NA	NA	-3.2
tap_da_SAN	-23.7	-24.8	-40.5	-49.2
toy_FlyReg	-18.2	-2.1	-38.0	-1.8
Trl_FlyReg	-147.2	-67.5	-inf	-53.7
ttk-PA_SOL	-11.5	NA	-12.3	NA
twi_da_SAN	NA	-5.8	NA	-19.1
Ubx_SOLEXA	-3.7	NA	-40.0	-2.8
vvl_SOLEXA	-1.3	NA	-7.1	NA
wor_SOLEXA	NA	NA	NA	-7.6
z_FlyReg_F	-19.7	-6.8	-30.7	-15.3

6.2 Supplementary figures
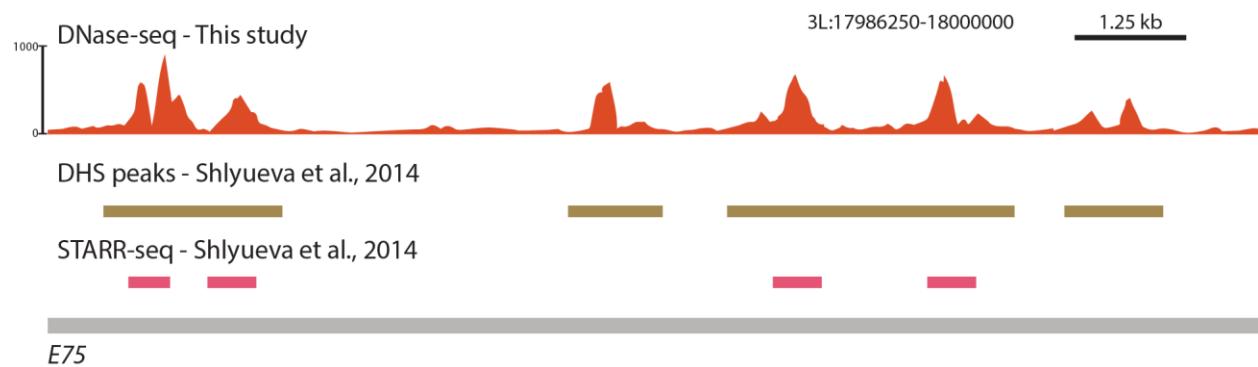
Figure S1: DNase-seq data obtained in this study show greater resolution compared to previous approaches. Genome browser screenshot of a DNase-seq coverage track over the *E75* locus. Track was compared to DHS peaks (brown boxes) and STARR-seq peaks (pink boxes) detected in Shlyueva et al., 2014.
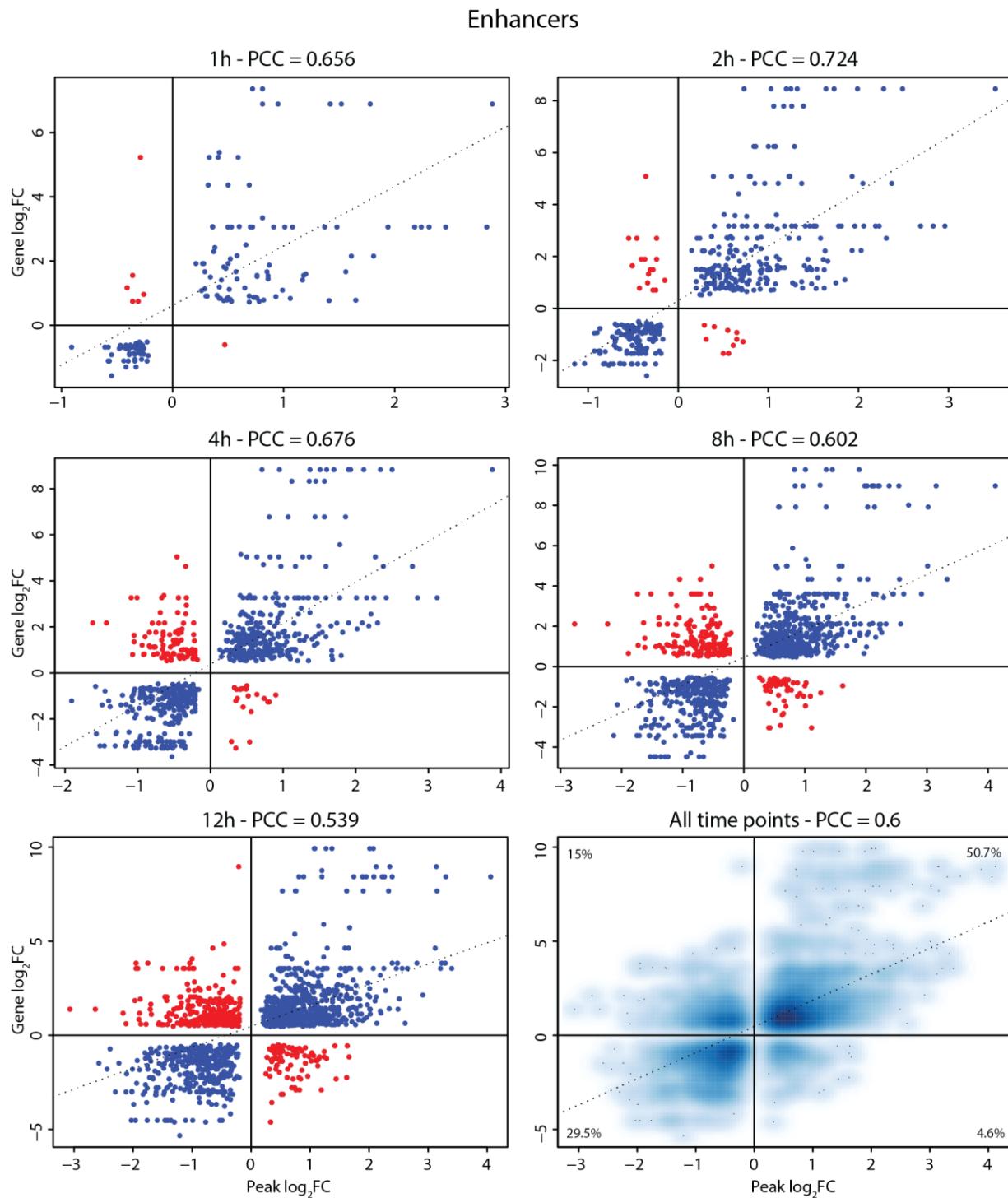
Figure S2: Quantitative correlation between enhancers and target genes FC. Log$_2$FC of enhancers (x-axis) and their associated target genes (y-axis) are illustrated. Each dot represents an enhancer-target gene association. Scatter plots for individual time points and all time points (merged time points) are displayed. Dashed lines represent linear regression. Blue dots indicate opening enhancers associated to up-regulated genes (positive values), or closing enhancers associated to down-regulated genes (negative values). Red dots represent opening enhancers associated to down-regulated genes, or closing enhancers associated to up-regulated genes. PCC values of correlation between enhancer and gene log$_2$FC are illustrated.
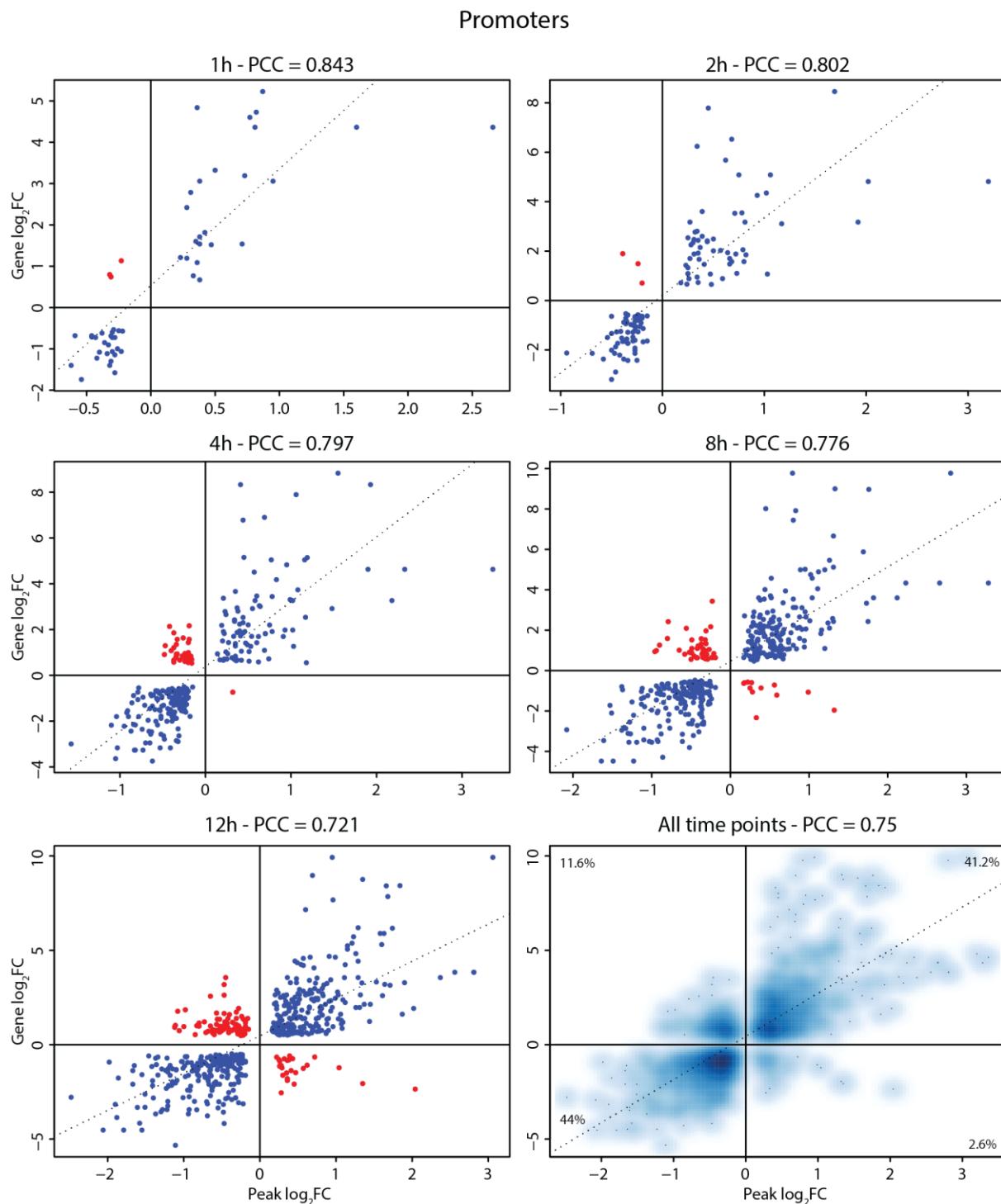
Promoters

Figure S3: Quantitative correlation between promoters and target genes FC. Log$_2$FC of promoters (x-axis) and their associated target genes (y-axis) are illustrated. Each dot represents a promoter-target gene association. Scatter plots for individual time points and all time points (merged time points) are displayed. Dashed lines represent linear regression. Blue dots indicate opening promoters associated to up-regulated genes (positive values), or closing promoters associated to down-regulated genes (negative values). Red dots represent opening promoters associated to down-regulated genes, or closing promoters associated to up-regulated genes. PCC values of correlation between promoter and gene log$_2$FC are illustrated.
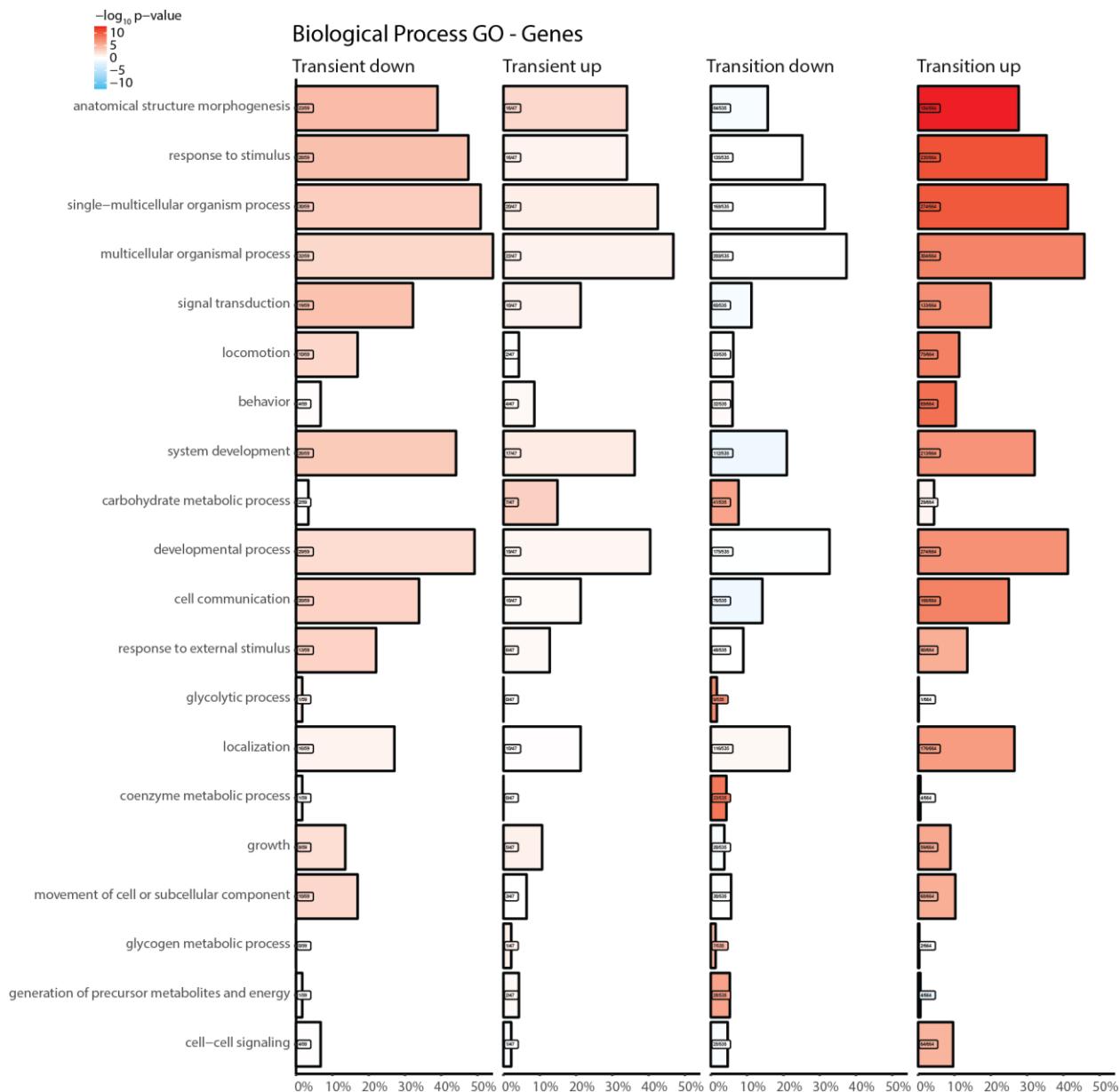
Figure S4: GO analysis based on differential genes shows enrichment for terms indicating morphological changes and response to ecdysone in 3 categories. Genes were examined based on their ImpulseDE2 categories. GO analysis was carried out with a hypergeometric test. Biological processes terms are illustrated. Box sizes indicate the percentage of genes included in each term compared to the total number of genes in the corresponding category (percentage scale illustrated at the bottom). The sum of percentages in each category is >100% as many genes are included in multiple terms. P-values (-$\log_{10}$) are color-coded and represent a measure of enrichment. Red values represent high enrichment. Blue values represent depletion.
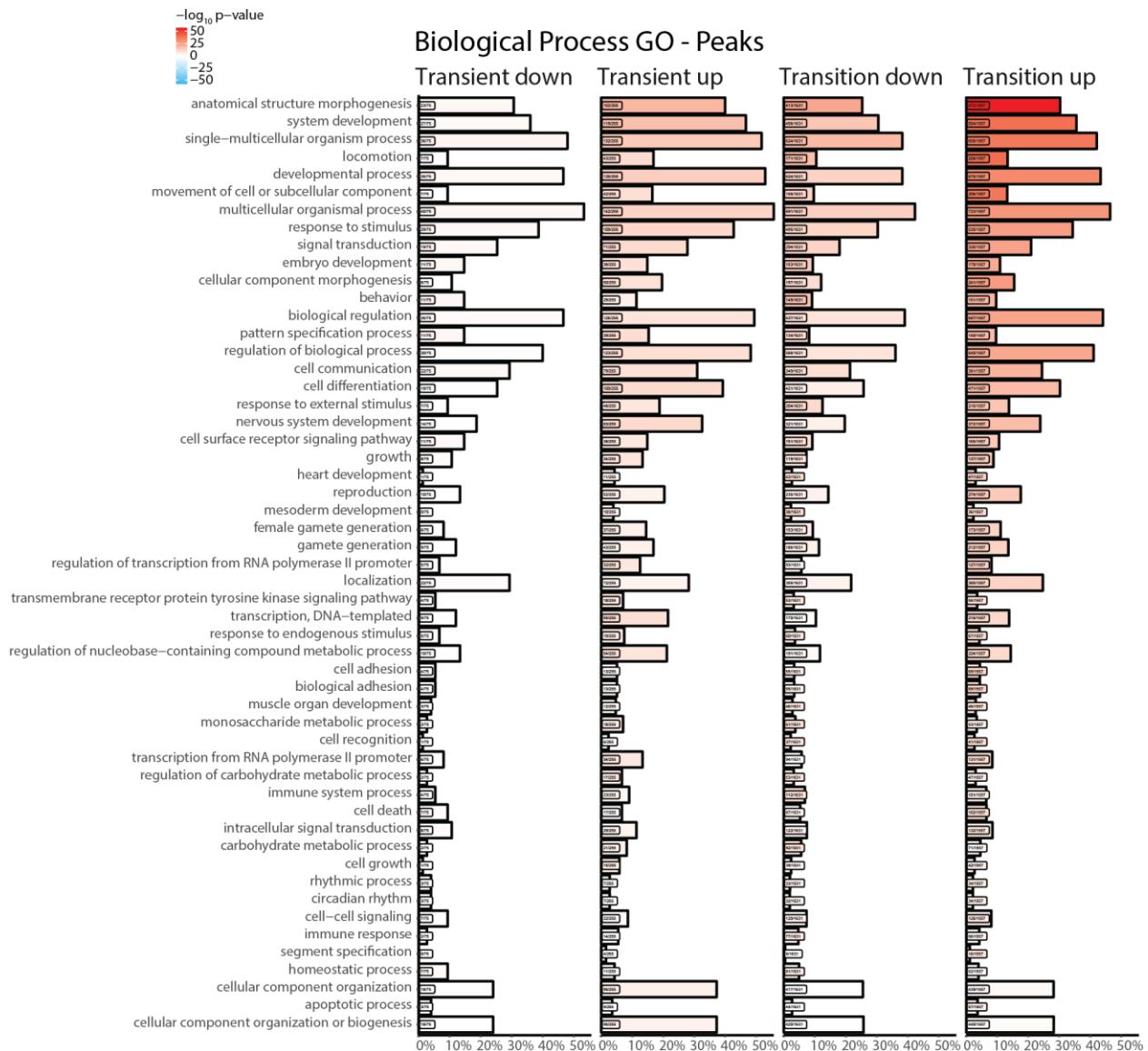
Figure S5: GO analysis based on differential DHSs shows enrichment for terms indicating morphological changes and response to ecdysone, and lack of depleted terms. DHSs were examined based on their ImpulseDE2 categories and associated to target genes. GO analysis was carried out with a hypergeometric test. Biological processes terms are illustrated. Box sizes indicate the percentage of DHSs included in each term compared to the total number of DHSs in the corresponding category (percentage scale illustrated at the bottom). The sum of percentages in each category is >100% as many DHSs are included in multiple terms. P-values (-$\log_{10}$) are color-coded and represent a measure of enrichment. Red values represent enrichment. Blue values represent depletion.
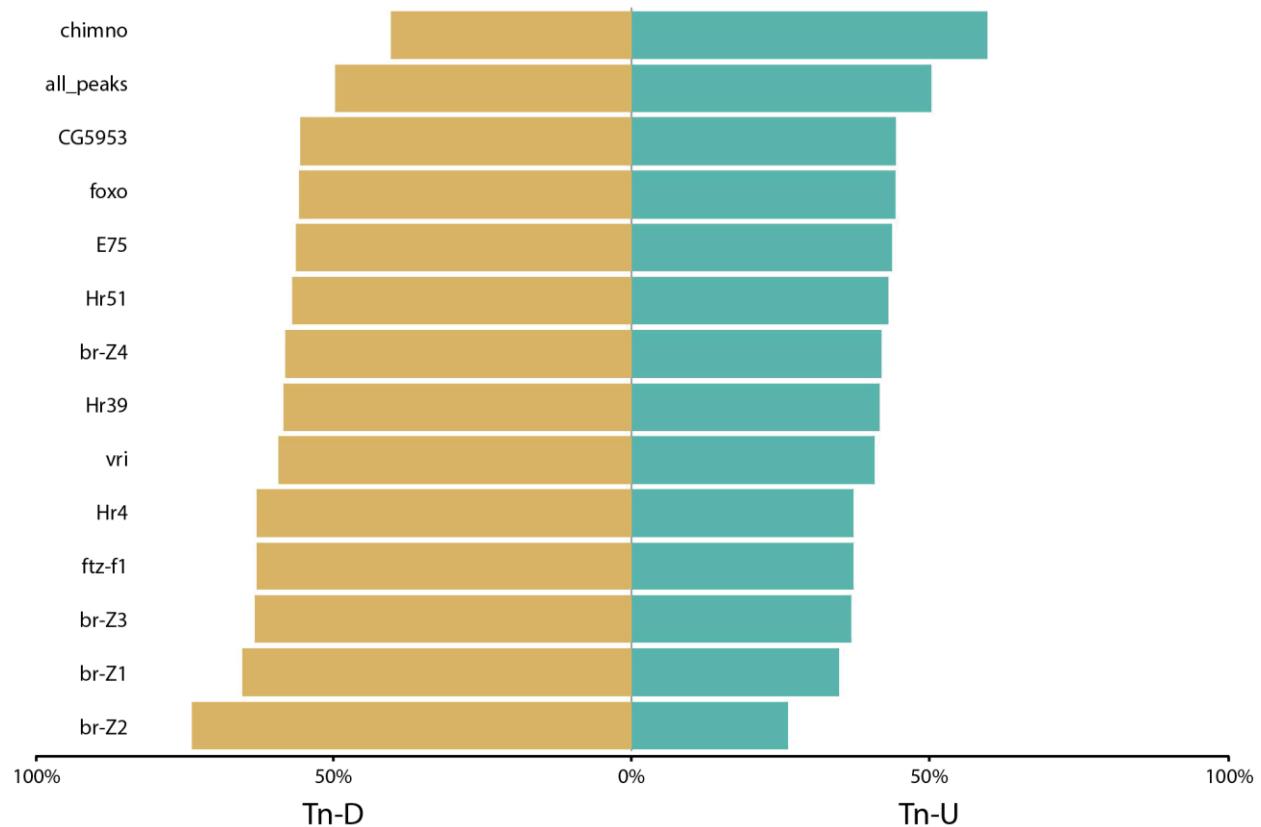
Figure S6: Distribution of TFBSs in Tn-D and Tn-U DHSs reveals preference of br isoforms for closing DHSs.
TFBSs identified either in Tn-D or in Tn-U DHSs were examined. The relative TFBS distributions in those two categories are displayed. TFBSs in Tn-D are in yellow. TFBSs in Tn-U are in green. The relative percentage of all DHSs is shown as reference (second term from the top).

A

Typical digestion

Chromatin

Mono-nucleosomes (>140 bp and <160 bp)
Sub-nucleosomes (>75 bp and <140 bp)

B

Peak correlation - PCC = 0.88

Coverage correlation - PCC = 0.98
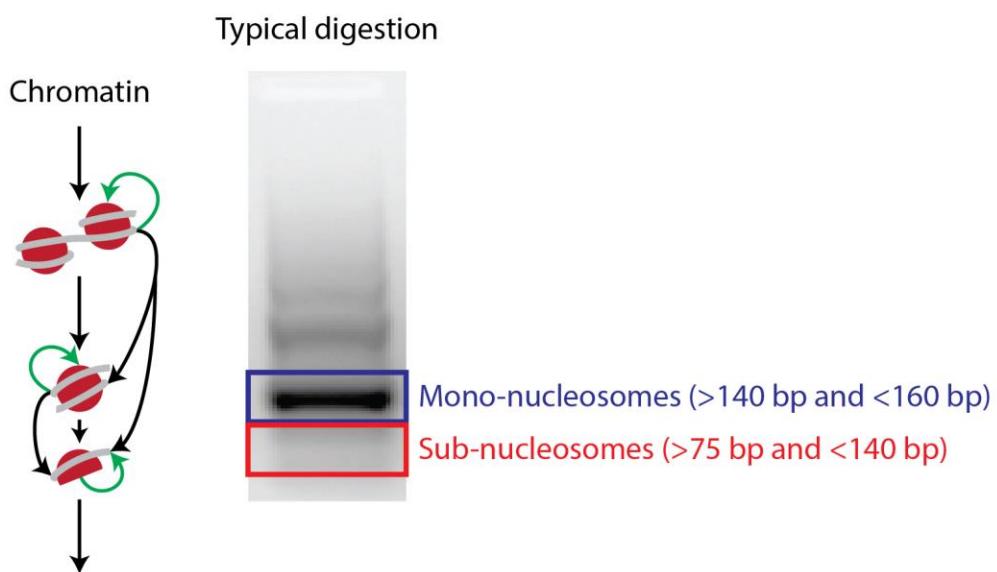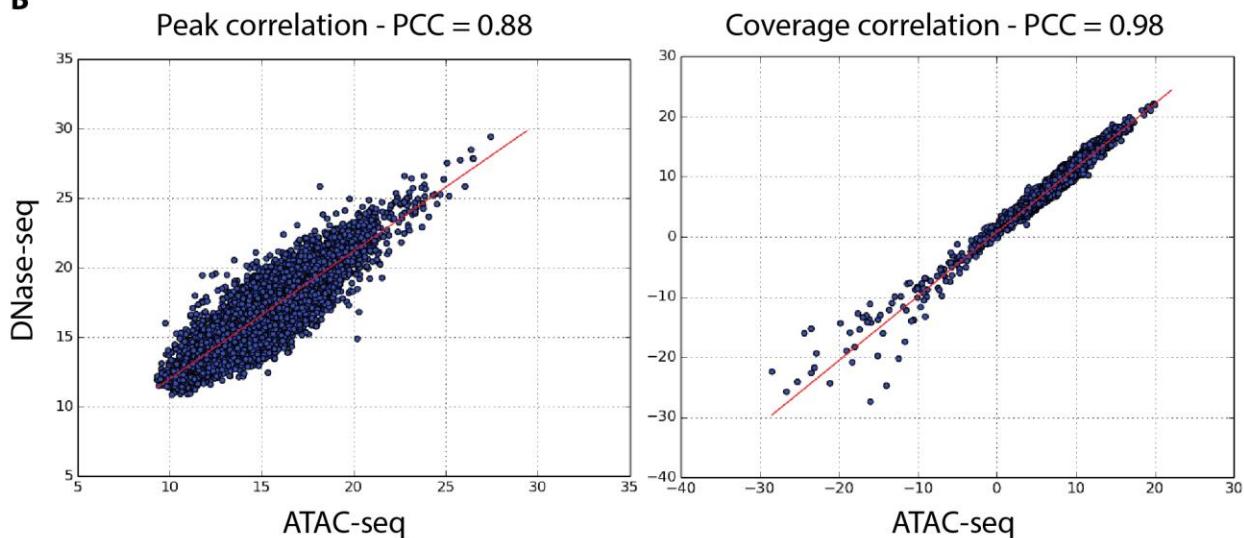
Figure S7: (A) Illustration of MNase progressive digestion on di-, mono- and sub-nucleosomes. In a typical digestion, MNase-sensitive nucleosomes undergo a transition to sub-nucleosomes (>75 bp and <140 bp) earlier than MNase-resistant nucleosomes, which are visible as mono-nucleosomal band (>140 bp and <160 bp). (B) DNase-seq (y-axis) and ATAC-seq (x-axis) correlate quantitatively. Left panel: correlation at peak level. Average coverage within individual peaks is illustrated. Right panel: correlation at coverage level genome-wide. Average coverage of 10 kb bins is illustrated.
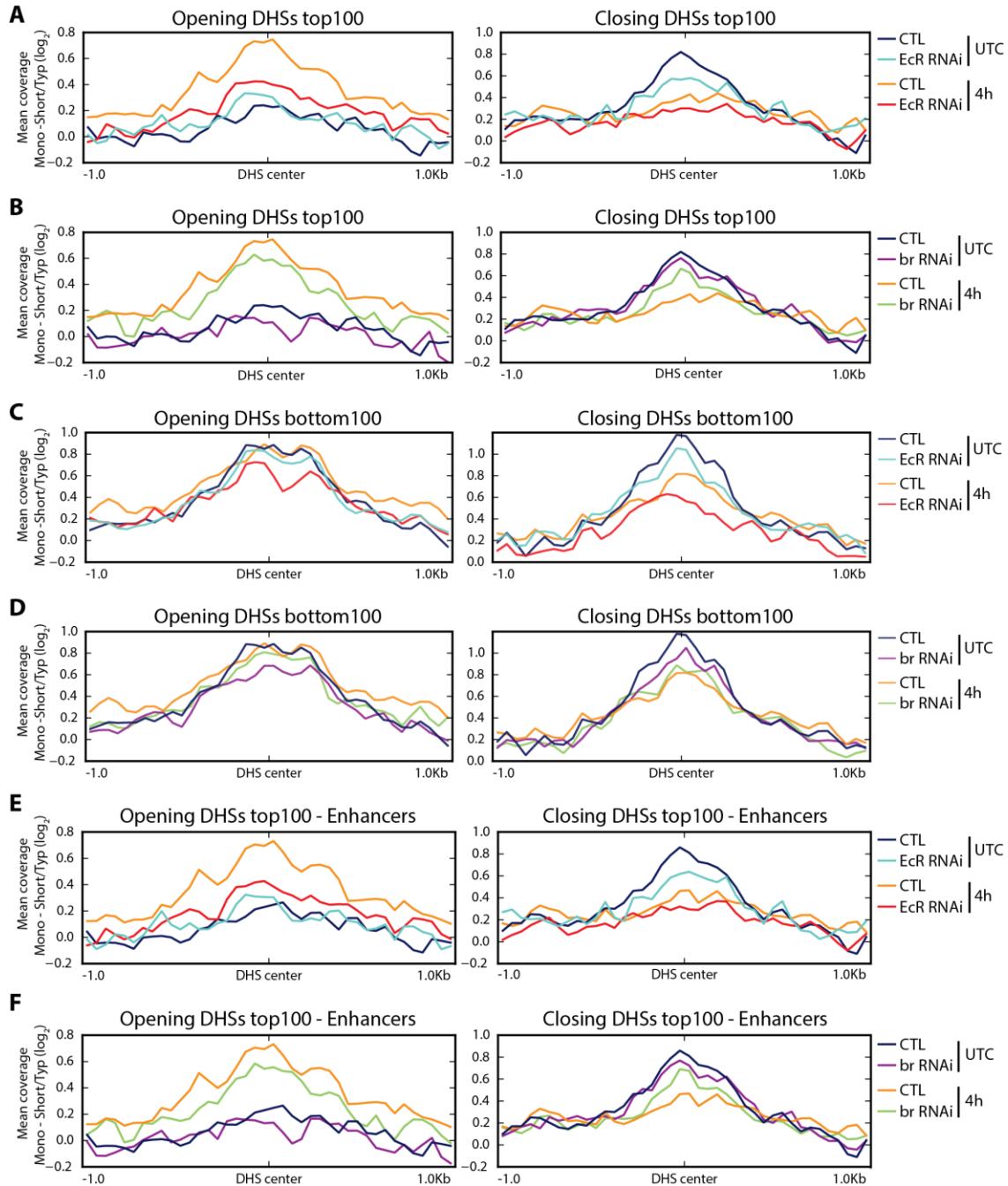
Figure S8: Ratios between Mono short and Mono typ mean dyad coverage mirror results obtained in typ Sub/Mono ratios. (A-B) Log_2 ratio between Mono short and Mono typ mean dyad coverage over top100 opening DHSs (left panels) and top100 closing DHSs (right panels). Ratios in CTL and EcR RNAi (A), and CTL and br RNAi (B) are shown. (C-D) Log_2 ratio between Mono short and Mono typ mean dyad coverage over bottom100 opening DHSs (left panels) and bottom100 closing DHSs (right panels). Ratios in CTL and EcR RNAi (C), and CTL and br RNAi (D) are shown. (E-F) Log_2 ratio between Mono short and Mono typ mean dyad coverage over top100 opening enhancers (left panels) and top100 closing enhancers (right panels). Ratios in CTL and EcR RNAi (E), and CTL and br RNAi (F) are shown. For CTL, dark blue and yellow lines indicate UTC and 4h, respectively. For EcR RNAi, light blue and red lines indicate UTC and 4h, respectively. For br RNAi, violet and green lines indicate UTC and 4h, respectively. Regions ±1 kb away from the DHS center were considered.
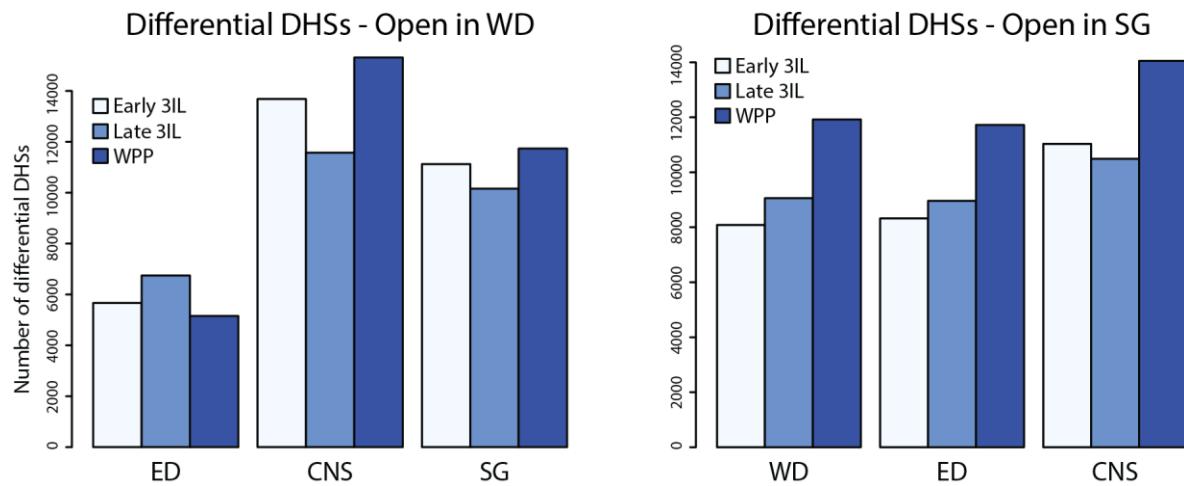
Figure S9: Comparison of differential DHSs among tissues indicates similar cell fates. Open DHSs in WD (left panel) or SG (right panel) were compared to the other tissues along time. DHSs that were open in the reference tissue but were not in the comparison tissue, were called as differential.
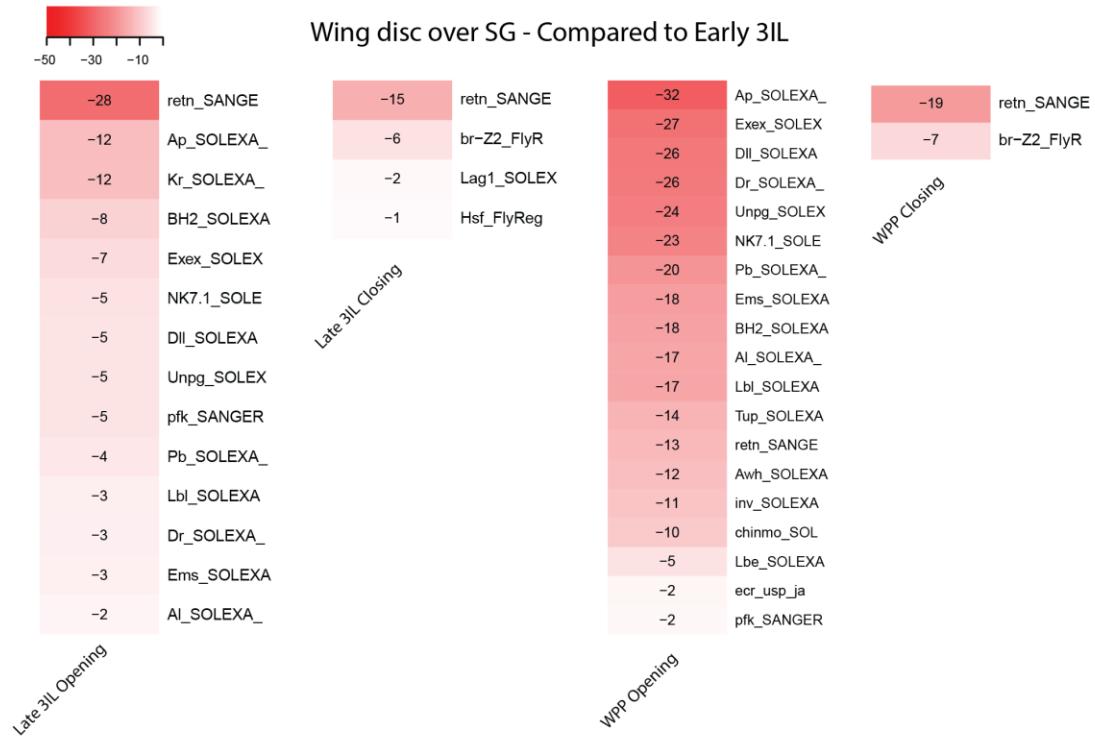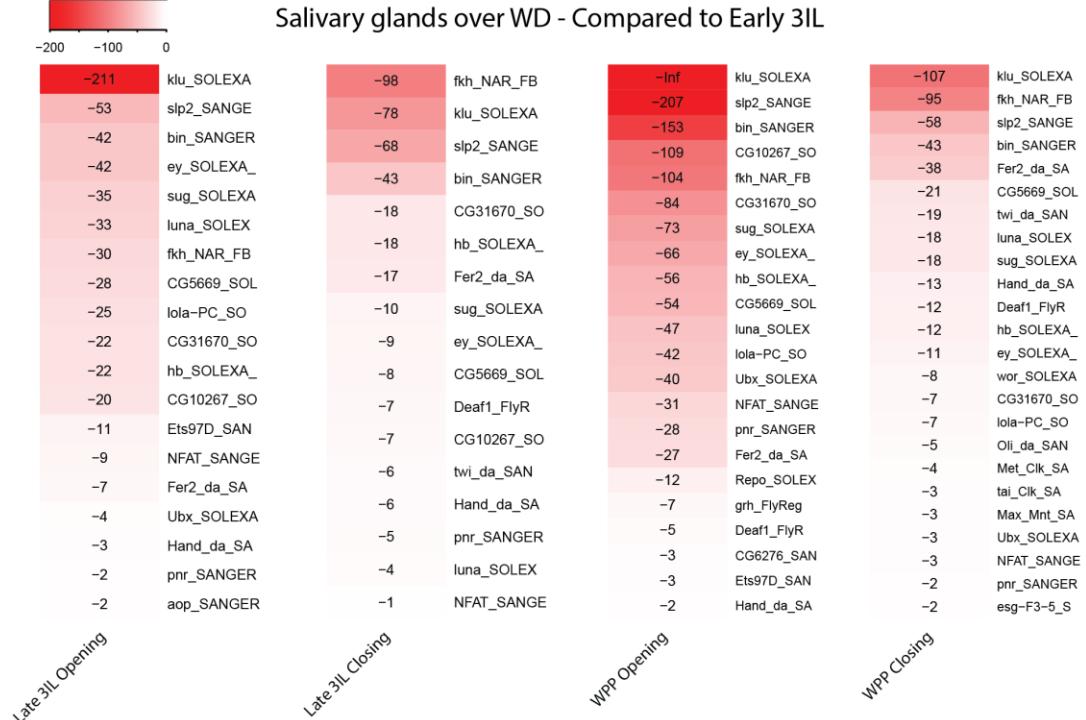
Figure S10: WD and SG motif enrichment in opening and closing DHSs over time. (A) WD TFs which did not show enrichment in SG set. (B) SG TFs which did not show enrichment in WD set. Differential DHSs were detected in comparison to Early 3IL in individual tissues. Heatmaps show TF adjusted p-values ($\log_{10}$) as measure of motif enrichment in each DHS set. Corresponding values are displayed.

6.3 Abbreviations

3rd IL	third instar larva
4uS	4-thiouridine
APF	after puparium formation
ATAC-seq	assay for transposase-accessible chromatin using sequencing
B1H	bacterial one-hybrid
ChIP-seq	chromatin immunoprecipitation followed by sequencing
CNS	central nervous system
CRE	*cis*-regulatory element
CTL	no RNAi control
DBD	DNA-binding domain
DHS	DNase I hypersensitive site
DPE	downstream promoter elements
DTA	dynamic transcriptome analysis
Early 3Il	early third instar larva
EcRE	ecdysone-responsive element
ED	eye disc
FAIRE-seq	formaldehyde-assisted isolation of regulatory elements using sequencing
FC	fold change
gDNA	genomic DNA
GO	gene ontology
GTF	general transcription factor
HMM	hidden Markov model
HT-SELEX	high-throughput SELEX
Late 3IL	late third instar larva
LBD	ligand-binding domain
modENCODE	model organism encyclopedia of DNA elements
NDR	nucleosome-depleted region
NGS	next-generation sequencing
p-adj	adjusted p-value

PBM	protein-binding microarrays
PCC	Pearson correlation coefficient
PCD	programmed cell death
Pol II	RNA polymerase II
PWM	position weight matrix
S2	Schneider 2
SG	salivary gland
SNR	signal-to-noise ratio
STARR-seq	self-transcribing active regulatory region sequencing
TF	transcription factor
TFBE	transcription factor binding event
TFBS	transcription factor binding site
Tn-D	transition down
Tn-U	transition up
TSS	transcriptional start site
Tt-D	transient down
Tt-U	transient up
UTC	untreated control
WD	wing disc
WPP	white prepupa

7 REFERENCES

Ables, E.T., and Drummond-Barbosa, D. (2010). The steroid hormone ecdysone functions with intrinsic chromatin remodeling factors to control female germline stem cells in *Drosophila*. Cell Stem Cell *7*, 581–592.

Allan, J., Fraser, R.M., Owen-Hughes, T., and Keszenman-Pereyra, D. (2012). Micrococcal nuclease does not substantially bias nucleosome mapping. J. Mol. Biol. *417*, 152–164.

Arbeitman, M.N., and Hogness, D.S. (2000). Molecular chaperones activate the *Drosophila* ecdysone receptor, an RXR heterodimer. Cell *101*, 67–77.

Arbeitman, M.N., Furlong, E.E.M., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., and White, K.P. (2002). Gene expression during the life cycle of *Drosophila melanogaster*. Science *297*, 2270–2275.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., Stark, A., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science (80-.). *339*, 1074–1077.

Arnold, C.D., Gerlach, D., Spies, D., Matts, J. a., Sytnikova, Y. a., Pagani, M., Lau, N.C., and Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. Nat. Genet. *46*, 685–692.

Arnold, C.D., Zabidi, M.A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T., and Stark, A. (2016). Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. Nat. Biotechnol.

Arnosti, D.N., Barolo, S., Levine, M., and Small, S. (1996). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development *122*, 205–214.

Ashburner, M. (1973). Sequential gene activation by ecdysone in polytene chromosomes of *Drosophila melanogaster*. I. Dependence upon ecdysone concentration. Dev. Biol. *35*, 47–61.

Badenhorst, P., Xiao, H., Cherbas, L., Kwon, S.Y., Voas, M., Rebay, I., Cherbas, P., and Wu, C. (2005). The *Drosophila* nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. Genes Dev. *19*, 2540–2545.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720–1723.

Baehrecke, E.H. (1996). Ecdysone signaling cascade and regulation of *Drosophila* metamorphosis. Arch. Insect Biochem. Physiol. *33*, 231–244.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: Tools for motif discovery and searching. Nucleic Acids Res. *37*.

References

Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014). Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers. Mol. Cell *54*, 844–857.

Beckendorf, S.K., and Kafatos, F.C. (1976). Differentiation in the salivary glands of *Drosophila melanogaster*: Characterizaion of the glue proteins and their developmental appearance. Cell *9*, 365–373.

Beckstead, R.B., Lam, G., and Thummel, C.S. (2005). The genomic response to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. Genome Biol. *6*, R99–R99.

Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.H., Trump, S., Lightman, S.L., et al. (2011). Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. Mol. Cell *43*, 145–155.

Blatti, C., Kazemian, M., Wolfe, S., Brodsky, M., and Sinha, S. (2015). Integrating motif, DNA accessibility and gene expression data to build regulatory maps in an organism. Nucleic Acids Res. *43*, 3998–4012.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. *44*, 148–156.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell *132*, 311–322.

Brodu, V., Mugat, B., Roignant, J.Y., Lepesant, J.A., and Antoniewski, C. (1999). Dual requirement for the EcR/USP nuclear receptor and the dGATAb factor in an ecdysone response in *Drosophila melanogaster*. Mol Cell Biol *19*, 5732–5742.

Brodu, V., Mugat, B., Fichelson, P., Lepesant, J. a, and Antoniewski, C. (2001). A UAS site substitution approach to the in vivo dissection of promoters: interplay between the GATAb activator and the AEF-1 repressor at a *Drosophila* ecdysone response unit. Development *128*, 2593–2602.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Burtis, K.C., Thummel, C.S., Jones, C.W., Karim, F.D., and Hogness, D.S. (1990). The *Drosophila* 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. Cell *61*, 85–99.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33–43.

References

Chao, A.T., and Guild, G.M. (1986). Molecular analysis of the ecdysterone-inducible 2B5 "early' puff in *Drosophila melanogaster*. EMBO J. *5*, 143–150.

Cherbas, L., Hu, X., Zhimulev, I., Belyaeva, E., and Cherbas, P. (2003). EcR isoforms in *Drosophila*: testing tissue-specific requirements by targeted blockade and rescue. Development *130*, 271–284.

Chereji, R.V., Kan, T.W., Grudniewska, M.K., Romashchenko, A. V., Berezikov, E., Zhimulev, I.F., Guryev, V., Morozov, A. V., and Moshkin, Y.M. (2015). Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in *Drosophila melanogaster*. Nucleic Acids Res. *44*, 1036–1051.

Chereji, R.V., Ocampo, J., and Clark, D.J. (2017). MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. Mol. Cell *65*, 565–577.e3.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Couderc, J.-L., Godt, D., Zollman, S., Chen, J., Li, M., Tiong, S., Cramton, S.E., Sahut-Barnola, I., and Laski, F. a (2002). The bric à brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in *Drosophila*. Development *129*, 2419–2433.

Cranna, N., Quinn, L., Fristrom, J., Fristrom, D., Church, R., Robertson, F., Zitnan, D., Kim, Y., Zitnanova, I., Roller, L., et al. (2009). Impact of steroid hormone signals on *Drosophila* cell cycle during development. Cell Div. *4*, 3.

Dogan, N., Wu, W., Morrissey, C.S., Chen, K.-B., Stonestrom, A., Long, M., Keller, C.A., Cheng, Y., Jain, D., Visel, A., et al. (2015). Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. Epigenetics Chromatin *8*, 16.

Emery, I.F., Bedian, V., and Guild, G.M. (1994). Differential expression of Broad-Complex transcription factors may forecast tissue-specific developmental fates during *Drosophila* metamorphosis. Development *120*, 3275–3287.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. *28*, 817–825.

Fischer, D.S., Theis, F.J., and Yosef, N. (2017). Impulse model-based differential expression analysis of time course sequencing data. bioRxiv 113548.

Frank, C.L., Liu, F., Wijayatunge, R., Song, L., Biegler, M.T., Yang, M.G., Vockley, C.M., Safi, A., Gersbach, C.A., Crawford, G.E., et al. (2015). Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum. Nat. Neurosci. *18*, 647–656.

Frith, M.C., Valen, E., Krogh, A., Hayashizaki, Y., Carninci, P., and Sandelin, A. (2008). A code for transcription initiation in mammalian genomes. Genome Res. *18*, 1–12.

References

Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. *5*, 3157–3170.

Gallo, S.M., Li, L., Hu, Z., and Halfon, M.S. (2006). REDfly: A Regulatory Element Database for *Drosophila*. Bioinformatics *22*, 381–383.

Gauhar, Z., Sun, L. V, Hua, S., Mason, C.E., Fuchs, F., Li, T., Boutros, M., and White, K.P. (2009). Genomic mapping of binding regions for the Ecdysone receptor protein complex Genomic mapping of binding regions for the Ecdysone receptor protein complex. 1006–1013.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: A platform for interactive large-scale genome analysis. Genome Res. *15*, 1451–1455.

Gomez, N.C., Hepperla, A.J., Dumitru, R., Simon, J.M., Fang, F., and Davis, I.J. (2016). Widespread Chromatin Accessibility at Repetitive Elements Links Stem Cells with Human Cancer. Cell Rep. *17*, 1607–1620.

Handler, A.M. (1982). Ecdysteroid titers during pupal and adult development in *Drosophila melanogaster*. Dev. Biol. *93*, 73–82.

He, H.H., Meyer, C.A., Hu, S.S., Chen, M.W., Zang, C., Liu, Y., Rao, P.K., Fei, T., Xu, H., Long, H., et al. (2014). Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. Nat. Methods *11*, 73–78.

Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res. *19*, 460–469.

Hooper, S.D., Boué, S., Krause, R., Jensen, L.J., Mason, C.E., Ghanim, M., White, K.P., Furlong, E.E., and Bork, P. (2007). Identification of tightly regulated groups of genes during *Drosophila melanogaster* embryogenesis. Mol. Syst. Biol. *3*.

Huet, F., Ruiz, C., and Richards, G. (1995). Sequential gene activation by ecdysone in *Drosophila melanogaster*: the hierarchical equivalence of early and early late genes. Development *121*, 1195–1204.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat. Genet. *43*, 27–33.

Istrail, S., and Davidson, E.H. (2005). Logic functions of the genomic cis-regulatory code. Proc. Natl. Acad. Sci. U. S. A. *102*, 4954–4959.

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol. Cell *62*, 79–91.

Jiang, C., Lamblin, A.-F.F., Steller, H., and Thummel, C.S. (2000). A Steroid-triggered Transcriptional Hierarchy Controls Salivary Gland Cell Death during *Drosophila* Metamorphosis. Mol. Cell *5*, 445–455.

References

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.

Juven-Gershon, T., Hsu, J.Y., and Kadonaga, J.T. (2008). Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. Genes Dev. *22*, 2823–2830.

von Kalm, L., Crossgrove, K., Von Seggern, D., Guild, G.M., and Beckendorf, S.K. (1994). The Broad-Complex directly controls a tissue-specific response to the steroid hormone ecdysone at the onset of *Drosophila* metamorphosis. EMBO J. *13*, 3505–3516.

Kaplan, T., Li, X.-Y., Sabo, P.J., Thomas, S., Stamatoyannopoulos, J.A., Biggin, M.D., and Eisen, M.B. (2011). Quantitative Models of the Mechanisms That Control Genome-Wide Patterns of Transcription Factor Binding during Early *Drosophila* Development. PLoS Genet *7*, e1001290.

Karim, F.D., and Thummel, C.S. (1992). Temporal coordination of regulatory gene expression by the steroid hormone ecdysone. EMBO J. *11*, 4083–4093.

Karim, F.D., Guild, G.M., and Thummel, C.S. (1993). The *Drosophila* Broad-Complex plays a key role in controlling ecdysone-regulated gene expression at the onset of metamorphosis. Development *118*, 977–988.

Kharchenko, P. V, Alekseyenko, A. a, Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A. a, Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. Nature *471*, 480–485.

King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. Elife *6*.

Koelle, M.R., Talbot, W.S., Segraves, W.A., Bender, M.T., Cherbas, P., and Hogness, D.S. (1991). The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell *67*, 59–77.

Koohy, H., Down, T.A., and Hubbard, T.J. (2013). Chromatin Accessibility Data Sets Show Bias Due to Sequence Specificity of the DNase I Enzyme. PLoS One *8*, 1–9.

Koyama, T., Rodrigues, M.A., Athanasiadis, A., Shingleton, A.W., and Mirth, C.K. (2014). Nutritional control of body size through FoxO-Ultraspiracle mediated ecdysone biosynthesis. Elife *3*.

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

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes

in Yeast. Mol. Cell *60*, 422–434.

Kubik, S., Bruzzone, M.J., and Shore, D. (2017). Establishing nucleosome architecture and stability at promoters: Roles of pioneer transcription factors and the RSC chromatin remodeler. BioEssays *1600237*, 1600237.

Kulkarni, M.M., and Arnosti, D.N. (2003). Information display by transcriptional enhancers. Development *130*, 6569–6575.

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

Lee, C., Wendel, D.P., Reid, P., Lam, G., Thummel, C.S., Baehrecke, E.H., Park, C., Park, C., and Hughes, H. (2000). E93 Directs Steroid-Triggered Programmed Cell Death in *Drosophila*. *6*, 433–443.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.

Li, T.R., and White, K.P. (2003). Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. Dev. Cell *5*, 59–72.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L., et al. (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. PLoS Biol. *6*, 0365–0388.

Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K.H. (2012). Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell *151*, 1608–1616.

Lin, G.G.H., Kozaki, T., and Scott, J.G. (2011). Hormone receptor-like in 96 and broad-complex modulate phenobarbital induced transcription of cytochrome P450 cyp6d1 in *Drosophila* S2 cells. Insect Mol. Biol. *20*, 87–95.

Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A. V., Moresco, J.J., Blokhina, T., Wolle, D., Aoki, T., Ryabykh, V., Yates, J.R., et al. (2017). The GAGA factor regulatory network: Identification of GAGA factor associated proteins. PLoS One *12*.

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.

Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., and Segal, E. (2015). Core promoter sequence in yeast is a major determinant of expression level. Genome Res. *25*, 1008–1017.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J.,

References

Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the *Drosophila* genome. Nature *453*, 358–362.

McKay, D.J., and Lieb, J.D. (2013). A Common Set of DNA Regulatory Elements Shapes *Drosophila* Appendages. Dev. Cell *27*, 306–318.

McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics *11*, 165.

Merika, M., and Thanos, D. (2001). Enhanceosomes. Curr. Opin. Genet. Dev. *11*, 205–208.

Mieczkowski, J., Cook, A., Bowman, S.K., Mueller, B., Alver, B.H., Kundu, S., Deaton, A.M., Urban, J.A., Larschan, E., Park, P.J., et al. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat. Commun. *7*, 11485.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dumcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dolken, L., et al. (2014). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol. Syst. Biol. *7*, 458–458.

Moshkin, Y.M., Chalkley, G.E., Kan, T.W., Reddy, B. a., Ozgur, Z., van Ijcken, W.F.J., Dekkers, D.H.W., Demmers, J. a., Travers, a. a., and Verrijzer, C.P. (2012). Remodelers Organize Cellular Chromatin by Counteracting Intrinsic Histone-DNA Sequence Preferences in a Class-Specific Manner. Mol. Cell. Biol. *32*, 675–688.

Mugat, B., Brodu, V., Kejzlarova-Lepesant, J., Antoniewski, C., Bayer, C.A., Fristrom, J.W., and Lepesant, J.A. (2000). Dynamic expression of broad-complex isoforms mediates temporal control of an ecdysteroid target gene at the onset of *Drosophila* metamorphosis. Dev. Biol. *227*, 104–117.

Negre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., et al. (2011). A cis-regulatory map of the *Drosophila* genome. Nature *471*, 527–531.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Noyes, M.B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M.H., and Wolfe, S.A. (2008). A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. Nucleic Acids Res. *36*, 2547–2560.

Ou, Q., and King-Jones, K. (2013). What Goes Up Must Come Down. Transcription Factors Have Their Say in Making Ecdysone Pulses (Elsevier Inc.).

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. Cell *129*, 1111–1123.

Papatsenko, D., Goltsev, Y., and Levine, M. (2009). Organization of developmental enhancers in the *Drosophila* embryo. Nucleic Acids Res. *37*, 5665–5677.

Pfreundt, U., James, D.P., Tweedie, S., Wilson, D., Teichmann, S.A., and Adryan, B. (2009).

FlyTF: Improved annotation and enhanced functionality of the *Drosophila* transcription factor database. Nucleic Acids Res. *38*.

Pique-regi, R., Degner, J.F., Pai, A. a, Boyle, A.P., Song, L., Lee, B., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). sequence and chromatin accessibility data Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. *21*, 447–455.

Qu, K., Zaba, L.C., Satpathy, A.T., Giresi, P.G., Li, R., Jin, Y., Armstrong, R., Jin, C., Schmitt, N., Rahbar, Z., et al. (2017). Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors. Cancer Cell 1–15.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). DeepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*.

Robinow, S., Talbot, W.S., Hogness, D.S., and Truman, J.W. (1993). Programmed cell death in the *Drosophila* CNS is ecdysone-regulated and coupled with a specific ecdysone receptor isoform. Development *119*, 1251–1259.

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem. Sci. *21*, 327–335.

Rogers, S.L., and Rogers, G.C. (2008). Culture of *Drosophila* S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. Nat. Protoc. *3*, 606–611.

Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, D91-4.

Schmidt, F., Gasparoni, N., Gasparoni, G., Gianmoena, K., Cadenas, C., Polansky, J.K., Ebert, P., Nordstrom, K., Barann, M., Sinha, A., et al. (2017). Combining transcription factor binding affinities with open-chromatin data for accurate gene expression prediction. Nucleic Acids Res. *45*, 54–66.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. Nature *451*, 535–540.

Segraves, W.A., and Hogness, D.S. (1990). The E75 ecdysone-inducible gene responsible for the 75B early puff in *Drosophila* encodes two new members of the steroid receptor superfamily. Genes Dev. *4*, 204–219.

Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A. V., and Zaret, K.S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. *23*, 804–809.

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.

Sinha, S. (2006). On counting position weight matrix matches in a sequence, with application to discriminative motif finding. In Bioinformatics, p.

Slattery, M., Ma, L., Spokony, R.F., Arthur, R.K., Kheradpour, P., Kundaje, A., Nègre, N., Crofts, A., Ptashkin, R., Zieba, J., et al. (2014). Diverse patterns of genomic targeting by transcriptional regulators in *Drosophila melanogaster*. Genome Res. *24*, 1224–1235.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annu. Rev. Biochem. *72*, 449–479.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., Gräf, S., Huss, M., Keefe, D., et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. *21*, 1757–1767.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature *528*, 147–151.

Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. Science *254*, 1385–1387.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N., et al. (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. Nature *450*, 219–232.

Stavreva, D.A., Coulon, A., Baek, S., Sung, M.H., John, S., Stixova, L., Tesikova, M., Hakim, O., Miranda, T., Hawkins, M., et al. (2015). Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing. Genome Res. *25*, 845–857.

Stormo, G.D., Schneider, T.D., Gold, L., and Ehrenfeucht, A. (1982). Use of the "perceptron" algorithm to distinguish translational initiation sites in E. coli. Nucleic Acids Res. *10*, 2997–3011.

Sung, M.H., Guertin, M.J., Baek, S., and Hager, G.L. (2014). DNase footprint signatures are dictated by factor dynamics and DNA sequence. Mol. Cell *56*, 275–285.

Talbot, W.S., Swyryd, E.A., and Hogness, D.S. (1993). *Drosophila* tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. Cell *73*, 1323–1337.

Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res. *16*, 962–972.

Tautz, D. (2000). Evolution of transcriptional regulation. Curr. Opin. Genet. Dev. *10*, 575–579.

The modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V, Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C. a, Ma, L., et al. (2011). Identification of Functional

Elements and Regulatory Circuits by *Drosophila* modENCODE. October *330*, 1787–1797.

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.

Thummel, C.S., Burtis, K.C., and Hogness, D.S. (1990). Spatial and temporal patterns of E74 transcription during *Drosophila* development. Cell *61*, 101–111.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

Timmerman, C., Suppiah, S., Gurudatta, B. V., Yang, J., Banerjee, C., Sandstrom, D.J., Corces, V.G., and Sanyal, S. (2013). The *Drosophila* Transcription Factor Adf-1 (nalyot) Regulates Dendrite Growth by Controlling FasII and Staufen Expression Downstream of CaMKII and Neural Activity. J. Neurosci. *33*, 11916–11931.

Truman, J.W., Talbot, W.S., Fahrbach, S.E., and Hogness, D.S. (1994). Ecdysone receptor expression in the CNS correlates with stage-specific responses to ecdysteroids during *Drosophila* and Manduca development. Development *120*, 219–234.

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.

Vierstra, J., Wang, H., John, S., Sandstrom, R., and Stamatoyannopoulos, J. a (2014). Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. Nat. Methods *11*, 66–72.

Weber, C., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-Modulated barriers to RNA polymerase. Mol. Cell *53*, 819–830.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. *20*, 90–100.

West, J.A., Cook, A., Alver, B.H., Stadtfeld, M., Deaton, A.M., Hochedlinger, K., Park, P.J., Tolstorukov, M.Y., and Kingston, R.E. (2014). Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. Nat. Commun. *5*, 4719.

Wu, C. (1980). The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854–860.

Xi, Y., Yao, J., Chen, R., Li, W., and He, X. (2011). Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. Genome Res. *21*, 718–724.

References

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., and Wu, C. (2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. Mol. Cell *8*, 531–543.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.

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

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*, 111–117.